

25



UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

FACULTAD DE ESTUDIOS SUPERIORES  
CUAUTITLAN

ELABORACION DE UN MANUAL PRACTICO DE  
PROBABILIDAD Y ESTADISTICA PARA LA  
CARRERA DE INGENIERIA EN ALIMENTOS.  
(REVISION BIBLIOGRAFICA)

**T E S I S**

QUE PARA OBTENER EL TITULO DE  
INGENIERO EN ALIMENTOS  
P R E S E N T A N :  
BALMORI CONSUEGRA LUCIA DEL C.  
VAZQUEZ GARDUÑO OSCAR

ASESOR DE TESIS: I.A. ROSALIA MELENDEZ PEREZ.

CUAUTITLAN IZCALLI, EDO. DE MEXICO.

1999

**TESIS CON  
FALLA DE ORIGEN**

270139



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

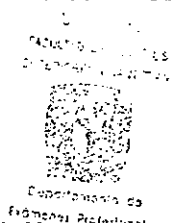
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN  
 UNIDAD DE LA ADMINISTRACION ESCOLAR  
 DEPARTAMENTO DE EXAMENES PROFESIONALES



UNIVERSIDAD NACIONAL  
 AUTÓNOMA DE  
 MÉXICO

ASUNTO: VOTOS APROBATORIOS

DR. JUAN ANTONIO MONTARAZ CRESPO  
 DIRECTOR DE LA F.E.S.-CUAUTITLAN  
 P R E S E N T E



ATN: Ing. Jaime de Anda Montañez  
 Jefe del Departamento de Exámenes  
 Profesionales de la F.E.S.-C

Con base en el art. 28 del Reglamento General de Exámenes, nos permitimos comunicar a usted que revisamos la TESIS:

Elaboración de un manual practico de Probabilidad y Estadística  
para la carrera de Ingeniería en Alimentos. (Revisión bibliográfica)

que presenta la pasante: Palmeri Consuegra Lucia del Carmen  
 con número de cuenta: 8737837- para obtener el TITULO de:  
Ingeniera en Alimentos.

Considerando que dicho tesis reúne los requisitos necesarios para ser discutida en el EXAMEN PROFESIONAL correspondiente, otorgamos nuestro VOTO APROBATORIO.

A T E N T A M E N T E  
 "POR MI RAZA HABLARA EL ESPIRITU"

Cuautitlan Izcalli, Edo. de Mexico, a 9 de Julio de 199 8

PRESIDENTE	<u>R.O. J. JALBA</u>	<u>eres Minutti</u>	
VOCA.	<u>Jose</u>	<u>roostro Rodriguez</u>	
SECUN: TARIO	<u>L.A. Rosalia</u>	<u>Manuel Perez</u>	
RIN: S SUPLENTE	<u>L.O. CUCUA</u>	<u>de rosa Villarreal</u>	
SEC. SÍG SUPLENTE	<u>de</u>	<u>arduno</u>	<u>H. Cas</u>

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN  
UNIDAD DE LA ADMINISTRACIÓN ESCOLAR  
DEPARTAMENTO DE EXÁMENES PROFESIONALES



ASUNTO: VOTO APROBATORIO

DR. JUAN ANTONIO MONTARAZ CRESPO  
DIRECTOR DE LA FES-CUAUTITLÁN  
P R E S E N T E

ATN: Ing. Jaime de Anda Montalvo  
Jefe del Departamento de Exámenes  
Profesionales de la FES

Con base en el art. 28 del Reglamento General de Exámenes, nos permitimos comunicarle a usted que revisamos la TESIS:

Elaboración de un manual práctico de Probabilidad y Estadística  
para la carrera de Ingeniería en Alimentos. (Revisión 1998)

que presenta el pasante: Vazquez Garduño Oscar  
con número de cuenta: 9256409-3 para obtener el TÍTULO de:  
Ingeniero en Alimentos

Considerando que dicho tesis reúne los requisitos necesarios para ser discutida en el EXÁMEN PROFESIONAL correspondiente, otorgamos nuestro VOTO APROBATORIO.

A T E N T A M E N T E.  
"POR MI RAZA HABLARA EL ESPIRITU"

Cuautitlán Izcalli, Edo. de México, a 9 de Julio de 1998

PRESIDENTE	<u>I.B.O. J. Jaime Flores Minutti</u>
VOCAL	<u>Ing. José L. Buenrostro Rodríguez</u>
SECRETARIO	<u>L.A. Rosalía Meléndez Pérez</u>
RIMER SUPLENTE	<u>I.B.Q Leticia Figueroa Villareal</u>
SEGUNDO SUPLENTE	<u>O.F.B. Hector Coss Garduño</u>

# DEDICATORIA

**DEDICAMOS ESTE TRABAJO A:**

**NUESTROS PADRES:**

IRMA CONSUEGRA REYES                      SILVIA GARDUÑO SOTELO  
GUILLERMO BALMORI BARRAGAN      ANGEL VAZQUEZ CARRIZALEZ

**NUESTROS HERMANOS:**

GUILLERMO, CLAUDIA, ELIZABETH,      SILVIA, ANGEL, ARTURO Y CESAR.  
FABIOLA Y JESSICA.

**NUESTROS AMIGOS DE VIDA:**

FAMILIA BARRERA AVILEZ Y FAMILIA ELIAS CASTRO, QUIENES NOS HAN DADO SU AMISTAD.

**NUESTROS AMIGOS DE LA FACULTAD:**

ARACELI DIAZ TORIBIO, OMAR CASTRO, JORGE UGALDE Y FAMILIA, Y HECTOR CHAVEZ POR SU APOYO DURANTE ESTE TIEMPO.

**NUESTROS AMIGOS DE HOY, MAÑANA Y SIEMPRE**

MARICELA ARAGÓN CALDERAS E ICARO CHAGOYA ARZATE.

**A NUESTRO ASESOR DE TESIS:**

LA ROSALIA MELÉNDEZ PÉREZ POR SU APOYO Y PACIENCIA.

*A TODOS USTEDES GRACIAS POR SU CONFIANZA, AMOR, SABIDURIA Y LEALTAD.*

**EN ESPECIAL DEDICAMOS ESTA TESIS**

**ADELA REYES OLIVARES**

**POR QUE ME ENSEÑASTE LAS COSAS BUENAS Y MALAS DE LA VIDA, POR HABERME ENSEÑADO A LUCHAR POR MIS IDEALES, POR QUE ME ENSEÑASTE A SER FUERTE Y PERSEVERANTE Y SOBRE TODO POR HABERME DADO TODO TU AMOR INCONDICIONAL, GRACIAS POR HABER EXISTIDO EN MI VIDA, SIEMPRE TE AMARE.**

**BALMORI.**

**CESAR VAZQUEZ GARDUÑO**

**POR HABER SIDO MI COMPAÑERO EN LAS TRAVESURAS Y LOS JUEGOS DE LA INFANCIA, POR QUE FUISTE MI DEFENSOR, POR TU ALEGRIA Y SABIDURIA, POR ENSEÑARME QUE LA VIDA HAY QUE VIVIRLA EN VIDA, NUNCA TE OLVIDARE PORQUE ERES PARTE DE MI CORAZON.**

**GRACIAS POR CREER EN MI. OSCAR VAZQUEZ.**

**GRACIAS DIOS POR HABERME DADO LA VIDA Y PERMITIR QUE LLEGARA HASTA AQUI**

# INDICE

<u>TEMAS</u>	<u>PAGINAS</u>
❖ INTRODUCCIÓN	I
❖ ANTECEDENTES	III
❖ SIMBOLOS	VI

## **CAPITULO PRIMERO**

<b><u>I.- TÉRMINOS BÁSICOS UTILIZADOS EN ESTADÍSTICA</u></b> .....	1
--	---

<b><u>II.- ESTADÍSTICA DESCRIPTIVA</u></b> .....	3
--	---

2.1).- <b><u>DATOS NO AGRUPADOS</u></b> .....	3
---	---

2.1.1) <b>MEDIDAS DE TENDENCIA CENTRAL:</b> .....	3
---	---

- \* Media Aritmética
- \* Mediana
- \* Moda
- \* Media Geométrica
- \* Media Armónica
- \* Media Ponderada o Podada
- \* Cuartiles y Percentiles

2.1.2) <b>MEDIDAS DE DISPERSIÓN O VARIABILIDAD:</b> .....	6
---	---

- \* Rango
- \* Varianza poblacional y muestral
- \* Desviación Estándar
- \* Coeficiente de Variación

2.2).- <b><u>DATOS AGRUPADOS</u></b> .....	9
--	---

- \* Amplitud
- \* Frecuencia
- \* Frecuencia Acumulada
- \* Frecuencia Relativa
- \* Frecuencia relativa acumulada
- \* Clase
- \* Marca de clase
- \* Media aritmética
- \* Mediana
- \* Moda
- \* Varianza
- \* Desviación Estándar
- \* Coeficiente de Variación

2.2.1) <b>REPRESENTACIÓN GRÁFICA DE DATOS AGRUPADOS:</b> .....	14
--	----

- a) Histograma
- b) Polígono de frecuencias

c) Curva suavizada	
d) Asimetría	
i) Media Podada	
ii) Media Windsor	
2.3).- PROBLEMAS RESUELTOS .....	16
<b>III.- PROBABILIDAD .....</b>	<b>25</b>
3.1).- REGLAS DE CONTEO .....	27
• Permutación	
• Arreglo	
• Combinación	
3.2).- OPERACIONES BÁSICAS CON CONJUNTOS .....	29
3.3).- AXIOMAS DE PROBABILIDAD .....	30
3.4).- PROBABILIDAD CONDICIONAL .....	30
3.5).- PROBLEMAS RESUELTOS .....	33
<b>IV.- DISTRIBUCIONES TEÓRICAS DE PROBABILIDAD .....</b>	<b>38</b>
4.1).- DISTRIBUCIÓN PROBABILÍSTICA .....	38
4.1.1) DISTRIBUCIÓN DE VARIABLE CONTINUA .....	38
1) Distribución Normal	
2) Distribución Gama	
3) Distribución Exponencial	
4) Distribución Ji-Cuadrada	
5) Distribución Weibull	
4.1.2) CUADRO RESUMEN DE VARIABLES CONTINUAS .....	43
4.1.3) DISTRIBUCIÓN DE VARIABLES ALEATORIAS DISCRETAS .....	44
1) Distribución Uniforme	
2) Distribución Binomial	
3) Distribución Multinomial	
4) Distribución de Poisson	
5) Distribución Geométrica	
6) Distribución Hipergeométrica	
4.1.4) CUADRO RESUMEN DE VARIABLES DISCRETAS .....	49
4.2).- COMPARACIÓN ENTRE LAS DISTRIBUCIONES BINOMIAL, NORMAL, POISSON, MULTINOMIAL .....	50
4.3).- PROBLEMAS RESUELTOS .....	51
<b>V.- MUESTREO .....</b>	<b>59</b>
5.1).- DEFINICIONES BÁSICAS .....	60
5.2).- BASES TEORÍCAS DEL MUESTREO .....	61
5.3).- TIPOS DE MUESTREO .....	61

5.3.1) MUESTREO ALEATORIO SIMPLE.....	62
5.3.2) MUESTREO ESTRATIFICADO .....	63
5.3.3) MUESTREO CONGLOMERADO .....	65
5.3.4) MUESTREO SISTEMATICO .....	65
5.3.5) MUESTREO DOBLE Y MULTIPLE .....	65
5.4).- <i>DISTRIBUCIONES MUESTRALES</i> .....	66
5.5).- <i>DISTRIBUCIÓN DE MEDIA Y VARIANZA DE UNA MUESTRA</i> .....	66
5.6).- <i>PROBLEMAS RESUELTOS</i> .....	68
<b>VI.- <u>TEOREMA DEL LIMITE CENTRAL</u></b> .....	70
6.1).- <i>PROBLEMAS RESUELTOS</i> .....	71
<b>VII.- <u>INFERENCIA ESTADÍSTICA O ESTIMACIÓN</u></b> .....	74
7.1).- <i>CARACTERÍSTICAS DE UN BUEN ESTIMADOR PUNTUAL</i> .....	75
1) Insesgabilidad	
2) Eficiencia	
3) Consistencia	
4) Suficiencia	
7.2).- <i>ESTIMACIÓN POR INTERVALOS DE CONFIANZA</i> .....	75
7.3).- <i>PRECISIÓN Y CONFIANZA</i> .....	76
7.4).- <i>DISTRIBUCIÓN t-STUDENT</i> .....	77
7.5).- <i>PROBLEMAS RESUELTOS</i> .....	81
<b>* RECAPITULACIÓN DE LOS TEMAS V, VI, VIII</b> .....	84
<b>VIII.- <u>PRUEBA DE HIPÓTESIS</u></b> .....	88
8.1).- <i>TIPOS DE ERRORES</i> .....	89
8.2).- <i>SECUENCIA DE APLICACIÓN</i> .....	92
8.2.1) <i>PRUEBAS SOBRE MEDIAS CON VARIANZA CONOCIDA</i> .....	93
8.2.2) <i>PRUEBAS SOBRE MEDIAS DE DISTRIBUCIONES NORMALES CON VARIANZA DESCONOCIDA</i> .....	94
8.3).- <i>PROBLEMAS RESUELTOS</i> .....	94

## CAPITULO SEGUNDO

<b>I.- ANÁLISIS DE VARIANZA (ANOVA)</b> .....	97
1.1).- <i>ANÁLISIS DE VARIANZA PARA UN SOLO FACTOR</i> .....	98
1.1.1) <i>ETAPAS PARA EL ANÁLISIS DE VARIANZA</i> .....	99
1.2).- <i>ANÁLISIS DE VARIANZA PARA DOS FACTORES</i> .....	101



1.3).- LA INTERPRETACIÓN DE "F" .....	103
A) Diferencia Mínima Significativa (DMS)	
B) Diferencia Mínima Significativa Honesta (DMSH)	
1.4).- PROBLEMAS RESUELTOS .....	106
<b>II.- BLOQUES ALEATORIOS</b> .....	119
2.1).- ANÁLISIS ESTADÍSTICO .....	120
2.1.1) ETAPAS PARA EL CÁLCULO DE BLOQUES ALEATORIOS.....	121
2.2).- PROBLEMAS RESUELTOS .....	124
<b>III.- ANÁLISIS FACTORIAL</b> .....	131
3.1).- ANÁLISIS ESTADÍSTICO .....	131
3.2).- INTERPRETACIÓN DE LOS EFECTOS PRINCIPALES Y DE LA INTERACCIÓN .....	133
3.3).- ETAPAS PARA EL ANÁLISIS FACTORIAL .....	134
3.3.1) VENTAJAS Y DESVENTAJAS DEL ANÁLISIS FACTORIAL .....	135
3.4).- DISEÑO FACTORIAL EN DOS FACTORES	
3.4.1) INTRODUCCIÓN Y NOTACIÓN .....	136
3.4.2) ANÁLISIS ESTADÍSTICO .....	137
3.4.3) SUBDIVISIÓN DE GRADOS DE LIBERTAD .....	138
3.5).- PROBLEMAS RESUELTOS .....	139
<b>IV.- REGRESIÓN LINEAL</b> .....	145
4.1).- CURVA DE AJUSTE .....	146
4.2).- REGRESIÓN LINEAL SIMPLE .....	147
4.2.1) ANÁLISIS DE CORRELACIÓN .....	147
4.3).- ANÁLISIS DE COVARIANZA .....	149
4.3.1) COEFICIENTE DE DETERMINACIÓN Y ANÁLISIS DE COVARIANZA..	149
4.4).- REGRESIÓN LINEAL MÚLTIPLE .....	151
4.4.1) ETAPAS PARA LA OBTENCIÓN DE LA ECUACIÓN LINEAL TRIPLE..	152
4.5).- MÉTODO DE MÍNIMOS CUADRADOS .....	153
4.5.1) ETAPAS PARA LA OBTENCIÓN DE MÍNIMOS CUADRADOS .....	154
4.6).- PROBLEMAS RESUELTOS .....	156

## **CAPITULO TERCERO**

### **I.- PROGRAMA EXCEL**

1.1).- **INTRODUCCIÓN** .....165

1.2).- **ALGUNOS ERRORES QUE GENERAN LAS FORMULAS** .....165

1.3).- **INTRODUCCIÓN DE FORMULAS CON ASISTENTE DE FUNCIONES** ..166

1.4).- **PROBLEMAS RESUELTOS** .....169

1.5).- **COMPARACIÓN ENTRE DOS PROBLEMAS** .....171  
(uno resuelto manualmente y el otro por el programa)

**CONCLUSIONES** ..... 173

• **APENDICE** .....174

• **BIBLIOGRAFÍA**

## INTRODUCCIÓN

La estadística o métodos estadísticos (como a veces se llama), está desempeñando un importante papel ascendente en casi todas las facetas del progreso humano; ya que anteriormente solo era aplicada a los asuntos del estado (de donde proviene su nombre) pero ahora su influencia se extiende a la Biología, Ingeniería, Negocios etc.

Debido al crecimiento que la estadística ha tenido, existe una gran información bibliográfica pero que en la mayoría de los casos los términos utilizados son presentados de manera compleja, por lo que surge la necesidad de facilitar la comprensión de las bases teóricas-prácticas de la estadística.

El objetivo de este manual<sup>1</sup> ó epítome<sup>2</sup> es que sirva como apoyo durante el proceso de enseñanza-aprendizaje para la materia de Probabilidad y Estadística, así como en todas las asignaturas experimentales llevadas dentro de la carrera de Ingeniería en Alimentos, y también en cualquier curso a nivel profesional en el área Químico-Biológica.

La elaboración de éste, fue realizada mediante una recopilación bibliográfica extensa, abarcando tanto el estudio de la Probabilidad, así como, de la Estadística y la relación que existe entre ambas disciplinas.

Durante su realización se trató de buscar ejemplos de aplicación correspondientes a la carrera, de tal manera que se motive al alumno a adquirir los conocimientos que aquí se exponen, cabe mencionar que no se eluden fórmulas ni expresiones matemáticas, pero se hizo que toda fórmula parezca natural, sencilla y razonable.

Sin embargo, aquellos que no deseen involucrarse en las definiciones de cada fórmula podrán usar los cuadros resumen que se presentan al final de cada tema.

El manual consta de 13 temas divididos en 3 capítulos. Esta división se debe a que en el primer capítulo se ven algunos de los temas que se imparten en la asignatura de Probabilidad y Estadística, siendo éstos necesarios para la comprensión de los siguientes temas. En el segundo capítulo se ven temas que no se imparten en la asignatura, pero, sirven para el análisis y resolución de diversos problemas. En el último capítulo se aborda un tema que sirve como herramienta para la resolución de problemas.

En el primer capítulo se dan los conceptos básicos utilizados en la estadística, la estadística descriptiva tanto para datos agrupados como para no agrupados y el estudio de la probabilidad donde se usaron diagramas para facilitar su comprensión. Otros temas que se estudian son las distribuciones de probabilidad discretas y continuas, en general; y algunos casos particulares como las distribuciones: Binomial, Poisson, Multinomial, Hipergeométrica y la Normal. También dentro de este capítulo se ven los diferentes tipos de muestreo así como las distribuciones muestrales. Por otra parte también se ilustra la Estadística inferencial donde se estudia la estimación por intervalo así como las pruebas de hipótesis.

<sup>1</sup> Libro en que se compendia lo más substancial de una materia

<sup>2</sup> Resumen o compendio de una obra extensa, abreviando y resumiendo cuanto es posible la materia tratada en ella y exponiendo únicamente lo más fundamental o preciso

De los temas que acaban de ser mencionados es importante señalar que solamente algunos son los que se estudian dentro del curso de Probabilidad y Estadística.

En el segundo capítulo se tratan temas más completos como son ANOVA, Factoriales, Bloques Aleatorios etc. Los cuales son de gran ayuda para poder realizar un tratamiento estadístico más completo. Estos temas son presentados de una manera sencilla y práctica ya que generalmente en la bibliografía los presentan de una forma enredosa provocando que para el estudiante se conviertan en temas complejos cuando en realidad no lo son.

En el tercer capítulo se maneja un programa de computación, el cual sirve para realizar los cálculos con mayor rapidez, exactitud y a su vez disminuir el porcentaje de error que se pudiera tener al realizar una operación manualmente. El objetivo de este capítulo es orientar al estudiante sobre algunas herramientas de apoyo, que conozca los beneficios que ofrecen pero a su vez darse cuenta que es necesario tener un conocimiento básico de lo que es la estadística.

En el Apéndice de este manual se presentan todas las tablas que pueden ser utilizadas para la solución de problemas.

## ANTECEDENTES

La Estadística comprende tanto la teoría como su aplicación práctica que incluye los métodos de recolección de datos y los de análisis e interpretación de los mismos. Mucha gente piensa en la Estadística como si se tratase únicamente en reunir números, de hecho éste fue el significado original pero con el tiempo las ideas se han modificado por lo que a continuación, se describen diferentes términos de la Estadística.

- REAL ACADEMIA ESPAÑOLA.- (de Estadística). Censo o recuento de la población, de los recursos naturales e industriales, del tráfico o cualquier otra forma de manifestación del estado, provincia, pueblo, clase etc. // Estudio de los hechos morales o físicos del mundo que se presentan a numeración o recuento.
- IDEA POPULAR.- La Estadística esta intimamente ligada a las actividades de recuento y enumeración que tradicionalmente han sido realizadas con el Estado con diversas finalidades: conocer el número de habitantes, registros de volúmenes de cosecha de productos agrícolas etc.
- DEFINICIÓN DE UTILIDAD.- La Estadística es un conjunto de técnicas para la colección, manejo, descripción y análisis de información, de manera que las conclusiones obtenidas de ella tengan un grado de confiabilidad especificado. (29)
- ESTADÍSTICA.- Esta ligada con los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones, como para tomar decisiones razonables de acuerdo con tales análisis. (1)
- LA ESTADÍSTICA es considerada como el suministro de un conjunto de herramientas sumamente útiles en la investigación, la ciencia de la estadística trata con:
  1. Colección y compendio de datos.
  2. Diseño de experimentos y reconocimientos.
  3. Medición de la variación tanto de datos experimentales como de reconocimiento.
  4. Estimación de parámetros de población y suministro de varias medidas de la exactitud y precisión de esas estimaciones.
  5. Ensayo de hipótesis respecto a poblaciones.
  6. Estudio de la relación entre dos ó más variables. (2)

De uno u otro modo, las técnicas Estadísticas se usan en casi todas las ramas de la Ciencia moderna, así como en muchos otros campos de la actividad humana, como por ejemplo, en la Medicina, Biología, Economía, Política, Ingeniería etc.

Por lo tanto la Estadística es capaz de proporcionarnos métodos para una gran cantidad de problemas prácticos importantes.

Debido a esto la Estadística puede servir para:

1. Encontrar de acuerdo a los objetivos perseguidos, de acuerdo a las condiciones experimentales y de acuerdo a la naturaleza de las características que se observan por el experimentador, aquel diseño que optimiza la cantidad de información necesaria y la utiliza, ó utiliza estos recursos para alcanzar sus metas.
2. Para saber todos los detalles estructurales del experimentador, ya que muchas veces este proceso lleva al experimentador a darse cuenta de factores que no había considerado.
3. Una vez que han procesado los datos estadísticamente, se puede interpretar o traducir los números a un lenguaje común, pudiendo así considerarse una decisión en la resolución de un problema particular.
4. Nos permite manejar los números a nuestro antojo, manipulando los números de acuerdo con ciertas reglas establecidas y se puede probar lo que se desee. (5)

Pero también tiene una ubicación importante dentro de la experimentación.

En el *METODO CIENTÍFICO* que es la búsqueda o planteamiento de un problema donde se formulan hipótesis y se obtienen principios y/o leyes; y pueden seguir dos caminos:

a) Método inductivo: Donde se buscan hechos a través de la experimentación y la observación y es por medio de la confrontación de las hipótesis planteadas y de lo observado que se llega a una conclusión.

b) Método deductivo: Consiste en clasificar y ordenar los hechos por medio de una relación, si ésta es constante de una manera que se pueda predecir un hecho y confirmarlo mediante la experimentación, puede ser general y formarse un postulado básico o una ley.

(3)

En el *MÉTODO EXPERIMENTAL* donde el científico “manipula” las condiciones en las que se reproduce un fenómeno para observar sus consecuencias. La función característica aunque no exclusivo de éste método es la contrastación de hipótesis, por ello se considera a la experimentación como un momento del método científico en el que la recolección de datos bajo condiciones controladas nos permite una evaluación rigurosa de las hipótesis.

El *MÉTODO ESTADÍSTICO* que es el conjunto de reglas algorítmicas que permiten realizar cálculos estadísticos en forma eficiente para dar respuestas satisfactorias, estos métodos pueden ser:

a) Exploratorios.- Tratan por un lado de traducir cualquier información en un discurso que pueda ser probabilizado, ajustando los datos empíricos a una variable aleatoria, estos métodos son por un lado herederos de la estadística descriptiva, y por otro lado recogen todos los instrumentos estadísticos inferenciales y decisorios.

b) Confirmatorios.- No solo tratan de determinar el mejor modelo de distribución ajustable a los datos que tenemos o tratamos de explicar, sino que también tratan de conocer el modelo procesual (Estocástico- Probabilístico)

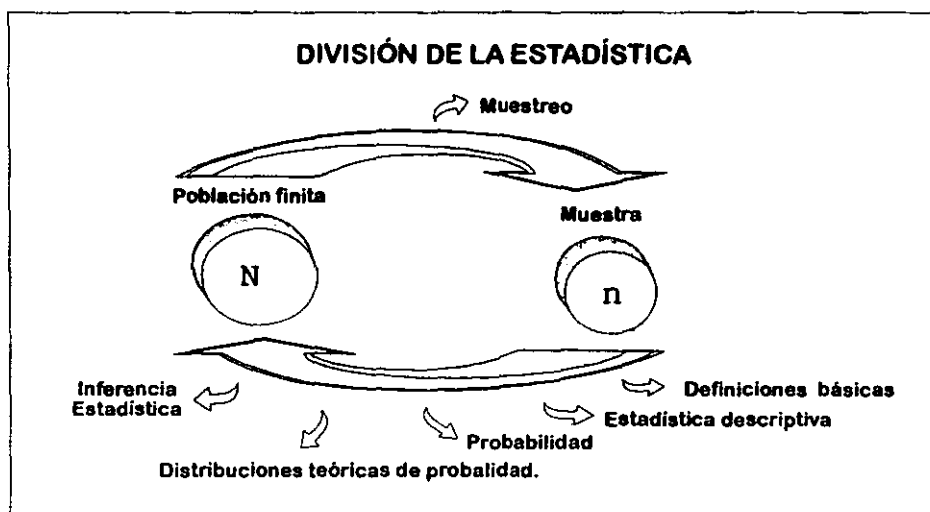
Dentro del *DISEÑO EXPERIMENTAL* ya que es una descripción previa de un plan de investigación y es también un procedimiento utilizado en investigaciones experimentales cuyo objetivo es la verificación de hipótesis determinadas a través de conclusiones finales acerca de las relaciones entre variables dependientes e independientes. El diseño experimental incluye la secuencia completa de pasos tomados de antemano para asegurar que los datos apropiados se obtendrán de modo que permitan un análisis objetivo que conduzca a deducciones válidas con respecto al problema establecido.

En general responde en síntesis a un intento de optimizar (maximizar ó minimizar) el proceso de investigación, incluye normalmente:

- \* Formulación del problema.
- \* Formulación de conceptos precisos y estimadores óptimos.
  - que es exactamente lo que queremos
  - como se medirá el efecto
  - que factores pueden afectar las características o efectos a estudiar etc. (8)

En la figura N°1 se presenta el diagrama de División de la Estadística que es una representación breve de los temas que se tratan en el primer capítulo de éste manual. Este diagrama nos indica que a partir de un conjunto de datos, elementos, cosas etc. Es necesario tomar una muestra que sea representativa de dicho conjunto (población finita), para poder realizar un tratamiento estadístico determinado, con el fin de conocer el comportamiento y características de la población finita, cabe mencionar que si se desea hacer un muestreo éste se realizará a partir de dicha población.

Fig.N°1



## SIMBOLOS

$\Sigma$	Sumatoria.
$R$	Conjunto de los números reales.
$Z$	Conjunto de los números enteros.
$\emptyset$	Conjunto vacío.
$U$	Conjunto universal.
$\in$	Pertenece a.
$\notin$	No pertenece a.
$\subset$	Contenido en, subconjunto de.
$\not\subset$	No está contenido en.
$\subseteq$	Suconjunto de.
$\not\subseteq$	No es subconjunto de.
$\mathbb{N}$	Conjunto de números naturales.
$\Leftrightarrow$	Si y solo si.
$\cap$	Intersección de conjuntos.
$\cup$	Unión de conjuntos.
$A^c$	Complemento del conjunto A.
$>$	Mayor que.
$\geq$	Mayor o igual a.
$<$	Menor que.
$\leq$	Menor o igual a.
$n!$	Factorial de un entero positivo.
$n_{Cr} = \binom{n}{r}$	Combinaciones de n elementos tomados de r en r.
$S$	Espacio muestral.
$P(A)$	Probabilidad del evento A.
$P(B A)$	Probabilidad condicional.
$X$	Variable aleatoria x.
$x_i$	Valor de la variable aleatoria.
$F(X)$	Función de probabilidad o función densidad de probabilidad de la variable aleatoria X.
$f(x_i)$	Función de probabilidad para un valor $x_i$ de la variable aleatoria.
$\mu$	Media.
$\sigma^2$	Varianza de una distribución de probabilidad (de la población).
$\sigma$	Desviación estandar.
$p$	Probabilidad de éxito de un evento.
$q=1-p$	Probabilidad de fracaso en un evento.
$\lambda$	Media y varianza de la distribución de Poisson.
$\approx$	Aproximadamente igual a.
$Z$	Variable aleatoria normal estándar.
$g.l.$	Grados de libertad.
$\Gamma$	Función gamma.
$\bar{x}$	Media muestral.
$Me$	Mediana.
$Mo$	Moda o modo.



- $\prod_{i=1}^n x_i$  Producto de los valores  $x_i$  desde  $i=1$  hasta  $n$ .  
 R Rango o recorrido.  
 $||$  Valor absoluto.  
 $s^2$  Varianza muestral.  
 $s$  Desviación estandar muestral.  
 C.V. Coeficiente de variación.  
 $c$  Amplitud de clase.  
 $n$  Tamaño de muestra.  
 $H_0$  Hipótesis nula.  
 $H_a$  Hipótesis alterna.  
 $1-\alpha$  Nivel de confianza.  
 $\alpha$  Nivel de significación, probabilidad de error de tipo I.  
 $\beta$  Probabilidad de error de tipo II.  
 $\mu_0$  Valor supuesto de  $\mu$ .  
 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$  Error estándar de la media.  
 $\pi$  Proporción binomial poblacional.  
 $\pi_0$  Valor supuesto de  $\pi$ .  
 $t$  Variable aleatoria  $t$  de student.  
 $s_{\bar{x}} = \frac{s}{\sqrt{n}}$  Estimación del error estándar de la media.  
 $\sigma_0^2$  Valor supuesto de  $\sigma^2$ .  
 $\tau_j$  Efecto de tratamiento.  
 $\epsilon_{ij}$  Error residual.  
 $X_{ij}$   $i$ -ésima observación que recibe el  $j$ -ésimo tratamiento.  
 $X_{\cdot j}$  Suma del  $j$ -ésimo tratamiento.  
 $\bar{x}_{\cdot j}$  Media del  $j$ -ésimo tratamiento.  
 $X_{\cdot \cdot}$  Suma de todas las observaciones.  
 $\bar{x}$  Gran media.  
 D.S.M. Diferencia significativa mínima de Fisher.  
 D.S.H. Diferencia significativa honesta de Tukey.  
 $\beta_i$  Efecto de bloques.  
 $\mu_{Y/X}$  Valor medio de  $Y$  correspondiente a un valor de  $X$ .  
 $\hat{Y}$  Estimador de  $\mu_{Y/X}$ .  
 $A$  Ordenada al origen de la recta de regresión poblacional.  
 $B$  Ordenada de la recta de regresión poblacional.  
 $a$  Estimador de  $A$ .  
 $b$  Estimador de  $B$ .  
 $\sigma_{Y/X}^2$  Varianza de regresión de  $Y$  sobre  $X$ .  
 $s_{Y/X}^2$  Estimador de la varianza de regresión de  $Y$  sobre  $X$ .  
 $\hat{Y}_0$  Valor de  $\hat{Y}$  correspondiente a  $X_0$ .  
 $\rho$  Coeficiente de correlación lineal poblacional.  
 $r$  Estimador de  $\rho$ .  
 $r^2$  Coeficiente de determinación.

## CAPÍTULO PRIMERO

Antes de profundizar sobre los temas que se expondrán posteriormente es importante comprender el significado de ciertos términos que serán empleados con frecuencia a través del texto.

De igual manera este material puede servir de ayuda aquellos lectores, para quienes el campo es completamente nuevo, por lo tanto este material será presentado en forma sencilla.

### I.- TÉRMINOS UTILIZADOS EN ESTADÍSTICA (conceptos básicos)

- **POBLACIÓN.**- La mayor colección de elementos por los cuales se tiene cierto interés en un instante particular, como complejo de objetos, medidas o individuos que poseen una característica particular, esta puede ser finita o infinita. (25)
- **PARÁMETRO.**- Cualquier característica de la población que sea medible. (5)
- **MUESTRA.**- Subconjunto de mediciones seleccionadas de la población de interés. (5) Su selección será de acuerdo con una regla o un plan, las cosas importantes que debemos saber son: 1) que estamos tratando con una muestra, 2) que población ha sido muestreada. (9)
- **ESTADÍSTICO.**- Número resultante de la manipulación de ciertos datos iniciales de acuerdo con determinados procedimientos específicos. Comúnmente utilizamos un estadístico que se calcula a partir de una muestra para estimar el parámetro de una población, por ejemplo: una muestra de envases de plástico se emplea para estimar la proporción de envases defectuosos en la población total. (5)
- **VARIABLE.**- Es un símbolo, tal como  $x$ ,  $y$ ,  $h$ , que puede tomar un valor cualquiera de un conjunto determinado de ellos. Una variable se diferencia de una constante ya que el valor de esta nunca varía. (1)
- **VARIABLE CUALITATIVA.**- Característica o fenómeno que solo puede catalogarse o nominarse. (13)
- **VARIABLE CUANTITATIVA.**- Que se mide o se cuenta. (13)
- **VARIABLE ALEATORIA.**- Los valores que son obtenidos y que provienen del azar. (2)
- **VARIABLE ALEATORIA CONTINUA.**- Cuando es posible que la variable tome todos los valores a lo largo de una recta numérica, o sea que no hay saltos ni interrupciones (teóricamente no debe haber saltos). (10)

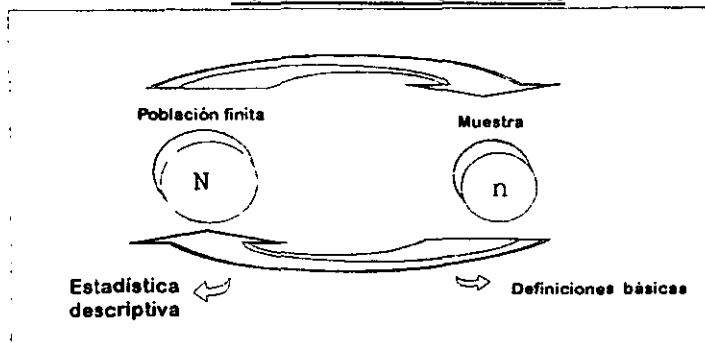
- **VARIABLE ALEATORIA DISCRETA.**- Cuando la variable puede tomar solo algunos valores, o sea que hay saltos e interrupciones. Valores unitarios o enteros, ejemplo:
  - 1.- En una familia el número N de hijos puede tomar cualquiera de los valores 0, 1, 2, etc (por lo tanto es una variable discreta), pero no puede ser 2.5 o 3.84.
  - 2.- La altura de un evaporador puede ser 15 mts, 16.5 mts ó 17.835 mts por lo tanto es una variable continua. (8)
- **DATOS.**- Números o medidas que han sido recopilados como resultado de observaciones, pueden ser recuentos como los datos de frecuencias. (5)
- **NIVEL DE COMPARACIÓN.**- Utilizando una variable continua, al efectuarse una medición a un conjunto de barras, del número de lote de fabricación podemos utilizar un metro, posiblemente encontremos que no hay diferencia entre ellas, pero si usamos un nivel de comparación más exacto (flexometro) podemos encontrar 1% o 5% de población diferente. Por lo tanto, al aumentar el nivel de comparación podemos encontrar mayor porcentaje de diferencia, diciendo entonces, que la diferencia observable será tanto más grande cuanto más extrico es el nivel de comparación usado. (8)
- **PRECISIÓN.**- Función de compactamiento de valores observados. (7)
- **EXACTITUD.**- Función del estándar, especificación u objeto fijado. (1)

Dentro del nivel de comparación usualmente encontramos términos que se relacionan con la medida dada por el instrumento utilizado, por ejemplo: para el caso de la precisión en un conjunto de lecturas encontramos que si los valores son muy similares pueden presentar alto valor, aunque no se encuentren cercanos al valor verdadero (exactitud), en el caso de la figura N°2 se esquematizan eventos con las diferentes formas de visualizar estas "definiciones"

Fig. N°2



## II.- ESTADÍSTICA DESCRIPTIVA



En la Estadística deben diferenciarse dos funciones del método estadístico como son: técnicas de estadística descriptiva y técnicas de inferencia estadística.

La *Inferencia Estadística* se ocupa de generalizar la información, más específicamente hace inferencia acerca de poblaciones a partir de las muestras extraídas de la misma

La *Estadística Descriptiva* abarca medidas de tendencia central, medidas de dispersión y representación gráfica, el objetivo central es presentar información en forma conveniente, útil y comprensible, por lo tanto es útil cuando se tiene una recopilación de datos como recuentos, frecuencias, datos de calificaciones, etc.. Al realizar la función descriptiva el estadístico formula reglas y procedimientos para la presentación de los datos en una forma más significativa y pueden ser representados gráficamente.

La estadística descriptiva se divide en dos partes que son datos agrupados y datos no agrupados. (12)

### 2.1).- DATOS NO AGRUPADOS

Si los datos son cuantitativos, dependiendo del tamaño de la muestra se tratarán en forma agrupada o no. Si se tienen muchos datos diferentes es conveniente agruparlos en clases ó intervalos ya que su distribución de frecuencia y gráficas resultan muy complicadas y hasta confusas. Los datos no agrupados se dividen en dos tipos que son: medidas de tendencia central y medidas de dispersión o variabilidad. (12)

#### 2.1.1 MEDIDAS DE TENDENCIA CENTRAL

Es una medida que indica la ubicación del centro de la distribución, el valor central o típico del conjunto en el sentido que es el más representativo de un conjunto de datos, también se le llama promedio.

Hay muchos promedios cada uno de los cuales posee propiedades particulares y cada uno es típico en alguna forma única. A los promedios se les llama *Medidas de Tendencia Central* y los más frecuentemente encontrados son la media aritmética, la mediana y la moda; otros menos usados son el centro de amplitud, la media armónica, la media geométrica y la media aritmética ponderada. (27)

\* **MEDIA ARITMÉTICA.**- Sencilla de calcular y usada para cualquier dato, puede servir para cálculos posteriores y para inferencia estadística, medida única en un grupo de datos, relativamente viable ya que si se sacan las medidas de muchas muestras de una misma población no varía tanto como otros estimadores de la media poblacional ( $\mu$ ).

Es afectada por valores aislados o alejados del conjunto de datos y por lo tanto en estos casos se puede desviar demasiado.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots(1)$$

Tiene ventajas y desventajas como son:

- Es el promedio más comúnmente empleado, es fácil de calcular, se entiende fácilmente, se presta a manipulaciones algebraicas
- Es uno de los miembros del sistema matemático que permite su empleo en análisis estadísticos más avanzados
- Es el más estable de las medidas de tendencia central
- La mayor desventaja es que se ve afectada por los valores extremos y por consiguiente puede estar muy lejos de ser una representación de la muestra. (19)

\* **MEDIANA.**- Valor central de una serie de datos ordenados, sus ventajas son las siguientes:

- No se ve afectada por valores extremos o lejanos
- Es única en conjunto de datos
- Divide a los datos en dos partes de modo que la primera mitad contiene valores menores a la mediana y la segunda mitad valores mayores
- Se puede usar para definir el punto medio de un cierto número de objetos que por sus cualidades no sean cuantitativos
- Divide el área bajo la curva en dos partes iguales. (4)

Desventajas:

- Difícil de calcular si hay muchos datos.
- No tiene propiedades algebraicas. (12)

Me = valor central de datos ordenados .....(2)

- **MODA.**- La moda se denota por M o Mo y es por el valor que con más frecuencia se presenta en un conjunto de datos. Es muy fácil de determinar, basta con observar detenidamente el conjunto de datos y ver cual es el que repite; sin embargo, no es muy útil por que puede ocurrir que una distribución tenga dos o más valores que se repitan con la misma frecuencia, en tal caso se tiene dos o más modas. También puede ocurrir que ningún valor se repita y entonces no habrá moda. Por otra parte puede ser un valor extremo el de mayor frecuencia y difícilmente podría ser considerado una medida de tendencia central. (12)

Ventajas:

- Fácil de calcular.
- Valor "típico", más valores reunidos en este punto que cualquier otro.

Desventajas:

- No se presta para el análisis estadístico.
- Puede haber más de una moda o ninguna.

$M_o =$  Valor que se presenta con mayor frecuencia. ....(3)

\* **MEDIA GEOMÉTRICA.**- Se utiliza cuando los valores individuales sean razones de proporción de otros (rendimiento de combustible de coches, número de bacterias en un cultivo microbiológico, etc..) Cuando se tiene constante el denominador. (8)

$$mg = \sqrt{(x_1)(x_2)...(x_n)} \quad \log mg = \frac{1}{n} \sum_{i=1}^n \log x_i \quad \dots (4)$$

Ejemplo: La media geométrica de los números 2, 4, 8, es

$$mg = \sqrt{(2)(4)(8)} = \sqrt[3]{64} = 4$$

\* **MEDIA ARMÓNICA.**- Para promediar razones o proporciones pero bajo ciertas condiciones específicas, es decir con el numerador fijo, ejemplo (125 huevos/ \$100)

$$ma = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad \dots (5)$$

La relación existente entre la media aritmética, geométrica y armónica es la siguiente:

La media geométrica de una serie de números positivos es menor o igual que su media aritmética pero es mayor o igual que su media armónica. (26)

\* **MEDIA PONDERADA.**- Es el promedio de otros promedios, a veces se asocia a los números  $x_1, x_2, \dots, x_n$  ciertos factores o pesos  $w_1, w_2, \dots, w_k$  que dependen de la significación o importancia de cada uno de los números. (25)

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_k x_n}{w_1 + w_2 + \dots + w_k} = \frac{\sum wx}{\sum w} \quad \dots (6)$$

- **CUARTILES Y PERCENTILES.**- Se define al primer *cuartil* como el valor de un conjunto de observaciones tal que ni exceda al 25% de las mismas, ni que sea excedido por más del 75%. Al segundo cuartil debe considerarse como la mediana, el tercer cuartil es el valor de un conjunto de datos u observaciones tal que ni exceda el 75% de las observaciones ni que sea excedido por más del 25% de las mismas.

Son útiles cuando se desea dividir un conjunto de datos en determinado número de grupos, de modo que cada uno de ellos contenga el mismo número de valores. (5)

El *percentil* se puede definir como el valor correspondiente a un punto tal que no exceda el porcentaje determinado por el mismo, ni sea excedido por más del (100% - el % determinado) de las observaciones, así el percentil 10 corresponderá al valor de un conjunto de observaciones que ni exceda el 10% ni sea excedido por más del 90% (o sea 100-90). Los percentiles pueden ser determinados con exactitud mediante una tabla de frecuencias relativas acumuladas de datos no agrupados y en forma aproximada mediante una tabla de datos agrupados. (6)

$$\text{Rango percentil} = (f_{\text{acum}} / N) 100 \dots(7)$$

Así una calificación abstracta no tiene significado, tampoco lo tiene un rango percentil, ya que debe expresarse siempre en relación a algún grupo de referencia. (5) Habiendo localizado el centro de una distribución de datos, nuestro siguiente paso es proporcionar una medida de la variabilidad o dispersión de datos, esta medida puede ser la amplitud pero no es completamente satisfactoria como medida de variación por lo que introducimos los cuartiles y percentiles y la aplicación de uno y otro modo es en gran medida a la cantidad de datos presentes.

### 2.1.2 MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Además de la tendencia de los valores a agruparse en las cercanías de un valor promedio, es necesario saber cuanto se dispersan o varían los valores, es decir, si están uno cerca de otro o alejados, por ejemplo, si fabricásemos latas, una variación excesiva en el diámetro de estas implicaría un alto % de productos defectuosos. Las medidas de este acercamiento o alejamiento se conocen como medidas de variabilidad o de dispersión, siendo las más usadas: Amplitud total o Rango, Varianza, Desviación estándar y Coeficiente de variación.

\* **RANGO**.- (Conocido también como horquilla, recorrido, gama, horizonte, amplitud, etc.). Es la medida de dispersión más simple, pero no toma en cuenta todas las observaciones ya que solamente considera a los valores extremos, por lo tanto no utiliza toda la información disponible. Conforme aumenta el número de observaciones, aumentará posiblemente el valor del rango o en el menor de los casos será constante pero no disminuirá.

Puede darse el caso de que se presenten uno o dos valores excepcionalmente alejados de la mayoría restante, por lo tanto esta lejanía nos mueve a dudar que dichos valores extremos sean verdaderamente representativos de los valores centrales.

Por lo tanto *se recomienda que se utilice cuando el tamaño de la muestra es pequeño* (10 datos como máximo)

$$R = \text{valor mayor} - \text{valor menor} \dots(8)$$

\* **VARIANZA Y DESVIACIÓN ESTÁNDAR**.- Se pueden considerar como las distancias de cualquier medida del conjunto de datos con respecto a la media de los mismos, la diferencia entre la varianza y la desviación estándar estriba en que la primera enmarca a esta distancia o diferencia en unidades cuadráticas y la segunda en unidades originales.

La *varianza* toma en cuenta la distancia de cada medida con respecto a su media, a esta distancia se le llama desviación, por lo tanto podemos tener desviaciones tanto positivas

como negativas, y esto es mucho más claro al determinar la raíz cuadrada a la varianza dando como resultado a la *desviación estándar*. (2)

Cuando es necesario distinguir la desviación típica de una población, de la desviación típica de una muestra sacada de esta población se emplea el símbolo  $s$  para la última y  $\sigma$  para la primera. Así la  $s^2$  y  $\sigma^2$  representarían la varianza muestral y poblacional respectivamente. (1)

Ejemplo: Supongamos que tenemos dos muestras: muestra #1  $x_1 = 17, x_2 = 18, x_3 = 19$   
muestra #2  $x_1 = 20, x_2 = 19, x_3 = 15$

La media aritmética para cada muestra es la misma  $\bar{x} = 18$ . Para determinar la medida de dispersión que nos indique cuanto se alejan los datos individuales de su media aritmética, determinamos la diferencia:

muestra #1	muestra #2
$(x_i - \bar{x}) =$	$(x_i - \bar{x}) =$
$17 - 18 = -1$	$20 - 18 = 2$
$18 - 18 = 0$	$19 - 18 = 1$
$19 - 18 = 1$	$15 - 18 = -3$
$\Sigma = 0$	$\Sigma = 0$

Para eliminar que el resultado de la sumatoria de las diferencias sea cero, elevamos al cuadrado cada una de estas diferencias

muestra #1	muestra #2
$(x_i - \bar{x}) =$	$(x_i - \bar{x}) =$
$(17 - 18)^2 = 1$	$(20 - 18)^2 = 4$
$(18 - 18)^2 = 0$	$(19 - 18)^2 = 1$
$(19 - 18)^2 = 1$	$(15 - 18)^2 = 9$
$\Sigma = 2$	$\Sigma = 14$

Por lo que esta suma se conoce como la suma de cuadrados (SC) o sea, la suma de cuadrados de las desviaciones respecto a la media.

La Varianza de una población de  $n$  observaciones, se define como el promedio del cuadrado de las desviaciones de las observaciones con respecto a su media ( $\mu$ ). (14)

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \dots(9)$$

donde  $N =$  número de observaciones de la población



La Varianza de una Muestra de  $n$  observaciones, se define como la suma de los cuadrados de las desviaciones con respecto a su media y dividida entre  $(n-1)$ . (20)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots(10)$$

Donde  $(n-1)$  son los **grados de libertad** que se refieren al número de valores que están libres para variar después de haber impuesto ciertas restricciones a los datos.

**\* DESVIACIÓN ESTÁNDAR.**- Representa dispersión de datos por lo que la variabilidad de diferentes distribuciones pueden compararse en términos de la desviación estándar y se representa con la letra  $S$ .

La  $S$  permite la interpretación precisa de los datos dentro de una distribución, la  $S$  como la media, es miembro de un sistema matemático que permite su utilización en consideraciones estadísticas más avanzadas. La  $S$  es por lo general más conveniente que la varianza para interpretar la variabilidad de un conjunto de datos, ya que la varianza está expresada en unidades al cuadrado, en tanto que la desviación tiene las mismas unidades que los datos originales. (23)

$$S = \sqrt{s^2} \dots(11)$$

**\* COEFICIENTE DE VARIACIÓN.**- Es la medida de variación relativa y no absoluta como la desviación estándar, representa a la desviación estándar como un porcentaje de la media. Es muy útil para demostrar el porcentaje de variación entre muestras y estas pueden estar en diferentes unidades por ejemplo una comparación entre alturas y variación de peso de latas de conservas. Un inconveniente del coeficiente de variación es que deja de ser útil cuando la media está próxima a cero. (19)

$$C.V = \frac{s}{\bar{x}} \times 100 = \% \dots(12)$$

## 2.2).- DATOS AGRUPADOS

Se entiende por datos agrupados la utilización de un método que permite organizar y presentar observaciones de un gran conjunto de datos, de tal forma que los aspectos más sobresalientes de los mismos sean más rápidos y fácilmente realizables. Por lo tanto se requiere que éstos métodos proporcionen:

- máxima información contenida y que sea fácilmente visualizada.
- que resuma o describa grandes masas de datos sin pérdida o distorsión de las características de la información
- que posean sencillez operativa.
- que se presenten los datos en forma estética. (14)

Ejemplo: La siguiente tabla es una distribución de frecuencias de pesos (gr) de 100 latas de conserva

Peso (gr)	Número de latas
60-62*	5
63-65	18
66-68	42
69-71	27
°72-74	8
	$\Sigma = 100$

\* Limite superior, ° Limite inferior de la clase ó 5ª clase.

La primera clase o categoría comprende los pesos de 60-62 gr, puesto que 5 latas tienen un peso perteneciente a esta clase, la correspondiente frecuencia de la clase es 5, y a esto se le llama datos agrupados. (1)

Para ordenar los datos agrupados se selecciona un conjunto de intervalos contiguos, que no se traslapen, tal que cada valor en el conjunto de observaciones pueda colocarse en uno y solo uno de los intervalos. Son los **intervalos de clase**. Por lo general, se recomienda de cinco a veinte clases dependiendo de la cantidad de datos, pero pocas clases no son convenientes puesto que hay más pérdida de información y muchas clases hacen perder la sintetización.

Para poder agrupar los datos obtenidos, es necesario conocer los términos que se utilizan para realizar una distribución de frecuencias, y son los siguientes:

\* **NÚMERO DE CLASES.**- Se puede recomendar lo siguiente: (14)

Si el número de valores es:	Las clases apropiadas son:
10-100	4-8
100-1000	8-11
1000-10000	11-14

Pero hay algunas fórmulas que pueden utilizarse para definir estas clases: (14)

$$\text{I.M.E.C.C.A.: } k = \sqrt{n}$$

$$\text{Sturges: } k = 1 + 3.322 \log_{10}(n) \dots (13)$$

$$\text{Reyes Castañeda: } k = 2.5\sqrt{n}$$

Donde:  $k$  = número de clases y  $n$  = número de datos.

En la utilización de cualquiera de las fórmulas anteriores, se recomienda cerrar el número de clases al entero más próximo.

\* **AMPLITUD ( $w$ ).**- Es la abertura que tendrá cada uno de los intervalos de clase. Y esta amplitud dependerá si la variación es del tipo continuo (cuando también se utilizan decimales) o discreto (cuando el número es entero).

$$w = \frac{R}{k} \dots(14)$$

Donde:  $w$  = amplitud,  $R$  = rango, y  $k$  = número de clases.

En el caso de trabajar con variable de tipo continuo, se debe de cumplir que el producto de  $k * w$  sea mayor al rango, determinado así de donde partirá la primera clase, asegurando contener a los valores o datos.

\* **FRECUENCIA ( $f$ ).** - El número de veces que aparece un valor  $x$  en una clase (4)

\* **FRECUENCIA ACUMULADA( $f_n$ ).** - Se acumulan las frecuencias para facilitar la obtención de la información. Representa a la frecuencia de los valores dentro de dos o más intervalos de clase.

\* **FRECUENCIA RELATIVA( $f_r$ ).** - Para conocer la proporción de valores que caen dentro de un intervalo particular. Es la frecuencia de la clase dividida por el total de frecuencias de todas las clases y se expresa generalmente como %. (1)

$$fr = f / ft \times 100 \dots(15)$$

\* **FRECUENCIA RELATIVA ACUMULADA( $f_{ra}$ ).** - Es la división de la frecuencia acumulada entre el tamaño  $n$  de la muestra. (4)

\* **CLASE.**- Es una agrupación dentro de la cual situamos todos los valores que son semejantes, basándose en la amplitud de los datos, se elige un intervalo de clase o ancho de la clase conveniente, con márgenes de clase que definen los límites inferior y superior de cada clase. (7)

\* **MARCA DE CLASE( $x_f$ ).**- Es el valor medio de un intervalo de clase. Se determina como la media aritmética de los límites inferior y superior de cada intervalo de clase:

$$x_f = \frac{L_i + L_s}{2} \dots(16)$$

Donde  $x_f$  = marca de clase,  $L_i$  = límite inferior de clase y  $L_s$  = límite superior de clase.

\* **MEDIA ARITMÉTICA.** - Tiene las mismas características que se revisaron en datos no agrupados, pero como se trabaja con frecuencias dentro de cada intervalo de clase, es necesario considerar dichas frecuencias, por lo tanto la fórmula para su cálculo queda de la siguiente forma: (18)

$$\bar{x} = \sum_{i=1}^n \frac{f_i x_i}{n} \dots (17)$$

Donde:  $\bar{x}$  = media aritmética,  $\sum_{i=1}^n$  representa a la sumatoria desde el primer valor ( $i=1$ ) hasta el último valor ( $n$ ),  $f_i$  = frecuencia de la clase deseada y  $x_i$  = marca de clase de la clase deseada.

\* **MEDIANA.** - Debe de estar en una clase tal que la frecuencia relativa acumulada hasta la clase que la precede sea menor o igual al 50% (ó 0.5) y la frecuencia que está después tenga 0.5 o más. La mediana de una colección de datos ordenados en orden de magnitud es el valor medio o la media aritmética de los dos valores medios. (13)

Para datos agrupados la mediana se puede obtener de diferentes maneras, aquí se muestran dos ecuaciones las cuales se consideran las más sencillas:

$$\text{Mediana} = L_1 + \left( \frac{N/2 - (\sum f)_1}{f_{\text{mediana}}} \right) C \dots (18)$$

Donde:  $L_1$  = limite real inferior de la clase mediana (es decir la clase que contienen la mediana)

$N$  = número total de datos (es decir la frecuencia total)

$(\sum f)_1$  = suma de las frecuencias de todas las clases por debajo de la clase mediana

$f_{\text{mediana}}$  = frecuencia de la clase mediana

$C$  = tamaño del intervalo de la clase mediana

$$Me = a + \frac{(b - a)(0.5 - c)}{d} \dots (19)$$

Donde:  $a$  = limite inferior de la clase de la mediana

$b$  = limite superior de la clase de la mediana

$c$  = frecuencia relativa acumulada hasta la clase que precede a la clase de la mediana

$d$  = frecuencia relativa de la clase de la mediana

0.5 = corresponde al valor del 50%

Hay que recordar que es el valor central la mediana y que por lo tanto determina que antes de ella se encuentra el 50% de los valores, y que después de ella se encuentra el 50% restante. (1)

Ejemplo: sean los números 3, 4, 4, 5, 6, 8, 8, 8, 10 → tiene una mediana de 6

sean los números 5, 5, 7, 9, 11, 12, 15, 18 → su mediana será  $\frac{1}{2}(9+11) = 10$

\* **MODA.** - Es el valor medio de la clase cuya frecuencia tiene el valor numérico mayor, la cual recibe el nombre de *frecuencia modal*. Se puede decir que es la marca de clase de la clase modal. Es el dato que ocurre con mayor frecuencia. (5) Es la medida más útil para los datos en los que se presentan frecuentemente valores semejantes y con pocos valores distanciados de los anteriores. (7)

$$M_o = \frac{L_1 + L_2}{2} \dots (20)$$

\* **VARIANZA.**- La diferencia con datos no agrupados, es que en ella no se considera a la frecuencia, por lo tanto la fórmula para su cálculo se determina:

$$s^2 = \sum_{i=1}^n \frac{f_i(x_i - \bar{x})^2}{n-1} \dots(21)$$

La razón para dividir entre  $n-1$  en lugar de  $n$ , es que solo hay  $n-1$  desviaciones independientes  $x_i - \bar{x}$ . Debido a que siempre su suma es cero, el valor de cualquiera de ellas siempre es igual al negativo de la suma de las demás  $n-1$  desviaciones. (6)

\* **DESVIACIÓN ESTÁNDAR.**- Sigue siendo la raíz cuadrada de la varianza para tener las unidades originales. La desviación estándar y la varianza son medidas de variación absolutas, esto es, miden la cantidad real de variación presente en un conjunto de datos y dependen de la escala de medición. (6)

$$s = \sqrt{s^2} \dots(22)$$

\* **COEFICIENTE DE VARIACIÓN.**- Se calcula igual que en datos no agrupados. Nos sirve para comparar la variación en diferentes conjuntos de datos, es una medida de variación relativa el cual da la desviación estándar como un porcentaje de la media. (6)

$$C.V = \frac{S}{\bar{x}} \times 100 = \% \dots(23)$$

### "REGLAS GENERALES PARA FORMAR LAS DISTRIBUCIONES DE FRECUENCIAS"

Se sabe que la Distribución de frecuencia es una tabla que divide un conjunto de datos en un número de clases ( categorías ) apropiadas, mostrando también el número de elementos de cada clase. Los pasos a seguir para obtener una distribución de frecuencias son:

- 1.- Determinar el número mayor y menor entre los datos registrados y así encontrar el rango.
- 2.- Dividir el rango en un número conveniente de intervalos de clases del mismo tamaño, si ésto no es posible utilizar intervalo de clase de diferente tamaño, el número de intervalo de clase se toma generalmente entre 5 y 20 dependiendo de los datos. (6)
- 3.- Determinar el número de observaciones que caen dentro de cada intervalo de clase, es decir, encontrar las frecuencias de clase. (1)

A continuación se presenta un cuadro comparativo entre las fórmulas utilizadas para Datos Agrupados como para Datos No Agrupados.

CUADRO COMPARATIVO

MEDIDAS	DATOS NO AGRUPADOS	DATOS AGRUPADOS
Media	$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$	$\bar{x} = \sum_{i=1}^n \frac{f_i x_i}{n}$
Mediana	Me = valor central de datos ordenados	$Mediana = L_1 + \left( \frac{N/2 - (\sum f)_1}{f_{mediana}} \right) C$ $Me = a + \frac{(b-a)(0.5-c)}{d}$
Moda	Mo = número repetido con mayor frecuencia	$Mo = \frac{Li + Ls}{2}$
Media geométrica	$mg = \sqrt[n]{(x_1)(x_2)(x_3)...(x_n)}$	$mg = \sqrt[n]{(x_1)(x_2)(x_3)...(x_n)}$
Media armónica	$ma = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	
Media ponderada	$\bar{x} = \frac{\sum WX}{\sum W}$	
Cuartil		Rango de 1 <sup>er</sup> cuartil $(1/4)N+1/4$ Rango de 2 <sup>o</sup> cuartil $(1/2)N+1/2 = \text{media}$ Rango de 3 <sup>er</sup> cuartil $(3/4)N+1/4$ N = números ordenados de mayor a menor
Percentil		Rango percentil = $(\text{facum}/N)100$
Rango	R = valor mayor - Valor menor	R = Ls - Li
Varianza	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$	$s^2 = \frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{n-1}$
Desviación estándar	$s = \sqrt{s^2}$	$s = \sqrt{s^2}$
Coficiente de variación	$C.V = \frac{s}{\bar{x}} \times 100 = \%$	$C.V = \frac{s}{\bar{x}} \times 100 = \%$

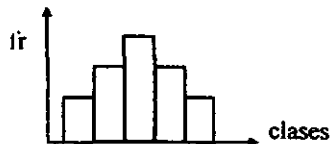
Por lo tanto la diferencia que existe entre ambos, es que para realizar los cálculos de datos agrupados siempre estará presente la frecuencia.

### 2.2.1 REPRESENTACIÓN GRÁFICA DE DATOS AGRUPADOS:

Con la información anterior se ha examinado alguno de los procedimientos empleados para hacer que adquiera sentido un conjunto de datos no organizados (datos agrupados); siendo este el primer paso para la organización de datos. El siguiente paso, por lo común, es presentar los datos en forma de dibujo de tal forma que el lector pueda percibir fácilmente los hechos esenciales de una distribución de frecuencias y compararlos con otras si se desea. Estos dibujos reciben el nombre de gráficas, y no deben considerarse como sustituto de un tratamiento estadístico de los datos, si no más bien, como ayuda visual para pensar e interpretar problemas estadísticos.

Existen diferentes tipos de gráficos, y son los siguientes:

a) **HISTOGRAMA.**- Es una representación gráfica de una tabla de frecuencias. Se construye por medio de rectángulos unidos cuyos anchos son los intervalos de clase que ellos representan y cuyas alturas representan las frecuencias o las frecuencias relativas. Sobre el eje horizontal (x) van los intervalos de clase y sobre el eje vertical (y) van las frecuencias de clase. (1)



b) **POLÍGONO DE FRECUENCIA.**- Es una forma geométrica obtenida de segmentos de recta que unen los puntos medios de intervalos de clase adyacentes en un histograma. Se cierra el polígono de frecuencias al prolongar los segmentos de recta de los extremos de tal forma que encuentren el eje horizontal en el punto medio de la clase hipotética siguiente.

Si consideramos que los intervalos de clase tienen la misma amplitud, y que las marcas de clase muestran el punto medio de ellos, por consiguiente el polígono de frecuencias unirá a las marcas de clase de cada intervalo.

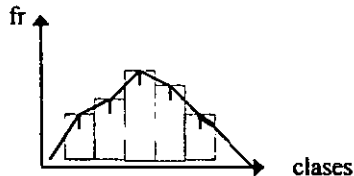
En el caso de trabajar con variables de tipo discreto, se tiene la necesidad de manejar los **límites reales de clase** que no son otra cosa que el promedio del límite superior de una clase más el límite inferior de la clase posterior entre dos, por lo general se habla de que al límite inferior de la primera clase se le resta la mitad de la unidad y al límite superior se le suma 0.5.

Sin embargo no es necesario elaborar un Histograma antes de construir un Polígono de frecuencias, todo lo que debe hacerse es marcar un punto en el sitio correspondiente a las cimas de las barras y luego unir esos puntos, proporciona una visión más clara

Los Polígonos de frecuencias pueden tomar un número ilimitado de formas diferentes, pero la forma más particular es la curva normal en "forma de campana". (5)

Se trabaja con datos continuos; 1) se trazan líneas que unan a los puntos medios de las clases, 2) suponemos en todos los cálculos que el punto medio de la clase es el promedio valor de todas las observaciones situadas en esta clase, 3) las clases extremas a ambos lados de los datos observados presentan una frecuencia cero, 4) el área bajo el polígono representa la totalidad de los datos observados, 5) si el número de clases incrementase

considerablemente y si el tamaño de la muestra fuese también considerablemente aumentado, el polígono se aproximará a una línea curva muy suave (curva normal de distribución). (7)



c) **CURVA SUAVIZADA.**- Es el proceso de "alisar" las esquinas del polígono de frecuencias de tal forma que se obtenga una curva suave, asemejando a una curva normal.

d) **ASIMETRÍA.**- Cuando la media aritmética es igual a la mediana y a la moda se habla de que tenemos una **simetría** en los datos, cuando la media aritmética es mayor a la mediana y a la moda hablamos de una **asimetría derecha** fácilmente identificable después de la representación gráfica de los datos, ya que se ve un histograma con cola derecha, ( la mayor densidad de los valores se encuentra de lado izquierdo). Cuando la media es menor a la mediana y a la moda, se tiene **asimetría izquierda**.

Se puede calcular el coeficiente de asimetría de la siguiente forma:

$$a_x = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \dots (24)$$

De esta forma, si el valor del coeficiente calculado es igual a cero, se tendrá un comportamiento simétrico, si el valor es positivo, las observaciones serán muy grandes con relación a la media; si es negativo, las observaciones serán pequeñas o menores a la media.

En el caso de tener distribuciones asimétricas se puede recomendar:

1. Reportar los resultados con el valor de la mediana, puesto que sólo representará a los valores centrales ordenados.
2. Utilizar la media podada o la media Windsor, ya que se consideran un "compromiso" entre la media y la mediana, (cuando se tienen distribuciones muy asimétricas o valores muy extremos).

e) **MEDIA PODADA.**- Se ordenan los datos de forma creciente, se elimina el 25% inferior de las observaciones y el 25% superior, y finalmente se calcula la media aritmética de las observaciones restantes.

f) **MEDIA WINDSOR.**- Se reemplaza el valor de cada observación en el 25% inferior de los datos por el valor de las observaciones no eliminados en el cálculo de la media podada. Se reemplaza en el 25% superior de los datos por el valor mayor de las observaciones no eliminados en el cálculo de la media podada. Calcular finalmente la media de todas las observaciones después de haber realizado la modificación.



### 2.3).- PROBLEMAS RESUELTOS:

(estadística descriptiva)

1) Los siguientes datos representan la cantidad de proteína (g) encontrados en 30 muestras de coliflor,

**Tabla 1**

(g/100g)				
1.5	2.6	1.4*	2.0	1.4
1.8	2.1	2.6	2.0	1.6
2.4	2.5	2.2	2.0	1.9
2.2	2.0	1.9	2.5	2.9*
2.1	2.3	2.0	2.2	2.4
2.2	2.3	1.7	2.2	1.6

El análisis rápido de los datos indica que el límite inferior de la muestra es 1.4 y el límite superior es 2.9.

Para distribuir las otras 28 observaciones que se encuentran cerca de 1.4 y cerca de 2.9, dividimos el intervalo en un número arbitrario de sub-intervalos de igual longitud, dependiendo de este número de sub-intervalos de la cantidad de datos disponibles (como regla general el número de sub-intervalos varía entre 5 y 20. A mayor cantidad de datos disponibles mayor el número de sub-intervalos).

Por ejemplo podríamos usar los sub-intervalos 1.35 a 1.55, 1.55 a 1.75 ..., note que los puntos extremos se han escogido de modo que sea imposible que una observación coincida con uno de los puntos extremos, eliminando así cualquier ambigüedad relacionada con la ubicación de una observación particular.

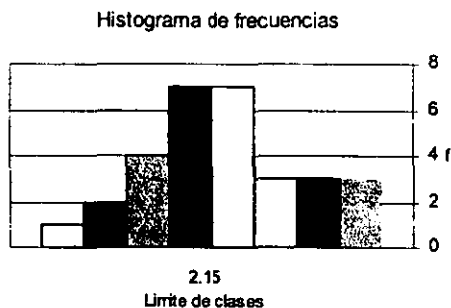
Estos sub-intervalos se llaman **clases** en el lenguaje estadístico forman celdas. Cada una de las 30 observaciones cae en una de las 8 clases a las que numeraremos para propósitos de identificación

**Tabla 2** Frecuencias Relativas

CLASE (i)	LIMITES DE CLASE	CONTEO	FRECUENCIA DE CLASE (fi)	FRECUENCIAS RELATIVAS (fi/n)
1	1.35-1.55		3	3/30
2	1.55-1.75		3	3/30
3	1.75-1.95		3	3/30
4	1.95-2.15	/	7	7/30
5	2.15-2.35	/	7	7/30
6	2.35-2.55		4	4/30
7	2.55-2.75		2	2/30
8	2.75-2.95		1	1/30
			$\sum n = 30$	1

El número de identificación aparece en la primera columna y los correspondientes límites de clase se dan en la 2a columna, la 3a se utiliza para conteo, colocando una marca en la clase correspondiente por cada observación que cae en dicha clase, es decir 3 de las 30 observaciones caen en la clase 1, 3 en la clase 2 etc., a este número de observaciones que caen en una clase particular se llama **Frecuencia de clases**, la última columna de esta tabla presenta la fracción total de observaciones que caen en cada clase.

La tabla resultante se puede representar graficando en la forma de un *Histograma de Frecuencias* sobre cada intervalo de clase se construye un rectángulo con altura proporcional al número de observaciones (Frecuencia de clase) que caen en dicho intervalo de clase.



Consideremos el Histograma de la muestra en mayor detalle.

1) Que fracción de muestras obtuvieron promedios iguales o mayores que 2.15 ?

Del Histograma de frecuencias vemos que la fracción incluirá todas las clases a la derecha de 2.15.

Usando la tabla 2 vemos que 14 muestras de proteína consiguieron promedios iguales o mayores que 2.15 por lo tanto la fracción es  $14 / 30$ , aproximadamente 47%. (3)

2) A continuación se presentan los contenidos netos (en gramos) de 65 sobres de Kool-Aid  
Tabla 1

g/ sobre												
7.13	7.36	6.71	6.58	7.25	6.71	6.46	7.28	7.79*	6.79	6.62	6.68	7.29
7.44	7.46	6.69	6.81	7.15	7.12	7.39	7.35	7.12	6.91	7.11	6.88	6.86
7.30	7.64	7.66	6.88	6.93	6.93	6.93	6.98	6.93	6.87	7.00	7.08	7.17
6.82	6.96	6.78	7.01	6.73	6.73	6.60	6.53	6.50	6.47	6.56	6.41	6.28
7.30	6.44	6.56	7.00	6.57	6.57	6.85	6.70	6.23*	6.63	6.32	6.55	7.29

- Agrupar en 7 clases
- Determinar frecuencias relativas de 1<sup>a</sup>, 3<sup>a</sup>, 5<sup>a</sup> y 7<sup>a</sup> clases.
- Determinar media aritmética, varianza y CV.
- Construir histograma y/o polígono de frecuencia.

Solución:

1° Para facilitar los cálculos se recomienda que los datos de la Tabla 1 se acomoden en forma ascendente

Tabla 2

g/ sobre

6.23	6.46	6.56	6.60	6.70	6.78	6.86	6.93	6.96	7.11	7.17	7.30	7.44
6.28	6.47	6.56	6.62	6.71	6.79	6.87	6.93	7.00	7.12	7.25	7.30	7.46
6.32	6.50	6.57	6.63	6.71	6.81	6.88	6.93	7.00	7.12	7.28	7.35	7.64
6.41	6.53	6.57	6.68	6.73	6.82	6.88	6.93	7.01	7.13	7.29	7.36	7.66
6.44	6.55	6.58	6.69	6.73	6.85	6.91	6.96	7.08	7.15	7.29	7.39	7.79

2° Primero buscamos la amplitud o rango, restando la observación máxima menos la mínima

$$R = Ls - Li$$

$$= 7.79 - 6.23 = 1.56$$

$$W = R/k = \text{Rango} / \text{N}^\circ \text{ de clases}$$

$$= 1.56 / 7 = 0.22285 \approx 0.23$$

Regla: Si  $R < Wk$  la amplitud que se utilizará para obtener los límites de clase será de tipo continuo

$$1.56 < (0.23)(7)$$

$$1.56 < 1.61 \rightarrow \text{se utilizara variable continua}$$

3° Se construye la Tabla de Distribución de frecuencias agrupando en 7 clases

Clases	Limites de Clases	fi	Xi	fi Xi	$f(X_i - \bar{x})^2$	fr
1	6.22 - 6.45	5	6.335	31.675	1.7041	0.076
2	6.45 - 6.68	13	6.565	85.345	0.1251	
3	6.68 - 6.91	15	6.795	101.925	0.0153	0.230
4	6.91 - 7.14	15	7.025	105.375	0.1127	
5	7.14 - 7.37	11	7.255	79.805	0.1130	0.169
6	7.37 - 7.60	3	7.485	22.455	0.3205	
7	7.60 - 7.83	3	7.715	23.145	0.6339	0.046
		$\Sigma n = 65$		$\Sigma = 449.725$	$\Sigma = 2.9388$	

Los datos de la columna (fi) son el número de datos que caen entre cada intervalo  
Ejemplo: Clase 1 = 6.22 - 6.45 tiene 5 frecuencias de clase que son las siguientes: 6.23, 6.28, 6.32, 6.41, 6.44

La columna Xi es la marca de clase  $X_i = Li + Ls / 2$

$$X_i = 6.22 + 6.45 / 2 = 6.335$$

b) Las frecuencias relativas se obtienen de  $f/n$

$$1^\circ - 5 / 65 = 7.69 \%$$

$$3^\circ - 15 / 65 = 23.07 \%$$

$$5^\circ - 11 / 65 = 16.92 \%$$

$$7^\circ - 3 / 65 = 4.61 \%$$
 Estas son las frecuencias relativas de cada clase

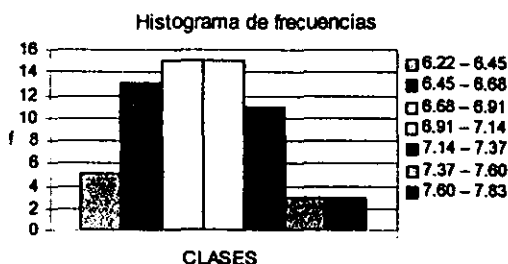
$$c) \quad \bar{x} = \frac{f_i x_i}{n} = \frac{449.725}{65} = 6.9188$$

$$S^2 = \frac{f_i (X_i - \bar{x})^2}{n-1} = \frac{2.9388}{64} = 0.04591$$

$$S = 0.2142$$

$$CV = \frac{S}{\bar{x}} \cdot 100 = 3.0971\%$$

d) Histograma de frecuencias



Como se puede ver en el histograma de frecuencias la mayor agrupación de los datos se tiene en la 3ª y 4ª clase, se tiene una dispersión entre los datos pequeña, siguiendo el comportamiento que tienen los datos en el gráfico observamos que es normal (forma de campana) pudiéndose usar una distribución normal para un análisis más completo.

3) Los siguientes datos se refieren al peso en gramos de racimos pequeños de 12 uvas variedad Conichon. Agrupar la información en 7 clases, determinar la media aritmética, desviación estándar y el coeficiente de variación. Determinar el intervalo con mayor frecuencia y la frecuencia relativa de la 5ª a la 7ª clases, construir histograma y polígono de frecuencias. Comenzar en el valor 7.47.

10.46	10.42	9.46	9.71	8.36	12.27
11.39	12.46	9.56	10.38	9.86	9.49
11.37	9.49	10.06	10.59	8.06	10.16
11.39	11.68	10.91	7.47*	10.66	8.58
9.72	9.74	9.32	8.96	11.49	10.86
10.20	7.99	11.79	9.85	11.18	
10.77	10.21	8.81	9.64	8.62	
9.67	10.08	10.16	11.18	9.89	
10.63	11.28	10.38	10.27	10.66	

Solución:

$$R = 12.46 - 7.48 = 4.98$$

$$W = 4.98 / 7 = 0.7114 \approx 0.71$$

$$R < Wk$$

$4.98 < 0.71 (7) = 4.97 \rightarrow$  como es menos 4.97 aumentamos la amplitud con el objeto de que todos los datos queden adentro

Por lo tanto tenemos  $0.72 (7) = 5.04$  y la amplitud a utilizar para los limites de clase será 0.72

Clases	Limite de clase	fi	Xi	fi Xi	fr	fra	$f(X_i - \bar{x})^2$
1	7.47 - 8.19	3	7.83	23.49	0.06	0.06	16.5280
2	8.19 - 8.91	4	8.55	34.20	0.08	0.14	10.5911
3	8.91 - 9.63	6	9.27	55.62	0.12	0.26	4.9381
4	9.63 - 10.35	15	9.99	149.85	0.30	0.56	0.5257
5	10.35 - 11.7	11	10.71	117.81	0.22	0.78	3.126
6	11.07 - 11.79	8	11.43	91.44	0.16	0.94	12.5561
7	11.79 - 12.51	3	12.15	36.45	0.06	1.00	11.6758
		n = 50		$\Sigma 508.86$			$\Sigma 59.9374$

$$\bar{x} = 508.86/50 = 10.1772 \text{ gr}$$

$$S^2 = 59.9374 / 49 = 1.2232 \text{ gr}^2 \quad S = 1.1060 \text{ gr}$$

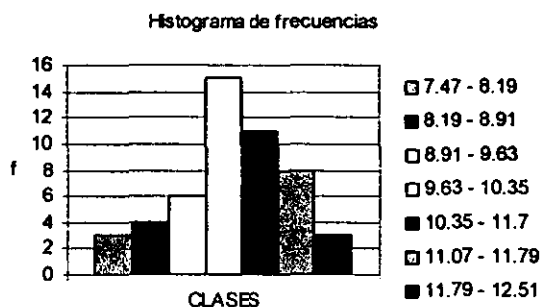
$$CV = (1.1060 / 10.1772) * 100 = 10.86 \%$$

$$Me = 9.4860$$

El intervalo con mayor frecuencia es la 4ª clase ya que cuenta con una agrupación de 15 datos

La frecuencia relativa de la 5ª a la 7ª clase es:  $0.22 + 0.16 + 0.06 \Rightarrow 44\%$

El histograma y polígono de frecuencias se muestra a continuación



INICIO 7.11

FINAL 12.

Como se puede ver en el gráfico los datos se encuentran en su mayoría acumulados a la derecha del gráfico por lo que se tiene un comportamiento de asimetría negativa (de cola izquierda prolongada), por lo que la media, moda y mediana se encuentran en diferentes puntos. También se observa que el coeficiente de variación es significativo, debido a una

dispersión que presentan los pesos de los racimos con respecto a la media de éstos mismos, indicando que hay racimos de diferentes pesos

4) Los siguientes datos muestran los porcentajes de proteína de algunas variedades de Arroz Palay

g / 100g					
1.3	1.8	2.0	2.2	2.7	3.6
1.4	1.8	2.0	2.3	2.9	3.8
1.4	1.9	2.1	2.4	2.9	4.1
1.4	1.9	2.1	2.4	3.0	4.5
1.7	1.9	2.1	2.5	3.2	4.7

a) Forme una distribución de frecuencias completa con amplitud de 0.8% de proteína comenzando en el valor 1.2.

b) Calcule  $\bar{x}$ , S y CV.

c) Dibuje histograma y polígono de frecuencias. Con variable discreta

Clases	Limite de clases	fi	Xi	fr(%)	fi Xi	$f(X_i - \bar{x})^2$
1	1.2-2.0	10	1.6	33.33	16.0	9.216
2	2.0-2.8	11	2.4	36.66	26.4	0.2816
3	2.8-3.6	4	3.2	13.33	12.8	1.6384
4	3.6-4.4	3	4.0	10.00	12.0	6.2208
5	4.4-5.2	2	4.8	6.66	9.6	10.0352
					$\Sigma$ 76.8	$\Sigma$ 27.392

**Solución:**

$$\bar{x} = 76.8 / 30 = 2.56 \text{ \% proteína}$$

$$S = 0.9718 \text{ \% proteína}$$

$$S^2 = 27.392 / 29 = 0.9445 \text{ (\% proteína)}^2$$

$$CV = 37.96 \text{ \%}$$

Histograma de frecuencias



INICIO POLÍGONO 1.6-0.8=0.8      FINAL POLÍGONO 4.8+0.8=5.6

5) La muestra siguiente es el conjunto de 16 mediciones en milímetros de los diámetros de zanahorias destinadas para el diseño de un preenfriador por aspersión. 52,62,68,74,74,74,76,76,78,78,82,84,88,88,92,96. Determinar las medidas de tendencia central y de dispersión para esta muestra y seleccionar cuál o cuáles recomendaría para dar mayor información sobre la muestra.

$X_i$	$(X_i - \bar{x})^2$
52	656.6406
62	244.1406
68	92.6406
74	13.1406
74	13.1406
74	13.1406
76	2.6406
76	2.6406
78	0.1406
78	0.1406
82	19.1406
84	40.6406
88	107.6406
88	107.6406
92	206.6406
96	337.6406
$\Sigma = 1242$	$\Sigma = 1857.7496$

**Solución:**

Medidas de tendencia central

$$\bar{x} = 1242 / 18 = 77.625 \text{ mm}$$

$$Me = 77 \text{ mm}$$

$$Mo = 74 \text{ mm}$$

Medidas de dispersión

$$R = 96 - 52 = 44 \text{ mm}$$

$$S^2 = 1857.7496 / 17 = 123.8499 \text{ mm}^2$$

$$S = 11.1287 \text{ mm}$$

$$CV = 14.3366 \%$$

Se recomendaría a la media aritmética ya que considera a todos los datos, la mediana ya que hay valores extremos de las medidas de dispersión y el coeficiente de variación.

6) En una prueba de penetración de calor en 35 alimentos, los cuales tenían diferente composición, se obtuvieron los siguientes valores de su coeficiente convectivo (faltan unidades)

6.72	6.77	6.82	6.70	6.78	6.70	6.62
6.75	6.66	6.66	6.64	6.76	6.73	6.80
6.72	6.76	6.76	6.68	6.66	6.62	6.72
6.76	6.70	6.78	6.76	6.67	6.70	6.72
6.74	6.81	6.79	6.78	6.66	6.76	6.72

- Obtener una tabla de frecuencia con 6 clases.
- Trace histograma y polígono de frecuencia.
- Determine media aritmética, mediana, moda, desviación estándar y coeficiente de variación.
- Que cantidad de alimentos se encuentran entre la 3ª y la 5ª clases?
- Determine el 2º cuartil y el percentil 50.

Solución:

$$R = 6.82 - 6.62 = 0.20$$

$$W = 0.20 / 6 = 0.033 \approx 0.03$$

$$WK > R$$

$0.03(6) < 0.20 \Rightarrow 0.18 < 0.20$  por lo tanto se aumenta el valor de la amplitud a  $W \approx 0.04$  para lograr que el conjunto de datos queden dentro del rango  $0.04(6) > 0.20 \Rightarrow 0.24 > 0.20$

	CLASES	f	Xi	fXi	f(Xi-x) <sup>2</sup>	fr	Fra
1	6.61-6.65	3	6.63	19.89	0.02828	8.5714	8.5714
2	6.65-6.69	6	6.67	40.02	0.1956	17.1428	25.7142
3	6.69-6.73	9	6.71	60.39	0.0026	25.7142	51.4284
4	6.73-6.77	9	6.75	60.75	0.0047	25.7142	77.1416
5	6.77-6.81	6	6.79	40.74	0.0237	17.1428	94.2854
6	6.81-6.85	2	6.83	13.66	0.0254	5.7140	99.9994
Σ		35		235.45	0.1043	100	



$$\bar{x} = 235.45 / 35 = 6.7271 \quad S^2 = 0.0030701 \quad S = 0.055 \quad CV = 0.8236$$

La cantidad de alimento presente entre la 3ª y 5ª clase es la siguiente

$$= 25.7142 + 25.7142 + 17.1428 = 68.5712$$

100 % → 35 alimentos

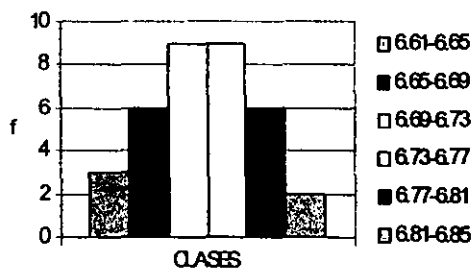
68.5712 → X ⇒ X = 24 alimentos

$$Me = 6.69 + ((6.73-6.69)(0.5-0.257142)) / 25.7142 = 6.7277$$

$$Mo = 6.71$$

2º cuartil = Me    Percentil 50 = Me

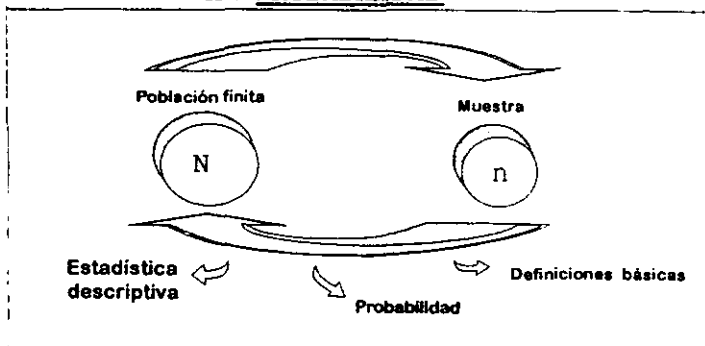
Histograma de frecuencias



INICIO 6.89    FINAL 6.87

Como se observa en el gráfico, éste presenta un comportamiento normal ya que tiene una distribución simétrica donde los valores de la media, mediana y la moda coinciden en un mismo punto. La media no es influenciada por los valores extremos ya que la variación de los datos es mínima. Las clases donde se encuentran la mayor cantidad de datos es la tercera y cuarta clase.

### III.- PROBABILIDAD



La probabilidad es el vehículo que permite usar la información contenida en una muestra para hacer inferencias o para describir la población de la cual se ha obtenido la muestra, está relacionada con el azar y con el grado de incertidumbre, y se puede definir o entender como los modelos que describen la posibilidad de que algunos resultados ocurran en una muestra tomada al azar, y esto es semejante a una tabla de frecuencias revisada en Estadística descriptiva. Se puede considerar tres tipos de probabilidades:

1) La del camino clásico: No se sabe que va a ocurrir, por ejemplo si se tiene un dado no cargado, puede salir en una tirada cualquiera de sus seis caras; por lo tanto cada cara tiene la misma oportunidad de caer.

$$n(A) / \text{total} \approx \text{Predicción}$$

2) La de la frecuencia relativa: Cuantificación, ya se sabe que va a pasar, por ejemplo cuando el dado está "cargado" se cuenta, en un número "x" de repeticiones (o sea un gran número de eventos) la cantidad de veces que se repite una misma cara, por lo tanto se puede construir una tabla de frecuencias relativas.

3) La probabilidad subjetiva: Esta es por un posible juicio de antecedentes.

La forma en que se puede expresar matemáticamente es:

$$P(A) = \frac{n(A)}{S}$$

Donde:  $P(A)$  es la probabilidad de que ocurra el evento,  $(A)$  es igual a la cantidad de veces que sucede  $(A)$  entre su espacio muestral. Al conjunto de todos los resultados posibles de un experimento estadístico se le llama Espacio muestral y se representa con la letra  $S$ .<sup>(1)</sup>

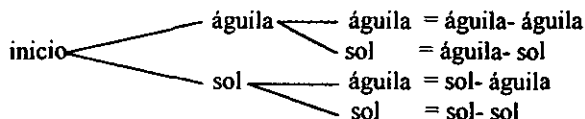
Si  $S$  contiene solamente un número finito de puntos, entonces a cada punto puede asociarse un número no negativo llamado *probabilidad* tal que la suma de todos los números correspondientes a todos los puntos de  $S$  sumen 1. <sup>(1)</sup>

El espacio muestral enlista todos los posibles sucesos de un experimento.

El **Experimento** es un proceso definido y repetido cuyos sucesos están determinados por el azar, depende del tamaño de la muestra o del tamaño de la población y si las muestras son tomadas con o sin reemplazo.

El espacio muestral es solo para experimentos que tienen un número finito de sucesos posibles; se puede hacer un listado de lo que contiene éste espacio muestral por ejemplo:

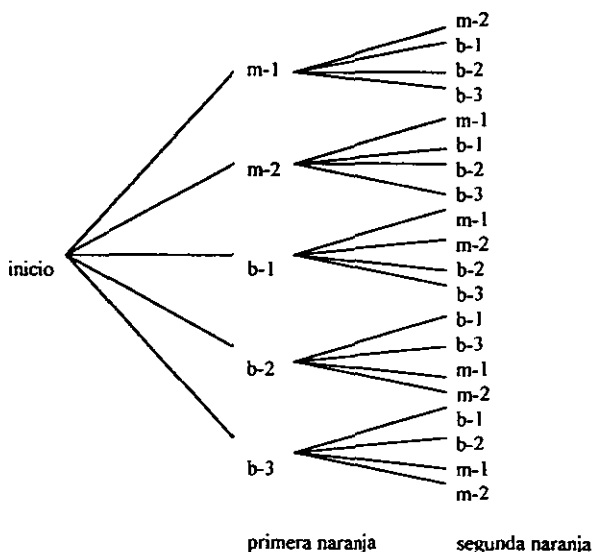
- si se lanza una moneda una vez, se tiene un suceso  $S = \text{sol-águila}$
- pero si se lanza dos veces:  $S = \text{águila-sol, águila-águila, sol-águila y sol-sol}$
- esto se puede hacer en un **diagrama de árbol**



Por lo tanto se van a tener eventos simples y compuestos. Un evento es un fenómeno observable que forma parte de un experimento. el evento simple solo va a tener un resultado mientras que los eventos compuestos son un conjunto de eventos simples, existen diferentes tipos de eventos y son los siguientes:

- a) Eventos independientes: El primer evento no se relaciona con el que sigue
- b) Eventos imposibles: Son aquellos eventos que no van a suceder ( $\Phi$ )
- c) Eventos seguros: Son eventos que sí suceden ( $S$ )
- d) Eventos mutuamente excluyentes: No se pueden tener dos cosas diferentes al mismo tiempo.
- e) Eventos con igual probabilidad: Son los que tienen la misma probabilidad de ser elegidos

Ejemplo de diagrama de árbol: Tengo 5 naranjas, 2 de ellas están malas y tres buenas, si escojo 2 para comerlas como se puede conformar el diagrama de árbol?



$$S = m-1, m-2; m-2, m-1; b-1, m-1; b-2, b-1; b-3, b-1; \\ m-1, b-1; m-2, b-1; b-1, m-2; b-2, b-3; b-3, b-2; \\ m-1, b-2; m-2, b-2; b-1, b-2; b-2, m-1; b-3, m-1; \\ m-1, b-3; m-2, b-3; b-1, b-3; b-2, m-2; b-3, m-2.$$

Los eventos compuestos son:

- a) Ambas naranjas escogidas son malas. Los eventos simples de estos compuestos son:  $m-1, m-2$  y  $m-2, m-1$ .
- b) Solo una de las naranjas es mala, los eventos simples de este compuesto son:  $m-1, b-2; m-1, b-2; m-1, b-3; m-2, b-1; m-2, b-2; m-2, b-3; b-1, m-1; b-1, m-2; b-2, m-1; b-2, m-2; b-3, m-1; b-3, m-2$ .
- c) Ninguna de las naranjas es mala:  $b-1, b-2; b-1, b-3; b-2, b-1; b-2, b-3; b-3, b-1; b-3, b-2$ .  
El total de eventos=20

### 3.1).- REGLAS DE CONTEO.

En general el espacio muestral  $S$  correspondiente a juegos de azar, a métodos de muestreo o a problemas de orden y ocupación es finito, y se asigna la probabilidad a todos sus puntos, sabemos que la probabilidad es  $1/m$  donde  $m$  es el número de puntos de  $S$ . (4)

Dado que en muchas situaciones no es obvia la determinación del número total de resultados posibles de un experimento, se han desarrollado algunas reglas para calcularlo, las cuales reciben el nombre de *Reglas de conteo* y son las siguientes. (8)

- **PERMUTACION.-** Se entiende por permutación diferentes “arreglos” u ordenaciones que se pueden realizar con “ $n$ ” elementos, tomando todos a la vez o bien parte de estos, pero no se pueden repetir los mismos elementos:

$${}_nP_r = \frac{n!}{(n-r)!}$$

Ejemplo: Supongamos que tenemos 3 latas diferentes  $L_1, L_2$  y  $L_3$ . ¿ De cuantas formas se pueden colocar las latas en un estante, tomando 2 cada vez, enumeramos las posibilidades colocando todas las combinaciones de dos latas en la primera columna y un reordenamiento de cada una de ellas en la segunda

Combinaciones de 2	Reordenamiento de combinaciones
$L_1 L_2$	$L_2 L_1$
$L_1 L_3$	$L_3 L_1$
$L_2 L_3$	$L_3 L_2$

El número de permutaciones es 6, resultado que se puede obtener usando la regla mn, la primer lata se puede seleccionar de  $m = 3$  formas distintas y la segunda lata se puede escoger de  $n = 2$  formas distintas, por lo tanto  $mn = 6$ . (3)

- **ARREGLO.-** Número de “permutaciones” que se pueden formar con “ $n$ ” elementos tomando “ $r$ ” elementos a la vez, el número de arreglos es igual al número de permutaciones:

$${}_nA_r = n^r$$

- **COMBINACIÓN.-** Número de subconjuntos que se pueden formar con “n” elementos tomando “r” elementos a la vez sin tomar en consideración el orden de los elementos. (1)

$$nC_r = \frac{n!}{(n-r)!r!}$$

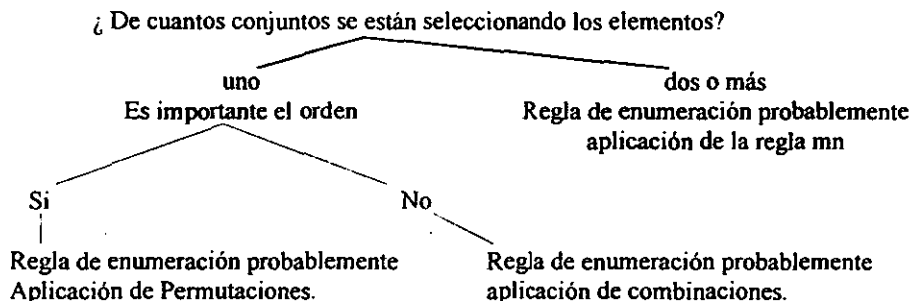
Una combinación de cosas dadas es cualquier selección de una ó más de las cosas sin considerar el orden de los elementos, Hay dos tipos:

- 1.- El número de combinaciones de n cosas diferentes tomando k a la vez **con** repeticiones es el número de conjuntos que se pueden formar de k cosas escogidas de las n dadas usándose cada una de ellas tantas veces como se quiera.
- 2.- El número de combinaciones de n cosas diferentes tomando k a la vez **sin** repeticiones es el número de conjuntos que se pueden formar con las n cosas, conteniendo cada conjunto k cosas diferentes y sin que dos conjuntos contengan exactamente las mismas k cosas. (4)

#### Sugerencias para la solución de Problemas:

Muchos estudiantes tienen dificultad para decidir cual de las tres reglas se debe utilizar en un problema dado. Las siguientes sugerencias pueden ser útiles:

- 1.- Analizar si se forma un suceso simple al:
  - (a) Seleccionar elementos de dos ó más conjuntos (situación que sugiere el uso de la regla mn)
  - (b) o al seleccionar “r” elementos de un solo conjunto de n elementos (situación que sugiere el uso de permutaciones o combinaciones)
- 2.- Si la situación es 1(b) hay que decidir si usar combinaciones o permutaciones, si cada ordenamiento diferente de los “r” elementos de un grupo conduce a un suceso simple diferente entonces se debe usar permutaciones. Si el ordenamiento no produce un nuevo suceso simple, entonces se deben usar combinaciones. Así el proceso de diagnóstico debe seguir las indicaciones del árbol de decisión mostrado en el diagrama siguiente:



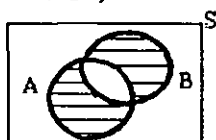
- 3.- Si se tiene problema para visualizar la regla de enumeración apropiada para un problema que involucra un gran número de puntos muestrales, se puede construir una versión en miniatura del problema de tal forma que se puedan contar manualmente los puntos, esto puede ayudar a ver cómo se puede resolver la versión más complicada. (3)

### 3.2).- OPERACIONES BÁSICAS CON CONJUNTOS

Ya que hablamos de espacios muestrales podemos hablar de los conjuntos con respecto a la probabilidad, por lo tanto, un conjunto es una colección de elementos, el conjunto universal es el que contiene a todos los elementos en discusión. (1) La mayoría de los sucesos de interés en situaciones prácticas son sucesos compuestos que requieren la enumeración de un gran número de puntos muestrales; los sucesos compuestos como el nombre sugiere se forman por alguna composición de dos o más sucesos. La composición tiene lugar en una de dos formas o en alguna combinación de éstas, como son una unión o una intersección. (3)

\*Definición de unión:

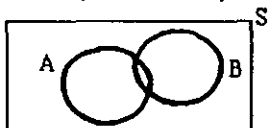
$$\text{UNIÓN: } \cup \Rightarrow A \cup B = \{x \in A, o, x \in B\}$$



"Sean A y B dos sucesos en un espacio muestral S. La unión de A y B se define como el suceso que contiene todos los puntos muestrales que están en A en B o en ambos y ocurre si y solo si ocurre A, ocurre B o ambos". (4)

\*Definición de intersección:

$$\text{INTERSECCIÓN: } \cap \Rightarrow A \cap B = \{x \in A, y, x \in B\}$$

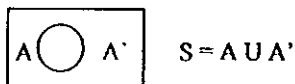


"Sean A y B dos sucesos en un espacio muestral. La intersección de A y B es el suceso compuesto por todos los puntos muestrales que existen en A y B, que ocurre sí, y solo sí, ocurren A y B simultáneamente, en la misma ejecución del experimento en consideración". (4)

Pero existen también relaciones entre sucesos como son los sucesos complementarios, independientes y mutuamente excluyentes, Habrá muchas ocasiones en las que desearemos si dos o más sucesos tienen entre sí alguna relación particular, la forma de saberlo está inherente en la definición de cada relación.

\* Definición de complemento:

$$\text{COMPLEMENTO: } A' = \{x \in S, x \notin A\}$$



$$S = A \cup A'$$

"El complemento de un suceso A es la colección de todos los puntos muestrales en S que no estén en A". (3)

\* MUTUAMENTE EXCLUYENTE.  $A \cap B = \emptyset$  Si no es posible que A y B ocurran simultáneamente. (4)

Un símbolo especial  $\emptyset$  se emplea a veces para indicar la ausencia de puntos en un sistema o conjunto llamado *sistema nulo o conjunto vacío*. Es decir  $P\{\emptyset\} = 0$ . (11) O bien si A y B no tienen elementos en común. (11)

Un procedimiento útil para la ilustración de las propiedades del álgebra de los conjuntos es el diagrama de Venn.

### 3.3).- AXIOMAS DE PROBABILIDAD. (de Kolmogorov)

La gran mayoría de los experimentos aleatorios que tienen importancia práctica exhiben su estabilidad, por esto podemos sospechar que prácticamente será cierto que la frecuencia relativa de un evento A en un gran número de ejecuciones de tal experimento aleatorio es aproximadamente igual a un determinado número  $P(A)$ .

Por esta razón se supone la existencia de un número  $P(A)$  que se llamara la probabilidad de E en el experimento aleatorio, entonces el decir "A tiene la probabilidad  $P(A)$ " significa que si efectuamos el experimento muy a menudo es prácticamente cierto que la frecuencia relativa es aproximadamente igual a  $P(A)$ .

Sin embargo después de establecer una forma de determinar la probabilidad experimentalmente, se pueden deducir leyes o propiedades de la probabilidad en forma lógica bajo ciertas suposiciones llamados *axiomas de la probabilidad*:

AXIOMA 1.-  $0 = P(A) = 1$

La probabilidad de que el evento A ocurra, se encuentra entre cero (imposible de ocurrir) y uno (siempre ocurre).

AXIOMA 2.-  $P(S) = 1$

La probabilidad del espacio muestral siempre será igual a uno. Asumiendo que la suma de las probabilidades individuales será igual a uno.

AXIOMA 3.-  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Pero si son eventos mutuamente excluyentes A y B entonces:

$$P(A \cup B) = P(A) + P(B).$$

### 3.4).- PROBABILIDAD CONDICIONAL.

Esta probabilidad sucede cuando dos sucesos se relacionan frecuentemente de manera tal que la probabilidad de la ocurrencia de uno depende si el segundo ha ocurrido o no. La probabilidad condicional de A dado de B, ha ocurrido, se denota en la forma  $P(A/B)$  donde la barra vertical dentro del paréntesis se lee "dado" y los sucesos que aparecen a la derecha de la barra son los sucesos que han ocurrido.

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad \text{donde:}$$

- a)  $P(A \cap B) = P(A) * P(B/A)$ . La probabilidad de que ocurran ambos eventos  $A$  y  $B$  es igual a la probabilidad de que ocurra  $A$  multiplicada por la probabilidad de que  $B$  ocurra a condición que  $A$  ocurra. Los símbolos  $P(B/A)$  y  $P(A/B)$  reciben el nombre de probabilidades condicionales, este término toma en cuenta la posibilidad de que la probabilidad de  $B$  pueda depender del hecho de que el evento  $A$  haya o no ocurrido, e inversamente para  $P(A/B)$  (regla de la multiplicación). (5)
- b)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  Si  $A$  y  $B$  son sucesos, la probabilidad de obtener cualquiera de los dos es igual a la probabilidad de  $A$  más la probabilidad de  $B$  menos la probabilidad de su ocurrencia conjunta. Si los eventos  $A$  y  $B$  son mutuamente exclusivos es  $P(A \cup B) = P(A) + P(B)$  (regla de la suma). (5)
- c)  $P(A \cap B) = P(A) * P(B)$ . Son independientes si  $P(A/B)$  es igual a  $P(A)$  o  $P(B/A)$  es igual a  $P(B)$ . Que el valor de la probabilidad de que  $A$  ocurra no es influido por la ocurrencia previa de  $B$ . Si esto no ocurre decimos que los sucesos son dependientes. (3)

Esto es: cuando son eventos independientes, que es una propiedad de las probabilidades y no de los eventos (si la ocurrencia de uno de ellos no influye sobre la probabilidad de ocurrencia de otro):

- Decimos que dos sucesos  $A$  y  $B$  son mutuamente excluyentes si el suceso  $AB$  no contiene ningún punto muestral  $P(A/B) = 0$ .
- Cuando son independientes  $P(A/B) = P(A)$ .

Dos o más sucesos se llaman Mutuamente excluyentes si la ocurrencia de uno o cualquiera de ellos imposibilita la ocurrencia de los otros  $P\{A+B\} = P\{A\} + P\{B\}$ . (1)

### LEYES DE PROBABILIDAD:

1.- Ley aditiva de Probabilidad: La probabilidad de una unión ( $A \cup B$ ) es igual a:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

2.- Ley multiplicativa de Probabilidad: Dados dos sucesos  $A$  y  $B$  la probabilidad de la intersección  $AB$  es:

$$\begin{aligned} P(AB) &= P(A) P(B/A) \\ &= P(B) P(A/B) \end{aligned}$$

Si  $A$  y  $B$  son independientes entonces  $P(AB) = P(A) * P(B)$

A continuación se presenta un cuadro resumen de todas las formulas utilizadas dentro del tema de probabilidad



Cuadro resumen de Formulas

Tabla 3.4.1

Permutación	${}_n P_r = \frac{n!}{(n-r)!}$	
Arreglo	${}_n A_r = n^r$	
Combinación	${}_n C_r = \frac{n!}{(n-r)!r!}$	
P (A)	$P(A) = \frac{n(A)}{S}$	Probabilidad subjetiva
P (B/A)	$P(B/A) = \frac{P(B \cap A)}{P(A)}$	Probabilidad Condicional
P (A∩B)	$P(A \cap B) = P(A) * P(B/A)$ $P(B \cap A) = P(B) * P(A/B)$	Regla de Multiplicación
P (A∪B)	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Probabilidad de Adición (regla de la suma)
P (AB) ó P (A∩B)	$P(A \cap B) = P(A) * P(B)$	Regla de Multiplicación cuando son independientes
P (A∪B) P (A∩B)	$P(A) + P(B)$	Regla de la suma cuando son independientes (mutuamente excluyentes)
P (A')	$1 - P(A)$	Probabilidad de Complemento

### 3.5).- PROBLEMAS RESUELTOS:

1) Los envases de refrescos de cierta compañía refresquera se encuentran separados en tres divisiones sucias, limpias, y rotas. La siguiente tabla indica el número de envases en cada división clasificados por tamaño.

División	Grandes	Chicos	$\Sigma$
Sucios	20	30	50
Limpios	60	140	200
Rotos	100	50	150
$\Sigma$	180	220	400

Si se elige aleatoriamente un envase; determinar las siguientes probabilidades:

- Que sea el envase grande
- Que el envase esté roto
- Que el envase sea chico y esté sucio (es una intersección)
- Que estén limpios si es que son envases grandes (es una condición)
- Que sea grande si el envase está limpio (es condicional)
- 220/240 es la probabilidad de que evento?

#### Solución:

Primero es necesario sacar las sumatorias tanto horizontales como verticales y ambas tienen que ser iguales, como en este caso la  $\Sigma$  es igual a 400 para ambos casos y este será el Espacio Muestral (S). A partir de este momento se puede hacer uso de las formulas para probabilidad (Tabla 3.4.1)

- $P(A) = n(A) / S = 180 / 400 = 0.45 \approx 45\%$  por lo tanto hay el 45% de probabilidad de que en la muestra de 400 envases 1 de ellos sea grande.
- $P(A) = n(A) / S = 150 / 400 = 0.375$  por lo tanto el 37.5% del total de los envases están rotos.
- $P(A) = n(A) / S = 30 / 400 = 0.075 \approx 7.5\%$  por lo tanto 30 envases de 400 son chicos y están sucios.
- $P(B/A) = P(A \cap B) / P(A) = (60 / 400) / (180 / 400) = 60 / 180 = 0.333 \approx 33.3\%$  por lo tanto de 180 envases grandes solo el 33.3% están limpios.
- $P(A \cap B) / P(B) = (60 / 400) / (200 / 400) = 60 / 200 = 0.3 \approx$  el 30% de los envases son grandes y están limpios.
- La probabilidad de este evento no existe ya que 220 es el total de envases chicos y 240 no existe.

2) Los productos alimenticios nuevos son sometidos a pruebas de sabor mediante un número de panelistas o jueces que asignen su preferencia al sabor sobre una escala nominal. Supóngase que se presentan cuatro nuevos productos de una bebida en polvo que saldrá próximamente al mercado. La siguiente tabla muestra la preferencia de los nuevos productos sometidos a ésta prueba. Los cuales fueron clasificados según el sabor.

Producto	Muy malo	Malo	Regular	Bueno	Muy bueno	$\Sigma$
Naranja	2	1	3	7	10	23
Toronja	3	3	5	9	12	32
Durazno	2	4	6	10	13	35
Manzana	4	5	7	11	14	41
$\Sigma$	11	13	21	37	49	131

Determinar las siguientes probabilidades:

- Probabilidad de seleccionar aleatoriamente un producto de buen sabor.
- Producto de durazno y que su sabor sea bueno. (intersección)
- Producto de manzana o de regular sabor. (unión)
- Producto de durazno (ó toronja) con sabor regular. (unión)
- De sabor bueno o muy bueno. (unión)
- De muy mal sabor si sabemos que es de naranja. (condicional)
- De durazno si sabemos que es bueno (ó malo). (condicional)

Solución:

A)  $P(A) = n(A) / S = 37 / 131 = 0.284 \approx 28.4\%$ , con lo que existe el 28.4% de probabilidad de que en una muestra de 131 productos algunos de ellos sean de buen sabor.

B)  $P(A) = n(A) / S = 10 / 35 = 0.285 \approx 28.5\%$ , o sea, el 28.5% de productos considerados buenos son de durazno.

C)  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 41 / 131 + 21 / 131 - 7 / 131$   
 $= 55 / 131 = 0.419 \approx 41.9\%$  son manzana o regular sabor.

D)  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 35 / 131 + 21 / 131 - 6 / 131$   
 $= 50 / 131 = 0.381 \approx 38.1\%$  son de durazno con regular sabor.  
 $= 32 / 131 + 21 / 131 - 5 / 131$   
 $= 48 / 131 = 0.366 \approx 36.6\%$  son de toronja con regular sabor.

E)  $P(A \cup B) = P(A) + P(B) - 0 = 37 / 131 + 49 / 131 - 0 = 86 / 131 = 0.656 \approx 65.6\%$  de 131 productos son de sabor bueno o muy bueno.

F)  $P(B/A) = P(A \cap B) / P(A) = (2 / 131) / (23 / 131) = 2 / 23 = 0.086$ , el 8.6% de 23 productos de naranja son de mal sabor.

G)  $P(B/A) = P(A \cap B) / P(A) = (10 / 131) / (37 / 131) = 10 / 37 = 0.270 \approx$  el 27% de 37 productos buenos son de durazno.

ó  $(4 / 10) / (13 / 131) = 4 / 13 = 0.307 \approx$  el 30.7% de 13 productos malos son de durazno.

3) Del cuadro que se presenta, calcular las probabilidades de:

	SABOR	DE	REFRESCO	
Tipo de Refresco	Naranja	Manzana	Tamarindo	$\Sigma$
Jarrito	20	30	25	75
Boing	24	26	19	69
$\Sigma$	44	56	44	144

- a) Boing de sabor manzana.
- b) Que sea jarrito.
- c) Que sea de sabor tamarindo o manzana.
- d) Jarrito y Naranja.
- e) Boing si es que es de naranja.
- f) De naranja si es que es Boing.
- g) Que sea Boing o de naranja.
- h) Jarrito o Boing.
- i) Jarrito y de sabor tamarindo.
- j) De sabor tamarindo si es que es Jarrito.

Solución:

- A)  $26 / 144 = 0.180 \approx$  por lo tanto el 18% es la probabilidad de tener un boing sabor manzana.
- B)  $75 / 144 = 0.52 \approx$  el 52% es la probabilidad de tener un jarrito sin importar el sabor.
- C)  $44 / 144 + 56 / 144 - 0 = 100 / 144 = 0.694 \approx$  el 69.4% es la probabilidad de tener un refresco sabor tamarindo o manzana sin importar que sea jarrito o boing.
- D)  $20 / 144 = 0.138 \approx$  13.8% es la probabilidad de tener un refresco jarrito y de sabor naranja.
- E)  $(24 / 144) / (44 / 144) = 24 / 44 = 0.545 \approx$  54.5% es la probabilidad de sacar un boing y que sea de naranja.
- F)  $(24 / 144) / (69 / 144) = 24 / 69 = 0.347 \approx$  por lo tanto el 34.7% es la probabilidad de sacar un refresco de sabor naranja y que sea boing.
- G)  $69 / 144 + 44 / 144 - 24 / 144 = 89 / 144 = 0.618 \approx$  61.8% es la probabilidad de tener un refresco que sea boing o que sea de sabor naranja.
- H)  $75 / 144 + 69 / 144 - 0 = 144 / 144 = 1 \approx$  el 100% es la probabilidad de tener un refresco boing o jarrito sin importar el sabor.
- I)  $25 / 144 = 0.173 \approx$  17.3% es la probabilidad de tener un jarrito y de sabor tamarindo.
- J)  $(25 / 144) / (75 / 144) = 25 / 75 = 0.333 \approx$  33.3% se tiene de probabilidad de tener un refresco de sabor tamarindo si es que es jarrito.

4) La tabla siguiente muestra las latas de jugo de frutas que se encuentran en un supermercado en la ciudad según su clasificación por defectos y la determinación de su tamaño real.

Defectos	Chica	Mediana	Grande	Dudoso	$\Sigma$
Etiqueta maltratada	35	42	57	60	194
Golpeada	12	8	6	6	32
Oxidada	1	3	62	35	101
Otros	8	8	7	6	29
$\Sigma$	56	61	132	107	356

- a) Probabilidad de que una lata elegida al azar sea de un tamaño mediano.
- b) Una lata con tamaño dudoso con un defecto de oxidación.
- c) Con etiqueta maltratada con tamaño grande.
- d) De tamaño chico con etiqueta maltratada.
- e) Tamaño grande si sabemos que tiene golpes

Solución:

- A)  $61 / 356 = 0.171 \approx 17.1\%$  es la probabilidad de elegir una lata de tamaño mediano sin importar sus defectos.
- B)  $35 / 356 = 0.098 \approx 9.8\%$  es la probabilidad de elegir una lata de tamaño dudoso.
- C)  $57 / 356 = 0.160 \approx 16\%$  es la probabilidad de elegir una lata grande con la etiqueta maltratada.
- D)  $35 / 356 = 0.098 \approx 9.8\%$  es la probabilidad de elegir una lata chica con la etiqueta maltratada.
- E)  $(6 / 356) / (32 / 356) = 6 / 32 = 0.187 \approx 18.7\%$  es la probabilidad de elegir una lata grande si sabemos que tiene golpes.

5) En un grupo de frutas cítricas se determinó que la distribución de los sabores era:

	FRUTAS		
Sabores	Naranja	Mandarina	$\Sigma$
Dulce	113	123	236
Amargo	103	103	206
Agrio	25	15	40
Sin sabor	10	15	25
$\Sigma$	251	256	507

Si se elige al azar una fruta de éste grupo, ¿Cuáles son las probabilidades de encontrar?:

- Sabor amargo.
- Sabor agrio.
- Que la fruta sea mandarina.
- Mandarina sin sabor.
- Que sea una naranja ó de sabor dulce.
- Que sea una naranja y sabor agrio.
- De sabor amargo ó sin sabor.
- Que sea una mandarina si sabemos que tiene sabor dulce.

Solución:

- A)  $206 / 507 = 0.406 \approx 40.6\%$  es la probabilidad de elegir una fruta amarga sin importar que sea naranja o mandarina.
- B)  $40 / 507 = 0.078 \approx 7.8\%$  es la probabilidad de elegir una fruta de sabor agrio.
- C)  $256 / 507 = 0.504 \approx 50.4\%$  es la probabilidad de elegir una mandarina sin importar su sabor.
- D)  $15 / 507 = 0.029 \approx 2.9\%$  es la probabilidad de elegir una fruta mandarina sin sabor.
- E)  $251 / 507 + 236 / 507 - 113 / 507 = 374 / 507 = 0.737 \approx 73.7\%$  es la probabilidad de elegir una naranja ó una fruta de sabor dulce.
- F)  $25 / 507 = 0.049 \approx 4.9\%$  es la probabilidad de elegir una naranja agria.
- G)  $206 / 507 + 25 / 507 - 0 = 231 / 507 = 0.455 \approx 45.5\%$  es la probabilidad de elegir una fruta amarga ó sin sabor.
- H)  $(123 / 507) / (256 / 507) = 123 / 256 = 0.480 \approx 48\%$  es la probabilidad de elegir una mandarina sabiendo que es dulce.

6) La probabilidad de que un cargamento de camarones programado en forma regular salga a tiempo es  $P(D) = 0.83$ ; la probabilidad de que llegue a tiempo es  $P(A) = 0.92$  y, la probabilidad de que salga y llegue a tiempo es  $P(D \cap A) = 0.78$ . Encuentre la probabilidad de que un camión:

- a) Llegue a tiempo dado que salió a tiempo.
- b) Haya salido a tiempo dado que llegó a tiempo.

Solución:

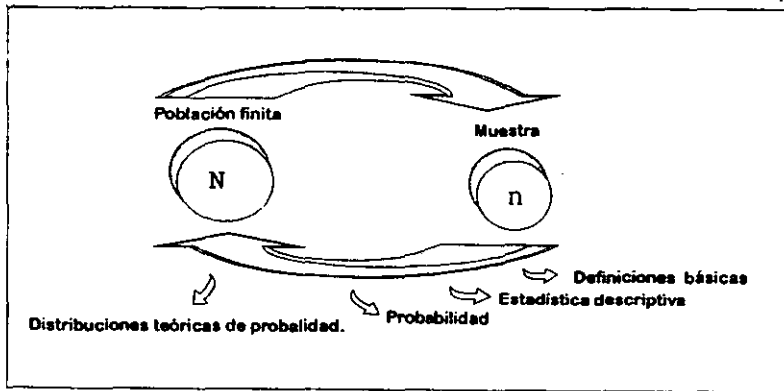
A)  $P(A/D) = P(D \cap A) / P(D) = 0.78 / 0.83 = 0.939$  por lo tanto el 93.9% es la probabilidad de que el cargamento de camarones llegue a tiempo.

B)  $P(D/A) = P(D \cap A) / P(A) = 0.78 / 0.92 = 0.847 \approx 84.7\%$  es la probabilidad de que el cargamento salga a tiempo.

#### IV.- DISTRIBUCIONES TEÓRICAS DE PROBABILIDAD

La Teoría de Probabilidad, se puede considerara como la teoría que tiene que ver con los posibles resultados de los experimentos, éstos deben ser potencialmente repetitivos, es decir, debemos ser capaz de reproducirlos bajo condiciones similares. (5)

La noción de la distribución de probabilidad surge al considerar un experimento aleatorio, existen dos tipos de distribuciones que son importantes en las aplicaciones prácticas y son las distribuciones discretas y continuas.(4)



#### 4.1.- DISTRIBUCIÓN PROBABILÍSTICA

Es la distribución de frecuencias relativas respecto a los resultados del espacio muestral y señala la proporción de veces que la variable aleatoria tiende a adoptar diversos valores, esta distribución se divide en variables aleatorias continuas y en variables aleatorias discretas.

Si arrojamamos dos dados, sabemos que la suma  $x$ , de los puntos que cae hacia arriba debe de ser un entero entre 2 y 12, pero no podemos predecir que valor de  $x$  aparecerá en el siguiente ensayo y podemos decir que  $x$  depende del azar. Si las observaciones no se dan en términos de números, podemos asignarles números y reducir las observaciones cualitativas al caso cuantitativo; así tenemos que la función que asigna números o valores a cada uno de los elementos del espacio muestra con una probabilidad definida, se denomina *variable aleatoria.*(11)

#### 4.1.1) DISTRIBUCIÓN DE VARIABLE CONTINUA

Si la variable aleatoria es continua, tiene un número incontable de valores posibles. La función de distribución acumulada será continua en el sentido de que su gráfica es suave es decir, no hay saltos en la probabilidad en cualquiera de los valores de la variable debido a la continuidad, la probabilidad de  $x$  en un punto cualquiera de sus valores posibles es cero por lo tanto no da resultados útiles.

En su lugar, representamos la función de  $x$ ,  $f(x)$  como una función densidad de probabilidades, por tanto es una medida de la concentración de probabilidad dentro de un intervalo. Esta probabilidad puede interpretarse como un área bajo la curva (una integral) de  $f(x)$ .(11)

Si graficamos las frecuencias ya que no se pueden tabular, tenemos la función de la densidad,  $f(x)$ , con cualquiera de las curvas que se muestran a continuación:



En donde  $x$  está definida sobre un espacio muestral continuo.<sup>(3)</sup> Existen diferentes tipos de distribuciones continuas y son las siguientes:

- 1.- Distribución Normal
- 2.- Distribución Gamma
- 3.- Distribución Exponencial
- 4.- Distribución Ji- cuadrada
- 5.- Distribución Weibull

### 1.- DISTRIBUCIÓN NORMAL

Sin lugar a duda el modelo de mayor uso de todas las distribuciones continuas es la distribución normal o llamada distribución Gaussiana (atribuida a K.F Gauss quien primero hizo referencia en el año de 1809 en relación con los errores de medidas físicas).

Si graficamos un polígono de frecuencias con una curva simétrica en donde "n" (tamaño de la muestra) sea grande, tendremos una agrupación alrededor del centro con una disminución gradual de los extremos, conocida como curva suave con una sola pendiente en cada punto. Consideramos que muchas de las características de la naturaleza se distribuyen de esta manera.

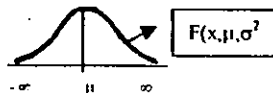
La distribución normal se define mediante su función densidad de probabilidad donde la función de la variable aleatoria  $x$  su media  $\mu$  y varianza  $\sigma^2$  es:

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Donde:  $\pi = 3.1416$ ,  $e = 2.71828$ ,  $\mu =$  media,  $\sigma =$  desviación típica

Esta fórmula tan aparentemente complicada, especialmente si deseamos calcular densidades y probabilidades con ella, puesto que su función de distribución acumulada es una integral que no se puede evaluar por métodos elementales, no debemos preocupar por que se puede calcular fácilmente usando las tablas que se han publicado para tal efecto (ver Apéndice Tabla de Distribución Normal).

La curva normal tiene forma de campana (campana de Gauss) con los extremos que tienden hacia el eje de las abscisas, los extremos de la curva son asíntoticos con respecto al eje de las  $x$ 's y no se tocan, se dirigen hacia un punto en el infinito.



Cuando la variable  $x$  viene expresada en unidades de desviación,  $z = (y - \mu) / \sigma$  la ecuación queda sustituida por la llamada *tipificada*

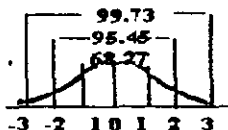
$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-1/2z^2}$$

En este caso se dice que  $z$  se distribuye normalmente con media cero y varianza uno, por lo tanto  $z = 0$  cuando  $y = \mu$ .



$Z$  = es positiva cuando  $y$  es mayor que la media  
 es negativa cuando  $y$  es menor que la media. (3)

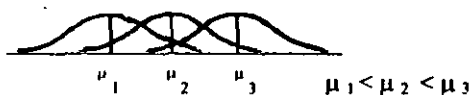
Un gráfico de esta curva normal tipificada se muestra en la figura siguiente, donde se han indicado las áreas incluidas entre  $z = -1y + 1$ ;  $z = -2y + 2$ ;  $z = -3y + 3$ ; que son respectivamente el 68.27%, 95.45% y el 99.73% del área total que vale uno. (1)



Distribución Normal donde  $\mu = 0$

### Características de la Distribución normal:

1. La media aritmética, la mediana y la moda son iguales. La media está representada por la perpendicular en el centro de la curva, como divide a la curva en dos partes iguales representa a la mediana y se observa que la densidad de frecuencia alcanza su valor máximo en este punto que representa a la moda.
2. El área total debajo de la curva por encima del eje de las  $x$ 's, es una unidad cuadrática, esto se deduce del hecho de que la distribución normal es una distribución de probabilidad.
3. Es simétrica a su media  $\mu$ , 50% está a la derecha y 50% a la izquierda
4. Entre la media y 1 desviación estándar por encima de la media se encuentra el 34.13% de todos los casos y análogamente por debajo de la media. Entre la media y 2 desviaciones estándar se encuentra el 47.72% de los casos, puesto que la curva normal es simétrica 47.72% de área también está comprendida entre la media y  $-2$  desviaciones estándar. Finalmente entre la media y 3 desviaciones estándar por encima de la media se encuentra el 49.87% de los casos y análogamente entre  $-3$  desviaciones estándar.
5. La distribución normal queda completamente determinada por los parámetros de  $\mu$  y  $\sigma$ , o sea, que hay una distribución normal específica para diferentes valores de  $\mu$  y  $\sigma$ . Los valores diferentes de  $\mu$  trasladan a la gráfica de distribución a lo largo del eje  $x$ . Los valores diferentes de  $\sigma$ , determinan el grado de alargamiento de la gráfica, esto se muestra a continuación:



6. Si se levantan perpendiculares del  $\sigma$  de la media en ambas direcciones, el área comprendida por estas perpendiculares será aproximadamente del 68% del área total. A  $2\sigma$ , el área es igual al 95% y con  $3\sigma$  a 99.7%



$$\sigma_1 < \sigma_2 < \sigma_3$$

La distribución normal es la distribución continua más importante por las siguientes razones:

1.- Muchas variables aleatorias que aparecen en relación con experimentos u observaciones prácticas están distribuidas normalmente

2.- Otras variables están distribuidas normalmente en forma aproximada

3.- Algunas veces una variable no está distribuida normalmente, ni siquiera en forma aproximada, pero se puede convertir en una variable con distribución normal por medio de una transformación sencilla.

4.- Ciertas distribuciones más complicadas se pueden aproximar mediante la distribución normal (como el caso de la distribución Binomial cuando  $n$  es grande y  $p$  no necesariamente pequeña y de Poisson)

5.- En estadística teórica, muchos problemas pueden ser resueltos fácilmente en el supuesto de una población normal. En el trabajo aplicado, encontramos que métodos elaborados según la ley de probabilidad normal dan resultados satisfactorios, aunque no se cumpla totalmente el supuesto de una población normal.

6.- Ciertas variables que son básicas para justificar pruebas estadísticas están distribuidas normalmente. (4)

### DETERMINACIÓN DEL ÁREA BAJO LA CURVA

Para determinar el área bajo la curva desde un punto  $x_1$  a un punto  $x_2$  y encontrar si:

- Probabilidad de que una "z" elegida al azar de la población contenga valores entre  $x_1$  y  $x_2$
- Frecuencia relativa de ocurrencia entre  $x_1$  y  $x_2$
- Proporción de valores entre  $x_1$  y  $x_2$
- Porcentaje de "z" con un valor entre  $x_1$  y  $x_2$

Es necesario resolver la ecuación que determina a una curva de distribución normal o bien, utilizar las tablas de Distribución Normal que se presentan en el Apéndice.

Para la utilización de la tabla debemos seguir los siguientes pasos:

- Transformar los valores de  $x$  a valores de  $z$  con la formula:  $z = \frac{(x - \mu)}{\sigma}$
- Consultar la tabla de distribución normal estándar para encontrar el valor de  $z$ . El número de la tabla que corresponde a  $z$ , es la probabilidad de encontrar valores entre  $\mu$  y  $x$ . (Nota: se pueden utilizar tablas de distribución normal con áreas que van de 0 a  $z$  y tablas que determinan áreas de  $-\infty$  a  $z$ )



Ejemplo: Se desea encontrar el área sombreada debajo de la curva

Por lo tanto es necesario usar la tabla de distribución normal.

La Tabla 1 es una versión abreviada de la Tabla 3 del libro Spiegel & Murray. Obsérvese que el valor de  $z$ , hasta el décimo más cercano se encuentra en la columna de la izquierda.

La segunda cifra decimal de  $z$  correspondiente a las centésimas se da en la fila superior. Para encontrar el área a  $z = .57$  desviaciones estándar a la derecha de la media, en la columna de la izquierda buscamos la fila que tiene el valor 5, luego en la fila superior buscamos la columna con el valor .07. La intersección de esta fila y columna nos da el área aproximada de .2157. Así el área entre la media y un punto situado a  $z = 0.7$  desviaciones estándar a la derecha, dada en la segunda columna de la tabla al lado de  $z = .7$  es igual a .2580. (3)

Tabla 1  
VERSIÓN ABREVIADA DE LA TABLA 3 APÉNDICE II

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	...09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	...	
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675		
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064		
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1440		
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1800		
0.5	.1915	.1950	.1985	.2010	.2054	.2088	.2123	.2157		
0.6	.2557									
0.7	.2580									
.										
1.0	.3413									
.										
2.0	.4772									

- 3) Si lo que se desea obtener es la probabilidad de valores inferiores a  $x$ , se determinará directamente multiplicando el valor de la tabla por 100



- 4) Ahora bien, si lo que se desea conocer es la probabilidad entre dos valores de  $x$ , al valor el área de  $x_2$  se le resta el valor de  $x_1$ .



Y el número encontrado es la probabilidad que hay entre  $\mu$  y  $x$

## 2.- DISTRIBUCIÓN GAMMA

Algunas densidades de probabilidad importante, son casos de esta distribución

$$f(x) = \begin{cases} \beta^\alpha \Gamma(\alpha)^{-1} x^{\alpha-1} e^{-x/\beta} \end{cases}$$

Para  $x > 0$ ,  $\alpha > 0$ ,  $\beta > 0$  en los demás puntos donde  $\Gamma(\alpha)$  es el valor de la función gamma, definido por:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

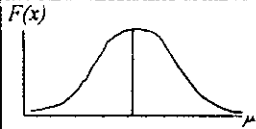
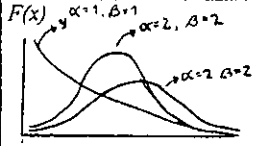
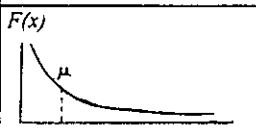
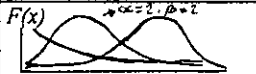
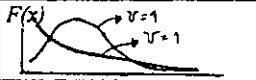

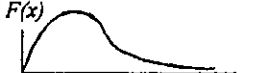
## 3.- DISTRIBUCIÓN Ji-CUADRADA

La distribución Ji-cuadrada fue introducida por F.R. Helmert en 1876 y constituye la base de varias pruebas estadísticas, sean  $x_1, x_2, \dots, x_n$  variables aleatorias normales independientes, cada una con media cero y varianza 1. La suma de cuadrados se representa por  $\delta^2$  (ji-cuadrada) al hablar de esta distribución el parámetro  $\gamma$  es llamado grados de libertad que es un entero positivo. (2)

$$f(u) = \frac{u^{(\gamma/2)-1} e^{-u/2}}{2^{\gamma/2} \Gamma(\gamma/2)}$$

4.1.2) CUADRO RESUMEN DE VARIABLES CONTINUAS  
DISTRIBUCIONES TEORICAS DE PROBABILIDAD

*Variables aleatorias continuas*

<b>Distribución</b>	<b>Ecuación que la define</b>	<b>Parámetros</b>	<b>Representación</b>	<b>Aplicaciones</b>
Normal	$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-1/2z^2}$	$\mu, \sigma$		<ul style="list-style-type: none"> <li>*Cálculo de probabilidades</li> <li>*Especificaciones en c.c</li> <li>*Planes de muestreo</li> <li>*Intervalos de confianza</li> <li>*Pruebas de hipótesis</li> </ul>
Gama	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} = x^{\alpha-1} e^{-x/\beta}$	$\mu = \alpha\beta$ $\sigma^2 = \alpha\beta^2$ $\alpha = \text{parámetro}$ $\beta = \text{parámetro de curva patrón}$		<ul style="list-style-type: none"> <li>*Tiempo de espera en llegadas consecutivas</li> <li>*Entrega de materia prima</li> <li>*Piezas que se rompen en 2 ciclos de espera con <math>f=2/100</math></li> <li>*<math>\beta = 2 = x^2</math></li> <li><math>\beta = 0, 1 = \text{exponencial}</math></li> </ul>
Exponencial	$f(x) = \frac{1}{\beta} e^{-x/\beta}$	$\mu = \beta$ $\sigma^2 = \beta^2$		<ul style="list-style-type: none"> <li>*Tiempos de espera y especificaciones</li> <li>*Confiabilidad de producto</li> <li>*Intervalos de confianza en diseño de productos</li> <li>*Tiempo de falla = <math>\mu</math>, probabilidad de que halla 2 fallas en 20 hr.</li> </ul>
Weibull	$f(x) = \frac{\alpha}{\beta} x^{\alpha-1} e^{-x^\alpha/\beta}$	$X = \text{parámetro}$		<ul style="list-style-type: none"> <li>*<math>\beta = 3.5, \alpha = 1, x = 1</math> dist. Normal</li> <li>*Durabilidad de productos</li> <li>*Para fallas</li> </ul>
Ji-cuadrada	$f(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1}$	$\mu = v$ $\sigma^2 = 2v$		<ul style="list-style-type: none"> <li>*Tiende a ser normal cuando n es grande</li> <li>*Inferencia estadística</li> <li>*Bondad de ajuste</li> </ul>
"t" Student	$f(x) = \frac{\Gamma((v+1)/2)}{\sqrt{v} \Gamma(v/2)} (1+(x^2/v))^{-(v+1)/2}$	$\mu = 0$ $\sigma^2 = 2v$		<ul style="list-style-type: none"> <li>*Teorema del limite central</li> <li>*Pruebas de hipótesis</li> <li>*Intervalos de confianza</li> </ul>
"F" Fisher	$f(x) = \frac{(v_1)^{v_1/2} (v_2)^{v_2/2} \Gamma(v_1/2 + v_2/2)}{\Gamma(v_1/2) \Gamma(v_2/2)} F$	$\mu = v_2/v_2 - 1$ $\sigma^2 = 2v_2(v_2 - 1)$		<ul style="list-style-type: none"> <li>*Diseño de experimentos</li> </ul>

#### 4.2.1).- DISTRIBUCIÓN DE VARIABLES ALEATORIAS DISCRETAS

Una distribución de variables discreta se identifica fácilmente examinando los valores que pueden asumir, si se pueden enumerar entonces debe ser discreta, ejemplo típico de variable aleatoria discreta: el número de latas defectuosas en una muestra de 10 extraída de una producción industrial, el número de fallas de una embotelladora en un periodo de tiempo etc.. (3)

Existen algunas distribuciones de variables aleatorias discretas que tienen gran aplicación práctica, entre ellas veremos las siguientes:

- 1.- Distribución Uniforme
- 2.- Distribución Binomial
- 3.- Distribución Multinomial
- 4.- Distribución de Poisson
- 5.- Distribución Geométrica
- 5.- Distribución Geométrica
- 6.- Distribución Hipergeométrica

Sea  $p$  la probabilidad de un evento  $A$  en un experimento aleatorio. Supóngase que el experimento se efectúa varias veces, digamos  $n$  ocasiones, entonces puede ser que estemos interesados en el número de veces que ocurren  $A$  en esos  $n$  ensayos. Esto nos conduce a la distribución Binomial si suponemos que  $p$  es constante y que los resultados de los ensayos no influyen uno sobre otro.

Si  $n$  es grande y  $p$  es pequeño, entonces la distribución puede aproximarse mediante la distribución de Poisson que es más sencilla. Mientras la distribución Binomial es importante en experimentos en que se extrae *con reemplazo*, la distribución Hipergeométrica interviene en forma semejante en experimentos en donde se extrae *sin reemplazo*, estos experimentos son un poco más complicados. (4)

##### 1.- DISTRIBUCIÓN UNIFORME

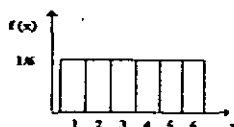
Una distribución de probabilidad de una variable discreta  $x$ , que toma valores  $x_1, x_2, \dots, x_n$  con igual probabilidad se llama distribución uniforme discreta, es decir su función de probabilidad está dada por:

$$f(x) = 1/n$$

La gráfica de una distribución uniforme es un conjunto de líneas con la misma altura y la media y la varianza son las siguientes:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}, \quad \sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Ejemplo: al lanzar un dado legal, cada elemento del espacio muestral  $S = \{1, 2, 3, 4, 5, 6\}$  ocurre con igual probabilidad, por lo tanto tenemos una distribución uniforme cuya función de probabilidad es  $f(x) = 1/6$  para toda  $x \in \{1, 2, 3, 4, 5, 6\}$  y su histograma es: (11)



## 2.- DISTRIBUCIÓN BINOMIAL

El modelo Binomial es de gran utilidad práctica, se puede considerar como la suma de  $n$  modelos de Bernoulli independientes (llamados así en honor a Jackes Bernoulli, matemático suizo que vivió en la última mitad del siglo XVII). (12)

La distribución Binomial ocurre si estamos interesados en el número de veces que sucede un evento  $A$  en  $n$  ejecuciones independientes de un experimento aleatorio, suponiendo que  $A$  tienen probabilidad  $P(A) = p$  en un solo ensayo. Entonces  $1 - p$  es la probabilidad de que un solo ensayo no ocurra  $A$ , hacemos  $q = 1 - p$ .

Considerando la variable aleatoria  $x =$  número de veces que ocurre el evento  $A$  y puede tomar valores de  $0$  ó  $1$  con probabilidad  $p$  y  $q$  ( $p + q = 1$ ), dependiendo respectivamente de que  $A$  no suceda o si suceda.

Si  $p$  es la probabilidad de ocurrencia de un suceso en un solo ensayo (llamada probabilidad de éxito) y  $q = 1 - p$  es la probabilidad de que el suceso no ocurra en un solo ensayo (llamada probabilidad de fallo o fracaso), entonces la probabilidad de que el suceso se presente exactamente  $x$  veces en  $N$  ensayos (es decir,  $x =$  éxitos y  $N - x =$  fallos) viene dada por: (4)

$$p(x) = NC_x p^x q^{N-x} = \frac{N!}{x!(N-x)!} p^x q^{N-x}$$

$$p(x) = (n, x) = p^x (1-p)^{n-x}$$

Donde:  $x = 0, 1, 2, \dots, N$

$N! = N(N-1)(N-2)$

$0! =$  por definición es  $1$

Un experimento Binomial debe cumplir con las siguientes propiedades:

- 1) El experimento consiste en  $n$  ensayos estadísticamente independientes y repetidos
- 2) Cada ensayo tiene dos resultados posibles. A uno de ellos lo llamaremos éxito ( $E$ ) y al otro fracaso ( $F$ ).
- 3) La probabilidad de éxito en un ensayo es igual a  $p$  y permanece igual en todos los ensayos. La probabilidad de fracaso es igual a  $q = (1 - p)$
- 4) Existe una variable aleatoria discreta  $x$ , asociada al experimento que cuenta el número de éxitos en los  $n$  ensayos ó el número de ensayos con éxito, entonces los valores de  $x$  son  $1, 2, 3, \dots, n$ .
- 5) Estamos interesados en el número de éxitos observados durante los  $n$  ensayos. (3)
- 6) Todos los experimentos deben contar con  $N =$  número de elementos en la muestra,  $P =$  probabilidad de éxito y  $x =$  variable discreta.
- 7) Con estos parámetros se puede obtener la probabilidad Binomial con ayuda de la tabla de Distribución acumulativa Binomial ó Distribución Binomial.

A continuación se presenta la manera de utilizar dicha tabla. La Tabla 2 es una versión abreviada de la tabla de Distribución acumulativa Binomial que se encuentra en el Apéndice, donde, en la fila superior aparecen las probabilidades de éxito ( $P$ ), en la primera columna aparece  $n =$  número de muestra y en la 2ª columna aparece la variable aleatoria  $x$ .

Ejemplo: para encontrar la probabilidad de una muestra dada con  $P = 0.15$ ,  $n = 4$  y  $x = 2$  se hace lo siguiente, en la fila superior buscamos el valor de  $0.15$ , localizamos en la primera columna  $n = 4$  con una  $x = 2$  y la intersección de esta fila con la columna de  $0.15$  nos da la probabilidad que se está buscando.

Tabla 2

n	x	P							
		.01	.05	.10	.15	...	... .40	.45	.50
2	0	.9801	.9025	.8100	.7225	.6306	.5325	.4375	.3500
	1	.9999	.9975	.9900	.9775	.8400	.7975	.7500	.7500
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0	.9703	.8574	.7290	.6141	.5160	.4364	.3664	.3025
	1	.9997	.9928	.9720	.9392	.8480	.7748	.7148	.6600
	2	1.000	.9999	.9990	.9966	.9360	.9089	.8750	.8500
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0	.9606	.8145	.6561	.5220	.4126	.3215	.2462	.1825
	1	.9994	.9860	.9477	.8905	.7752	.6910	.6225	.5675
	2	1.000	.9995	.9963	.9880	.8208	.7585	.7075	.6675
	3	1.000	1.000	.9999	.9995	.9744	.9590	.9375	.9225
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0	.9510	.7738	.5905	.4437	.3378	.2503	.1813	.1262
	1	.9990	.9774	.9185	.8352	.7370	.6262	.5000	.3775
	2	1.000	.9998	.9914	.9734	.8262	.7531	.6600	.5500
	3	1.000	1.000	.9995	.9978	.9130	.8688	.8125	.7625
	4	1.000	1.000	1.000	.9999	.9898	.9815	.9688	.9588
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0	.9415	.7351	.5314	.3771	.2667	.1877	.1256	.0812
	1	.9985	.9672	.8857	.7765	.6333	.4636	.3104	.1825
	2	1.000	.9978	.9842	.9527	.8443	.7415	.6438	.5500
	3	1.000	.9999	.9987	.9941	.8208	.7447	.6563	.5625
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
15	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
20	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

### 3.- DISTRIBUCIÓN MULTINOMIAL:

El modelo multinomial se refiere a ensayos independientes con más de dos resultados si los sucesos  $E_1, E_2, \dots, E_k$  pueden ocurrir con probabilidades  $p_1, p_2, \dots, p_k$  respectivamente, entonces la probabilidad de que  $E_1, E_2, \dots, E_k$  ocurran  $x_1, x_2, \dots, x_k$  veces respectivamente:

$$f(x) = \frac{N!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

Donde:  $x_1 + x_2 + \dots + x_k = N$

Esta distribución es una generalización de la distribución Binomial, se llama distribución Multinomial puesto que la ecuación es el término general del desarrollo multinomial  $(p_1 + p_2 + \dots + p_k)^N$ . (1)

Por ejemplo, un artículo de producción puede asignarse a una de las cuatro clases: buena, de pequeño defecto, de defecto grande y de defecto crítico, entonces mediante inspección los "n" artículos de la muestra se clasificarían en los 4 grupos; donde el número del primer grupo se indicaría con  $x_1$ , el número del segundo grupo con  $x_2$  y así sucesivamente. (2)

#### 4.- DISTRIBUCIÓN DE POISSON:

La distribución de Poisson es otra distribución discreta, cuyo nombre se debe al matemático francés Simeón Denis Poisson (1781-1840), esta distribución tiene aplicaciones importantes, de hecho es una aproximación conveniente de la distribución Binomial en casos en donde existe un gran número de  $n$  ensayos y una probabilidad pequeña de éxito ( $p$ ) en un solo ensayo. (4)

$$P(x) = \frac{x^e e^{-\lambda}}{x!}$$

Donde:  $X = 0, 1, 2, \dots, n$

$e = 2.71828$  es una constante

$\lambda$  (lambda) = es el parámetro de esta distribución y es el promedio de ocurrencias del evento aleatorio en el intervalo.

Un experimento de Poisson debe cumplir con las siguientes condiciones:

- 1) Los eventos ocurren en forma independiente, es decir, la ocurrencia de un evento en un intervalo de tiempo o espacio no afecta la probabilidad de una segunda ocurrencia del evento en el mismo u otro intervalo.
- 2) Teóricamente es posible que el evento pueda ocurrir infinitas veces en el intervalo.
- 3) La probabilidad de que ocurra un evento en un intervalo es proporcional a la longitud del intervalo.
- 4) Para determinar la probabilidad requerida de una muestra dada, es necesario utilizar como herramienta las tablas de Distribución acumulativa de Poisson ó Distribución de Poisson.
- 5) Es necesario que el experimento tenga una muestra total ( $P$ ), una muestra parcial ( $n$ ) que será a la cual se le determine una probabilidad "x" y una variable ( $x$ ).
- 6) Para el uso de la tabla es necesario obtener  $\lambda$ , mediante la multiplicación de  $n$  y  $P$ , por lo tanto  $\lambda = n P$ .

A continuación se presenta la manera de utilizar dicha tabla. La Tabla 3 es una versión abreviada de la tabla de Distribución acumulativa de Poisson del libro Canavos

Tabla 3

x	$\lambda$							
	0.1	0.2	0.3...	... 1.6	1.7 ...	... 2.4	2.5 ...	... 3.0 ...
0	.9048	.8187	.7408	.2019	.1827	.0907	.0821	<b>.0498</b>
1	.9953	.9825	.9631	.5249	.4932	.3084	.2873	.1991
2	.9998	.9989	.9964	.7834	.7572	.5697	.5438	.4232
3	1.000	.9999	.9997	.9212	.9068	.7787	.7576	.6472
4	1.000	1.000	1.000	.9763	.9704	.9041	.8912	.8153
5	1.000	1.000	1.000	.9940	.9920	.9643	.9580	.9161
6	1.000	1.000	1.000	.9987	.9981	.9884	.9858	.9665
7	1.000	1.000	1.000	.9997	.9996	.9967	.9958	.9881
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
13	.	.	.	.	.	.	.	.
14	.	.	.	.	.	.	.	.

Donde en la fila superior aparece el parámetro de la distribución  $\lambda$  (lambda), y en la primera columna aparece la variable aleatoria  $x$ .



Ejemplo: se tiene una muestra total de 50 / 500 cosas dadas (P) y se desea encontrar cierta probabilidad de la muestra para 30 cosas solamente (n) sabiendo que no se tendrá ni un solo error (x) por lo tanto se hace lo siguiente:

Primero encontramos el valor de lambda  $\lambda = n P \rightarrow \lambda = (30)(50 / 500) = 3.0$ , en la fila superior buscamos el valor de 3.0 y localizamos en la primera columna  $x = 0$  y la intersección de la columna de  $x = 0$  con  $\lambda = 3$  nos da la probabilidad que se está buscando que es igual a 0.0498.

### APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL MEDIANTE LA DISTRIBUCIÓN DE POISSON.

Quando el número de ensayos es muy grande los cálculos de las probabilidades Binomiales se vuelven tediosos y largos, por lo que es conveniente tener una aproximación de estas probabilidades, una de ellas es la aproximación de Poisson y la otra es la Normal

La aproximación de la distribución Binomial mediante la distribución de Poisson se usa cuando n es muy grande ( $n \rightarrow \infty$ ) y p es muy pequeño ( $n \rightarrow 0$ ) de tal manera que  $\mu = np$  permanezca constante, por lo tanto tenemos: (11)

$$\begin{aligned} f(x) &= \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)(n-x+1)}{x!} p^x (1-p)^{n-x} \end{aligned}$$

#### 5.- DISTRIBUCIÓN GEOMÉTRICA:

Si consideramos el caso especial de la distribución Binomial negativa donde x representa el número de ensayos en que ocurre el primer éxito ( $k = 1$ ), tenemos una distribución de probabilidad para el número de ensayos en que ocurre un solo éxito con función de probabilidad. (11)

$$\begin{aligned} f(x) &= p(1-p)^{x-1} \\ \text{Donde: } x &= 1, 2, 3, \dots, n \\ \mu &= 1/p \end{aligned}$$

Se utiliza cuando se tiene la probabilidad de algo en cierto número de unidades, cosas, etc... En teoría de colas, la distribución geométrica tiene importantes aplicaciones relacionadas con el número de unidades que son o esperan ser atendidas en un periodo más o menos breve. (6)

#### 6.- DISTRIBUCIÓN HIPERGOMÉTRICA:

Al haber extracciones con reemplazo los resultados son independientes por lo que la extracción con reemplazo lleva a la distribución Binomial. La extracción sin reemplazo es de gran importancia también pero conduce a una distribución más complicada, la distribución hipergeométrica se relaciona con situaciones con dos o más resultados, en los que la probabilidad de tener éxito cambia de ensayo a ensayo (no independientes).

*Ésta es la distribución en la que caen prácticamente todos los muestreos de aceptación por atributos, donde un elemento de la producción es clasificado bien como defectuoso o bien como no defectuoso.*

De este modo el número promedio de elementos defectuosos en una muestra es igual al tamaño de la muestra multiplicada por la fracción de elementos defectuosos en el lote. (2)

$$\mu = n(K/N)$$

Consideremos una población de  $N$  unidades,  $K$  de las cuales poseen ciertas características y  $N - K$  de las cuales no las poseen. Si se hace una elección al azar de una unidad entre tal población, el resultado debe ser una de las  $K$  (éxitos) o una de las  $N - K$  (fracasos). Pero si se hacen  $n$  elecciones al azar, sin reemplazo, entre la población, cada extracción subsecuente es dependiente y la probabilidad de éxito cambia a cada extracción. En estas condiciones si deseamos obtener  $x$  unidades del tipo  $K$ , en una muestra al azar de tamaño  $n$ , el número de éxitos en esta situación se llama variable hipergeométrica y es generada por tres números fijos  $N$ ,  $n$  y  $K$ .

Por lo tanto tenemos la siguiente ecuación hipergeométrica:

$$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$$

Donde:  $\binom{N}{n}$  es el número total de formas de elegir  $n$  artículos de los  $N$

$\binom{K}{x}$  es el número de formas diferentes de elegir  $x$  artículos que posean la característica  $K$ .

$\binom{N-K}{n-x}$  es el número de formas diferentes de escoger  $n - x$  artículos de los  $N - K$  que no poseen la característica. (11)

#### 4.1.4.- CUADRO RESUMEN DE VARIABLES DISCRETAS DISTRIBUCIONES TEORICAS DE PROBABILIDAD Variables aleatorias discretas

Distribución	Ecuación que define	Parámetros	Aplicaciones
Binomial	$f(x) = \frac{N!}{x!(N-x)!} p^x q^{n-x}$	$\mu = np$ $\sigma = npq$	*Cantidad de defectos y fracción defectuosa *Evaluación sensorial
Multinomial	$f(x) = \frac{N!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$		* 3 niveles de calidad * Llegadas a tiempo y a destiempo * Cuando las sumas de $x_i$ son = $n$
Poisson	$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$	$\lambda = \mu$ $\sigma^2 = \lambda$	* Curvas características de operación * Está en función al tiempo, espacio y volumen
Geométrica	$f(x) = p(1-p)^{x-1}$	$\mu = q/p$ $\sigma^2 = q/p^2$	*Para ensayos independientes y definidos * Cuando $n$ no es fija * C.C en el número de unidades atendidas en un tiempo más o menos breve
Hipergeométrica	$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$	$\mu = n \frac{K}{N}$ $\sigma^2 = \frac{N-n}{N-1} \mu \frac{K}{N} (1 - \frac{K}{N})$	* Cuando no se utiliza Binomial o Poisson * "r" ocurrencias en $n$ ensayos del lote "N"
Normal negativa	$f(x) = C_k^{x-1} p^x q^{k-x}$	$\mu = kq/p$ $\sigma^2 = kq/p^2$	* Número de ocurrencias en tiempo y espacio y cuando la frecuencia de está no es constante

4.2). COMPARACIÓN ENTRE LAS DISTRIBUCIONES BINOMIAL, NORMAL, POISSON, MULTINOMIAL, HIPERGEOMÉTRICA.

Distribución	Función de probabilidad	Número de resultados	Parámetros	Observaciones	Media	Varianza y Desviación	Coficiente de sesgo
Normal	$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$	Un solo resultado	$\mu$	Curva asintótica, tamaño de muestra grande	$\mu$	$\sigma^2$ $\sigma$	$\alpha_3 = 0$
Binomial	$P(x) = \frac{N!}{x!(N-x)!} p^x q^{N-x}$	2 resultados enteros entre 0 y n (un éxito y fracaso)	N y p	Independientes con probabilidad constante. Muestras con reemplazo	$\mu = Np$	$\sigma^2 = Npq$ $\sigma = \sqrt{Npq}$	$\alpha_3 = \frac{q-p}{\sqrt{Npq}}$
Poisson	$P(x) = \frac{\lambda^x e^{-\lambda}}{X!}$	Solamente éxitos. Enteros de 0 a $\infty$	$\lambda$	Independientes con probabilidad constante	$\mu = \lambda$	$\sigma^2 = \lambda$ $\sigma = \sqrt{\lambda}$	$\alpha_3 = 1/\sqrt{\lambda}$
Multinomial	$\frac{N!}{X_1! X_2! \dots X_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$	Más de 2 resultados		Muestra con reemplazo, independientes con probabilidad cte.			
Hipergeométrica	$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$	2 o más resultados	N, n, K	Observaciones dependientes. Muestreo sin reemplazo	$\mu = n\left(\frac{K}{N}\right)$	$\sigma^2 = \frac{nK(N-n)(N-K)}{N^2(N-1)}$	

4.3).- PROBLEMAS RESUELTOS**DISTRIBUCIÓN NORMAL**

1) Una industria produce caramelos con un diámetro medio de 0.75 in. Hay una cierta cantidad de variabilidad asociada con el proceso de manufactura, esto es, no todos los caramelos tienen exactamente 0.75in de diámetro. En el diámetro hay una desviación estándar de 0.02in. Suponga que el control de calidad exige que los caramelos tengan diámetros entre 0.745 y 0.755, cualquier cantidad fuera de este rango debe ser desechada de nuevo o vendida como desperdicio. ¿Qué proporción de caramelos será desechada si suponemos que los diámetros de los caramelos están normalmente distribuidos?

Solución:

$$\mu = 0.75 \text{ in}$$

$$\sigma = 0.02 \text{ in}$$

$$x_1 = 0.745 \text{ y } x_2 = 0.755 \text{ diámetros necesarios}$$

Siguiendo la secuencia antes mencionada tenemos que transformar los valores de  $x$  a valores de  $z$ , por lo tanto utilizando la fórmula tenemos:

$$z = \frac{(x - \mu)}{\sigma}$$

$$z_1 = \frac{(0.745 - 0.75)}{0.02} = -0.25 \quad \text{y} \quad z_2 = \frac{(0.755 - 0.75)}{0.02} = 0.25$$

Consultando la tabla de distribución normal con el valor de  $z_1 = -0.25$  tenemos  $\Rightarrow$  0.4013 pero como la curva es simétrica donde 50% está a la derecha y 50% a la izquierda, esto se observa en los datos obtenidos para  $z_1$  y  $z_2$ . Tenemos  $(0.4013) = 0.8026$  por lo tanto tenemos el 80.26% de caramelos defectuosos.

2) Un rancho tiene 500 vacas productoras de leche con una media de 18lt y una varianza de 4 lt<sup>2</sup>.

- Encontrar una proporción de vacas que producen entre 18 y 23lt
- Cuántas vacas producen entre 19 y 20lt.
- Proporción de vacas que producen más de 14lt
- Proporción de vacas con producción inferior a 14lt
- Proporción de vacas con producción entre 16 y 20lt.
- Cual es la producción por abajo de la cual se encuentra el 90% de la población
- Cual es la producción por arriba de la cual se encuentra el 95% de la población.
- Cuales son las 2 proporciones entre las cuales se encuentra el 99% de la población.

Solución:

$$\mu = 18 \text{ lt}$$

$$\sigma^2 = 4 \text{ lt}^2 \quad \text{y una } \sigma = 2$$



- $z = 23 - 18 / 2 = 2.5$  buscando en la tabla tenemos 0.9938  
 $(0.9938 - 0.5) = 0.4938$  por lo tanto como 500 vacas son el 100% tenemos que

500 vacas  $\rightarrow$  100%

$x \rightarrow 49.38\% = 246.9 \approx 247$  vacas producen entre 18 y 23lt de leche

b)  $z_1 = 19 - 18 / 2 = 0.5$  buscando en la tabla tenemos 0.6915

$z_2 = 20 - 18 / 2 = 1$  buscando en la tabla tenemos 0.8413



tenemos  $(0.8413 - 0.6915) = 0.1498$

500 vacas  $\rightarrow$  100%

$x \rightarrow 14.98\% = 74.9 \approx 75$  vacas producen entre 98 y 203lt de leche

c)  $z = 14 - 18 / 2 = -2$  buscando en la tabla tenemos 0.0228



como toda el área bajo la curva equivale a uno hacemos lo siguiente  $1 - 0.0228 = 97.72\%$  por lo tanto este es el porcentaje de vacas que producen más de 14 litros de leche.

d)  $z = 14 - 18 / 2 = -2$  buscando en la tabla tenemos 0.0228 por lo tanto sí

500 vacas  $\rightarrow$  100%

$x \rightarrow 2.28\% = 11.4 \approx 11$  vacas que tienen una producción menor a 14lt de

leche



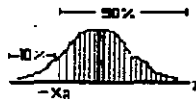
e)  $P(16 \leq x \leq 20)$  por lo tanto sacamos los valores de  $z_1$  y  $z_2$

$z_1 = 16 - 18 / 2 = -1$ ;  $z_2 = 20 - 18 / 2 = 1$



Buscando en la tabla el valor de  $z$  tenemos que  $z_1 = -1 \Rightarrow 0.1587$  y  $z_2 = 1 \Rightarrow 0.8413$   
 $0.8413 - 0.1587 = 0.6826$  es decir tenemos el 68.26% de vacas que producen entre 16 y 20 litros de leche, o sea, 341 vacas de 500

f)  $P(x < 90\%)$  por lo tanto es necesario encontrar el valor de  $x$ , despejando de la formula tenemos que  $x = z \sigma + \mu$  así que primero se tiene que buscar el valor de  $z$

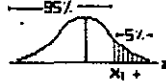


Como lo que se quiere es toda el área sombreada y no tenemos ningún valor de  $x$  para poder obtener  $z$ , lo que se hace es lo siguiente con el 90% se busca en la tabla el número 0.90 o el más cercano, pero no se va a buscar ni por la columna de  $z$  ni por la fila superior, si no que será buscado por todo el conjunto de números que se encuentran dentro de éstos dos. Una vez que se encuentra el número se intercepta la columna y la fila superior.

El número encontrado es 0.9015 interceptando con la columna de  $z$  y la fila superior tenemos  $z = 1.29$  ahora sustituimos en la fórmula.

$X = (-1.29)(2) + 18 = 15.42$  lt es la producción que se encuentra por abajo del 90% de la población.

*Nota: recuerde que: Valores (+) arriba del 90% y abajo del 10%  
Valores (-) abajo del 90% y arriba del 10%*

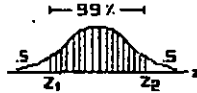


g)  $P(x > 95\%)$  hacemos lo mismo que en el inciso (e)

Encontramos el valor de  $z = 1.65$  sustituyendo en la fórmula tenemos

$x = (1.65)(2) + 18 = 21.3$ , por lo tanto el 21.3% de la producción se encuentra por arriba del 95%

h) Buscamos el valor de  $z$  con el 99%,  $z = 2.33$ , sustituimos en la fórmula para obtener el valor de  $x_1$  y  $x_2$



$$x_1 = (-2.33)(2) + 18 = 13.34 \text{ lt}$$

$x_2 = (2.33)(2) + 18 = 22.66$  lt por lo tanto éstas son las dos proporciones en las que se encuentra el 99% de la población.

3) El número de Ingenieros en Alimentos que entran a un seminario de Desarrollo Industrial en cualquier día dado está aproximadamente distribuido en forma normal con media de 40 y una desviación de 11. Hallar la probabilidad de que durante un día dado:

- Un mínimo de Ingenieros llegue
- 35 Ingenieros lleguen
- Entre 35 y 42 Ingenieros inclusive

Solución:

$$\mu = 40$$

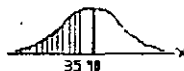
$$\sigma = 11$$

a)  $P(x < 30) z = 29.5 - 40 / 11 = -0.95$  buscando en la tabla se tiene 0.1711



$(1 - 0.1711) = 0.8289 \Rightarrow 82.89\%$  es la probabilidad de que lleguen 30 Ingenieros

b)  $P(x = 35), z_1 = 34.5 - 40 / 11 = -0.5$  y  $z_2 = 35.5 - 40 / 11 = -0.4$



Buscamos en tabla los valores de las  $z$ 's y tenemos que  $z_1 = -0.5 \Rightarrow 0.3085$  y  $z_2 = -0.4 \Rightarrow 0.3446$

$(0.3446 - 0.3085) = 0.0388$  por lo tanto tiene el 3.88% de probabilidad de que lleguen 35 Ingenieros

$$\begin{aligned} \text{c) } P(35 \leq x \leq 42), z_1 &= 35 - 40 / 11 = -0.45 \\ z_2 &= 42 - 40 / 11 = 0.18 \end{aligned}$$



Buscando en la tabla tenemos  $z_1 = 0.3264$  y  $z_2 = 0.5714$   
 $(0.5714 - 0.3264) = 0.245$  por lo tanto el 24.5% es la probabilidad de que entren 35 y 42 Ingenieros

### DISTRIBUCIÓN DE POISSON

1) Supóngase que una cepa particular de estafilococos produce un determinado síntoma en el 2% de las personas infectadas, en un día dado 200 personas comieron alimentos contaminados con dicho microorganismo. Cuales son las probabilidades que tengan síntomas:

- Exactamente 4 personas
- Ninguno
- Más de 4 personas
- 10 ó menos personas

#### Solución:

Datos:  $p = 20\% = 0.02$ ,  $N = 200$

$$\begin{aligned} \text{a) } \lambda &= Np = (200)(0.02) = 4 \\ x &= 4 \end{aligned}$$

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(4)^4 (2.71)^{-4}}{4 \times 3 \times 2 \times 1} = .1953 \approx 19.53\%$$

Es la probabilidad de que exactamente 4 personas tengan síntomas.

Es necesario mencionar que el resultado será el mismo si se realiza por la fórmula o en tablas la cual resulta más rápida (esa es la diferencia de usar un método u otro).

Por tablas de Distribución de Poisson obtenemos una probabilidad de 0.1953  $\approx$  19.53%

$$\text{b) } x = 0$$

Buscando en la tabla de distribución acumulativa de Poisson tenemos 0.0183 es decir el 1.83% de probabilidad de que ninguna persona se contagiara.

$$\text{c) } x = 3, 4, 5, 6, 7, \dots, 20 \text{ buscamos en la tabla a partir de } x = 3 \text{ puesto que estamos}$$

tratando de datos acumulativos.

$1 - p(3) = 1 - 0.4335 = 0.5665$  por lo tanto tiene el 56.65% de probabilidad de que más de 4 personas tengan los síntomas.

$$\text{d) } x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \dots, 19, 20 \text{ buscamos en la tabla con una } x = 10$$

tenemos 0.9972 por lo que el 99.72% es la probabilidad de que 10 o menos personas tengan los síntomas.

2) Al inspeccionar un almacén de cereales por el personal de control de calidad, se descubren en promedio 0.2 granos germinados por costal. Calcular la probabilidad de descubrir:

- un grano germinado por 3 costales
- al menos 2 granos germinados por 5 costales
- a lo mucho un grano germinado en 15 costales

Solución:

Datos:  $p = 0.2$

a)  $\lambda = Np$

$\lambda = (3)(0.2) = 0.6$  donde  $x = 1$  resolvemos por ecuación

$P(x = 1) = (0.6)^1(2.71)^{-0.6} = 32.93\%$  es la probabilidad de encontrar 1 grano germinado por tres costales. Si buscamos en tablas de distribución de Poisson se encontrará el mismo resultado.

b)  $\lambda = Np$

$\lambda = (5)(0.2) = 1.0$  donde  $x = 2, 3, 4, 5, \dots, \infty$  resolvemos por Tablas de Distribución acumulativa de Poisson donde se elegirá  $x = 1$ , por lo tanto tenemos:  $(1 - 0.7358) = 0.2642 \approx 26.42\%$  es la probabilidad de encontrar de 2 o más granos germinados en 5 costales.

c)  $\lambda = Np$

$\lambda = (15)(0.2) = 3.0$  donde  $x = 0, 1$  buscamos en la tabla de distribución acumulativa de Poisson tenemos el 19.91% de probabilidad de encontrar a lo mucho 1 grano germinado en 15 costales.

3) En promedio 120 personas/hora consultan a los especialistas en la industria de refrigeración en México. Cual es la probabilidad de que 3 personas se acerquen al especialista en 10 min.

Solución:

$\lambda = Np \Rightarrow \lambda = 120 / 60 = 2$  personas por minuto buscamos en la Tabla de distribución de Poisson donde  $x = 0, x = 1, x = 2$  por lo tanto tenemos:

$x(0) = 0.1353$

$x(1) = 0.2706$

$x(2) = 0.2706 \Rightarrow \Sigma = 0.6765$   $(1 - 0.6765) = 32.85\%$  es la probabilidad de que 3 personas consulten al especialista en 10 minutos



## DISTRIBUCIÓN BINOMIAL

1) La probabilidad de que un microorganismo sobreviva a un tratamiento térmico es 0.4 si se sabe que 15 m.o han recibido dicho tratamiento. Cual es la probabilidad de que:

- a) Sobrevivan a lo más 10
- b) Sobrevivan entre 3 y 8 inclusive
- c) Sobrevivan exactamente 5

### Solución:

Datos:  $p = 0.4$ ,  $q = (1 - p) = 0.6$ ,  $N = 15$

a)  $P(x < 10)$  = buscamos en la Tabla de Distribución acumulativa Binomial y tenemos que con una  $x = 10 \Rightarrow 0.9662$  por lo tanto  $(1 - 0.9662) = 0.0338 \approx 3.38\%$  es la probabilidad de que sobrevivan al menos 10 m.o

b)  $P(8) - P(3)$  = buscamos en la misma tabla con una  $x = 8$  y  $x = 2$ :

$$X(8) = 0.9050$$

$X(2) = 0.0271 \Rightarrow 0.9050 - 0.0271 = 0.8779 \approx 87.79\%$  es la probabilidad de que sobrevivan entre 3 y 8 m.o.

c)  $P(x=5)$  = Buscamos en la tabla con una  $x = 5$  y  $x = 4$  puesto que estamos tratando con una tabla de distribución acumulativa.

$$X(5) = 0.4032$$

$X(4) = 0.2173 \Rightarrow 0.4032 - 0.2173 = 18.59\%$  como probabilidad de que sobrevivan 5 m.o

*Nota: Si este dato del inciso © se busca directamente en una tabla de Distribución Binomial con una  $x = 5$  nos dará el mismo resultado.*

2) Si el 20% de envases tetra pak sellados por una máquina son defectuosos, determinar la probabilidad de que 4 envases sellados por la máquina y elegidos al azar:

- a) uno sea defectuoso
- b) ninguno sea defectuoso
- c) a lo más 2 sean defectuosos

### Solución:

Datos:  $p = 0.20$ ,  $q = 0.80$ ,  $N = 4$

a)  $x = 1$  utilizamos nuestras formulas anteriormente mencionadas

$$P(x) = C_x^N p^x q^{N-x}$$

$P(1) = C_1^4 \cdot 0.20^1 \cdot 0.80^3$  primero es necesario obtener el valor de C para utilizar la ecuación, por lo tanto utilizamos la siguiente formula:

$$C_x^N = \frac{N!}{x!(N-x)!} = \frac{4 \times 3 \times 2 \times 1}{1(3 \times 2 \times 1)} = \frac{24}{6} = 4 \text{ ahora ya podemos sustituir en la primera ecuación para poder obtener la probabilidad deseada}$$

$P(1) = (4) (0.20)^1 (0.80)^3 = 0.4096 \approx 40.96\%$  es la probabilidad de que un solo envase sea defectuoso.

b)  $x = 0$

$$P(x) = C_x^N p^x q^{N-x}$$

$$P(0) = C_0^4 0.20^0 0.80^4$$

$$C_x^N = \frac{N!}{x!(N-x)!} = \frac{4 \times 3 \times 2 \times 1}{1(3 \times 2 \times 1)} = 4, \text{ por lo tanto tenemos: } P(0) = (4) (0.20)^0 (0.80)^4 =$$

$0.4096 \approx 40.96\%$  es la probabilidad de que ninguno sea defectuoso.

Si se busca por tablas de Distribución acumulativa Binomial obtenemos el mismo resultado.

c)  $x = 0, 1, 2$  buscamos en tabla acumulativa Binomial y tenemos:  $0.9728 \approx 97.28\%$  es la probabilidad de que a lo más 2 envases tetra pak sean defectuosos de un total de 4.

3) Tres almacenes para la conservación de hortalizas comienzan simultáneamente y su éxito o quiebra es independiente uno del otro. Si la probabilidad de éxito es 0.5, determinar cual es la:

- probabilidad de un éxito y 2 quiebras
- probabilidad de 2 éxitos y una quiebra
- probabilidad de que a lo más una tenga éxito
- probabilidad de que la firma A tenga éxito mientras B y C quiebren

#### Solución:

$$p = 0.5, q = 0.5, N = 3$$

a)  $x = 2$  Buscamos en la tabla acumulativa Binomial 0.8750 por lo tanto el 87.50% es la probabilidad de tener 2 quiebras y un éxito.

b)  $x = 1$  por la misma tabla obtenemos .5000  $\approx 50.0\%$  es la probabilidad de que un almacén quiebre y uno tenga éxito.

c)  $x = 0, 1$  Buscamos en tabla y tenemos 0.5000  $\approx 50.0\%$  es la probabilidad de que a lo más uno tenga éxito.

d)  $P(A)(B)(C) = (0.6)(0.4)(0.4) = 0.096 \approx 9.6\%$  es la probabilidad de que la firma A tenga éxito mientras que B y C quiebren.

### *DISTRIBUCIÓN MULTINOMIAL*

Un Ingeniero sabe que el 80% de la producción de una máquina formadora de cajas es aceptable, el 15% de la producción requiere de cierto trabajo adicional y el 5% se debe descartar. En una muestra de  $n = 10$ ; cual es la probabilidad de obtener 8 partes buenas, 2 que necesiten ser repetidas y ninguna se deba descartar.

Solución:

$$f(x_1, x_2, \dots, x_k) = \frac{N!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$f(8, 2, 10) = \frac{10!}{8! 2! 0!} (0.8)^8 (0.15)^2 (0.05)^0 = 0.17 \approx 17\%$  es la probabilidad de obtener 8 cajas bien formadas, 2 que necesitan ser repetidas y ninguna que sea inservible.

### *DISTRIBUCIÓN HIPERGEOMETRICA*

Una serie de 8 termopares se conectan a un cuarto frío, de tal forma que si uno de ellos falla, el sistema no funcionará. Si dos termopares fallan:

- Cuál es la probabilidad de que el primer termopar que se inspeccione, sea una de los que falló
- Cuál es la probabilidad de encontrar los dos termopares que fallaron si se inspeccionan cuatro de ellos.
- Cuántos termopares se deben inspeccionar para tener un 70% de probabilidades de encontrar los dos termopares defectuosos.

Solución:

Datos:  $N = 8$ ,  $k = 2$ , utilizando la formula  $f(x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$

a)  $n = 1, x = 1$

$f(1) = \frac{\binom{2}{1} \binom{6}{0}}{\binom{8}{1}} = \frac{2}{8} = 0.25 \approx 2.5\%$  es la probabilidad de que el primer termopar que se inspeccione sea el que fallo.

b)  $n = 4, x = 2$

$f(2) = \frac{\binom{2}{2} \binom{6}{2}}{\binom{8}{4}} = \frac{15}{70} = 0.2143 \approx 21.43\%$  es la probabilidad de encontrar los 2 termopares que fallan al inspeccionar 4 de ellos.

c)  $n = ?, x = 2$

$f(2) = \frac{\binom{2}{2} \binom{6}{n-2}}{\binom{8}{n}} = 0.70$  como no se tiene el valor de "n" es necesario despejarlo de la

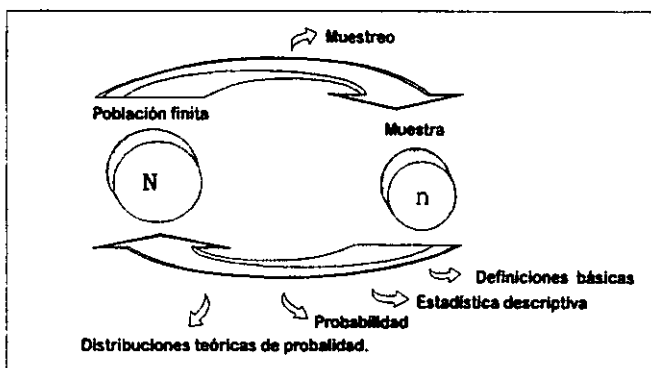
formula:

$$\frac{6!}{(n-2)!(8-n)!} = 0.70 \Leftrightarrow \frac{6!n!}{8!(n-2)!} = 0.70 \Leftrightarrow \frac{n(n-1)}{8 \cdot 7} = 0.70 \Leftrightarrow n^2 - n = 39.2$$

$n = 6.78 \approx 7$ , ya que se encontró el valor de n, se sustituye en la primera ecuación.

## V.- MUESTREO

Cuando se hace una investigación estadística, se deben seguir ciertos pasos o etapas; después de definir el problema en términos estadísticos y antes de recolectar los datos tenemos que decidir de que forma se van a coleccionar dichos datos; es decir, que procedimiento se va a utilizar para obtener la muestra, de manera que podamos aprender algo acerca de la población sobre la base de los datos extraídos de una parte de ella.



Las razones por las cuales se utiliza el muestreo y no toda la población son varias; los grandes progresos hechos en la teoría del muestreo durante las últimas décadas hacen posible medir las propiedades de cantidades masivas de datos con precisión calculada sobre la base de muestras.

Además del hecho de que pueden obtenerse resultados fidedignos de procedimientos de muestreo correctos, hay otras importantes razones para su amplia adopción. Primero, las poblaciones que se investigan pueden ser infinitas, y en tales casos la muestra es el único procedimiento posible. Segundo, aún en el caso de poblaciones finitas, muy frecuentemente el muestreo es el único procedimiento práctico, pues, una población finita puede constar de millares o millones de elementos y su enumeración completa es prácticamente imposible.

En tercer lugar, la medición de las propiedades de una población a veces requiere de la destrucción de sus elementos, por último aunque financiera, práctica y físicamente sea posible observar toda la población, el *muestreo* puede ser aún el procedimiento más eficiente.

Los resultados obtenidos por el estudio de una muestra pueden ser iguales o aún más precisos que los hallados de una cuenta completa del conjunto.

Cualquier encuesta estadística, muestreo o censo, siempre contiene cierto error. Los errores estadísticos son de dos clases: no de muestreo y de muestreo.

Por estas razones no solo puede esperarse que el error total sea menor en un estudio de muestra, sino que los resultados de ella también pueden ser usados con un mayor grado de confianza por nuestro conocimiento del tamaño probable del error. (11)

### 5.1).- DEFINICIONES BÁSICAS

**\*MUESTRA.-** Es la parte de la población obtenida de acuerdo a una regla determinada y debe ser obtenida a partir de un experimento aleatorio, con tamaño "n". Debe ser representativa de la población, por lo tanto es necesario diseñar el tamaño el cual debe ser de una muestra suficientemente grande, de tal manera que contenga todas las variedades de la característica que se trata. Esto puede traer consigo la desventaja de tener altos costos para su obtención por lo que debe de tener un tamaño adecuado, es decir, no muy reducido ni tampoco muy grande. (8)

**\*MUESTRA REPRESENTATIVA.-** Tiene como finalidad ser útil al inferir estadísticamente sobre la población.

**\*MUESTRA ALEATORIA.-** Procedimiento mediante el cual se puede obtener muestras representativas, en el sentido de que cada miembro de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra.

Matemáticamente se considera una variable aleatoria "x", asociada a una población en donde se supone que está la variable A, en algún tipo de distribución de densidad. Para que un muestreo sea aleatorio se requiere que:

- a) Las pruebas para obtener una muestra sean independientes una de otra.
- b) La distribución de probabilidad de la variable aleatoria "x" permanezca siendo la misma de prueba a prueba. (6)

**\*MUESTREO AL AZAR.-** Una técnica para obtener una muestra al azar es asignando números a cada miembro de la población, escritos en pequeños papeles, se introducen en una urna y después se extraen números de la urna, teniendo cuidado de mezclarlos bien antes de cada extracción, pero esto puede ser sustituido por el empleo de una tabla de números aleatorios construida especialmente para tales propósitos. (1)

*Nota:* Ver APÉNDICE Para el uso de la Tabla de Números aleatorios.

**\*NÚMEROS ALEATORIOS.-** Se obtienen mediante la utilización de una urna o de una tabla. Esta tabla es en realidad una muestra aleatoria de la población formada por los dígitos del 0 al 9.

#### Uso de la Tabla de Números Aleatorios:

- Se enumeran los datos, por ejemplo si se tiene una muestra de 200, comenzando con el 001, 002, 003, ... hasta el 200, o bien 000 hasta 199

- Localizar el punto de partida aleatoriamente (se puede comenzar en cualquier renglón, en cualquier columna, en cualquier sentido por ejemplo de arriba a bajo, de derecha a izquierda, diagonalmente, etc...)
- Una vez localizado el punto, se debe tomar los números que se encuentran entre los valores propuestos hasta completar el tamaño deseado de la muestra.

### 5.2).- BASES TEORICAS DEL MUESTREO

Aunque la diversidad es una cualidad universal de los datos, no hay virtualmente ninguna población estadística en la práctica cuyos elementos varíen entre sí sin límite.

Los hechos de que cualquier población tiene propiedades características y que las variaciones en sus elementos son claramente limitadas hacen posible que escojamos una muestra relativamente pequeña y al azar que puede reflejar bastante bien las características de la población.

Debido a la uniformidad estadística, si se escoge una muestra grande al azar, la característica de esta muestra difiere muy poco de las características de la población, los promedios son más estables que los valores individuales, además resultan más estables cuantas más observaciones haya en la muestra. (11)

La Teoría del muestreo tiene los siguientes objetivos:

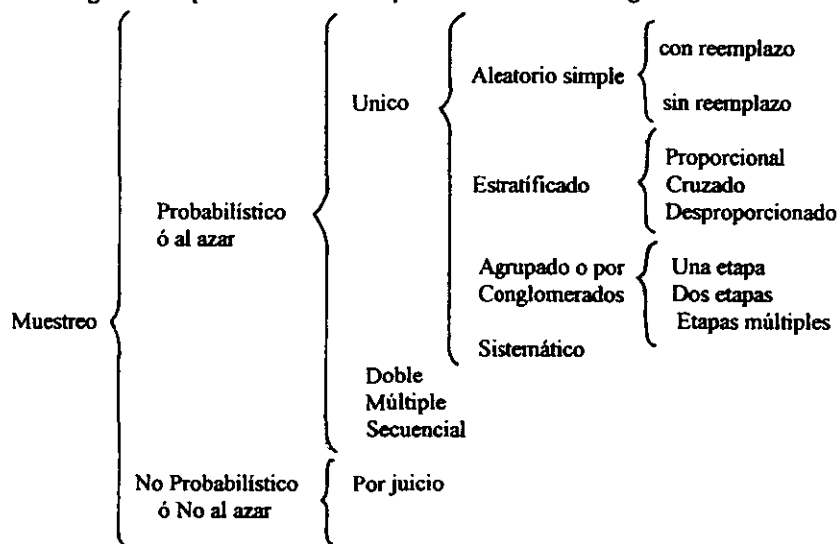
- 1) Obtención de muestras representativas
- 2) Relación de población y muestras
- 3) Es el punto de unión entre la Estadística Descriptiva, la Probabilidad y la Inferencia Estadística ya que se puede “estimar” a la población a partir de la muestra.
- 4) Se puede emplear para tomar decisiones.
- 5) Es un estudio de las relaciones existentes entre una población y muestras extraídas de la misma, permite estimar cantidades desconocidas de la población (media poblacional, varianza, etc. llamadas parámetros poblacionales) a partir del conocimiento de las correspondientes cantidades muestrales (media muestral, varianza, etc. llamados estadísticos muestrales).
- 6) También es útil para determinar las diferencias que se puedan observar entre dos muestras si son debidas a la aleatoriedad de las mismas o si por el contrario son realmente significativas. (1)

### 5.3).- TIPOS DE MUESTREO

Los modelos de muestreo pueden agruparse convenientemente como *muestreo al azar* ó *muestreo probabilístico*, porque si el proceso de muestreo es al azar, puede aplicarse las leyes de probabilidad y *muestreo no al azar* o *por juicio* es un proceso de selección de muestras sin el uso del azar. En otras palabras se escoge una muestra no al azar sobre una base distinta de consideraciones de probabilidad, tal como el juicio de un experto, conveniencia o algún otro criterio.

En consecuencia la probabilidad, de que cada elemento individual sea extraído de la población es desconocida y la fidelidad de sus resultados no pueden ser objeto del análisis de probabilidades, sino que debe depender del juicio personal del investigador.

Los siguientes tipos de muestreo se pueden resumir en el siguiente cuadro:



### 5.3.1) MUESTREO ALEATORIO SIMPLE:

Se puede describir este método fundamental de selección de la muestra de la siguiente manera. A partir de una población de  $N$  unidades seleccionese una, dando igual probabilidad a todas las unidades. La mejor forma de proceder es con la ayuda de una tabla de números aleatorios, se puede demostrar el número de maneras de seleccionar una muestra de la siguiente manera: <sup>(3)</sup>

$$C_n^N = \frac{N!}{n!(N-n)!}$$

Donde:  $N$  = población,  $n$  = muestra.

El muestreo aleatorio simple es ciertamente un procedimiento práctico si la población no es grande y si es relativamente fácil y barato encontrar las unidades de muestreo. También podría ser un procedimiento práctico para poblaciones grandes cuyos elementos están concentrados dentro de una área pequeña.

Un inconveniente de aplicar el muestreo aleatorio simple a poblaciones grandes es que la población debe ser enumerada. <sup>(11)</sup>

Muestra aleatoria simple con reemplazo. Si la operación se realiza  $n$  veces Todo miembro de la población está disponible en cada extracción es decir puede elegirse más de una vez. <sup>(1)</sup>

Muestra aleatoria simple sin reemplazo: El registro extraído no se regresa con los demás, hasta después de extraer la muestra completa, por lo tanto un registro dado aparecerá solo una vez.

### 5.3.2) MUESTREO ESTRATIFICADO:

Si la población es muy heterogénea y las consideraciones de costo limitan el tamaño de la muestra, podría ser imposible obtener un estimador lo suficientemente preciso tomando una muestra aleatoria simple de toda la población. Pero, supongamos que sea posible encontrar una forma de subdividir la población de tal manera que se reduzca considerablemente la variación dentro de las subdivisiones, entonces podrá hacerse una mejor estimación de la población total.

El proceso de estratificación requiere que la población sea dividida en grupos o clases llamados estratos. Entonces se toma una muestra de cada estrato por métodos simples al azar o aleatorios simples, y la muestra resultante se llama *muestra estratificada*.

Una muestra estratificada puede ser proporcional o desproporcionada.

En el Muestreo estratificado proporcional, el número de unidades extraídas de cada estrato es proporcional al tamaño de éste, por ejemplo si la población es dividida en cuatro estratos, siendo sus tamaños respectivos 10, 20, 30 y 40 por ciento de la población y ha de extraerse una muestra de tamaño 500, la muestra proporcional deseada deberá obtenerse de la siguiente manera:

Estrato	I	500 (0.10)	= 50 unidades
Estrato	II	500 (0.20)	= 100 unidades
Estrato	III	500 (0.30)	= 150 unidades
Estrato	IV	500 (0.40)	= 200 unidades
	Tamaño	de la muestra	= 500 unidades

El Muestreo estratificado desproporcionado comprende los procedimientos de tomar un número igual de unidades de cada estrato, sin tener en cuenta su tamaño.

La eficiencia de la estratificación puede aumentar aún más cuando la naturaleza de los datos lo permite, los que, a su vez, pueden ser subdivididos en grupos más pequeños. Este refinamiento en la estratificación de grupos de población con respecto a varias características relevantes simultáneamente se conoce como estratificación cruzada.

La estratificación es más eficaz con poblaciones heterogéneas o altamente asimétricas, en tales situaciones podemos estratificar la población de tal manera que:

- 1) dentro de cada estrato haya la mayor uniformidad posible
- 2) entre los distintos estratos las diferencias sean lo más grandes posibles.

Así, antes de decidir sobre la estratificación, debemos poseer algún conocimiento de las características de la población, tal conocimiento puede basarse en datos pasados, observación preliminar de estudios piloto, juicio experto o simplemente intuición o buenas conjeturas. (11)

La ventaja de esta procedimiento consiste en que, si es posible identificar los subconjuntos homogéneos de la población, solo se necesita un número relativamente pequeño de observaciones para determinar las características de cada subconjunto. (8)



Ejemplo: Existen tres estratos en los cuales dentro de ellos sus características son homogéneas, pero entre estratos son heterogéneas.

$N_1$	$N_2$	$N_3$	
* * * * *	# # # # #	o o o o o o o o	Dentro de cada estrato se tiene que manejar el muestreo aleatorio simple
*			
↓	↓	↓	
$\sigma^2_1$	$\sigma^2_2$	$\sigma^2_3$	Varianza especificada para cada estrato
\$	\$	\$	Costo específico para cada estrato

Donde:  $K$  = Número de estratos

Se requiere determinar cuanta muestra de cada estrato se necesita, así que puede hacerse de diferentes maneras:

Utilizando un (a) estratificado igual; (b) estratificado proporcional; (c) estratificado de Neyman y un (d) estratificado óptimo. A continuación se muestra un cuadro con los diferentes tipos de estratificación, sus características y la ecuación que debe ser utilizada.

Muestra Estratificada	Características	Ecuación
Igual	Todo es igual $N_i, \sigma^2, C_i$ ó sin costo	$n_i = \frac{n}{K}$
Proporcional	$N_i \neq, \sigma^2 =, C_i$ ó sin costo	$n_i = \frac{N_i}{N} n$
Neyman	$N_i \neq, \sigma^2 \neq, C_i$ ó sin costo	$n_i = \frac{N_i \sigma_i}{\sum_{k=1}^k N_k \sigma_k} n$
Óptimo	$N_i \neq, \sigma^2 \neq, C_i$	$n_i = \frac{N_i \sigma_i / \sqrt{c_i}}{\sum_{i=1}^k N_i \sigma_i / \sqrt{c_i}} n$

Donde:  $N_i$  = tamaño de cada estrato,  $N$  = tamaño de la población,  $n$  = tamaño de muestra buscada,  $n_i$  = tamaño de muestra de cada estrato,  $K$  = número de estratos,  $\sigma^2$  = varianza,  $C_i$  = costo (si el costo es elevado disminuye el tamaño de la muestra).

### 5.3.3) MUESTREO POR CONGLOMERADO:

Para grandes poblaciones, se pierde aleatoriedad ya que se van haciendo elecciones. En el se subdivide la población en grupos con el propósito de obtener una muestra eficiente. Cada subdivisión se llama conglomerado y cada uno tiene las mismas características de la población. <sup>(8)</sup>

Cuando los grupos o regiones que son las unidades primarias son extraídos, podemos incluir en la muestra todas las unidades elementales de las regiones escogidas o tomar una muestra de unidades primarias más pequeña o unidades elementales de las regiones escogidas.

Cuando se han observado todas las unidades elementales de las regiones, tenemos como lo que se conoce muestreo de una sola etapa. Cuando se extrae una muestra de unidades elementales de las regiones, tenemos un tipo de diseño llamado muestreo en dos etapas o submuestreo. En ambas etapas se escoge una muestra simple al azar.

Cuando el muestreo por agrupación supone más de dos etapas para escoger la muestra final, se llama muestreo en múltiples etapas.

En el muestreo por conglomerados es eficiente tener:

1. Diferencias entre las unidades elementales del mismo grupo lo más grandes posibles
2. Diferencias entre los grupos lo más pequeñas posibles.

### 5.3.4) MUESTREO SISTEMÁTICO:

Otro diseño usado frecuentemente es el sistemático; para obtener una *muestra sistemática* al azar podemos numerar también las unidades de muestreo de la población serialmente de 1 a N y determinar primero lo que se conoce por intervalo de muestreo,  $k = N/n$ . Luego se escoge al azar un número del primer intervalo de muestreo. Si este número es por ejemplo  $a$ , con  $a < k$ , entonces la muestra con tamaño  $n$  tendría como su miembro las unidades de muestreo cuyos números de serie corresponden a:  $a, a+k, a+2k, a+3k$ . <sup>(11)</sup>

Ejemplo:  $N = 1000, n = 8$

1) Seleccionar un número entre 1 y  $N/n$

2) Se toman sucesivamente las unidades faltantes de  $N/n$  en  $N/n$ , si  $N = 1000, n = 8$

$N/n = 1000/8 = 125$  por lo tanto

\*  $005 + 125 = 130$

$130 + 125 = 155$  etc..

Nota: \* el primer número se toma al azar

### 5.3.5) MUESTREO DOBLE Y MÚLTIPLE:

Todos los tipos de muestreo anteriores se conocen como muestreo único, pues cada diseño se usa para obtener una sola muestra de la cual se hace una estimación o una prueba de hipótesis. Sin embargo en últimas fechas se ha generalizado el uso del muestreo doble, múltiple y secuencial.

El *Muestreo doble* es el proceso por el cual se observa primero una muestra pequeña y se toma una decisión sobre aceptar, rechazar o tomar una segunda muestra. Así por ejemplo, en el muestreo doble podemos formular esta regla de decisión. Escoja una muestra al azar de 40 unidades, si se encuentran dos o menos unidades defectuosas acepte el lote; si se encuentran 4 o más defectuosas, rechace el lote; si se encuentran 3 defectuosas, tome una segunda muestra de 60 unidades, el total muestreado es ahora de 100. El procedimiento que utiliza una serie de muestras pequeñas tal que el número acumulado de defectos es comparado con un criterio de aceptar y uno de rechazar después de extraída cada muestra hasta que finalmente puede tomarse una decisión, se llama *Muestreo Múltiple*.

#### 5.4).- DISTRIBUCIONES MUESTRALES:

Se llama distribución muestral a la distribución de todos los valores posibles que pueden ser tomados por algún estadístico, calculados a partir de muestras del mismo tamaño extraídas aleatoriamente de la misma población. (1)

Para construir una distribución muestral se procede de la siguiente manera:

1) De una población finita discreta, de tamaño  $N$ , se extraen aleatoriamente todas las muestras de tamaño  $n$ . En general para muestreo con reemplazo, el número de muestras posibles de tamaño  $n$  de una población de tamaño  $N$ , es igual a  $N^n$ , y para muestras sin reemplazo e ignorando el orden, el número de muestras de tamaño  $n$  es  $C_n^N$  ó  $\binom{N}{n}$ .

2) Se calcula el estadígrafo (es un valor que describe en forma global a una muestra) ó estadístico de interés para cada muestra

3) Se construye una tabla de los valores del estadígrafo de interés con sus respectivas probabilidades.

4) Se calculan los parámetros de  $\mu$  y  $\sigma^2$  de la distribución

#### 5.5).- DISTRIBUCIÓN DE MEDIA Y VARIANZA DE UNA MUESTRA:

Si una muestra aleatoria de tamaño  $n$  se elige de una población que tiene media  $\mu$  y varianza  $\sigma^2$ , entonces la  $x$  es un valor de una variable aleatoria cuya distribución tiene media poblacional.

Para muestras tomadas de poblaciones infinitas la varianza de cada distribución es  $\sigma^2 / n$ ; para muestras extraídas de poblaciones finitas de tamaño  $N$  la varianza es  $\frac{\sigma^2}{n} \frac{N-n}{N-1}$ . (6)

Supóngase una población:

$N = 4$  por lo tanto  $X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 4$

Calculando la media y la varianza poblacional:

$$\mu = \sum_{i=1}^N \frac{x_i}{N} = \frac{1+2+3+4}{4} = 2.5$$

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{N} = \frac{(1-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2}{4} = 1.25$$

Seleccionando todas las muestras posibles de tamaño 2 y determinando las medias de las muestras: Muestras de tamaño 2 ( cantidad de muestras =  $N^n$  )

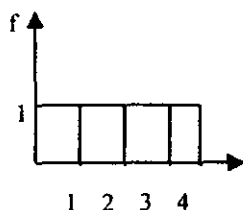
1.1	2.1	3.1	4.1
1.2	2.2	3.2	4.2
1.3	2.3	3.3	4.3
1.4	2.4	3.4	4.4
<hr/>			
1	1.5	2	2.5
1.5	2	2.5	3
2	2.5	3	3.5
2.5	3	3.5	4

----- Media de las muestras

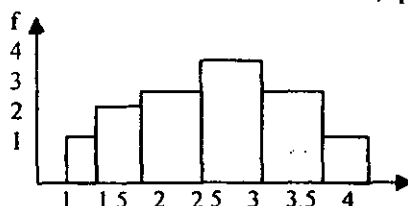
Sin Reemplazo

Comparando los Histogramas de la población y de la muestra:

a) Población donde se presenta distribución uniforme



b) distribución de medias de las muestras, aproximadamente normal:



Determinando la media de la distribución muestral ( media de las medias ) y la varianza de la misma:

$$\mu = \sum_{i=1}^n \frac{X_i}{N^n} = \frac{1+1.5+2+\dots+4}{16} = 2.5$$

$$\sigma_x^2 = \sum_{i=1}^N \frac{(x_i - \mu_x)^2}{N^n} = \frac{(1-2.5)^2 + 2(1.5-2.5)^2 + \dots + (4-2.5)^2}{16} = 0.625$$

Comparando los datos poblacionales y muestrales:

POBLACIÓN

$$\mu = 2.5$$

$$\sigma^2 = 1.25$$

MUESTRA

$$\mu_x = 2.5 \Rightarrow 2.5 = 2.5$$

$$\sigma_x^2 = 0.625 \Rightarrow 1.25/2^* = 0.625$$

\* Tamaño de la muestra

5.6). - PROBLEMAS RESUELTOS:

**\* MUESTREO ALEATORIO SIMPLE**

1) En una planta industrial se tiene el envasado de 15000 lts/hr de cierto producto. Dentro de ésta área de envasado se tiene una pérdida de 3700 envases sin producto por día, esto representa un 7% de pérdida en el área. Una de las posibles causas de la pérdida de envases puede ser que el material no este llegando a la planta con la cantidad señalada en las cajas. Cada caja contiene 300 envases, acomodados en dos filas.

Para determinar si las cajas contienen el número de piezas señaladas por el proveedor , se procede a realizar un muestreo aleatorio al azar del total de cajas que se encuentren en la planta.

Se sabe que en el almacén se tiene una existencia de 21 tarimas acomodadas de 3\*7, cada tarima contiene 40 cajas, las cuales se encuentran en 8 camas y cada cama contiene 5 cajas.

Solución:

Datos: población finita 840 cajas que se encuentran en la planta.

Muestra representativa del 5%= 42 cajas son las que se van a muestrear.

Primero es necesario enumerar las cajas por tarima, comenzando con el 001 hasta el 280, para saber que cajas se van a seleccionar de las 840.

Como se sabe el acomodo de las tarimas es de 3\*7, es decir, por cada fila hay 7 tarimas que representan 280 cajas, de las cuales es necesario tomar 14 cajas por fila.

Para saber que cajas son las que se van a seleccionar utilizamos las tablas de números aleatorios donde se va a localizar el punto de partida aleatoriamente.

La selección fue la siguiente: tomando la primera columna de la tabla en forma horizontal descendente se toman las tres últimas cifras de cada número (es importante mencionar que la cantidad de cifras seleccionadas está en función del número de dígitos que se tenga que muestrear, ejemplo 54441).

Por lo que se van a seleccionar los 14 primeros números menores de 280. Los números seleccionados son:

067, 187, 130, 208, 003, 032, 047, 166, 052, 090, 149, 064, 148, 260.

*Nota: todos los números seleccionados no se tienen que repetir.*

Se hace lo mismo hasta completar la muestra requerida. Una vez que se tiene la muestra seleccionada se realiza el tratamiento estadístico deseado.

**\* MUESTREO ESTRATIFICADO**

1) Un muestreador propone tomar una muestra aleatoria estratificada. Espera que sus costos de campo serán  $\sum C_i * n_i$ . Sus estimaciones adelantadas a las cantidades son como siguen:

Estrato	$N_i/N_i$	$S_i$	$C_i$
1	0.4	10	4
2	0.6	20	9

a) Encuentre los valores de  $n_1/n$  y  $n_2/n$  que minimizan los costos de campo totales para un valor dado de  $\sigma^2 x$ .

Solución:

Datos:  $k=2$  estratos,  $n=1$ ,  $N_1=0.4$   $\sigma_1=10$   $\$=4$   
 $N_2=0.6$   $\sigma_2=20$   $\$=9$

Como lo que queremos realizar es minimizar los costos de campo utilizamos un muestreo estratificado óptimo.

Por lo tanto utilizamos la siguiente ecuación:  $n_i = \frac{N_i \sigma_i / C_i}{\sum N_i \sigma_i / C_i} n$

$$N_1 \sigma_1 / C_1 = (0.4)(10)/4 = 1$$

$$N_2 \sigma_2 / C_2 = (0.6)(20)/9 = 1.3/2.3$$

Sustituimos en la ecuación de muestreo estratificado óptimo:

$$n_1 = 1 / 2.3(1) = 0.43$$

$$n_2 = 1.3 / 2.3(1) = 0.56$$

Como podemos ver el estrato que minimiza los costos es el estrato 1, por lo que conveniente elegirlo.

2) En una población con  $N=6$  y  $N=2$  los valores son 0,1,2 en el estrato 1 y 4,6,11 en el estrato 2. Se va a tomar una muestra con  $n=4$ .

a) demostrar que las  $n_i$  óptimas para la asignación de Neyman, aproximando a números enteros, son  $n_i=1$  en el estrato 1 y  $n_i=3$  en el estrato 2.

Solución:

$N=6$  Estrato 1:  $N_1=0$ ,  $N_2=1$ ,  $N_3=2$

$K=2$  Estrato 2:  $N_1=1$ ,  $N_2=4$ ,  $N_3=6$

$n=4$

Sustituimos los datos en la ecuación de Neyman  $n_i = \frac{N_i \sigma_i / C_i}{\sum N_i \sigma_i / C_i} n$

$$n_{i1} = 0.868 \approx 1 \quad n_{i2} = 3.13 \approx 3$$

$$\sigma_{i1} = 0.816 \quad \sigma_{i2} = 2.943$$

$$N_i \sigma_{i1} = 20.448 \quad N_i \sigma_{i2} = 8.829 \quad \sum = N_i \sigma = 11.277$$

$$n-1 \Rightarrow \sigma_{i1} = 1 \quad \sigma_{i2} = 3.605$$

## VI.- TEOREMA DEL LIMITE CENTRAL

El Teorema de Límite Central (TLC) además de ser una distribución de probabilidad importante por si misma, la distribución normal se puede usar frecuentemente para aproximar las distribuciones de otras variables aleatorias como consecuencia del TLC.

El TLC establece que, bajo condiciones bastante generales las sumas y medias de muestras, de observaciones aleatorias extraídas de una población tienden a poseer, aproximadamente una distribución acampanada cuando se repite el muestreo muchas veces.

“Si se seleccionan al azar, muestras de  $n$  observaciones de una población con media finita  $\mu$  y  $\sigma$  entonces, cuando  $n$  es grande, la media muestral tienen una distribución aproximadamente normal con media igual a  $\mu$  y desviación estándar  $\frac{\sigma}{\sqrt{n}}$  la aproximación se vuelve más precisa a medida que  $n$  aumenta. (28)

El TLC nos dice que la media y la desviación estándar de la distribución de las medias muestrales están ciertamente relacionadas con la media y desviación estándar de la población de donde se obtiene la muestra, tanto como con el tamaño de la muestra ( $n$ ) las dos distribuciones tienen la misma media  $\mu$  y la desviación estándar de la distribución de las medias muestrales es igual a la desviación de la población dividida entre  $\sqrt{n}$ .

Otra contribución del TLC y la más importante es la inferencia estadística. Muchos estimadores o estadísticos que se usan para hacer inferencias acerca de los parámetros de una población son sumas o promedios de las observaciones muestrales. Cuando esto es cierto y cuando el tamaño de la muestra  $n$  es grande podemos esperar que el estimador o estadístico posea una distribución normal cuando se repite el muestreo muchas veces.

Una característica del *Teorema de Limite Central* es que debemos tener alguna idea de qué tamaño  $n$  debe tener la muestra, para que la aproximación de resultados útiles. Desgraciadamente no hay una respuesta precisa a esta pregunta, ya que el valor apropiado de  $n$  depende tanto de la distribución de probabilidad de la población como del uso que se va a hacer de la aproximación.

En general, el teorema de límite central funciona muy bien aun para muestras pequeñas. Pero tomando en cuenta que se trabaja con poblaciones finitas sin reemplazo tenemos:

$$\mu_x = N \frac{\sum_{i=1}^N x_i}{C_n} \text{ donde } \Rightarrow C_n^N = \frac{N!}{n!(N-n)!}$$

$$\sigma_x^2 = N \frac{\sum_{i=1}^N (x_i - \mu_x)^2}{C_n^N}$$

Por lo tanto se requiere aplicar el factor de corrección para población finita,  $\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)$

que puede ignorarse cuando la muestra o el tamaño de ella es pequeño. (2)

Para determinar el tamaño de la muestra es necesario entender la estimación por intervalo de confianza que se basa en:

$$\mu = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ donde } \Rightarrow z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ es el margen de error}$$

$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ despejando } n$$

$$n = \frac{z^2 \sigma^2}{e^2} \quad \text{Muestreo con reemplazo de poblaciones finitas}$$

$$e = \frac{z\sigma}{\sqrt{n}} \left[ \sqrt{\frac{N-n}{N-1}} \right] \quad \text{Factor de corrección}$$

$$n = \frac{N z^2 \sigma^2}{e^2 (N-1) + z^2 \sigma^2} \quad \text{Muestreo sin reemplazo de poblaciones finitas}$$

$e$  = Cantidad de error (+/-) que se está dispuesto a aceptar en el uso estadístico muestral para estimar el parámetro de la población. (23)

Cuando se desconoce  $\sigma^2$  se recomienda:

- 1.- Extraer muestras piloto y puede usarse  $S^2$  calculada a partir de la muestra como estimación de  $\sigma^2$ .
- 2.- Puede contarse con estimaciones de  $\sigma^2$  obtenidas de estudios previos.
- 3.- Si se sabe que la población se distribuye normalmente:

$$\sigma \approx \frac{R}{6} \quad \text{donde } R = \text{rango}$$

### 6.1).- PROBLEMAS RESUELTOS:

1) Para evitar dificultades con las agencias de protección al consumidor, una embotelladora de bebidas debe estar razonablemente segura de que las botellas de 12 onzas realmente contengan 12 onzas de bebida. Para inferir si una maquina embotelladora está trabajando satisfactoriamente, se seleccionan 10 botellas al azar cada hora y se mide la cantidad de bebida en cada botella. La media de las 10 mediciones se usa para decidir si se debe o no reajustar la cantidad de bebida por botella dispensada por la máquina. Si la experiencia pasada demuestra que la cantidad de bebida por botella tiene una desviación estándar de 0.2 onzas y si la máquina está ajustada para producir una cantidad media por botella de 12.1 onzas ¿cual es la probabilidad de que la media muestral de las 10 botellas seleccionadas sea menor que 12 onzas?



Solución:

La media de distribución de probabilidad de la media muestral  $\bar{x}$  es idéntica a la muestra de la población formada por el contenido de las botellas, es decir,  $\mu = 12.1$  onzas y la desviación estándar de la distribución de las medias muestrales, denotada por el símbolo  $\sigma_{\bar{x}}$  es

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{10}} = 0.063$$

Donde  $\sigma$  es la desviación estándar de la población formada por el contenido de las botellas y  $n$  es el número de botellas en la muestra.

Aun cuando  $n$  es tan pequeño como 10, es probable, para este tipo de datos, que la distribución de probabilidad de  $\bar{x}$  sea aproximadamente normal, debido al TLC. Entonces la distribución de probabilidad de  $\bar{x}$  tendrá la forma mostrada en la figura 6.1. La probabilidad de que la media sea menor que 12 onzas es aproximadamente igual al área sombreada debajo de la curva normal.

Figura 6.1



Por lo tanto esta área es igual a  $(.5 - A)$ , donde  $A$  es el área entre 12 y la media  $\mu = 12.1$ , expresando esta distancia en términos de  $z$ :

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{12 - 12.1}{0.063} = -1.59$$

Hay que observar que debemos usar  $\sigma_{\bar{x}}$  y no  $\sigma$  en la fórmula para  $z$  porque estamos calculando un área bajo la distribución de probabilidad de  $\bar{x}$ , y no bajo la distribución de probabilidad de  $x$ .

Entonces, el área  $A$  sobre el intervalo  $12 \leq \bar{x} \leq 12.1$  la buscamos en la tabla de distribución normal (ver APÉNDICE) y encontramos que es igual a .4441 por lo que la probabilidad de que  $\bar{x}$  sea menor que 12 onzas es:

$$\begin{aligned} P(\bar{x} < 12) &= .5 - A = .5 - .4441 \\ &= .0559 \\ &\approx .056 \end{aligned}$$

En otras palabras, si la máquina está ajustada para producir un contenido medio de 12.1 onzas, la media de la muestra de 10 botellas es menor que 12 onzas con probabilidad aproximadamente igual a .056.

Cuando ocurre esta señal de peligro se debe tomar una muestra mayor para verificar el ajuste de la máquina. Nótese que el TLC juega un rol en la solución de este problema porque justifica la normalidad aproximada de la distribución de probabilidad de la media muestral.

2) Extraigamos todas las muestras de tamaño 3 con reemplazo de la población 5, 7, 8, 10 cuya media es  $\mu = 7.50$  y varianza  $\sigma^2 = 3.25$  y obtengamos las distribuciones muestrales de la media, de la mediana y de la varianza.

### Solución:

El número de muestras de tamaño 3 con reemplazo es  $4^3 = 64$ , después de ordenar las medias, medianas y varianzas de todas las muestras, tenemos las distribuciones muestrales siguientes:

X	f(X)	Md	f(Md)	S <sup>2</sup>	f(S <sup>2</sup> )
5	1/64	5	10/64	0	4/64
5.6667	3/64	7	22/64	0.3333	6/64
6	3/64	8	22/64	1.3333	12/64
6.3333	3/64	10	10/64	2.3333	12/64
6.6667	9/64		64/64	3	12/64
7	4/64			6.3333	12/64
7.3333	9/64			8.3333	6/64
7.6667	9/64				64/64
8	4/64				
8.3333	9/64				
8.6667	3/64				
9	3/64				
9.3333	3/64				
10	1/64				
	64/64				

De las 3 distribuciones obtenidas tenemos que:

a)  $\mu_x = 7.50$

$$\sigma_x^2 = 1.0833 = \frac{3.25}{3} = \frac{\sigma^2}{n}$$

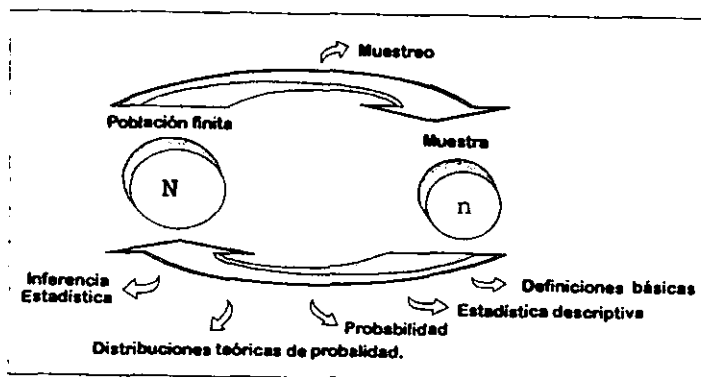
b)  $\mu_{Md} = 7.50$

$$\sigma_{Md}^2 = 2.13 > 1.0833 = \sigma_x^2$$

c)  $\mu_{S^2} = 3.25 = \sigma^2$

$$\sigma_{S^2}^2 = 6.52$$

## VII.- INFERENCIA ESTADÍSTICA O ESTIMACIÓN



El término *ESTIMACIÓN* se encuentra definido como el proceso de inferir valores de alguna característica de la población sobre la base de la información contenida en una muestra aleatoria.

Un *Estimador* es una regla que nos dice cómo calcular la estimación basándose en la información contenida en la muestra, generalmente se expresa como una fórmula, por ejemplo la media muestral  $\bar{x}$  que es un estimador de la media poblacional  $\mu$ , y explica exactamente como se puede obtener el valor numérico de la estimación, una vez que se conocen los valores de la muestra.<sup>(3)</sup> Se dice que un estimador es imparcial, si el valor esperado del estimador es igual a la cantidad de la población en estimación.<sup>(2)</sup>

La diferencia entre una estimación particular y el parámetro que se está estimando se llama error de estimación podemos considerar a  $2\sigma_x$  como una cota aproximada de este error, esto significa que al menos  $\frac{3}{4}$  de las estimaciones y muy probablemente el 95% de ellas, se alejan de la media una distancia menor que  $2\sigma_x$ .

Estimador puntual de la media de una población:

Estimador:  $\bar{x}$

$$\text{Cota de error: } 2\sigma_x = \frac{2\sigma}{\sqrt{n}}$$

Existen diferentes tipos de estimadores y son los siguientes:

- a) **PUNTUALES.**- Es sólo un valor numérico utilizado para estimar el parámetro correspondiente a la población ( $\bar{x}$ ,  $s$ ). Por ejemplo se desea estimar el promedio de un estudiante de Universidad, la estimación podría ser un número como 8.9, por que el número se puede asociar con un punto en una línea recta.<sup>(3)</sup>
- b) **ESTIMADORES POR INTERVALO.**- Son dos valores numéricos que definen un intervalo, el cual con algún grado de confianza, considera que se incluya al parámetro que se está estimando.

### 7.1).- CARACTERÍSTICAS DE UN BUEN ESTIMADOR PUNTUAL

Puede estimarse un parámetro por más de un estimador, por ejemplo, para estimar la media de una población puede usarse a la media, a la mediana de la muestra o a la moda, por lo tanto, ¿Cómo puede seleccionarse un buen estimador?, cuando éste cumple con las siguientes características:

1. Insesgabilidad: Se puede decir que un estimador  $\hat{\theta}$  es un estimador insesgado de  $\theta$  si el valor esperado de  $\hat{\theta}$  es igual a  $\theta$ .  $E(\hat{\theta}) = \theta$ . ( Para la media, cuando tomamos muchas muestras,  $\bar{x}$  es un estadístico insesgado). En promedio el valor de la estimación debe ser igual al parámetro poblacional que se está estimando. En condiciones ideales, el sesgo de un estimador es igual a cero = insesgado, se dice que un estimador es insesgado si su valor esperado es igual al verdadero valor del parámetro que se está estimando. Esto es  $E[\text{estimador}] = \text{parámetro poblacional}$ . Un inconveniente de la propiedad de insesgamiento consiste en que el criterio únicamente requiere que sea el valor promedio del estimador lo que iguale al parámetro poblacional. No se precisa que la mayor parte o incluso cualquiera de los valores del estimador, esté razonablemente próximo al parámetro poblacional, como será deseable un buen estimador, por esta razón es importante la propiedad de eficiencia. (8)
2. Eficiencia: Se dice que un estimador  $\hat{\theta}_1$ , es más eficiente que otro estimador  $\hat{\theta}_2$  para  $\theta$  si el primero tiene menor varianza que el segundo, por lo tanto cuando menor es la varianza, tanto más concentrada es la distribución del estimador alrededor de su propia media tanto mejor será el estimador siempre que sea insesgado: (12)

$$E(\hat{\theta} - E(\hat{\theta}))^2 < E(\hat{\theta}^\Delta - E(\hat{\theta}^\Delta))^2 \quad \text{donde } \hat{\theta}^\Delta \text{ es otro estimador de } \theta.$$

3. Consistencia: Generalmente un estimador no es idéntico al parámetro debido al error de muestreo que es la diferencia  $(\hat{\theta} - \theta)$ , por lo tanto la consistencia se refiere a que le estimador se acerca más al valor del parámetro al crecer el tamaño de la muestra: (12)

$$E(\hat{\theta} - \theta)^2 \rightarrow 0 \text{ cuando "n"} \rightarrow \infty$$

4. Suficiencia: Si se transmite tanta información de la muestra como es posible acerca del parámetro, de modo que no se verá proporcionada más información para cualquier otro estimador calculado a partir de la misma muestra, es decir se dice que un estimador es suficiente, si incluye toda la información que la muestra puede proporcionar acerca del parámetro poblacional. (13)

### 7.2).- ESTIMACIÓN POR INTERVALOS DE CONFIANZA

La probabilidad de que un intervalo de confianza incluya el parámetro estimado se llama *coeficiente de confianza*.

Supóngase que se desea estimar la media de alguna población normalmente distribuida, por lo tanto se extrae una muestra aleatoria de tamaño "n" de la población y se calcula, la cual es una estimación puntual de  $\mu$ . Aunque  $\bar{x}$  posee todas las cualidades de un buen estimador de  $\mu$  no puede esperarse que  $\bar{x}$  sea igual a  $\mu$ , por esto es más útil estimar a  $\Psi\forall$  mediante un intervalo de confianza, ya que esta estimación poseerá más información ya que marca la precisión y el nivel de confianza con el cual se asegura que el parámetro a estimar está contenido en ese intervalo.

Se puede encontrar un intervalo de confianza para la media de una población, es posible que  $\bar{x}$  sea mayor o menor que la media de la población, aunque esperamos que no se aleje de la media  $\mu$  más de aproximadamente  $2\sigma_x$  por lo tanto si tomamos  $(\bar{x} - 2\sigma_x)$  como el extremo inferior del intervalo llamado límite de confianza inferior y  $(\bar{x} + 2\sigma_x)$  como el extremo superior llamado límite de confianza superior muy probablemente el intervalo contendrá a la verdadera media de la población  $\mu$ . De hecho, si n es grande y si la distribución de x es aproximadamente normal, esperaríamos que el 95% de los intervalos obtenidos al repetir el muestreo contengan a la media de la población, el intervalo de confianza descrito se llama "intervalo de confianza para muestras grandes" porque n debe ser suficientemente grande para que el TLC sea aplicable y por lo tanto para que la distribución de  $\bar{x}$  sea aproximadamente normal.

Al establecer el intervalo, dentro del cual suponemos que cae la media de la población, no se ha establecido ninguna probabilidad de que la media obtenida sea la correcta, estas afirmaciones son válidas solamente respecto al intervalo y no con relación a cualquier valor particular de la media de la muestra. Además puesto que la media de la población es un valor fijo y no tiene una distribución, las declaraciones de probabilidad nunca se refieren a  $\mu$ , se afirma la probabilidad con respecto al intervalo, o sea, la probabilidad de que el intervalo contenga  $\mu$ .<sup>(5)</sup>

Intervalos de confianza para la media de la población  $\mu$  en el caso de muestras grandes.<sup>(3)</sup>

$$\bar{x} \pm \frac{z\alpha/2\sigma}{\sqrt{n}}$$

### 7.3).- PRECISIÓN Y CONFIANZA.

A medida que aumentamos el tamaño de la muestra  $\bar{x}$  se hace más precisa  $\mu$ .

Al aumentar la confianza, aseguramos que  $\mu$  caiga dentro de ese intervalo.

Pero al aumentar el intervalo se disminuye la precisión con igual tamaño de muestra (n).

En el muestreo de una población normalmente distribuida, la distribución muestral será normal con:

$$\begin{aligned} \text{Media } \bar{x} \text{ (media muestral)} &= \mu \text{ (media poblacional)} \\ \text{Varianza } \sigma_x^2 &= \sigma^2 / n \end{aligned}$$

Lo que se conoce de distribución normal es aplicable para  $\bar{x}$ . Por ejemplo si conocemos que aproximadamente el 95% de los valores posibles de  $\bar{x}$  que constituyen la distribución están entre los límites de 1.96 (localizando en tablas de distribución normal con el valor de 0.975) respecto a la media, por lo tanto el intervalo quedará:

$$\mu \pm 1.96 \sigma_{\bar{x}}$$

Que contendrá el 95% de los valores de  $\bar{x}$ .

Como se conoce el valor de  $\mu$ , no se logra mucho por medio de la expresión, sin embargo se tiene una estimación del punto  $\mu$ . Por lo tanto si se construyeran intervalos alrededor de todo el valor posible de  $\bar{x}$ , calculados a partir de todas las muestras de tamaño  $n$  de la población de interés, se tendría un número grande de intervalos de la forma:

$$\bar{x} \pm 1.96 \sigma_{\bar{x}}$$

con anchos todos iguales al ancho del intervalo de  $\mu$  desconocido. El 95% de los intervalos contendrá a  $\mu$ .

Una estimación de intervalo para  $\mu$  queda:

$$\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$$

donde  $z_{\alpha/2} \sigma_{\bar{x}}$  es el margen de error.

Si designamos el área total bajo la curva de  $\bar{x}$  que queda fuera del intervalo de confianza deseado como  $\alpha$ , entonces el área dentro del intervalo será  $1-\alpha$  y puede darse la siguiente interpretación al intervalo de confianza:

-**Probabilística:** En el muestreo repetido, a partir de una población normalmente distribuida, 100(1- $\alpha$ ) por ciento de todos los intervalos de la forma  $\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$  incluirían, a la larga, a la media de la población.

-**Práctica:** En el muestreo repetido, a partir de una población normalmente distribuida, se tiene 100(1- $\alpha$ ) por ciento de confianza que el intervalo único calculado,  $\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$ , contenga la media de la población.

Sin embargo, no siempre la población se distribuye normalmente pero se puede aplicar el **teorema de límite central** cuando "n" es grande y cuando "n" es pequeña usamos distribución **t-Student**. (23)

#### 7.4).- DISTRIBUCIÓN t DE "STUDENT"

Para construir el intervalo de confianza se requiere conocer la variación de la población, de la cual se extraen las muestras, pero puede darse el caso que se desconozca tanto la media como la varianza, cuando ésto suceda usaremos la **desviación estándar** de la muestra para reemplazar a  $\sigma^2$  cuando "n" es grande ( $n > 25$ ) nuestra confianza es s, como una aproximación de  $\sigma$ .

Pero cuando el tamaño de muestra es pequeño se utiliza un procedimiento alternativo para la construcción de los intervalos, se tiene la **distribución t-Student** que recibe este nombre en honor a su descubridor Gosset, que publicó sus trabajos con el pseudónimo de "Student" a principios del siglo XX.

Sea el estadístico  $t = \frac{(\bar{x} - \mu)}{(s/\sqrt{n})}$  que es análogo al estadístico z

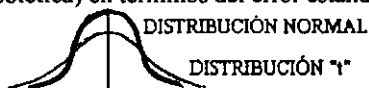
Los ensayos de hipótesis y significación se extienden fácilmente al caso de pequeñas muestras, siendo la única diferencia que el valor de z se sustituye por un adecuado valor de t.

*Medias.* - para ensayar la  $H_0$  de que una población normal tiene de media  $\mu$ , se utiliza el valor de t

$$t = \frac{\bar{x} - \mu}{s} = \frac{\bar{x} - \mu}{\hat{s}} \sqrt{N-1} = \frac{\bar{x} - \mu}{\hat{s}} \sqrt{N} \quad \text{donde } \bar{x} \text{ es la media de una muestra de tamaño } N$$

Esto es análogo al utilizar el valor de z:  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$  para N grandes. Excepto que se utiliza  $\hat{S} = S\sqrt{N/(N-1)}$  en lugar de  $\sigma$ , se advierte que la aproximación de  $\hat{S}$  a  $\sigma$  es pobre sobre todo para las muestras pequeñas, esta falta de aproximación se debe al hecho de que con muestras pequeñas  $\hat{S}$  tenderá a subestimar a la  $\sigma$  más de la mitad de las veces, en consecuencia el estadístico tenderá a estar más espaciado hacia afuera que la distribución normal. La diferencia estriba en que mientras z se distribuye normalmente, t sigue una distribución de Student, cuando aumenta N ambas tienden a coincidir.

El estadístico t es similar en muchos aspectos a la distribución normal ambos están expresados como la desviación de la media de la muestra con respecto a la media de la población (conocida o hipotética) en términos del error estándar de la media.



Para el estadístico existe una familia de distribución que varía como función de los grados de libertad (g.l). (5)

Donde el término "grados de libertad" se refiere al número de valores que están libres para variar después de haber impuesto ciertas restricciones a nuestros datos (N-1).

Existen muchas distribuciones t, una para cada tamaño de la muestra; por consiguiente una para cada g.l. Entre mayor es el tamaño de la muestra, menos dispersa es la distribución t y más confiable es la S estimada y en consecuencia menos variable es t. Por último a medida que el tamaño de la muestra tiende al infinito,  $S^2$  corresponde por completo a  $\sigma^2$  y, entonces la distribución t coincide con la normal. Pero en todos los demás casos, los valores de t son mayores que los de z. (9)

### Características de las Distribuciones t.<sup>(12)</sup>

- Media igual a cero.
- Simétrica con respecto a una media igual a cero.
- Tiene varianza mayor a 1, pero tiende a 1 cuando  $n$  aumenta.
- Toma valores de  $-\infty$  a  $\infty$ .
- En realidad es una familia de distribuciones, ya que se tiene una distribución diferente para cada valor de  $n-1$  (**grados de libertad**).
- Sin embargo mientras mayor sean los grados de libertad, mayor será el parecido de las distribuciones  $t$  y la curva normal
- Comparada con la distribución normal,  $t$  es menos alta en el centro y tiene colas más altas.

### USO DE LA TABLA t-Student.

En contraste con nuestro uso de la curva normal, los valores tabulados son valores críticos, o sea, aquellos valores que limitan las zonas críticas de rechazo correspondientes a diferentes niveles de significancia. Así empleando la tabla para las distribuciones  $t$ , localizamos el número adecuado de grados de libertad en la primera columna del lado izquierdo, y en la fila superior localizamos la  $\alpha$  elegido.

Los valores tabulados representan el cociente  $t$  requerido para la significación, si se obtiene un cociente  $t$  igual o que exceda estos valores tabulados, debemos rechazar  $H_0$  (hipótesis nula).

- Seleccionar la columna con la confianza deseada.
- Localizar el renglón con los grados de libertad a utilizar.
- La intersección de columna y renglón mostrará el área bajo la curva deseada. (ver apéndice para el uso de la Tabla de Distribución  $t$ -Student).

A continuación se muestra en la Tabla 4.1.1 la diferencia entre la simbología de un estimador y un parámetro:

Tabla 4.1.1

Estimador	Parámetro
$\bar{x}$ = Media muestral	$\mu$ = Media poblacional
$s^2$ = Varianza muestral	$\sigma^2$ = Varianza poblacional
$s$ = Desviación estándar	$\sigma$ = Desviación estándar. Poblacional



A continuación se presenta un cuadro resumen de las formulas que son utilizadas para realizar los cálculos de la distribución t-student.

RESUMEN DE LAS FÓRMULAS UTILIZADAS PARA LA  
DISTRIBUCIÓN T-STUDENT

Intervalos de confianza para la media de la población en el caso de muestras grandes (mayor de 25 elementos)	$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$
Intervalos de confianza para la media de la población en el caso de muestras pequeñas (menor de 25 elementos)	$\bar{x} - \frac{\sigma}{\sqrt{n}} t_{/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} t_{/2}$
Varianza poblacional	$\sigma_x^2 = \sigma^2 / n$
Estimación de intervalo para la media poblacional	$\mu = \bar{x} \pm z_{\alpha/2} \sigma$
Distribución t de Student	$t = \frac{(\bar{x} - \mu)}{S/\sqrt{n}}$
Media de t de Student	$t = \frac{\bar{x} - \mu}{\hat{S}} \sqrt{N}$
Aproximación de $\hat{S}$ a $\sigma$	$\hat{S} = S \sqrt{N/(N-1)}$

### 7.5).- PROBLEMAS RESUELTOS

Supongamos que se desea estimar la producción diaria de un producto elaborado en una planta industrial, la producción diaria para  $n = 50$  días, tuvo media y desviación estándar iguales a:  $\bar{y} = 871$  Ton ;  $S = 21$  ton. Estimar la producción diaria promedio  $\mu$ .

#### Solución

La estimación de la producción diaria es  $\bar{y} = 871$  Ton, la cota del error de estimación es

$2\sigma_y = \frac{2\sigma}{\sqrt{n}} = \frac{2\sigma}{\sqrt{50}}$  Aunque  $\sigma$  es desconocida, podemos aproximar su valor usando el estimador  $S$  del parámetro  $\sigma$ . Así la cota de error de estimación es aproximada.

$$\frac{2S}{\sqrt{n}} = \frac{2(21)}{\sqrt{50}} = \frac{42}{7.07} = 5.94$$

Podemos sentirnos bastante confiados en que nuestra estimación de 871 Ton se aleja menos de 5.94 Ton. de la verdadera producción promedio.

2) Se quiere determinar la eficiencia de un secador por aspersión, al cual se alimento durante 28 días 30 lt de leche, el promedio de leche deshidratada era de 8.25 kg con una desviación estándar de 0.6 kg.

- cual será la estimación de la media de la población para la cantidad de leche deshidratada obtenida durante los 28 días con una confianza del 90, 95 y 99%
- si solamente se hubiesen trabajado 5 días cuales serían estos intervalos.

#### Solución:

Datos:  $\bar{x} = 8.25$ kg

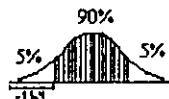
$$\sigma = 0.6$$
kg

- $n = 28$ , intervalos de confianza del 90, 95 y 99%, como tenemos una muestra de

mayor de 25 elementos utilizamos la siguiente ecuación:  $\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$

y utilizamos la tabla de distribución normal

- Intervalo de confianza del 90%.

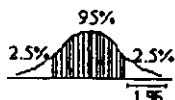


$$8.25 - \frac{0.6}{\sqrt{28}} (1.64) \leq \mu \leq 8.25 + \frac{0.6}{\sqrt{28}} (1.64)$$

$$8.06 \leq \mu \leq 8.43$$

∴ se asume que con un 90% de confianza, la cantidad de leche en polvo obtenida será de 8.06 a 8.43 en donde se contendría la media de la población.

- Intervalo de confianza del 95%



$$8.25 - \frac{0.6}{\sqrt{28}}(1.96) \leq \mu \leq 8.25 + \frac{0.6}{\sqrt{28}}(1.96)$$

$$8.02 \leq \mu \leq 8.47$$

- Intervalo de confianza del 99%



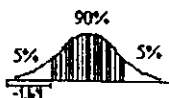
$$8.25 - \frac{0.6}{\sqrt{28}}(2.58) \leq \mu \leq 8.25 + \frac{0.6}{\sqrt{28}}(2.58)$$

$$7.95 \leq \mu \leq 8.54$$

Como puede observarse, al abrir la confianza, los contenidos promedio, caerán en intervalos mayores, manejando el mismo tamaño de muestra.

b)  $n = 5$  intervalos de confianza del 90, 95 y 99%, como tenemos una muestra de menor de 25 elementos utilizamos la siguiente ecuación:  $\bar{x} - \frac{\sigma}{\sqrt{n}}t_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}}t_{\alpha/2}$  y utilizamos la tabla t - Student

- Intervalo de confianza del 90%



Con grados de libertad de 1000 y una  $t_{0.050}$  unimos la fila superior y la columna y obtenemos el dato  $-1.646$  sustituimos en la ecuación y tenemos:

$$7.80 \leq \mu \leq 8.69$$

- Intervalo de confianza del 95%

$$7.72 \leq \mu \leq 8.77$$

Dentro de este intervalo, al 95%, se asume que el valor medio de la población será contenido entre 7.72 y 8.77, que al comparar con tamaño de muestra mayor éste, a su vez será mayor.

- Intervalo de confianza del 99%  
 $7.55 \leq \mu \leq 8.94$

3) En una industria se tienen la producción diaria de harina para pastel, el jefe encontró que en un día "x" la cantidad promedio de harina fue de 1362 kg por caja, para comprobarlo toma una muestra de 10 cajas y se determinó que su peso era de 1360 kg con una desviación estándar de 5 gr por lo tanto supuso que la producción reportada era correcta. ¿Que opina usted?  $\alpha = 5\%$

Solución:

Datos:  $\bar{x} = 1362$  kg por caja

$$n = 10$$

$$\sigma = 5 \text{ gr}$$

$$\alpha = 5\% \text{ y una } \mu = 1360 \text{ kg}$$

Establecemos las hipótesis a comprobar:  $H_0: \mu > 1360$

$$H_1: \mu \leq 1360$$

$$1 - \alpha$$

$1 - 0.05 = .95$ , por lo tanto buscamos en la tabla de t – Student el área de la curva del .05 que será la zona de rechazo.



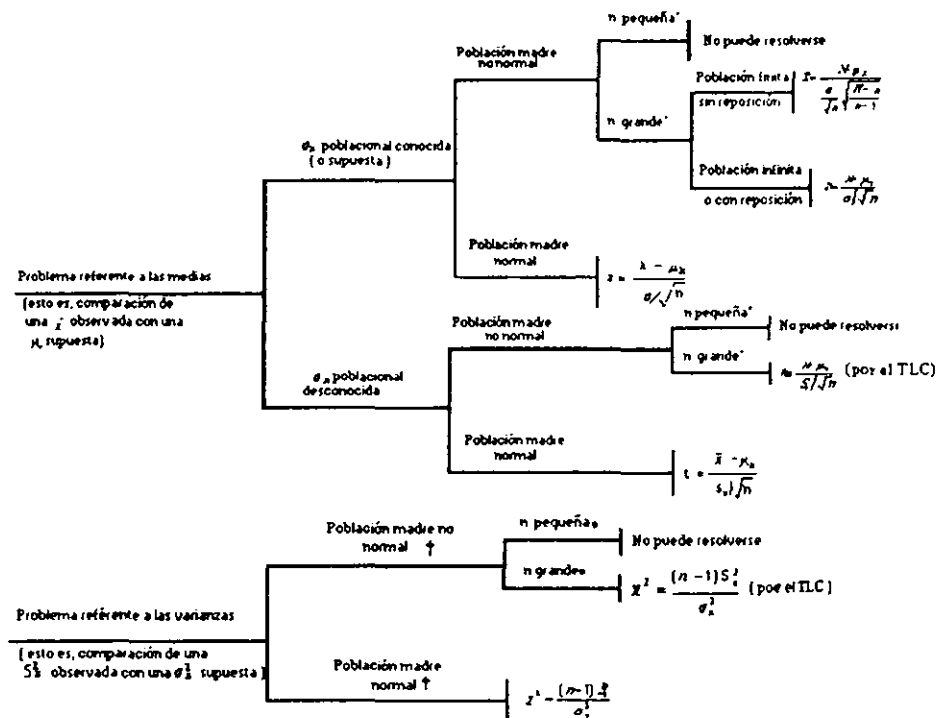
Por lo tanto si  $t < t_{\alpha}$  se acepta la  $H_0$ , donde  $t_{\alpha} = 1.646$  (valor obtenido de la tabla)

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{1362 - 1360}{5 / \sqrt{10}} = 1.26$$

$1.26 < 1.646$  aceptamos la hipótesis nula, es decir la cantidad de harina en promedio dentro de las cajas es mayor a 1360kg y su producción diaria es correcta.

## RECAPITULACIÓN DE LOS TEMAS V, VI, VII

*Diagrama de árbol para el uso de las distribuciones probabilísticas*



\* Una muestra se considera "grande" o "pequeña" según la exactitud deseada y la forma de la población donde proviene. Cuando se trata de una distribución simétrica, puede bastar que  $n \geq 10$ . En los demás casos, cuando  $n \geq 30$  puede considerarse que la muestra es "grande".

† Recuérdese que la  $t$  y la  $\chi^2$  son bastante "robustas" lo cual significa que con ellas se obtienen resultados suficientemente precisos, aun cuando la población no tenga una distribución exactamente normal, con tal de que presente algunas de sus características aunque sea en forma aproximada (unimodal y sin sesgo exagerado).

## Resumen de las distribuciones muestrales

Variable aleatoria	Situación	Distribución resultante para la resolución de problemas	Media	Varianza
$\bar{X}$	Población normal, $\sigma$ conocida, tamaño muestral $n$	$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	0	1
$\bar{X}$	Población normal, $\sigma$ desconocida, tamaño muestral $n^*$	$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$	0	$v(v-2)$ donde $v = n-1$
$\bar{X}$	Población desconocida, $\sigma$ conocida, $n$ "grande"	$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	0	1
$S^2$	Población normal, tamaño muestral $n$	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	$v$	$2v$ donde $v = n-1$

\* Si  $n$  es "grande",  $\bar{X}$  tendrá una distribución aproximadamente normal, y  $s$  tendrá un valor próximo a  $\sigma$ . Por tanto, la distribución de  $(\bar{X} - \mu) / (s / \sqrt{n})$  será aproximadamente  $N(0, 1)$ .

1. **diseño muestral:** plan en el cual se especifica la forma de obtener una muestra antes de haber recogido ningún dato.
2. **errores muestrales y no muestrales:** errores muestrales son los que se refieren al caso en que una muestra, a pesar de estar perfectamente diseñada, no representa exactamente a la población de la cual proviene. Los errores no muestrales son los humanos o evitables.
3. **número aleatorio:** número seleccionado de una población de modo que tenga la misma probabilidad de ocurrir que cualquier otro.
4. **muestreo probabilístico:** diseño muestral que se basa principalmente en una selección aleatoria, a esta categoría pertenecen:
  - **el muestreo aleatorio simple:** cada elemento y cada grupo de elementos tiene la misma probabilidad de pertenecer a la muestra.
  - **el muestreo sistemático:** a partir de un punto aleatorio se toman los elementos que ocupan el  $k$ -ésimo lugar.
  - **el muestreo estratificado:** los elementos se seleccionan de capas o estratos.
  - **el muestreo por conglomerados:** los elementos se seleccionan de grupos o conglomerados que tienen características similares.
  - **el muestreo secuencial:** los elementos muestrales no se seleccionan simultáneamente, sino de forma secuencial

5. *muestreo no probabilístico*: se basa principalmente en una selección no aleatoria. A esta categoría pertenecen:

- *el muestreo por cuotas*: cada una de las personas encargadas recolecta un número específico de datos.
- *el muestreo de conveniencia*: se toman los datos en donde se presenten, y de acuerdo a las conveniencias.
- *el muestreo de criterio*: lo fundamental es la apreciación de la persona que tiene a su cargo la responsabilidad del muestreo.

6. *estadístico muestral*: determinada característica de una muestra.

7. *media muestral*:  $\bar{x} = \frac{1}{n} \sum x_i f_i$

8. *varianza muestral*:  $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 f_i$

9. *distribución muestral*: la distribución de probabilidades de un estadístico muestral.

10. *valor esperado de  $\bar{x}$* :  $\mu_{\bar{x}} = \mu$

11. *varianza de  $\bar{x}$* :  $\sigma_{\bar{x}}^2 = \sigma^2/n$

12. *error estándar de la media*:  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

13. *estandarización de  $\bar{x}$* :  $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

19. *distribución de  $(n-1)s^2/\sigma^2$* : corresponde a una variable aleatoria que sigue una distribución ji-cuadrada con  $n-1$  grados de libertad.



### VIII) PRUEBA DE HIPÓTESIS

Las pruebas de significación o de hipótesis permiten verificar la veracidad de alguna hipótesis establecida acerca de una población, determinando si los valores difieren significativamente de los esperados por la hipótesis o si las diferencias observadas son debidas solo al azar. (12)

Una hipótesis es una suposición que se hace acerca de la distribución de una variable aleatoria. (4) Según Webster es “una teoría tentativa o suposición adoptada provisionalmente para explicar ciertos hechos y guiar la investigación de otros”. (2)

Una prueba estadística de una hipótesis es un procedimiento en el cual se usa una muestra con el fin de determinar cuando podemos “no rechazar” (“aceptar”) la hipótesis, es decir actuar como si fuera cierta, o cuando debemos “rechazar” ésta, es decir, actuar como si fuera falsa.

En las pruebas de hipótesis se distinguen la *Hipótesis nula*, designada por  $H_0$  conocida también como hipótesis de no diferencia, es cualquier hipótesis estadística que se establece en principio con el único propósito de rechazarla o anularla; y la *Hipótesis alternativa*,  $H_1$  que es cualquier suposición que difiera de la nula.

La hipótesis nula generalmente sé especifica en una forma opuesta a la que se supone cierta y la hipótesis nula y alternativa deben ser mutuamente exclusivas. (12)

Para dar una idea de la forma en que se trabaja esto, daremos un ejemplo típico:

Para el control de calidad en una fábrica, se toma una pequeña muestra de productos fabricados en cada hora y se mide cualquier cualidad en la que tengamos interés. Cualquiera de estas cualidades se debe considerar como una variable aleatoria, ya que ésta varía debido a un gran número de pequeños factores cuya influencia no se puede predecir. En muchos casos estos factores no se pueden numerar.

Supongamos que, por ejemplo, el diámetro debería ser  $\mu=5.08$  cm. Entonces se tiene que probar la hipótesis  $\mu=5.08$  cm. Mientras las medias de las muestras que se tomen no se desvíen “demasiado” del valor de  $\mu=5.08$  cm que se necesita, se permite que la producción continúe. Si la desviación es “demasiado grande”, se rechaza la hipótesis, se detiene la producción, y se comienza a buscar la fuente de la perturbación, tales desviaciones reciben el nombre de *desviaciones significantes*. Significa que se necesita alguna otra teoría que explique los resultados de los experimentos, y estas desviaciones están en contraste con las pequeñas desviaciones que son provocadas únicamente por factores aleatorios.

En muchos casos deseamos comparar dos cosas, por ejemplo, dos drogas diferentes, dos métodos para efectuar un cierto trabajo, la exactitud de dos métodos de medición, o la calidad de productos producidos por dos máquinas; dependiendo del resultado de una prueba conveniente, decidimos cual de las drogas se usa, se adopta el mejor método de trabajo etc.

Pero es necesario señalar que aunque hayamos aceptado o rechazado una hipótesis, no la hemos aprobado o desaprobado.

Al estudiar hipótesis, es conveniente clasificarlas en uno de dos tipos por medio de la siguiente definición

Si una hipótesis clasifica completamente la distribución, esto es, si especifica su forma funcional al igual que los valores de todos los parámetros, se le llama **hipótesis simple**; si no es así, se le llama **hipótesis compuesta**.

En el caso de una distribución normal, una hipótesis acerca de la media de la forma  $\mu = \mu_0$  ( $\mu_0$  un número dado) es simple si la variancia de la distribución se conoce, y es compuesta si no se conoce la variancia; en el último caso, la variancia queda sin especificarse, de manera que existe un número infinito de posibilidades para la distribución de la población; una hipótesis simple es aquella que especifica todos los parámetros de la distribución (es decir asigna un valor numérico a cada uno de ellos), mientras que una hipótesis compuesta no especifica todos estos parámetros. (4)

### 8.1) TIPOS DE ERRORES

Los procedimientos que facilitan el decidir si una hipótesis se acepta o se rechaza o el determinar si las muestras observadas difieren significativamente de los resultados esperados se llama *ensayos de hipótesis*. En el ensayo de hipótesis hay dos tipos de errores:

*Error tipo I.- Rechazo de una hipótesis que es verdadera ( $\alpha$ )*

*Error tipo II.- Aceptación de una hipótesis que es falsa ( $\beta$ )*

En el error tipo II, dejamos de rechazar la  $H_0$  cuando en realidad es falsa, este tipo de error es más frecuente que el error de tipo I. Esta probabilidad  $\beta$  depende de la alternativa  $\eta=1-\beta$

Se llama potencia de la prueba, es obvio que ésta es la probabilidad de evitar un error tipo II.

Tabla 8.1.1 Errores tipo I y II al probar una hipótesis  $\theta=\theta_0$  contra una alternativa  $\theta=\theta_1$

No Rechazada ( Aceptada)	Verdad Desconocida	
	$\theta=\theta_0$	$\theta=\theta_1$
$\theta=\theta_0$	Decisión verdadera $P = 1-\alpha$	Error tipo II $P = \beta$
$\theta=\theta_1$	Error tipo I $P = \alpha$	Decisión verdadera $P = 1-\beta$

Esta probabilidad ( $1- \beta$ ) se llama potencia de la dócima, puesto que la capacidad que tiene la dócima de reconocer correctamente que la hipótesis nula es falsa (y que por tanto debe rechazarse  $H_0$ ). Así siempre será deseable que la dócima tenga una potencia grande (cercana a uno) o, lo que es lo mismo, un valor pequeño de  $\beta$ , cuando  $H_0$  es falsa.

En la Tabla 8.1.2 se indican las probabilidades correspondientes a los cuatro casos.

Nótese que cada una de estas probabilidades es *condicional* y que la suma de las probabilidades de cada columna es igual a uno, ya que los sucesos con los que están

asociados son *complementarios*. La suma de  $\alpha$  y  $\beta$  no tiene que ser necesariamente igual a uno, ya que estas dos probabilidades no son complementarias, si se cambia el valor de  $\alpha$  en determinada cantidad, no tiene porqué producirse el mismo cambio en  $\beta$ , y viceversa.

Sin embargo  $\alpha$  y  $\beta$  no son independientes entre sí, y tampoco lo son del tamaño muestral  $n$ . Cuando disminuye el valor de  $\alpha$ , normalmente aumenta el de  $\beta$ , y viceversa (si el tamaño  $n$  permanece fijo). Cuando  $n$  crece, es posible que los valores de  $\alpha$  y  $\beta$  disminuyan, debido a que hay un decrecimiento potencial del error muestral, puesto que el incremento de  $n$  determina por lo general un gasto adicional

Tabla 8.1.2

	La verdadera situación puede ser:	
	Ho es verdadera	Ho es falsa
Se acepta Ho	$1 - \alpha$ (Nivel de confianza)	$\beta$ (beta)
Se rechaza Ho	$\alpha$ (alfa)	$1 - \beta$ (Potencia de la d6cima)
Suma	1.0	1.0

Debe quedar claro que las probabilidades de los errores tipo I y II son probabilidades condicionales ya que la primera depende de la condici6n de que  $H_0$  sea verdadera y la segunda de que  $H_1$  sea verdadera.

La probabilidad correspondiente a un intervalo de confianza es  $(1 - \alpha) =$  nivel de significancia y representa el complemento de  $P$  (error tipo I), l6gicamente al construir una d6cima, desearíamos que fuese peque1a la probabilidad de cometer un error de tipo I, por tanto es necesario que  $\alpha$  sea peque1a.

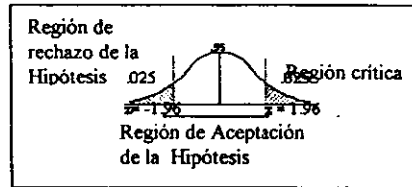
La hip6tesis que se quiere probar recibe el nombre de *hip6tesis nula* ( $H_0$ ) y una suposici6n contraria se llama *hip6tesis alternativa* ( $H_1$ ). Al n6mero  $\alpha$  (o  $100\alpha\%$ ) se le llama *nivel de significancia* de la prueba.

La regi6n que contiene los valores para los que se rechaza la hip6tesis se llama **regi6n de rechazo** o **regi6n cr6tica**. La regi6n de los valores en que no se rechaza la hip6tesis se llama **regi6n de aceptaci6n**. Una elecci6n frecuente para  $\alpha$  es 5%. (4)

La probabilidad  $\alpha$ , aumenta o disminuye dependiendo del tama1o de la regi6n de rechazo. Para un tama1o de muestra  $n$  fijo  $\alpha$  y  $\beta$  est1n en relaci6n inversa a medida que una aumenta la otra disminuye. Un aumento del tama1o de la muestra, proporciona m1s informaci6n para tomar la decisi6n y por tanto reduce tanto  $\alpha$  como  $\beta$  en una situaci6n experimental las probabilidades de los errores de tipo I y tipo II miden el riesgo de tomar una decisi6n incorrecta. (2)

Para que esto pueda quedar m1s claro se ejemplifica de la siguiente manera: como se indica en la figura N<sup>o</sup> 8.1.3 se puede estar con el 95% de confianza de que si la hip6tesis es cierta, el valor de  $z$  obtenido de una muestra real para el estadístico  $S$  se encontrara entre (-1.96 y 1.96).

Figura 8.1.3



El área total sombreada es 0.05 es el nivel de significación del ensayo. Representa la probabilidad de cometer error al rechazar la hipótesis, es decir, la probabilidad de cometer error tipo I, así pues se dice que la hipótesis se rechaza al nivel de significación del 0.05 o que la  $z$  obtenida del estadístico muestral dado es significativa al nivel de significación del 0.05.

En este ensayo se tienen los valores a ambos lados de la media por esta razón se llaman ensayos de dos colas ó bilaterales. Sin embargo se puede estar interesado en los valores extremos a un solo lado de la media es decir ensayos de una cola de la distribución, por ejemplo cuando se está ensayando la hipótesis de que un proceso es mejor que otro. (1)

La simbología utilizada es la siguiente:

1.- Medias = Aquí =  $S = \bar{x}$

Media muestral  $\mu_s = \mu \bar{x} = \mu$

Media poblacional  $\sigma_s = \sigma \bar{x} = \sigma / \sqrt{N}$  donde  $\sigma$  = desviación poblacional  
 $N$  = tamaño muestral.

por lo tanto el valor de  $z$  viene dado por  $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}}$

2.- Proporciones = Aquí =  $S = P$

$\rho$  = proporción de éxitos en la población

$N$  = tamaño de la muestra

$\sigma_s = \sigma_p = \sqrt{\rho q / N}$

$q = 1 - \rho$  proporción de fracasos

por lo tanto el valor de  $z$  viene dado por  $z = \frac{P - \rho}{\sqrt{\rho q / N}}$  (1)

3.- Nivel de significación ( $N_s$ )

- Cuando los sucesos ocurren 5% de las veces o menos ( por azar algunos investigadores afirman que los resultados se deben a factores ajenos al azar) su  $N_s = 0.05$
- Cuando los sucesos ocurren 1% o menos su  $N_s = 0.01$
- La Hipótesis Nula ( $H_0$ ): especifica valores hipotéticos para uno o más de los parámetros de la población, en nuestros días se utiliza principalmente para designar cualquier hipótesis formulada para ver si puede ser rechazada.
- La Hipótesis Alterna ( $H_1$ ): afirma que el parámetro de la población es algún valor diferente al hipotético

Un análisis cuidadoso de la lógica de la inferencia estadística revela que la  $H_0$  no puede probarse nunca ni tampoco podemos probar directamente la  $H_1$ . Sin embargo si podemos rechazar la  $H_0$ , o sea que el parámetro de la población es un valor diferente al valor hipotético.

#### ¿ Las condiciones para rechazar la $H_0$ ?

Si se está empleando el  $N_s = 0.05$ , se rechaza la  $H_0$  cuando un resultado dado ocurre por azar 5% del tiempo o menos, cuando se emplea el  $N_s = 0.01$  se rechaza la  $H_0$  cuando un resultado dado ocurre por azar el 1% ó menos de las veces y por la tanto se acepta la  $H_1$ . (5)

### 8.2) SECUENCIA DE APLICACIÓN

Para abordar problemas consistentes en probar hipótesis en manera sistemática sería útil proceder de la siguiente manera:

- 1.- Formulamos una hipótesis nula y una hipótesis alterna apropiada que será aceptada cuando la  $H_0$  debe ser rechazada.
- 2.- Estadística de prueba donde escogemos la muestra ( $n$ ) y se fija el nivel de significancia ( $\alpha$ ) que se usa para saber si la  $H_0$  es aceptada o rechazada.
- 3.- Establecer la estadística de prueba, calculo de  $z$  o  $t$  dependiendo del tamaño de la muestra.
- 4.- Especificamos la probabilidad de un error tipo I, si es posible, conveniente o necesario, podemos especificar también las probabilidades de errores de tipo II para alternativas particulares.

Después de que la  $H_0$ , la  $H_1$  y el  $\alpha$  han sido especificados las restantes etapas son:

5.- Con base en la distribución muestral de un estadístico apropiado, construimos un criterio para probar la  $H_0$  contra la alternativa determinada.

6.- Decidimos rechazar la  $H_0$ , aceptar la  $H_0$ , o bien nos abstenemos de tomar una decisión (aplicar regla de decisión) y dar una conclusión al problema planteado

#### Regla para Seleccionar la $H_0$ :

Cuando la meta de un experimento sea establecer una afirmación, la negación de la afirmación se debe tomar como  $H_0$ , la afirmación es entonces  $H_1$ .

**REGLA DE DECISIÓN:** Define las condiciones que llevan a la aceptación o rechazo de la  $H_0$  por comparación de los valores obtenidos de la prueba de hipótesis.

$H_0$	$H_1$	Rechazo	Aceptación
$\mu \leq \mu_0$	$\mu > \mu_0$	$Z \geq Z_{\alpha} \dots t \geq t_{\alpha}$	$Z < Z_{\alpha} \dots t < t_{\alpha}$
$\mu > \mu_0$	$\mu < \mu_0$	$Z \leq -Z_{\alpha} \dots t \leq -t_{\alpha}$	$Z > -Z_{\alpha} \dots t > -t_{\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z \geq Z_{\alpha/2} \dots t \geq t_{\alpha/2}$ $Z \leq -Z_{\alpha/2} \dots t \leq -t_{\alpha/2}$	$-Z_{\alpha/2} < Z < Z_{\alpha/2}$ $-t_{\alpha/2} < t < t_{\alpha/2}$

Cuadro resumen donde se muestran las regiones críticas para hipótesis alternas tanto unilaterales como bilaterales.

### 8.2.1) Pruebas sobre medias con variancia conocida

Hipótesis	Estadística de prueba	Criterios de rechazo
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$ Z_0  > Z_{\alpha/2}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$		$Z_0 < -Z_{\alpha}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$Z_0 > Z_{\alpha}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$ Z_0  > Z_{\alpha/2}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$		$Z_0 < -Z_{\alpha}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$		$Z_0 > Z_{\alpha}$

## 8.2.2) Pruebas sobre medias de distribuciones normales con variancia desconocida

Hipótesis	Estadística de prueba	Criterios de rechazo
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$ t_0  > t_{\alpha/2, n-1}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	$t_0 < -t_{\alpha, n-1}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_0 > t_{\alpha, n-1}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $v = n_1 + n_2 - 2$	$ t_0  > t_{\alpha/2, v}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 < -t_{\alpha, v}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$t_0 = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$	$t_0 > t_{\alpha, v}$

## 8.3).- PROBLEMAS RESUELTOS

1) Consideremos el estudio de la pasteurización de leche y la prueba estadística basada en la región de rechazo ( $y = 8, 9, 10$ )

- Enuncie la hipótesis nula y alterna para esta prueba
- Encuentre  $\alpha$  para esta prueba
- Encuentre  $\beta$

La probabilidad de aceptar la  $H_0$  es cuando la probabilidad de que un lote de leche no salga contaminado  $p = 0.9$

Solución:

a) La hipótesis nula establece que  $p = 0.5$ , en otras palabras que la pasteurización no es efectiva.

La hipótesis alterna es que  $p > 0.5$ , es decir que la pasteurización es efectiva

b) La probabilidad de rechazar la hipótesis nula siendo cierta es  $p = 0.5$  (*error tipo I*)

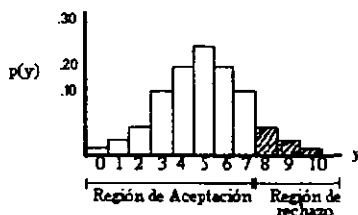
$$\alpha = P(y = 8, 9, 10 \text{ dado que } p = 0.5) = \sum_{y=8}^{10} p(y)$$

donde  $p(y)$  es la probabilidad de distribución Binomial con  $p = 0.5$ . Entonces usando la tabla de Distribución Binomial, tenemos que:

$$\alpha = \sum_{y=8}^{10} C_y^{10} (0.5)^y (0.5)^{10-y} = 0.055$$

La distribución de probabilidad para  $n = 10$  y  $p = 0.5$  se muestra en la figura 8.3.1 donde  $\alpha$  está representada por la porción sombreada de la distribución.

Fig.8.3.1



Distribución de Probabilidad Binomial con  $n = 10$  y  $p = 0.5$

c) Deseamos encontrar el valor de  $\beta$  cuando la hipótesis nula es falsa ( $p$  no es igual a 0.5) y en realidad  $p = 0.9$ . Cuando  $p = 0.9$  la distribución de probabilidad del número  $y$  de lotes de leche que no sufren ninguna contaminación aparece en la figura 8.3.2, puesto que:

$$\begin{aligned} \beta &= P(\text{aceptar la hipótesis nula cuando } p = 0.9) \\ &= P(y = 0, 1, 2, \dots, 6, 7 \text{ dado que } p = 0.9) \end{aligned}$$

entonces  $\beta$  es igual a la porción sombreada en la figura 8.3.2, es decir, en la porción que cae en la región de aceptación. Así:

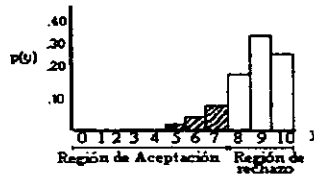
$$\beta = \sum_{y=0}^7 p(y) \text{ dado } p = 0.9$$

donde  $p(y)$  es la distribución Binomial, esta cantidad la obtenemos de la tabla de distribución Binomial (ver Apéndice)

$$\beta = \sum_{y=0}^7 C_y^{10} (0.9)^y (1)^{10-y} = 0.07$$



Fig.8.3.2



Distribución de probabilidad cuando la hipótesis alterna es cierta:  $n = 10$  y  $p = 0.9$

Para resumir las aplicaciones de los incisos a, b y c del ejercicio los valores  $\alpha = 0.055$  y  $\beta = 0.07$  dan medidas de los riesgos de cometer los dos tipos de errores para esta prueba estadística.

Donde la probabilidad de que caiga en rechazo, es cuando la hipótesis nula es cierta, la probabilidad de que la pasteurización es eficaz cuando en realidad no lo es, es solo el  $0.055 \approx 5.5\%$

Pero supongamos que la pasteurización es realmente eficaz y que la probabilidad de pasar el tiempo de conservación sin ninguna contaminación para un lote de leche es  $0.9$ , la probabilidad de aceptar la hipótesis nula que establece que "la pasteurización es eficaz" y el riesgo de cometer este error es solo  $\beta = 0.07$

2) Un fabricante de termopares anota el número de horas de servicio de 36 de ellos antes de fallar, encontró que su media es de 460 horas y su desviación estándar de 42, se puede decir validamente que la vida media de los termopares sobrepasa las 450hr con un nivel de significancia del 0.01.

### Solución:

Datos:  $\bar{x} = 460$ ,  $\sigma = 42$ ,  $\alpha = 0.01$

1.- Formulación de hipótesis:

$$H_0: \mu \leq 450$$

$$H_1: \mu > 450 \rightarrow \text{se desea comprobar}$$

2.-  $n = 36$

$$\alpha = 0.01 \quad 1 - \alpha = 0.99 \text{ zona de aceptación}$$



De tablas de distribución acumulativa Normal buscamos el área sombreada de la curva que es 0.99 y tenemos que  $z_\alpha = 2.33$

3.- Estadística de prueba 
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{460 - 450}{42 / \sqrt{36}} = 1.428$$

4.- Regla de decisión, si  $z < z_\alpha$  se acepta la hipótesis nula

$1.428 < 2.33$  por lo tanto se acepta la hipótesis nula y se rechaza la hipótesis alterna.

Llegando a una conclusión de que estadísticamente no hay una evidencia suficiente de que la vida media de los termopares no sobrepasa las 450 horas.

## CAPITULO SEGUNDO

Dentro de este capítulo, veremos tratamientos estadísticos más completos, para un conjunto de datos, estos tratamientos tratarán de describirse lo más detallado posible con el objetivo de que sirvan de ayuda para los estudiantes que apenas incursionan en la materia.

### 1.- ANÁLISIS DE VARIANZA (ANOVA)

*Libros 5,6,9,10,12,16,27, ...*

*Se utiliza por lo general cuando se involucran 3 o más grupos tiene su máxima potencia cuando se estudia al mismo tiempo 2 o más variables independientes, para contrastar las diferencias entre las medias de las poblaciones*

Como el nombre lo indica, el procedimiento de análisis de varianza trata de analizar la variación de una respuesta y de asignar porciones de esta variación a cada variable de un conjunto de variables independientes. El razonamiento es que las variables de respuesta varían sólo debido a la variación de un conjunto de variables independientes desconocidas; puesto que el experimentador difícilmente podrá incluir todas las variables que afectan la respuesta en su experimento, la respuesta presentará una variación aleatoria, aun si todas las variables aleatorias independientes consideradas se mantienen constantes.

El análisis de varianza se basa en una descomposición de la “suma de cuadrados” (suma de los cuadrados de las desviaciones de los valores de la muestra y la media de la misma), en un arreglo experimental sencillo o experimento para un factor, por ejemplo clase de alimento, tipo de aceite etc.. y por lo tanto, descomponemos la suma en dos partes, una que corresponde al factor cuyo efecto deseamos investigar y el otro que corresponde a la variación aleatoria por lo que su hipótesis se expresa  $H_0: \mu_1 = \mu_2$  y la  $H_i: \mu_1 \neq \mu_2$ . Si estamos interesados en dos factores, tenemos que diseñar un arreglo experimental doble (en dos direcciones) o experimento para dos factores, en este experimento la suma de cuadrados se descompone en 3 partes, que corresponden a los 2 factores que se investigan y a la varianza aleatoria, por ejemplo en un experimento de alimentación de un cerdo, puede desearse cuándo resulta afectada la ganancia mensual en peso por clase de alimento (factor A) y por la cantidad de alimento (factor B), entonces los niveles de A son las diferentes clases de alimento y los niveles B son las diferentes raciones que se han dado durante el experimento

El objetivo del análisis de varianza es identificar las variables independientes importantes en un estudio, y determinar como interactúan y afectan a la respuesta, también sirve para la comparación de medias de muestras independientes, para realizar el análisis estadístico de un diseño completamente al azar, es decir, para probar la  $H_0: \mu_1 = \mu_2 = \dots = \mu_t$  contra la  $H_i$ :

no todas las medias son iguales. Se tendrá que realizar un Análisis de Varianza ó también conocido como ANOVA, con el fin de descomponer la varianza total en una parte, debido a los distintos tratamientos y en otra parte al error experimental.

### 1.1).- ANÁLISIS DE VARIANZA PARA UN SOLO FACTOR

En nuestro modelo de un factor la hipótesis nula dice que las medias de las diferentes poblaciones son iguales, para efectuar esta dócima tomaremos una muestra de cada una de las poblaciones. En el método del análisis de la varianza las diferencias entre las medias de las poblaciones se estudian examinando la cantidad de variación dentro de cada una de estas muestras en relación con la cantidad de variación entre las muestras.

A continuación se presenta la forma de obtener dicho análisis.

Tabla ANOVA para el Diseño completamente al Azar  
(Caso balanceado:  $r_1 = r_2 = \dots r_t$ )

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Fc
Tratamiento	$t - 1$	$SCTR = \sum_{i=1}^t X_i^2 / r - C$	$CMTR = SCTR / (t - 1)$ $CMER = SCER / (t (r - 1))$	$CMTR / CMER$
Error	$t (r - 1)$	$SCER = SCTL - SCTR$		
Total	$r = t - 1$	$SCTL = \sum_{i=1}^t \sum_{j=1}^r x_{ij}^2$		

Donde:  $t$  = número de tratamientos

$r$  = número de repeticiones

$$G = X.. \text{ gran total } G = \sum_{i=1}^t \sum_{j=1}^r x_{ij}$$

$$C = \text{factor de corrección } C = G^2 / (r \cdot t)$$

$X_i$  = total del  $i$ -ésimo tratamiento

Fc = el cociente F consiste en la división de la estimación de la varianza entre grupos entre la estimación de la varianza dentro de grupos, por lo que podemos determinar cuándo o cuándo no la  $H_0$  de la población puede ser razonablemente aceptada. En el caso de tener un coeficiente significativo, puede concluirse que los grupos no fueron extraídos a partir de poblaciones con las mismas medias

Los valores de SCT y SCE miden dos clases de error, la variación entre las medias de las muestras y la variación de las observaciones dentro de las muestras.

\* Cuando la hipótesis nula es cierta ( $\mu_1 = \mu_2$  es decir que las muestras fueron extraídas a partir de la misma población) la CMER y CMTR estiman la misma cantidad y deben ser más o menos de la misma magnitud.

Regla de decisión: Si Fc es mayor que Ft rechazar la  $H_0$

\* Cuando la hipótesis nula es falsa ( $\mu_1 \neq \mu_2$ ) el CMTR probablemente será mayor que el CMER.

El diseño completamente al azar es aquél en el cual los tratamientos son asignados en forma aleatoria a las unidades experimentales.

También se llama análisis de clasificación en un solo sentido porque solo se investiga un factor, además se requiere que el experimento se realice en orden aleatoria, de manera que el medio ambiente en el que se usan los tratamientos (llamados a menudo unidades experimentales) sea lo más uniforme posible

Tabla ANOVA para el Diseño completamente al Azar  
(caso desbalanceado, el número de repeticiones no es el mismo para cada tratamiento)

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medias	Fc
Tratamientos	t - 1	$SCTR = \sum_{i=1}^t X_i^2 / r_i - C$	CMTR = SCTR / (t-1)	CMTR / CMER
Error	N - t	SCER = SCTL - SCTR	CMER = SCER / (N-t)	
Total	N - 1	$SCTL = \sum_{i=1}^t \sum_{j=1}^{r_i} X_{ij}^2 - C$		

Donde:  $r_i$  = número de repeticiones en el i-ésimo tratamiento

$$N = \text{tamaño total de todas las muestras } N = \sum_{i=1}^t r_i$$

$$C = \text{factor de corrección } C = X^2 \dots / N = G^2 / N$$

Las ventajas que este tiene son las siguientes.

- Permite gran flexibilidad es decir, puede usarse cualquier número de tratamientos y repeticiones, se puede variar el número de repeticiones de un tratamiento a otro.
- El análisis estadístico es sencillo, aun si el número de repeticiones no es el mismo para cada tratamiento

### 1.1.1).- ETAPAS PARA EL ANÁLISIS DE VARIANZA

Es importante mencionar que los pasos que se describen para el análisis de varianza no se encuentran en ningún libro, éstos fueron obtenidos de la recopilación bibliográfica. A continuación se presentará el ANOVA de una manera muy sencilla.

Cabe mencionar que las ecuaciones que se muestran nos darán los mismos resultados que cualquier ecuación que se encuentra en los libros.

Para el análisis de varianza es necesario tener los siguientes puntos:

- Hipótesis nula:  $\mu_1 = \mu_2 = \mu_3$ , Hipótesis alterna: las  $\mu$  no son iguales
- Nivel de significancia:  $\alpha \approx 0.05$  la cual esta en función de lo propuesto

- Criterio
- Cálculos; el ANOVA se desarrolla como sigue:

1) Calcular el factor de corrección C elevando al cuadrado la suma de los totales de los tratamientos G y dividiendo por el número total de observación

$$C = G^2 / (r \cdot t)$$

2) Sumar los cuadrados de cada una de las observaciones restándole el factor de corrección C. A esto se le denomina Suma de Cuadrados Total (SCTL)

$$SCTL = \sum_{i=1}^I \sum_{j=1}^r X_{ij}^2 - C$$

3) Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_i$ , dividase por el número de repeticiones r y réstese el factor de corrección. A esto se le llama Suma de Cuadrados de Tratamientos (SCTR)

$$SCTR = \sum_{i=1}^I X_i^2 / r - C$$

4) Restar la suma de cuadrados de los tratamientos (SCTR) de la suma de cuadrados totales (SCTL) y resultará la Suma de Cuadrados debido al Error (SCER), que en este caso es una medida de la variación.

$$SCER = SCTL - SCTR$$

5) Determinar los Grados de Libertad (g.l), para cada SC se determina como sigue:

Para los tratamientos:  $(t - 1)$

Para el error:  $t (r - 1)$

Para el total:  $r \cdot t - 1$

6) Calcular los Cuadrados Medios de los Tratamientos y del Error (cmtr y CMER) dividiendo sus respectivas sumas de cuadrados entre sus respectivos grados de libertad

$$CMTR = SCTR / g.l$$

$$CMER = SCER / g.l$$

7) Calcular el valor de  $F_c$  dividiendo CMTR entré CMER, a esto se le llama F Calculada

$$F_c = CMTR / CMER$$

8) Se formula la tabla ANOVA

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medias	$F_c$
Tratamientos				
Error				
Total				

9) Con el Nivel de Significancia ( $\alpha$ ) y los g.l se encuentra el valor de Ft en tablas de la Distribución F (ver APENDICE), basándose con la regla de decisión

Si  $F_c > F_t$  se rechaza la  $H_0$

10) Para resumir la información obtenida sobre los tratamientos es útil formar la siguiente tabla

Tabla de Medias y Errores Estándar

Tratamientos	A	B	C
Media $\bar{x}_i$			
Error estándar $S_{\bar{x}_i} = \sqrt{CMER / r_i}$			

### 1.2).- ANÁLISIS DE VARIANZA PARA 2 FACTORES (arreglo experimental doble)

En un arreglo experimental doble tenemos dos factores A y B, de tal manera que A tiene  $r$  niveles  $A_1, A_2, \dots, A_r$  y B tiene  $p$  niveles  $B_1, B_2, \dots, B_p$ , para estudiar el efecto de A y B sobre una variable aleatoria  $x$  tomamos una muestra de  $n = rp$  valores. Podemos arreglar esta muestra en la forma de un arreglo rectangular que conste de  $r$  renglones y  $p$  columnas verticales

	$r$ columnas $x_1, x_2, x_3, \dots, x_p$ $\vdots$ $x_{11}, x_{12}, x_{13}, \dots, x_p$
$p$ renglones	

A continuación se mencionan los pasos de la prueba en el caso de un arreglo experimental doble con una observación por celda, donde la hipótesis es que las medias son iguales

1° Se calcula el valor medio del  $i$ -ésimo renglón es comúnmente representado por  $\bar{x}_i$

$$\bar{x}_i = \frac{1}{p} \sum_{k=1}^p x_{ik} = \text{Suma de los valores en el } i\text{-ésimo renglón} / \text{Número de valores por renglón}$$

2° Se calcula el valor medio de la  $k$ -ésima columna y se representa por  $\bar{x}_{.k}$

$$\bar{x}_{.k} = \frac{1}{r} \sum_{i=1}^r x_{ik} = \text{Suma de los valores en la } k\text{-ésima columna} / \text{Número de valores por columna}$$

3° Se calcula el valor medio de toda la muestra

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{k=1}^p x_{ik}$$

4° Se calcula la suma de cuadrados, recordamos que en el caso de un arreglo experimental sencillo descomponemos la suma de cuadrados en dos partes. En el caso que ahora tratamos de un arreglo experimental doble descomponemos la suma de cuadrados (representada por la letra  $q$ ) en tres partes  $q_1$ ,  $q_2$ ,  $q_3$

$$* q = q_1 + q_2 + q_3$$

\*  $q_1$  = suma de cuadrados del factor A, esta representa la variación en los  $x_{ik}$  debido a efectos del factor A (hay que recordar que los  $r$  renglones corresponden a los  $r$  niveles del factor A)

$$q_1 = p \sum_{i=1}^r (\bar{x}_i - \bar{x})^2$$

\*  $q_2$  = suma de cuadrados del factor B, representa la variación en los  $x_{ik}$  debido a los efectos del factor B.

$$q_2 = r \sum_{k=1}^p (\bar{x}_{.k} - \bar{x})^2$$

\*  $q_3$  = suma de cuadrado del error, esta suma representa la variación en los  $x_{ik}$  que queda después de quitar los efectos de los factores A y B

$$q_3 = \sum_{i=1}^r \sum_{k=1}^p (x_{ik} - \bar{x}_i - \bar{x}_{.k} + \bar{x})^2$$

5° Se calcula los cuadrados medios de cada suma de cuadrados

$$S_1^2 = q_1 / (r - 1)$$

$$S_2^2 = q_2 / (p - 1)$$

$$S_3^2 = q_3 / (r - 1)(p - 1)$$

Se arreglan los resultados como se muestra en la siguiente tabla

Fuente de Variación	Grados de libertad	Suma de cuadrados	Cuadrado medio
Entre renglones (efecto A)	$r - 1$	$q_1$	$S_1^2 = q_1 / (r - 1)$
Entre columnas (efecto B)	$p - 1$	$q_2$	$S_2^2 = q_2 / (p - 1)$
Desviaciones	$(r - 1)(p - 1)$	$q_3$	$S_3^2 = q_3 / (r - 1)(p - 1)$
Total	$n - 1$	$q$	

6° Calcular los cocientes

$$v_1 = \frac{S_1^2}{S_3^2} \quad \text{y} \quad v_2 = \frac{S_2^2}{S_3^2}$$

7° Se elige un nivel de significancia  $\alpha$  (5% o 1%)

8° Se determina la solución  $c_1$  de la ecuación

$$P(v \leq c_1) = 1 - \alpha$$

Con ayuda de la tabla de la distribución F con  $[r - 1, (r - 1)(r - p)]$  grados de libertad.

Sí  $(v_1 \leq c_1)$  se asegura que no hay diferencia significativa entre los renglones

Sí  $(v_1 > c_1)$  se asegura que si existe esa diferencia y rechazamos la hipótesis

9° Se determina la solución  $c_2$  de la ecuación

$$P(v \leq c_2) = 1 - \alpha$$

De la tabla de la distribución F con  $[p - 1, (r - 1)(p - 1)]$  grados de libertad

Sí  $(v_2 \leq c_2)$  se asegura que no existe diferencia significativa entre las columnas

Sí  $(v_2 > c_2)$  se asegura que existe dicha diferencia

10° Sí  $v_1 \leq c_1$  y  $v_2 \leq c_2$ , no se rechaza la hipótesis de que todo los  $n$  valores medios  $\mu_{ik}$  son iguales. En cualquier otro caso se rechaza la hipótesis

### 1.3).- INTERPRETACIÓN DE F

El hallazgo de un cociente F es significativo para el experimento en su totalidad, nos permite la investigación de hipótesis específicas, si no hay significación en el cociente F cualquier diferencia significativa entre ciertas comparaciones específicas deberá considerarse como sospechosa, representará probablemente una diferencia casual.

En los últimos años, se ha desarrollado un gran número de pruebas que permiten al investigador el estudio específico de hipótesis relacionados con los parámetros de la población. En líneas generales hay dos tipos de estas pruebas:

1. *Comparaciones planeadas o a priori*: cuando se planean las comparaciones anticipadamente, el empleo de pruebas a priori es apropiado para pruebas a priori que no es necesario que el cociente F sea significativo
2. *Comparaciones a posteriori*: cuando las comparaciones no se planean anticipadamente, es apropiado el empleo de una prueba a posteriori

#### 2.) Comparación a posteriori

Cuándo en el Análisis de Varianza se encuentra que un valor de F es significativo, el interrogante es ¿ Cual de las medias de tratamientos son significativamente diferentes? A continuación se presentan dos de las pruebas más comúnmente usadas para responder esta cuestión.



### A) DIFERENCIA MÍNIMA SIGNIFICATIVA

(de Fisher)

Esta es una prueba múltiple utilizando t-Student y deberá ser usada de la manera acostumbrada únicamente cuando se requieran probar algunas comparaciones de interés, cuando el análisis de varianza indica la existencia de una diferencia significativa, deseamos conocer cual de los pares de medias causa la diferencia. Cuando las muestras son de igual tamaño el DMS nos ayuda a localizar esta fuente

La Hipótesis a probar será:

$$H_0 = \mu_i - \mu_j = 0 \quad \text{vs} \quad H_1 = \mu_i - \mu_j \neq 0$$

La fórmula de esta prueba es:

$$DMS = t_t \cdot S_{\bar{x}_i - \bar{x}_j}$$

Donde: DMS = diferencia mínima significativa

$t_t$  = valor obtenido de tablas, con los grados de libertad del error y el nivel de significancia deseado

$S_{\bar{x}_i - \bar{x}_j}$  = desviación estándar de la diferencia entre dos medias, por lo tanto

tenemos:

$$S_{\bar{x}_i - \bar{x}_j} = \sqrt{2s^2/r} = \sqrt{2CMER/r}$$

Entonces el criterio de la prueba es que si la diferencia entre  $\bar{x}_i - \bar{x}_j$  excede el valor del DMS se pueden considerar estadísticamente diferentes

### B) DIFERENCIA MÍNIMA SIGNIFICATIVA HONESTA

(Prueba de Tukey)

Es una prueba de comparación de medias bastante estricta y mantiene la probabilidad de que cualquier diferencia de medias haya sido declarada falsamente significativa, en el nivel  $\alpha$  fijado. La prueba DMSH al igual que el DMS solo se puede usar después de que se ha rechazado la  $H_0$  en el análisis de varianza y cuando los tamaños de la muestra son todos iguales

Su procedimiento es el siguiente:

1°.- Calcular el valor de la diferencia mínima significativa honesta ( DMSH ) a partir de la siguiente fórmula

$$DMSH = q_{\alpha, t, g, t} \cdot S_{\bar{x}}$$

Donde:  $q_{\alpha, t, g, t} = q$ , factor obtenido de las tablas de Rango Estudentizado (ver APENDICE) con un nivel de significancia dado; t tratamientos y g.l los grados de libertad del error

$$S_{\bar{x}} = \sqrt{s^2/r} = \sqrt{CMER/r} = \text{desviación estándar de la media}$$

2°.- Cualquier diferencia entre dos medias se declara estadísticamente significativa si excede el valor de la DMSH

## RESUMEN

El ANOVA es aplicable a situaciones en las cuales deseamos comprobar el efecto de más de un factor; de hecho es en estos casos cuando el ANOVA resulta más útil, para el caso de factor único:

- El ANOVA es un método de repartir la variabilidad total de una serie de observaciones entre las posibles fuentes de variabilidad
- El ANOVA incorpora pruebas F para comparar las medias de los cuadrados
- El ANOVA es válido únicamente si los datos han sido adecuadamente obtenidos, la hipótesis a comprobar y el nivel de significación deberían establecerse anticipadamente a la recogida de los datos. Puesto que el ANOVA supone estas fuentes de variación, distintas de los factores ensayados, son imputables al error, es necesario que las observaciones se hagan de forma que las fuentes de variación no ensayadas presenten sus efectos aleatoriamente

*Como el Análisis de Varianza es un tema muy amplio que incluye métodos y modelos extensos desarrollados durante las décadas pasadas, solamente introducimos ideas básicas considerando alguno de los diseños experimentales más simples y su análisis estadístico, los experimentos se pueden diseñar para investigar el efecto de muchas variables cuantitativas y cualitativas sobre una respuesta; éstas pueden ser tanto las variables de principal interés para el experimentador como otras variables sin importancia, por lo tanto a los estudiantes que se interesen por más métodos se les remite a Anderson y Bancroft, Cochran y Scheffe. (28)*

1.4).- PROBLEMAS RESUELTOS:

(Problema desbalanceado)

1) En un proceso de fabricación de pan Danés, se cuenta con tres tipos diferentes de máquina, las cuales son las siguientes: A = manual, B = semimecánica y C = automática. Se mide la pérdida de humedad (%) de los productos en una semana y se obtienen los siguientes datos

A	B	C
3.09	2.19	3.16
3.18	1.92	4.92
0.92	4.65	3.01
2.10	2.16	4.38
1.95	3.15	5.04
3.27	2.88	3.59
$\Sigma = 14.51$	4.23	4.62
$n = 6$	4.35	2.21
	3.90	3.69
	3.13	2.83
	4.02	4.17
	$\Sigma = 37.08$	3.62
	$n = 11$	3.81
		4.65
		3.97
		3.61
		$\Sigma = 61.28 = X_c$
		$n = 16$

Pruebe con un nivel de significancia del 5%, si existe diferencia en la pérdida de humedad del pan Danés fabricado en cada máquina.

Solución:

Tabla 1:

Máquina	A	B	C
$r_i$	6	11	16
$X_i$	14.51	37.08	61.28
$\bar{x}_i$	2.41	3.37	3.83

$\Sigma N = 33$  número total de observaciones  
 $\Sigma X_{..} = 112.87 = G$  el gran total

Donde: n es la cantidad de datos presentes en cada columna

$X_i$  es la suma de los datos presentes en cada columna

$\bar{x}_i$  es la división de  $X_i/n$  de cada columna

→ indica el sentido, para la obtención de la sumatoria

1° Calcular el factor de corrección

$$C = G^2 / N = (112.87)^2 / 33 = 386.050$$

2° Sumar los cuadrados de cada observación y réstese el factor de corrección para obtener la suma de los cuadrados totales

$SCTL = \sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 - C$  = esta fórmula indica que es necesario elevar al cuadrado cada uno de los datos que se encuentran en las columnas A, B y C desde 3.09 hasta 3.61

$$\begin{aligned} SCTL = & (3.09)^2 + (3.18)^2 + (0.92)^2 + (2.10)^2 + (1.95)^2 + (3.27)^2 + (2.19)^2 + (1.92)^2 + (4.65)^2 \\ & + (2.16)^2 + (3.15)^2 + (2.88)^2 + (4.23)^2 + (4.35)^2 + (3.90)^2 + (3.13)^2 + (4.02)^2 + (3.16)^2 + (4.92)^2 \\ & + (3.01)^2 + (4.38)^2 + (5.04)^2 + (3.59)^2 + (4.62)^2 + (2.21)^2 + (3.69)^2 + (2.83)^2 + (4.17)^2 + (3.62)^2 \\ & + (3.81)^2 + (4.65)^2 + (3.97)^2 + (3.61)^2 - 386.050 \\ SCTL = & 8.736 \end{aligned}$$

3° Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_i$ , dividir por el número de repeticiones  $r_i$  y réstese el factor de corrección para obtener la suma de cuadrados de tratamientos

$SCTR = \sum_{i=1}^I X_i^2 / r_i - C$  = la fórmula indica que es necesario elevar al cuadrado las sumatorias

de cada columna A, B y C, y cada dato se divide entre el número de repetición que le corresponde, y a ese resultado se le resta el factor de corrección que se obtuvo en el punto 1

$$\begin{aligned} SCTR = & [(14.51)^2 / 6 + (37.08)^2 / 11 + (61.28)^2 / 16] - 386.050 \\ SCTR = & 8.736 \end{aligned}$$

4° Por diferencia se calcula la suma de cuadrados del error

$SCER = SCTL - SCTR$  = estos son datos que ya se obtuvieron en los pasos 2 y 3

$$SCER = 32.93 - 8.736$$

$$SCER = 24.201$$

5° Calcular los grados de libertad

\*Tratamientos:  $(t - 1)$  = los tratamientos son el número de columnas que contienen todos los datos como con las columnas A, B y C

\*Error:  $N - t = N$  es el número total de las observaciones, que se obtuvo con la sumatoria de las  $r_i$ , este dato se encuentra en la Tabla 1

\* Total:  $N - 1$ , y a estos datos obtenidos se les llama grados de libertad

$$\text{Tratamientos: } 3 - 1 = 2$$

$$\text{Error: } 33 - 3 = 30$$

$$\text{Total: } 33 - 1 = 32$$

6° Calcular los cuadrados medios para los tratamientos y para el error

$CMTR = SCTR / \text{tratamientos}$  = es decir el dato obtenido en el paso 3 entre el dato obtenido en el paso 5

$CMER = SCER / \text{error}$  = el dato obtenido en el paso 4 entre el paso 5

$$\begin{aligned} \text{CMTR} &= 8.736 / 2 = 4.368 \\ \text{CMER} &= 24.201 / 30 = 0.8067 \end{aligned}$$

7° Calcular el valor de  $F_c$

$F_c = \text{CMTR} / \text{CMER}$  que son los datos obtenidos anteriormente

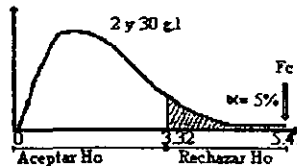
$$F_c = 4.368 / 0.8067 = 5.4147$$

8° Se resumen los cálculos obtenidos en la tabla de ANOVA

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados media	$F_c$
Tratamientos	2	SCTR = 8.736	CMTR = 4.368	5.4147
Error	30	SCER = 24.201	CMER = 0.8067	
Total	32	SCTL = 32.937		

9° Comparar la  $F_c = 5.4147$  con la  $F_t$  para  $\alpha = 5\%$ ,  $2\%$  y  $30\%$  g.l;  $F_t = 3.32$

Como la  $F_c = 5.4147 > F_t = 3.32$  rechazamos la hipótesis nula, es decir, el contenido de humedad en el pan Danés no es la misma en todas las máquinas



10° Para resumir la información obtenida sobre los tratamientos es útil formar la siguiente tabla

Tabla de Medias y Errores estándar de % de humedad

Tipo de Máquina	Manual	Semiautomática	Automática
Media $\bar{x}_i$	2.42%	3.37%	3.83%
Error estándar $S_{\bar{x}_i} = \sqrt{\text{CMER}/r_i}$	$\pm 0.37\%$	$\pm 0.27\%$	$\pm 0.22\%$

Como se puede observar en esta tabla el contenido de humedad en el pan Danés no es el mismo en los tres tipos de máquinas, por lo que podemos decir que se elegirá la máquina dependiendo de nuestras necesidades, es decir si queremos un pan con un porcentaje de humedad mayor, podemos utilizar una máquina automática y por el contrario si deseamos un pan Danés con un 2.42% de humedad es necesario trabajar con una máquina manual.

2) Un experto en agricultura interesado en la producción de frutas en ciertos huertos colindantes que son análogos en todo, excepto en los fungicidas y pesticidas que se les aplica. Toma una muestra aleatoria de árboles frutales y determina su producción, obteniéndose los siguientes resultado.

Tratamiento 1	Tratamiento 2	Tratamiento 3
130	126	131
126	127	128
128	129	133
132	124	130
126	126	130
	130	132
	125	
	126	
	128	
	126	

Una manera de hacer que el trabajo sea menos laborioso en ANOVA consiste en restar un valor arbitrario a todas las observaciones para hacer que los números sean más fáciles de manejar. Este procedimiento no modifica la variación entre o dentro de las columnas puesto que se resta el mismo número a todos los datos obtenidos.

En la tabla 2 aparecen todos los datos de la tabla anterior después de restarle el número 120 a cada uno de los números también aparecen las sumas muestrales, las medias y los valores estimados (que suelen llamarse *datos codificados*).

Tabla 2. Datos codificados

Tratamiento 1	Tratamiento 2	Tratamiento 3
10	6	11
6	7	8
8	9	13
12	4	10
6	6	10
	10	12
	5	
	6	
	8	
	6	

Solución:

Tabla 1:

Producción	Tratamiento 1	Tratamiento 2	Tratamiento 3
$r_i$	5	10	6 →
$X_i$	42	67	64 →
$\bar{X}_i$	8.4	6.7	10.7

$\Sigma N = 21$  número total de observaciones

$\Sigma X_i = 173 = G$  el gran total

1° Calcular el factor de corrección

$$C = G^2 / N = (173)^2 / 21 = 1425.19$$

$$2^\circ \text{ SCTL} = \sum_{i=1}^3 \sum_{j=1}^7 X_{ij}^2 - C = (10^2 + 6^2 + 8^2 + \dots + 10^2 + 10^2 + 12^2) - 1425.19$$

$$= 1557 - 1425.19 = 131.81$$

$$3^\circ \text{ SCTR} = \sum_{i=1}^3 X_i^2 / r_i - C = [(42)^2 / 5 + (67)^2 / 10 + (64)^2 / 6] - 1425.19$$

$$= [1764 / 5 + 4489 / 10 + 4096 / 6] - 1425.19 = 59.18$$

$$4^\circ \text{ SCER} = \text{SCTL} - \text{SCTR} = 131.81 - 59.18 = 72.63$$

5° Grados de libertad: Tratamientos = (3 - 1) = 2

$$\text{Error} = 21 - 3 = 18$$

$$\text{Total} = 21 - 1 = 20$$

$$6^\circ \text{ CMTR} = \text{SCTR} / \text{tratamientos} = 59.18 / 2 = 29.59$$

$$\text{CMER} = \text{SCER} / \text{error} = 72.63 / 18 = 4.035 \approx 4.04$$

$$7^\circ \text{ Fc} = \text{CMTR} / \text{CMER} = 29.59 / 4.04 = 7.32$$

8° Se formula la tabla de ANOVA

F. V	g.l	SC	CM	Fc
Tratamientos	2	59.18	29.59	7.32
Error	18	72.63	4.04	
Total	20	131.81		

9° Puesto que el valor crítico correspondiente a  $\alpha = 0.01$  con ayuda de las tablas  $F_{(2,18)} = 6.01$  como  $F_c > F_t$  es necesario rechazar la hipótesis nula de que los tratamientos no determinan una diferencia en la producción.

Hay una diferencia significativa entre las medias de las poblaciones que fueron sometidas a los diferentes tratamientos, con referencia a las medias muestrales que aparecen en la tabla 2, encontramos que el tratamiento 3 es el más efectivo para aumentar la producción

(Problema balanceado)

3) Para investigar la acción o el efecto que tiene un lubricante sobre determinada parte metálica de una máquina envasadora, se experimento con 3 diferentes tipos de lubricantes bajo las mismas condiciones controladas, midiendo el peso perdido debido a la fricción (en mg) de 8 partes metálicas para cada tipo de lubricante que se anotan enseguida:

Lubricante	x = peso perdido							
A	11.2	11.8	13.1	11.0	3.9	4.1	10.3	8.4
B	10.9	5.7	13.5	9.4	11.4	15.7	10.8	14.2
C	12.7	19.9	13.9	11.7	18.3	14.3	22.8	20.4

con 5% de significación, probar si el tipo de lubricante tiene efecto sobre el desgaste de la parte metálica de la máquina envasadora.

**Solución:**

Planteamos las siguientes hipótesis:  $H_0: \mu_A = \mu_B = \mu_C$

$H_1$ : no todas las medias son iguales

Con la regla de decisión: Si  $F_c$  es menor que  $F_t$ , aceptar la  $H_0$  y efectuamos los siguientes cálculos, con  $t = 3$  tipos de lubricante (tratamientos)

$r = 8$  repeticiones

	A	B	C
	11.2	10.9	12.7
	11.8	5.7	19.9
	13.1	13.5	13.9
	11.0	9.4	11.7
	3.9	11.4	18.3
	4.1	15.7	14.3
	10.3	10.8	22.8
	8.4	14.2	20.4

Lubricantes	A	B	C
$r_i$	8	8	8
$X_i$	73.8	91.6	134
$X_i$	9.225	11.45	16.75

$\Sigma N = 24$  número total de observaciones

$\Sigma X_{..} = 299.4$  el gran total

$$1^\circ C = G^2 / N = (299.4)^2 / (24) = 3735.015$$

$$\text{ó bien } C = X^2_{..} / (r * t)$$

$$2^\circ \text{ SCTL} = \sum_{i=1}^t \sum_{j=1}^r X_{ij}^2 - C = 11.2^2 + 11.8^2 + \dots + 22.8^2 + 20.4^2 - 3735.015 = 510.365$$

$$3^\circ \text{ SCTR} = \sum_{i=1}^t X_i^2 / r_i - C = [(73.8^2 + 91.6^2 + 134^2) / 8] - 3735.015 = 239.11$$

$$4^\circ \text{ SCER} = \text{SCTL} - \text{SCTR} = 510.365 - 239.11 = 271.255$$

$$5^\circ \text{ Grados de libertad: Tratamientos} = 3 - 1 = 2$$

$$\text{Error} = 3(8 - 1) = 21$$

$$\text{Total} = (3)(8) - 1 = 23$$

$$6^\circ \text{ CMTR} = \text{SCTR} / \text{tratamientos} = 239.11 / 2 = 119.555$$

$$\text{CMER} = \text{SCER} / \text{error} = 271.255 / 21 = 12.9169$$

$$7^\circ F_c = \text{CMTR} / \text{CMER} = 119.555 / 12.9169 = 9.256$$



8° Se formula la tabla de ANOVA

F. V	g.l	SC	CM	Fc
Tipo de lubricante	2	239.11	119.555	9.256
Error	21	271.255	12.9169	
Total	23	510.365		

9°  $\alpha = 5\%$     2 y 21 g.l     $F_t = 3.47$

Como  $F_c = 9.256$  es mayor que  $F_t = 3.47$ , rechazamos la hipótesis nula y podemos afirmar que el tipo de lubricante si tiene efecto sobre el desgaste de la pieza metálica, es decir, su variación no se debe al azar.

Para ver exactamente cuales son las diferencias entre los lubricantes es necesario continuar con la prueba a posteriori utilizando las técnicas de diferencia mínima significativa y diferencia mínima significativa honesta.

Basándose en los datos y resultados obtenidos de los lubricantes se pregunta, cuáles son las medias diferentes estadísticamente con las pruebas de:

a) DMS

b) DMSH

con un nivel de significancia del 1%

Solución:

Tratamiento	A	B	C
$\bar{x}_i$	9.225	11.45	16.75

a) DMS

Datos:  $\alpha = 1\%$  ; 21 g.l ;  $t_1 = 2.831$  este valor se obtiene de tablas con los grados de libertad del error y el nivel de significancia.

$$DMS = t_1 * S_{\bar{x}_i - \bar{x}_j}$$

$$S_{\bar{x}_i - \bar{x}_j} = \sqrt{2s^2/r} = \sqrt{2(CMER)/r}$$

$$S_{\bar{x}_i - \bar{x}_j} = \sqrt{2(12.917)/8} = 1.797$$

$$DMS = (2.831)(1.797) = 5.087$$

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
C - B	5.3	C mayor que B
C - A	7.525	C mayor que A
B - A	2.225	ns

Anteponiendo que todo el valor que exceda el DMS (diferencia mínima significativa) presenta una diferencia significativa, por lo que la comparación entre B y A no tiene diferencia significativa, entre C y B ó C y A existe una diferencia significativa. Por lo tanto se elegirá el lubricante dependiendo de las necesidades, es decir, si se quiere un lubricante que disminuya el peso de las partes metálicas se escogerá el lubricante C y por el contrario si se desea un lubricante que no disminuya el peso se elegirá la A.

#### b) DMSH

Datos:  $\alpha = 1\%$  ;  $t = 3$  ;  $g.l = 21$  ;  $q_t = 4.6075$  (interpolando entre 4,68 y 4.54) de las tablas de rango estudentizado utilizando el nivel de significancia, el número de tratamientos y los grados de libertad.

$$DMSH = q_{t,\alpha,g.l} * S_{\bar{x}}$$

$$S_{\bar{x}} = \sqrt{s^2/r} = \sqrt{CMER/r}$$

$$S_{\bar{x}} = \sqrt{12.917/8} = 1.270$$

$$DMSH = (4.6075)(1.270) = 5.855$$

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
C - B	5.3	ns
C - A	7.525	C mayor que A
B - A	2.225	ns

Con la prueba DMSH, la comparación C - B resulta no significativa, pero era significativa con la prueba DMS.

Para estar más seguros se considera solamente la comparación C - A como significativa, es decir que el lubricante A es mejor que el lubricante C, puesto que propicia menor desgaste, aunque se encuentra cierta evidencia de que B supera a C.

También se acostumbra a representar estas comparaciones anotando los tratamientos y uniendo con una línea recta los tratamientos en los que no se encontró diferencia significativa, por ejemplo:

Tratamientos	$\bar{x}_i$
C	16.75
B	11.45
A	9.225

4) Suponga que se ha diseñado un experimento para determinar la eficiencia de cuatro métodos distintos de secado de carne, cada método se repitió cinco veces en un diseño completamente al azar, midiéndose el porcentaje de agua residual después de aplicado el método.

Los datos son los que se presentan a continuación:

% de Agua residual			
A	B	C	D
1.9	1.7	1.6	1.1
1.7	1.6	1.7	1.3
2.0	1.5	1.9	1.6
1.6	2.0	1.6	1.1
1.9	1.6	1.4	1.2

Pruebe con un nivel de significancia del 1% si hay diferencia en el porcentaje de agua residual encontrado con los distintos métodos de secado.

**Solución:**

Para probar la  $H_0: \mu_A = \mu_B = \mu_C = \mu_D$ , se desarrolla la tabla de ANOVA

	A	B	C	D	
$r_i$	5	5	5	5	→ $\sum N = 20$ número total de observaciones
$\bar{x}_i$	9.1	8.4	8.2	6.3	
$\bar{\bar{x}}$	1.82	1.68	1.64	1.26	→ $\sum X_{..} = 32 = G$ el gran total

El ANOVA se desarrolla como sigue:

$$1^\circ C = G^2 / N = (32)^2 / (20) = 51.2$$

$$\text{ó bien } C = X^2_{..} / (r \cdot t)$$

$$2^\circ \text{SCTL} = \sum_{i=1}^t \sum_{j=1}^r x_{ij}^2 - C = (1.9^2 + 1.7^2 + 2.0^2 \dots + 1.1^2 + 1.2^2) - 51.2 = 1.42$$

$$3^\circ \text{SCTR} = \sum_{i=1}^t x_i^2 / r_i - C = [(9.1^2 + 8.4^2 + 8.2^2 + 6.3^2) / 5] - 51.2 = 0.86$$

$$4^\circ \text{SCER} = \text{SCTL} - \text{SCTR} = 1.42 - 0.86 = 0.56$$

$$5^\circ \text{Grados de libertad: Tratamientos} = 4 - 1 = 3$$

$$\text{Error} = 4(5 - 1) = 16$$

$$\text{Total} = (4)(5) - 1 = 19$$

$$6^\circ \text{CMTR} = \text{SCTR} / \text{tratamientos} = 0.86 / 3 = 0.287$$

$$\text{CMER} = \text{SCER} / \text{error} = 0.56 / 16 = 0.035$$

$$7^\circ F_c = \text{CMTR} / \text{CMER} = 0.287 / 0.035 = 8.19$$

8° Se formula la tabla de ANOVA

F. V	g.l	SC	CM	Fc
Tratamientos	3	0.86	0.287	8.19
Error	16	0.56	0.035	
Total	19	1.42		

9° Con un nivel de significancia  $\alpha = 1\%$ ; 3 g.l de tratamientos y 16 g.l de error se encuentra el valor de Ft en tablas, por lo tanto tenemos que  $F_t = 5.29$ .

Como  $F_c = 8.19 > F_t = 5.29$  se rechaza la  $H_0$ , por lo tanto hay diferencia en los métodos de secado, por lo que es necesario saber cuales medidas son estadísticamente iguales y cuales no, utilizando las pruebas DMS y DMSH con un nivel de significancia del 5%

Solución:

Tratamiento	A	B	C	D
$\bar{x}_i$	1.82	1.68	1.64	1.26

a) utilizando la prueba de la diferencia mínima significativa, se tendrá que calcular el valor del DMS

$$DMS = t_t * S_{\bar{x}_i - \bar{x}_j}$$

Localice el valor de  $t_t$  con 16 g.l de error y  $\alpha = 5\%$  se tiene:

$$t_t = 2.12$$

Calcule ahora el valor de  $S_{\bar{x}_i - \bar{x}_j} = \sqrt{2s^2/r} = \sqrt{2(CMER)/r}$

$$S_{\bar{x}_i - \bar{x}_j} = \sqrt{2(0.035)/5} = 0.118$$

entonces  $DMS = (2.12)(0.116) = 0.25$

La diferencia entre dos medias que exceda el valor de 0.25 se considera estadísticamente diferente

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
A - B	0.14	ns
A - C	0.18	ns
A - D	0.56	A mayor que D
B - C	0.04	ns
B - D	0.42	B mayor que D
C - D	0.38	C mayor que D

Como se puede ver en el cuadro entre las comparaciones de A - B, A - C y B - C no existe una diferencia estadística, mientras que en A - D, B - D y C - D se consideran estadísticamente diferentes.

b) Utilizando la prueba de la diferencia media significativa honesta, calculamos el valor de la DMSH

$$DMSH = q_{t,\alpha,g,l} * S_{\bar{x}}$$

Localizando el valor de  $q$  de la tabla de rango estudentizado (ver APÉNDICE) con  $\alpha = 5\%$ ,  $t = 4$  (que es el número de tratamientos) y los grados de libertad del error = 16

$$\text{Calcule ahora el valor de } S_{\bar{x}} = \sqrt{s^2/r} = \sqrt{CMER/r}$$

$$q_{0.05,4,16} = 4.05$$

$$S_{\bar{x}} = \sqrt{0.035/5} = 0.084$$

entonces  $DMSH = (4.05)(0.084) = 0.3402$

La diferencia entre dos medias que excede este valor se considera estadísticamente diferentes

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
A - B	0.14	ns
A - C	0.18	ns
A - D	0.56	A mayor que D
B - C	0.04	ns
B - D	0.42	B mayor que D
C - D	0.38	C mayor que D

Como los resultados obtenidos utilizando la prueba del DMS concuerdan con los resultados obtenidos utilizando la prueba DMSH, se puede estar bastante seguro en estos resultados.

Lo que se aprecia al observar estos resultados es que no se encuentra diferencia significativa entre los métodos A, B, C y se detecta que estos métodos dan un promedio mayor en % de agua residual que con el método D, por lo tanto este es el método de secado más recomendado.

Si se quisiera escoger algún método entre el A, B y C en el cual no se aprecia diferencia significativa, se escogería de acuerdo a otras razones como costo, tiempo de secado etc..

*(Arreglo experimental doble)*

5) (Experimento de alimentación de cerdos). Se agrupó a 12 cerdos en  $r = 4$  (renglones) y grupos de  $p = 3$  (columnas) cerdos cada uno, de acuerdo a su peso en el principio del experimento, de manera que el factor A en el sentido de la teoría, es el peso inicial de un cerdo y tiene 4 niveles, en cada grupo se dio 3 dietas diferentes  $B_1$ ,  $B_2$  y  $B_3$ , a los tres cerdos. Después de cierto intervalo de tiempo, la ganancia en peso fue según la Tabla 1.

Con esta prueba deseamos probar la hipótesis de que ni el peso inicial ni el tipo de dieta afecta a la ganancia.

Tabla 1

	Dieta			X =	$\bar{x} =$
	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>		
Grupo A <sub>1</sub>	7.0	14.0	8.5	29.5	9.8
Grupo A <sub>2</sub>	16.0	15.5	16.5	48.0	16.0
Grupo A <sub>3</sub>	10.5	15.0	9.5	35.0	11.67
Grupo A <sub>4</sub>	13.5	21.0	13.5	48.0	16.0
	40.70	65.5	48.0	160.5	53.47
	11.75	16.38	12.0		40.13

Solución:

1° Calcular la suma de los valores en el i-ésimo renglón entre el número de valores por renglón

$$\bar{x}_i = \frac{1}{p} \sum_{k=1}^p x_{ik} = \frac{160.5}{3} = 53.5$$

2° Calcular el valor medio de la k-ésima columna

$$\bar{x}_k = \frac{1}{r} \sum_{i=1}^r x_{ik} = \frac{160.5}{4} = 40.1$$

3° Calcular el valor medio de toda la muestra

$$\bar{x} = \frac{1}{n} \sum_{k=1}^p \sum_{i=1}^r x_{ik} = \frac{160.5}{12} = 13.37$$

4° Calcular la suma de cuadrados  $q_1, q_2, q_3, q$

$$\begin{aligned} * q_1 &= 3 [(9.8 - 13.37)^2 + (16 - 13.37)^2 + (11.67 - 13.37)^2 + (16 - 13.37)^2] \\ &= 3 (12.74 + 6.91 + 2.89 + 6.91) \\ &= 88.35 \end{aligned}$$

$$\begin{aligned} * q_2 &= 4 [(11.75 - 13.37)^2 + (16.38 - 13.37)^2 + (12 - 13.37)^2] \\ &= 4 (2.62 + 9.06 + 1.88) \\ &= 54.2 \end{aligned}$$

$$* q_3 = 28.21$$

$$* q = 88.35 + 54.2 + 28.21 = 170.76$$

5° Calcular el cuadrado medio de cada suma de cuadrados

$$S_1^2 = q_1 / (r - 1) = 88.35 / (4 - 1) = 29.42$$

$$S_2^2 = q_2 / (p - 1) = 54.2 / (3 - 1) = 27.10$$

$$S_3^2 = q_3 / (r - 1)(p - 1) = 28.21 / 6 = 4.70$$

Agrupar los datos en la siguiente tabla:

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio
Entre renglones	3	$q_1 = 88.35$	29.42
Entre columnas	2	$q_2 = 54.2$	27.10
Desviaciones	$(r - 1)(p - 1) = 6$	$q_3 = 28.21$	4.70
Total	$n - 1 = 11$	$q = 170.76$	

$$6^\circ \quad v_1 = \frac{29.42}{4.70} = 6.25 \quad ; \quad v_2 = \frac{27.10}{4.70} = 5.76$$

7° Elegimos el nivel de significancia  $\alpha = 0.05$

8° Para la distribución F con 3.6 g.l la ecuación

$$P(v \leq c_1) = 1 - \alpha = 0.95 \\ c_1 = 4.76$$

Vemos que  $v_1 = 6.25 > c_1 = 4.76$ . Por lo tanto aseguramos que existe una diferencia significativa entre los grupos, esto es, aseguramos que el peso inicial afecta la ganancia, en consecuencia rechazamos la hipótesis

9° Para la distribución F con 2.6 g.l la ecuación

$$P(v \leq c_2) = 1 - \alpha = 0.95 \\ c_2 = 5.14$$

Tenemos que  $v_2 = 5.76 > c_2 = 5.14$ . Por lo tanto como conclusión tenemos que si existe una diferencia significativa entre las columnas, esto es, aseguramos que el tipo de dieta y el peso inicial de los cerdos afecta la ganancia.

## II.- BLOQUES ALEATORIOS (ANDEVA)

*Libros 2, 5, 9, 10, 13, 15, 17, 18*

*El diseño de Bloques Aleatorios se utiliza para comparar las medias de cualquier número de tratamientos, dentro de bloques relativamente homogéneos de p unidades experimentales. Se puede usar cualquier número de bloques, pero cada tratamiento debe aparecer exactamente una vez en cada bloque, también sirve para reducir el error experimental*

En el diseño de Bloques Aleatorios primero se forman grupos (*bloques*) de unidades experimentales (u.e) similares y después todos los tratamientos son probados dentro de cada bloque con la asignación de los tratamientos a las u.e hechas aleatoriamente.

Por ejemplo, supóngase que se tienen 4 tratamientos A, B, C, y D y cuatro bloques, un posible diseño en el que los cuatro tratamientos estén asignados aleatoriamente en cada bloque es:

Unidad Experimental en Bloque

		1	2	3	4
Bloque	I	D	B	A	C
	II	B	A	C	D
	III	B	C	D	A
	IV	C	A	D	B

En muchos problemas de investigación es necesario diseñar experimentos en los que pueda controlarse sistemáticamente la variabilidad producida por diversas fuentes extrañas, por ejemplo: un experimento con un solo factor (es decir lo que se quiere probar).

Lápiz → factor es la punta del lápiz, si se tienen 4 puntas con 4 observaciones resulta que son  $4 \times 4 = 16$  ensayos pero también es necesario considerar el lugar donde va a ser probada la punta del lápiz como puede ser una hoja que a su vez puede ser de diferentes materiales, por lo tanto tenemos un diseño aleatorizado por bloques completos, es decir que todos los tratamientos son probados en cada bloque (tipos de hoja).

Para el análisis se supone que no existe interacción entre los bloques y los tratamientos. Esta técnica es una generalización de la prueba de comparación de medias de muestras apareadas donde cada pareja es un bloque

### Las Ventajas de Bloques aleatorios:

- 1 La ventaja principal de este diseño es que toma en cuenta el efecto de los bloques, mientras que el diseño completamente aleatorio incluye este efecto en el error, es decir, se puede obtener mayor sensibilidad al utilizar el diseño en bloques aleatorios
- 2 El análisis de este diseño es sencillo, aun cuando exista el problema de datos perdidos

### Las Desventajas:

1. La principal desventaja es que el número de tratamientos tiene que ser el mismo que el número de unidades experimentales en cada bloque



## 2.1).- ANÁLISIS ESTADÍSTICO

El diseño de bloques aleatorizados y el diseño en cuadro latino son ilustraciones de experimentos que contienen una, dos y tres variables cualitativas independientes respectivamente. El análisis de varianza divide la suma total de los cuadrados de las desviaciones de las observaciones respecto a su media en partes asociadas con cada variable independiente y con el error experimental. La suma de cuadrados asociados con las variables independientes se pueden comparar con la suma de cuadrados de error, usando los cuadrados medios y el estadístico F, para ver si los cuadrados medios de las variables independientes son grandes, indicando así la existencia de un efecto sobre la respuesta

Supóngase que existen "t" tratamientos y "r" bloques (repeticiones), entonces se puede tener la siguiente representación:

Tratamientos $i = 1, \dots, t$	Bloques $j = 1, \dots, r$			Suma $X_i$	Media $\bar{X}_i$
	1	j	r		
1	$x_{11}$	$x_{1j}$	$x_{1r}$	$x_{1.}$	$\bar{x}_{1.}$
.	.	.	.	.	.
y	$x_{i1}$	$x_{ij}$	$x_{ir}$	$x_{i.}$	$\bar{x}_{i.}$
.	.	.	.	.	.
t	$x_{t1}$	$x_{tj}$	$x_{tr}$	$x_{t.}$	$\bar{x}_{t.}$
				$\Sigma$	$\Sigma$

Suma  $X_j$

$x_{.1} \dots x_{.j} \dots x_{.r}$

$x_{..}$

Media  $\bar{X}_j$

$\bar{x}_{.1} \dots \bar{x}_{.j} \dots \bar{x}_{.r}$

$\bar{x}_{..}$

Donde:

\*  $X_{ij}$  = La observación de la variable de respuesta medida en el i-ésimo tratamiento y el i-ésimo bloque

\*  $\longrightarrow$  la flecha nos indica el sentido hacia donde se hace la sumatoria de cada renglón ( $x_{i.}$ ), el dato obtenido será dividido entre la cantidad de datos que contenga cada renglón para poder obtener su media

\*  $\downarrow$  es el sentido para obtener la sumatoria por cada columna y sucesivamente obtener su media, dividiéndola entre la cantidad de datos que contenga cada columna

\*  $X_{..}$  es la suma total (gran total) de los renglones y de las columnas, es decir cada suma que se obtuvo por renglón ( $X_{i.}$ ) se vuelven a sumar todas, y cada suma que se obtuvo por

columna también se suman todas y tanto la sumatoria de columnas como de renglones tienen que dar el mismo resultado.

\*  $\bar{x} \dots$  es la media de las medias de todos los renglones y todas las columnas, es decir las medias obtenidas en cada renglón se suman y se dividen entre la cantidad de renglones y para las columnas es lo mismo y ambos datos tienen que ser iguales

\*  $r$  = número total de bloques (columnas)

\*  $t$  = número total de tratamientos (renglones)

### 2.1.1) ETAPAS PARA EL CALCULO DE BLOQUES ALEATORIOS

Al igual que el análisis de varianza los pasos que a continuación se muestran fueron obtenidos de la recopilación bibliográfica, tratando de presentarlo de una manera sencilla.

- Hipótesis nula:  $\mu_1 = \mu_2 = \mu_3$ ; Hipótesis alterna: las  $\mu$  no son iguales
- Nivel de significancia:  $\alpha = 0.05, 0.01\%$  la cual está en función de lo propuesto
- Criterio
- Cálculos; la tabla de ANDEVA se desarrolla como sigue:

1) Calcular el factor de corrección (C), elevar al cuadrado la suma total y dividir entre el número total de bloques y el número total de renglones

$$C = X^2 \dots / (r \cdot t)$$

2) Sumar los cuadrados de cada una de las observaciones y restar el factor de corrección C. A esto se le denomina Suma de Cuadrados Total (SCTL)

$$SCTL = \sum_{i=1}^t \sum_{j=1}^r x_{ij}^2 - C$$

3) Se eleva al cuadrado la sumatoria de cada renglón y se suman

$$\sum_{i=1}^t x_i^2$$

4) Se eleva al cuadrado la sumatoria de cada columna y se suman

$$\sum_{j=1}^r x_j^2$$

5) Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_i$  (dato obtenido en el paso 3), dividirse por el número total de bloques  $r$  y réstese el factor de corrección. A esto se le llama Suma de Cuadrados de Tratamientos (SCTR)

$$\sum_{i=1}^t x_i^2 / r - C$$

6) Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_j$  (dato obtenido en el paso 4), dividir entre el número total de renglones  $t$  y restar el factor de corrección. A esto se le llama Suma de Cuadrados de Bloques ( SCBL)

$$\sum_{j=1}^r x_j^2 / t - C$$

7) A la suma de cuadrados totales (SCTL) restar la suma obtenida entre la suma de cuadrados de tratamiento (SCTR) y la suma de cuadrado de bloques (SCBL) y se obtiene la Suma de Cuadrados debido al Error (SCER), que en este caso es una medida de la variación.

$$SCER = SCTL - (SCTR + SCBL)$$

8) Determinar los Grados de Libertad (g.l), para cada SC y se determina como sigue:

Para los tratamientos:  $(t - 1)$   
 Para los bloques :  $(r - 1)$   
 Para el error :  $(t - 1)(r - 1)$   
 Para el total :  $r \cdot t - 1$

9) Calcular los Cuadrados Medios de los Tratamientos y del Error (CMTR, CMER y CMBL) dividiendo sus respectivas sumas de cuadrados entre sus respectivos grados de libertad

$$\begin{aligned} CMTR &= SCTR / (t - 1) \\ CMBL &= SCBL / (r - 1) \\ CMER &= SCER / [(t - 1)(r - 1)] \end{aligned}$$

10) Calcular el valor de  $F_c$  dividiendo CMTR entre CMBL, a esto se le llama F Calculada

$$F_c = CMTR / CMBL$$

11) Se formula la tabla ANDEVA. La tabla de análisis de varianza para este diseño es la siguiente:

Tabla de Análisis para Bloques Aleatorios  
(ANDEVA)

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	$F_c$
Tratamientos	$t - 1$	$SCTR = \sum_{j=1}^t x_j^2 / r - C$	$CMTR = SCTR / (t - 1)$	$F_c = CMTR / CMER$
Bloques o Repeticiones	$r - 1$	$SCBL = \sum_{j=1}^r x_j^2 / t - C$	$CMBL = SCBL / (r - 1)$	

Error	$(t - 1)(r - 1)$	$SCER = SCTL - (SCTR + SCBL)$	$CMER = SCER / [(t - 1)(r - 1)]$
Total	$rt - 1$	$SCTL = \sum_{j=1}^t \sum_{i=1}^r x_{ij}^2 - C$	

12) La prueba F se aplica en forma similar, es decir, si  $F_c > F_t$  se rechaza la hipótesis nula, donde  $F_t$  se obtiene de las tablas de Distribución F con un nivel de significancia  $\alpha$  para los tratamientos con  $(t - 1)$  y  $(t - 1)(r - 1)$  grados de libertad. En caso de que se rechace la  $H_0$ , las comparaciones múltiples entre las medias se hace en la misma forma del Análisis de Varianza utilizando la Prueba a Posteriori.

## RESUMEN

Hay dos puntos que deben tomarse en cuenta al usar diseños con bloques. Primero no se debe utilizar un diseño de bloques aleatorizados para investigar los efectos de dos factores sobre una variable de respuesta (tomando los tratamientos como niveles de un factor y los bloques como niveles de otro factor). Si los efectos de los dos factores sobre la respuesta no son independientes entre sí, esto es, si el efecto de un factor sobre la respuesta depende del nivel del otro factor.

Un análisis de varianza de bloques aleatorizados podría conducir a conclusiones muy erróneas a cerca de la relación entre los factores y la respuesta. Recuérdese que el objetivo de los diseños de bloques es eliminar la variación introducida por variables no importantes (bloques) y reducir así SCE.

El segundo punto es que la introducción de bloques no siempre es beneficiosa, ya que produce una ganancia de información si la variación entre bloques es mayor que la variación dentro de bloques. En este caso la formación de bloques separa esta fuente mayor de variación de SCE y  $s^2$  toma un valor menor debido al diseño, al mismo tiempo se pierde información ya que la formación de bloques reduce el número de grados de libertad asociados con SCE.

Por consiguiente para que el uso de bloques sea beneficioso, la información de ganancia debida a la eliminación de la variación de los bloques debe ser mayor que la pérdida debida a la reducción del número de grados de libertad asociados con SCE. A menos que el uso de bloques ocasione que el número de grados de libertad de SCE sea pequeño, la pérdida de grados de libertad causa solo una pequeña reducción de la información en el experimento.

1.2).- PROBLEMAS RESUELTOS

1) Cuatro diferentes cultivos (agar) A, B, C y D son utilizados en un experimento para comparar los rendimientos de un microorganismo. Supóngase que un lote de material es suficiente para efectuar cuatro pruebas; si se utilizan cuatro lotes de material (como los cuatro bloques), cada lote es dividido en cuatro partes y a cada parte se le asigna aleatoriamente un cultivo, de tal forma que en cada lote estén asignados los cuatro tipos de cultivos. Los rendimientos, en tanto por cierto, son los siguientes

Tratamiento	Bloques			
	I	II	III	IV
A	89	90	92	86
B	86	88	90	90
C	89	92	93	91
D	88	90	94	92

Con un nivel de significancia del 5%

- a) ¿ Existe alguna diferencia en el rendimiento de la reacción utilizando diferentes tipos de cultivos?  
 b) En caso de que la prueba F sea significativa, realice las comparaciones de medias utilizando DMS y DMSH.

Solución:

Tratamientos	Bloques				Suma $X_i$	Media $\bar{x}_i$
	I	II	III	IV		
A	89	90	92	86	357	89.25
B	86	88	90	90	354	88.50
C	89	92	93	91	365	91.25
D	88	90	94	92	364	91.00
Suma $X_j$	352	360	369	359	$X_{..} = 1440$	
Media $\bar{x}_j$	88	90	92.25	89.75		$\bar{x}_{..} = 90$

$$r = 4, t = 4$$

Donde:  $X_i$  es la suma de cada renglón

$X_j$  es la suma de cada columna

$\bar{x}_i$  es la media de cada renglón

$\bar{x}_j$  es la media de cada columna

$X_{..}$  es la suma total de los tratamientos

$\bar{x}$ . es la media de todos los tratamientos

→ indica el sentido de la sumatoria

1) Calcular el factor de corrección

$$C = X^2_{..} / (r \cdot t)$$

$$C = (1440)^2 / (4 \times 4) = 129,600$$

2) Sumar los cuadrados de cada una de las observaciones restándole el factor de corrección C.

$$SCTL = \sum_{i=1}^t \sum_{j=1}^r x_{ij}^2 - C$$

$$SCTL = (89)^2 + (86)^2 + (89)^2 + (88)^2 + (90)^2 + (88)^2 + (92)^2 + (90)^2 + (92)^2 + (90)^2 + (93)^2 + (94)^2 + (86)^2 + (90)^2 + (91)^2 + (92)^2 - (129,600) =$$

$$SCTL = 129680 - 129600 = 80.0$$

3) Elevar al cuadrado la sumatoria de cada renglón y se suman

$$\sum_{i=1}^t x_i^2 = (357)^2 + (354)^2 + (365)^2 + (364)^2 = 518486$$

4) Elevar al cuadrado la sumatoria de cada columna y se suman

$$\sum_{j=1}^r x_{.j}^2 = (352)^2 + (360)^2 + (369)^2 + (359)^2 = 518546$$

5) Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_i$  (dato obtenido en el paso 3), dividase por el número total de bloques r y réstese el factor de corrección.

$$SCTR = \sum_{i=1}^t x_i^2 / r - C$$

$$SCTR = (518486 / 4) - 129600 = 21.5$$

6) Sumar los cuadrados de cada uno de los totales de los tratamientos  $X_{.j}$  (dato obtenido en el paso 4), dividase entre el número total de renglones t y réstese el factor de corrección.

$$SCBL = \sum_{j=1}^r x_{.j}^2 / t - C$$

$$SCBL = (518546 / 4) - 129600 = 36.5$$

7) Calcular la suma de cuadrados de error.

$$SCER = SCTL - (SCTR + SCBL)$$

$$SCER = 80 - (21.5 + 36.5) = 22.0$$

8) Determinar los Grados de Libertad (g.l), para cada SC

$$\text{Para los tratamientos: } (t - 1) = (4 - 1) = 3$$

$$\text{Para los bloques : } (r - 1) = (4 - 1) = 3$$

$$\text{Para el error : } (t - 1)(r - 1) = (4 - 1)(4 - 1) = 9$$

9) Calcular los Cuadrados Medios de los Tratamientos y del Error ( CMTR, CMER y CMBL) dividiendo sus respectivas sumas de cuadrados entre sus respectivos grados de libertad

$$\text{CMTR} = \text{SCTR} / (t - 1) = 21.5 / (4 - 1) = 7.166$$

$$\text{CMBL} = \text{SCBL} / (r - 1) = 36.5 / (4 - 1) = 12.166$$

$$\text{CMER} = \text{SCER} / [(t - 1)(r - 1)] = 22 / [(4 - 1)(4 - 1)] = 2.444$$

10) Calcular el valor de  $F_c$

$$F_c = \text{CMTR} / \text{CMER}$$

$$F_c = 7.166 / 2.444 = 2.93$$

11) Formular la tabla ANDEVA. La tabla de análisis de varianza para este diseño es la siguiente:

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	$F_c$
Tratamientos	3	21.5	7.166	2.93
Bloques	3	36.5	12.166	
Error	9	22.0	2.444	
Total	15	80.0		

12) Con la tabla de Distribución F se obtienen la  $F_t$  con un nivel de significancia de 5%, 3 y 9 g.l

$$F_t = 3.86$$

- a) Como  $F_c$  es menor que  $F_t$  aceptamos la  $H_0$ , es decir, no existe diferencia significativa entre los cultivos es decir el rendimiento de los microorganismos será el mismo utilizando cualquier tipo de agar.
- b) Debido a que la prueba F no resulto significativa, no se realizan las comparaciones de las medias ( pruebas a posteriori)

2) Un Ingeniero de Control de Calidad para la manufactura de componentes electrónicos, intuitivamente siente que hay mucha variabilidad entre los 36 hornos usados por la compañía manufacturera para la cual trabaja, en las pruebas de duración de los diversos componentes. Para determinar si tiene razón, escoge un solo tipo de componente y obtiene los siguientes datos para las dos temperaturas (T) normalmente usadas para la prueba de duración de estas unidades:  $T_1 = 650^\circ\text{F}$  y  $T_2 = 600^\circ\text{F}$ . El componente se pone a funcionar en un horno hasta que falla, en este experimento se utilizaron 3 hornos (H) seleccionados aleatoriamente y se midió los minutos de duración del componente electrónico

	T	T <sub>1</sub>	T <sub>2</sub>
H			
	H <sub>1</sub>	246	180
	H <sub>2</sub>	191	144
	H <sub>3</sub>	187	134

- a) Con 5% de significación, pruebe si el Ingeniero tiene razón en su creencia de la variabilidad entre los 36 hornos, a partir de los datos anteriores.
- b) Si la prueba anterior le resulta significativa, efectúe las pruebas a posteriori DMS y DMSH con el 5% de significación.

Solución:

a) T = temperatura  $\Rightarrow r = 2$  Temperaturas (Bloques)

H = Horno  $\Rightarrow t = 3$  Hornos (Tratamientos)

(t)(r) = 6 Componentes electrónicos (unidades experimentales [u.e])

X<sub>ij</sub> = Duración en minutos del componente electrónico probado en el horno i-ésimo, con la temperatura j-ésima.

	T	T <sub>1</sub>	T <sub>2</sub>	Suma Xi	Media $\bar{X}_j$
H					
	H <sub>1</sub>	246	180	426	213
	H <sub>2</sub>	191	144	335	167.5
	H <sub>3</sub>	187	134	321	160.5
	Suma X <sub>j</sub>	624	458	X <sub>..</sub> = 1082	
	Media $\bar{x}_j$	208	152.6		$\bar{x}_{..} = 180.3$

1) Calcular el factor de corrección

$$C = X_{..}^2 / (r \cdot t)$$

$$C = (1082)^2 / (2 \times 3) = 195,120.67$$

$$2) SCTL = \sum_{i=1}^t \sum_{j=1}^r x_{ij}^2 - C = [(246)^2 + (191)^2 + \dots + (134)^2] - 195,120.67$$

$$SCTL = 203,058 - 195,120.67 = 7937.33$$

$$3) \sum_{i=1}^t x_i^2 = (426)^2 + (335)^2 + (321)^2 = 396,742$$

$$4) \sum_{j=1}^r x_j^2 = (624)^2 + (458)^2 = 599,140$$



$$5) SCTR = \sum_{i=1}^t x_i^2 / r - C$$

$$SCTR = (396,742 / 2) - 195,120.67 = 3250.33$$

$$6) SCBL = \sum_{j=1}^r x_j^2 / t - C$$

$$SCBL = (599,140 / 3) - 195,120.67 = 4592.67$$

7) Calcular la suma de cuadrados de error.

$$SCER = SCTL - (SCTR + SCBL)$$

$$SCER = 7937.33 - (3250.33 + 4592.67) = 94.33$$

8) Determinar los Grados de Libertad (g.l), para cada SC

$$\text{Para los tratamientos: } (t - 1) = (3 - 1) = 2$$

$$\text{Para los bloques : } (r - 1) = (2 - 1) = 1$$

$$\text{Para el error : } (t - 1)(r - 1) = (3 - 1)(2 - 1) = 2$$

9) Calcular los Cuadrados Medios de los Tratamientos y del Error

$$CMTR = SCTR / (t - 1) = 3250.33 / (3 - 1) = 1625.165$$

$$CMBL = SCBL / (r - 1) = 4592.67 / (2 - 1) = 4592.67$$

$$CMER = SCER / [(t - 1)(r - 1)] = 94.33 / [(3 - 1)(2 - 1)] = 47.165$$

10) Calcular el valor de  $F_c$

$$F_c = CMTR / CMER$$

$$F_c = 1625.165 / 47.165 = 34.45$$

11) Formular la tabla ANDEVA.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	$F_c$
Tratamientos (Hornos)	2	3,250.33	1,625.165	34.45
Bloques (Temperatura)	1	4,592.67	4,592.67	
Error	2	94.33	47.165	
Total	5	7,937.33		

12) Con la tabla de Distribución F se obtienen la  $F_t$  con un nivel de significancia de 5%, 2 y 2 g.l

$$F_t = 19.0$$

a) Por lo tanto como  $F_c = 34,45 > F_t = 19$ , se puede decir que la variabilidad entre los 36 hornos de la compañía si es significativa.

b) Por lo tanto es necesario realizar las pruebas a posteriori

**Solución:**

i) utilizando la prueba de la diferencia mínima significativa, se tendrá que calcular el valor de la DMS.

$$DMS = t_1 * S_{\bar{x}_i - \bar{x}_j}$$

Localizar el valor de  $t_1$  con 2 g.l ( g.l de error) y  $\alpha = 5\%$  se tiene:

$$t_1 = 4.303$$

$$\text{Calcular ahora el valor de } S_{\bar{x}_i - \bar{x}_j} = \sqrt{2s^2/r} = \sqrt{2(CMER)/r}$$

$$S_{\bar{x}_i - \bar{x}_j} = \sqrt{2(47.167/2)} = 6.86$$

$$\text{entonces } DMS = ( 4.303 ) ( 6.86 ) = 29.553$$

La diferencia entre dos medias que exceda este valor se considera estadísticamente diferente

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
$H_1 - H_3$	52.5	$H_1$ mayor que $H_3$
$H_1 - H_2$	45.5	$H_1$ mayor que $H_2$
$H_2 - H_3$	7.0	ns

ii) Utilizando la prueba de la diferencia media significativa honesta, calculamos el valor de la DMSH

$$DMSH = q_{t, \alpha, g.l} * S_{\bar{x}}$$

Localizando el valor de  $q$  de la tabla de rango estudentizado con  $\alpha = 5\%$ ,  $t = 43$  (que es el número de tratamientos) y los grados de libertad del error = 2

$$\text{Calcular ahora el valor de } S_{\bar{x}} = \sqrt{s^2/r} = \sqrt{CMER/r}$$

$$S_{\bar{x}} = \sqrt{47.167/2} = 4.856$$

$$\text{entonces } DMSH = ( 8.33 ) ( 4.856 ) = 40.453$$

La diferencia entre dos medias que excede este valor se considera estadísticamente diferentes

Comparación	Diferencia $\bar{x}_i - \bar{x}_j$	Significancia
$H_1 - H_3$	52.5	$H_1$ mayor que $H_3$
$H_1 - H_2$	45.5	$H_1$ mayor que $H_2$
$H_2 - H_3$	7.0	ns

Se puede concluir que en el horno  $H_1$  el componente electrónico dura más que en los hornos  $H_2$  y  $H_3$ , lo que corrobora que si existe variabilidad entre los hornos utilizados para las pruebas.

### III.- ANÁLISIS FACTORIAL

*Libros 6, 9, 15, 16, 17, 19, 27, 28.*

*El diseño factorial es un diseño de los tratamientos que se caracteriza por utilizar como tratamientos a todas las combinaciones posibles de las diferentes necesidades (conocidas como niveles) de cada una de dos más variables (conocidas como factores) que probablemente afecten la variable de respuesta  $x$  en un experimento*

El Análisis Factorial es un efecto producido por 2 o más factores siendo los más eficientes en este tipo de experimentos, es decir en un experimento en que se interesa estudiar dos factores A y B con a niveles y b niveles, respectivamente se ensayarán todas las  $a*b$  posibles combinaciones de niveles, como tratamientos. Por diseño factorial se entiende aquel en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento, a menudo se dice que los factores están cruzados, cuando éstos se arreglan en un diseño factorial.

El efecto principal es el cambio de la respuesta producida por un cambio en el nivel del factor.

Los diseños factoriales son necesarios cuando alguna interacción puede estar presente, con el propósito de hacer conclusiones que son engañosas.

- 1) Se puede estudiar los efectos de diversos factores en el mismo conjunto de experimentos
- 2) Podemos comprobar el efecto de cada factor en todos los niveles de los otros factores y podemos describir si este efecto cambia o no al cambiar aquellos
- 3) Podemos comprobar no sólo los efectos de los factores separadamente, si no también las interacciones-efectos conjuntos de dos o más factores combinados
- 4) Cada juicio que formulemos acerca de los efectos de los factores está basado en todas las observaciones acumuladas en todo el conjunto de experimentos y no solo sobre unas pocas observaciones. Así los experimentos factoriales son más sensibles en la detección de efectos reducidos.

El diseño factorial de los tratamientos se puede utilizar con cualquier diseño experimental de las unidades experimentales, por ejemplo, con el diseño completamente al azar ó con el diseño de bloques aleatorios etc.

A continuación se vera el caso más sencillo, en el cual cada uno de los factores A y B tienen dos niveles. En este caso se denomina como diseño experimental.

#### 3.1).- ANÁLISIS ESTADÍSTICO

Se tienen  $4r$  observaciones de la variable de respuesta  $X$  que se puede arreglar de la siguiente manera:

		Factor B	
		$b_1$	$b_2$
Factor A	$a_1$	$X_{111}$	$X_{121}$
		$X_{112}$	$X_{122}$
		.	.
	$a_2$	$X_{211}$	$X_{221}$
		$X_{212}$	$X_{222}$
		$X_{21r}$	$X_{22r}$

Según el diseño experimental utilizado, se calcula la tabla de ANOVA como de costumbre, sin embargo la suma de cuadrados (SCTR) se acostumbra en desglosar en 3 sumas de cuadrados:

$SC_A$  asociada con el efecto principal del factor A

$SC_B$  que se asocia con el efecto principal del factor B

$SC_{A \times B}$  que se asocia con la interacción de A y B. Cada SC tiene 1 g.l. en el caso de experimento factorial  $2 \times 2$ .

Una manera de calcular cada efecto y SC en un diseño  $2 \times 2$  es subdividir la SCTR, utilizando los siguientes 3 contrastes ortogonales:

Tabla 1

Efecto factorial (contraste lineal)	Multiplicador $\lambda$ para los totales del tratamiento $a_1b_1, a_2b_2, a_1b_2, a_2b_1$	Total para el efecto factorial
A	-1    1    -1    1	$[A] = -a_1b_1 + a_2b_1 - a_1b_2 + a_2b_2$
B	-1    -1    1    1	$[B] = -a_1b_1 - a_2b_1 + a_1b_2 + a_2b_2$
A x B	1    -1    -1    1	$[A \times B] = a_1b_1 - a_2b_1 - a_1b_2 + a_2b_2$

El efecto de un factor se define como el cambio en la respuesta producida por un cambio en el nivel del factor. Con frecuencia éste se conoce como *efecto principal* porque se refiere a los factores de interés primordial del experimento.

El efecto principal para cada factor se calcula dividiendo el total para el efecto factorial entre  $2r$  para A:

$$\text{Efecto principal de A} = \frac{[A]}{2r}$$

Por ejemplo consideremos los datos de la tabla 2. El efecto principal del factor A podría interpretarse como la diferencia entre la respuesta promedio en el primero y segundo nivel de ese factor. Numéricamente.

**Tabla 2**  
Experimento factorial

		Factor B	
		B <sub>1</sub>	B <sub>2</sub>
Factor A	A <sub>1</sub>	20	30
	A <sub>2</sub>	40	52

$$A = \frac{[A]}{2r}$$

De la tabla I  $[A] = -a_1b_1 + a_2b_1 - a_1b_2 + a_2b_2 = (a_2b_1 + a_2b_2) - (a_1b_1 + a_1b_2)$

$$A = (a_2b_1 + a_2b_2) - (a_1b_1 + a_1b_2) / 2r$$

$$A = \frac{40 + 52}{2} - \frac{20 + 30}{2} = 21$$

En otras palabras, incrementar el factor A del nivel 1 y 2 produce un cambio en la respuesta promedio de 21 unidades. Sucede lo mismo en el factor B.

Usualmente, para obtener conclusiones acerca del efecto de una interacción significativa el experimentador debe examinar los niveles de dicho factor, manteniendo fijos los niveles de los otros factores.

La contribución a la SCTR se calcula dividiendo el cuadrado de cada total para el efecto factorial entre  $4r$ :

$$SC_A = [A]^2 / 4r$$

$$SC_B = [B]^2 / 4r$$

$$SC_{A*B} = [A*B]^2 / 4r$$

Y la tabla de ANOVA queda como sigue (para un diseño experimental con  $r$  bloques aleatorios).

F.V	g.l	S.C	C.M	F
Tratamientos	3	SCTR	CMTR	CMTR/CMER
A	1	SC <sub>A</sub>	CM <sub>A</sub>	CMA/CMER=F <sub>A</sub>
B	1	SC <sub>B</sub>	CM <sub>B</sub>	CMB/CMER=F <sub>B</sub>
Interacción A*B	1	SC <sub>A*B</sub>	CM <sub>A*B</sub>	CM <sub>A*B</sub> /CMER=F <sub>A*B</sub>
Bloques	$r-1$	SCBL	CMBL	
Error	$3(r-1)$	SCER	CMER	
Total	$4r-1$	SCTL		

### 3.2). - INTERPRETACION DE LOS EFECTOS PRINCIPALES Y DE LA INTERACCIÓN

\*El efecto simple de A cuando B toma nivel  $b_j$ :

∴  $(A)b_j$  es igual al cambio en la variable de respuesta X cuando se cambia de tratamiento  $a_1b_j$  al tratamiento  $a_2b_j$ .

\*El efecto simple de A cuando B toma nivel  $b_2$ :

=  $(A)b_2$  es igual al cambio en X cuando se cambia del tratamiento  $a_1b_2$  al tratamiento  $a_2b_2$ .

\*El efecto principal de A es el promedio de los dos efectos simples de A.

\*El efecto principal de B se define de manera similar.

La interacción de A y B es igual a la diferencia del efecto simple de A cuando B toma nivel  $b_2$  menos el nivel simple de A cuando B toma nivel  $b_1 = (A)b_2 - (A)b_1$ . Es decir, si el efecto sobre X debe cambiar del nivel  $a_1$  al nivel  $a_2$  del factor A es igual para los dos niveles de B, entonces la interacción es cero ("no hay interacción" ó "los factores operan independientemente"). Si el efecto simple de A depende del nivel de B entonces hay interacción. Dependiendo del signo de la interacción, ésta se conoce como "un efecto sinergizante" o "potencialización" (signo positivo) y como "efecto antagónico" o interferencia (signo negativo). También se puede detectar e interpretar la interacción de manera gráfica como se verá más adelante.

### 3.2.1) LAS HIPOTESIS QUE SE PRUEBAN.

Las pruebas F de la tabla de ANOVA corresponden a las siguientes hipótesis:

$F_A$	$H_0$ : El efecto principal de A=0
	$H_1$ : El efecto principal de A $\neq$ 0
$F_B$	$H_0$ : El efecto principal de B=0
	$H_1$ : El efecto principal de B $\neq$ 0
$F_{A*b}$	$H_0$ : No hay interacción
	$H_1$ : Hay interacción

### 3.3.- ETAPAS PARA EL ANÁLISIS FACTORIAL

Los pasos que se siguen para poder obtener un análisis factorial son cuatro, éstos fueron obtenidos de la recopilación bibliográfica. Y son los siguientes:

1.- Formar la tabla de ANOVA como se mostró anteriormente, indicando cuales resultados son estadísticamente significativos y a que nivel.

2.- Presentar una Tabla de medias y Errores:

		FACTOR B		
		Efecto simple		Efecto principal
$a_1$	$\bar{X}_{11}$	$\bar{X}_{12}$	$\bar{X}_{12}$	$((B)a_1 + (B)a_2) / 2 \pm S / \sqrt{r}$
			$(B)a_1 = -\bar{X}_{11}$	
$a_2$	$\bar{X}_{21}$	$\bar{X}_{22}$	$\bar{X}_{22}$	
			$(B)a_2 = -\bar{X}_{21}$	

### FACTOR A

Efecto simple  $= \bar{X}_{21} - \bar{X}_{11}, \bar{X}_{22} - \bar{X}_{12}$  donde  $s = \sqrt{CMER}$

Efecto principal  $((A)b_1 + (A)b_2) / 2 \pm S / \sqrt{r}$

Interacción de A\*B  $((A)b_2 - (A)b_1) / 2 \pm S / \sqrt{r}$

#### Observación:

Recuérdese que para el cálculo de los errores estándar de los efectos principales se tienen los resultados de la Técnica sobre  $S_L$ . Por ejemplo el error estándar del efecto principal del Factor A se basa en la combinación lineal.

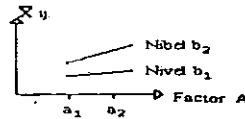
$$L = \bar{X}_{22} / 2 - \bar{X}_{21} / 2 + \bar{X}_{12} / 2 - \bar{X}_{11} / 2$$

y por lo tanto:

$$\begin{aligned} s_L &= \sqrt{\sum \frac{l_i^2}{X} s^2} \\ &= \sqrt{(1/2)^2 + (-1/2)^2 + (1/2)^2 + (-1/2)^2} s / \sqrt{r} \\ &= s / \sqrt{r} \end{aligned}$$

#### 3. Dibujar la gráfica correspondiente.

Valor promedio  
de la variable de  
respuesta



#### Nota:

Como se mencionó anteriormente en el caso de no-intersección habrá paralelismo esencialmente entre las dos líneas en esta gráfica, y en el caso en el que hay interacción habrá un no-paralelismo "significativo".

4.- Redactar un reporte completo y claro, pero al mismo tiempo conciso de los resultados mostrados en los puntos 1, 2 y 3.

### 3.3.1) VENTAJAS Y DESVENTAJAS DEL ANÁLISIS FACTORIAL

#### VENTAJAS

Se puede decir que los diseños experimentales factoriales poseen algunas ventajas. Son más eficientes que los experimentos de un factor a la vez. Más aun, los diseños factoriales son necesarios cuando alguna interacción puede estar presente, con el propósito de evitar hacer conclusiones que son engañosas. Finalmente, los diseños factoriales, permiten estimar los efectos de un factor en diversos niveles de los otros factores, produciendo conclusiones que son válidas sobre todo la extensión de las condiciones experimentales.

#### DESVENTAJAS

Si el número de factores y/o niveles es grande, cada replica indica muchas unidades experimentales que pueden no estar fácilmente disponible o resultan ser muy heterogéneas.



3.4).- DISEÑO FACTORIAL EN DOS FACTORES Y SUBDIVISIÓN DE LOS GRADOS DE LIBERTAD.

3.4.1) INTRODUCCIÓN Y NOTACIÓN

En la técnica anterior se trató en detalle el caso del diseño factorial en 2 factores con dos niveles cada uno, el diseño  $2 \times 2$ . Aquí se verá la generalización al caso en que los factores A y B tienen un número arbitrario de niveles los cuales denotamos por a y b respectivamente, (el diseño factorial  $a \times b$ ). Se hará hincapié en los puntos donde el diseño  $a \times b$  difiere del diseño  $2 \times 2$ .

Los  $a \cdot b \cdot r$  observaciones se pueden representar de la siguiente manera (denotando con N a las observaciones de la variable de respuesta como se ha acostumbrado):

$$X_{ijk} \quad \text{con} \quad \begin{aligned} i &= 1, 2, 3, \dots, n. \\ j &= 1, 2, 3, \dots, n. \\ k &= 1, 2, 3, \dots, n. \end{aligned}$$

donde:

i denota el nivel del factor A

j denota el nivel del factor B

y k denota el número de la replicación.

Nota

Se considerará solo el caso de igual número de repeticiones r para cada uno de los a, b tratamientos, el caso balanceado como se conoce. En el caso desbalanceado existen varias dificultades serias.

Estos datos se arreglan en una tabla de dos sentidos como sigue:

		FACTOR B			
		1	2	...	b
1		$X_{111}$	$X_{121}$		$X_{1b1}$
		$X_{112}$	$X_{122}$		$X_{1b2}$
		:	:		:
		$X_{11r}$	$X_{12r}$		$X_{1br}$
2		$X_{211}$	$X_{221}$		$X_{2b1}$
		$X_{212}$	$X_{222}$		$X_{2b2}$
		:	:		:
		$X_{21r}$	$X_{22r}$		$X_{2br}$
a		$X_{a11}$	$X_{a21}$		$X_{ab1}$
		$X_{a12}$	$X_{a22}$		$X_{ab2}$
		:	:		:
		$X_{a1r}$	$X_{a2r}$		$X_{abr}$

La notación  $X_{ij}$  sigue denotando la suma sobre el índice (ó índices) que están con  $\cdot$ , p. ej.

$$X_{ij} = \sum_{k=1}^r X_{ijk} \quad ; \quad X_{ij\cdot} = \sum_{i=1}^a \sum_{k=1}^r X_{ijk}$$

Y con el símbolo  $\bar{X}_{\cdot j}$  (por ejemplo) se denota la media de la suma  $X_{\cdot j}$ , es decir:

$$\bar{X}_{\cdot j} = \frac{\sum_{i=1}^a \sum_{k=1}^r X_{ijk}}{a \cdot r}$$

### 3.4.2) ANÁLISIS ESTADÍSTICO.

Se calculan las sumas de cuadrados (SC) correspondientes a los Tratamientos, el Total y el Error (y a los Bloques si es que se utilizó un diseño en bloques aleatorios) de la manera acostumbrada:

$$C = \text{Factor de corrección} = X^2 \dots / (a \cdot b \cdot r)$$

$$SCTR = \sum_{j=1}^b \sum_{i=1}^a X_{ij}^2 / r - C$$

$$SCTL = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r X_{ijk}^2 - C$$

(En el caso de Diseño de bloques:

$$SCBL = \sum_{k=1}^r X_{\cdot k}^2 / (a \cdot b) - C$$

y la SCER se modifica a

$$SCER = SCTL - SCTR - SCBL.$$

La SCTR se subdivide en  $SC_A$ ,  $SC_B$  y  $SC_{A \cdot B}$  como sigue:

$$SC_A = \sum_{i=1}^a X_i^2 / (b \cdot r) - C$$

$$SC_B = \sum_{j=1}^b X_j^2 / (a \cdot r) - C$$

$$Y \quad SC_{A \cdot B} = SCTR - SC_A - SC_B.$$

#### Nota:

Si  $a = b = 2$  estas fórmulas dan el mismo resultado que las fórmulas dadas en la técnica anterior.

Se resumen las sumas de cuadrados en una tabla de ANOVA como sigue (se presenta el caso del Diseño en Bloques Aleatorios).

F.V	g.l	S.C	C.M	F
Tratamientos	$a \cdot b - 1$	SCTR		
A	$a - b$	$SC_A$	$CM_A = SC_A / (a - 1)$	$CM_A / CMER = F_A$
B	$b - 1$	$SC_B$	$CM_B = SC_B / (b - 1)$	$CM_B / CMER = F_B$
A*B	$(a - 1)(b - 1)$	$SC_{A \cdot B}$	$CM_{A \cdot B} = SC_{A \cdot B} / ((a - 1)(b - 1))$	$CM_{A \cdot B} / CMER = F_{A \cdot B}$
Bloques	$r - 1$	SCBL	$CMBL = SCBL / (r - 1)$	
Error	$(a \cdot b - 1)(r - 1)$	SCER	$CMER = SCER / ((a \cdot b - 1)(r - 1))$	
Total	$a \cdot b \cdot r - 1$	SCTL		

[En el caso del Diseño completamente Aleatorizado se suprime el renglón de bloques y se aumenta los g.l del error a  $(a*b)(r-1)$ ]

Las hipótesis que se prueban y el método de probarlos son iguales que en el caso del diseño  $2*2$  (técnica anterior). Los resultados se presentan de la misma manera que se describió en el caso  $2*2$  hasta donde sea posible. La tabla de medias y errores estándar se tiene que modificar si  $a>2$  ó  $b>2$  ya que los efectos principales son más complicados (tienen más de 1 g.l) y su interpretación no es tan clara. Sin embargo muchas veces es posible simplificar y esclarecer la situación utilizando la idea siguiente.

### 3.4.3).- SUBDIVISIÓN DE LOS G.L

En la técnica anterior se vio que cuando se tienen formuladas, al planear el experimento, hipótesis sobre comparaciones lineales de los tratamientos, éstas pruebas de hipótesis dan mucha información adicional sobre como operan los diferentes tratamientos. Esto es el caso siempre y cuando el estadístico F este asociado con más de un grado de libertad (g.l) y por lo tanto se puede partir la S.C en varios contrastes lineales. Esto se puede efectuar en el diseño factorial cuando la S.C de A ó B ó  $A*B$  tienen más de 1 g.l.

Las reglas de cómo partir las S.C en contrastes lineales para los diferentes efectos factoriales son las mismas que se dieron en la técnica anterior. Pero la selección de cuales contrastes son de interés queda a juicio del investigador y un uso óptimo de ésta técnica requiere de experiencia. Antes de remitir al lector a los ejemplos resueltos se ofrecen algunas indicaciones generales como guía:

1. De preferencia se debe utilizar contrastes ortogonales.
2. Si los niveles de un factor son numéricos, es conveniente subdividir los g.l asociados con el efecto principal e este factor en su componente lineal, cuadrático, cúbico, etc.
3. Si los niveles de un factor son nominales es recomendable subdividir los g.l asociados con el efecto principal en componentes del tipo: control v.s los casos, entre marcas, entre diferentes formulaciones con un mínimo principio activo, etc.
4. Se pueden determinar los coeficientes (lí's) de los contrastes para subdividir la S.C asociada a la interacción de  $A*B$ , multiplicando los coeficientes de un contraste para A con los coeficientes de un contraste para B. Por ejemplo el efecto lineal de A por control v.s el resto de B representa una comparación entre el efecto lineal de A para el control de B y el efecto lineal de A promediado sobre el resto de los tratamientos de B.

3.5).- PROBLEMAS RESUELTOS*Diseño factorial en dos factores con dos niveles cada uno*

1) En un experimento con remolacha se consideraban dos factores: A= profundidad de labranza con  $a_1=18\text{cm}$  contra  $a_2=28\text{cm}$ ; y B= nitrógeno con  $b_1= \text{sin él}$ , contra  $b_2=336.33\text{kg}$  de sulfato de amonio por ha. La labranza se hizo a finales de enero; el nitrógeno se aplicó a fine de abril y la semilla se sembró a principios de mayo. La variable de respuesta X es igual a rendimiento de azúcar en toneladas por ha. El diseño experimental fue en dos bloques aleatorios.

Rendimiento de azúcar (toneladas) de remolacha

		B= Nitrógeno	
		$b_1$	$b_2$
A= Profundidad	$a_1=18$ cm	4.51	5.47
	$a_2=28$ cm	4.67	5.38
		4.78	5.74
		4.68	5.55

Efectuar el análisis de varianza, probar cuales efectos son significativos e interpretar los resultados.

Solución:  
Calcular

TOTALES	$a_1b_1$	$a_2b_1$	$a_1b_2$	$a_2b_2$	TOTAL PARA EL EFECTO FACTORIAL
EFECTO	9.18	9.46	10.55	11.29	
A= PROFUNDIDAD	-1	1	-1	1	[A]= 0.72
B= NITROGENO	-1	-1	1	1	[B]= 3.50
A*B= INTERACCIÓN DE PROFUNDIDAD Y NITROGENO	1	-1	-1	1	[A*B]= 0.16

$$a_1b_1 = 4.51 + 4.67 = 9.18 \quad a_2b_1 = 4.78 + 4.68 = 9.46$$

$$a_1b_2 = 5.47 + 5.38 = 10.85 \quad a_2b_2 = 5.74 + 5.55 = 11.29$$

De la tabla 1

$$[A] = -a_1b_1 + a_2b_1 - a_1b_2 + a_2b_2 = -9.18 + 9.46 - 10.85 + 11.29 = 0.72$$

$$[B] = -a_1b_1 - a_2b_1 + a_1b_2 + a_2b_2 = -9.18 - 9.46 + 10.85 + 11.29 = 3.50$$

$$[AxB] = a_1b_1 - a_2b_1 - a_1b_2 + a_2b_2 = 9.18 - 9.46 - 10.85 + 11.29 = 0.16$$

Entonces las sumas de cuadrados (S.C) son:

$$SC_A = [A]^2 / 4r = [0.72]^2 / (4 \cdot 2) = 0.065$$

$$SC_B = [B]^2 / 4r = [3.50]^2 / (4 \cdot 2) = 1.531$$

$$SC_{A \cdot B} = [A \cdot B]^2 / 4r = [0.16]^2 / (4 \cdot 2) = 0.003$$

donde  $r=2$

Y calculando las otras sumas de cuadrados como de costumbre en los diseños en bloques aleatorios, se tiene:

$$\text{Donde } c = (9.18+9.46+10.85+11.29)^2 / 8 = 207.876$$

$$\text{SCTR} = \sum_{i=1}^2 \sum_{j=1}^2 x_{ij}^2 / r - c = (9.18^2 + \dots + 11.29^2) / 2 - 207.876 = 1.599$$

$$\text{SCBL} = \sum_{k=1}^r x^2 \dots_k / a - c = (20.502 + 20.282) / 4 - 207.876 = 0.006$$

$$\text{SCTL} = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^r x_{ijk}^2 - c = (4.51)^2 + \dots + (5.55)^2 - 207.876 = 1.639$$

$$\text{SCER} = \text{SCTL} - \text{SCTR} - \text{SCBL}$$

Y se tiene:

### 1. TABLA DE ANOVA

F.V	g.l	S.C	C.M	F
Tratamientos	3	1.599	<u>.11mCMTR</u>	-8.5
A= Profundidad	1	0.065	0.065	5.9(0.06 < F < 0.10)
B= Nitrógeno	1	1.531	1.531	139.2*** (F < 0.006)
A*B = Interacción	1	0.003	0.003	0.3 n.s
Bloques	1	0.006	0.006	
Error	3	0.034	0.011	
Total	4r-1	1.639		

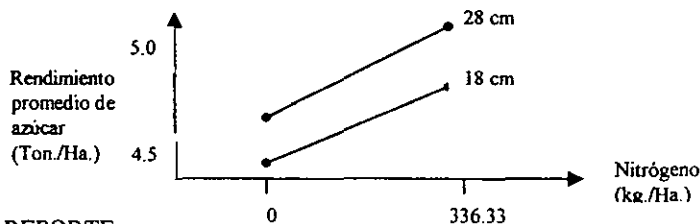
### 2. Tabla de medias y errores estándar para rendimiento de remolacha (Ton/Ha)

		NITROGENO			
		0 K	335.33 K	Efecto simple	Efecto principal
18 cm	4.59	5.43	0.84	0.88***(±0.007)	
28 cm	4.73	5.65	0.92		

Profundidad

Efecto simple	0.14	0.22
Efecto principal	0.18(± 0.07)	
Interacción	0.04 n.s.(± 0.07)	

## 3. GRAFICA



## 4. REPORTE

En base, a la tabla de ANOVA y la tabla de media se ve que la interacción es pequeña y no significativamente estadísticamente. Por lo tanto, los efectos principales sirven para resumir la información: "se espera un aumento de 880 kg ( $\pm 70$ ) de azúcar por la aplicación de nitrógeno casi importar la profundidad de la labranza. El surco de 28 cm es superior al surco de 18 cm en promedio en 180kg por hectárea ". El efecto del nitrógeno es altamente significativo ( $P < 0.005$ ) mientras que en la tabla de ANOVA no es claro si hay efecto estadísticamente de la profundidad de la labranza. Aquí nos puede auxiliar la tabla de medias, si es que queremos probar la hipótesis alternativa de un aumento por la labranza profunda ya que no sólo da la magnitud del efecto sino también de su signo. Entonces una prueba t de Student ( $t = 0.18 / 0.007 = 2.57$ ) nos da un resultado mayor que el valor crítico unilateral de 2.35 ( $\alpha = 5\%$ ) y se tiene una evidencia respetable para apoyar su hipótesis que la labranza profunda aumenta el rendimiento. Un intervalo de confianza a 90% para el aumento debido a la labranza profunda es de 15kg a 345kg por hectárea.

Por último convendría que los investigadores encargados el proyecto esperaban una intersección entre la profundidad del surco y el efecto del nitrógeno, ya que el surcado profundo debería permitir un sistema más vigoroso de raíces las cuales podrían aprovechar mejor el nitrógeno. Ahora la interacción de 40kg/Ha apoya su hipótesis, sin embargo dista mucho de ser estadísticamente significativa. Por lo tanto hay que recordar que el simple hecho de aceptar  $H_1$  ( de que no hay interacción) no demuestra que no existe una interacción pequeña, que no se pudo detectar con sete tamaño de experimento.

#### *Diseño factorial en dos factores y un número arbitrario de niveles.*

1) En un experimento para estudiar el rendimiento para dos variedades mejoradas "enanas" en comparación con una variedad tradicional "alta" bajo diferentes niveles de fertilización, se utilizó un diseño experimental en bloques aleatorios con tres repeticiones para cada tratamiento. El diseño de los tratamientos fue un factorial  $3 \times 3$  con el factor A= cantidad de nitrógeno (proveniente de urea) con tres niveles ( $a_1 = 0$  kg/Ha de nitrógeno,  $a_2 = 80$  kg/Ha de nitrógeno y  $a_3 = 160$  kg/Ha de nitrógeno) y el factor B= variedad con los niveles, las tres variedades arriba mencionadas. Los resultados en arroz palay en T/H fueron los siguientes:

REANDIMIENTOS EN TONELADAS POR HECTAREA.

		Variedad =B		
		b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>
		ENANA 1	ENANA 2	ENANA 3
a <sub>1</sub> = 0 kg/Ha	Bloque I	4.71	4.91	3.10
	Bloque II	4.00	2.87	1.43
	Bloque III	3.47	4.04	2.65
a <sub>2</sub> = 80 kg/Ha	Bloque I	5.90	5.67	3.67
	Bloque II	5.02	4.42	3.60
	Bloque III	4.42	4.90	3.44
a <sub>3</sub> = 160 kg/Ha	Bloque I	6.49	6.62	4.99
	Bloque II	5.71	5.40	4.40
	Bloque III	6.02	6.54	5.18

A= nivel de nitrógeno

Efectuar el análisis de varianza, subdividir la S.C de los efectos principales y de la interacción, determinar cuales efectos son significativos, graficar resultados y formular conclusiones.

Solución:

Se forma una tabla con los totales para cada tratamiento, los totales marginales y los totales de los bloques.

	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	Σ
a <sub>1</sub>	12.18	11.82	7.18	31.18
a <sub>2</sub>	15.34	14.99	10.71	41.04
a <sub>3</sub>	18.22	18.56	14.57	51.35
Σ	45.74	45.37	32.46	123.57 = X <sub>..</sub> = G

Bloque	I	II	III
	46.06	36.85	40.66

$$C = (123.57)^2 / 27 = 565.54$$

$$SCTR = (12.18^2 + 11.82^2 + \dots + 14.57^2) / 3 - C = 35.48$$

$$SCTR = (46.06^2 + 36.85^2 + 40.66^2) / 9 - C = 4.76$$

$$SCTR = (4.71)^2 + (4.00)^2 + \dots + (5.18)^2 - C = 43.35$$

$$SCER = SCTL - SCTR - SCBL = 3.11$$

$$SC_A = (31.18^2 + 41.04^2 + 51.35^2) / 9 - C = 22.60$$

$$SC_B = (45.74^2 + 45.37^2 + 32.46^2) / 9 - C = 12.71$$

$$SC_{A*B} = SCTR - SC_A - SC_B = 0.17$$

Se reúnen las sumas de cuadrados y se hacen los cálculos adicionales en una tabla de ANOVA.

TABLA DE ANOVA

F.V	g.l	S.C	C.M	F
Tratamientos	8	35.48	4.44	
A	2	22.60	11.30	59.47***
L <sub>1</sub>	1	22.60	22.60	118.95***
L <sub>2</sub>	1	0.004	0.004	< 1 n.s
B	2	12.71	6.35	32.42***
L <sub>3</sub>	1	0.01	0.01	< 1 n.s
L <sub>4</sub>	1	12.70	12.70	66.85***
A*B	4	0.17	0.04	< 1 n.s
Bloques	2	4.76	2.38	
Error	16	3.11	0.19	
Total	26	43.35		

1ª SUBDIVISION DE LOS GRADOS DE LIBERTAD

		Tratamiento									
		A <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>1</sub> b <sub>3</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>3</sub>	a <sub>3</sub> b <sub>1</sub>	a <sub>3</sub> b <sub>2</sub>	a <sub>3</sub> b <sub>3</sub>	
Contraste		12.18	11.82	7.18	15.34	14.99	10.71	18.22	18.56	14.57	
Subdivisión del efecto de A	L <sub>1</sub> = efecto lineal del N	-1	-1	-1	0	0	0	1	1	1	20.17
	L <sub>2</sub> = efecto lineal del N	+1	+1	+1	-0	-2	-2	1	1	1	0.45
	L <sub>3</sub> =rend promedio de variedad b <sub>1</sub> vs b <sub>2</sub>	1	-1	0	1	-2	0	1	-1	0	0.37
Subdivisión del efecto de B	L <sub>4</sub> =rend promedio de las "enanitas" vs variedad b <sub>2</sub> "alta"	1	1	-2	1	1	-2	1	1	-2	26.19

$$SC_{L_1} = [L_1] / r \cdot \sum 1_i^2 = (20.17)^2 / (3 \cdot 6) = 22.50$$

$$SC_{L_2} = (0.45)^2 / (3 \cdot 18) = 0.004$$

$$SC_{L_3} = (0.37)^2 / (3 \cdot 6) = 0.01$$

$$SC_{L_4} = (20.19)^2 / (3 \cdot 18) = 12.70$$

Nota:

No se subdividió la  $SC_{A \cdot B}$  ya que su valor (0.17) es tan bajo en comparación a la SCER que aun en el caso en que en la subdivisión todo el valor 0.17 resultará asignado a un solo contraste (y por lo tanto a un solo grado de libertad) aun así no sería estadísticamente significativo.

2. TABLA DE MEDIAS Y ERRORES ESTANDAR PARA EL RENDIMIENTO (T/Ha) DE TRES VARIEDADES DE ARROZ

		Variedad		
		Enana 1	Enana 2	Alta
0		4.06	3.94	2.39
		(±0.25)	(±0.25)	(±0.25)

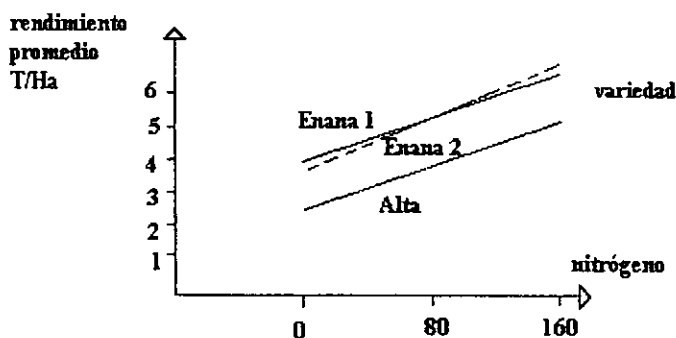


Nitrógeno (Kg/Ha)	80	5.11 (±0.25)	5.00 (±0.25)	3.57 (±0.25)
	160	8.07 (±0.25)	5.19 (±0.25)	4.86 (±0.25)

Observación:

Recuérdese que para el cálculo de los errores estándar de las medias de los tratamientos se tiene que:

$$s_x = \sqrt{\text{CMER} / r} = \sqrt{0.19/3} = 0.25$$

GRAFICA:REPORTE:

Debido a la no significancia de la interacción (reflejado en el paralelismo de las líneas en la gráfica), se puede hablar de la diferencia entre las variedades independientemente del nivel de nitrógeno. Entonces del contraste L3 se ve que para las variedades enanas no hay diferencias significativas en rendimiento, pero ambas superan significativamente ( $P < 0.005$ ) a la variedad alta. El efecto del nitrógeno es altamente significativo ( $P < 0.005$ ) y en las a la significancia de L1 y la no significancia de L2 se puede asegurar que el efecto es lineal en este rango experimental para las tres variedades.

En resumen las variedades enanas superan a las altas en aproximadamente 1.46 T/Ha siendo lineal la res puesta de las tres variedades al nitrógeno aumentando a razón de 14 kg de rendimiento por kg adicional de nitrógeno.

#### IV.- REGRESIÓN LINEAL

*Libros 2,8,12,13, 15, 17, 23, 25, 26*

Frecuentemente en investigación estamos interesados en estudiar la relación entre dos variables como cantidad de fertilizante y producción, dureza de los plásticos tratados con calor durante diferentes periodos de tiempo. La naturaleza y grado de relación entre las variables como estas, pueden ser analizadas por dos técnicas: *Regresión y Correlación* que aunque están relacionadas tienen propósitos e interpretaciones diferentes.

Hay una distinción entre estos dos procedimientos y no deben sustituirse uno por el otro en una situación experimental dada.

El **Análisis de Regresión** es útil para determinar la forma probable de la relación entre las variables cuando hay un fenómeno de causa y efecto; y su objetivo principal es el de predecir o estimar el valor de una variable dependiente (Y) correspondiente al valor dado de la otra variable independiente (X). En otras palabras el investigador decide cuales valores tomará la variable independiente, mientras que los valores de la variable dependiente están determinados por la relación que existe, si la hay, entre la v.dependiente y la independiente. Por lo tanto, debe emplearse el análisis de regresión en situaciones experimentales en las cuales el investigador controla la variable independiente.

Cuando representamos gráficamente unos datos (de algún experimento), designamos una línea que utilizamos para representarlos y recibe el nombre de *línea de regresión*.

Los criterios para trazar una línea se suelen expresar de la siguiente manera:

- 1) Tantos puntos como sea posible deberían caer en la línea que se dibuje.
- 2) Tantos puntos deberían caer por encima como por debajo de la línea.
- 3) Las desviaciones de los puntos no alineados deberían sumar cero, considerando las desviaciones por encima de la línea como positivas y como negativas las que caigan debajo.

Una crítica de este método es que resulta difícil apreciar el grado en que la línea trazada representa los datos, tenemos como una técnica mejor, un método objetivo de ajustar líneas a los datos conocidos como el método de "**Mínimos Cuadrados**". Con este método definimos la línea de mejor ajuste como una línea que hace que la suma de las desviaciones de las puntas respecto a la línea sea mínima.

El **Análisis de Correlación** por otra parte, consiste en la medición del grado o intensidad de *asociación* entre dos variables sin importar cual es causa y cual es efecto.

Cuando se puede demostrar que la variación de una variable esta de algún modo asociada con la variación de otra, entonces se puede decir que las dos variables están *correlacionadas*, una correlación puede ser positiva (cuando al aumentar una variable la otra también aumenta) o negativa (cuando al aumentar una variable la otra disminuye). Por otro lado si la variación de una variable no corresponde en absoluto a la variación de la otra, entonces no existe ninguna asociación y, por lo tanto, ninguna correlación entre las dos variables.

Es importante enfatizar que en el análisis de correlación, ni X, ni Y representan una variable independiente; en oras palabras tanto X como Y son variables aleatorias.

#### 4.1).- CURVA DE AJUSTE (diagramas de dispersión)

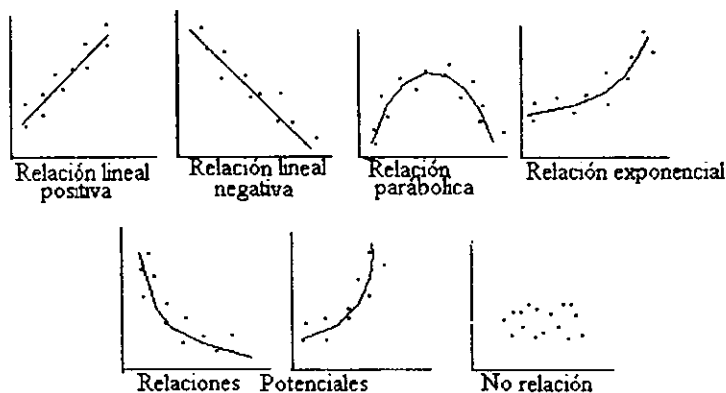
Convencionalmente al hacer la representación gráfica de los datos cuantitativos los valores de la variable independiente (si solo hay una) se sitúa sobre el eje horizontal (de un sistema cartesiano de dos dimensiones) y los valores de respuesta sobre el eje vertical. Todo error es atribuible a al variable de respuesta.

Un gráfico de la línea de regresión es más que una simple representación de los datos. Trazar dicha línea representa nuestra apreciación de que:

- 1) Realmente existe una relación funcional entre las variables
- 2) La línea expresa esta relación en términos cuantitativos

Debido a la variación del muestreo, los puntos estarán dispersos, si los puntos muestran una tendencia lineal positiva o negativa se le puede ajustar una línea recta que servirá entre otras cosas para predecir valores de Y correspondientes a valores de X.

A continuación se muestran algunos diagramas de dispersión que se pueden obtener a partir de un conjunto de datos y sus ecuaciones. Todas las letras distintas a X y Y representan constantes, las variables X e Y se conocen a menudo como la *variable independiente* y *dependiente*:



- |   |                             |
|---|-----------------------------|
| (1) $Y = a_0 + a_1 X$                               | Línea recta                 |
| (2) $Y = a_0 + a_1 X + a_2 X^2$                     | Parábola o curva cuadrática |
| (3) $Y = a_0 + a_1 X + a_2 X^2 + a_3 X^3$           | Curva cúbica                |
| (4) $Y = a_0 + a_1 X + a_2 X^2 + a_3 X^3 + a_4 X^4$ | Curva cuártica              |
| (5) $Y = a_0 + a_1 X + a_2 X^2 + \dots + a_n X^n$   | Curva de grado n            |

Las ecuaciones anteriores se llaman polinomiales de primero, segundo, tercero, cuarto y n grados respectivamente. Las funciones definidas por las cuatro primeras de estas ecuaciones se llaman a veces funciones lineal, cuadrática, cúbica y cuártica respectivamente.

Otras muchas ecuaciones (entre muchas) que en la práctica aparecen, son las siguientes.

$$(6) Y = \frac{1}{a_0 + a_1 X} \text{ o } \frac{1}{Y} = a_0 + a_1 X \quad \text{Hipérbola}$$

$$(7) Y = ab^X \text{ o } \log Y = \log a + (\log b)X = a_0 + a_1 X \quad \text{Curva exponencial}$$

$$(8) Y = aX^b \text{ o } \log Y = \log a + b \log X$$

Curva geométrica

$$(9) Y = ab^x + g$$

Curva exponencial modificada

Para decidir qué curva deberá utilizarse es de gran ayuda obtener los diagramas de dispersión de las variables transformadas, por ejemplo, si un diagrama de dispersión de los ejes  $\log Y$ ,  $X$  muestra una relación lineal la ecuación tiene la forma (7), mientras que si en los ejes  $\log Y$ ,  $\log X$  muestra una relación lineal la ecuación tiene la forma (8).

Una vez que se han dibujado los puntos, se puede saber que los puntos siguen un patrón, mismo que indicará el modelo matemático a utilizarse en el análisis. Otra forma de llegar al modelo puede ser por consideraciones teóricas o por que se sabe por experiencia o por referencia como se comportan las variables

La relación funcional resultante puede ser una línea recta ó una curva y también lineal simple o lineal múltiple. Si la representación es una línea recta, tendremos un modelo de la regresión lineal y si la relación difiere de una recta tendremos un modelo de la regresión curvilínea.

#### 4.2).- REGRESIÓN LINEAL SIMPLE

Después que se ha determinado el modelo matemático a utilizar y se conoce que es lineal se procede a ajustar una recta llamada recta de regresión o recta de ajuste.

##### Suposiciones para la regresión lineal:

- 1) Los valores de la variable independiente  $X$  son fijos y se le llama variable no aleatoria.
- 2) Para cada valor de  $X$  hay una subpoblación de valores de  $Y$  y cada subpoblación de valores de  $Y$  debe estar normalmente distribuida.
- 3) Las varianzas de las subpoblaciones de  $Y$  deben ser iguales.
- 4) Las medias de las subpoblaciones de  $Y$  todas están sobre una recta (suposición de linealidad).
- 5) Los valores de  $Y$  son estadísticamente independientes, es decir, los valores de  $Y$  correspondientes a un valor de  $X$  no dependen de los valores de  $Y$  para otro valor de  $X$

#### 4.2.1 ) ANÁLISIS DE CORRELACIÓN.

Como mencionamos al inicio de éste capítulo usamos el análisis de regresión cuando nos interesa establecer el tipo de relación que hay entre dos variables; pero, cuando solo nos interesa establecer *el grado de relación* entre dos variables aleatorias usamos *el análisis de correlación*.

La medida del grado de relación entre dos variables se llama *coeficiente de correlación*, representado universalmente por  $\rho$ . En el modelo de correlación, se asume que  $X$  y  $Y$  varían conjuntamente en una distribución conjunta. Si esta distribución está distribuida normalmente, entonces es llamada distribución normal bivariable, para el cual se estima  $\rho$ , se describen a continuación:

1.  $Y$  y  $X$  son variables aleatorias. Como tales, no deben ser designadas como dependiente e independiente; cualquier designación dará el mismo resultado.

2. La población bivariable es normal. Una población normal bivariable es, entre otras cosas, aquella en la que Y y X están normalmente distribuidas.
3. La relación entre Y y X es, en cierto sentido, lineal. Este supuesto implica que todas las medias de Y asociadas con valores de X,  $\mu_{Y|X}$ , caen sobre una línea recta, que es la línea de regresión de Y sobre X:  $\mu_{Y|X} = A + BX$ . Igualmente todas las medias de X asociadas con valores de Y,  $\mu_{X|Y}$  caen sobre una línea recta, que es la línea de regresión de X sobre Y:  $\mu_{X|Y} = A' + B'Y$ .

Según los supuestos anteriormente, el coeficiente de correlación Pearson de la población se define de la siguiente manera:

$$\rho = \frac{C_{XY}}{\sigma_X \sigma_Y} = \frac{\sum X_i Y_i - n \mu_X \mu_Y}{\sqrt{(\sum X_i^2 - n \mu_X^2)(\sum Y_i^2 - n \mu_Y^2)}}$$

Por lo tanto puede observarse lo siguiente:

\* Primero, es una ecuación que contiene los cinco parámetros de la población normal bivariable:  $\mu_X$ ,  $\sigma_X$ ,  $\mu_Y$ ,  $\sigma_Y$ ,  $\rho$ , este último, como se mencionó anteriormente es el coeficiente de correlación para la población normal bivariable.

\* Segundo,  $\rho$  es simétrico con respecto a Y y a X; es decir, el intercambio entre X y Y no cambia al coeficiente de correlación.

\* Tercero, cuando la varianza es cero,  $\rho$  es cero, indicando que no hay relación entre las variables. Cuando hay covarianza perfecta entre X y Y, y ambas varían en la misma dirección,  $\rho = 1$ . Análogamente cuando hay varianza perfecta, pero si Y y X varían en sentidos opuestos,  $\rho = -1$ . Por otra parte, cuando existe cierto grado de covarianza entre X y Y, tenemos

$$-1 < \rho < 0 \quad \text{y} \quad 0 < \rho < 1$$

En general:  $-1 \leq \rho \leq 1$

Cuando se extrae una muestra de n valores, donde cada valor X es una observación al azar de la población X y cada valor Y es una observación al azar de la población Y; pero las dos no son necesariamente independientes. Además, cuando se cumple el supuesto de una población normal bivariable, el estimador de  $\rho$  es r, definido por:

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{(\sum X_i^2 - n \bar{X}^2)(\sum Y_i^2 - n \bar{Y}^2)}} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{(n-1)S_X S_Y}$$

que puede variar, como  $\rho$ , entre  $-1$  y  $1$ . ( $-1 \leq r \leq 1$ ), pero su interpretación se hará por medio de su cuadrado ( $r^2$ ) el coeficiente de determinación.

Este valor puede emplearse como interpretación de la intensidad de la asociación entre las dos variables que parecen estar correlacionadas. Específicamente, *el coeficiente de determinación indica el porcentaje de la variación de X que está asociada con (o "es explicada por") la variación de Y, o viceversa*. Por ejemplo, si la correlación muestral entre dos variables tales como la clorofila y la biomasa es  $r = 0.50$ , elevando al cuadrado este coeficiente da un coeficiente de determinación  $r^2 = 0.25$ . Esto sugiere que el 25% de la variación de una de las dos variables, está asociada con o "es explicada por" la variación de la otra. No podemos decir cual "explica" cual, porque ambas son consideradas variables aleatorias.

Puede señalarse que  $r$  es un estimador sesgado de  $\rho$ , excepto cuando  $\rho = 0$ . Cuando  $\rho = 0$ , la distribución muestral de  $r$  sigue una distribución  $t$  de student con  $n-2$  grados de libertad

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

#### 4.3).- ANÁLISIS DE COVARIANZA

En el modelo de regresión lineal cada conjunto de datos apareados ( $X_i$  y  $Y_i$ ) representa un valor fijo de  $X$  (nivel) y un valor de la variable de respuesta  $Y$ . El valor de  $Y_i$  es dependiente si existe una línea de regresión del nivel de  $X$  y del error aleatorio presente.

Si los valores  $Y$  están relacionados linealmente con los de  $X$ , formulamos una hipótesis de que la verdadera línea de regresión es:  $Y = \beta_0 + \beta_1 X$

Donde:  $\beta_0$  es el valor de  $Y$  en el origen, esto es, en el punto en el que  $X = 0$ ,

$\beta_1$  define la contribución al valor  $Y$  por el efecto de la variable independiente  $X$ . Para cualquier punto  $X_i$  y  $Y_i$  existirá alguna contribución debida al error de forma que

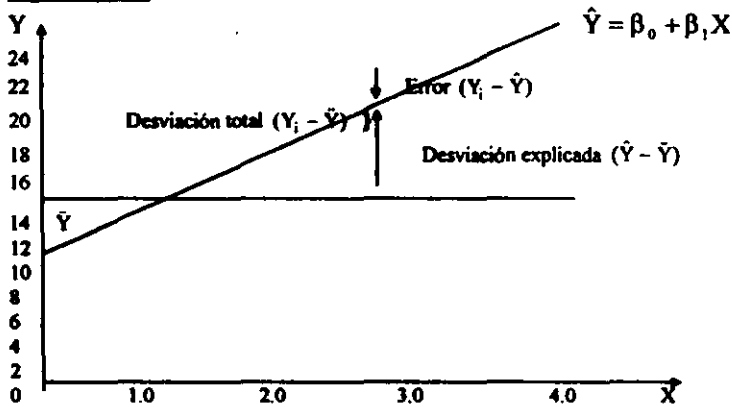
$$\hat{Y} = \beta_0 + \beta_1 X_i + \epsilon_i$$

La ecuación precedente es el modelo matemático para el análisis de la regresión lineal, donde  $\epsilon_i$  es el error

##### 4.3.1 COEFICIENTE DE DETERMINACIÓN Y ANÁLISIS DE COVARIANZA

Habiendo obtenido la ecuación de regresión de la muestra  $\hat{Y} = \beta_0 + \beta_1 X$ , comparando las desviaciones de los puntos con relación a la recta de regresión con las desviaciones al rededor de la recta  $\bar{Y}$ , ( $\bar{Y}$  = valor medio de los valores observados de  $Y$ ). Usando el diagrama de dispersión podremos visualizar lo expuesto en forma gráfica

Figura 4.2.1.1



Como vemos en la figura 4.2.1.1 si consideramos un punto cualquiera  $(X_i, Y_i)$ , la distancia vertical del valor observado  $Y_i$  a  $\bar{Y}$  es llamada *desviación total* y la designamos por  $(Y_i - \bar{Y})$ . Si medimos la distancia vertical desde la recta de regresión  $\hat{Y}_i$  a la recta  $\bar{Y}$ , obtenemos  $(\hat{Y}_i - \bar{Y})$  lo que llamamos *desviación explicada* y finalmente, a la distancia vertical del punto  $(X_i, Y_i)$  a la recta de regresión  $\hat{Y}_i$ ,  $(Y_i - \hat{Y}_i)$  que es el error, la llamamos *desviación inexplicada*.

Entonces tenemos:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

Desviación total	Desviación explicada	Desviación inexplicada
---------------------	-------------------------	---------------------------

Si se miden estas desviaciones para todos los puntos, se elevan al cuadrado y se suman se obtiene:

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

Suma de cuadrados total	Suma de cuadrados explicada	Suma de cuadrados inexplicada
-------------------------------	-----------------------------------	-------------------------------------

La suma de cuadrados explicada entre la suma de cuadrados total es el *coeficiente de determinación*  $r^2$  es obvio que  $r^2$  varía entre 0 y 1, ( $0 \leq r^2 \leq 1$ ). Si  $r^2 = 1$ ,  $(Y_i - \hat{Y}_i) = 0$  y  $e_i = 0$  por lo tanto todos los puntos están sobre la recta de regresión. Si  $r^2 = 0$ , la línea de regresión  $\hat{Y}_i$  y la recta  $\bar{Y}$  coinciden y por lo tanto no hay variación explicada por la regresión.

Usando las tres sumas de cuadrados y los grados de libertad asociadas a ellas, podemos construir una tabla de análisis de varianza para la regresión lineal. En general, los grados de libertad asociados con la suma de cuadrados debida a la regresión es igual al número de coeficientes en la ecuación de regresión menos 1. En el caso de regresión lineal simple hay dos constantes a y b, por lo tanto, *el número de grados de libertad debidos a regresión lineal simple es  $2 - 1 = 1$* .

Tabla de ANDEVA para Covarianza

Fuente	Grados de libertad	Suma de Cuadrados	Cuadrados medios	F calc.
Regresión	2 - 1 = 1	SCexplicada	$S^2_{\text{expl}} = \frac{SC_{\text{expl}}}{1}$	$F = \frac{S^2_{\text{expl}}}{S^2_{\text{in expl}}}$
Error	n - 2	SCinexplicada	$S^2_{\text{in expl}} = \frac{SC_{\text{in expl}}}{n - 2}$	
Total	n - 1	SCtotal		

La prueba F es equivalente a la prueba t bilateral sobre la pendiente de regresión B. Esto significa que la hipótesis nula a contrastar es:

Ho: Y no depende de X ( $B = 0$ )

Ha: Y depende de X ( $B \neq 0$ )

La diferencia entre las dos pruebas es que F puede ser generalizada a análisis de regresión múltiple (más de una variable independiente), mientras que la prueba t solo puede usarse para un coeficiente de regresión (B) cada vez.

#### 4.4.- REGRESIÓN LINEAL MÚLTIPLE (triple)

En este modelo tenemos una variable dependiente Y, y dos independientes  $X_1$  y  $X_2$  así en vez de obtener una función lineal en dos variables (línea recta), tendremos una función lineal en tres variables (ecuación en el plano tridimensional)

El modelo matemático para tal caso es  $\hat{Y}_i = a + b_1 X_1 + c_2 X_2 + \epsilon$

Donde  $\hat{Y}_i$  = valor de la variable de respuesta, para los valores dados de  $X_1$  y  $X_2$

a = valor asociado con el nivel general de Y

$X_1$  y  $X_2$  = variables independientes

$b_1$  = el coeficiente de regresión parcial (pendiente) de Y sobre  $X_1$  con  $X_2$  constante

$b_2$  = el coeficiente de regresión parcial (pendiente) de Y sobre  $X_2$  con  $X_1$  constante

$\epsilon$  = es la parte de la variación total no explicada o debida a errores aleatorios

En general el modelo matemático de la regresión lineal múltiple debe cumplir con las siguientes suposiciones:

- Las variables independientes  $X_1$  y  $X_2$  son fijas o no aleatorias
- Pueden existir relaciones significativas de dependencia lineal entre dos cualesquiera de las variables independientes, pero su correlación no debe ser perfecta
- Los errores  $\epsilon_i$  tienen distribuciones normales con  $\mu_{\epsilon_i} = 0$
- La varianza de la regresión es constante e igual a las varianzas de los  $\epsilon_i$

$$S^2_{Y/X_1X_2} = S^2_{\epsilon_i} = S^2$$





Habiendo obtenido la ecuación de regresión de la muestra, debemos establecer si los resultados de la muestra son estadísticamente significativos para poder usar la ecuación de regresión como instrumento de predicción.

Por lo tanto, usamos el análisis de varianza descomponiendo la variación total en dos partes independientes

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

Suma de cuadrados total
Suma de cuadrados explicada
Suma de cuadrados de error

$$SC_{total} = SC_{expl.} + SC_{error}$$

$$\text{Donde: } SC_{total} = \sum (Y_i - \bar{Y})^2 = m_{11}$$

$$SC_{explicada} = \sum (\hat{Y}_i - \bar{Y})^2 = bm_{12} + cm_{13}$$

$$SC_{error} = SC_{total} - SC_{expl.}$$

Los grados de libertad debidos a la regresión son el número de variables menos 1,  $3 - 2 = 1$  Los grados de libertad del error son el tamaño de muestras menos el número de variables  $n - 3$

Los grados de libertad totales  $(n - 1)$ . Así tenemos la tabla de ANDEVA para la regresión triple como sigue:

Tabla de ANDEVA para la Regresión Triple

Fuente	Grados de libertad	Suma de Cuadrados	Varianza	F calc.
Regresión	$3 - 1 = 2$	SCexplicada	$S^2_{expl} = \frac{SC_{expl}}{2}$	$F = \frac{S^2_{expl}}{S^2_{error}}$
Error	$n - 3$	SCerror	$S^2_{error} = \frac{SC_{error}}{n - 3}$	
Total	$n - 1$	SCtotal		

La hipótesis nula en este caso es:  $H_0: B=C=0$

$H_1: B \neq 0 \text{ ó } C \neq 0$

El coeficiente de determinación triple se define al igual que el simple como la razón de la variación explicada a la total, esto es:

$$r^2 = \frac{SC_{expl}}{SC_{total}}$$

El coeficiente de correlación lineal triple es la raíz cuadrada positiva del coeficiente de determinación

$$r\sqrt{r^2}$$

#### 4.5).- MÉTODO DE "MÍNIMOS CUADRADOS"

El procedimiento estadístico para encontrar la recta de "mejor ajuste" para un conjunto de puntos podría verse en muchos aspectos como una formalización del procedimiento que se emplea al ajustar una recta visualmente, tratando de minimizar las desviaciones de los puntos para escoger la recta de mejor ajuste, es decir que sea razonable, objetivo y que bajo ciertas condiciones produzca la mejor predicción de Y para un valor dado de X.

Tómese la recta de “mejor ajuste” aquella que minimiza la suma de los cuadrados de las desviaciones de los valores observados de Y, respecto a los valores predichos.

El criterio utilizado es el siguiente:

Primero se elevan al cuadrado las desviaciones para hacer que todas sean positivas y a continuación se suman para obtener el criterio de mínimos cuadrados que es minimizar, este criterio selecciona una línea recta de mejor ajuste.

Se acostumbra a utilizar la definición anterior cuando X es la variable independiente y la variable dependiente. Si X es la variable dependiente, la definición se modifica considerando las desviaciones horizontales en lugar de las verticales, lo que equivale a un intercambio de los ejes X e Y por lo que conduciría a diferentes rectas de mínimos cuadrados. Al menos que se especifique de otro modo, se considera Y como la variable dependiente y X como la independiente.

#### 4.5.1 ETAPAS PARA LA OBTENCIÓN DE MÍNIMOS CUADRADOS

1° El primer paso en un análisis de regresión es el de representar los puntos en un diagrama de dispersión para saber que relación lineal siguen los puntos.

2° Expresado matemáticamente deseamos escoger valores para  $\hat{\beta}_0, \hat{\beta}_1$ , por lo tanto los datos que se nos den tendrán que ser simplificados en otra tabla de la siguiente manera.

Y <sub>i</sub>	X <sub>i</sub>	X <sub>i</sub> <sup>2</sup>	X <sub>i</sub> Y <sub>i</sub>	Y <sub>i</sub> <sup>2</sup>
.	.	.	.	.
n	.	.	.	.
Σ	.	.	.	.

A partir de esta tabla se realizan los siguientes cálculos:

$$1) \text{ Suma de cuadrados de X: } SC_x = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

$$2) \text{ Suma de cuadrados de XY: } SC_{xy} = \frac{x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

$$3) \text{ Media de } \bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$4) \text{ Media de } \bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$5) \hat{\beta}_1 = \frac{SC_{xy}}{SC_x}$$

$$6) \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \text{Suma de cuadrados de error: } SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ donde}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X$$

7) De acuerdo al principio de mínimos cuadrados, se encuentra la recta de mejor ajuste con la siguiente ecuación y los datos obtenidos anteriormente:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X$$

Donde:  $\hat{\beta}_0$  = intersección de la recta en el eje y (ordenada al origen) cuando  $x = 0$

$\hat{\beta}_1$  = pendiente (donde la pendiente de una recta es el cambio en la altura Y, al moverse una unidad hacia la derecha, en la dirección X)

## RESUMEN

Al tratar las regresiones curvilíneas generalmente se hace una transformación inicial de datos, por medio de logaritmos del tal forma que la relación entre las variables transformadas sean aproximadamente lineal, pero las transformaciones logarítmicas no son las únicas, pudiera darse el caso en que la transformación recíproca fuese más adecuada.

Las transformaciones en regresión lineal se hacen principalmente por dos razones. Primera por que este procedimiento consiste en facilitar los cálculos, segunda, las transformaciones en regresión lineal son convenientes porque la teoría estadística está especialmente bien desarrollada bajo el supuesto de linealidad.

El análisis de regresión lineal se realiza descomponiendo la variación por la variable o las variables independientes y la variable del error. El método de mínimos cuadrados tiene la función de minimizar la suma de cuadrados del error y maximizar la suma de cuadrados debida a la regresión.

El grado de ajuste en cada caso es medido por  $r^2$ , el coeficiente de determinación que es la razón de la suma de cuadrados debida a la regresión entre la suma de cuadrados total.

Finalmente, existen paquetes de computadora para realizar análisis de regresión lineal simple, múltiple y no lineal, de manera que los cálculos aquí señalados se pueden hacer en forma rápida y por consiguiente solo tenemos que interpretar los resultados.

4.6). - PROBLEMAS RESUELTOS*(Análisis de covarianza)*

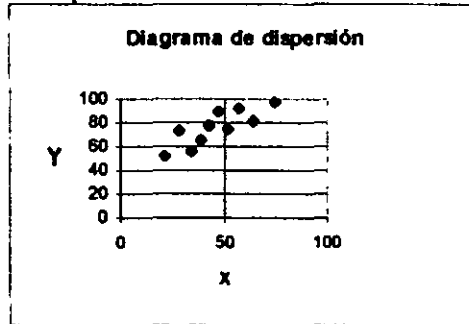
1) Obtenga la recta de predicción de mínimos cuadrados para los datos de la siguiente tabla:

Tabla 1

Lata de conserva	Peso inicial	Peso final
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
n = 10	52	75

Solución:

Realizar el diagrama de dispersión

El cálculo de  $\hat{\beta}_0, \hat{\beta}_1$  para los datos de la tabla 1, se simplifican usando la tabla 2, colocando las sumas apropiadas de la tabla 2 en las ecuaciones de cuadrados mínimos:Tabla 2

$Y_i$	$X_i$	$X_i^2$	$Y_i^2$	$X_i Y_i$
65	39	1521	4225	2535
78	43	1849	6084	3354
52	21	441	2704	1092
82	64	4096	6724	5248
92	57	3249	8464	5244
89	47	2209	7921	4183
73	28	784	5329	2044
98	75	5625	9604	7350
56	34	1156	3136	1904
75	52	2704	5625	3900
$\Sigma = 760$	460	23,634	59,816	36,854

$$1) \text{ Suma de cuadrados de X: } SC_x = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = 23,634 - \frac{(460)^2}{10} = 2474$$

$$2) \text{ Suma de cuadrados de XY: } SC_{xy} = \frac{\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

$$= 36,854 - \frac{(460)(760)}{10} = 1894$$

$$3) \text{ Media de } \bar{Y} = \frac{\sum_{i=1}^n y_i}{n} = 760 / 10 = 76$$

$$4) \text{ Media de } \bar{X} = \frac{\sum_{i=1}^n x_i}{n} = 460 / 10 = 46$$

$$5) \hat{\beta}_1 = \frac{SC_{xy}}{SC_x} = 1894 / 2474 = 0.765 \approx 0.77$$

$$6) \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 76 - [(0.77)(46)]$$

$$= 76 - 35.42 = 40.58$$

7) De acuerdo al principio de mínimos cuadrados, la recta de mayor ajuste que relaciona el peso final con los pesos iniciales de las latas es.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$\hat{y} = 40.58 + 0.77X \Rightarrow \text{ecuación de predicción}$$

La intersección de la recta en el eje  $y = 40.58$ , cuando  $x = 0$ , la pendiente es 0.77 que da el cambio estimado en  $y$  por cada unidad de cambio en  $x$ .

Ahora ya con la ecuación podemos predecir  $Y$  para un valor de  $X$ , sustituyendo  $X$  en la ecuación de predicción o con la ayuda del gráfico.

Por ejemplo si una lata de conservas pesa inicialmente  $x = 50$ , su peso final en cálculo sería:  $\hat{y} = 40.58 + 0.77(50) = 79.08$

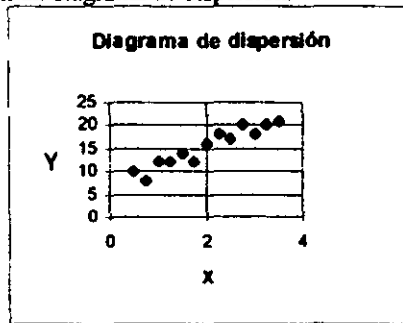
2) Se llevó a cabo un experimento para estudiar el efecto de ciertos conservadores influyendo estos en la perecibilidad del alimento. La variable independiente es la cantidad de conservador en gramos y la variable dependiente es la diferencia entre el menor tiempo de perecibilidad después de la incorporación del conservador y el tiempo antes de su incorporación (control). Los datos se muestran a continuación

Cantidad de conservados (gr) X	Tiempo de vida (horas) Y
0.50	10
0.75	8
1.00	12
1.25	12
1.50	14

1.75	12
2.00	16
2.25	18
2.50	17
2.75	20
3.00	18
3.25	20
3.50	21

**Solución:**

Como se menciona anteriormente el primer paso en un análisis de regresión es el de representar los puntos en un diagrama de dispersión.



Como se puede observar los puntos siguen una relación lineal positiva, por lo tanto se procede a determinar la recta de ajuste.

$X_i$	$Y_i$	$X_i^2$	$Y_i^2$	$X_i Y_i$
0.50	10	0.2500	100	5.0
0.75	8	0.5625	64	6.0
1.00	12	1.0000	144	12.0
1.25	12	1.5625	144	15.0
1.50	14	2.2500	196	21.0
1.75	12	3.0625	144	21.0
2.00	16	4.0000	256	32.0
2.25	18	5.0625	324	40.5
2.50	17	6.2500	289	42.5
2.75	20	7.5625	400	55.0
3.00	18	9.0000	324	54.0
3.25	20	10.5625	400	65.0
3.50	21	12.2500	441	73.5
$\Sigma = 26$	198	63.375	3226	442.5

$$1) \text{ Suma de cuadrados de X: } SC_x = 63.375 - \frac{(26)^2}{13} = 11.37$$

$$2) \text{ Suma de cuadrados de XY: } SC_{xy} = 442.5 - \frac{(26)(198)}{13} = 46.5$$

$$3) \text{ Media de } \bar{Y} = \frac{\sum_{i=1}^n y_i}{n} = 198 / 13 = 15.23$$

$$4) \text{ Media de } X = \frac{\sum_{i=1}^n x_i}{n} = 26 / 13 = 2$$

$$5) \hat{\beta}_1 = 46.5/11.37 = 4.089$$

$$6) \hat{\beta}_0 = 15.23 - [(4.089)(2)] = 7.052$$

7) La ecuación de la recta de ajuste es  $\hat{y} = 7.052 + 4.089X$

Al sustituir X por los valores 1.0 y 3.0 se obtienen los puntos (1.0 , 11.14) y (3.0 , 19.32) que se utilizan para representar la recta sobre el diagrama de dispersión.

3) Usando los datos del ejemplo 2 tenemos lo siguiente:

$$\hat{y} = 7.052 + 4.089X$$

$X_i$	$Y_i$	$\hat{Y}_i$
0.50	10	9.096
0.75	8	10.1187
1.00	12	11.1410
1.25	12	12.1632
1.50	14	13.1855
1.75	12	14.2077
2.00	16	15.2300
2.25	18	16.2522
2.50	17	17.2745
2.75	20	18.2967
3.00	18	19.3190
3.25	20	20.3412
3.50	21	21.3635
	198	

Los datos de la tercera columna se obtienen a partir de la ecuación de la recta obtenida anteriormente.  $\hat{Y} = \beta_0 + \beta_1 X_i + \epsilon_i \rightarrow \hat{y} = 7.052 + 4.089X$  sustituyendo el valor de  $X_i$  de la primera columna en la ecuación

1° Calcular la media de Y ;  $\bar{Y} = 198/3 = 15.2308$

2° Calcular la suma de cuadrados total

$$\begin{aligned} SC_{\text{total}} &= \sum (Y_i - \bar{Y})^2 \\ &= (10 - 15.2308)^2 + (8 - 15.2308)^2 + (12 - 15.2308)^2 + (12 - 15.2308)^2 + (14 - 15.2308)^2 \\ &+ (12 - 15.2308)^2 + (16 - 15.2308)^2 + (18 - 15.2308)^2 + (17 - 15.2308)^2 + (20 - 15.2308)^2 \\ &+ (18 - 15.2308)^2 + (20 - 15.2308)^2 + (21 - 15.2308)^2 = 210.3072 \end{aligned}$$

3° Calcular la varianza de Y

$$\begin{aligned} S^2_y &= \sum (Y_i - \bar{Y})^2 / n - 1 \\ S^2_y &= 210.3072 / 12 = 17.5256 \end{aligned}$$

4° Calcular la suma de cuadrados explicada

$$\begin{aligned} SC_{\text{expl.}} &= \sum (\hat{Y} - \bar{Y})^2 \\ &= (9.096 - 15.2308)^2 + (10.1187 - 15.2308)^2 + (11.1410 - 15.2308)^2 + \dots + (19.3190 - 15.2308)^2 \\ &+ (20.3412 - 15.2308)^2 + (21.3635 - 15.2308)^2 \\ &= 190.0753 \end{aligned}$$



5° Calcular la suma de cuadrados inexplicada

$$\begin{aligned} \text{SCinex.} &= \text{SCtotal} - \text{SCexpl.} \\ &= 210.3072 - 190.0753 = 20.2319 \end{aligned}$$

Construir la tabla de análisis de varianza para la regresión lineal simple

Fuente	g.l	Suma de Cuadrados	Cuadrados medios	F calc.
Regresión	1	SCexpl. = 190.0753	$S_{expl}^2 = 190.1753$	$F = \frac{S_{expl}^2}{S_{inexpl}^2} = 103.3432$
Error	11	SCerror = 20.2319	$S_{inexpl}^2 = 1.8393$	
Total	12	SCtotal = 210.3072		

Sacando la F teórica de tablas de distribución F tenemos  $F_{0.95, 1, 11} = 4.84$

Esto significa que la hipótesis nula a contrastar es:

$$H_0: Y \text{ no depende de } X (\beta_1 = 0)$$

$$H_a: Y \text{ depende de } X (\beta_1 \neq 0)$$

Por lo tanto se concluye que Y depende significativamente de X.

*(Regresión lineal múltiple)*

4) Para comprobar al efecto de la velocidad de rotación en la cantidad de jugo filtrado, en un filtro de tambor rotatorio, se probó con diferentes espesores de capa a varias velocidades. Los datos para tres espesores y cinco velocidades se presentan en la tabla 1.

Aquí las variables independientes son la velocidad de rotación y el espesor de la capa (son no aleatorias) y la variable dependiente es la cantidad de jugo filtrado. El problema es expresar la variable dependiente de manera aproximada como una función lineal de las dos variables independientes. Las estimaciones de  $a_1$ ,  $b_2$  y  $b_3$  se harán con la ayuda de la tabla 2.

Tabla 1

Cantidad de jugo filtrado para 3 espesores y 5 velocidades

Espesor de capa $X_2$	Jugo filtrado a velocidades seleccionadas $X_3$				
	30 rpm	40 rpm	50 rpm	60 rpm	70 rpm
2.29	21.55	20.07	19.11	17.83	16.72
3.98	18.25	20.00	16.32	15.77	13.61
5.25	18.33	19.28	15.62	14.22	12.74

Tabla 2:

$X_2$	$X_3$	Y	$YX_2$	$YX_3$	$X_2 X_3$
2.29	30	21.55	49.3495	646.50	68.70
2.29	40	20.07	45.9603	802.80	91.60
2.29	50	19.11	43.7619	955.50	114.50
2.29	60	17.83	40.8307	1069.80	137.40
2.29	70	16.72	38.2888	1170.40	160.30
3.98	30	18.25	72.6350	547.50	119.40
3.98	40	20.00	79.6000	800.00	159.20
3.98	50	16.32	64.9536	816.00	199.00
3.98	60	15.77	62.7646	946.20	238.80
3.98	70	13.61	54.1678	952.70	278.60

5.25	30	18.33	96.2325	549.90	157.50
5.25	40	19.28	101.2200	771.20	210.00
5.25	50	15.62	82.0050	781.00	262.50
5.25	60	14.22	74.6550	853.20	315.00
5.25	70	12.74	66.8850	891.80	367.50
$\Sigma = 57.60$	750	259.42	973.3097	12554.50	2880.00

1° Calcular las medias de cada una de las variables con su respectiva varianza

$$\bar{x}_2 = 3.84 \quad , \quad \bar{x}_3 = 50 \quad , \quad \bar{Y} = 17.2947$$

$$S^2_{x_2} = 1.5751 \quad , \quad S^2_{x_3} = 1.5751 \quad , \quad S^2_Y = 1.5751$$

2° Calcular la ecuación de regresión  $\hat{Y}_i = a + b_1 X_1 + c_2 X_2$  con las siguientes ecuaciones:

$$* b = \frac{m_{12}m_{13} - m_{13}m_{16}}{m_{14}m_{15} - m_{16}^2}$$

$$* c = \frac{m_{13}m_{14} - m_{12}m_{16}}{m_{14}m_{15} - m_{16}^2}$$

$$* a = \bar{Y} - b\bar{x}_2 - c\bar{x}_3$$

3° Calcular las m's

$$m_{12} = \sum Y_i X_{i2} - n \bar{Y} \bar{X}_2 = 973.3097 - [15(17.2947)(3.84)]$$

$$= 973.3097 - 996.1747 = -22.865$$

$$m_{13} = \sum Y_i X_{i3} - n \bar{Y} \bar{X}_3 = 12,554.5 - [15(17.2947)(50)]$$

$$= 12,554.5 - 12,971.025 = -416.525$$

$$m_{14} = (n-1)S^2_{x_2} = (15-1)(1.5751)$$

$$= 22.0514$$

$$m_{15} = (n-1)S^2_{x_3} = (15-1)(214.2854)$$

$$= 3000$$

$$m_{16} = \sum X_{i2} X_{i3} - n \bar{X}_2 \bar{X}_3 = 2880.00 - [15(3.84)(50)]$$

$$= 2880.00 - 2880.00 = 0.0$$

Obtener:

$$* b = \frac{(-22.865)(3000) - (-416.525)(0)}{(22.0514)(3000) - (0)^2} = \frac{-68151.475}{66,154.2} = -1.0369$$

$$* c = \frac{(-416.525)(22.0514) - (-22.865)(0)}{(22.0514)(3000) - (0)^2} = \frac{-9162.094}{66,154.2} = -0.1388$$

$$* a = 17.2947 - (-1.0369)(3.84) - (-0.1388)(50)$$

$$= 28.2164$$

Así la ecuación de predicción es :

$$\hat{Y} = 28.22 - 1.04X_2 - 0.14X_3$$

Para un espesor de capa de 4.0 cm a una velocidad de 55 rpm, la cantidad de jugo filtrado debe ser alrededor de:

$$\hat{Y} = (4.0, 55) = 28.22 - 1.04(4) - 0.14(55)$$

$$= 28.22 - 4.16 - 7.7 = 16.36 \text{ lt}$$

4° Habiendo obtenido la ecuación de regresión de la muestra, construir la tabla de ANDEVA para la regresión triple, donde la hipótesis nula en este caso es:

$$H_0: B = C = 0$$

$$H_1: B \neq 0 \text{ ó } C \neq 0$$

$$\begin{aligned} SC_{\text{total}} &= \sum (Y_i - \bar{Y})^2 = (21.55 - 17.2947)^2 + (20.07 - 17.2947)^2 + (19.11 - 17.2947)^2 + \\ &\quad (17.83 - 17.2947)^2 + \dots + (14.22 - 17.2947)^2 + (12.74 - 17.2947)^2 \\ &= 92.82 \end{aligned}$$

$$\begin{aligned} SC_{\text{explicada}} &= \sum (\hat{Y}_i - \bar{Y})^2 = bm_{12} + cm_{13} = (-1.0369)(-22.865) + (-0.1388)(- \\ 416.525) \\ &= 23.7087 + 57.8136 = 81.52 \end{aligned}$$

$$\begin{aligned} SC_{\text{error}} &= SC_{\text{total}} - SC_{\text{expl.}} = 92.82 - 81.52 \\ &= 11.30 \end{aligned}$$

El análisis de varianza se muestra en la tabla 3

**Tabla 3**

Fuente	Grados de libertad	Suma de Cuadrados	Varianza	F calc.
Regresión	3 - 1 = 2	81.52	40.76	F = 40.76 / 0.9416 = 43.2941
Error	15 - 3 = 12	11.30	0.9416	
Total	15 - 1 = 14	92.82		

Obteniendo la F teórica tenemos  $F_{0.95, 2, 12} = 3.89$

La hipótesis  $H_0: B = C = 0$  es rechazada, por lo tanto, deducimos que la cantidad de jugo filtrado es significativamente dependiente del espesor de la capa y la velocidad de rotación del tambor rotatorio.

5° El coeficiente de determinación triple es:

$$r^2 = \frac{SC_{\text{expl}}}{SC_{\text{total}}} = \frac{81.52}{92.82} = 0.8783$$

Por lo tanto concluimos que 87.83% de la variación de Y (cantidad de jugo filtrado) es explicada por la variación de  $X_2$  y  $X_3$  (el espesor y la velocidad de rotación).

Y su coeficiente de correlación lineal triple es:

$$r\sqrt{r^2} = 0.9372$$

*(Análisis de correlación)*

5) Supongamos que estamos interesados en investigar si hay o no correlación entre el peso del cuerpo y la concentración de colesterol en la sangre. Seleccionamos aleatoriamente 15 sujetos de una población de varones adultos entre 50 y 55 años de edad con una estatura entre 1.75 y 1.78m. Medimos entonces, el peso y la concentración de colesterol en cada sujeto. Los datos se muestran en la tabla 1. Aunque, tanto el peso como el contenido de colesterol son obviamente variables aleatorias, por conveniencia asignamos los símbolos tradicionales X y Y a las variables.

Tabla 1

Sujeto	Peso (lb) X	Colesterol (mg/100ml) Y	Sujeto	Peso (lb) X	Colesterol (mg/100ml) Y
1	146	181	9	151	112
2	205	228	10	188	241
3	157	182	11	181	225
4	165	249	12	163	223
5	184	259	13	198	257
6	153	201	14	193	337
7	220	339	15	157	197
8	181	224			

Solución:

Con la ayuda de una calculadora obtenemos:

$$\begin{aligned}
 n &= 15 & \Sigma XY &= 622508 \\
 \bar{X} &= 176.1333 & \bar{Y} &= 230.3333 \\
 S_X &= 22.1581 & S_Y &= 57.3544
 \end{aligned}$$

El coeficiente de correlación es entonces,

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{(n-1)S_X S_Y} = \frac{622508 - 15(176.1333)(230.3333)}{14(22.1581)(57.3544)}$$

$$r = 0.785$$

El coeficiente de determinación es:

$$r^2 = 0.6163$$

lo que significa que 61.63% de la variación de una de las variables es explicada por la variación de la otra. Con esto se sabe que el 61.63% del peso de las personas está relacionado por el colesterol, por lo tanto, el 38.37% del peso tiene que ver con otros factores.

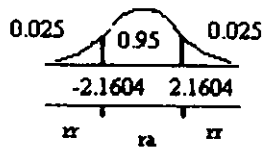
Probemos ahora la hipótesis de linealidad con  $\alpha=0.05$

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$\text{g.l.} = n-2 = 13$$



$$t = \frac{0.785\sqrt{13}}{\sqrt{1-0.6163}} = 4.569$$

Se rechaza  $H_0$ , por lo tanto existe correlación entre las variables X y Y.

## CAPITULO TRES

### PROGRAMA

El objetivo de éste capítulo, es facilitar al estudiante los cálculos que realicen durante su experimentación y/o cuando sea necesario. Esto se va a obtener con la ayuda del programa *Excel*. Lo anterior es con el objetivo de disminuir el tiempo de resolución de un problema y al mismo tiempo los errores de cálculo que se pudieran presentar.

### 1.- PROGRAMA EXCEL

#### 1.1).- INTRODUCCIÓN

Como se menciona anteriormente, este programa tiene la finalidad de ayudar a resolver problemas de una manera más rápida y eficiente.

Excel es una hoja de cálculo para Windows que además permite trabajar con gráficos, bases de datos y macros. Una hoja de cálculo es una tabla compuesta de filas y columnas. La intersección de una fila y una columna se le llama celda. Las filas están identificadas por números y las columnas por letras.

Cada celda tiene una dirección que está dada por la combinación de la letra de la columna y el número de la fila.

Excel ofrece gran cantidad de funciones para realizar cálculos utilizando fórmulas y además cuenta con la ayuda Asistente de funciones, que facilita la creación de fórmulas, y en el caso particular de la suma, tiene el botón Autosuma.

Sumar filas y columnas es, quizá, la operación que más se ejecuta en una hoja de cálculo.

Una fórmula es una ecuación que realiza cálculos en una celda. Puede crearse una fórmula que realice operaciones matemáticas, como la suma o la multiplicación. Cuando se introduzca una fórmula a una hoja de cálculo, es necesario poner el signo = al inicio de la fórmula. (ésto le indica a la hoja de cálculo que se está introduciendo una fórmula) Se puede crear una serie de fórmulas, las cuales, tienen la función de una secuencia de cálculo para la resolución de diversos problemas.

#### 1.2).- ALGUNOS ERRORES QUE GENERAN LAS FÓRMULAS

Cuando una fórmula está mal redactada o no puede calcularse un valor, Excel envía un mensaje de error a la celda activada.

Los valores de error siempre comienzan con el signo de número (#).

Algunos de los errores que se presentan son los siguientes:

#¡DIV/0!	Esto significa que se intenta dividir por cero.
#¡NULO!	Especifica una intersección no válida en dos áreas.
#N/A	Significa que no hay ningún valor en algunas de las celdas utilizadas en la fórmula.
#¡NOMBRE?	Excel no reconoce el nombre utilizado en la fórmula.
#¡NUM!	Se está utilizando un número incorrecto.
#¡REF!	Se refiere a una celda no válida.

#¡VALOR! Se utiliza un argumento o un operando equivocado.  
 ##### No es precisamente un error, significa que el resultado no cabe en la celda. El valor no se pierde; puede aumentar el tamaño de la celda hasta que aparezca el dato correspondiente.

### 1.3).- INTRODUCCION DE FORMULAS CON ASISTENTE DE FUNCIONES

Para la introducción de fórmulas, Excel ofrece el **Asistente para funciones** que puede generar desde una fórmula muy sencilla hasta fórmulas complejas.

Una función es una fórmula que realiza una operación con base en datos. Por ejemplo, =SUMA(C31:C36) es una función que suma los datos desde la celda C31 hasta la C36, que es lo mismo que hacer la operación = (C31+C32+C33+C34+C35+C36).

Hay funciones que no necesitan parámetros, es decir, dentro del paréntesis no va ningún argumento, por ejemplo, la función =AHORA( ) devuelve el número de serie de la fecha y hora actuales.

#### **Asistente para funciones**

El asistente para funciones permite ejecutar fórmulas por medio de funciones facilitando las operaciones en una hoja de cálculo. Asistente para funciones contiene diferentes categorías, dentro de las cuales se encuentra una llamada **Estadísticas**.

Dentro de la categoría de estadística se encuentran diferentes funciones, que a continuación se mencionan:

<u>FUNCION</u>	<u>DEFINICION</u>
BINOM.CRIT	Devuelve el valor mínimo para el cual la desviación binomial acumulativa es menor o igual que un valor de un criterio.
COEF.DE.CORREL	Devuelve el coeficiente de correlación entre dos conjuntos de datos.
COEFICIENTE.ASIMETRIA	Devuelve el sesgo de una distribución.
COEFICIENTE.R2	Devuelve el cuadrado del coeficiente de correlación del momento del producto Pearson.
CONTARA	Cuenta cuántos valores hay en la lista de argumentos.
CONTAR	Cuenta cuántos números hay en la lista de argumentos.
COVAR	Devuelve la covarianza, el promedio de los productos de las desviaciones pareadas.
CRECIMIENTO	Devuelve valores su curva exponencial.
CUARTIL	Devuelve el cuartil de un conjunto de datos.
CURTOSIS	Devuelve la curtosis de un conjunto de datos.
DESVESTA	Calcula la desviación estándar de una muestra, incluidos números, texto y valores lógicos.
DESVESTPA	Calcula la desviación estándar de toda la población, incluidos números, texto y valores lógicos.
DESVESTP	Calcula la desviación estándar de toda la población.
DESVEST	Calcula la desviación estándar de una muestra.
DESVIA2	Devuelve la suma de los cuadrados de las desviaciones.
DESVPROM	Devuelve el promedio de las desviaciones absolutas de los puntos

DIST.WEIBULL	de datos partiendo de su media. Devuelve la distribución Weibull.
DISTR.BETA.INV	Devuelve el valor inverso de la función de densidad de probabilidad beta acumulativa.
DISTR.BETA	Devuelve la función de densidad de probabilidad beta acumulativa.
DISTR.BINOM	Devuelve la probabilidad de distribución binomial de un término individual.
DISTR.CHI	Devuelve la probabilidad de una cola de la distribución chi cuadrado.
DISTR.EXP	Devuelve la distribución exponencial.
DISTR.F.INV	Devuelve el valor inverso de la función de distribución de probabilidad F.
DISTR.F	Devuelve la distribución de probabilidad F.
DISTR.GAMMA. INV	Devuelve el valor inverso de la función gamma acumulativa.
DISTR.GAMMA	Devuelve la distribución gamma.
DISTR.HIPERGEOM	Devuelve la distribución hipergeométrica.
DISTR.LOG.INV	Devuelve el valor inverso de la función de distribución logarítmica normal.
DISTR.LOG.NORM	Devuelve la distribución logarítmica normal acumulativa.
DISTR.NORM. ESTAND.INV	Devuelve el valor inverso de la distribución normal acumulativa estándar.
DISTR.NORM. ESTAND	Devuelve la distribución estándar normal acumulativa.
DISTR.NORM.INV	Devuelve el valor inverso de la función de distribución normal acumulativa.
DISTR.NORM	Devuelve la distribución normal acumulativa.
DISTR.T.INV	Devuelve el valor inverso de la distribución t de Student.
DISTR.T	Devuelve la distribución t de Student.
ERROR.TIPICO.XY	Devuelve el error estándar del valor y predicho de cada valor X de la curva de regresión.
ESTIMACION. LINEAL	Devuelve los parámetros de una tendencia lineal.
ESTIMACION. LOGARITMICA	Devuelve los parámetros de una tendencia exponencial.
FISHER	Devuelve la transformación de Fisher.
FRECUENCIA	Devuelve una distribución de frecuencia como una matriz vertical.
GAMMA.LN	Devuelve el logaritmo natural (neperiano) de la función gamma, $G(x)$ .
INTERSECCION.EJE	Devuelve la intersección de la línea de regresión lineal.
INTERVALO. CONFIANZA	Devuelve el intervalo de confianza de la media de un población.
JERARQUIA	Devuelve la jerarquía (rango) de un número en una lista de números.



K.ESIMO.MAYOR	Devuelve el valor máximo k-ésimo en un conjunto de datos.
K.ESIMO.MENOR	Devuelve el valor k-ésimo mínimo en un conjunto de datos.
MAXA	Devuelve el valor máximo de una lista de argumentos, incluidos números, texto y valores lógicos.
MAX	Devuelve el valor máximo de una lista de argumentos.
MEDIA.ACOTADA	Devuelve la media del interior de un conjunto de datos.
MEDIA.ARMO	Devuelve la media armónica.
MEDIA.GEOM	Devuelve la media geométrica.
MEDIANA	Devuelve la mediana de los números dados.
MINA	Devuelve el valor mínimo de una lista de argumentos, incluidos números, texto y valores lógicos.
MIN	Devuelve el valor mínimo de una lista de argumentos.
MODA	Devuelve el valor más común en un conjunto de datos.
NEGBINOMDIST	Devuelve la distribución binomial negativa.
NORMALIZACION	Devuelve un valor normalizado.
PEARSON	Devuelve el coeficiente de correlación del momento del producto Pearson.
PENDIENTE	Devuelve la pendiente de la línea de regresión lineal.
PERCENTIL	Devuelve el percentil k-ésimo de valores de un rango.
PERMUTACIONES	Devuelve el número de permutaciones de un número de objetos dado.
POISSON	Devuelve la distribución Poisson.
PROBABILIDAD	Devuelve la probabilidad de que los valores de un rango estén comprendidos entre dos límites.
PROMEDIOA	Devuelve el promedio de sus argumentos, incluidos números, texto y valores lógicos.
PROMEDIO	Devuelve el promedio de sus argumentos.
PRONOSTICO	Devuelve un valor junto con una tendencia lineal.
PRUEBA.CHI.INV	Devuelve el valor inverso de la probabilidad una de una cola de la distribución chi cuadrado.
PRUEBA.CHI	Devuelve la prueba de independencia.
PRUEBA.FISHER. INV	Devuelve el valor inverso de la transformación de Fisher.
PRUEBA.F	Devuelve el resultado de una prueba F.
PRUEBA.T	Devuelve la probabilidad asociada a una prueba t de Student.
PRUEBA.Z	Devuelve el valor P de dos colas de una prueba Z.
RANGO.PERCENTIL	Devuelve el rango del percentil de un valor en un conjunto de datos.
TENDENCIA	Devuelve los valores y su tendencia lineal.
VARA	Calcula la varianza de una muestra, incluidos números, texto y valores lógicos.
VARPA	Calcula la varianza de toda la población, incluidos números, texto y valores lógicos.
VARP	Calcula la varianza de toda la población.
VAR	Calcula la varianza de una muestra.

Cabe mencionar que en Excel, así como también en todo el ambiente Windows, se cuenta con **Asistente de ideas** ó **Ayudante de Office**, el cual proporciona consejos útiles sobre lo que se esté ejecutando.

#### 1.4).- PROBLEMAS RESUELTOS

Para entender mejor lo anterior, a continuación se resolverán varios problemas, los cuales fueron seleccionados arbitrariamente.

##### 1) DISTRIBUCION BINOMIAL.

Cuando lanza una moneda al aire, sólo puede obtener dos resultados: cara o cruz. La probabilidad de que la primera vez que la lance resulte en cara es 0,5 y la probabilidad de que exactamente 6 de 10 intentos resulten en cara es:

Solución:

Para resolver el problema utilizaremos el **Asistente para funciones**, por lo que se necesitan los siguientes datos:

- ❖ Número de éxitos. Es el número de éxitos en el ensayo
- ❖ Ensayos. Es el número de ensayos independientes.
- ❖ Probabilidad de éxito. Es la probabilidad de éxito en cada ensayo.
- ❖ Acumulado. Es un valor lógico que determina la forma de la función. Si el argumento es VERDADERO, devuelve la función acumulada; si es FALSO, devuelve la densidad de probabilidad. Si el argumento acumulado es VERDADERO, DISTR.BINOM devuelve la función de distribución acumulada, que es la probabilidad de que exista el máximo número de éxitos; si es FALSO, devuelve la función de masa de probabilidad, que es la probabilidad de que un evento se reproduzca un número de veces igual al argumento núm\_éxito.

La información anterior fue obtenida del **Ayudante de Office**, en caso de necesitar más información de la función, acudir a éste.

Entonces:

Número de éxito: 6

Ensayos: 10

Probabilidad de éxito: 0.5

Acumulado: falso

Por lo que la **sintaxis** (ó manera de escribirse) será la siguiente:

DISTR.BINOM(núm\_éxito;ensayos;prob\_éxito;acumulado)

Sustituyendo los datos en la fórmula de la sintaxis, queda:

DISTR.BINOM(6;10;0.5;FALSO)

y el resultado es : 0.205078125  $\approx$  20.50% es la probabilidad que de 10 tiros, 6 de ellos sean cara.

*Nota: no es necesario escribir la sintaxis con letras mayúsculas.*

## 2) DESVIACION ESTANDAR

Supongamos que toma una muestra aleatoria de 10 herramientas forjadas por la misma máquina durante un ciclo de producción y cuya resistencia a la rotura desea medir. Los valores de la muestra son 1345, 1301, 1368, 1322, 1310, 1370, 1318, 1350, 1303, 1299.

### Solución:

Se introducen los valores de la muestra en la hoja de cálculo, acomodándolos en forma de columna o de fila.

La **sintaxis** de esta función es:

DESVEST(número1; número2; ...)

Sustituyendo los valores en la fórmula de la sintaxis, queda:

DESVEST(1345;1301;1368;1322;1310;1370;1318;1350;1303;1299)

ó se puede escribir el rango en el cual se colocaron los datos, por ejemplo, si los datos fueron introducidos en las celdas A1,A2,A3,A4,A5,A6,A7,A8,A9,A10; entonces.

Desvest(A1:A10)

y el resultado es: 27.46

## 3) INTERVALO DE CONFIANZA

Supongamos que observa una muestra de 50 personas que realizan diariamente un trayecto, y que la duración media de dicho trayecto es de 30 minutos, con una desviación estándar de la población de 2,5. Puede tener una confianza del 95 por ciento de que la media de la población se encuentra en el intervalo:

### Solución:

Para resolver el problema se necesitan los siguientes datos:

- ❖ **Alfa**. es el nivel de significación empleado para calcular el nivel de confianza. El nivel de confianza es igual a  $100(1 - \text{alfa})\%$ , es decir, un alfa de 0,05 indica un nivel de confianza de 95%.
- ❖ **Desv estándar**. Es la desviación estándar de la población y se asume que es conocida
- ❖ **Tamaño**. Es el tamaño de la muestra.

La información anterior fue obtenida del **Ayudante de Office**, en caso de necesitar más información de la función, acudir a éste.

Entonces:

Alfa: 0.05

Desviación estándar: 2.50

Tamaño: 50 personas

Por lo que la **sintaxis** será la siguiente:

INTERVALO.CONFIANZA(alfa;desv\_estándar;tamaño)

Sustituyendo los datos en la fórmula de la sintaxis, queda:

Intervalo de confianza(0.05;2.5;50)

y el resultado es igual a 0,692951. En otras palabras, la longitud media de desplazamiento es  $30 \pm 0,692951$  minutos, o de 29,3 a 30,7 minutos.

1.5).- COMPARACION ENTRE UN PROBLEMA RESUELTO MANUALMENTE Y POR EL PROGRAMA EXCEL.

Se retomará el problema No. 1 planteado en el tema de distribución binomial (Capítulo primero, tema IV). En este problema se determinó la probabilidad de que un microorganismo(m.o.) sobreviva a un tratamiento térmico. Las probabilidades a determinar fueron:

- Sobrevivan a lo más 10.
- Sobrevivan entre 3 y 8 inclusive.
- Sobrevivan exactamente 5.

Datos:  $p=0.4$  y  $N=15$

Solución manual:

- Con una  $x$  menor a 10 buscamos la probabilidad en la tabla de distribución acumulativa binomial. El resultado fue 0.9662 y se le resta a 1 el valor obtenido, por lo que el resultado es 0.0338, o sea, 3.38% es la probabilidad de que sobrevivan al menos 10 m.o.
- Se busca en la misma tabla el valor de la probabilidad. Primero con una  $x=8$  y después con una  $x=2$ . Los resultados que se obtuvieron se restaron, teniendo como resultado final 0.8779, o sea, 87.79% es la probabilidad de que sobrevivan entre 3 y 8 m.o.
- El procedimiento que se utilizó fue el mismo que el anterior, pero con una  $x=5$  y una  $x=4$ . El resultado es 0.1859, o sea, 18.59% como probabilidad que sobrevivan 5 m.o.

Solución con el programa Excel:

- Se tiene acceso a Excel y después al asistente de funciones. Se elige Estadísticas en categorías de funciones y Distribución binomial en Nombre de la función. Los datos a introducir en esta función son:

Número de éxito: 9

Ensayos: 15

Probabilidad de éxito: 0.4

Acumulado: verdadero

Se introducen los datos y el resultado es 0.96616. Al resultado se le resta 1, teniendo 0.03384, o sea, 3.384% es la probabilidad de que sobrevivan 10 m.o.

- Se lleva a cabo el mismo procedimiento que el anterior. Los datos son:

Para 8: Número de éxito: 8

Para 2: Número de éxito: 2

Ensayos: 15

Ensayos: 15

Probabilidad de éxito: 0.4

Probabilidad de éxito: 0.4

Acumulado: verdadero

Acumulado: verdadero

Se restan los resultados, 0.90495-0.0271 respectivamente. Teniendo 0.87785, o sea, 87.785 es la probabilidad que sobrevivan entre 3 y 8 m.o.

c) Se lleva a cabo el mismo procedimiento que el anterior. Los datos son:

Para 5: Número de éxito: 5

Ensayos: 15

Probabilidad de éxito: 0.4

Acumulado: falso

El resultado es 0.1859, o sea, 18.59% como probabilidad que sobrevivan 5 m.o.

#### *Conclusión:*

Para los cálculos que se realizan a mano es necesario tener las tablas de distribución binomial, o bien utilizar la ecuación que se muestra en el Capítulo Segundo, por lo que son indispensables para la resolución del problema. Ya que las tablas presentan gran cantidad de valores, se puede cometer el error de copiar un valor en lugar de otro, un punto importante es que toma tiempo encontrar los valores en las tablas.

Para los cálculos realizados por el programa no es necesario el uso de las tablas ni mucho menos de la ecuación, ya que vienen incorporados los valores en el mismo. Solo es necesario introducir 4 números y automáticamente se tiene el resultado; es más, si se desea se puede mantener abierta la ventana de asistente de funciones en donde se pueden estar realizando un número ilimitado de operaciones referentes al tema y solamente se va escribiendo los resultados.

Las ventajas del uso del programa para la resolución de problemas, es que se tiene una disminución del tiempo, no se comete error, es más práctico.

Como desventaja, es necesario tener un conocimiento de estadística para poder trabajar en el programa

## CONCLUSIONES

- ❖ La Estadística es una herramienta que nos puede servir para interpretar, describir y conocer un evento (este evento puede ser tan sencillo o tan complejo como uno quiera), así como también, nos permite analizar a un conjunto de datos de una población en particular.
- ❖ En este manual se puede observar que el uso de la estadística no es tan complejo como parece, pero, se requiere que la persona ó interesado tenga los conocimientos o herramientas básicas de la estadística para poder realizar un análisis más complejo a los datos que lo requieran.
- ❖ El introducir problemas de estadística relacionados con la carrera de Ingeniero en Alimentos fue de vital importancia, ya que, relaciona los términos teóricos de la estadística con eventos involucrados con los Alimentos.
- ❖ Otros de los grandes objetivos de este manual, fue el de orientar al interesado en la materia a comprender la estadística de una manera más fácil e interesante y a su vez facilitar el uso de ecuaciones, tablas y gráficos en aplicaciones cotidianas. Donde todo éste trabajo se puede realizar con ayuda de un programa de computación, con la finalidad de aprovechar la tecnología de hoy en día.
- ❖ Esperando que ésta información le sea de gran utilidad a sus intereses o fines, y provocando con esto despertar un interés por la materia y sus aplicaciones prácticas.

# APENDICE

- TABLA A :VALORES DE LA FUNCION DE DISTRIBUCION ACOMULATIVA BINOMIAL.
- TABLA B :VALORES DE LA FUNCION DE DISTRIBUCION ACOMULATIVA DE POISSON.
- TABLA C :VALORES DE LAS FUNCIONES DE PROBABILIDAD Y DISTRIBUCION ACOMULATIVA PARA LA DIST. HIPERGEOMETRICA.
- TABLA D :VALORES DE LA FUNCION DE DISTRIBUCION ACOMULATIVA NORMAL ESTANDAR.
- TABLA E :VALORES DE CUARTILES DE LA DISTRIBUCION CHI-CUADRADA.
- TABLA F :VALORES DE CUARTILES DE LA DISTRIBUCION t-STUDENT.
- TABLA G :VALORES DE CUARTILES DE LA DISTRIBUCION F.
- TABLA H :VALORES PARA LOS LIMITES DE TOLERANCIA BILATERALES CUANDO HAGA UN MUESTREO DE DIST. NORMALES.
- TABLA I :VALORES PARA LOS LIMITES DE TOLERANCIA UNILATERALES CUANDO HAGA UN MUESTREO DE DIST. NORMALES.
- TABLA J :DISTRIBUCION DE POISSON.
- TABLA K :DISTRIBUCION NORMAL
- TABLA L :TABLA DE NUMEROS ALEATORIOS.
- TABLA M :DISTRIBUCION JI-CUADRADA.
- TABLA N :PROBABILIDADES BINOMIALES.
- TABLA O :VALORES PARA LA PRUEBA DEL RANGO MULTIPLE DE DUNCAN

TABLA A VALORES DE LA FUNCION DE DISTRIBUCION ACUMULATIVA BINOMIAL.

$$F(x; n, p) = P(X \leq x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$$

n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.0001	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
	1	0.9999	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	0.0001	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.0000	0.0013	0.0044	0.0100	0.0188	0.0313	0.0488	0.0713	0.0988	0.1313	0.1700
	1	0.9999	0.9987	0.9956	0.9899	0.9812	0.9688	0.9512	0.9288	0.9012	0.8688	0.8200
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	3	0.0000	0.0013	0.0044	0.0100	0.0188	0.0313	0.0488	0.0713	0.0988	0.1313	0.1700
4	0	0.0000	0.0006	0.0025	0.0061	0.0133	0.0244	0.0394	0.0583	0.0813	0.1083	0.1383
	1	0.9999	0.9994	0.9975	0.9939	0.9867	0.9756	0.9606	0.9417	0.9183	0.8917	0.8617
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	4	0.0000	0.0006	0.0025	0.0061	0.0133	0.0244	0.0394	0.0583	0.0813	0.1083	0.1383
5	0	0.0000	0.0003	0.0013	0.0033	0.0070	0.0133	0.0225	0.0344	0.0488	0.0650	0.0825
	1	0.9999	0.9997	0.9991	0.9983	0.9973	0.9956	0.9933	0.9906	0.9875	0.9833	0.9783
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	5	0.0000	0.0003	0.0013	0.0033	0.0070	0.0133	0.0225	0.0344	0.0488	0.0650	0.0825
6	0	0.0000	0.0002	0.0006	0.0017	0.0040	0.0083	0.0144	0.0225	0.0325	0.0444	0.0575
	1	0.9999	0.9998	0.9996	0.9993	0.9988	0.9981	0.9973	0.9963	0.9950	0.9933	0.9913
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	6	0.0000	0.0002	0.0006	0.0017	0.0040	0.0083	0.0144	0.0225	0.0325	0.0444	0.0575
7	0	0.0000	0.0001	0.0004	0.0011	0.0025	0.0050	0.0094	0.0156	0.0233	0.0325	0.0425
	1	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993	0.9991	0.9989	0.9987	0.9985	0.9983
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	7	0.0000	0.0001	0.0004	0.0011	0.0025	0.0050	0.0094	0.0156	0.0233	0.0325	0.0425
8	0	0.0000	0.0001	0.0003	0.0008	0.0017	0.0033	0.0061	0.0100	0.0150	0.0213	0.0283
	1	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9992	0.9991
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	8	0.0000	0.0001	0.0003	0.0008	0.0017	0.0033	0.0061	0.0100	0.0150	0.0213	0.0283
9	0	0.0000	0.0001	0.0002	0.0005	0.0011	0.0021	0.0037	0.0056	0.0078	0.0104	0.0133
	1	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	9	0.0000	0.0001	0.0002	0.0005	0.0011	0.0021	0.0037	0.0056	0.0078	0.0104	0.0133
10	0	0.0000	0.0001	0.0001	0.0003	0.0006	0.0013	0.0025	0.0040	0.0056	0.0075	0.0094
	1	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	10	0.0000	0.0001	0.0001	0.0003	0.0006	0.0013	0.0025	0.0040	0.0056	0.0075	0.0094





TABLA A (continuación) Valores de la función de distribución acumulativa Normal

17	0	0.8427	0.4181	0.1648	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.9817	0.7923	0.4818	0.1182	0.0361	0.0133	0.0047	0.0016	0.0005	0.0001	0.0000
	2	0.9994	0.9477	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	1.0000	0.9912	0.9174	0.7516	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	1.0000	0.9988	0.9779	0.9113	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	1.0000	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	1.0000	0.9992	0.9917	0.9623	0.8929	0.7732	0.6188	0.4478	0.2902	0.1682
	7	1.0000	1.0000	0.9999	0.9983	0.9891	0.9598	0.8934	0.7873	0.6405	0.4743	0.3143
	8	1.0000	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6853
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9632	0.9174	0.8358
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9323
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9723
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9981	0.9959
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9996
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.8343	0.3972	0.1504	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.9862	0.7733	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9993	0.9419	0.7238	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0009
	3	1.0000	0.9891	0.9018	0.7282	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0040
	4	1.0000	0.9985	0.9718	0.9194	0.8394	0.7184	0.5887	0.4327	0.2886	0.1811	0.1020
	5	1.0000	0.9998	0.9916	0.9581	0.8672	0.7175	0.5344	0.3530	0.2068	0.1277	0.0720
	6	1.0000	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1382
	7	1.0000	1.0000	0.9998	0.9973	0.9837	0.9431	0.8293	0.7283	0.5634	0.3915	0.2402
	8	1.0000	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4082
	9	1.0000	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7471	0.5822
	10	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7590
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9938	0.9797	0.9463	0.8800
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9528
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9950	0.9870
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9980
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

		$\mu$										
$\sigma$	$z$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.20
19	0	0.8262	0.3774	0.1351	0.0464	0.0141	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.9837	0.7567	0.4201	0.1985	0.0829	0.0310	0.0104	0.0031	0.0010	0.0002	0.0000
	2	0.9991	0.9315	0.7051	0.4433	0.2469	0.1111	0.0462	0.0170	0.0055	0.0015	0.0004
	3	1.0000	0.9868	0.8650	0.6833	0.4933	0.2611	0.1312	0.0591	0.0210	0.0077	0.0022
	4	1.0000	0.9980	0.9648	0.8356	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	1.0000	0.9998	0.9914	0.9463	0.8369	0.6678	0.4719	0.2968	0.1629	0.0777	0.0318
	6	1.0000	1.0000	0.9981	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0833
	7	1.0000	1.0000	0.9999	0.9967	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	1.0000	0.9992	0.9911	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9123	0.8139	0.6710	0.3300
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9937	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9974	0.9909	0.9884	0.9638	0.9163
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9911	0.9882
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.8179	0.3585	0.1216	0.0385	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.9811	0.7358	0.4917	0.2876	0.0927	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9990	0.9245	0.6769	0.4034	0.2061	0.0911	0.0355	0.0121	0.0034	0.0009	0.0002
	3	1.0000	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	1.0000	0.9974	0.9568	0.8298	0.6796	0.4148	0.2575	0.1182	0.0510	0.0189	0.0039
	5	1.0000	0.9997	0.9887	0.9127	0.8042	0.6172	0.4114	0.2454	0.1256	0.0533	0.0207
	6	1.0000	1.0000	0.9976	0.9781	0.9131	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	1.0000	0.9996	0.9943	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	1.0000	0.9999	0.9987	0.9909	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	1.0000	0.9998	0.9976	0.9934	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	1.0000	0.9994	0.9961	0.9820	0.8668	0.7525	0.5917	0.4119



TABLA B VALORES DE LA FUNCION DE DISTRIBUCION ACOMULATIVA DE POISSON.

$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$$

λ										
λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9964
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

λ										
λ	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.3129	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.6950	0.6626	0.6268	0.5718	0.5278	0.4932	0.4628	0.4328	0.4030	0.3732
2	0.9048	0.8795	0.8371	0.8335	0.8088	0.7814	0.7572	0.7306	0.7017	0.6707
3	0.9741	0.9662	0.9369	0.9463	0.9144	0.9212	0.9068	0.8913	0.8747	0.8571
4	0.9916	0.9923	0.9893	0.9837	0.9814	0.9763	0.9704	0.9636	0.9559	0.9473
5	0.9981	0.9983	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868	0.9834
6	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966	0.9956
7	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992	0.9989
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

λ										
λ	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0741	0.0672	0.0614	0.0560	0.0510
1	0.1746	0.1546	0.1369	0.1208	0.1071	0.0954	0.0857	0.0778	0.0714	0.0654
2	0.4446	0.4227	0.3980	0.3697	0.3378	0.3030	0.2662	0.2284	0.1906	0.1530
3	0.8156	0.8194	0.7993	0.7728	0.7506	0.7320	0.7161	0.7020	0.6896	0.6787
4	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8163
5	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9911	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9965	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
8	0.9992	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
9	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

λ										
λ	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1847	0.1712	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916
2	0.4812	0.3799	0.3394	0.3197	0.3028	0.3027	0.2854	0.2689	0.2531	0.2381
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4533	0.4335
4	0.7982	0.7806	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6484	0.6288
5	0.9057	0.8946	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851
6	0.9612	0.9554	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893
7	0.9758	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489
8	0.9893	0.9913	0.9911	0.9917	0.9921	0.9923	0.9923	0.9920	0.9915	0.9909
9	0.9966	0.9982	0.9982	0.9982	0.9981	0.9980	0.9979	0.9977	0.9974	0.9971
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972
11	0.9999	0.9999	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991
12	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

λ										
λ	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0.0186	0.0156	0.0130	0.0107	0.0087	0.0070	0.0056	0.0045	0.0036	0.0027
1	0.0845	0.0720	0.0613	0.0523	0.0448	0.0386	0.0336	0.0296	0.0264	0.0231
2	0.1746	0.1546	0.1369	0.1208	0.1071	0.0954	0.0857	0.0778	0.0714	0.0654
3	0.4446	0.4227	0.3980	0.3697	0.3378	0.3030	0.2662	0.2284	0.1906	0.1530
4	0.8156	0.8194	0.7993	0.7728	0.7506	0.7320	0.7161	0.7020	0.6896	0.6787
5	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8163
6	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
7	0.9911	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
8	0.9965	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
9	0.9992	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
10	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
11	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

TABLE B (Continued): Values of the function of distribution cumulative of Pareto

	1	2	3	4	5	6	7	8	9	10
0	0.0001	0.0015	0.0050	0.0165	0.0491	0.1401	0.0313	0.0030	0.0027	0.0023
1	0.0112	0.0342	0.0914	0.2259	0.4526	0.6216	0.7236	0.7896	0.8189	0.8274
2	0.0165	0.0489	0.1016	0.2314	0.4582	0.6272	0.7292	0.7952	0.8245	0.8330
3	0.0218	0.0642	0.1374	0.2719	0.5000	0.6700	0.7720	0.8380	0.8673	0.8758
4	0.0271	0.0840	0.1809	0.3165	0.5319	0.6919	0.7939	0.8600	0.8893	0.8978
5	0.0324	0.1038	0.2160	0.3621	0.5638	0.7238	0.8258	0.8919	0.9212	0.9297
6	0.0377	0.1236	0.2514	0.4073	0.5957	0.7557	0.8577	0.9238	0.9531	0.9616
7	0.0430	0.1434	0.2860	0.4525	0.6276	0.7876	0.8896	0.9557	0.9850	0.9935
8	0.0483	0.1632	0.3207	0.4977	0.6595	0.8195	0.9215	0.9876	1.0000	1.0000
9	0.0536	0.1830	0.3549	0.5429	0.6914	0.8514	0.9534	1.0000	1.0000	1.0000
10	0.0589	0.2028	0.3891	0.5881	0.7233	0.8833	0.9853	1.0000	1.0000	1.0000
11	0.0642	0.2226	0.4233	0.6333	0.7552	0.9152	0.9972	1.0000	1.0000	1.0000
12	0.0695	0.2424	0.4575	0.6785	0.7871	0.9471	1.0000	1.0000	1.0000	1.0000
13	0.0748	0.2622	0.4917	0.7237	0.8190	0.9790	1.0000	1.0000	1.0000	1.0000
14	0.0801	0.2820	0.5259	0.7689	0.8509	0.9909	1.0000	1.0000	1.0000	1.0000
15	0.0854	0.3018	0.5601	0.8141	0.8828	0.9928	1.0000	1.0000	1.0000	1.0000
16	0.0907	0.3216	0.5943	0.8593	0.9147	0.9947	1.0000	1.0000	1.0000	1.0000
17	0.0960	0.3414	0.6285	0.9045	0.9466	0.9966	1.0000	1.0000	1.0000	1.0000
18	0.1013	0.3612	0.6627	0.9497	0.9785	0.9985	1.0000	1.0000	1.0000	1.0000
19	0.1066	0.3810	0.6969	0.9949	0.9904	0.9999	1.0000	1.0000	1.0000	1.0000
20	0.1119	0.4008	0.7311	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
21	0.1172	0.4206	0.7653	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0072	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0157	0.0146	0.0134	0.0123	0.0113	0.0103	0.0095	0.0087	0.0080	0.0073
2	0.0242	0.0236	0.0224	0.0213	0.0203	0.0193	0.0185	0.0177	0.0170	0.0163
3	0.0327	0.0322	0.0310	0.0300	0.0290	0.0280	0.0272	0.0264	0.0257	0.0250
4	0.0412	0.0407	0.0395	0.0384	0.0374	0.0364	0.0356	0.0348	0.0341	0.0334
5	0.0497	0.0492	0.0480	0.0470	0.0460	0.0450	0.0442	0.0434	0.0427	0.0420
6	0.0582	0.0577	0.0565	0.0554	0.0544	0.0534	0.0526	0.0518	0.0511	0.0504
7	0.0667	0.0662	0.0650	0.0640	0.0630	0.0620	0.0612	0.0604	0.0597	0.0590
8	0.0752	0.0747	0.0735	0.0724	0.0714	0.0704	0.0696	0.0688	0.0681	0.0674
9	0.0837	0.0832	0.0820	0.0810	0.0800	0.0790	0.0782	0.0774	0.0767	0.0760
10	0.0922	0.0917	0.0905	0.0894	0.0884	0.0874	0.0866	0.0858	0.0851	0.0844
11	0.1007	0.1002	0.0990	0.0980	0.0970	0.0960	0.0952	0.0944	0.0937	0.0930
12	0.1092	0.1087	0.1075	0.1064	0.1054	0.1044	0.1036	0.1028	0.1021	0.1014
13	0.1177	0.1172	0.1160	0.1150	0.1140	0.1130	0.1122	0.1114	0.1107	0.1100
14	0.1262	0.1257	0.1245	0.1234	0.1224	0.1214	0.1206	0.1198	0.1191	0.1184
15	0.1347	0.1342	0.1330	0.1320	0.1310	0.1300	0.1292	0.1284	0.1277	0.1270
16	0.1432	0.1427	0.1415	0.1404	0.1394	0.1384	0.1376	0.1368	0.1361	0.1354
17	0.1517	0.1512	0.1500	0.1490	0.1480	0.1470	0.1462	0.1454	0.1447	0.1440
18	0.1602	0.1597	0.1585	0.1574	0.1564	0.1554	0.1546	0.1538	0.1531	0.1524
19	0.1687	0.1682	0.1670	0.1660	0.1650	0.1640	0.1632	0.1624	0.1617	0.1610

	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0004
1	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0013	0.0013
2	0.0026	0.0025	0.0025	0.0024	0.0024	0.0023	0.0023	0.0022	0.0022	0.0022
3	0.0035	0.0034	0.0034	0.0033	0.0033	0.0032	0.0032	0.0031	0.0031	0.0031
4	0.0044	0.0043	0.0043	0.0042	0.0042	0.0041	0.0041	0.0040	0.0040	0.0040
5	0.0053	0.0052	0.0052	0.0051	0.0051	0.0050	0.0050	0.0049	0.0049	0.0049
6	0.0062	0.0061	0.0061	0.0060	0.0060	0.0059	0.0059	0.0058	0.0058	0.0058
7	0.0071	0.0070	0.0070	0.0069	0.0069	0.0068	0.0068	0.0067	0.0067	0.0067
8	0.0080	0.0079	0.0079	0.0078	0.0078	0.0077	0.0077	0.0076	0.0076	0.0076
9	0.0089	0.0088	0.0088	0.0087	0.0087	0.0086	0.0086	0.0085	0.0085	0.0085
10	0.0098	0.0097	0.0097	0.0096	0.0096	0.0095	0.0095	0.0094	0.0094	0.0094
11	0.0107	0.0106	0.0106	0.0105	0.0105	0.0104	0.0104	0.0103	0.0103	0.0103
12	0.0116	0.0115	0.0115	0.0114	0.0114	0.0113	0.0113	0.0112	0.0112	0.0112
13	0.0125	0.0124	0.0124	0.0123	0.0123	0.0122	0.0122	0.0121	0.0121	0.0121
14	0.0134	0.0133	0.0133	0.0132	0.0132	0.0131	0.0131	0.0130	0.0130	0.0130
15	0.0143	0.0142	0.0142	0.0141	0.0141	0.0140	0.0140	0.0139	0.0139	0.0139
16	0.0152	0.0151	0.0151	0.0150	0.0150	0.0149	0.0149	0.0148	0.0148	0.0148
17	0.0161	0.0160	0.0160	0.0159	0.0159	0.0158	0.0158	0.0157	0.0157	0.0157
18	0.0170	0.0169	0.0169	0.0168	0.0168	0.0167	0.0167	0.0166	0.0166	0.0166
19	0.0179	0.0178	0.0178	0.0177	0.0177	0.0176	0.0176	0.0175	0.0175	0.0175
20	0.0188	0.0187	0.0187	0.0186	0.0186	0.0185	0.0185	0.0184	0.0184	0.0184
21	0.0197	0.0196	0.0196	0.0195	0.0195	0.0194	0.0194	0.0193	0.0193	0.0193

	R 1	R 2	R 3	R 4	R 5	R 6	R 7	R 8	R 9	R 10
0	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006	0.0006	0.0006
2	0.0017	0.0017	0.0017	0.0016	0.0016	0.0016	0.0016	0.0015	0.0015	0.0015
3	0.0026	0.0026	0.0026	0.0025	0.0025	0.0025	0.0025	0.0024	0.0024	0.0024
4	0.0035	0.0035	0.0035	0.0034	0.0034	0.0034	0.0034	0.0033	0.0033	0.0033
5	0.0044	0.0044	0.0044	0.0043	0.0043	0.0043	0.0043	0.0042	0.0042	0.0042
6	0.0053	0.0053	0.0053	0.0052	0.0052	0.0052	0.0052	0.0051	0.0051	0.0051
7	0.0062	0.0062	0.0062	0.0061	0.0061	0.0061	0.0061	0.0060	0.0060	0.0060
8	0.0071	0.0071	0.0071	0.0070	0.0070	0.0070	0.0070	0.0069	0.0069	0.0069
9	0.0080	0.0080	0.0080	0.0079	0.0079	0.0079	0.0079	0.0078	0.0078	0.0078
10	0.0089	0.0089	0.0089	0.0088	0.0088	0.0088	0.0088	0.0087	0.0087	0.0087
11	0.0098	0.0098	0.0098	0.0097	0.0097	0.0097	0.0097	0.0096	0.0096	0.0096
12	0.0107	0.0107	0.0107	0.0106	0.0106	0.0106	0.0106	0.0105	0.0105	0.0105
13	0.0116	0.0116	0.0116	0.0115	0.0115	0.0115	0.0115	0.0114	0.0114	0.0114
14	0.0125	0.0125	0.0125	0.0124	0.0124	0.0124	0.0124	0.0123	0.0123	0.0123
15	0.0134	0.0134	0.0134	0.0133	0.0133	0.0133	0.0133	0.0132	0.0132	0.0132
16	0.0143	0.0143	0.0143	0.0142	0.0142	0.0142	0.0142	0.0141	0.0141	0.0141
17	0.0152	0.0152	0.0152	0.0151	0.0151	0.0151	0.0151	0.0150	0.0150	0.0150
18	0.0161	0.0161	0.0161	0.0160	0.0160	0.0160	0.0160	0.0159	0.0159	0.0159
19	0.0170	0.0170	0.0170	0.0169	0.0169	0.0169	0.0169	0.0168	0.0168	0.0168
20	0.0179	0.0179	0.0179	0.0178	0.0178	0.0178	0.0178	0.0177	0.0177	0.0177
21	0.0188	0.0188	0.0188	0.0187	0.0187	0.0187	0.0187	0.0186	0.0186	0.0186



TABLA C. VALORES DE LAS FUNCIONES DE PROBABILIDAD Y DISTRIBUCION ACOMULATIVA PARA LA DIST. HIPERGEOMETRICA.

$$P(X = x) = \frac{r! (n - r)! (N - n)!}{(n!) (N - n)!} \sum_{j=0}^x \frac{\binom{r}{j} \binom{N-r}{n-j}}{\binom{N}{n}}$$

N	n	r	x	P(x)	F(x)	N	n	r	x	P(x)	F(x)
2	1	1	0	0.500000	0.500000	6	3	2	2	1.000000	0.966667
2	1	1	1	0.500000	1.000000	6	3	2	2	1.000000	0.966667
3	1	1	0	0.666667	0.666667	6	3	2	0	0.200000	0.200000
3	1	1	1	0.333333	1.000000	6	3	2	1	0.800000	0.600000
3	2	1	0	0.666667	0.666667	6	3	2	2	1.000000	0.200000
3	2	1	1	0.333333	1.000000	6	3	2	3	1.000000	0.400000
3	2	2	2	0.333333	1.000000	6	3	3	1	0.500000	0.400000
3	2	2	1	0.666667	0.666667	6	3	3	2	0.500000	0.400000
3	2	2	0	0.333333	0.333333	6	3	3	3	1.000000	0.800000
4	1	1	0	0.750000	0.750000	6	4	1	0	0.333333	0.333333
4	1	1	1	0.250000	1.000000	6	4	1	1	0.666667	0.666667
4	2	1	0	0.666667	0.666667	6	4	2	0	0.200000	0.200000
4	2	1	1	0.333333	1.000000	6	4	2	1	0.600000	0.333333
4	2	1	2	0.000000	1.000000	6	4	2	2	1.000000	0.400000
4	2	2	2	0.250000	0.250000	6	4	2	3	1.000000	0.200000
4	2	2	1	0.750000	0.750000	6	4	2	4	1.000000	0.200000
4	3	1	0	0.750000	0.750000	6	4	3	2	0.300000	0.200000
4	3	1	1	0.250000	1.000000	6	4	3	3	1.000000	0.200000
4	3	1	2	0.000000	1.000000	6	4	3	4	1.000000	0.200000
4	3	2	2	0.125000	0.125000	6	4	4	2	0.400000	0.400000
4	3	2	1	0.375000	0.375000	6	4	4	3	0.933333	0.333333
4	3	2	0	0.500000	0.500000	6	4	4	4	1.000000	0.666667
4	3	3	3	1.000000	1.000000	6	4	4	4	1.000000	0.666667
5	1	1	0	0.800000	0.800000	6	5	1	0	0.166667	0.166667
5	1	1	1	0.200000	1.000000	6	5	1	1	0.833333	0.833333
5	2	1	0	0.600000	0.600000	6	5	2	1	0.333333	0.333333
5	2	1	1	0.400000	1.000000	6	5	2	2	1.000000	0.666667
5	2	2	2	0.200000	0.200000	6	5	2	3	1.000000	0.500000
5	2	2	1	0.800000	0.800000	6	5	2	4	1.000000	0.666667
5	2	2	0	0.200000	0.200000	6	5	3	4	1.000000	0.666667
5	3	1	0	0.800000	0.800000	6	5	4	4	1.000000	0.333333
5	3	1	1	0.200000	1.000000	6	5	4	5	1.000000	0.333333
5	3	2	2	0.000000	1.000000	6	5	5	5	1.000000	0.166667
5	3	2	1	0.200000	0.200000	7	1	0	0	0.857143	0.857143
5	3	2	1	0.800000	0.800000	7	1	1	1	1.000000	0.428571
5	3	3	1	0.300000	0.300000	7	2	1	0	0.714286	0.714286
5	3	3	2	0.700000	0.700000	7	2	1	1	1.000000	0.285714
5	3	3	3	1.000000	1.000000	7	2	2	0	0.476190	0.476190
5	3	3	4	1.000000	1.000000	7	2	2	1	0.952381	0.476190
5	4	1	0	0.800000	0.800000	7	2	2	2	1.000000	0.952381
5	4	1	1	0.200000	1.000000	7	2	2	3	1.000000	0.952381
5	4	2	2	0.400000	0.400000	7	3	1	0	0.714286	0.714286
5	4	2	1	0.600000	0.600000	7	3	1	1	1.000000	0.428571
5	4	3	2	0.600000	0.600000	7	3	2	0	0.285714	0.285714
5	4	3	1	0.400000	0.400000	7	3	2	1	0.857143	0.714286
5	4	3	0	0.200000	0.200000	7	3	3	0	1.000000	0.428571
6	1	1	0	0.933333	0.933333	7	3	3	1	0.628571	0.314286
6	1	1	1	0.066667	1.000000	7	3	3	2	0.971428	0.428571
6	2	1	0	0.666667	0.666667	7	3	3	3	1.000000	0.285714
6	2	1	1	0.333333	1.000000	7	3	3	4	1.000000	0.285714
6	2	2	0	0.400000	0.400000	7	4	1	1	1.000000	0.714286
6	2	2	1	0.600000	0.600000	7	4	1	2	0.142857	0.428571
7	4	2	1	0.714286	0.714286	8	3	2	2	0.982143	0.367857
7	4	2	2	0.666667	0.666667	8	3	2	3	1.000000	0.632143
7	4	3	0	0.252371	0.252371	8	4	1	0	0.500000	0.500000
7	4	3	1	0.747629	0.747629	8	4	1	1	1.000000	0.300000
7	4	3	2	0.857143	0.857143	8	4	2	0	0.314286	0.314286
7	4	3	3	1.000000	1.000000	8	4	2	1	0.985714	0.991429
7	4	4	1	0.142857	0.142857	8	4	2	2	1.000000	0.192857
7	4	4	2	0.528571	0.528571	8	4	3	0	0.971429	0.628571
7	4	4	3	0.714286	0.714286	8	4	3	1	1.000000	0.428571
7	4	4	4	1.000000	1.000000	8	4	3	2	0.952371	0.428571
7	5	1	0	0.283714	0.283714	8	4	4	3	1.000000	0.971429
7	5	1	1	0.000000	0.000000	8	4	4	4	0.814286	0.814286
7	5	2	0	0.047619	0.047619	8	4	4	5	0.285714	0.285714
7	5	2	1	0.328571	0.328571	8	4	4	6	0.737143	0.714286
7	5	2	2	1.000000	0.416190	8	4	4	7	1.000000	0.285714
7	5	3	1	0.142857	0.142857	8	4	4	8	1.000000	0.114286
7	5	3	2	0.714286	0.714286	8	5	1	0	0.333000	0.333000
7	5	3	3	1.000000	0.283714	8	5	1	1	1.000000	0.133000
7	5	4	2	0.283714	0.283714	8	5	2	0	0.891429	0.107143
7	5	4	3	0.857143	0.714286	8	5	2	1	0.642857	0.337143

ABLA C (continuación)

N°	n	k	s	F(s)	p(s)	N	n	k	s	F(s)	p(s)
7	1	4	4	1 000000	0.142857	8	5	2	2	1.000000	0.357143
7	5	5	3	0.476190	0.476190	8	5	3	0	0.017857	0.017857
7	5	5	4	0.952381	0.476190	8	5	3	1	0.267857	0.267857
7	5	5	5	1.000000	0.017619	8	5	3	2	0.821429	0.357143
7	6	1	0	0.142857	0.142857	8	5	3	3	1.000000	0.178571
7	6	1	1	1.000000	0.857143	8	5	4	1	0.071429	0.071429
7	6	1	1	0.285714	0.285714	8	5	4	2	0.500000	0.428571
7	6	2	2	1.000000	0.714286	8	5	4	3	0.928571	0.428571
7	6	3	2	0.428571	0.428571	8	5	4	4	1.000000	0.071429
7	6	3	3	1.000000	0.571429	8	5	5	2	0.178571	0.178571
7	6	4	3	0.571429	0.571429	8	5	5	3	0.714286	0.357143
7	6	4	4	1.000000	0.428571	8	5	5	4	0.982143	0.267857
7	6	5	4	0.714286	0.714286	8	5	5	5	1.000000	0.078571
7	6	5	5	1.000000	0.285714	8	6	1	0	0.250000	0.250000
7	6	6	5	0.857143	0.857143	8	6	1	0	1.000000	0.750000
7	6	6	6	1.000000	0.142857	8	6	2	0	0.035714	0.035714
8	1	1	0	0.875000	0.875000	8	6	2	1	0.464286	0.428571
8	1	1	1	1.000000	0.125000	8	6	2	2	1.000000	0.357143
8	2	1	0	0.750000	0.750000	8	6	3	1	0.107143	0.107143
8	2	1	1	1.000000	0.250000	8	6	3	2	0.642857	0.535714
8	2	2	0	0.535714	0.535714	8	6	3	3	1.000000	0.357143
8	2	2	1	0.964286	0.428571	8	6	4	2	0.214286	0.214286
8	2	2	2	1.000000	0.035714	8	6	4	3	0.785714	0.321429
8	3	1	0	0.671000	0.625000	8	6	4	4	1.000000	0.214286
8	3	1	1	1.000000	0.375000	8	6	5	4	1.000000	0.357143
8	3	2	0	0.357143	0.357143	8	6	5	5	0.892857	0.357143
8	3	2	1	0.892857	0.535714	8	6	5	5	1.000000	0.107143
8	3	2	2	1.000000	0.107143	8	6	6	4	0.535714	0.535714
8	3	3	0	0.178571	0.178571	8	6	6	5	0.964286	0.428571
8	3	3	1	0.714286	0.535714	8	6	6	6	1.000000	0.035714
8	7	1	0	0.125000	0.125000	9	5	3	1	0.404762	0.357143
8	7	1	1	1.000000	0.875000	9	5	3	2	0.809524	0.476190
8	7	2	1	0.250000	0.250000	9	5	3	3	1.000000	0.119048
8	7	2	2	1.000000	0.750000	9	5	4	0	0.007936	0.007936
8	7	3	2	0.315000	0.315000	9	5	4	1	0.166667	0.150730
8	7	3	3	1.000000	0.425000	9	5	4	2	0.424257	0.076190
8	7	4	3	0.500000	0.500000	9	5	4	3	0.901517	0.117660
8	7	4	4	1.000000	0.500000	9	5	4	4	1.000000	0.089643
8	7	5	4	0.625000	0.425000	9	5	5	1	0.019683	0.079683
8	7	5	5	1.000000	0.125000	9	5	5	2	0.357143	0.317660
8	7	6	5	0.750000	0.750000	9	5	5	3	0.333333	0.476190
8	7	6	6	1.000000	0.250000	9	5	5	4	0.979663	0.187200
8	7	7	6	0.875000	0.875000	9	5	5	5	1.000000	0.007936
8	7	7	7	1.000000	0.125000	9	6	1	0	0.333333	0.333333
9	1	1	0	0.888889	0.888889	9	6	1	1	1.000000	0.666667
9	1	1	1	1.000000	0.111111	9	6	2	0	1.000000	0.053333
9	2	1	0	0.777778	0.777778	9	6	2	1	0.303333	0.300000
9	2	1	1	1.000000	0.222222	9	6	2	2	0.416667	0.416667
9	2	2	0	0.303333	0.303333	9	6	3	0	0.811905	0.011905
9	2	2	1	0.972222	0.388889	9	6	3	1	0.206190	0.214286
9	2	2	2	1.000000	0.027778	9	6	3	2	0.761905	0.357143
9	3	1	0	0.666667	0.666667	9	6	3	3	1.000000	0.280952
9	3	1	1	1.000000	0.333333	9	6	4	1	0.047619	0.047619
9	3	2	0	0.416667	0.416667	9	6	4	2	0.327143	0.327143
9	3	2	1	0.916667	0.500000	9	6	4	3	0.889922	0.476190
9	3	2	2	1.000000	0.083333	9	6	4	4	1.000000	0.190480
9	3	3	0	0.238095	0.238095	9	6	5	2	0.119048	0.119048
9	3	3	1	0.773809	0.537714	9	6	5	3	0.395238	0.476190
9	3	3	2	0.980952	0.714286	9	6	5	4	0.972280	0.357143
9	3	3	3	1.000000	0.011905	9	6	5	5	1.000000	0.067619
9	4	1	0	0.555556	0.555556	9	6	6	1	0.238095	0.238095
9	4	1	1	1.000000	0.444444	9	6	6	2	0.773809	0.337143
9	4	2	0	0.277778	0.277778	9	6	6	3	0.968095	0.214286
9	4	2	1	0.833333	0.533336	9	6	6	4	1.000000	0.061905
9	4	2	2	1.000000	0.166667	9	7	1	0	0.221222	0.222222
9	4	3	0	0.119048	0.119048	9	7	1	1	1.000000	0.777778
9	4	3	1	0.295238	0.476190	9	7	2	0	0.027778	0.027778
9	4	3	2	0.933333	0.371429	9	7	2	1	0.416667	0.300000
9	4	3	3	1.000000	0.047619	9	7	2	2	1.000000	0.333333
9	4	4	0	0.039683	0.039683	9	7	3	1	0.061333	0.061333



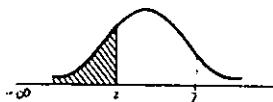
TABLE C (continued)

N	n	k	i	F(x)	f(x)	N	n	k	i	F(x)	f(x)
9	4	4	1	0.157143	0.117160	9	7	2	2	0.583333	0.500000
9	4	4	2	0.833113	0.476190	9	7	3	3	1.000000	0.416667
9	4	4	3	0.992063	0.158730	9	7	4	2	0.166667	0.166667
9	4	4	4	1.000000	0.007936	9	7	4	3	0.727272	0.555556
9	5	1	0	0.444444	0.444444	9	7	4	4	1.000000	0.277778
9	5	1	1	1.000000	0.555556	9	7	5	3	0.277778	0.277778
9	5	2	0	0.166667	0.166667	9	7	5	4	0.833333	0.555556
9	5	2	1	0.727272	0.555556	9	7	5	5	1.000000	0.166667
9	5	2	2	1.000000	0.277778	9	7	6	4	0.416667	0.416667
9	5	3	0	0.047619	0.047619	9	7	6	5	0.916667	0.500000
9	7	6	8	1.000000	0.833333	10	5	1	0	0.500000	0.500000
9	7	7	5	0.583333	0.583333	10	5	1	1	1.000000	0.500000
9	7	7	6	0.922222	0.388889	10	5	2	0	0.222222	0.222222
9	7	7	7	1.000000	0.027778	10	5	2	1	0.777778	0.555556
9	8	1	0	0.111111	0.111111	10	5	2	2	1.000000	0.222222
9	8	1	1	1.000000	0.888889	10	5	3	0	0.083333	0.083333
9	8	2	1	0.222222	0.222222	10	5	3	1	0.500000	0.416667
9	8	2	2	1.000000	0.777778	10	5	3	2	0.916667	0.916667
9	8	3	2	0.333333	0.333333	10	5	3	3	1.000000	0.083333
9	8	3	3	1.000000	0.666667	10	5	4	0	0.023810	0.023810
9	8	4	3	0.444444	0.444444	10	5	4	1	0.261905	0.238095
9	8	4	4	1.000000	0.555556	10	5	4	2	0.738095	0.476190
9	8	5	4	0.555556	0.555556	10	5	4	3	0.976190	0.238095
9	8	5	5	1.000000	0.444444	10	5	4	4	1.000000	0.023810
9	8	6	5	0.666667	0.666667	10	5	5	0	0.003968	0.003968
9	8	6	6	1.000000	0.531111	10	5	5	1	0.101775	0.099206
9	8	7	6	0.777778	0.777778	10	5	5	2	0.500000	0.396825
9	8	7	7	1.000000	0.222222	10	5	5	3	0.896825	0.396825
9	8	8	7	0.888889	0.888889	10	5	5	4	0.996032	0.099206
9	8	8	8	1.000000	0.111111	10	5	5	5	1.000000	0.003968
10	1	1	0	0.500000	0.500000	10	6	1	0	0.400000	0.400000
10	1	1	1	1.000000	0.100000	10	6	1	1	1.000000	0.600000
10	2	1	0	0.900000	0.800000	10	6	2	0	0.133333	0.133333
10	2	1	1	1.000000	0.700000	10	6	2	1	0.666667	0.533333
10	2	2	0	0.622222	0.622222	10	6	2	2	1.000000	0.333333
10	2	2	1	0.977778	0.555556	10	6	3	0	0.053333	0.053333
10	2	2	2	1.000000	0.022222	10	6	3	1	0.333333	0.300000
10	3	1	0	0.700000	0.700000	10	6	3	2	0.833333	0.500000
10	3	1	1	1.000000	0.300000	10	6	3	3	1.000000	0.166667
10	3	2	0	0.466667	0.466667	10	6	4	0	0.047622	0.047622
10	3	2	1	0.933333	0.466667	10	6	4	1	0.119048	0.114286
10	3	2	2	1.000000	0.066667	10	6	4	2	0.547619	0.428571
10	3	3	0	0.291667	0.291667	10	6	4	3	0.928571	0.380952
10	3	3	1	0.816667	0.529000	10	6	4	4	1.000000	0.071429
10	3	3	2	0.991667	0.173000	10	6	5	1	0.023810	0.023810
10	3	3	3	1.000000	0.008333	10	6	5	2	0.261905	0.238095
10	4	1	0	0.600000	0.600000	10	6	5	3	0.718095	0.476190
10	4	1	1	1.000000	0.400000	10	6	5	4	0.976190	0.238095
10	4	2	0	0.333333	0.333333	10	6	5	5	1.000000	0.023810
10	4	2	1	0.846667	0.573333	10	6	6	2	0.071429	0.071429
10	4	2	2	1.000000	0.133333	10	6	6	3	0.452381	0.380952
10	4	3	0	0.166667	0.166667	10	6	6	4	0.809712	0.428571
10	4	3	1	0.666667	0.500000	10	6	6	5	0.995238	0.142868
10	4	3	2	0.966667	0.300000	10	6	6	6	1.000000	0.004762
10	4	3	3	1.000000	0.033333	10	7	1	0	0.300000	0.300000
10	4	4	0	0.071429	0.071429	10	7	1	1	1.000000	0.700000
10	4	4	1	0.452381	0.380952	10	7	2	0	0.066667	0.066667
10	4	4	2	0.809523	0.428571	10	7	2	1	0.333333	0.466667
10	4	4	3	0.995238	0.114286	10	7	2	2	1.000000	0.466667
10	4	4	4	1.000000	0.004762	10	7	3	0	0.008333	0.008333

Source: Taken from Tables of the Hypergeometric probability distribution, por Gerald J. Lieberman y Donald R. Owen, con permiso de los editores. Standard University Press. Copyright © 1961 by the Board of Trustees of the Leland Stanford Junior University.

TABLA D VALORES DE LA FUNCION DE DISTRIBUCION ACOMULATIVA NORMAL ESTANDAR.

$$P(Z \leq z) = F(z, 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-t^2/2) dt$$

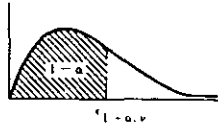


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0733	0.0717	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.2	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.3	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.4	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.5	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.6	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.7	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.8	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.9	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389



TABLA E VALORES DE CUARTILES DE LA DISTRIBUCION CHI-CUADRADA.

$$P(X \leq x) = F(x) = \frac{1}{\Gamma(r/2) 2^{r/2}} \int_0^x t^{r/2-1} \exp(-t/2) dt = 1 - \alpha$$



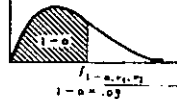
$\alpha$	$\chi_{0.99}$	$\chi_{0.95}$	$\chi_{0.90}$	$\chi_{0.85}$	$\chi_{0.80}$	$\chi_{0.75}$	$\chi_{0.70}$	$\chi_{0.65}$	$\chi_{0.60}$	$\chi_{0.55}$	$\chi_{0.50}$
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.64	7.90	9.21
2	0.01	0.02	0.05	0.10	0.21	4.60	5.99	7.38	9.22	10.59	12.02
3	0.07	0.11	0.22	0.35	0.58	6.25	7.82	9.36	11.32	12.82	14.16
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.15	13.28	14.82	16.15
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.84	15.09	16.76	18.15
6	0.67	0.87	1.24	1.63	2.20	10.65	12.60	14.46	16.81	18.55	20.27
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.02	18.47	20.27	22.32
8	1.34	1.64	2.18	2.73	3.49	13.36	15.51	17.55	20.08	21.94	24.00
9	1.73	2.09	2.70	3.32	4.17	14.69	16.93	19.03	21.65	23.56	25.19
10	2.15	2.55	3.24	3.94	4.86	15.99	18.31	20.50	23.19	25.19	27.15
11	2.60	3.05	3.81	4.57	5.58	17.28	19.68	21.93	24.75	26.71	28.30
12	3.06	3.57	4.40	5.22	6.30	18.55	21.03	23.35	26.25	28.25	30.19
13	3.56	4.10	5.01	5.89	7.04	19.81	22.37	24.75	27.72	29.88	32.01
14	4.07	4.65	5.62	6.57	7.79	21.07	23.69	26.13	29.17	31.38	33.80
15	4.59	5.23	6.26	7.26	8.55	22.31	25.00	27.50	30.61	32.86	35.58
16	5.14	5.81	6.90	7.96	9.31	23.55	26.30	28.86	32.03	34.32	37.15
17	5.69	6.40	7.56	8.67	10.08	24.77	27.59	30.20	33.43	35.77	38.58
18	6.25	7.00	8.23	9.39	10.86	25.99	28.88	31.54	34.83	37.21	40.00
19	6.82	7.61	8.90	10.11	11.65	27.21	30.15	32.87	36.22	38.63	41.41
20	7.42	8.25	9.59	10.85	12.44	28.42	31.42	34.18	37.59	40.05	42.91

$\alpha$	$\chi_{0.50}$	$\chi_{0.40}$	$\chi_{0.30}$	$\chi_{0.20}$	$\chi_{0.10}$	$\chi_{0.05}$	$\chi_{0.025}$	$\chi_{0.01}$	$\chi_{0.005}$	$\chi_{0.001}$
21	8.02	8.89	10.28	11.59	13.24	29.62	32.68	35.49	38.96	41.45
22	8.62	9.53	10.98	12.34	14.04	30.82	33.93	36.79	40.31	42.84
23	9.25	10.19	11.69	13.09	14.85	32.01	35.18	38.09	41.66	44.23
24	9.87	10.85	12.40	13.84	15.66	33.20	36.42	39.38	43.00	45.60
25	10.50	11.51	13.11	14.61	16.47	34.38	37.66	40.66	44.34	46.97
26	11.13	12.19	13.84	15.38	17.29	35.57	38.89	41.94	45.66	48.33
27	11.79	12.87	14.57	16.15	18.11	36.74	40.12	43.21	46.99	49.69
28	12.44	13.55	15.30	16.92	18.94	37.92	41.34	44.47	48.30	51.04
29	13.09	14.24	16.04	17.70	19.77	39.09	42.56	45.74	49.61	52.38
30	13.77	14.94	16.78	18.49	20.60	40.26	43.78	46.99	50.91	53.71
35	17.16	18.49	20.56	22.46	24.79	46.06	49.81	53.22	57.36	60.31
40	20.67	22.14	24.42	26.51	29.06	51.80	55.75	59.34	63.71	66.80
45	24.28	25.88	28.36	30.61	33.36	57.50	61.65	65.41	69.98	73.20
50	27.96	29.68	32.35	34.76	37.69	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	74.39	79.08	83.30	88.40	91.96
70	43.25	45.42	48.75	51.74	55.33	83.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.91	82.36	118.49	124.14	129.36	133.82	140.19



TABLA G VALORES DE CUARTILES DE LA DISTRIBUCION F.

$$F(x) = \frac{1}{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2})} \int_0^x \frac{t^{\frac{\nu_1}{2}-1} (x-t)^{\frac{\nu_2}{2}-1} dt}{(1+t^2)^{\frac{\nu_1+\nu_2}{2}} (1+t^2+x^2-t^2)^{\frac{\nu_2}{2}}}$$



$\alpha = 0.05$

$\nu_1 = \text{grados de libertad del numerador}$

$\nu_2$	1	2	3	4	5	6	7	8	9	10
1	39.86	49.50	53.59	55.83	57.74	58.20	58.91	59.24	59.66	60.19
2	8.33	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.79	2.75	2.72	2.70
8	3.46	3.14	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.03	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.96
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.06	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.03	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.15	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
35	2.85	2.46	2.25	2.11	2.01	1.95	1.89	1.85	1.82	1.79
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66
200	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63
500	2.72	2.31	2.09	1.96	1.86	1.79	1.73	1.68	1.64	1.61
1000	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64	1.61

$\alpha = 0.05$

$\nu_1 = \text{grados de libertad del numerador}$

$\nu_2$	11	12	15	20	30	40	50	100	1000
1	69.87	68.71	61.22	47.74	62.56	62.56	62.55	62.69	63.00
2	9.40	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48
3	5.22	5.22	5.20	5.19	5.17	5.17	5.16	5.15	5.14
4	3.91	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.78
5	3.28	3.27	3.24	3.23	3.19	3.17	3.16	3.15	3.13
6	2.92	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.72
7	2.68	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.47
8	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.30
9	2.40	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.16
10	2.30	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.06
11	2.23	2.21	2.17	2.12	2.10	2.08	2.05	2.04	1.98
12	2.17	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.94
13	2.12	2.10	2.05	2.01	1.98	1.96	1.94	1.92	1.85
14	2.07	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.80
15	2.04	2.02	1.97	1.92	1.89	1.87	1.85	1.81	1.79
16	2.01	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.76
17	1.99	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.73
18	1.97	1.93	1.89	1.84	1.80	1.78	1.75	1.74	1.66
19	1.95	1.91	1.86	1.81	1.78	1.76	1.73	1.71	1.64
20	1.93	1.89	1.84	1.79	1.76	1.74	1.71	1.69	1.65
21	1.90	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.63
22	1.88	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.61
23	1.87	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.59
24	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.58
25	1.84	1.81	1.77	1.72	1.68	1.66	1.63	1.61	1.56
26	1.83	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.55
27	1.82	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.54
28	1.81	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.53
29	1.80	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.52
30	1.79	1.77	1.72	1.67	1.63	1.61	1.57	1.55	1.51
35	1.76	1.74	1.68	1.63	1.60	1.57	1.54	1.51	1.48
40	1.74	1.71	1.66	1.61	1.57	1.54	1.51	1.48	1.44
50	1.70	1.68	1.63	1.57	1.53	1.50	1.46	1.44	1.39
60	1.68	1.66	1.61	1.54	1.50	1.48	1.44	1.41	1.36
80	1.65	1.63	1.57	1.51	1.47	1.44	1.40	1.38	1.32
100	1.64	1.61	1.56	1.49	1.45	1.42	1.38	1.35	1.29
200	1.60	1.56	1.52	1.46	1.41	1.38	1.34	1.31	1.24
500	1.58	1.56	1.50	1.44	1.39	1.36	1.31	1.28	1.21
1000	1.58	1.55	1.49	1.43	1.38	1.35	1.30	1.27	1.20

$\alpha = 0.95$

$\nu_1 = \text{grados de libertad del numerador}$

$\nu_2$	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.34	241.88
2	18.31	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.97
5	6.61	5.70	5.41	5.19	5.05	4.93	4.84	4.82	4.77	4.73
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.52	3.66	3.26	3.03	2.88	2.77	2.68	2.62	2.57	2.54
16	4.46	3.60	3.20	2.97	2.82	2.71	2.62	2.56	2.51	2.48
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.01	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11
40	4.08	3.23	2.83	2.61	2.45	2.33	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88
500	3.86	3.01	2.62	2.39	2.23	2.11	2.03	1.96	1.90	1.85
1000	3.85	3.01	2.61	2.38						

TABLA C (continuación) Valores de cuantiles de la distribución F

1 - $\alpha$ = 0.99										
$\nu_2$	$\nu_1$ = grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	10
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.38	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.50	27.34	27.22
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.02	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.67	5.49	4.60	4.10	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
45	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
50	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.64
60	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.56
70	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
80	6.76	4.71	3.85	3.41	3.11	2.89	2.73	2.60	2.50	2.41
90	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36
100	6.64	4.61	3.80	3.34	3.04	2.82	2.66	2.53	2.41	2.34

1 - $\alpha$ = 0.99										
$\nu_2$	$\nu_1$ = grados de libertad del numerador									
	11	12	15	20	25	30	40	50	100	1000
2	99.41	99.42	99.43	99.45	99.46	99.46	99.47	99.48	99.49	99.51
3	27.12	27.01	26.85	26.67	26.58	26.50	26.41	26.35	26.24	26.14
4	14.45	14.37	14.19	14.02	13.91	13.84	13.75	13.69	13.58	13.48
5	9.96	9.89	9.72	9.55	9.45	9.38	9.30	9.24	9.13	9.03
6	7.79	7.72	7.56	7.40	7.29	7.23	7.15	7.09	6.99	6.89
7	6.94	6.87	6.71	6.56	6.46	6.40	6.32	6.26	6.15	6.06
8	5.73	5.67	5.52	5.36	5.26	5.20	5.12	5.07	4.96	4.87
9	5.18	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.41	4.32
10	4.77	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.01	3.92
11	4.46	4.40	4.25	4.10	4.00	3.94	3.86	3.81	3.71	3.61
12	4.22	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.47	3.37
13	4.02	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.27	3.18
14	3.86	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.11	3.02
15	3.73	3.67	3.52	3.37	3.28	3.21	3.13	3.08	2.98	2.89
16	3.62	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.86	2.78
17	3.52	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.76	2.68
18	3.43	3.37	3.23	3.08	2.98	2.92	2.84	2.79	2.68	2.59
19	3.36	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.60	2.50
20	3.29	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.54	2.45
21	3.24	3.17	3.03	2.88	2.78	2.72	2.64	2.58	2.48	2.37
22	3.18	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.42	2.32
23	3.14	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.37	2.27
24	3.09	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.33	2.22
25	3.06	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.29	2.18
26	3.02	2.96	2.81	2.66	2.56	2.50	2.42	2.36	2.25	2.14
27	2.99	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.22	2.11
28	2.96	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.19	2.08
29	2.94	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.16	2.05
30	2.91	2.84	2.70	2.55	2.45	2.39	2.30	2.24	2.13	2.02
35	2.83	2.74	2.60	2.44	2.35	2.28	2.19	2.14	2.02	1.90
40	2.75	2.66	2.52	2.37	2.27	2.20	2.11	2.06	1.94	1.82
50	2.62	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.82	1.70
60	2.56	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.75	1.61
80	2.48	2.42	2.27	2.12	2.01	1.94	1.85	1.79	1.65	1.51
100	2.43	2.37	2.22	2.07	1.97	1.89	1.80	1.74	1.60	1.45
200	2.34	2.27	2.13	1.97	1.87	1.79	1.69	1.63	1.48	1.30
500	2.29	2.22	2.07	1.92	1.81	1.74	1.63	1.57	1.41	1.20
1000	2.27	2.20	2.06	1.90	1.79	1.72	1.61	1.54	1.38	1.16

TABLA H : VALORES PARA LOS LIMITES DE TOLERANCIA BILATERALES CUANDO HAGA UN MUESTREO DE DIST. NORMALES.

n	γ = 0.75					γ = 0.90				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
6	1.704	2.427	2.869	3.779	4.632	2.196	3.131	3.733	4.870	6.188
7	1.624	2.318	2.757	3.611	4.393	2.034	2.902	3.432	4.521	5.750
8	1.568	2.238	2.663	3.491	4.244	1.921	2.743	3.264	4.278	5.466
9	1.525	2.178	2.591	3.400	4.130	1.839	2.626	3.125	4.098	5.220
10	1.492	2.131	2.537	3.328	4.041	1.775	2.535	3.018	3.959	5.048
11	1.465	2.093	2.493	3.271	3.969	1.724	2.463	2.933	3.849	4.908
12	1.443	2.062	2.456	3.223	3.910	1.683	2.404	2.863	3.758	4.792
13	1.425	2.037	2.424	3.183	3.859	1.648	2.355	2.805	3.682	4.697
14	1.409	2.013	2.398	3.148	3.816	1.619	2.314	2.756	3.618	4.615
15	1.395	1.994	2.375	3.118	3.779	1.594	2.278	2.713	3.562	4.545
16	1.383	1.977	2.355	3.092	3.746	1.572	2.246	2.676	3.514	4.484
17	1.372	1.962	2.337	3.069	3.717	1.552	2.219	2.643	3.471	4.430
18	1.361	1.948	2.321	3.048	3.691	1.535	2.194	2.614	3.433	4.382
19	1.351	1.936	2.307	3.030	3.667	1.520	2.172	2.588	3.399	4.339
20	1.347	1.927	2.294	3.013	3.646	1.506	2.152	2.564	3.368	4.300
21	1.340	1.915	2.282	2.998	3.627	1.493	2.135	2.543	3.340	4.264
22	1.334	1.906	2.271	2.984	3.609	1.482	2.118	2.524	3.315	4.232
23	1.328	1.898	2.261	2.971	3.593	1.471	2.103	2.506	3.292	4.203
24	1.322	1.891	2.252	2.959	3.578	1.462	2.089	2.489	3.270	4.176
25	1.317	1.885	2.244	2.948	3.564	1.453	2.077	2.474	3.251	4.151
26	1.313	1.877	2.236	2.938	3.551	1.444	2.065	2.460	3.232	4.127
27	1.309	1.871	2.229	2.929	3.540	1.437	2.054	2.447	3.215	4.106
28	1.305	1.865	2.222	2.920	3.530	1.430	2.044	2.435	3.199	4.085
29	1.301	1.860	2.216	2.911	3.521	1.423	2.034	2.424	3.184	4.066
30	1.297	1.855	2.210	2.904	3.513	1.417	2.025	2.413	3.170	4.049
31	1.294	1.850	2.204	2.896	3.509	1.411	2.017	2.403	3.157	4.032
32	1.291	1.846	2.199	2.890	3.500	1.405	2.009	2.393	3.145	4.016
33	1.288	1.842	2.194	2.883	3.492	1.400	2.001	2.385	3.133	4.001
34	1.285	1.838	2.189	2.877	3.484	1.395	1.994	2.376	3.122	3.987
35	1.283	1.834	2.185	2.871	3.477	1.390	1.988	2.368	3.112	3.974
36	1.280	1.830	2.181	2.866	3.469	1.386	1.981	2.361	3.102	3.961
37	1.278	1.827	2.177	2.860	3.463	1.381	1.975	2.355	3.092	3.949
38	1.275	1.824	2.173	2.855	3.457	1.377	1.969	2.346	3.083	3.938
39	1.273	1.821	2.169	2.850	3.451	1.374	1.964	2.340	3.075	3.927
40	1.271	1.818	2.166	2.846	3.445	1.370	1.959	2.334	3.066	3.917
41	1.269	1.815	2.162	2.841	3.439	1.366	1.954	2.328	3.059	3.907
42	1.267	1.812	2.159	2.837	3.434	1.363	1.949	2.322	3.051	3.897
43	1.266	1.810	2.156	2.833	3.429	1.360	1.944	2.316	3.044	3.888
44	1.264	1.807	2.153	2.829	3.424	1.357	1.940	2.311	3.037	3.879
45	1.262	1.805	2.150	2.826	3.419	1.354	1.935	2.306	3.030	3.871
46	1.261	1.802	2.148	2.822	3.415	1.351	1.931	2.301	3.024	3.863
47	1.259	1.800	2.145	2.819	3.411	1.348	1.927	2.297	3.018	3.855
48	1.258	1.799	2.143	2.815	3.406	1.345	1.924	2.292	3.012	3.847
49	1.256	1.796	2.140	2.812	3.402	1.343	1.920	2.288	3.006	3.840
50	1.255	1.794	2.138	2.809	3.398	1.340	1.916	2.284	3.001	3.833

n	γ = 0.95				γ = 0.99					
	0.75	0.90	0.95	0.99	0.75	0.90	0.95	0.99		
6	2.604	3.732	4.414	5.725	7.337	3.743	5.337	6.345	8.301	10.548
7	2.361	3.369	4.007	5.248	6.676	3.233	4.613	5.488	7.187	9.142
8	2.197	3.136	3.732	4.891	6.226	2.905	4.147	4.936	6.468	8.234
9	2.078	2.967	3.532	4.631	5.895	2.677	3.822	4.550	5.966	7.601
10	1.987	2.839	3.379	4.433	5.649	2.508	3.582	4.265	5.594	7.129
11	1.916	2.737	3.259	4.277	5.452	2.378	3.397	4.045	5.368	6.766
12	1.858	2.655	3.162	4.150	5.291	2.274	3.250	3.870	5.079	6.477
13	1.810	2.587	3.081	4.044	5.158	2.193	3.130	3.727	4.893	6.244
14	1.770	2.529	3.012	3.955	5.045	2.120	3.029	3.608	4.737	6.043
15	1.735	2.480	2.954	3.878	4.949	2.061	2.935	3.507	4.603	5.876
16	1.705	2.437	2.903	3.812	4.865	2.009	2.852	3.421	4.492	5.732
17	1.679	2.400	2.858	3.754	4.791	1.965	2.783	3.345	4.393	5.607
18	1.655	2.366	2.819	3.702	4.725	1.926	2.733	3.279	4.307	5.497
19	1.635	2.337	2.784	3.656	4.667	1.891	2.701	3.221	4.230	5.399
20	1.616	2.310	2.752	3.615	4.614	1.860	2.659	3.168	4.161	5.312
21	1.599	2.286	2.723	3.577	4.567	1.833	2.620	3.121	4.101	5.234
22	1.584	2.264	2.697	3.543	4.523	1.808	2.584	3.078	4.044	5.163
23	1.570	2.244	2.673	3.512	4.484	1.785	2.551	3.040	3.991	5.098
24	1.557	2.225	2.651	3.483	4.447	1.764	2.522	3.004	3.947	5.039
25	1.545	2.208	2.631	3.457	4.413	1.745	2.494	2.972	3.904	4.985
26	1.534	2.193	2.612	3.432	4.382	1.727	2.469	2.941	3.865	4.935
27	1.523	2.178	2.595	3.408	4.353	1.711	2.446	2.914	3.828	4.888
28	1.514	2.164	2.579	3.384	4.326	1.695	2.424	2.888	3.794	4.845
29	1.505	2.152	2.564	3.368	4.301	1.681	2.404	2.864	3.763	4.805
30	1.497	2.140	2.549	3.350	4.278	1.668	2.385	2.841	3.733	4.768
31	1.489	2.129	2.536	3.332	4.256	1.656	2.367	2.820	3.706	4.732
32	1.481	2.118	2.524	3.316	4.235	1.644	2.351	2.801	3.680	4.699
33	1.475	2.108	2.512	3.300	4.215	1.633	2.335	2.782	3.655	4.668
34	1.468	2.099	2.501	3.286	4.197	1.623	2.320	2.764	3.632	4.639
35	1.462	2.090	2.490	3.272	4.179	1.613	2.306	2.748	3.611	4.611
36	1.455	2.081	2.479	3.258	4.161	1.604	2.293	2.732	3.590	4.585
37	1.450	2.073	2.470	3.246	4.146	1.595	2.281	2.717	3.571	4.560
38	1.446	2.066	2.464	3.237	4.134	1.587	2.269	2.703	3.552	4.537
39	1.441	2.060	2.458	3.226	4.120	1.579	2.257	2.690	3.534	4.514
40	1.435	2.052	2.445	3.213	4.104	1.571	2.247	2.677	3.518	4.493
41	1.430	2.045	2.437	3.202	4.090	1.564	2.236	2.665	3.502	4.472
42	1.426	2.039	2.429	3.192	4.077	1.557	2.227	2.653	3.486	4.453
43	1.422	2.033	2.422	3.183	4.065	1.551	2.217	2.642	3.472	4.434
44	1.418	2.027	2.415	3.175	4.053	1.545	2.208	2.631	3.458	4.416
45	1.414	2.021	2.408	3.165	4.042	1.539	2.201	2.621	3.444	4.399
46	1.410	2.016	2.402	3.156	4.031	1.533	2.192	2.611	3.431	4.383
47	1.406	2.011	2.396	3.148	4.021	1.527	2.184	2.602	3.419	4.367
48	1.403	2.006	2.390	3.140	4.011	1.522	2.176	2.593	3.407	4.352
49	1.400	2.001	2.384	3.133	4.002	1.517	2.168	2.584	3.396	4.337
50	1.396	1.999	2.379	3.126	3.993	1.512	2.162	2.576	3.385	4.323

Source: C. Freudenthal, M. W. Hooley, and W. A. Wallis, *Techniques of statistical analysis*, McGraw-Hill, New York, 1947. Publicado con permiso.



TABLA I VALORES PARA LOS LIMITES DE TOLERANCIA UNILATERALES CUANDO HAGA UN MUESTREO DE DIST. NORMALES.

n/d	$\gamma = 0.75$					$\gamma = 0.90$				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
6	1.087	1.860	2.336	3.243	4.273	1.540	2.494	3.091	4.242	5.356
7	1.043	1.791	2.250	3.126	4.188	1.435	2.333	2.894	3.972	5.201
8	1.011	1.742	2.190	3.042	4.098	1.360	2.219	2.755	3.783	4.955
9	0.984	1.707	2.141	2.977	3.974	1.302	2.131	2.649	3.641	4.772
10	0.964	1.671	2.103	2.927	3.838	1.257	2.063	2.568	3.532	4.629
11	0.947	1.646	2.073	2.885	3.804	1.219	2.012	2.503	3.444	4.515
12	0.933	1.624	2.048	2.851	3.760	1.188	1.966	2.448	3.371	4.420
13	0.919	1.606	2.026	2.822	3.722	1.162	1.928	2.403	3.310	4.341
14	0.909	1.591	2.007	2.796	3.690	1.139	1.893	2.363	3.257	4.274
15	0.899	1.577	1.991	2.776	3.661	1.119	1.866	2.329	3.212	4.215
16	0.891	1.566	1.977	2.756	3.637	1.101	1.842	2.299	3.172	4.164
17	0.883	1.554	1.964	2.739	3.615	1.085	1.820	2.272	3.136	4.118
18	0.876	1.544	1.951	2.724	3.595	1.071	1.800	2.249	3.106	4.078
19	0.870	1.536	1.942	2.710	3.577	1.058	1.781	2.228	3.074	4.041
20	0.865	1.528	1.933	2.697	3.561	1.046	1.763	2.208	3.052	4.009
21	0.859	1.520	1.923	2.686	3.545	1.035	1.756	2.190	3.035	3.979
22	0.854	1.514	1.916	2.675	3.532	1.025	1.750	2.174	3.020	3.952
23	0.849	1.508	1.910	2.665	3.520	1.016	1.744	2.159	2.987	3.923
24	0.845	1.502	1.901	2.656	3.509	1.007	1.738	2.145	2.969	3.900
25	0.842	1.496	1.893	2.647	3.497	0.999	1.702	2.132	2.952	3.882
30	0.825	1.475	1.869	2.613	3.454	0.966	1.657	2.080	2.884	3.794
35	0.812	1.458	1.849	2.588	3.421	0.942	1.623	2.041	2.833	3.730
40	0.803	1.443	1.834	2.568	3.395	0.923	1.598	2.010	2.793	3.679
45	0.795	1.435	1.821	2.552	3.375	0.908	1.577	1.966	2.762	3.638
50	0.788	1.426	1.811	2.538	3.358	0.894	1.560	1.967	2.735	3.604

n/d	$\gamma = 0.95$					$\gamma = 0.99$				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
6	1.895	3.006	3.707	5.062	6.612	2.849	4.408	5.499	7.334	9.340
7	1.712	2.755	3.399	4.641	6.061	2.490	3.878	4.730	6.411	8.348
8	1.617	2.592	3.188	4.353	5.656	2.252	3.496	4.287	5.811	7.566
9	1.532	2.454	3.031	4.143	5.414	2.085	3.242	3.971	5.389	7.014
10	1.465	2.335	2.911	3.961	5.203	1.954	3.048	3.739	5.075	6.603
11	1.411	2.235	2.815	3.832	5.036	1.854	2.897	3.557	4.828	6.284
12	1.366	2.150	2.736	3.747	4.900	1.771	2.773	3.410	4.615	6.032
13	1.329	2.155	2.670	3.659	4.787	1.702	2.677	3.290	4.472	5.826
14	1.296	2.108	2.614	3.585	4.690	1.645	2.592	3.189	4.336	5.651
15	1.268	2.100	2.566	3.520	4.607	1.598	2.524	3.102	4.274	5.507
16	1.242	2.035	2.521	3.463	4.534	1.553	2.458	3.028	4.124	5.374
17	1.220	2.011	2.486	3.415	4.471	1.514	2.405	2.962	4.038	5.260
18	1.201	1.974	2.453	3.350	4.415	1.481	2.357	2.906	3.961	5.167
19	1.184	1.949	2.423	3.311	4.364	1.450	2.315	2.851	3.893	5.078
20	1.167	1.926	2.396	3.295	4.319	1.424	2.275	2.807	3.832	5.013
21	1.152	1.905	2.371	3.262	4.276	1.397	2.241	2.768	3.776	4.932
22	1.138	1.887	2.350	3.233	4.238	1.376	2.208	2.729	3.727	4.866
23	1.126	1.869	2.329	3.206	4.204	1.355	2.179	2.693	3.680	4.806
24	1.114	1.851	2.309	3.181	4.171	1.336	2.154	2.663	3.638	4.755
25	1.103	1.838	2.292	3.158	4.143	1.318	2.129	2.642	3.601	4.706
30	1.059	1.778	2.229	3.064	4.022	1.249	2.029	2.516	3.440	4.508
35	1.025	1.732	2.166	2.994	3.934	1.195	1.957	2.431	3.334	4.364
40	0.999	1.697	2.120	2.941	3.866	1.151	1.902	2.365	3.250	4.255
45	0.978	1.669	2.092	2.897	3.811	1.122	1.857	2.313	3.181	4.188
50	0.961	1.646	2.065	2.863	3.766	1.096	1.821	2.290	3.124	4.136

Source: G. J. Lehmann, Table for one-sided statistical tolerance limits, Industrial Quality Control, XIV, 1958, p. 7. Reprinted with permission.

TABLA J DISTRIBUCION DE POISSON.

TABLA II Función de distribución de Poisson  $F(x;\lambda) = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$

$\lambda \backslash x$	0	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	0.999	1.000						
0.20	0.819	0.982	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.994	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.55	0.577	0.894	0.982	0.998	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.65	0.522	0.861	0.972	0.996	0.999	1.000				
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.75	0.472	0.827	0.959	0.993	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.85	0.427	0.791	0.945	0.989	0.998	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
0.95	0.387	0.754	0.929	0.984	0.997	1.000				
1.00	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.998	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.833	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.995	0.999	1.000	
2.2	0.111	0.355	0.623	0.819	0.928	0.975	0.993	0.998	1.000	
2.4	0.091	0.308	0.570	0.779	0.904	0.964	0.988	0.997	0.999	1.000
2.6	0.074	0.267	0.518	0.736	0.877	0.951	0.983	0.995	0.999	1.000
2.8	0.061	0.231	0.469	0.692	0.848	0.935	0.976	0.992	0.998	0.999
3.0	0.050	0.199	0.423	0.647	0.815	0.916	0.966	0.988	0.996	0.999

$\lambda \backslash x$	0	1	2	3	4	5	6	7	8	9
3.2	0.041	0.173	0.380	0.603	0.781	0.895	0.955	0.983	0.994	0.998
3.4	0.033	0.147	0.340	0.558	0.744	0.871	0.942	0.977	0.992	0.997
3.6	0.027	0.126	0.303	0.515	0.706	0.844	0.927	0.969	0.986	0.996
3.8	0.022	0.107	0.269	0.473	0.668	0.816	0.909	0.960	0.984	0.994
4.0	0.018	0.092	0.238	0.433	0.629	0.785	0.889	0.949	0.979	0.992
4.2	0.015	0.078	0.210	0.395	0.590	0.753	0.867	0.936	0.972	0.989
4.4	0.012	0.066	0.185	0.359	0.551	0.720	0.844	0.921	0.964	0.985
4.6	0.010	0.056	0.163	0.326	0.513	0.686	0.818	0.905	0.955	0.980
4.8	0.008	0.048	0.143	0.294	0.476	0.651	0.791	0.887	0.944	0.975
5.0	0.007	0.040	0.125	0.265	0.440	0.616	0.762	0.867	0.932	0.968
5.2	0.006	0.034	0.109	0.238	0.406	0.581	0.732	0.845	0.918	0.960
5.4	0.005	0.029	0.095	0.213	0.373	0.546	0.702	0.822	0.903	0.951
5.6	0.004	0.024	0.082	0.191	0.342	0.512	0.670	0.797	0.886	0.941
5.8	0.003	0.021	0.072	0.170	0.313	0.478	0.638	0.771	0.867	0.925
6.0	0.002	0.017	0.062	0.151	0.285	0.446	0.606	0.744	0.847	0.916
10	11	12	13	14	15	16				
2.8	1.000									
3.0	1.000									
3.2	1.000									
3.4	0.999	1.000								
3.6	0.999	1.000								
3.8	0.998	0.999	1.000							
4.0	0.997	0.999	1.000							
4.2	0.996	0.999	1.000							
4.4	0.994	0.998	0.999	1.000						
4.6	0.992	0.997	0.999	1.000						
4.8	0.990	0.996	0.999	1.000						
5.0	0.986	0.995	0.998	0.999	1.000					
5.2	0.982	0.993	0.997	0.999	1.000					
5.4	0.977	0.990	0.996	0.999	1.000					
5.6	0.972	0.988	0.995	0.998	0.999	1.000				
5.8	0.965	0.984	0.993	0.997	0.999	1.000				
6.0	0.957	0.980	0.991	0.996	0.999	0.999	1.000			

X	0	1	2	3	4	5	6	7	8	9
6.2	0.002	0.015	0.054	0.134	0.259	0.414	0.574	0.716	0.826	0.902
6.4	0.002	0.012	0.046	0.109	0.235	0.384	0.542	0.687	0.803	0.886
6.6	0.001	0.010	0.040	0.095	0.213	0.355	0.511	0.658	0.780	0.869
6.8	0.001	0.009	0.034	0.093	0.192	0.327	0.480	0.628	0.755	0.850
7.0	0.001	0.007	0.030	0.082	0.173	0.301	0.450	0.599	0.729	0.830
7.2	0.001	0.006	0.025	0.072	0.156	0.276	0.420	0.569	0.703	0.810
7.4	0.001	0.005	0.022	0.063	0.140	0.253	0.392	0.539	0.676	0.788
7.6	0.001	0.004	0.019	0.055	0.125	0.231	0.365	0.510	0.648	0.765
7.8	0.000	0.004	0.016	0.048	0.112	0.210	0.338	0.481	0.620	0.741
8.0	0.000	0.003	0.014	0.042	0.100	0.191	0.313	0.453	0.593	0.717
8.5	0.000	0.002	0.009	0.030	0.074	0.150	0.256	0.386	0.523	0.653
9.0	0.000	0.001	0.006	0.021	0.055	0.116	0.207	0.324	0.456	0.587
9.5	0.000	0.001	0.004	0.015	0.040	0.089	0.165	0.269	0.392	0.522
10.0	0.000	0.000	0.003	0.010	0.029	0.067	0.130	0.220	0.333	0.458
	10	11	12	13	14	15	16	17	18	19
5.2	0.949	0.975	0.989	0.993	0.998	0.999	1.000			
5.4	0.939	0.969	0.986	0.994	0.997	0.999	1.000			
6.6	0.927	0.963	0.982	0.992	0.997	0.999	0.999	1.000		
6.8	0.915	0.955	0.978	0.990	0.996	0.998	0.999	1.000		
7.0	0.901	0.947	0.973	0.987	0.994	0.998	0.999	1.000		
7.2	0.887	0.937	0.967	0.984	0.993	0.997	0.999	0.999	1.000	
7.4	0.871	0.926	0.961	0.980	0.991	0.996	0.998	0.999	1.000	
7.6	0.854	0.915	0.954	0.976	0.989	0.995	0.998	0.999	1.000	
7.8	0.835	0.902	0.943	0.971	0.986	0.993	0.997	0.999	1.000	
8.0	0.816	0.888	0.936	0.966	0.983	0.992	0.996	0.998	0.999	1.000
8.5	0.763	0.849	0.909	0.949	0.973	0.986	0.993	0.997	0.999	0.999
9.0	0.706	0.803	0.876	0.926	0.959	0.978	0.989	0.995	0.998	0.999
9.5	0.645	0.752	0.836	0.898	0.940	0.967	0.982	0.991	0.996	0.998
10.0	0.583	0.697	0.792	0.864	0.917	0.951	0.973	0.986	0.993	0.997
	20	21	22	23	24	25	26	27	28	29
8.5	1.000									
9.0	1.000									
9.5	0.999	1.000								
10.0	0.998	0.999	1.000							
10.5	0.997	0.999	0.999	1.000						
11.0	0.995	0.998	0.999	1.000						
11.5	0.992	0.996	0.998	0.999	1.000					
12.0	0.988	0.994	0.997	0.999	0.999	1.000				
12.5	0.983	0.991	0.995	0.998	0.999	0.999	1.000			

X	0	1	2	3	4	5	6	7	8	9
10.5	0.000	0.000	0.002	0.007	0.021	0.050	0.102	0.179	0.279	0.397
11.0	0.000	0.000	0.001	0.005	0.015	0.038	0.079	0.143	0.232	0.341
11.5	0.000	0.000	0.001	0.003	0.011	0.028	0.060	0.114	0.191	0.285
12.0	0.000	0.000	0.001	0.002	0.008	0.020	0.046	0.090	0.155	0.242
12.5	0.000	0.000	0.000	0.002	0.005	0.015	0.033	0.070	0.125	0.201
13.0	0.000	0.000	0.000	0.001	0.004	0.011	0.026	0.054	0.100	0.166
13.5	0.000	0.000	0.000	0.001	0.003	0.008	0.019	0.041	0.079	0.135
14.0	0.000	0.000	0.000	0.000	0.002	0.006	0.014	0.032	0.062	0.105
14.5	0.000	0.000	0.000	0.000	0.001	0.004	0.010	0.024	0.048	0.088
15.0	0.000	0.000	0.000	0.000	0.001	0.003	0.008	0.018	0.037	0.070
	10	11	12	13	14	15	16	17	18	19
10.5	0.521	0.639	0.742	0.825	0.888	0.932	0.960	0.978	0.988	0.994
11.0	0.460	0.579	0.689	0.781	0.854	0.907	0.944	0.968	0.982	0.991
11.5	0.402	0.520	0.633	0.733	0.815	0.878	0.924	0.954	0.974	0.985
12.0	0.347	0.462	0.576	0.682	0.772	0.844	0.899	0.937	0.963	0.979
12.5	0.297	0.406	0.519	0.628	0.725	0.806	0.869	0.916	0.948	0.965
13.0	0.252	0.353	0.463	0.573	0.675	0.764	0.835	0.890	0.930	0.957
13.5	0.211	0.304	0.409	0.518	0.623	0.718	0.798	0.861	0.908	0.942
14.0	0.176	0.260	0.358	0.464	0.570	0.669	0.756	0.827	0.883	0.922
14.5	0.145	0.220	0.311	0.413	0.518	0.619	0.711	0.790	0.853	0.901
15.0	0.118	0.185	0.268	0.363	0.466	0.568	0.664	0.749	0.819	0.875
	20	21	22	23	24	25	26	27	28	29
13.0	0.975	0.986	0.992	0.996	0.998	0.999	1.000			
13.5	0.965	0.980	0.989	0.994	0.997	0.998	0.999	1.000		
14.0	0.952	0.971	0.983	0.991	0.995	0.997	0.999	0.999	1.000	
14.5	0.936	0.960	0.976	0.986	0.992	0.996	0.998	0.999	0.999	1.000
15.0	0.917	0.947	0.967	0.981	0.989	0.994	0.997	0.998	0.999	1.000
	4	5	6	7	8	9	10	11	12	13
16.0	0.000	0.001	0.004	0.010	0.022	0.043	0.077	0.127	0.193	0.275
17.0	0.000	0.001	0.002	0.005	0.013	0.026	0.049	0.085	0.135	0.201
18.0	0.000	0.000	0.001	0.003	0.007	0.015	0.030	0.053	0.092	0.143
19.0	0.000	0.000	0.001	0.002	0.004	0.009	0.018	0.035	0.061	0.098
20.0	0.000	0.000	0.000	0.001	0.002	0.005	0.011	0.021	0.039	0.066
21.0	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.013	0.025	0.043
22.0	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.008	0.015	0.028
23.0	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.009	0.017
24.0	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.011
25.0	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.006

$\lambda \backslash x$	14	15	16	17	18	19	20	21	22	23
16.0	0.368	0.467	0.566	0.659	0.742	0.812	0.868	0.911	0.942	0.963
17.0	0.281	0.371	0.468	0.564	0.655	0.736	0.805	0.861	0.905	0.937
18.0	0.208	0.287	0.375	0.469	0.562	0.651	0.731	0.799	0.855	0.899
19.0	0.150	0.215	0.292	0.378	0.469	0.561	0.647	0.725	0.793	0.849
20.0	0.105	0.157	0.221	0.297	0.381	0.470	0.559	0.644	0.721	0.787
21.0	0.072	0.111	0.163	0.227	0.302	0.384	0.471	0.558	0.640	0.716
22.0	0.048	0.077	0.117	0.169	0.232	0.306	0.387	0.472	0.556	0.637
23.0	0.031	0.052	0.082	0.123	0.175	0.238	0.310	0.389	0.472	0.555
24.0	0.020	0.034	0.056	0.087	0.128	0.180	0.243	0.314	0.392	0.473
25.0	0.012	0.022	0.038	0.060	0.092	0.134	0.185	0.247	0.318	0.394
	24	25	26	27	28	29	30	31	32	33
16.0	0.978	0.987	0.993	0.996	0.998	0.999	0.999	1.000		
17.0	0.959	0.975	0.985	0.991	0.995	0.997	0.999	0.999	1.000	
18.0	0.932	0.955	0.972	0.983	0.990	0.994	0.997	0.998	0.999	1.000
19.0	0.893	0.927	0.951	0.969	0.980	0.988	0.993	0.996	0.998	0.999
20.0	0.843	0.888	0.922	0.948	0.966	0.978	0.987	0.992	0.995	0.997
21.0	0.782	0.838	0.883	0.917	0.944	0.963	0.976	0.985	0.991	0.994
22.0	0.712	0.777	0.832	0.877	0.913	0.940	0.959	0.973	0.983	0.989
23.0	0.635	0.708	0.772	0.827	0.873	0.908	0.936	0.956	0.971	0.981
24.0	0.554	0.632	0.704	0.768	0.823	0.868	0.904	0.932	0.953	0.969
25.0	0.473	0.553	0.629	0.700	0.763	0.818	0.863	0.900	0.929	0.950
	34	35	36	37	38	39	40	41	42	43
19.0	0.999	1.000								
20.0	0.999	0.999	1.000							
21.0	0.997	0.998	0.999	0.999	1.000					
22.0	0.994	0.996	0.998	0.999	0.999	1.000				
23.0	0.988	0.993	0.996	0.997	0.999	0.999	1.000			
24.0	0.979	0.987	0.992	0.995	0.997	0.998	0.999	0.999	1.000	
25.0	0.966	0.978	0.985	0.991	0.994	0.997	0.998	0.999	0.999	1.000

NUMEROS ALZATORIOS\*

Table with 10 columns of numerical data under the heading 'NUMEROS ALZATORIOS\*'. The table contains multiple rows of numbers, some with leading zeros.

NUMEROS ALZATORIOS\*

Table with 10 columns of numerical data under the heading 'NUMEROS ALZATORIOS\*'. The table contains multiple rows of numbers, some with leading zeros.

\* Tomados por los miembros de la Union Republicana para el 100.000 Aniversario de la Independencia de la Republica de Chile.

TABLA L TABLA DE NUMEROS ALZATORIOS

NUMEROS ALATORIOS

98297	33613	33409	33673	91954	24531	40222	51574	31670	41659
93394	73381	21912	24973	24772	31644	41274	10191	8124	09291
78643	24544	61944	101117	40441	37274	44204	41219	24705	41414
31104	70295	42174	101117	40441	37274	44204	41219	24705	41414
04437	10065	33741	79416	72457	35952	46956	09053	70483	19179
30698	30818	90051	76320	43453	28141	40959	70495	93247	04114
27142	48154	32174	41897	42174	16795	84312	91520	44743	24644
36725	60728	70544	41104	44204	51415	33314	21870	50445	01111
02560	31679	29040	32297	36314	17174	04104	02711	03949	39959
16244	46697	08331	30201	36303	09171	33993	60232	31105	30689
64481	04102	20770	44201	04498	42575	90118	84406	81403	21154
74123	41174	49300	57972	29254	94364	96398	00419	60448	97335
84722	77241	49434	41712	21401	09465	27159	21218	02940	10794
59013	71612	43040	39180	99773	72323	46314	39083	60213	27664
32347	04440	45678	34976	92012	16396	28394	13493	40734	44760
09037	07404	40439	73320	47155	71128	24498	67455	04320	42004
77744	39416	37111	01918	71444	42545	71115	68620	87339	77977
20511	34541	42106	60931	42718	40958	18290	40334	10741	89447
34074	12523	68723	13934	62818	64137	34409	99761	70109	04355
60679	43862	42675	03653	21060	31096	71332	28930	44207	08334
49416	38110	63718	87515	39290	76554	34110	21490	10943	57330
65714	34447	11740	92444	41334	35835	49882	36431	08091	01981
17114	77711	63131	42919	40910	29184	74434	58007	34873	23916
00111	13129	09448	07544	81439	08075	34982	05971	27365	60177
68274	79013	82773	83412	97328	81003	63918	43310	78347	29316
44714	91496	45448	92281	73854	47452	52141	41582	81549	47434
47974	11496	37048	40753	44754	50711	04241	61454	34718	94157
18273	46797	33320	41155	44190	94860	42074	11178	27947	12644
56003	24170	79649	23431	91012	41102	53498	78494	54902	21363
94993	30681	08410	07244	42067	76669	19656	44064	67141	10170
31481	31607	40016	74472	91244	30241	14039	94481	33767	71077
82947	34120	04432	52545	47114	55445	04436	61791	67304	57204
09441	40182	40031	60204	04446	72254	47115	87507	03112	89189
75791	34746	70475	09401	18178	12075	36323	36843	07037	74129
27633	13175	30711	45814	92222	16704	00197	57247	33224	40274
90092	40991	12771	71733	45214	33855	32939	30723	88987	69761
07204	93333	25112	29410	30177	44936	18439	02235	67404	72930
88259	97234	02771	21341	64657	42113	12743	01023	24224	24912
30497	91207	72900	15699	58053	38661	27077	34444	84081	48400
09774	18075	45852	54468	43743	12030	78412	74456	93012	10684
85110	01985	24129	40714	42419	91907	25404	37097	93366	24043
09740	12148	05001	44923	66128	54146	67213	32234	48381	89442
01314	11447	15375	45831	84641	40702	47362	93444	42047	18044
11274	14774	48768	37195	36287	01046	27336	36657	91913	70040
71036	46762	80221	35863	27304	21121	94711	11607	60882	44359
34208	05174	40104	43178	97247	54875	26599	67452	14457	80354
47132	62919	42198	29445	60430	76219	02772	44651	40449	89213
56483	91102	43916	43441	45441	10106	12713	37248	81333	23440
17803	14744	10410	27494	94014	20524	41455	44219	70406	28474
55045	17219	61737	39480	74459	12626	60661	33333	70062	14789
09223	33423	99442	49791	81318	33425	74212	77062	01295	11051
07202	46746	21444	54425	17111	49944	44444	56444	33445	24978
84923	46742	94750	11071	54734	79211	04951	24113	78327	1326
44323	14446	17446	41178	27174	42719	60101	40331	41106	83441
80874	33313	97323	43064	66221	44447	37491	07947	12933	44191
44274	91979	26744	40741	36743	42290	21111	81016	41745	04109
30134	74121	75744	42104	43121	44444	44444	06104	47102	14679
56483	14252	24444	44712	14711	31154	47741	34743	82442	00121
31744	81212	20475	41144	44455	20723	01181	13674	40418	81429
81727	23512	93469	36825	41842	91083	68323	44955	34966	32327

NUMEROS ALATORIOS

30807	33338	92514	87198	21221	84500	32117	02771	43886	48234
73104	77447	78792	22718	45444	81200	33402	78713	01426	02310
06149	42714	34615	01224	23073	22792	21102	27144	33710	33617
22984	34834	64187	34282	47117	50454	48472	71499	04714	24671
13173	64761	20944	11924	18449	20211	77112	69641	96747	67047
73273	34104	55245	13441	22705	00714	27045	40148	01446	03177
98419	65714	44454	04412	44292	74244	47125	34414	03140	24359
91108	43542	18443	14870	49449	39477	77041	12654	54752	14941
51443	01542	17441	12144	92441	30444	42106	61216	34447	70184
02644	23364	63964	67447	92573	84177	83044	40434	20413	84272
38606	83374	74013	42141	09740	04274	10355	39053	31312	14476
43164	28908	84474	14472	83119	12114	12114	12114	63144	92740
80933	32930	45214	34474	05041	38471	27174	56473	44474	65823
47290	96766	47700	19114	47443	41444	11111	42474	54774	40494
46379	17014	23342	13114	40442	13114	24712	71113	43369	44776
46353	39333	44675	30143	24223	15474	11111	42474	14047	67003
30019	09517	32215	09144	20450	48170	47113	43270	43225	93150
87200	05722	42195	91474	39272	46442	77153	27713	03131	51692
58118	32942	14907	68214	34491	47244	24914	49744	32907	16673
97210	31027	70166	43232	98802	70713	30218	33386	18909	79638







Función de distribución binomial

$$B(x; n, p) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$$

	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 2	0	0.9025	0.8100	0.7225	0.6400	0.5675	0.4900	0.4225	0.3500	0.2075	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9325	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1684	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9968	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 4	0	0.8145	0.6581	0.5270	0.4096	0.3164	0.2401	0.1785	0.1290	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9378
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 5	0	0.7738	0.5808	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4294	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8985	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8175
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9978	0.9947	0.9898	0.9815	0.9687
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8957	0.7785	0.6354	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9841	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4315	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9325	0.8825	0.8208	0.7447	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0162	0.0078
	1	0.9556	0.8503	0.7168	0.5787	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9982	0.9743	0.9282	0.8520	0.7584	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9687	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9963	0.9912	0.9844	0.9737	0.9571	0.9347	0.9074
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9971	0.9948	0.9912	0.9863	0.9798
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9927
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0038
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9947	0.9819	0.9548	0.9066	0.8385	0.7518	0.6478	0.5364	0.4201	0.3045
	3	0.9998	0.9950	0.9788	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8283	0.7398	0.6387
	5	1.0000	1.0000	0.9998	0.9988	0.9968	0.9927	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9918	0.9810	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 9	0	0.6307	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9228	0.7748	0.5995	0.4262	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0858
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4828	0.3614	0.2520
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9806
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
n = 10	0	0.5987	0.3487	0.1968	0.1074	0.0563	0.0282	0.0125	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0880	0.0484	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5296	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3873	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7315	0.6331	0.5044	0.3720
	5	1.0000	0.9999	0.9988	0.9936	0.9803	0.9527	0.9051	0.8328	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9462	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
n = 11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0129	0.0059
	2	0.9848	0.9104	0.7789	0.6174	0.4582	0.3127	0.2001	0.1189	0.0537	0.0237
	3	0.9994	0.9715	0.9306	0.8389	0.7133	0.5698	0.4256	0.2963	0.1911	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6783	0.5328	0.3971	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7533	0.6331	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8282	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9876	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
n = 12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8818	0.6890	0.4475	0.2749	0.1504	0.0860	0.0474	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0804	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4825	0.3487	0.2283	0.1345	0.0730
	4	0.9998	0.9987	0.9781	0.9274	0.8474	0.7237	0.5833	0.4382	0.3064	0.1938
	5	1.0000	0.9995	0.9964	0.9806	0.9456	0.8822	0.7873	0.6652	0.5289	0.3872
	6	1.0000	0.9999	0.9993	0.9861	0.9585	0.9144	0.8154	0.6818	0.5393	0.3828
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8963	0.8262
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9963	0.9944	0.9847	0.9644
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9984
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
n = 13	0	0.5133	0.2542	0.1209	0.0580	0.0278	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.8618	0.6213	0.3963	0.2336	0.1287	0.0637	0.029			

n	x	p										
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	
n = 13	5	1.0000	0.9991	0.9975	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905	
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000	
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095	
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666	
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539	
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9958	0.9888	
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	n = 14	0	0.4877	0.2788	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
		1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
		2	0.9699	0.8416	0.6478	0.4481	0.2811	0.1608	0.0820	0.0398	0.0170	0.0085
		3	0.9958	0.9559	0.8523	0.6982	0.5212	0.3552	0.2205	0.1243	0.0632	0.0287
4		0.9998	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898	
5		1.0000	0.9985	0.9885	0.9661	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120	
6		1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953	
7		1.0000	1.0000	0.9997	0.9978	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047	
8		1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9755	0.9417	0.8811	0.7880	
9		1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102	
10		1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9888	0.9713	
11		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9925	
12		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9981	
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999		
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
n = 15	0	0.4633	0.2709	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000	
	1	0.8290	0.5490	0.3188	0.1871	0.0802	0.0353	0.0142	0.0052	0.0017	0.0008	
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037	
	3	0.9945	0.9444	0.8227	0.6482	0.4813	0.2989	0.1727	0.0905	0.0424	0.0178	
	4	0.9994	0.9911	0.9382	0.8358	0.6865	0.5165	0.3519	0.2113	0.1204	0.0592	
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5642	0.4032	0.2808	0.1929	
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8889	0.8149	0.6998	0.5522	0.3936	
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8968	0.7889	0.6525	0.5000	
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964	
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8481	
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408	
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824	
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9983	
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9995		
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
n = 18	0	0.4401	0.1853	0.0742	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000	
	1	0.8108	0.5147	0.2830	0.1407	0.0635	0.0281	0.0098	0.0033	0.0010	0.0003	
	2	0.9571	0.7692	0.5814	0.3518	0.1971	0.0994	0.0461	0.0183	0.0068	0.0021	
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2499	0.1339	0.0651	0.0281	0.0106	
	4	0.9991	0.9630	0.8729	0.7982	0.6302	0.4499	0.2892	0.1686	0.0853	0.0384	
	5	0.9999	0.9967	0.9765	0.9383	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051	
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3860	0.2672	
	7	1.0000	0.9999	0.9990	0.9930	0.9729	0.9256	0.8406	0.7181	0.5629	0.4018	
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982	
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7778	
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8948	

n	x	p									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 16	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	0.9912	0.9174	0.7958	0.6489	0.5330	0.2019	0.1028	0.0464	0.0184	0.0064
	4	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5868	0.4197	0.2639	0.1471	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.9998	0.8954	0.7872	0.6405	0.4743	0.3146
	8	1.0000	1.0000	0.9997	0.9974	0.9878	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6858
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9886	0.9652	0.9174	0.8238
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9854	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9931	0.9836
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
n = 18	0	0.3972	0.1501	0.0526	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.7725	0.4503	0.2247	0.0991	0.0395	0.0142	0.0046	0.0013	0.0002	0.0001
	2	0.9418	0.7338	0.4797	0.2710	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	0.9891	0.9218	0.7202	0.5013	0.3057	0.1648	0.0783	0.0378	0.0120	0.0038
	4	0.9995	0.9719	0.8794	0.7164	0.5187	0.3227	0.1886	0.0947	0.0411	0.0154
	5	0.9998	0.9936	0.9581	0.8571	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.8810	0.7217	0.5491	0.3743	0.2258	0.1169
	7	1.0000	0.9998	0.9973	0.9823	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7388	0.5778	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5822
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9474	0.8720	0.7597
	11	1.0000	1.0000	1.0000	0.9998	0.9998	0.9966	0.9928	0.9727	0.9423	0.8211
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9947	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.00					

X	p										
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	
n = 19	8	1.0000	1.0000	0.9997	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8129	0.6710	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6782
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9166
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904	
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9976	
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9998	
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
n = 20	0	0.3585	0.1278	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.2917	0.1756	0.0892	0.0243	0.0078	0.0021	0.0006	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0365	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2282	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6298	0.4146	0.2375	0.1182	0.0510	0.0189	0.0058
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0583	0.0207
	6	1.0000	0.9976	0.9781	0.8123	0.7858	0.8080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9879	0.8982	0.7723	0.6010	0.4159	0.2820	0.1318
	8	1.0000	0.9999	0.9987	0.9911	0.9867	0.8867	0.7824	0.5966	0.4143	0.2817
	9	1.0000	1.0000	0.9998	0.9974	0.9881	0.9520	0.8782	0.7553	0.6914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7807	0.6881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9425	0.8892	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9835	0.9423	0.8423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9762	0.9383	
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9841	
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
n = 21	0	0.3406	0.1094	0.0379	0.0092	0.0024	0.0006	0.0001	0.0000	0.0000	0.0000
	1	0.7170	0.2647	0.1580	0.0578	0.0190	0.0058	0.0014	0.0003	0.0001	0.0000
	2	0.9151	0.6484	0.3705	0.1787	0.0745	0.0271	0.0086	0.0024	0.0008	0.0001
	3	0.9811	0.8480	0.6113	0.3704	0.1917	0.0858	0.0331	0.0110	0.0031	0.0007
	4	0.9968	0.9478	0.8025	0.5860	0.3874	0.1984	0.0924	0.0370	0.0126	0.0030
	5	0.9996	0.9856	0.9173	0.7853	0.5668	0.3827	0.2009	0.0957	0.0398	0.0133
	6	1.0000	0.9987	0.9713	0.8915	0.7438	0.5605	0.3587	0.2002	0.0984	0.0392
	7	1.0000	0.9994	0.9817	0.9568	0.8701	0.7200	0.5385	0.3498	0.1971	0.0848
	8	1.0000	0.9999	0.9980	0.9856	0.9439	0.8523	0.7059	0.5237	0.3413	0.1917
	9	1.0000	1.0000	0.9998	0.9999	0.9794	0.9324	0.8377	0.6914	0.5117	0.3318
	10	1.0000	1.0000	1.0000	0.9999	0.9936	0.9728	0.8758	0.6790	0.5000	
	11	1.0000	1.0000	1.0000	0.9998	0.9963	0.9813	0.8872	0.6913	0.5117	0.3318
	12	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9892	0.8948	0.6982	
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9989	0.9977	0.9821	0.8064
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9984	0.9868	0.8068
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9983	0.8067
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.8064
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.8064
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8064

X	p										
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	
n = 21	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 22	0	0.3235	0.0985	0.0280	0.0074	0.0018	0.0004	0.0001	0.0000	0.0000	0.0000
	1	0.6982	0.2392	0.1367	0.0480	0.0149	0.0041	0.0010	0.0002	0.0000	0.0000
	2	0.9052	0.6200	0.3382	0.1845	0.0606	0.0207	0.0061	0.0018	0.0003	0.0001
	3	0.9778	0.8281	0.5752	0.3320	0.1824	0.0681	0.0245	0.0076	0.0020	0.0004
	4	0.9960	0.9378	0.7738	0.5429	0.3228	0.1645	0.0718	0.0266	0.0083	0.0022
	5	0.9994	0.9818	0.9001	0.7326	0.5168	0.3134	0.1629	0.0722	0.0271	0.0085
	6	0.9999	0.9956	0.9632	0.8870	0.6994	0.4942	0.3022	0.1584	0.0705	0.0282
	7	1.0000	0.9991	0.9888	0.9439	0.8385	0.6713	0.4738	0.2898	0.1518	0.0698
	8	1.0000	0.9998	0.9970	0.9799	0.9254	0.8135	0.6466	0.4540	0.2764	0.1431
	9	1.0000	1.0000	0.9993	0.9929	0.9705	0.9064	0.7918	0.6244	0.4350	0.2617
	10	1.0000	1.0000	0.9999	0.9984	0.9900	0.9613	0.8930	0.7720	0.6037	0.4159
	11	1.0000	1.0000	1.0000	0.9997	0.9971	0.9880	0.9526	0.8793	0.7543	0.5841
	12	1.0000	1.0000	1.0000	0.9999	0.9993	0.9957	0.9820	0.9449	0.8672	0.7383
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9989	0.9642	0.8785	0.8383
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9530	0.8787	0.8321
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9981	0.9520	0.8738
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9979	0.9918
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9978
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 23	0	0.3074	0.0886	0.0238	0.0059	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000
	1	0.6794	0.2151	0.1204	0.0298	0.0116	0.0030	0.0007	0.0001	0.0000	0.0000
	2	0.8948	0.5920	0.3080	0.1332	0.0492	0.0157	0.0043	0.0010	0.0002	0.0000
	3	0.9742	0.8073	0.5396	0.2965	0.1370	0.0538				



TABLA O VALORES PARA LA PRUEBA DEL RANGO MULTIPLE DE DUNCAN

Valores para la prueba del Rango Múltiple de Duncan ( $\alpha = 10\%$ )

$g, \ell$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929	8.929
2	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130	4.130
3	3.328	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330	3.330
4	3.015	3.074	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081	3.081
5	2.850	2.934	2.964	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970	2.970
6	2.748	2.846	2.890	2.908	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911	2.911
7	2.680	2.785	2.838	2.864	2.876	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878	2.878
8	2.630	2.742	2.800	2.832	2.849	2.857	2.858	2.858	2.858	2.858	2.858	2.858	2.858	2.858	2.858	2.858	2.858	2.858
9	2.592	2.708	2.771	2.808	2.829	2.840	2.845	2.847	2.847	2.847	2.847	2.847	2.847	2.847	2.847	2.847	2.847	2.847
10	2.563	2.682	2.748	2.788	2.813	2.827	2.835	2.839	2.839	2.839	2.839	2.839	2.839	2.839	2.839	2.839	2.839	2.839
11	2.540	2.660	2.730	2.772	2.799	2.817	2.827	2.833	2.835	2.835	2.835	2.835	2.835	2.835	2.835	2.835	2.835	2.835
12	2.521	2.643	2.714	2.759	2.789	2.808	2.821	2.828	2.832	2.833	2.833	2.833	2.833	2.833	2.833	2.833	2.833	2.833
13	2.505	2.628	2.701	2.748	2.779	2.800	2.815	2.824	2.829	2.832	2.832	2.832	2.832	2.832	2.832	2.832	2.832	2.832
14	2.491	2.616	2.690	2.739	2.771	2.794	2.810	2.820	2.827	2.831	2.832	2.833	2.833	2.833	2.833	2.833	2.833	2.833
15	2.477	2.605	2.681	2.731	2.765	2.789	2.805	2.817	2.825	2.830	2.833	2.834	2.834	2.834	2.834	2.834	2.834	2.834
16	2.469	2.596	2.673	2.723	2.759	2.784	2.802	2.815	2.824	2.829	2.833	2.835	2.836	2.836	2.836	2.836	2.836	2.836
17	2.460	2.588	2.665	2.717	2.753	2.780	2.798	2.812	2.822	2.829	2.834	2.836	2.838	2.838	2.838	2.838	2.838	2.838
18	2.452	2.580	2.659	2.712	2.749	2.776	2.796	2.810	2.821	2.828	2.834	2.838	2.840	2.840	2.840	2.840	2.840	2.840
19	2.445	2.574	2.653	2.707	2.745	2.773	2.793	2.808	2.820	2.828	2.834	2.839	2.841	2.842	2.843	2.843	2.843	2.843
20	2.439	2.568	2.648	2.702	2.741	2.770	2.791	2.807	2.819	2.828	2.834	2.839	2.843	2.845	2.845	2.845	2.845	2.845
24	2.420	2.550	2.632	2.688	2.729	2.760	2.783	2.801	2.815	2.827	2.835	2.842	2.848	2.851	2.854	2.856	2.857	2.857
30	2.400	2.532	2.615	2.674	2.717	2.750	2.776	2.796	2.813	2.826	2.837	2.846	2.853	2.859	2.863	2.867	2.869	2.871
40	2.381	2.514	2.600	2.660	2.705	2.741	2.769	2.791	2.810	2.825	2.838	2.849	2.859	2.866	2.873	2.878	2.883	2.887
60	2.361	2.497	2.584	2.646	2.694	2.731	2.761	2.785	2.807	2.825	2.839	2.853	2.864	2.874	2.883	2.890	2.897	2.903
120	2.344	2.481	2.568	2.632	2.682	2.722	2.754	2.781	2.804	2.824	2.842	2.857	2.871	2.883	2.893	2.903	2.912	2.920
-	2.326	2.462	2.552	2.619	2.670	2.712	2.746	2.776	2.801	2.824	2.844	2.861	2.877	2.892	2.905	2.918	2.929	2.939

( $\alpha = 5\%$ )

$g, \xi$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92	1.92
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.591	4.516	4.441	4.366	4.291	4.216	4.141	4.066	3.991	3.916	3.841	3.766	3.691	3.616	3.541	3.466	3.391	3.316
4	3.927	4.013	4.099	4.185	4.271	4.357	4.443	4.529	4.615	4.701	4.787	4.873	4.959	5.045	5.131	5.217	5.303	5.389
5	3.635	3.739	3.797	3.854	3.911	3.968	4.025	4.082	4.139	4.196	4.253	4.310	4.367	4.424	4.481	4.538	4.595	4.652
6	3.461	3.587	3.619	3.680	3.741	3.802	3.863	3.924	3.985	4.046	4.107	4.168	4.229	4.290	4.351	4.412	4.473	4.534
7	3.344	3.477	3.518	3.598	3.678	3.758	3.838	3.918	3.998	4.078	4.158	4.238	4.318	4.398	4.478	4.558	4.638	4.718
8	3.261	3.389	3.475	3.521	3.579	3.636	3.694	3.752	3.810	3.868	3.926	3.984	4.042	4.100	4.158	4.216	4.274	4.332
9	3.199	3.329	3.420	3.470	3.522	3.574	3.626	3.678	3.730	3.782	3.834	3.886	3.938	3.990	4.042	4.094	4.146	4.198
10	3.151	3.29	3.476	3.480	3.485	3.490	3.495	3.500	3.505	3.510	3.515	3.520	3.525	3.530	3.535	3.540	3.545	3.550
11	3.113	3.274	3.342	3.397	3.435	3.462	3.480	3.491	3.491	3.491	3.491	3.491	3.491	3.491	3.491	3.491	3.491	3.491
12	3.082	3.225	3.313	3.370	3.410	3.439	3.459	3.474	3.484	3.491	3.496	3.499	3.499	3.499	3.499	3.499	3.499	3.499
13	3.055	3.200	3.289	3.348	3.389	3.419	3.442	3.458	3.470	3.478	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481
14	3.033	3.178	3.268	3.329	3.372	3.403	3.426	3.444	3.457	3.467	3.474	3.479	3.482	3.484	3.484	3.485	3.485	3.485
15	3.014	3.160	3.250	3.312	3.356	3.389	3.413	3.432	3.446	3.457	3.465	3.471	3.476	3.478	3.480	3.481	3.481	3.481
16	2.998	3.144	3.235	3.298	3.343	3.376	3.402	3.422	3.437	3.449	3.458	3.465	3.470	3.473	3.477	3.478	3.478	3.478
17	2.984	3.130	3.222	3.285	3.331	3.365	3.392	3.412	3.429	3.441	3.451	3.459	3.465	3.469	3.473	3.475	3.476	3.476
18	2.971	3.118	3.210	3.274	3.321	3.356	3.383	3.405	3.421	3.435	3.445	3.454	3.460	3.465	3.470	3.472	3.474	3.474
19	2.960	3.107	3.199	3.264	3.311	3.347	3.375	3.397	3.415	3.429	3.440	3.449	3.456	3.462	3.467	3.470	3.472	3.473
20	2.950	3.097	3.189	3.255	3.303	3.339	3.368	3.391	3.409	3.424	3.436	3.445	3.453	3.459	3.464	3.467	3.470	3.472
24	2.934	3.086	3.180	3.226	3.276	3.315	3.345	3.370	3.390	3.406	3.420	3.432	3.441	3.449	3.456	3.461	3.465	3.469
30	2.908	3.061	3.156	3.199	3.250	3.290	3.322	3.349	3.371	3.389	3.405	3.419	3.431	3.441	3.449	3.457	3.464	3.469
40	2.878	3.032	3.128	3.171	3.221	3.261	3.294	3.321	3.344	3.363	3.379	3.393	3.405	3.416	3.425	3.433	3.441	3.448
50	2.852	2.976	3.073	3.116	3.166	3.207	3.241	3.270	3.294	3.314	3.331	3.347	3.361	3.374	3.386	3.397	3.407	3.416
120	2.800	2.947	3.045	3.116	3.172	3.213	3.248	3.276	3.301	3.322	3.340	3.357	3.373	3.388	3.401	3.413	3.424	3.435
*	2.772	2.918	3.017	3.089	3.146	3.187	3.222	3.250	3.274	3.300	3.318	3.335	3.351	3.366	3.380	3.394	3.407	3.419

( $\alpha = 5\%$ )

$g, \xi$	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814
6	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697
7	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626	3.626
8	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579	3.579
9	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547	3.547
10	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526	3.526
11	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510	3.510
12	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499	3.499
13	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490	3.490
14	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483	3.483
15	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481	3.481
16	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478	3.478
17	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476	3.476
18	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474	3.474
19	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473	3.473
20	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471	3.471
24	3.471	3.475	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477	3.477
30	3.470	3.479	3.481	3.484	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486	3.486
40	3.469	3.479	3.486	3.492	3.497	3.500	3.503	3.504	3.504	3.504	3.504	3.504	3.504	3.504	3.504	3.504	3.504
60	3.467	3.481	3.492	3.501	3.509	3.515	3.521	3.525	3.529	3.531	3.534	3.537	3.539	3.543	3.547	3.551	3.557
120	3.466	3.483	3.498	3.511	3.522	3.532	3.541	3.549	3.555	3.561	3.566	3.569	3.576	3.580	3.583	3.587	3.591
*	3.466	3.486	3.505	3.522	3.536	3.550	3.562	3.574	3.584	3.594	3.603	3.610	3.618	3.626	3.634	3.642	3.650

( $\alpha = 1\%$ )

g. l.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03
2	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04
3	8.261	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321
4	6.512	6.677	6.740	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756
5	5.702	5.893	5.989	6.040	6.065	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074
6	5.243	5.439	5.549	5.614	5.655	5.680	5.694	5.701	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703
7	4.949	5.145	5.260	5.334	5.383	5.416	5.439	5.454	5.464	5.470	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472
8	4.746	4.939	5.057	5.135	5.189	5.227	5.256	5.276	5.291	5.302	5.309	5.314	5.316	5.317	5.317	5.317	5.317	5.317
9	4.596	4.787	4.906	4.986	5.043	5.086	5.118	5.142	5.160	5.174	5.185	5.193	5.199	5.203	5.205	5.206	5.206	5.206
10	4.482	4.671	4.790	4.871	4.931	4.975	5.010	5.037	5.058	5.074	5.088	5.098	5.106	5.112	5.117	5.120	5.122	5.124
11	4.392	4.579	4.697	4.780	4.841	4.887	4.924	4.952	4.975	4.994	5.009	5.021	5.031	5.039	5.045	5.050	5.054	5.057
12	4.320	4.506	4.622	4.706	4.767	4.815	4.852	4.883	4.907	4.927	4.944	4.958	4.969	4.978	4.986	4.993	4.998	5.002
13	4.260	4.442	4.560	4.644	4.706	4.755	4.793	4.824	4.850	4.872	4.889	4.904	4.917	4.928	4.937	4.944	4.950	4.956
14	4.210	4.391	4.508	4.591	4.654	4.704	4.743	4.775	4.802	4.824	4.843	4.859	4.872	4.884	4.894	4.902	4.910	4.916
15	4.168	4.347	4.463	4.547	4.610	4.660	4.700	4.733	4.760	4.783	4.803	4.820	4.834	4.846	4.857	4.866	4.874	4.881
16	4.131	4.309	4.425	4.509	4.572	4.622	4.663	4.696	4.724	4.748	4.768	4.786	4.800	4.813	4.825	4.835	4.844	4.851
17	4.099	4.275	4.391	4.475	4.539	4.589	4.630	4.664	4.693	4.717	4.738	4.756	4.771	4.785	4.797	4.807	4.816	4.824
18	4.071	4.246	4.362	4.445	4.509	4.560	4.601	4.635	4.664	4.689	4.711	4.729	4.745	4.759	4.772	4.783	4.792	4.801
19	4.046	4.220	4.335	4.419	4.483	4.534	4.575	4.610	4.639	4.665	4.686	4.705	4.722	4.736	4.749	4.761	4.771	4.780
20	4.024	4.197	4.312	4.395	4.459	4.510	4.552	4.587	4.617	4.642	4.664	4.684	4.701	4.716	4.729	4.741	4.751	4.761
24	3.956	4.126	4.241	4.322	4.386	4.437	4.480	4.516	4.546	4.573	4.596	4.616	4.634	4.651	4.665	4.678	4.690	4.703
30	3.949	4.056	4.164	4.250	4.314	4.364	4.409	4.445	4.477	4.504	4.528	4.550	4.569	4.586	4.601	4.615	4.628	4.640
40	3.825	3.988	4.098	4.180	4.244	4.294	4.339	4.376	4.408	4.436	4.461	4.483	4.503	4.521	4.537	4.553	4.566	4.579
60	3.762	3.922	4.031	4.111	4.174	4.224	4.270	4.307	4.340	4.368	4.394	4.417	4.438	4.456	4.474	4.490	4.504	4.518
120	3.702	3.858	3.965	4.044	4.107	4.158	4.202	4.239	4.272	4.301	4.327	4.351	4.372	4.392	4.410	4.426	4.442	4.456
=	3.643	3.796	3.900	3.978	4.040	4.091	4.135	4.172	4.205	4.235	4.261	4.285	4.307	4.327	4.345	4.363	4.379	4.394

( $\alpha = 1\%$ )

g. l.	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03
2	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04
3	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321
4	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756	6.756
5	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074	6.074
6	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703	5.703
7	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472	5.472
8	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317	5.317
9	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206	5.206
10	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124	5.124
11	5.059	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061	5.061
12	5.006	5.010	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011	5.011
13	4.960	4.966	4.969	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972	4.972
14	4.921	4.929	4.934	4.938	4.940	4.941	4.941	4.941	4.941	4.941	4.941	4.941	4.941	4.941	4.941	4.941	4.941
15	4.887	4.897	4.903	4.907	4.910	4.912	4.913	4.913	4.913	4.913	4.913	4.913	4.913	4.913	4.913	4.913	4.913
16	4.856	4.868	4.874	4.878	4.881	4.883	4.884	4.884	4.884	4.884	4.884	4.884	4.884	4.884	4.884	4.884	4.884
17	4.827	4.841	4.848	4.852	4.855	4.857	4.858	4.858	4.858	4.858	4.858	4.858	4.858	4.858	4.858	4.858	4.858
18	4.800	4.815	4.822	4.826	4.829	4.831	4.832	4.832	4.832	4.832	4.832	4.832	4.832	4.832	4.832	4.832	4.832
19	4.774	4.790	4.812	4.821	4.826	4.831	4.834	4.836	4.838	4.840	4.841	4.842	4.843	4.844	4.845	4.845	4.845
20	4.769	4.784	4.795	4.805	4.811	4.816	4.821	4.825	4.830	4.832	4.833	4.833	4.833	4.833	4.833	4.833	4.833
24	4.710	4.727	4.741	4.752	4.762	4.770	4.777	4.783	4.788	4.791	4.794	4.802	4.802	4.802	4.802	4.802	4.802
30	4.650	4.669	4.685	4.699	4.711	4.721	4.730	4.738	4.744	4.750	4.755	4.772	4.772	4.772	4.772	4.772	4.772
40	4.591	4.611	4.630	4.645	4.659	4.671	4.682	4.692	4.700	4.708	4.715	4.740	4.754	4.761	4.764	4.764	4.764
60	4.530	4.553	4.573	4.591	4.607	4.620	4.633	4.645	4.655	4.665	4.673	4.707	4.730	4.745	4.755	4.761	4.765
120	4.469	4.494	4.516	4.535	4.552	4.568	4.583	4.596	4.609	4.619	4.630	4.673	4.703	4.727	4.745	4.759	4.770
=	4.408	4.434	4.457	4.478	4.497	4.514	4.530	4.545	4.559	4.572	4.584	4.635	4.675	4.707	4.734	4.756	4.776

( $\alpha = 0.5\%$ )

$g \cdot \xi$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1
2	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93
3	10.55	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63
4	7.916	8.126	8.210	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238
5	6.751	6.980	7.100	7.167	7.204	7.222	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228
6	6.105	6.334	6.466	6.547	6.600	6.635	6.658	6.672	6.679	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682
7	5.699	5.922	6.057	6.145	6.207	6.250	6.281	6.304	6.320	6.331	6.339	6.343	6.345	6.345	6.345	6.345	6.345	6.345
8	5.420	5.638	5.773	5.864	5.930	5.978	6.014	6.042	6.064	6.080	6.092	6.101	6.108	6.113	6.116	6.118	6.119	6.119
9	5.218	5.430	5.565	5.657	5.725	5.776	5.815	5.846	5.871	5.891	5.907	5.920	5.930	5.938	5.944	5.949	5.952	5.955
10	5.065	5.273	5.405	5.498	5.567	5.620	5.662	5.695	5.722	5.744	5.762	5.777	5.790	5.800	5.809	5.816	5.821	5.826
11	4.945	5.149	5.280	5.372	5.442	5.496	5.539	5.574	5.603	5.626	5.646	5.663	5.678	5.690	5.700	5.709	5.716	5.722
12	4.849	5.048	5.178	5.270	5.341	5.396	5.439	5.475	5.505	5.531	5.552	5.570	5.585	5.599	5.610	5.620	5.629	5.636
13	4.770	4.966	5.094	5.186	5.256	5.312	5.356	5.393	5.424	5.450	5.472	5.492	5.508	5.523	5.535	5.546	5.556	5.564
14	4.704	4.897	5.023	5.116	5.185	5.241	5.286	5.324	5.355	5.382	5.405	5.425	5.442	5.458	5.471	5.483	5.494	5.503
15	4.647	4.838	4.964	5.055	5.125	5.181	5.226	5.264	5.297	5.324	5.348	5.368	5.386	5.402	5.416	5.429	5.440	5.450
16	4.599	4.787	4.912	5.003	5.073	5.129	5.175	5.213	5.245	5.273	5.298	5.319	5.338	5.354	5.368	5.381	5.393	5.404
17	4.557	4.744	4.869	4.958	5.027	5.084	5.130	5.168	5.201	5.229	5.254	5.275	5.295	5.311	5.327	5.340	5.352	5.363
18	4.521	4.705	4.828	4.918	4.987	5.043	5.090	5.129	5.162	5.190	5.215	5.237	5.256	5.274	5.289	5.303	5.316	5.327
19	4.488	4.671	4.793	4.883	4.952	5.008	5.054	5.093	5.127	5.156	5.181	5.203	5.222	5.240	5.256	5.270	5.283	5.295
20	4.460	4.641	4.762	4.851	4.920	4.976	5.022	5.061	5.095	5.124	5.150	5.172	5.193	5.210	5.226	5.241	5.254	5.266
24	4.371	4.547	4.666	4.753	4.822	4.877	4.924	4.963	4.997	5.027	5.054	5.076	5.097	5.116	5.133	5.148	5.162	5.175
30	4.285	4.456	4.572	4.658	4.726	4.781	4.827	4.867	4.901	4.931	4.958	4.981	5.003	5.022	5.039	5.056	5.071	5.085
40	4.202	4.369	4.482	4.566	4.632	4.687	4.733	4.772	4.806	4.837	4.864	4.888	4.910	4.930	4.948	4.965	4.980	4.995
60	4.122	4.284	4.394	4.476	4.541	4.595	4.640	4.679	4.713	4.744	4.771	4.796	4.818	4.838	4.857	4.874	4.890	4.905
120	4.045	4.201	4.308	4.388	4.452	4.505	4.550	4.589	4.622	4.652	4.679	4.704	4.726	4.747	4.766	4.784	4.800	4.815
$\infty$	3.970	4.121	4.225	4.303	4.365	4.417	4.461	4.499	4.532	4.562	4.589	4.614	4.636	4.657	4.676	4.694	4.710	4.726

( $\alpha = 0.5\%$ )

$g \cdot \xi$	20	22	24	26	28	30	32	34	36	38	40	50	60	70	80	90	100
1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1	180.1
2	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93	19.93
3	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63	10.63
4	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238	8.238
5	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228	7.228
6	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682	6.682
7	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345	6.345
8	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119	6.119
9	5.956	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957	5.957
10	5.829	5.834	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836	5.836
11	5.727	5.735	5.740	5.743	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744	5.744
12	5.642	5.653	5.660	5.665	5.668	5.670	5.670	5.670	5.670	5.670	5.670	5.670	5.670	5.670	5.670	5.670	5.670
13	5.571	5.583	5.593	5.600	5.605	5.608	5.610	5.611	5.611	5.611	5.611	5.611	5.611	5.611	5.611	5.611	5.611
14	5.511	5.525	5.535	5.544	5.550	5.555	5.559	5.561	5.563	5.563	5.563	5.563	5.563	5.563	5.563	5.563	5.563
15	5.459	5.474	5.486	5.495	5.503	5.509	5.514	5.518	5.520	5.522	5.523	5.523	5.523	5.523	5.523	5.523	5.523
16	5.412	5.429	5.441	5.453	5.462	5.469	5.475	5.479	5.483	5.485	5.488	5.489	5.489	5.489	5.489	5.489	5.489
17	5.371	5.389	5.401	5.416	5.425	5.433	5.440	5.445	5.450	5.453	5.456	5.461	5.461	5.461	5.461	5.461	5.461
18	5.336	5.355	5.367	5.383	5.391	5.402	5.409	5.415	5.420	5.425	5.428	5.436	5.436	5.436	5.436	5.436	5.436
19	5.306	5.325	5.337	5.354	5.364	5.374	5.382	5.390	5.395	5.399	5.403	5.414	5.415	5.415	5.415	5.415	5.415
20	5.277	5.296	5.311	5.328	5.338	5.348	5.357	5.364	5.370	5.374	5.380	5.394	5.397	5.397	5.397	5.397	5.397
24	5.187	5.209	5.226	5.242	5.255	5.267	5.278	5.287	5.295	5.302	5.308	5.329	5.340	5.343	5.343	5.343	5.343
30	5.098	5.121	5.140	5.157	5.172	5.186	5.198	5.209	5.218	5.227	5.235	5.264	5.281	5.292	5.297	5.298	5.298
40	5.008	5.032	5.054	5.072	5.089	5.104	5.118	5.130	5.141	5.151	5.160	5.197	5.221	5.238	5.249	5.257	5.261
60	4.919	4.944	4.967	4.987	5.005	5.021	5.036	5.050	5.062	5.074	5.084	5.128	5.159	5.182	5.199	5.213	5.223
120	4.830	4.856	4.880	4.901	4.920	4.937	4.953	4.968	4.982	4.995	5.007	5.056	5.094	5.123	5.146	5.166	5.182
$\infty$	4.740	4.767	4.792	4.813	4.833	4.852	4.869	4.885	4.899	4.913	4.926	4.981	5.024	5.059	5.088	5.114	5.136





## BIBLIOGRAFÍA

1. Bricio Hernández Diego; **"PROBABILIDAD Y ESTADÍSTICA I"**, Ed. Limusa, México.
2. Canavos George C.; **"PROBABILIDAD Y ESTADÍSTICA"**, Ed. MacGrawhill, México.
3. Chao Lincoln L.; **"INTRODUCCIÓN A LA ESTADÍSTICA"**, Compañía Editorial Continental, México, 1985.
4. Collet D.; **"ESTADÍSTICA Y TEORÍA DE LA PROBABILIDAD"**, México, 1987.
5. Haber Audrey; **"ESTADÍSTICA GENERAL"**, Fondo Educativo Interamericano(FEI), Bogotá Colombia,
6. Hanett; **"INTRODUCCIÓN AL ANÁLISIS ESTADÍSTICO"**,
7. Hernández L.D.; **"APUNTES DE PROBABILIDAD Y ESTADÍSTICA"**, Publicación de Sociedad Matemática Mexicana-UAM, México, 1977.
8. Hines William W.; **"PROBABILIDAD Y ESTADÍSTICA PARA INGENIERÍA Y ADMINISTRACIÓN"**,
9. Kreyszig Erwin; **"INTRODUCCIÓN A LA ESTADÍSTICA MATEMÁTICA"**, Ed. Limusa, México, 1979.
10. Lipschutz Seymour; **"PROBABILIDAD"**, Ed. MacGrawhill, 1994.
11. Maisel Louis; **"PROBABILIDAD Y ESTADÍSTICA"**, FEI, Bogotá Colombia,
12. Marques María José; **"PROBABILIDAD Y ESTADÍSTICA PARA CIENCIAS QUIMICO BIOLÓGICAS"**, UNAM, México, 1988.
13. Mendenhall William; **"INTRODUCCION A LA PROBABILIDAD Y ESTADÍSTICA"**, Ed. Internacional Iberoamericana, Belmont California,
14. Méndez Ramírez Ignacio; **"PROBABILIDAD Y ESTADÍSTICA IV"**, Ed. Limusa, México,
15. Meyer Paul L.; **"PROBABILIDAD Y APLICACIONES ESTADÍSTICAS"**, FEI, Bogotá Colombia,
16. Miller Irwin; **"PROBABILIDAD Y ESTADÍSTICA PARA INGENIEROS"**, Ed. Reverté, México, 1992.
17. Montgomery C. Douglas; **"DISEÑO Y ANALISIS DE EXPERIMENTOS"**, Ed. Iberoamericana, México, 1991.
18. Moreno Bonnett A.; **"ELEMENTOS DE PROBABILIDAD Y ESTADÍSTICA"**, Representaciones y Servicios de Ingeniería, México,
19. Ostle Bernard; **"ESTADÍSTICA APLICADA"**, Ed. Limusa, México, 1974.
20. Phee Hurley D.; **"PROBABILIDAD Y ESTADÍSTICA III"**, Ed. Limusa, México,
21. Rascon Octavio; **"INTRODUCCION A LA TOERIA DE PROBABILIDADES"**, UNAM, Dirección General de Publicaciones, México, 1976.

22. Rickmers; **"INTRODUCCIÓN A LA ESTADÍSTICA"**, Ed. Continental, México, 1979.
23. Ruiz Moncayo A.; **"PROBABILIDAD Y ESTADÍSTICA II"**, Ed. Limusa México,
24. Scheaffer M.; **"ELEMENTOS DE MUESTREO"**, Grupo Editorial Iberoamericana,
25. Spiegel & Murray R.; **"TEORIA Y PROBLEMAS DE PROBABILIDAD Y ESTADÍSTICA"**, Ed. MacGrawhill, México,
26. Walpole Ronald E.; **"PROBABILIDAD Y ESTADÍSTICA PARA INGENIEROS"**, FEI, México, 1982.
27. Whitmore; **"PROBLEMAS DE ESTADÍSTICA"**, Compañía Editorial Continental, México, 1981.
28. Wonnacot Ronald; **"ESTADÍSTICA BÁSICA PRÁCTICA"**, Ed. Limusa, México,