18

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLÁN"

296635

"MEMORIA DEL DESEMPEÑO ACADÉMICO EN LAS ASIGNATURAS ESTADÍSTICA I Y ESTADÍSTICA II DEL COLEGIO DE CIENCIAS Y HUMANIDADES, PLANTEL NAUCALPAN, DURANTE EL CICLO ESCOLAR 98-I/98-II"

MEMORIA DE DESEMPEÑO PROFESIONAL

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ACTUARÍA

P R E S E N T A :

HUGO MAEL HERNÁNDEZ TREVETHAN

ASESORA: ACT. SILVIA MADRIGAL HERNÁNDEZ

SEPTIEMBRE, 2001

Memoria del desempeño académico en las asignaturas Estadística I y Estadística II del Colegio de Ciencias y Humanidades, plantel Naucalpan, durante el ciclo escolar 98-I/98-II

...siglos pasaron sin que el azar, duende juguetón, sus hilos moviera...

Joaquín Sabina.

A mis alumnos, por su invaluable ayuda
en la elaboración de este trabajo

Estela Barrera Noguez
Mayra López Mejía
Arlette Patricia Medina Arellano
Carlos Enrique Odriozola Aviña
Toanaméyotl Rodríguez Rayón
María Elena Saavedra Deciga

A todos los que de una u otra forma me apoyaron para hacer esto posible.

ÍNDICE

INTRODUCCION

Introducción

a) Introducción

El presente trabajo recoge las experiencias acumuladas como docente en las asignaturas de Estadística I y Estadística II en el plantel Naucalpan del Colegio de Ciencias y Humanidades de la UNAM, durante los semestres 97-I (de agosto de 1996 a enero de 1997) y 97-II (de febrero a mayo de 1997) respectivamente.

Dado que en agosto de 1996 entraron en vigor los nuevos Planes y Programas de Estudio para el Colegio de Ciencias y Humanidades para los alumnos de nuevo ingreso al bachillerato, cabe aclarar que las materias a las que hace referencia la presente memoria fueron impartidas a los alumnos de quinto y sexto semestre de acuerdo a los Planes y Programas vigentes desde 1971, ya que los nuevos programas para quinto y sexto semestres comenzaron a impartirse en agosto de 1998.

Las materias de Estadística tenían un carácter semioptativo, esto es, los alumnos debían cursar en estos últimos dos semestres exclusivamente una materia del área de Matemáticas, pero podían elegir libremente entre Lógica, Cálculo Diferencial e Integral (no se presentaba con este nombre sino como Matemáticas V y VI) y Estadística.

b) Contextualización

La enseñanza de la Matemática, desde siempre ha presentado un doble reto al docente: el de presentar una Ciencia en sí misma, y el de presentar una serie de modelos auxiliares a otras ciencias (el concepto de la Matemática como herramienta). Esta doble cualidad siempre nos lleva, en el marco de la didáctica, al tener que decidir si enseñar Matemáticas como una herramienta práctica, en donde no importan tanto los " por qué " de las cosas y sí los "para qué", o si enseñar una Matemática más teórica o "axiomática" con la idea de formar en nuestros estudiantes una capacidad de razonamiento lógico, ordenado y analítico y con la intención de que los estudiantes conozcan los sustentos teóricos de alguna rama de la Matemática en particular.

En el Colegio de Ciencias y Humanidades esto no era, ni es, la excepción, y más si se toma en cuenta la mecánica de los cursos del área de Matemáticas en los últimos dos semestres, en donde los alumnos tenían que elegir de entre Matemáticas (que comprende un programa de Cálculo Diferencial e Integral), Lógica y Estadística una sola materia, selección que los colocaba en la situación de inscribir la materia según con su perfil de egreso esperado de acuerdo a la carrera profesional que pensaran seguir (en el caso de Estadística, la mayoría de los alumnos inscritos en la materia estaban orientados profesionalmente a carreras de áreas Químico-Biológica y Económico-Administrativa mayoritariamente). Este concepto de "utilidad" de la Estadística dentro de una carrera profesional no matemática, y que es el que guiaba a los estudiantes a inscribir la materia en particular, me hizo concluir que la didáctica de esta materia debería ser enfocada principalmente al aspecto práctico, a la Estadística como herramienta, más que a cuestiones puramente (o mayoritariamente) teóricas, aunque sin perder la conciencia de que el aspecto práctico no puede ir desligado de las cuestiones teóricas, con el fin de cumplir la demanda que el perfil de egreso presentaba en mis estudiantes.

Con lo anterior podríamos decir que todo se reducía a enseñar Estadística mayoritariamente práctica para el caso particular de la materia, pero ocurría que no solamente esto no se daba casi nunca, sino que además no existía un programa general para la materia que se manejara en el Colegio de Ciencias y Humanidades en el plantel Naucalpan, en donde se recogieron las experiencias que le dieron forma al presente trabajo.

c) Problemática

En el ámbito del plantel Naucalpan del Colegio de Ciencias y Humanidades, al impartir por vez primera la materia de Estadística, encontré que si bien existían algunos programas y temarios, éstos eran diferentes en cada turno y además los profesores de cada materia no se sujetaban del todo a dichos temarios, y lo que se enseñaba en cada curso era más bien discrecional, aunque muy apegado a los aspectos teóricos y formales de la Estadística y, sobre todo, de la Probabilidad. Es decir, los contenidos impartidos estaban en función del criterio de cada profesor y el enfoque con el que se abordaban era sumamente teórico y poco didáctico.

A medida que avancé en mí trabajo con la materia, fui notando un rechazo general de los estudiantes hacía la Estadística, argumentando que de las tres opciones en Matemáticas para 5° semestre, era "la más difícil". Encontré que el problema principal radicaba en el hecho de que el enfoque vocacional de aquellos que habían elegido Estadística, era hacia áreas Químico-Biológicas o Económico-Administrativas en su mayoría, y esta necesidad de una Estadística práctica no siempre coincidía con el carácter teórico con el que se enseñaba, lo que causaba altos niveles de reprobación y deserción y rechazo hacia la materia, con los consecuentes prejuicios de los alumnos, ya que al cursarla tenían una fuerte convicción de que no iban a aprender y por ende de que iban a reprobar, predisposición que en nada les ayudaba a lo largo de su desempeño de los cursos; para el caso de los alumnos ya inscritos en Estadística, el objetivo de aprender algo que les fuera útil se perdía en el objetivo de aprobar por cualquier medio. Y así llegamos a la consecuencia más grave de todo lo anterior: el proceso de enseñanza-aprendizaje en la materia de Estadística resultaba deficiente, siendo dos de las posibles causas la falta de un programa homologado, tanto por cursos ordinarios como para exámenes extraordinarios, por una parte, y al carácter teórico casi en su totalidad con el que se abordaban los cursos, por otra parte, cuando éstos debían ser más prácticos en función de los requerimientos de los alumnos.

d) Objetivos

Al hacerme cargo de un grupo de Estadística, cuestionaba (y cuestiono) a los alumnos sobre los motivos que tuvieron para elegir la materia, y en la gran mayoría de los casos la respuesta era (y sigue siendo) la necesidad de la "herramienta estadística" para su carrera, además de una pequeña cantidad de respuestas más vanas, como lo eran acompañar a alguna persona en particular o el hecho de tener materias de Matemáticas no aprobadas de los anteriores semestres, en las cuales se encuentran nociones de Lógica y Cálculo, por lo que se buscaba intentar aprobar con otro tipo de Matemática. Lo anterior me llevó a plantear una serie de objetivos propios de los cursos, a partir de los cuales se desprenden los objetivos del presente trabajo:

Introducción

1. Objetivos de los dos cursos:
   - Que el alumno identifique los elementos descriptivos, numéricos o gráficos, de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.
   - Que el alumno calcule los elementos descriptivos numéricos de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.
   - Que el alumno trace los elementos descriptivos gráficos de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.
   - Que el alumno identifique los distintos arreglos en el Cálculo Combinatorio y resuelva problemas de conteo.
   - Que el alumno pueda aplicar elementos propios del Análisis Combinatorio al cálculo de probabilidades.
   - Que el alumno conozca las definiciones principales y algunos teoremas básicos en la Teoría de Probabilidad, tales como espacio muestral, evento, eventos independientes, eventos mutuamente excluyentes, eventos complementarios.
   - Que el alumno comprenda los Axiomas de Kolmogorov como base de la Teoría Moderna de Probabilidades.
   - Que el alumno conozca los conceptos de variable aleatoria discreta y de distribución de probabilidad de variable aleatoria discreta.
   - Que el alumno diferencie entre una variable aleatoria discreta y una variable aleatoria continua.
   - Que el alumno calcule el valor esperado, la variancia y la desviación estándar de una variable aleatoria discreta.
   - Que el alumno conozca los ensayos de Bernoulli e identifique algunas distribuciones de probabilidad de variable aleatoria discreta, algunas de ellas como consecuencia de los ensayos de Bernoulli, como modelos de solución para ciertos problemas de cálculo de probabilidades y que resuelva dichos problemas (Binomial, Binomial Negativa, Poisson, Multinomial).

   - Que el alumno conozca algunos conceptos básicos de Muestreo y Estimación, tales como estimación puntual, estimación por intervalos, muestreo, parámetro, estimador, como una introducción hacia la Estadística Inferencial.
   - Que el alumno conozca la necesidad de aplicar cálculo de probabilidades dentro de la Estadística Inferencial.
   - Que el alumno recuerde las características de una variable aleatoria discreta y sus diferencias con una variable aleatoria continua.
   - Que el alumno conozca el concepto de variable aleatoria continua y sus características.
   - Que el alumno identifique la probabilidad de ocurrencia de valores de variables aleatorias continuas como áreas bajo la curva, a partir de problemas que lleven a una distribución Uniforme.
   - Que el alumno identifique problemas de cálculo de probabilidades que involucran variables aleatorias continuas.

- Que el alumno conozca y aplique la distribución Uniforme como modelo de solución para ciertos problemas de cálculo de probabilidades, y que resuelva dichos problemas.
- Que el alumno conozca y aplique la distribución Exponencial como modelo de solución para ciertos problemas de cálculo de probabilidades, y que resuelva dichos problemas.
- Que el alumno conozca la distribución Normal y su estandarización a partir de la Regla Empírica.
- Que el alumno identifique la distribución Normal Estandarizada como un modelo de solución para ciertos problemas de cálculo de probabilidades de variable aleatoria continua y que resuelva dichos problemas.
- Que el alumno conozca el concepto de estimación por intervalos y de los elementos que lo componen.
- Que el alumno conozca los modelos de estimación por intervalos que utilizan las distribuciones muestrales Normal, t de Student Ji-Cuadrada y F de Snedecor y resuelva problemas de estimación de los parámetros media, proporción, diferencia de medias, diferencia de proporciones, variancia y cociente de variancias con el modelo correspondiente e interpretando las soluciones.
- Que el alumno conozca el concepto de Prueba de Hipótesis y de los elementos que la componen.
- Que el alumno aplique los modelos de prueba de hipótesis que utilizan las distribuciones muestrales Normal, t de Student Ji-Cuadrada y F de Snedecor para resolver problemas de prueba de hipótesis de los parámetros media, proporción, diferencia de medias, diferencia de proporciones, variancia y cociente de variancias con el modelo correspondiente e interpretando las soluciones.
- Que el alumno establezca una relación lineal, parabólica, exponencial o logarítmica entre dos variables aleatorias aparejadas involucradas en un problema estadístico, cuando esto sea posible, con el fin de resolver problemas de estimación por medio de regresión lineal o regresión no lineal, según sea el caso.
- Que el alumno conozca y utilice el método de mínimos cuadrados y los elementos que lo conforman (diagramas de dispersión, coeficiente de correlación) para predecir el comportamiento de variables aleatorias aparejadas.
- Que el alumno conozca algunos elementos básicos a considerar dentro del diseño de un experimento, a partir de retomar los elementos básicos de Muestreo y Estimación y relacionarlos con los conocimientos adquiridos en el curso.

2. Objetivos del trabajo:
   - Que los profesores del plantel Naucalpan del Colegio de Ciencias y Humanidades (CCH Naucalpan) dispongan de un material empíricamente avalado dentro el aula y en el contexto del mismo plantel, ya que el origen del trabajo se encuentra precisamente en la experiencia en el salón de clases.
   - Contar, en un solo documento, con ejercicios, actividades, conceptos, cuadros y tablas, fórmulas y algunos elementos de evaluación, a los que normalmente se tiene acceso a través de diferentes textos, para que puedan ser utilizados en posteriores cursos de la materia de Estadística y Probabilidad en el CCH Naucalpan.

Así mismo, en lo personal busco lograr, a lo largo de los dos cursos, los puntos siguientes y que también son base para los objetivos de este trabajo:

-Dotar a los alumnos de los requerimientos curriculares necesarios en la carrera profesional de su elección.
-Apoyar el proceso de enseñanza-aprendizaje en Estadística y Probabilidad a partir de presentar la materia de una forma accesible y estableciendo su utilidad práctica a lo largo de los dos cursos.
-Tratar de que la reprobación y la deserción en los grupos a mi cargo se lo menor posible, al presentar la materia como un elemento tanto útil como accesible para los estudiantes.
-Fomentar en lo posible la capacidad de abstracción y de formar un razonamiento ordenado, la capacidad para interpretar y explicar un resultado numérico dentro del contexto de un problema, y manejar simultáneamente el lenguaje matemático y el lenguaje común por medio de actividades y problemas enfocados a aplicaciones prácticas de la Estadística, y presentando los conceptos propios de la materia de forma sencilla y tratando de evitar tratamientos teóricos de un nivel superior al del bachillerato.

e) Propuesta de solución

El primer punto a resolver con el fin de mejorar la enseñanza de la Estadística consistió en buscar o generar al menos una guía de contenidos, que incluyera los mínimos elementos estadísticos necesarios para recabar, organizar, presentar e interpretar los datos numéricos de interés que pudieran obtenerse de la investigación de alguna o algunas características propias de una población; los planteamientos que justificaran la existencia de los elementos anteriores; y las cuestiones que ayudaran a formar un criterio y una noción de azar y probabilidad. Al mismo tiempo, guías con las que se trabajaría deberían ajustarse a los objetivos establecidos.

Así, de acuerdo con los lineamientos anteriores, con mis experiencias en la materia de Estadística como alumno en el bachillerato y en la carrera, y analizando qué temas de Estadística y Probabilidad elementales tenían más aplicación o eran más fáciles de adecuar a situaciones cotidianas, decidí impartir en el primer curso elementos de Estadística Descriptiva, Cálculo Combinatorio y Probabilidad y algunas funciones de probabilidad de variables aleatorias discretas; y en el segundo curso nociones básicas de Muestreo y Diseño de Experimentos, algunas funciones de probabilidad de variables aleatorias continuas, Intervalos de Confianza y Pruebas de Hipótesis utilizando tablas de distribución de probabilidades adecuadas a cada prueba y Estimación por Mínimos Cuadrados (en resumen, un curso básico de Estadística Inferencial). A continuación se presentan las guías de contenidos que se desarrollaron para cada curso:

Estadística I
1. Estadística Descriptiva
   a) Datos no agrupados y datos agrupados. Clases, marcas de clase, límites de clase, límites reales de clase, tamaño de clase. Tablas de distribución de frecuencias.
   b) Medidas de tendencia central. Media, mediana y moda.
   c) Gráficas. Histograma, polígono de frecuencias, ojiva, ojiva porcentual y gráfico circular.
   d) Medidas de dispersión. Amplitud, variancia y desviación estándar.

e) Medidas de posición o cuantiles. Cuartiles, quintiles, deciles y centiles. Método de interpolación.
f) Medidas de sesgo y apuntalamiento.

2. Técnicas de conteo
a) Operación factorial.
b) Tipos de arreglos en el Cálculo Combinatorio. Teorema Fundamental del Cálculo Combinatorio. Combinación, Permutación, Permutación con Objetos Repetidos, Ordenación, Ordenación con Repetición, Ordenación Cíclica y Combinación con Repetición.

3. Elementos de Probabilidad
a) Definiciones, axiomas de Kolmogorov, teoremas. Probabilidad marginal, conjunta y condicional. Eventos mutuamente excluyentes, eventos independientes.

4. Funciones de distribución de probabilidad de variable aleatoria discreta
a) Variables aleatorias discretas. Tablas de distribución de probabilidades de variables aleatorias discretas. Valor esperado y variancia de una función de distribución de probabilidades de variable aleatoria discreta.
b) Funciones de distribución de probabilidades Bernoulli, Binomial, Poisson, Geométrica, Hipergeométrica, Binomial Negativa, Multinomial e Hipergeométrica Multivariada.

Estadística II

1. Conceptos básicos de Muestreo y Estimación
a) Conceptos básicos.

2. Funciones de distribución de probabilidad de variables aleatorias continuas
a) Distribución Uniforme.
b) Distribución Exponencial.
c) Distribución Normal Estandarizada. Manejo de tablas de áreas bajo la curva Normal Estandarizada. Aproximación de la Binomial a la Normal.

3. Intervalos de confianza
a) Intervalos de confianza para la media, la diferencia de medias, la proporción y la diferencia de proporciones con muestras grandes, utilizando la tabla de áreas bajo la curva Normal estandarizada.
b) Intervalos de confianza para la media, y la diferencia de medias con muestras chicas, utilizando la tabla de áreas bajo la curva de la función de distribución t de Student.
c) Intervalos de confianza para la variancia, utilizando la tabla de áreas bajo la curva de la función de distribución Ji – cuadrada.
d) Intervalos de confianza para el cociente de dos variancias, utilizando la tabla de áreas bajo la curva de la función de distribución F de Snedecor.

4. Prueba de hipótesis
a) Pruebas de hipótesis para la media, la diferencia de medias, la proporción y la diferencia de proporciones con muestras grandes, utilizando la tabla de áreas bajo la curva Normal estandarizada.
b) Pruebas de hipótesis para la media y la diferencia de medias con muestras chicas, utilizando la tabla de áreas bajo la curva de la función de distribución t de Student.
c) Prueba de hipótesis para la variancia, utilizando la tabla de áreas bajo la curva de la función de distribución Ji – cuadrada.

c) Prueba de hipótesis para la variancia, utilizando la tabla de áreas bajo la curva de la función de distribución Ji – cuadrada.
d) Prueba de hipótesis para el cociente de dos variancias utilizando la tabla de áreas bajo la curva de la función de distribución F de Snedecor.

5. Estimación por mínimos cuadrados
   a) Nociones de ajuste de curvas. Diagramas de dispersión.
   b) La recta de mínimos cuadrados. Coeficiente de correlación. Estimaciones.
   c) La parábola de mínimos cuadrados. Estimaciones.
   d) La exponencial de mínimos cuadrados. Estimaciones.
   e) La logarítmica de mínimos cuadrados.
6. Algunas consideraciones en el Diseño de Experimentos
   a) Conceptos básicos.

En casi la totalidad del curso, las estrategias didácticas empleadas consistieron en presentar a los estudiantes alguna problemática por medio de un ejemplo o de un ejercicio práctico, para después ir generando los elementos necesarios para la solución, e incluso, en algunos casos, para generar también nuevos requerimientos de soluciones que deducir. La manera de ir creando estos modelos de solución se fue la lluvia de ideas de los estudiantes, ideas que se iban ordenando y guiando para lograr una adecuada construcción de los modelos dentro del ejercicio y su posterior generalización. Una vez resuelto el problema presentado y establecidas las formas generales de los modelos deducidos, se procedía a presentar a los alumnos una serie de distintos problemas en los que dichos modelos generados fueran aplicables, problemas que resolvían en clases por medio de trabajo cooperativo y contando siempre con la asistencia del profesor. Igualmente la presentación de algún problema se utilizó para recuperar conocimientos previos requeridos, como lo fueron la operación factorial y el Teorema de Tales.

Ahora, en algunos casos se emplearon otro tipo de estrategias. Estos fueron:

1. Presentación directa de algunos modelos específicos, por requerir su deducción de elementos más allá del alcance del conocimiento propio de los alumnos de bachillerato; por ejemplo, de los coeficientes de sesgo y de apuntalamiento.
2. En la unidad 3 de Estadística I, Elementos de Probabilidad, la estrategia empleada fue la de la exposición del tema por parte del profesor e ilustrando con algunos ejemplos las definiciones y los teoremas presentados, en donde algunos ejemplos también se les solicitaron a los alumnos. La razón para esta exposición fue que la unidad propiamente se pensó para introducir a los estudiantes al lenguaje y la simbolización propios de la probabilidad, con el fin de tener bases en las unidades 4 de Estadística I, Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta, y 2 de Estadística II Funciones de distribución de probabilidad de variables aleatorias continuas.
3. En las unidades 1 y 6 de Estadística II, Conceptos básicos de Muestreo y Estimación y Algunas consideraciones en el Diseño de Experimentos, se realizó una exposición de los temas por parte del profesor e ilustrando con algunos ejemplos, ya que la intención de estas unidades fue, respectivamente, introducir a los alumnos al lenguaje y la notación de algunos elementos propios del muestreo y la estimación, y la de presentar a las estudiantes la idea de que el muestreo, la estimación y, en general, el diseño de un experimento, presentan un panorama bastante más amplio

que lo visto en los temas que conformaron sus dos cursos de la materia de Estadística.

Un elemento a considerar dentro de los cursos es la evaluación. En este punto se dijo a los alumnos que se aplicarían dos exámenes parciales, y un examen final para aquellos que no aprobaran alguno de los parciales (o ambos) y para aquellos alumnos exentos interesados en conseguir una calificación más alta. Pero también se explicó que para efectos de una evaluación integral se tomaría en consideración todo el trabajo realizado a lo largo de cada curso, de modo tal que la dedicación y la actitud hacia la materia sería igualmente un factor en la calificación final.

De este modo, efectivamente se aplicaron dos exámenes parciales y un examen final, que fueron calificados y promediados de la manera tradicional (se presentan modelos de estos exámenes en los Anexos de este trabajo). Pero no fue esta la única vía para asignar una calificación final, ya que a lo largo del curso se observó y tomó en consideración el desempeño de cada estudiante dentro de las clases en lo que fue su participación en la solución de los problemas presentados en clase, en su capacidad de abstracción y deducción al momento de generar algún modelo dentro de la clase, a su participación con sus compañeros en el curso. Obviamente esta vía para complementar la evaluación tradicional implicó un mayor trabajo e incluso ciertos criterios más subjetivos que objetivos y acarreó una dificultad extra: el depender directamente de la asistencia regular de los alumnos, por lo que para aquellos estudiantes que decidieron solo presentar los exámenes y desatender el curso la evaluación dependió exclusivamente de lo que mostraran en dichos exámenes, teniéndose en la generalidad resultados no aprobatorios.

Finalmente, la bibliografía más adecuada y completa posible para este curso, enfocado principalmente a aspectos prácticos, fue:

Probabilidad y Estadística. Aplicaciones y métodos.- Canavos, George C. McGraw-Hill. México, 1988.

Estadística.- Hayslett Jr., H. T. Compañía General de Ediciones, S. A. México, 1978.

Estadística Elemental.- Johnson, Robert. Grupo Editorial Iberoamérica. México, 1990.

Estadística Matemática con aplicaciones.- Mendenhall; Scheaffer; Wackerly. Grupo Editorial Iberoamérica. México, 1986.

Técnicas de Conteo.- Olivera Salazar; Zúñiga Barrera. Noriega Editores. México, 1987.

Estadística.- Spiegel, Murray R. McGraw-Hill. Colombia, 1969.

Estadística para Administración y Economía.- Stevenson, William. Editorial Harla. México, 1994.

¿De cuántas formas?.- Vilenkin. Editorial Mir. URSS, 1972.

Estadística y Probabilidad.- Walpole; Myers. McGraw-Hill.

Introducción

Razonamiento Estadístico.- Williams, Frederick. Editorial Nuevo Mundo.

Estadística.- Yamane, Taro. Editorial Harla. México, 1979.

ESTADISTICA I

Unidad I, Estadística Descriptiva

Objetivo:
Que el alumno identifique los elementos descriptivos, numéricos o gráficos, de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.
Que el alumno calcule los elementos descriptivos numéricos de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.
Que el alumno trace los elementos descriptivos gráficos de una muestra (o de una población), tanto para presentaciones de datos no agrupados como de datos agrupados.

Como estrategia didáctica se recabarán una serie de datos entre los mismos estudiantes y se generará la necesidad de ordenar y presentar estos datos, y a partir de esto crear la necesidad de elementos que nos describan de manera precisa el comportamiento de los datos recabados. Los modelos para calcular estos elementos descriptivos se irán creando en el aula a partir de las aportaciones que hagan los estudiantes y la dirección que dé el profesor, bajo la idea de calcular primero el valor en particular para el elemento descriptivo buscado y generalizándolo después a partir de los procesos seguidos para su obtención. Los gráficos que se abordarán en la unidad serán generados de la misma manera, a partir de generar en el aula la necesidad de representar distintos comportamientos de la muestra de formas diferentes a la puramente numérica. Posteriormente los estudiantes resolverán ejercicios en el aula.

1. Estadística descriptiva

a) Datos no agrupados y datos atípicos

En este primer tema se trabajó con los elementos de Estadística Descriptiva, tanto para datos agrupados como no agrupados. Todos los temas enumerados en el temario se enseñaron a través de ejemplos prácticos y que fueran generados a partir de datos tomados de cuestiones lo más cercanas posibles a la vida cotidiana de los alumnos.

De esta manera se inició el tema para datos no agrupados, aclarando que más adelante se explicaría la diferencia de estos con los datos agrupados, y que se irían aprendiendo los distintos elementos descriptivos sobre ejemplos prácticos.

Así, comenzamos con el siguiente ejercicio: se preguntó a cada uno de los presentes en el salón por su número de hermanos (de modo que la Estadística se presentara dentro de su cotidianeidad), y se obtuvieron como respuestas:

7, 5, 5, 3, 4, 3, 3, 3, 6, 5, 6, 4, 5, 4, 5, 4, 1, 9.

Luego se presentó la necesidad de ordenar los datos obtenidos y la conveniencia de hacerlo del menor al mayor valor pensando en la finalidad de mostrar con más claridad el comportamiento de la muestra respecto a la característica número de hermanos. Pero el realizar una lista ordenada podría ser poco práctico pues

1
3
3
3
3
4
4
4
4
5
5
5
5
5
6
6
7
9

tiene el inconveniente de ser un tanto extensa, cuestión mucho más problemática en el caso de tener muestras grandes, como usualmente sucede en la Estadística. Por ello se recurre a una presentación un poco más práctica, consistente en anotar una sola vez cada valor numérico observado de la variable (número de hermanos en este caso) junto con el número de veces que este valor se repite:

x	**f**
1	1
2	0
3	4
4	4
5	5
6	2
7	1
8	0
9	1

lo que permite presentar los datos en una forma más reducida y cómoda. Y en este punto se definieron la columna marcada por **x** como la de variable puesto que en ella se describen todos los valores que se observaron para la variable de interés (**número de hermanos**), y la columna marcada por **f** como de frecuencias absolutas, ya que indica precisamente la frecuencia con la que se repite cada dato, y el cuadro como tabla de distribución de frecuencias, ya que indica en que cantidades de observaciones se repartieron o distribuyeron todos los datos recabados. Además, esta columna, al sumarse todos los datos, nos indicará el número de observaciones totales que se realizaron, valor que se conoce como tamaño de la muestra y que se simboliza por **n**. También se aprovechó esta tabla para presentar una cuestión más: cuando se presentan datos con frecuencia absoluta igual a cero, como son las observaciones 2 y 8 en el ejemplo, éstas pueden eliminarse de la presentación de datos:

x	**f**
1	1
3	4
4	4
5	5
6	2
7	1
9	1
Σ	
	18

donde puede verse que hay un dato, el 9, que tiende a "dispararse" del comportamiento, aclarando que cuando esto llega a suceder, sea por un dato demasiado pequeño o demasiado grande, a este dato se le denomina como ***dato atípico***.

b) Medidas de tendencia central

Dado que la mayoría de las observaciones se concentran entre los valores 3, 4 y 5, podríamos buscar una serie de valores que permitieran presentar esta característica en nuestra muestra de un modo más preciso, y que para ello precisamente es que se tienen las llamadas medidas de tendencia central (llamadas así por mostrarnos los valores hacia los cuales "tienden a centralizarse" los valores de nuestras observaciones): media aritmética o promedio, mediana y moda.

Y así, en primera instancia se preguntó al grupo la manera en que se calcula una media o promedio, recibiendo como respuesta que es la suma de todos los valores obtenidos dividida entre el total de observaciones realizadas. De este modo, y simbolizando la media como $\bar{x}$, se vio que bastaría con sumar los 18 valores observados y dividir el resultado entre 18, pero que la manera más práctica de hacerlo consiste no en sumar uno a uno, sino en multiplicar cada observación por su respectiva frecuencia absoluta y luego sumar estos resultados, cosa sencilla de realizar en la tabla de distribución de frecuencia añadiendo una columna nueva calculada como el producto de las observaciones x por las frecuencias absolutas f:

x	**f**	**fx**
1	1	1
3	4	12
4	4	16
5	5	25
6	2	12
7	1	7
9	1	9
Σ		
	18	82

con lo que la media sería $\bar{x} = \frac{82}{18} = 4.55$, dato que para efectos un poco más prácticos en el ámbito informativo, es válido redondear a un valor entero, como $\bar{x} \approx 5$, aunque para efectos de análisis se pediría mantenerlo tal y como se obtenga del cálculo.

Y en forma general definimos como fórmula para el cálculo de la media:

$$\bar{x} = \frac{\sum fx}{n}.$$

Para la mediana nuevamente se preguntó al grupo qué representa o cómo se calcula esta medida encontrándose como respuesta vaga que es el dato de la mitad, aclarando entonces que es el dato bajo el cual se encuentra el 50% de las observaciones realizadas y a partir del cual se encuentra el restante 50%, siempre y cuando estas se hayan ordenado de manera creciente. En este sentido, cuando se tiene que n (el número de observaciones) es un número impar podemos hallar la mediana efectivamente como el valor justo a la mitad de la "lista ordenada" de observaciones, y cuando se tiene una n par entonces la mediana será el promedio de las dos observaciones de la mitad de la lista, por lo que en este caso específico nuestra mediana sería el promedio entre las observaciones ordenadas novena y décima. Igualmente se comentó que la utilidad de esta medida consiste principalmente en la verificación de qué tanto datos atípicos pueden alterar considerablemente el valor de la media como una medida centralista. Para efectos de simbolización, la mediana se representaría en adelante como $\tilde{x}$. Para el cálculo se crea una columna que denominamos de frecuencias absolutas acumuladas, y que se genera primero repitiendo la primera frecuencia absoluta (1), y luego agregando cada vez la siguiente frecuencia absoluta de manera sucesiva hasta completar el total de los datos (18) en la última observación:

x	f	fx	f.a.a.
1	1	1	1
3	4	12	5
4	4	16	9
5	5	25	14
6	2	12	16
7	1	7	17
9	1	9	18
Σ	18	82	

y en la que puede verse que la novena observación ordenada ocurre en el valor de variable 4, y que la décima observación ordenada ocurre en el primero de los cinco valores 5, por lo que la mediana aquí será:

$$\tilde{x} = \frac{4+5}{2} = 4.5,$$

valor no muy alejado de la media, lo que nos lleva a concluir que los posibles datos atípicos en la muestra no alteran el carácter centralista de la media.

Finalmente, al preguntar al grupo sobre la moda, se coincidió en que es el valor que más se repite, o bien aquel valor de variable que presenta la frecuencia absoluta mayor. Se planteó la simbolización de la moda con $\hat{x}$, y para el ejercicio en particular la moda será:

$$\hat{x} = 5,$$

ya que cinco hermanos es el valor que aparece más veces.

c) Proporciones

La siguiente consideración fue que generalmente resulta conveniente presentar los correspondientes porcentajes para cada observación con respecto al total de la muestra, porcentajes que pueden calcularse y expresarse con sólo dividir la frecuencia absoluta de cada observación entre n (es decir, calculando una proporción):

x	f	fx	f.a.a.	f.r.
1	1	1	1	0.056
3	4	12	5	0.222
4	4	16	9	0.222
5	5	25	14	0.278
6	2	12	16	0.111
7	1	7	17	0.056
9	1	9	18	0.056
Σ	18	82		1.000

lo que nos genera una nueva columna a la que llamaremos de frecuencias relativas, y que invariablemente cumplirá con la condición de sumar 1 -o 100%-, con la salvedad de pequeñas pérdidas o ganancias por redondeo.

Ahora bien, del mismo modo en el que con las frecuencias absolutas se presentó la posibilidad de acumularlas, podemos generar una columna en la que se presenten estos porcentajes también de forma acumulada:

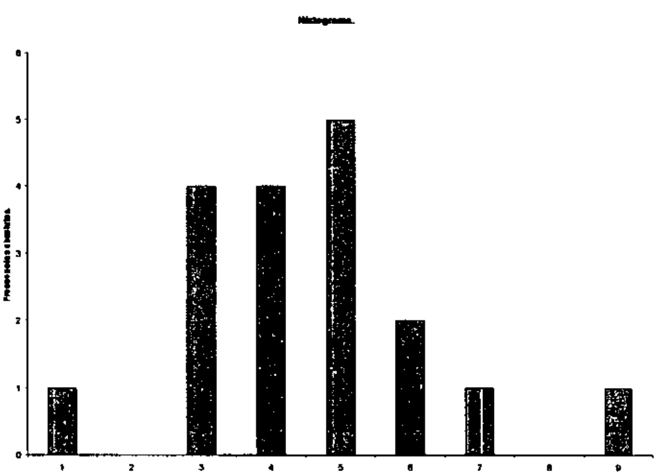
x	f	fx	f.a.a.	f.r.	f.r.a.
1	1	1	1	0.056	0.056
3	4	12	5	0.222	0.278
4	4	16	9	0.222	0.500
5	5	25	14	0.278	0.778
6	2	12	16	0.111	0.889
7	1	7	17	0.056	0.944
9	1	9	18	0.056	1.000
Σ	18	82		1.000	

Así, con estas dos últimas columnas podrá manejarse la información de la muestra, sea para cada observación o de manera acumulada, en términos porcentuales o de proporciones.

d) Gráficos

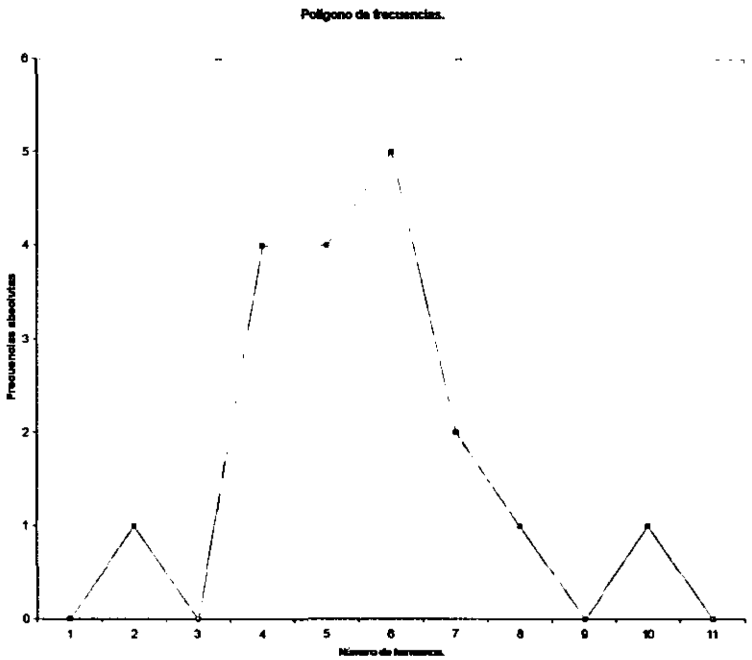
Ahora, el siguiente punto a considerar fue el de darle a los datos y resultados obtenidos una presentación un poco más accesible: los gráficos: Y de ellos se manejarán el gráfico de barras o histograma, el polígono de frecuencias, la ojiva, la ojiva porcentual y el gráfico circular (también llamado de pastel o de secciones).

Los dos primeros gráficos, el histograma y el polígono de frecuencias, nos muestran como están distribuidas todas nuestras observaciones para cada valor de variable, por lo que generaremos estos gráficos sobre la base de los valores de variable contra sus respectivas frecuencias absolutas. Para el histograma, trazaremos una recta vertical desde cada valor de variable y cuya altura corresponda a su frecuencia absoluta:



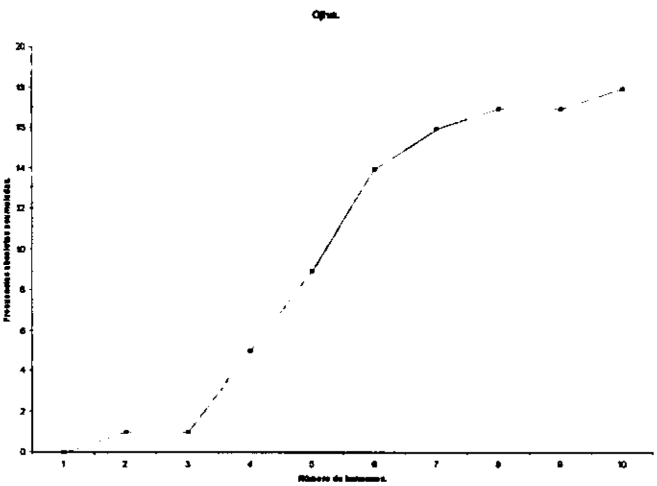
con lo que queda generado el gráfico. De este modo tenemos una presentación aún más sencilla que la numérica de cómo se presenta en la muestra la característica número de hermanos.

En el caso del polígono de frecuencias, el gráfico se obtiene uniendo mediante rectas los puntos cuyas coordenadas son *número de hermanos* y *frecuencias absolutas*, y que corresponden con la parte más alta de cada recta en el histograma, y cerrando el polígono con las hipotéticas observaciones anterior a la primera y posterior a la última, cuyas frecuencias absolutas serán de cero:



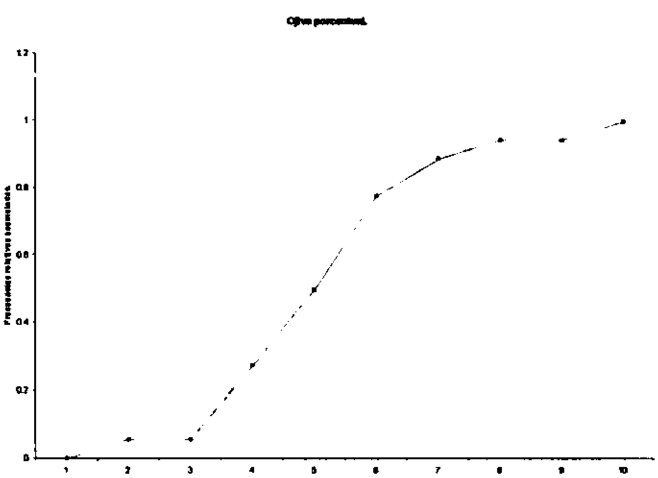
con lo que nos queda una representación de cómo se distribuyeron las 18 observaciones desde el menor hasta el mayor valor de la variable de estudio. Además se hizo el comentario de que el área bajo la curva generada corresponde al 100% de las observaciones hechas, esto con el propósito de facilitar más adelante la distribución Normal Estandarizada.

Los dos siguientes gráficos a tratar, la ojiva y la ojiva porcentual, servirán para mostrar la forma en que la muestra se fue completando de forma ordenada. Para el caso de la ojiva, tenemos que este gráfico muestra el crecimiento en términos absolutos, por lo que graficaremos los valores de la variable contra las frecuencias absolutas acumuladas punto por punto, y después uniremos estos con una serie de rectas, para finalmente cerrar el trazado en su inicio contra la hipotética observación anterior a la primera cuya frecuencia absoluta acumulada es de cero:



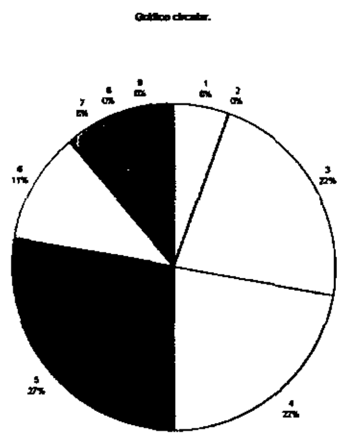
con lo que podemos observar que las observaciones que más contribuyeron a completar la muestra estuvieron entre 3 y 6 hermanos, ya que es en dónde la curva muestra un comportamiento más vertical, mientras que en aquellas partes en la que su comportamiento es más horizontal puede asumirse que se tuvieron pocas observaciones.

En el caso de la ojiva porcentual, este gráfico muestra también el crecimiento que describe la ojiva, pero en términos porcentuales, con lo que el gráfico se trazará con los valores de la variable contra las frecuencias relativas acumuladas, y siguiendo el mismo proceso visto en la ojiva:



donde puede apreciarse en que porcentaje contribuye cada valor de variable para completar el 100% de elementos de la muestra.

El último gráfico, el gráfico circular o de pastel, es una representación muy común y muy simple de las frecuencias relativas de cada valor de variable. Para esto se traza una circunferencia que será dividida en ángulos de diferentes medidas. Esta circunferencia encierra una superficie total que representa el 100% de las observaciones realizadas, y será "rebanada" en partes que llevarán una cierta parte porcentual de dicha superficie total, parte porcentual equivalente a cada frecuencia relativa observada, “rebanadas” que obtenemos multiplicando 360 por cada frecuencia relativa. Una vez obtenidas las secciones, sólo se requiere anotar el porcentaje que representa cada una y el valor de variable al que corresponde:



lo que da una presentación muy clara de la distribución de frecuencias relativas para cada valor de variable.

e) Medidas de dispersión y la Regla Empírica

Los siguientes elementos descriptivos que se trataron fueron las medidas de dispersión. En este apartado se comentó con el grupo que se manejarían sólo tres medidas de esta naturaleza, la amplitud o rango, la desviación estándar y la variancia.

Primeramente se estableció que el concepto de error dentro de Estadística corresponde a que tan desviados de un valor fijo pueden aparecer un valor de variable o una serie de ellos. Después se comentó que las medidas de dispersión sirven para medir que tan grandes o que tan pequeños pueden ser los errores que los datos de una muestra tengan de forma global respecto a alguna medida centralista, y que en especial trabajaríamos con la media como dicha medida centralista.

Ahora, en lo tocante a la primera de las medidas de dispersión, la amplitud, se dijo que en sí no se le podía considerar como una medida de dispersión sino como un dato auxiliar en la comprensión de aquéllas, ya que la amplitud simplemente nos señala el espacio numérico que ocupan todas las observaciones y por ende el espacio en el que van a "caber" todos los errores respecto a alguna medida de tendencia central. Para el cálculo basta con tomar el valor de variable mayor y restarle el valor de variable menor, así en nuestro ejercicio la amplitud, simbolizada como **Amp** será

Amp = 9 - 1 = 8,

lo que indica en términos de dispersión que, en el peor de los casos, una observación podrá estar alejada de una medida centralista, o tener un error respecto a ella, en cuando mucho 8 unidades. De forma general podemos decir que:

Amp = Mayor valor de variable - Menor valor de variable.

Para las siguientes medidas de error se planteó por una parte que la media se había obtenido al repartir equitativamente el total de elementos observados entre el número de observaciones hechas, y por otra, que sería lógico, si se habían repartido equitativamente las observaciones, que también los errores y las desviaciones se repartieran equitativamente, y que este reparto se haría de los errores con respecto a la media. Así, se procedió a medir los errores o diferencias con respecto a la media haciendo $x_i - \bar{x}$, en donde x_i representa cada uno de todos los valores recabados, pero para evitar el cálculo repetido de estas diferencias en el caso de valores que aparecían más de una vez podría hacerse con $f(x_i - \bar{x})$. Después se planteó que el reparto equitativo de todas estas diferencias tendría que hacerse primero sumándolas y luego dividiéndolas entre el total de observaciones. Así, en la tabla de distribución de frecuencias:

x	f	fx	f.a.a.	f.r.	f.r.a.	$x_i - \bar{x}$	$f(x_i - \bar{x})$
1	1	1	1	0.056	0.056	-3.55	-3.55
3	4	12	5	0.222	0.278	-1.55	-6.20
4	4	16	9	0.222	0.500	-0.55	-2.20
5	5	25	14	0.278	0.778	0.45	2.25
6	2	12	16	0.111	0.889	1.45	2.90
7	1	7	17	0.056	0.944	2.45	2.45
9	1	9	18	0.056	1.000	4.45	4.45
Σ	18	82		1.000			≈0

en donde tenemos que la suma de $f(x_i - \bar{x})$ es casi cero, y de hecho, tal y como se planteó al grupo, en términos reales es exactamente igual a cero, solo que en este caso fue un valor aproximado porque 4.55 es ya en sí un valor aproximado de $\bar{x}$. Aquí se explicó que este resultado se daba de así no porque no se tuvieran errores, sino por que al ser la media un reparto equitativo, los errores por valores menores a ella y por valores mayores a ella tendían a equilibrarse en forma justa. Para resolver el problema de cómo pueden medirse los errores totales en la muestra para repartirlos, se vieron dos alternativas para eliminar los valores negativos de las diferencias: una consistente en calcular el valor absoluto de las diferencias antes de sumarlas (lo que llevaría a la medida conocida como desviación de la media), y otra consistente en elevar al cuadrado las diferencias antes de sumarlas, lo que nos forzaría a calcular luego la raíz cuadrada del resultado final para regresar a los términos lineales en los que están dados los errores. Se tomó la segunda opción, y así, previo a multiplicarlas por sus frecuencias absolutas, se elevaron las diferencias al cuadrado; esto es, primero se hizo $(x_i - \bar{x})$, después $(x_i - \bar{x})^2$, y finalmente $f(x_i - \bar{x})^2$, resultados que se sumarían para posteriormente repartirlos; en la tabla de distribución de frecuencias:

x	f	fx	f.a.a.	f.r.	f.r.a.	$x_i - \bar{x}$	$f(x_i - \bar{x})$	$f(x_i - \bar{x})^2$
1	1	1	1	0.056	0.056	-3.55	-3.55	12.6025
3	4	12	5	0.222	0.278	-1.55	-6.20	9.6100
4	4	16	9	0.222	0.500	-0.55	-2.20	1.2100
5	5	25	14	0.278	0.778	0.45	2.25	1.0250
6	2	12	16	0.111	0.889	1.45	2.90	4.2050
7	1	7	17	0.056	0.944	2.45	2.45	6.0025
9	1	9	18	0.056	1.000	4.45	4.45	19.8025
Σ	18	82		1.000			≈0	54.4450

Teniendo la suma de $f(x_i - \bar{x})^2$, se dividió sobre n=18, obteniéndose

$$\frac{54.4450}{18} = 3.024722$$

aproximadamente. En este punto solo resta calcular la raíz cuadrada de este valor para "volver" a las medidas lineales de error. Pero antes de hacerlo se presentó al grupo este resultado como la medida de dispersión llamada variancia, y a la que podemos definir como la suma de los cuadrados de las diferencias de las observaciones con respecto a la media, dividida por el total de observaciones. Esta medida se representará por s^2 o por σ^2, y para calcularla podemos plantear la expresión

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n},$$

aunque aquí se hizo al grupo la siguiente aclaración: esta medida tiene ciertas características que exigen de un pequeño ajuste, consistente en dividir no sobre n, sino sobre n-1, si se desea que el cálculo sea más exacto cuando se tiene la condición de que la muestra sobre la que se realiza el cálculo consta de 30 o más elementos, lo que se conoce como muestra grande. Se planteó también que las razones corresponden a cuestiones de Estadística más avanzada que la de bachillerato, por lo que no se haría una demostración más formal del por qué de este ajuste. De este modo se tienen para la variancia las expresiones

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n}, \text{ si } n < 30$$

y

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n-1}, \text{ si } n \geq 30.$$

En el caso de nuestro ejercicio se tiene que n=18<30, por lo que la división realizada previamente sobre 18 nos da un resultado válido para la variancia.

Ahora regresamos aquí a la repartición del error. La última operación por realizar para hallar el tamaño del "error promedio" respecto a la media para cada observación había sido planteada como la raíz cuadrada de lo que acababa de definirse como variancia. Así, denotando este promedio con s o σ, tenemos que

$$s = \sqrt{3.024722} = 1739173$$

aproximadamente, y que podemos interpretar como el "promedio" del error con respecto a la media; es decir, cada una de nuestras observaciones estará desviada de la media por 1.739173 unidades en promedio, aproximadamente. Dado que esta medida reparte o estandariza los errores para todas las observaciones, se le conoce como error estándar, desviación típica, error típico o desviación estándar, y es la medida de error de uso más frecuente en Estadística. De esta forma pueden darse dos expresiones para la desviación estándar: una para cuando la muestra con la que se trabaje sea de 30 o más elementos, y otra para cuando dicha muestra conste de menos de 30 elementos:

$$s = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n}}, \text{ si } n < 30$$

y

$$s=\sqrt{\frac{\sum f(x_i-\bar{x})^2}{n-1}}\text{, si } n\geq 30.$$

Existe otra característica que mencionar con respecto a la desviación estándar, y es la llamada Regla Empírica. Esta regla nos dice que si a la media se le resta y se le suma una vez el valor de la desviación estándar, el intervalo generado contendrá a aproximadamente el 68% de la muestra; si restamos y sumamos a la media dos veces la desviación estándar, se generaría un intervalo en el que estará contenido aproximadamente al 95% de la muestra; si restamos y sumamos a la media tres veces la desviación estándar, el intervalo contendrá aproximadamente al 99% del total de la muestra; y finalmente, y para efectos prácticos, la Regla Empírica dice que al restar y sumar de la media tres y media veces la desviación estándar, se generará un intervalo que contenga al total de los elementos de la muestra. Para concluir con la Regla Empírica, se dijo que esta es de mucha importancia en el manejo de ciertos elementos de Probabilidad.

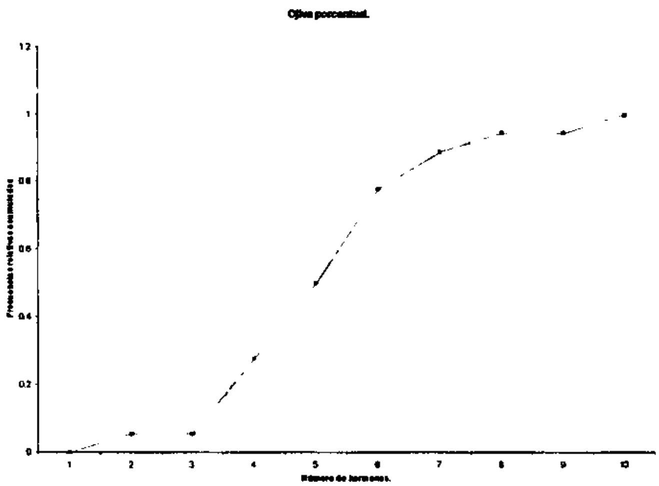
Para finalizar el apartado referente a las medidas de dispersión, se comentó que es el error precisamente lo que le da razón de ser a la Estadística, ya que si no hubiera desviaciones con respecto de algún valor, entonces todos los elementos de una población serían iguales y no sería necesario ningún análisis estadístico para llegar a conocer el comportamiento de dicha población con respecto a alguna característica en particular.

f) Cuantiles

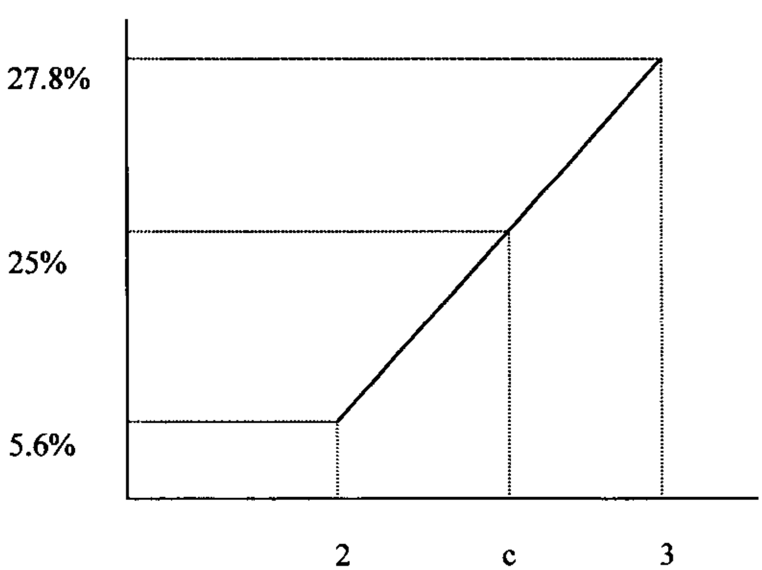
Las siguientes medidas descriptivas a tratar fueron las llamadas medidas de posición o *cuantiles*. Para ello se recordó que la columna de frecuencias relativas acumuladas nos muestra el porcentaje del total de la muestra que se acumula para cada uno de los valores de variable ordenados en forma creciente, pero que en dicha columna estamos limitados a los porcentajes que allí se establecen y no tenemos acceso directo a los valores de variable bajo los cuales se acumulan algunos porcentajes en particular, obtenidos a partir de seccionar la muestra ordenada en partes equitativas. Así, se presentaron las divisiones porcentuales que podrían parecer más usuales o más cómodas y se dijo que los valores de variable que seccionaran en esos porcentajes preestablecidos a la muestra serían precisamente los cuantiles, llamados medidas de posición en el sentido en que son los valores de la variable que están situados o posicionados precisamente en un determinado punto porcentual de la "lista" ordenada que pueda hacerse con cada una de las observaciones de la muestra. Las partes en las que se encontró práctico seccionar la muestra fueron en cuartas, quintas, décimas y centésimas, lo que permitió clasificar a los cuantiles en *cuartiles*, *quintiles*, *deciles* y *centiles* respectivamente. En términos porcentuales, el problema es calcular bajo qué valor de variable se tienen acumulados los siguientes porcentajes:

- En el caso de los cuartiles el 25%, el 50%, el 75% y el 100%, que son respectivamente el primero, segundo, tercer y cuarto cuartil.
- En el caso de los quintiles el 20%, el 40%, el 60%, el 80% y el 100%, que son el primer, segundo, tercer, cuarto y quinto quintil respectivamente.
- En el caso de los deciles el 10%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90% y el 100%, que son el primer, segundo, tercer, cuarto, quinto, sexto, séptimo, octavo, noveno y décimo decil respectivamente.
- En el caso de los centiles el 1%, el 2%, y hasta el 99% y el 100%, que son respectivamente el primer, segundo, y hasta el nonagésimo noveno y centésimo centil.

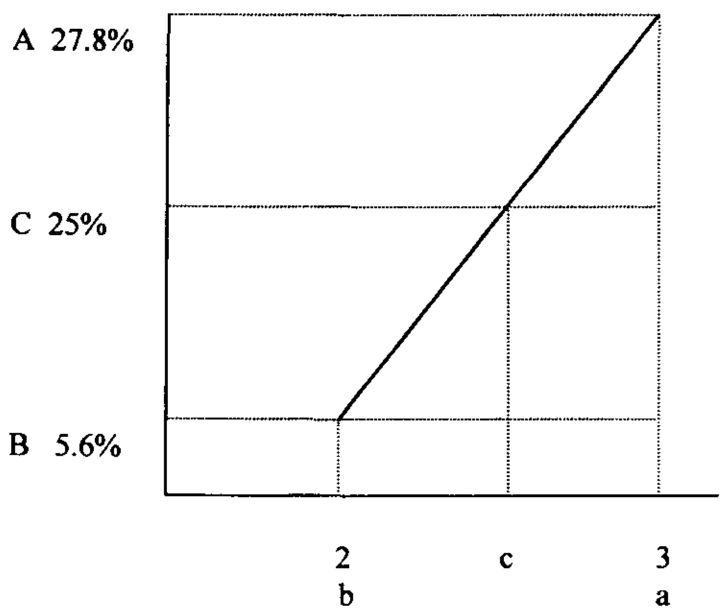
Para el cálculo del valor de variable bajo el que se acumula un determinado porcentaje de la muestra se planteó la existencia de un proceso estándar funcional para cualquier porcentaje -y consecuentemente para cualquier cuantil- que se conoce como interpolación. Dicho proceso fue explicado sobre la gráfica que involucra a los porcentajes acumulados: la ojiva porcentual:



El grupo se decidió por el cálculo del primer cuartil, así que primeramente se hizo notar que se estaba buscando la observación bajo la cual se tiene acumulado el 25% de las observaciones, y que, dado que bajo el valor 2 se tiene acumulado el 5.6% de la muestra y bajo la observación 3 se tiene el 27%, esta observación buscada tendría que ser un valor entre 2 y 3 (e incluso más cercano a tres). Luego se tomó la porción de la gráfica en la que se presentan estos datos y se hizo ver que del mismo modo en el que 25% se localiza entre 5.6% y 27.8%, el primer cuartil, denotado por **c**, estaría efectivamente entre 2 y 3:



Luego se simbolizaron los valores 2 y 3 (aquellos entre los que se localiza c) como b al menor (2), y como a al mayor (3), y sus correspondientes porcentajes acumulados con las mayúsculas A para el porcentaje de a, B para el porcentaje de b y C para el porcentaje de c, que es el porcentaje que marque el cuantil que se esté buscando. Aquí se subrayó que, dado que se generaría un modelo matemático de uso general con este desarrollo, la notación recién presentada no debería modificarse en lo más mínimo. También conjuntamente con la simbolización se extendieron las proyecciones de B y de C, formándose así dos triángulos rectángulos de bases paralelas:



En este punto se recordó con el grupo el Teorema de Tales, que plantea el hecho de que al seccionar un triángulo cualquiera con una recta paralela a alguno de sus lados, el nuevo triángulo generado será semejante con el original. De este modo el triángulo de vértices (2, 5.6%), (3, 27.8%) y (3, 5.6%) será semejante con el triángulo de vértices (c, 25%), (3, 27.8%) y (3, 25%), por lo que las proporciones que guarden entre sí los lados correspondientes serán las tres iguales. En este caso nos interesan sólo las proporciones entre las bases y las alturas, por lo que se dedujo que la base del triángulo mayor está dada por 3-2 -o **a** - **b**- y su altura por 27.8-5.6 -o **A** - **B**- obviando los porcentajes, y que la base y la altura del triángulo menor son c-2 -o **a** - **c**- y 27.8-25 -o **A** - **C**- respectivamente, y retirando también la notación porcentual. Luego, ya que los triángulos son semejantes, podemos decir que $\frac{27.8-5.6}{27.8-25}=\frac{3-2}{3-c}$, con lo que nos queda una ecuación de primer grado con una incógnita, incógnita que es precisamente el cuantil que estamos buscando. Así, despejando, tenemos que

$c = 2.874$,

que es el valor del primer cuartil y que, como se había intuido, es un valor entre 2 y 3, y más cercano a 3.

Ahora, generalizando el modelo recién planteado con la notación previamente establecida, se tiene que el cuantil **c** puede calcularse sobre la base del modelo

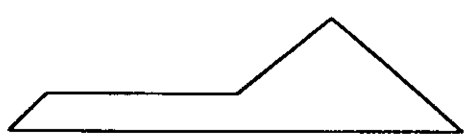
$$\frac{A-B}{A-C}=\frac{a-b}{a-c}$$

recalcando de nuevo que es indispensable no alterar el orden el que se asignaron las literales a los valores de variable ni a los porcentajes, a riesgo de obtener resultados distintos al cuantil buscado.

g) Coeficiente de sesgo y coeficiente de apuntalamiento

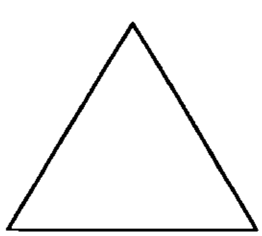
Los elementos restantes que se aprendieron en esta unidad fueron los coeficientes de sesgo y de apuntalamiento, elementos con los que se pretendía dar una descripción más teórica de cómo se presenta la distribución de frecuencias en la muestra con respecto a las medidas centralistas.

En el primer caso, con el coeficiente de asimetría o sesgo, se tendría una medida que nos indicaría si nuestra distribución de frecuencias tendía a agruparse más hacia alguno de los extremos de la muestra o hacia el centro de ésta. Para ello debería calcularse el valor numérico del coeficiente, denotado por α_3, y después compararlo contra cero para decidir si se tiene una distribución *asimétrica negativa* (si $\alpha_3 < 0$), *asimétrica positiva* (si $\alpha_3 > 0$) o *simétrica* (si $\alpha_3 = 0$). Esto se explicó en términos gráficos con bocetos de polígonos de frecuencias ilustrando los tres casos:



$\alpha_3 < 0$

$\alpha_3 > 0$



$\alpha_3 = 0$

en donde se puede apreciar que en los dos primeros casos se tiene un sesgo hacia alguno de los extremos de la distribución, o, dicho de otro modo, se tiene una concentración mayor de observaciones hacia uno de los extremos de la distribución, mientras que en el tercer caso se nota una distribución más pareja u homogénea con respecto a los extremos.

Para efectos del cálculo se presentó directamente la fórmula para el coeficiente de asimetría como:

$$\alpha_3 = \frac{\sum f(x_i - \bar{x})^3}{\left[\sum f(x_i - \bar{x})^2\right]^{3/2}},$$

dejando la explicación del por qué de esta expresión de lado, ya que el concepto de momentos de la media estaba fuera del conocimiento de los alumnos.

La fórmula se explicó apoyándola en el desarrollo en la tabla de distribución de frecuencias (ya sin anotar más que las columnas necesarias):

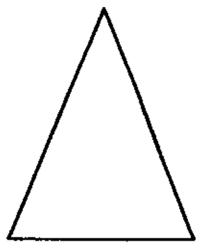
x	f	$(x_i - \bar{x})$	$f(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$f(x_i - \bar{x})^3$
1	1	-3.55	12.6025	-44.738875	-44.738875
3	4	-1.55	9.6100	-3.723875	-14.8955
4	4	-0.55	1.2100	-0.166375	-0.6655
5	5	0.45	1.0125	0.091125	0.455625
6	2	1.45	4.2050	3.048625	6.09725
7	1	2.45	6.0025	14.706125	14.706125
9	1	4.45	19.8025	88.121125	88.121125
Σ	18		54.4450		49.08025

Con lo que el coeficiente de asimetría sería:

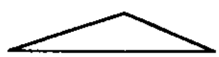
$$\alpha_3 = \frac{49.08025}{[54.4450]^{3/2}} = \frac{49.08025}{\sqrt{(54.4450)^3}} = 0.122171459748 \approx 0.122171$$

lo que indica que nuestra muestra tiene una distribución asimétrica positiva.

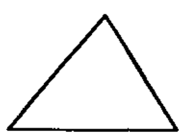
Para el segundo caso, el del coeficiente de apuntalamiento, se especificó que éste nos señala que tan grandes o qué tan pequeñas son las frecuencias absolutas en términos de las medidas centralistas, ó, dicho de otra forma, qué tan homogénea o que tan heterogénea es nuestra muestra en términos de frecuencias absolutas. Aquí denotaremos el coeficiente de apuntalamiento, también llamado curtosis, por α_4 y, de la misma forma que lo hicimos con α_3, tendremos que comparar su valor contra una constante para verificar en cuál de tres casos posibles de curtosis cae nuestra distribución, y esta constante será 3. Así, cuando $\alpha_4 > 3$ tendremos una distribución con demasiadas observaciones en los valores más cercanos a las medidas centralistas (gráficamente una distribución muy apuntalada o "picuda"), y la llamaremos leptocúrtica; cuando $\alpha_4 < 3$ tendremos una distribución muy dispersa en lo que a las frecuencias absolutas se refiere (gráficamente una distribución poco apuntalada o llana), y la llamaremos platicúrtica; y cuando $\alpha_4 = 3$ tendremos una distribución con una dispersión (y concentración) mediada (gráficamente una distribución ni muy llana ni muy apuntada), y la llamaremos mesocúrtica. En términos gráficos de nueva cuenta se presentaron los tres bocetos de los correspondientes polígonos de frecuencia:



$\alpha_4 > 3$



$\alpha_4 < 3$



$\alpha_4 = 3$

Nuevamente la fórmula para el cálculo de este valor se dio directamente, ésta como:

$$\alpha_4 = \frac{\sum f(x_i - \bar{x})^4}{\left[\sum f(x_i - \bar{x})^2\right]^2},$$

replanteando que el proceso de cálculo es análogo al de α_3. Para el cálculo nuevamente recurrimos a la tabla de distribución de frecuencias, y tenemos (nuevamente reduciendo la tabla a las columnas estrictamente necesarias):

x	f	$(x_i - \bar{x})$	$f(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$	$f(x_i - \bar{x})^4$
1	1	-3.55	12.6025	158	158.82300625
3	4	-1.55	9.6100	-3.723875	23.088025
4	4	-0.55	1.2100	-0.166375	0.366025
5	5	0.45	1.0125	0.091125	0.20503125
6	2	1.45	4.2050	3.048625	8.8410125
7	1	2.45	6.0025	14.706125	36.03000625
9	1	4.45	19.8025	88.121125	392.13900625
Σ	18		54.4450		619.4921125

de donde tenemos que el coeficiente de apuntalamiento es:

$$\alpha_4 = \frac{619.4921125}{[54.4450]^2} = 0.208987243106 \approx 0.208987,$$

lo que nos indica que tenemos una distribución platicúrtica.

Finalmente se planteó que para dar una información completa debería anotarse como título de la tabla la característica que se estaba presentando, en este caso "Número de hermanos por alumno en el grupo", o algo semejante.

En este punto se han completado todos los elementos descriptivos a tratar en esta unidad. Lo siguiente fueron ejercicios para el grupo con datos tomados nuevamente de entre los mismos alumnos sobre la base de cuestiones como el número de letras en sus nombres de pila, el número de monedas que tuvieran en ese momento, el último dígito de su número de cuenta o algunos ejercicios ya diseñados, como:

-Ejercicio:

x	f
0	1
1	3
2	8
3	5
4	3

Terminados estos ejercicios se procedió a enseñar al grupo los mismos elementos para datos agrupados.

h) Datos agrupados

Para el caso de los datos agrupados se planteó de nueva cuenta al grupo un ejercicio de recopilación, ordenación, presentación y análisis de datos. Una vez más, al igual que en la mayoría de los ejercicios vistos antes, se les encuestó acerca de alguna cuestión en particular, específicamente sobre el tiempo en minutos que ocupaba cada uno en llegar de su casa al plantel. Los datos fueron:

75, 60, 50,30,20,90, 45, 60. 60, 60, 40,40, 30, 90, 40, 45, 75, 75, 30, 50, 10, 75, 90, 40, 90, 40, 65, 110, 80, 75.

Para sortear las dificultades que puede crearnos el hecho de tener demasiados datos y muy heterogéneos entonces se propuso tratar las observaciones como datos agrupados, cuestión que consiste en manejar las observaciones dentro de subconjuntos o intervalos, llamados *clases*. En principio, se deben de crear estos intervalos de manera conveniente, y que puede ser una presentación en clases de entre 5 y 20 clases para que sea tanto manejable como confiable, aunque el número óptimo de clases depende del problema en particular y del criterio de quien lo resolviera. De este modo se halló conveniente agrupar nuestros datos en intervalos de 20 en 20 minutos:

0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
100 - 120

y no más allá, ya que en la última clase caben hasta las observaciones más grandes. Aquí se hizo ver que la última clase podría haberse terminado en 110 (o 112 en aras de dar un margen de error a 110 minutos), y que en general nada obligaba a construir las clases todas del mismo tamaño, pero que para efectos prácticos esto era lo más recomendable.

Una vez construidas las clases registramos sus frecuencias, y para ello se anota el número de observaciones que caen dentro de cada clase como la correspondiente frecuencia absoluta a esta. Así:

Clases	f
0 – 20	2
20 – 40	8
40 – 60	8
60 – 80	7
80 – 100	4
100 - 120	1

queda como nuestra tabla de distribución de frecuencias para los tiempos de recorrido manejada con datos agrupados. Y en este punto pudo hacerse evidente la diferencia entre tablas para valores agrupados y no agrupados.

En este punto se presentaron al grupo las características de las clases un poco más a fondo, con el fin de poder realizar adecuadamente los cálculos de los elementos descriptivos vistos en el tema de *Datos no agrupados*. Primero se vieron los extremos de cada clase, a los que se denominaron como *límite inferior* de clase al valor menor de esta y como *límite superior de clase* al valor mayor, haciendo un paréntesis aquí para explicar dos formas diferentes de construir las clases:

Clases	Clases
4 – 6	4 – 6
7 – 9	6 – 8
10 – 12	8 – 10
13 – 15	10 – 12
16 - 18	12 - 14

En ambos casos podemos observar ventajas y desventajas. El primer caso implica que si se están midiendo resultados que pueden ser cualquier número real, entonces una observación 6.8 o 12.5 o semejante no tendría dónde ser contabilizada, mientras que el segundo caso implica que si se están tomando observaciones que sólo arrojaran valores enteros entonces observaciones como 6, 8 o 10 podrían contabilizarse dentro de dos clases distintas. De este modo, deberá tenerse cuidado en la manera de construir las clases en función de la naturaleza de los resultados que pueda tomar la variable de estudio. Ahora, si se nos presenta una distribución como la primera de las dos anteriores con las clases ya construidas y resulta que los valores de la variable son continuos entonces se recurre a calcular lo que se conoce como límites reales de clase tomando el límite superior de la primera clase y promediándolo con el límite inferior de la clase siguiente, y el resultado pasa a ser el límite real superior de la primera clase y, al mismo tiempo, el límite real inferior de la siguiente clase; de este modo

Clases
4-6
7-9
10-12
13-15
16-18

pasa a ser

Clases
4-6.5
6.5-9.5
9.5-12.5
12.5-15.5
15.5-18

pero el límite superior de la última clase y el límite inferior de la primera clase también deberán ajustarse del mismo modo que los demás, por lo que sumaremos 0.5 al primero y restaremos 0.5 al segundo, en este caso, de modo que

Clases
3.5-6.5
6.5-9.5
9.5-12.5
12.5-15.5
15.5-18.5

pasa a ser nuestra construcción de clases en términos de límites reales, y ofrece ya la posibilidad de contar cualquier valor real entre 3.5 y 18.5.

La segunda característica que se manejó fue el *tamaño o ancho de clase*, elemento que, denotado como **c**, se obtiene para cada clase como
c = límite superior de clase - límite inferior de clase,
y que es el número de unidades que cubre cada clase. Como un ejemplo y tomando una de las clases de nuestro ejercicio tenemos que
c = 60 - 40 = 20.

La tercera característica a manejar tiene que ver directamente con el problema inicial de no tener los valores puntuales requeridos para los cálculos y trazos de los elementos descriptivos vistos con anterioridad, y es la que se conoce como *marca de clase*. Este elemento se obtiene para cada clase promediando sus dos límites, esto es, si la denotamos como M.C.:
M.C. = (límite inferior de clase + límite superior de clase) /2.

Este valor se considera como el valor hipotético que tomaron todas y cada una de las observaciones contenidas en la clase, así, para la clase 60-80, se tiene que

$$\text{M.C.} = \frac{60+80}{2} = \frac{140}{2} = 70,$$

lo que significa que, en teoría, todas y cada una de las 7 observaciones que se dieron en la clase 60 - 80 fueron de valor 70. De esta manera podemos tener valores de variable puntuales y no por intervalo, tomando la marca de clase como los elementos x_i. Esto nos deja la siguiente tabla de distribución de frecuencias:

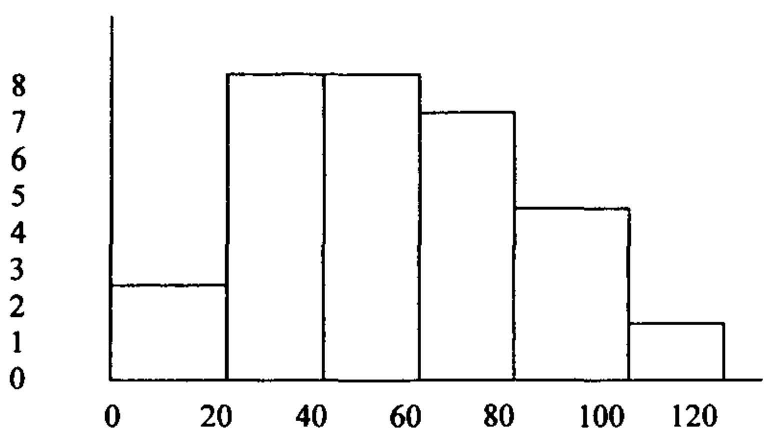
Clases	M.C. x	f
0-20	10	2
20-40	30	8
40-60	50	8
60-80	70	7
80-100	90	4
100-120	110	1

y a partir de aquí casi todos los elementos descriptivos se calcularán y trazarán igual que como se hizo con datos agrupados, con dos excepciones, y otras dos posibles excepciones: la amplitud y el histograma en el primer caso, y la mediana y la moda en el segundo.

Para el caso de la amplitud, nuevamente se manejará la diferencia entre el último valor menos el primero, pero esta vez serán tomados como el límite superior de la última clase y el límite inferior de la primera clase:

Amp = límite superior de la última clase - límite superior de la primera clase,
de modo que
Amp = 120 - 0=120.

En lo que toca al histograma, aquí el trazado se realiza sobre las clases contra sus frecuencias absolutas. El sentido del gráfico es el mismo, pero su formato cambiará ya que ahora en lugar de levantar rectas en cada valor, levantaremos rectángulos cuyos extremos de base serán los límites de cada clase y cuya altura corresponderá igualmente a cada frecuencia absoluta correspondiente a cada clase:



En el caso de que nuestras clases estén construidas de manera que cada límite superior no coincida con el límite inferior de cada clase, las barras del histograma estarán separadas unas de otras. En este caso es recomendable calcular los límites reales de clase y utilizar estos para hacer el gráfico, lo cual nos dejará un trazo igual al anterior.

Por lo que toca a la mediana y a la moda, se dijo que en el caso de los datos agrupados estas podrían manejarse igual que en los datos no agrupados, lo que nos deja una mediana de 50 y una moda doble: de 30 y de 50 (lo que permitió explicar que en una distribución puede haber más de una moda, y que dichas distribuciones se conocen como multimodales), o bien que podrían manejarse sobre la base de fórmulas. Para la mediana la fórmula que se tiene es:

$$\tilde{x} = L_i + \left(\frac{\frac{n}{2} - \sum f_{ant}}{f_{med}} \right) c,$$

en donde

- L_i es el límite real inferior de la clase en la que se encuentra la mediana. En nuestro ejercicio será la clase en la que se acumulan 15 observaciones, la mitad de las 30 hechas, y que es la clase 40 – 60, cuyo límite real inferior es directamente 40.
- n es el tamaño de la muestra. En nuestro ejercicio será 30.
- $\sum f_{ant}$ es la suma de todas las frecuencias absolutas anteriores a la de la clase en la que está la mediana, o bien la frecuencia absoluta acumulada para la clase anterior a la clase de la mediana. En el ejercicio será 10.
- f_{med} es la frecuencia de la clase mediana. Para el problema será 8.
- c es el ancho de la clase mediana. Aquí será de 20.

Así, nuestra mediana calculada por fórmula será:

$$\tilde{x} = 40 + \left(\frac{\frac{30}{2} - 10}{8} \right) 20 = 52.5 .$$

Para la moda se tiene la fórmula

$$\hat{x} = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

en donde

- L_i es el límite real inferior de la clase modal (aquella con la frecuencia absoluta mayor). En nuestro ejemplo tenemos dos clases modales posibles: 20 – 40 y 40 – 600, cuyos límites reales inferiores son 20 y 40 respectivamente.
- Δ_1 es la diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase inmediata anterior. En el ejemplo serán 6 y 10.
- Δ_1 es la diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase inmediata posterior. En el ejemplo serán 0 y 1.
- c es el ancho de la clase mediana. Aquí será de 20.

De esta manera tenemos que las modas para nuestro ejercicio, calculadas por fórmulas, serán:

$$\hat{x}_1 = 20 + \left(\frac{6}{6+0}\right)20 = 40$$

y

$$\hat{x}_2 = 40 + \left(\frac{0}{0+1}\right)20 = 40\,.$$

De cualquier manera, en el caso de estas dos medidas no es estrictamente necesario el uso de las fórmulas.

El cálculo de los restantes elementos descriptivos se dejó como ejercicio al grupo, aclarando que en el caso del gráfico circular es válido también anotar las clases y no las marcas de clase a un lado de cada "rebanada".

Finalmente, se dejaron algunos otros ejercicios al grupo, algunos de los cuales se generaron sobre la base de datos recabados entre los alumnos, y otros preestablecidos como:

-Ejercicios:

Clases	f
50 – 59	2
60 – 69	3
70 – 79	8
80 – 89	5
90 - 99	2

Clases	f
22 – 32	1
33 – 43	2
44 – 54	5
55 – 65	2
66 – 76	9
77 – 87	9
88 – 98	10
99 - 109	5
110 – 120	3
121 - 131	4

En esta unidad de Estadística Descriptiva se sugiere manejar los temas básicamente con ejercicios y sobre práctica, tratando de dar fundamento a todos aquellos elementos en el que los conocimientos previos y la capacidad de abstracción de los alumnos los permitan. El trabajar con valores tomados de entre los mismos alumnos y de cuestiones sencillas, ayuda a relajar un poco la reticencia inicial al trabajo numérico y deja a la Estadística como algo más tangible y cercano a la realidad cotidiana de los estudiantes.

Unidad II, Técnicas de conteo

Objetivo:
Que el alumno identifique los distintos arreglos en el Cálculo Combinatorio y resuelva problemas de conteo.
Que el alumno pueda aplicar elementos propios del Análisis Combinatorio al cálculo de probabilidades.

Como estrategia didáctica se presentará un problema sencillo que permita desarrollar el conteo a partir de diagramas de árbol; luego, al complicar el problema, se buscará deducir de manera informal el Teorema Fundamental del Cálculo Combinatorio. A partir de la necesidad de trabajar con productos, se recordará la operación factorial, y recuperado dicho conocimiento, se procederá a deducir los diferentes arreglos propios del Cálculo Combinatorio a partir de problemas. Posteriormente los alumnos resolverán ejercicios en el aula.

2. Técnicas de Conteo

Para comenzar a estudiar el comportamiento del conteo se partió de la herramienta más elemental con ejemplos sencillos: los diagramas de árbol. Se manejaron y mostraron primero dos ejemplos, en los que los alumnos trazaron los diagramas. Básicamente se buscó no sólo una alternativa para enunciar la Regla del Producto, sino para ir introduciendo las posibilidades de tener o no un orden, y de tener o no la posibilidad de la repetición. A partir de los diagramas de árbol se llegó al Teorema Fundamental del Cálculo Combinatorio, génesis todo lo correspondiente a las técnicas de conteo. Se aclaró también que a este Teorema se le conoce también como Teorema Fundamental de la Aritmética, Teorema del Producto o Regla del Producto.

Así, enunciamos el Teorema como: "Si una cierta acción puede realizarse de p maneras diferentes, y si una segunda acción puede realizarse de q maneras diferentes, entonces el número de formas en las que ambas acciones pueden llevarse a cabo de manera conjunta será **pq**. En caso de buscarse el número de formas en que ambas acciones pueden realizarse de manera excluyente, este será **p + q**. Los dos resultados anteriores pueden hacerse extensivos para n acciones distintas", aclarando los términos "de manera conjunta" y "manera excluyente".

Ahora bien, pueden involucrarse distintas condiciones al momento de plantear un problema de conteo. Y algunas de estas condiciones son de presencia tan usual que se han generado algunos resultados estándar, que se conocen como arreglos y que son: Permutación, Ordenación, Ordenación con Repetición, Ordenación Cíclica, Combinación, Combinación con Repetición, Permutación con Objetos Iguales y Combinación Múltiple, y cuyas expresiones pueden obtenerse basándose en el Teorema Fundamental del Cálculo Combinatorio dependiendo de las características que pudiéramos encontrar.

Antes de continuar con los arreglos, aquí abrimos un pequeño tema de suma utilidad: la Operación Factorial. Esta se denotará como n! (n factorial) y podrá ser aplicada a todos los números naturales y al cero. La operación en sí consiste en tomar el número entero n y multiplicarlo por todos los enteros inferiores a sí hasta 1: $\mathbf{n! = n(n-1)(n-2)\ldots(1)}$, y donde definimos $1!=1$ y $0!=1$.

Vista la Operación Factorial se presentaron al grupo los siguientes arreglos, deduciendo las correspondientes fórmulas a partir de ejemplos:

Permutación -ordenamiento del total de n elementos sin remplazo- cuyo número total de posibles formas de llevarlo a cabo se simboliza por $P_{(n,n)}$, y en el que $\mathbf{P_{(n,n)} = n!}$;

Ordenación –ordenamiento de r elementos tomados de entre un total de n sin remplazo- cuyo número de formas de llevarse a cabo se simboliza por $O_{(n,r)}$, donde $\mathbf{O_{(n,r)} = n(n-1)\ldots(n-r+1)}$ o $\mathbf{O_{(n,r)}} = \frac{n!}{(n-r)!}$.

Ordenación con Repetición –ordenamiento de r elementos tomados con reemplazo de entre un total de n- cuyo número de formas se simboliza $OR_{(n,r)}$, en donde $\mathbf{OR_{(n,r)} = n^r}$, y donde además r puede no solo ser igual o menor a n, sino también mayor, precisamente por la alternativa de tomar elementos repetidos.

Ordenación Cíclica -ordenamiento de n elementos en forma cerrada sin reemplazo- cuyo número de formas de realizarse se simboliza por $OC_{(n,n)}$ con $\mathbf{OC_{(n,n)}=(n-1)!}$.

Combinación -selección sin orden ni remplazo de r elementos entre un total de n- cuyo número de formas en que puede realizarse se simboliza como $C_{(n,r)}$, en donde $\mathbf{C}_{(n,r)}=\frac{n!}{(n-r)!r!}$.

Combinación con Repetición -selección de r elementos de entre un total de n sin orden y con reemplazo- cuyo número de formas en las que puede llevarse a cabo se simboliza como $CR_{(n,r)}$, con $\mathbf{CR}_{(n,r)}=\mathbf{C}_{(n+r-1,r)}$.

Permutación con Objetos Repetidos -permutaciones en un grupo de tamaño n, con k series de elementos iguales entre sí repartidos como r_1, r_2,..., r_k- en donde el número de permutaciones posibles con esos n elementos, denotada por $PR_{(n;r_1,r_2,\ldots,r_k)}$, estará dado como $\mathbf{PR}_{(n;r_1,r_2,\ldots,r_k)}=\frac{n!}{r_1!r_2!\ldots r_k!}$.

Por último, y dado que el problema principal al que se enfrentan los alumnos es la identificación de la fórmula a utilizar en un problema en particular, se presentó el siguiente cuadro en el que se resumen las características de cada arreglo (excepto los múltiples) y que a su vez pueden funcionar como una herramienta de identificación:

ARREGLO	CIRCULAR	REPITE	ORDENA	FÓRMULA	RESTRICCIÓN
Permutación	No	No	Sí	$P_{(n,n)}=n!$	$n=r$
Ordenación	No	No	Sí	$O_{(n,r)}=\frac{n!}{(n-r)!}$	$n>r$
Ordenación con Repetición	No	Sí	Sí	$OR_{(n,r)}=n^r$	$n\geq r$, $n<r$
Ordenación Cíclica	Sí	No		$OC_{(n,n)}=(n-1)!$	$n=r$
Combinación	No	No	No	$C_{(n,r)}=\frac{n!}{(n-r)!r!}$	$n\geq r$
Combinación con Repetición	No	Sí	No	$CR_{(n,r)}=C_{(n+r-1,r)}$	$n\geq r$, $n<r$

Aquí se aclaro a los alumnos que existe un arreglo más, el de Combinación Múltiple, pero que no sería manejado en este tema y que se abordaría más adelante, con la Función de Distribución de Probabilidad llamada Multinomial.

Finalmente se planteó la siguiente serie de problemas para que fueran resueltos por los alumnos, haciendo notar que algunos no sólo eran un arreglo sino dos o más, o el mismo varias veces, buscando la aplicación no únicamente de las fórmulas sino también y al mismo tiempo del Teorema Fundamental:

-¿De cuántas formas pueden acomodarse los números del 1 al 15 en una ruleta?

-¿De cuántas formas pueden sentarse 4 personas en una fila de 9 lugares?

-¿De cuántas maneras pueden colgarse 11 cuadros a lo largo de un muro?

-Una mujer que tiene 8 hijos quiere enviar a 3 de ellos a la tienda, 2 a lavar platos y 3 a jugar. ¿De cuántas maneras puede seleccionarlos?

-¿De cuántas formas puede llenarse una papeleta de "Melate"?

-¿Cuántas banderas tricolores de franjas pueden formarse con los colores del arcoiris?

-¿De cuántas formas puede llenarse una papeleta de Pronósticos Deportivos?

-Si se quiere sentar a 5 hombres y 5 mujeres alrededor de una mesa, ¿de cuántas formas puede hacerse si se busca sentarlos alternadamente?

-¿De cuántas formas puede llenarse una papeleta de Tris?

-Si tomamos un disco compacto en el modo "random", ¿de cuántas maneras podrían ser escuchadas las 13 canciones del disco?

-¿Cuántos números telefónicos de 7 dígitos es posible formar si el primer dígito no puede ser cero?

-Una puerta cuenta con 5 chapas, de las cuales algunas se dejan abiertas y otras cerradas para evitar robos. ¿De cuántas maneras puede dejarse asegurada la puerta?

-¿De cuántas formas pueden acomodarse los números en un disco del teléfono?

-En un barco se tiene 4 banderas de distintos colores para hacer señales colocándolas en el mástil. ¿Cuántas señales diferentes pueden hacerse con:
a) ninguna bandera?
b) una bandera?
c) dos banderas?
d) tres banderas?
e) cuatro banderas?

-Se tienen 3 libros de Algebra, 5 de Geometría y 8 de Estadística. ¿De cuántas formas pueden colocarse en un estante si:
a) pueden ir todos revueltos?
b) deben ir juntos los de cada materia?

-En una clase de Estadística se tienen 36 alumnos e igual número de bancas. Sí en cada clase los alumnos se distribuyen de manera diferente ¿cuántas clases se requieren para que los alumnos puedan sentarse de todas las formas?

-Con la intención de incrementar sus ventas una empresa de comida chatarra lanzó una promoción de una colección de 25 estampas. ¿De cuántas formas puede llenarse cada sobre de 5 estampas si no les preocupa que pueda haber repetidas?

-¿De cuántas formas pueden acomodarse los caracteres de la palabra:
a) acacias?
b) Ingeniería?
c) Laura?
d) electrolito?

Unidad III, Elementos de Probabilidad

Objetivo:
Que el alumno conozca las definiciones principales y algunos teoremas básicos en la Teoría de Probabilidad, tales como espacio muestral, evento, eventos independientes, eventos mutuamente excluyentes, eventos complementarios.
Que el alumno comprenda los Axiomas de Kolmogorov como base de la Teoría Moderna de Probabilidades.

Como estrategia didáctica el profesor expondrá ante el grupo algunas de las principales definiciones y teoremas de la Teoría de Probabilidad, ejemplificándolos en su momento; algunos de estos ejemplos serán dados por los alumnos por medio de lluvia de ideas. Después el profesor resolverá algún o algunos ejemplos y para terminar los alumnos resolverán ejercicios en el aula.

3. Elementos de Probabilidad

En esta unidad del curso se buscó plantear los elementos teóricos básicos necesarios para la comprensión del tema correspondiente a las Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta sobre la base de definiciones y algunos teoremas, ejemplificando cada uno con ejemplos sencillos.

La unidad se inicia comentando con el grupo la importancia de la Probabilidad dentro de la toma de decisiones, y la necesidad de aclarar terminología y simbología previo a la resolución de problemas prácticos, para lo cual se presentaran aquellos elementos que para los efectos resultarán útiles. Estos son las siguientes definiciones y teoremas:

-Definición.- Si un experimento qué está sujeto al azar resulta de n formas posibles y mutuamente excluyentes, y si n_A de estos resultados tiene un atributo A, la probabilidad de A es la proporción de n_A con respecto a n, esto es:

$$P(A)=\frac{n_A}{n}.$$

Esta es la *Definición Clásica de Probabilidad.*

Antes de continuar se aclara con los alumnos que un experimento sujeto al azar es aquel que en términos coloquiales depende de la suerte, aunque en Probabilidad nunca se utiliza el término suerte; y que más adelante se explicará más a fondo cuando se dice que se tienen resultados mutuamente excluyentes.

-Definición.- Si un experimento se repite n veces bajo las mismas condiciones, y n_B de los resultados son favorables a un atributo B, el límite del cociente de n_B con respecto a n conforme n se vuelve grande se define como la probabilidad del atributo B, esto es:

$$P(B)=\lim_{n\to\infty}\frac{n_B}{n}.$$

Esta es la *Definición de Probabilidad como Frecuencia Relativa*, la cual nos permite establecer una probabilidad de un evento como un porcentaje una frecuencia relativa entendida bajo la misma idea manejada en la unidad de Estadística Descriptiva, cuestión que será manejada más adelante en más de una ocasión.

Aquí se mencionó que existe una tercera definición de Probabilidad, la definición subjetiva, que se refiere a aquellas ocasiones en las que se asume que es poco probable o muy probable que algo suceda, basándose exclusivamente en la intuición, especificando que este tipo de Probabilidad no es algo muy "científico", pero que tiene cierta validez en términos de aleatoriedad.

-Definición.- El conjunto de todos los posibles resultados de un experimento aleatorio (que depende del azar) recibe el nombre de espacio muestral y se simboliza por S o por Ω.

-Definición.- Se dice que un espacio muestral es *discreto* si sus resultados pueden expresarse en términos de enteros positivos.

-Definición.- Se dice que un espacio muestral es *continuo* si sus resultados consisten de un intervalo de números reales.

-Definición.- Un evento del espacio muestral es un grupo de resultados contenidos en éste, cuyos miembros tienen una característica en común.

-Definición.- El evento que contiene a ningún elemento del espacio muestral recibe el nombre de evento nulo o vacío, y se simboliza por $\varnothing$ o $\{\ \}$.

Para las siguientes definiciones, los símbolos E, E_1 y E_2 representan eventos del espacio muestral.

-Definición.- El evento formado por todos los posibles resultados en E_1 o E_2 o en ambos, recibe el nombre de la unión de E_1 y E_2 y se denota por $\mathbf{E_1 \cup E_2}$.

-Definición.- El evento formado por todos los resultados comunes tanto a E_1 como a E_2, recibe el nombre de intersección de E_1 y E_2 y se denota por $\mathbf{E_1 \cap E_2}$.

-Definición.- Se dice que los eventos E_1 y E_2 son mutuamente excluyentes, o disjuntos, si no tienen resultados en común; en otras palabras, si $\mathbf{E_1 \cap E_2 = \varnothing}$.

-Definición.- Si cualquier resultado de E_2 también es un resultado de E_1, se dice que el evento E_2 está contenido en E_1 y se denota por $\mathbf{E_2 \subset E_1}$.

-Definición.- El complemento de un evento E con respecto al espacio muestral S, es aquel que contiene a todos los elementos de S que no se encuentran en E, y se denota por $\mathbf{E'}$, $\mathbf{E^c}$ o $\overline{E}$.

-Definición.- Sea S cualquier espacio muestral y E cualquier evento de éste. Se llamará función de probabilidad sobre el espacio muestral S a P(E) si satisface los siguientes axiomas:

1) $P(E) \geq 0$.

2) $P(S)=1$.

3) Si para los eventos $E_1, E_2, E_3 ..., E_i \cap E_j = \varnothing \ \forall \ i \neq j \Rightarrow$

$P(E_1 \cup E_2 \cup E_3 \cup ...)=P(E_1)+P(E_2)+P(E_3)+...$

Estos axiomas se conocen como **Axiomas de Kolmogorov** y en este punto se explicaron no solo los axiomas sino también el simbolismo matemático empleado en ellos.

Luego se comentó con el grupo que estos axiomas son de suma importancia en la teoría moderna de la Probabilidad (son recientes contra los años de vigencia de la Probabilidad clásica) y que a partir de ellos se establecen algunos teoremas que precisamente componen dicha teoría moderna. Algunos de los teoremas son:

-Teorema.- $\mathbf{P(\varnothing)=0}$ (la probabilidad del evento nulo es cero).

-Teorema.- Para cualquier evento $E \subset S$, se tiene que $\mathbf{0 \leq P(E) \leq 1}$.

-Teorema.- Sea S un espacio muestral que contiene a cualesquiera dos eventos A y B, entonces $\mathbf{P(A \cup B)=P(A)+P(B)-P(A \cap B)}$.

-Teorema.- $\mathbf{P(A)+P(A^c)=1}$ (la probabilidad de un evento más la probabilidad de su complemento, es igual a uno).

Para ejemplificar parte de lo anterior se planteó el siguiente problema: "Un sistema contiene dos componentes A y B, y se conecta de manera que éste funciona si cualesquiera de las dos componentes funcionan. Se sabe que la probabilidad de que A funcione es $P(A)=0.9$, la probabilidad de que B funcione es $P(B)=0.8$ y la probabilidad de que ambas funcionen es $P(A \cap B)=0.72$. Determinar la probabilidad de que el sistema funcione". Para resolverlo primero se "tradujo" a términos probabilísticos el evento de que el sistema funcione como el evento de que sucedan A o B o ambos, es decir $A \cup B$. Luego, sabemos por el tercer teorema que se presentó que $P(A \cup B)=P(A)+P(B)-P(A \cap B)$, en donde conocemos del problema todos los datos; así $P(A \cup B)=0.9+0.8-0.72=0.98$ es la probabilidad de que el sistema funcione.

Una vez vista esta aplicación se definieron de modo informal algunos elementos más específicos de probabilidad:

-Probabilidad marginal.- Es la probabilidad de la ocurrencia de un evento del espacio muestral sin involucrar ningún elemento más. Por ejemplo la probabilidad de sacar un as de una baraja inglesa. Se denota por **P(A)**, asumiendo A como el evento.

-Probabilidad conjunta.- Es la probabilidad de ocurrencia de dos eventos de un espacio muestral, sea de modo simultáneo o sucesivo. Por ejemplo la probabilidad de extraer de la baraja inglesa una carta que sea un as y además negra. Se denota como **$P(A\cap B)$**, siendo A y B los eventos.

-Probabilidad condicional.- Es la probabilidad de que ocurra un evento del espacio muestral bajo la condición de que previamente halla ocurrido otro evento de dicho espacio. Como ejemplos podemos mencionar la probabilidad de accidentarse en un coche dado que no se hizo el correspondiente ajuste de frenos, la probabilidad de enfermar dado que se comió en un puesto callejero, o la probabilidad de aprobar un examen dado que se estudió. Se denota por $\mathbf{P}\langle A|B\rangle$, siendo A el evento cuya probabilidad se busca calcular y B el evento cuya ocurrencia condiciona la probabilidad de A.

En este punto se plantearon algunas formas de manejar estos tipos de probabilidades sobre la base de características de los eventos. Así, replanteamos primero el tercer teorema que anteriormente se había dado:

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

y, realizando un despeje elemental, obtenemos una expresión para la probabilidad conjunta:

$\mathbf{P(A\cap B)=P(A)+P(B)-P(A\cup B)}$,

aunque aclarando que esta no es la única opción para la probabilidad conjunta, como se verá más adelante.

Para este momento se habían manejado ya los casos de probabilidad marginal con la Definición Clásica, y la de probabilidad conjunta al definir la intersección de eventos. Pero el caso de la probabilidad condicional aún resultaba un elemento nuevo, por lo que se repitieron algunos ejemplos como los anotados más arriba -y en los que se estableció que a los eventos involucrados se les podía denominar como "dependientes"- y se presentó al grupo la expresión de resolución de un planteamiento de probabilidad condicional para dos eventos como sigue:

$\mathbf{P}\langle A|B\rangle=\mathbf{P(A\cap B)/P(B)}$,

haciendo notar que se trata del cociente de la probabilidad conjunta entre la probabilidad marginal del evento *condicionante*, pues a fin de cuentas se está buscando la probabilidad de que ocurra el evento condicionado. Y aclarando también que la barra vertical entre A y B se lee como "dado" o "dado que".

Ahora, y de nuevo sobre la base de un despeje simple, podemos hallar otra expresión para la probabilidad conjunta, dependiente de la probabilidad marginal:

$\mathbf{P(A\cap B)=P}\langle A|B\rangle\mathbf{P(B)}$.

Inmediatamente de esto se plantearon algunos ejemplos de probabilidad condicional un tanto "absurdos": la probabilidad de accidentarse dado que se maneja un carro rojo, la probabilidad de enfermarse dado que se tiene un nombre en particular, o la probabilidad de aprobar un examen dado que es día soleado. Estos llevaron a la conclusión de que pueden llegarse a establecer eventos en los cuales la previa ocurrencia de uno no afecta en modo alguno la probabilidad de que se dé el otro; es decir, estos eventos son independientes entre sí, y cuando esto sucede entonces puede establecerse que

P$\langle A|B\rangle$=P(A),

lo que va a provocar que la probabilidad conjunta para eventos independientes se calcule como

P(A$\cap$B)=P(A)P(B),

basándonos sobre la sustitución de P$\langle A|B\rangle$ por P(A) en la expresión P(A$\cap$B)= P$\langle A|B\rangle$P(B).

Finalmente planteamos lo que ocurre con la probabilidad conjunta al tenerse eventos mutuamente excluyentes. Si dos eventos son mutuamente excluyentes, entonces esto significa que A$\cap$B=$\varnothing$, y nosotros sabemos que P($\varnothing$)=0; por lo que si dos eventos son mutuamente excluyentes entonces

P(A$\cap$B)=0,

lo que reduce la expresión P(A$\cup$B)=P(A)+P(B)-P(A$\cap$B) a

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

De este modo podemos tener distintas expresiones para la probabilidad condicional, para la probabilidad conjunta y para la probabilidad de la unión de eventos, dependiendo de las relaciones que los eventos involucrados guarden entre sí. En este punto se aclaró a los alumnos que para efectos del curso sólo se asumiría independencia o mutua exclusividad entre eventos cuando explícitamente se planteara alguna de estas características dentro de un problema o un desarrollo teórico.

Para concluir la parte teórica, se presentaron al grupo dos elementos más de teoría de conjuntos que resultan útiles en el manejo de la Probabilidad. Estos elementos no se explicaron ya que los alumnos los manejan desde el primer semestre del bachillerato, simplemente se recordaron:

-Propiedad conmutativa.- **A$\cap$B= B$\cap$A**, y
A$\cup$B= B$\cup$A.

-Leyes de DeMorgan.- **(A$\cup$B)^c=A^c$\cap$B^c**, y
(A$\cap$B)^c=A^c$\cup$ B^c.

Una vez vistos los elementos teóricos, se plantearon los siguientes problemas a los alumnos, problemas que tras un tiempo razonable serían resueltos también por el profesor con la finalidad de que en el grupo pudieran comprobar sus resultados y/o presentar sus dudas respecto al tema:

-Dados los eventos A y B tales que P(A)=0.5, P(B)=0.3 y P(A$\cap$B)=0.1, encontrar:
a) P(B^c).
b) P$\langle A|B\rangle$.
c) P$\langle B|A\rangle$.
d) P(A$\cup$B).
-Sean los eventos A y B tales que P(A)=0.2, P(B)=0.3 y P(A$\cup$B)=0.4; hallar:
a) P(A$\cap$B).
b) P(A^c $\cup$ B^c).
c) P$\langle A^c|B^c\rangle$.
-Si A y B son dos eventos independientes con P(A)=0.5 y P(B)=0.2, encontrar:
a) P(A$\cup$B).

b) $P(A^c \cap B^c)$.
c) $P(A^c \cup B^c)$.
-Si P(A)=0.35, P(B)=0.47 y $P(A \cup B)$=0.58, hallar:
a) $P(A \cap B)$.
b) $P\langle A|B\rangle$.
c) $P\langle B|A\rangle$.
d) $P(A^c)$.
-Si A y B son eventos independientes con P(A)=0.3 y P(B)=0.5, hallar:
a) $P(A \cup B)$.
b) $P(A^c \cap B^c)$.
-Si C y D son eventos mutuamente excluyentes con P(C)=0.6 y P(D)=0.2, encontrar:
a) $P(C \cup D)$.
b) $P(C^c \cup D^c)$.
-Si A y B son eventos mutuamente excluyentes con P(A)=0.7 y $P(B^c)$=0.25, hallar:
a) $P(A^c \cup B^c)$.
b) $P(A \cap B)$.
-Sean los eventos C y D tales que P(C)=0.4, P(D)=0.3 y $P(C \cap D)$=0.2. Encontrar:
a) $P(C \cup D)$.
b) $P\langle D|C\rangle$.

Unidad IV. Funciones de distribución de probabilidad de variable aleatoria discreta

Objetivos:
Que el alumno conozca los conceptos de variable aleatoria discreta y de distribución de probabilidad de variable aleatoria discreta.
Que el alumno diferencie entre una variable aleatoria discreta y una variable aleatoria continua.
Que el alumno calcule el valor esperado, la variancia y la desviación estándar de una variable aleatoria discreta.
Que el alumno conozca los ensayos de Bernoulli e identifique algunas distribuciones de probabilidad de variable aleatoria discreta, algunas de ellas como consecuencia de los ensayos de Bernoulli, como modelos de solución para ciertos problemas de cálculo de probabilidades y que resuelva dichos problemas (Binomial, Binomial Negativa, Poisson, Multinomial).

Como primera estrategia didáctica los alumnos generarán un concepto informal de variable aleatoria discreta y de distribución de probabilidad de variable aleatoria discreta por medio de lluvia de ideas; posteriormente se señalarán las similitudes entre estas distribuciones y las distribuciones de frecuencias vistas en la primera unidad, así como la posibilidad de calcular la "media" y la variancia de una variable aleatoria, con el fin de deducir las expresiones para el valor esperado y la variancia de una variable aleatoria discreta, así como las propiedades de una distribución de variable aleatoria discreta. Posteriormente resolverán ejercicios en el aula.
Como segunda estrategia didáctica se definirá en el aula el ensayo de Bernoulli, a partir de las distribuciones de probabilidad de variable aleatoria discreta. Después, para cada modelo generado a partir de repeticiones de ensayos de Bernoulli y por medio de lluvia de ideas, se irán deduciendo los modelos Binomial, Geométrico, Binomial Negativo, e Hipergeométrico, así como sus características, y se resolverán ejercicios en el aula; el modelo de Poisson será presentado en parte por el profesor y sus características deducidas por los alumnos por medio de lluvia de ideas y la dirección del profesor, para posteriormente resolver ejercicios en el aula. Finalmente, y a partir de las distribuciones Binomial e Hipergeométrica, los alumnos deducirán las distribuciones Multinomial e Hipergeométrica Multivariada a partir de problemas que se ajusten a cada modelo, resolviendo al final de cada deducción problemas en el aula.

4. Funciones de distribución de probabilidad de variable aleatoria discreta

INTRODUCCIÓN

En esta unidad se presentaron, en primer lugar, los conceptos generales de función de distribución de probabilidad y de variable aleatoria; y en segundo lugar, las funciones de distribución de probabilidad de variable aleatoria discretas de Bernoulli, Binomial, de Poisson, Geométrica, Hipergeométrica, Binomial Negativa, Multinomial e Hipergeométrica Multivariada.

Para iniciar se buscó primero generar conceptos propios del grupo respecto a variable aleatoria, a variable aleatoria discreta, a función de distribución de probabilidad y a función de distribución de probabilidad de variable aleatoria discreta. Para ello se repasaron los conceptos de cada término involucrado uno por uno, deduciendo que una función de distribución de probabilidad de variable aleatoria discreta es una regla matemática que asocia una serie de resultados discretos obtenidos de un experimento aleatorio (de hecho, de su espacio muestral) con sus respectivas probabilidades de ocurrencia. Igualmente se estableció, mediante un ejemplo sencillo la diferencia entre variables aleatorias discretas y continuas.

Una vez construidos los conceptos propios del grupo, se procedió a formalizar con definiciones ya establecidas que fueron dadas a los alumnos:

-Definición.- Una variable aleatoria es una función cuyos valores son números reales, definida en un espacio muestral.

Aquí se aclaró que esto se refiere básicamente a la representación de los resultados que pueden darse dentro de un experimento aleatorio. También se dijo que una variable aleatoria se simboliza con una letra mayúscula, y algún valor específico de ella con la misma letra pero en minúscula (i.e. Y es una variable aleatoria y y es un valor específico de aquella).

-Definición.- Una variable aleatoria Y se dice discreta si solamente puede tomar un conjunto numerable de valores.

-Definición.- La distribución de probabilidad para una variable aleatoria discreta Y se puede representar por una fórmula, una tabla o una gráfica que indique las probabilidades p(y) correspondientes a cada valor y de Y.

Igualmente, se planteó que toda distribución de probabilidad de variable aleatoria discreta, independientemente de la forma en la que se le represente, cumple con las siguientes propiedades:

1. $0 \leq p(y) \leq 1$ para toda y.

2. $\sum p(y) = 1$, donde la sumatoria se toma sobre todos valores de y con probabilidad diferente de cero.

Esta definición y las propiedades fueron ilustradas con el siguiente ejemplo:

-Experimento.- Lanzamiento de un dado de seis caras. Aquí tenemos como espacio muestral todos los valores de 1 a 6, en donde cada uno de ellos tiene una probabilidad de ocurrencia de 1/6, según la Definición Clásica de Probabilidad. Así, el experimento puede describirse como

y	1	2	3	4	5	6
p(y)	1/6	1/6	1/6	1/6	1/6	1/6

en donde es fácil verificar que para toda p(y) se tiene $0 \leq p(y) \leq 1$, y que Σ p(y) = 1, además de ser esta una tabla que efectivamente indica las probabilidades p(y) correspondientes a cada uno de los valores de y.

Aquí se aprovechó el momento para establecer las cuestiones de notación de la siguiente manera:
p(y) se refiere a la probabilidad de que la variable aleatoria Y tome específicamente el valor de y; por ejemplo 5. E igualmente puede escribirse como P(Y = y); así p(y) = P(Y = y). Por ejemplo, la probabilidad de obtener 5 al lanzar el dado es p(5) o P(Y = 5). Y del mismo modo en que se utilizó una igualdad en P(Y = y), podría utilizarse una desigualdad cualquiera, como P(Y > y); por ejemplo P(Y > 4), que es la probabilidad de obtener algún valor mayor a 4, y que dado que en este experimento en particular los resultados de cada cara son mutuamente excluyentes, podríamos decir que P(Y > 4) = p(5) + p(6), aunque estos planteamientos con desigualdades serían abordados más adelante.

Ahora, regresando con el ejemplo, se hizo una comparación de esta tabla con las tablas de distribución de frecuencias de Estadística Descriptiva, planteando que este nuevo cuadro podría tomarse como una tabla de distribución de frecuencias *relativas*, ya que, si se recordaba, la probabilidad de un evento también podía definirse como una frecuencia relativa. Además, si con las tablas de distribución de frecuencias había sido posible encontrar valores descriptivos de la muestra, parecería factible encontrar aquí valores descriptivos para el experimento, ó, más correctamente, para la variable aleatoria. Así se recordó que dos de los elementos que más carga presentan en la Estadística Descriptiva son la media y la variancia, de modo que podríamos pensar en el cálculo de específicamente esos valores para la distribución de probabilidades.

Para la media se vio la expresión

$$\bar{x} = \frac{\Sigma fx}{n},$$

en donde el cociente de cada f entre n es la frecuencia relativa, con lo que podríamos establecer una expresión para la media en la distribución que utilizara las frecuencias relativas o probabilidades de esta forma:

$$\bar{x} = \Sigma xp(x).$$

Así, se procedió a las siguientes definiciones, planteando al grupo que se buscaba formalizar de principio todas estas "suposiciones":

-Definición.- Sea Y una variable aleatoria discreta con función de probabilidad p(y). Entonces, el valor esperado de Y, E[Y], está definido por

$$E[Y] = \Sigma\, yp(y).$$

Aquí se dijo que E[Y] igualmente se representa por μ.

-Definición.- La variancia de una variable aleatoria Y, VAR[Y], está definida como el valor esperado de $(Y - \mu)^2$. Es decir

$$VAR[Y] = E[(Y - \mu)^2].$$

Se explicó también que a este valor se le conoce igualmente como esperanza matemática o valor esperado. Y que esta última denominación corresponde al hecho de que este resultado representa precisamente el valor que se esperaría obtener como resultado en el experimento aleatorio (en términos de juego, el valor al que hay que apostar al jugar con el dado).

Seguido a esta definición se planteó que del mismo modo que en Estadística Descriptiva se hallaba la desviación estándar como la raíz cuadrada positiva de la variancia, aquí la desviación estándar para la variable aleatoria es la raíz cuadrada positiva de VAR[Y]. Y también que existe una manera más sencilla de calcular la variancia, a partir del siguiente Teorema:

-Teorema.- $VAR[Y] = E[Y^2] - \mu^2$.

Con estos elementos se procedió entonces a calcular el valor esperado, la variancia y la desviación estándar del experimento de lanzar el dado de seis caras:

y	1	2	3	4	5	6
p(y)	1/6	1/6	1/6	1/6	1/6	1/6
y^2	1	4	9	16	25	36

E[Y] = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 21/6 = 7/2.
VAR[Y] = 1(1/6) + 4(1/6) + 9(1/6) + 16(1/6) + 25(1/6) + 36(1/6)
$- (7/2)^2 = 91/6 - 49/4 = 70/24 = 35/12$.
DS[Y] = $\sqrt{35/12}$,
con lo que se podría concluir que al lanzar el dado la apuesta segura era sobre 3.5. Evidentemente esto resultaría absurdo en términos prácticos, pero no en términos teóricos, y se recordó que en Estadística Descriptiva también se había obtenido un resultado no entero para el número promedio de hermanos. Ahora, para dar sentido al cálculo de la desviación estándar se recordó la Regla Empírica, y se generaron los intervalos correspondientes a restar y sumar una, dos, tres y tres y media veces la desviación estándar a la media:

- [1.79, 5.21], que aglutina a los 4 resultados enteros 2, 3, 4 y 5 en términos de redondeo, y que son el (4/6)100 = 67% de los resultados posibles.
- [0.08, 6.92], que aglutina a los 6 resultados enteros 1, 2, 3, 4, 5 y 6 en términos de redondeo y que son cerca del 95% de los resultados posibles (en términos reales son el 100%, pero se pretendía verificar la Regla Empírica, y de allí que se manejaran los resultados porcentuales aproximados a la regla).
- [-1.62, 8.62], que aglutina a aproximadamente el 99% de los resultados posibles.
- [-2.48, 9.48], que aglutina, para efectos prácticos al 100% de los resultados posibles.

Terminados estos cálculos se presentaron los siguientes elementos teóricos para completar la axiomatización del tema, pero sin intención de más allá en las aplicaciones ni en los desarrollos teóricos correspondientes:

-Definición.- Sea g(Y) una función de una variable aleatoria discreta Y, que tiene una función de probabilidad p(y). Entonces el valor esperado de g(Y) está definido por
$E[g(Y)] = \Sigma g(y)p(y)$.

-Teorema.- Sea c una constante. Entonces E[c] = c.

-Teorema.- Sea g(Y) una función de variable aleatoria Y y sea c una constante. Entonces E[cg(Y)] = cE[g(Y)].

-Teorema.- Sean $g_1(Y), g_2(Y), ..., g_k(Y)$ funciones de la variable aleatoria Y. Entonces

$E[g_1(Y) + g_2(Y) + ... + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + ... + E[g_k(Y)]$

Para finalizar con el tema se presentaron los siguientes ejercicios, en los que los alumnos habrían de calcular la media y la variancia de la variable aleatoria:

a)

Y	0	1	2	3
p(y)	1/8	1/4	3/8	1/4

b)

Y	1	2	3	4
p(y)	0.4	0.3	0.2	0.1

c)

Y	21	22	23	24	25	26
p(y)	0.05	0.20	0.30	0.25	0.15	0.05

d)

Y	0	1	2
p(y)	0.1	0.5	0.4

FUNCIONES DE DISTRIBUCION

Una vez vistas las propiedades de las distribuciones de probabilidad de variable aleatoria de discreta, se procedió al estudio de las funciones de distribución, entendidas como "fórmulas" para el cálculo de probabilidades de ciertos tipos de resultados bajo ciertas condiciones. Se presentaron al grupo las funciones que se verían en este tema como de Bernoulli, Binomial, de Poisson, Geométrica, Hipergeométrica, Binomial Negativa, Multinomial e Hipergeométrica Multivariada, aclarando que de la distribución de Bernoulli sólo se haría mención de sus características, pero que su aplicación sería en términos de las demás funciones (exceptuando la Multinomial y la Hipergeométrica Multivariada).

a) Distribución de Bernoulli

Para esta distribución simplemente se mencionó que se formaba de uno y solo un ensayo en el que únicamente podía haber dos resultados, los que genéricamente se denominan como éxito y fracaso. En este punto se aclaró que en Probabilidad los términos éxito y fracaso no tienen el mismo significado que en el lenguaje coloquial, en el que entendemos como éxito un resultado agradable o favorable, y como fracaso un resultado desagradable o desfavorable, sino que el éxito es la ocurrencia del resultado que se espera observar y el fracaso su no ocurrencia. Por ejemplo, estamos acostumbrados a tomar un deceso o un accidente como fracasos, pero si se está haciendo un estudio sobre pagos de primas de seguro o sobre seguridad, podrían interesarnos los decesos o los accidentes como observaciones, y por lo tanto los tomaríamos como éxitos.

En forma general, si al éxito se le asigna una probabilidad genérica de p y al fracaso de q (nótese que en todo momento debe cumplirse que p + q = 1) tendremos que una distribución de Bernoulli se puede ilustrar como

y	Exito	Fracaso
p(y)	p	q

Finalmente, se estableció que la real importancia de la distribución de Bernoulli es que, si se repite un ensayo de esta naturaleza, pueden generarse a partir de ella casi la totalidad de las distribuciones que conforman el resto de la unidad.

b) Distribución Binomial

Se presentó como aquella función que, obtenida a partir de realizar n repeticiones de un ensayo de Bernoulli, nos permite calcular la probabilidad de que entre esas n repeticiones se den un número específico de éxitos.

-Características:

1. Un experimento Binomial consta de n ensayos de Bernoulli iguales, repetidos e independientes.
2. Cada ensayo ofrece sólo dos posibles resultados: éxito y fracaso.
3. Las probabilidades de éxito y fracaso, denotadas por p y q respectivamente, se mantienen constantes en cada una de los n ensayos.
4. La variable aleatoria de estudio, denotada por Y, es el número de éxitos en los n ensayos.

Con estos elementos se pudo establecer como función de distribución Binomial la expresión

$$p(y) = C_{(n,y)}\, p^{y} q^{n-y}$$

y que genéricamente nos muestra las probabilidades de ocurrencia de todos y cada uno de los *y* posibles valores de Y.

Para acabar con la parte teórica, se establecieron las siguientes propiedades, que finalmente son las ya vistas en el tema de variable aleatoria visto anteriormente.

-Propiedades:

1. p + q = 1.
2. Σ p(y) = 1.

Ya con la parte teórica se presentó el siguiente ejemplo, en el que se establecieron las opciones del cálculo de un valor puntual y de valores acumulativos, estos últimos calculados a través de la segunda de las propiedades.

-Ejemplo.- La probabilidad de que un paciente se recupere de cierta enfermedad es de 0.8. Si 20 personas contraen tal afección, ¿cuál es la probabilidad de que sobrevivan:

a) exactamente 14?
b) a lo más 3?
c) al menos 18?
d) a lo más 18?
e) entre 6 y 8?

Una vez concluido el ejemplo y la serie de explicaciones correspondientes, el subtema se terminó con la resolución de los siguientes problemas por parte del grupo:

-Un examen de opción múltiple consta de 15 preguntas, cada una de las cuales tiene cinco posibles respuestas con sólo una correcta. Si un estudiante contesta el examen al azar ¿qué probabilidad tiene de responder correctamente al menos 10 preguntas?

-Un sistema de protección contra cohetes está constituido por cinco unidades de radar que funcionan independientemente, cada uno con probabilidad de 0.9 de detectar un cohete que ingrese a la zona que cubren todas las unidades. Si un cohete ingresa en la zona ¿cuál es la probabilidad de que exactamente cuatro unidades detecten el cohete? ¿Cuál es la probabilidad de que al menos una unidad detecte el cohete?

-Se sabe que de la producción de fusibles de una fábrica el 5% son defectuosos. ¿Qué probabilidad hay de que en un lote de 20 fusibles de esta fábrica haya menos de tres defectuosos?

-La probabilidad de que un avión de cierta línea aérea tenga un accidente es del 1%. Si la compañía cuenta con una flota de 17 aviones ¿qué probabilidad hay de que se accidenten entre 1 y 4 aviones?

-¿Qué probabilidad hay de sacar sol seis veces si lanzamos 10 "volados"?

-La probabilidad de que un microbús se accidente es de 65%. Si una persona toma seis microbuses diferentes en un día ¿qué probabilidad tiene de:
a) no sufrir ningún accidente?
b) sufrir seis accidentes?
c) sufrir tres accidentes?

-Una máquina de fax transmite la copia al primer intento el 92% de las veces. Si cierto día se envían seis faxes ¿qué probabilidad hay de que se transmitan a la primera cinco o menos?

-Se sabe que de los alumnos que ingresan a cierta carrera el 30% no la termina. Si ingresa una generación de 12 alumnos ¿qué probabilidad hay de que:
a) terminen más de 9?
b) ninguno termine?

-Una prueba de embarazo ofrece una certeza del 25% si se aplica durante el primer mes. Si se le aplica la prueba a ocho mujeres con un mes o menos de embarazo ¿qué probabilidad hay de que la prueba salga positiva en más de la mitad de los casos?

c) Distribución de Poisson

La distribución de Poisson se presentó como una alternativa a solución de problemas de naturaleza de Bernoulli que se encuentran situados en un espacio-tiempo. Esto es, pueden tenerse planteamientos probabilísticos que involucren cuestiones dimensionales como información o dato preliminar a la búsqueda del cálculo de la probabilidad de ocurrencia de un determinado evento.

Para poder resolver este tipo de problemas se recurre a la distribución de Poisson bajo las siguientes ideas. Primero, que la distribución efectivamente puede generarse sobre la base de ensayos de Bernoulli y de hecho con más semejanza a una distribución Binomial. ya que se busca la probabilidad de un cierto número de éxitos a lo largo de un determinado número de ensayos, si es que asumimos que cada espacio-tiempo puede subdividirse en n intervalos tan pequeños como sea necesario para que en cada uno de ellos sólo quepa un éxito y que este se dé con una probabilidad de **p**. Segundo, que el elemento promedio dentro del espacio-tiempo sería el dato a utilizar dentro de los cálculos de probabilidades, y que este promedio podría obtenerse como **np** si se partía del planteamiento establecido dentro de la primera consideración. Tercero, que el espacio-tiempo en el que se realizara la búsqueda de probabilidades de ocurrencia de determinado número de éxitos tendría que coincidir con el espacio-tiempo en el que se presentara distribuido el promedio genérico de éxitos, y que si esto no era así en un problema entonces el promedio debería ajustarse. Y cuarto, que el número de éxitos que podían caber dentro de cada espacio-tiempo es ilimitado. Cabe decir que se dejó al grupo el deducir esta última consideración.

Como parte final a esta introducción, se estableció que el dato promedio de éxitos en el espacio-tiempo se simbolizaría con la letra griega lambda en minúscula: λ.

Planteado lo anterior, se presentaron al grupo las características de esta distribución.

-Características:

1. Parte de n ensayos de Bernoulli iguales, repetidos e independientes.
2. Requiere del promedio de éxitos ocurridos dentro de un espacio-tiempo. Este promedio se denota por λ.

Puede utilizarse como una aproximación de un experimento Binomial en el que n sea mayor a 15 o 20 ensayos y menor a 30.[1]

3. La variable aleatoria de estudio es el número de éxitos en un espacio-tiempo en específico.

La función de distribución se presentó como:

$$\mathbf{p(y)} = \frac{\lambda^y}{y!}e^{-y}.$$

Antes de seguir se presentó el número e, llamado número de Napier, como un valor irracional aproximadamente igual a 2.718281828..., y a partir del cual se genera la función llamada exponencial, asignando la variable independiente como potencia de e. Se explicó también de modo muy superficial, que esta función tiene demasiada importancia dentro de las Matemáticas y que se manejaría en el segundo curso de Estadística en dos ocasiones más, y que parte de su importancia radica en el hecho de que dicha función describe elementos de crecimiento tanto poblacional como financiero. Igualmente se explicó cómo encontrar valores de e elevada a distintas potencias con la calculadora.

Una vez aclarado lo más elemental con respecto al número de Napier, se presentaron las propiedades de esta distribución:

[1] En este punto se estableció que esta aproximación se vería al final del subtema, y que los valores establecidos aquí para n eran relativamente discrecionales.

-Propiedades:
1. λ = np.
2. $\sum_{0}^{\infty} p(y) = 1$.

Terminados todos estos elementos teóricos se presentó este ejemplo para ilustrar el manejo práctico de la distribución de Poisson:
-Ejemplo.- En un almacén los clientes llegan al mostrador de caja en un promedio de siete por hora. En una hora dada ¿cuál es la probabilidad de que:
a) no lleguen más de tres clientes?
b) lleguen al menos dos clientes?
c) lleguen exactamente cinco clientes?
d) lleguen entre uno y cuatro clientes?
Para concluir, se dieron a los alumnos los siguientes ejercicios para su resolución:

-El número de nudos en un tipo particular de madera se presenta en un promedio de 1.5 nudos por cada 10 ft^3 de madera. Encuentre la probabilidad de que en un bloque de 10 ft^3 de esta madera se presente a lo más un nudo.

-Cierto tipo de árboles tiene retoños dispersos de manera aleatoria sobre un área extensa, con una densidad promedio de 5 por cada 100 m^2. Encuentre la probabilidad de que un guardabosque, al recorrer al azar una porción de 100 m^2 en esa área, no encuentre retoño alguno de dichos árboles.

-El número de errores cometidos por una mecanógrafa tiene una media de cuatro errores por página. Si una página dada tiene más de cuatro errores, la mecanógrafa tendrá que repetir la página entera. ¿Cuál es la probabilidad de que no se tenga que repetir cierta página?

-El promedio de automóviles que entran por un túnel es de un coche por período de 2 minutos. Si un número excesivo de coches entra al túnel por un período corto, se produce una situación de peligro. Encuentre la probabilidad de que el número de coches que entran al túnel en un período de dos minutos exceda de tres.

-Se sabe que los alumnos de CCH, al finalizar sexto semestre, deben un promedio de cinco materias. ¿Qué probabilidad tiene un alumno de nuevo ingreso de finalizar sexto semestre debiendo:
a) 20 materias?
b) ninguna materia?
c) más de dos materias?
d) entre ocho y diez materias?

-El número de defectos en el tejido de cierta tela se presenta con una media de 4 por m^2.
a) Encuentre la probabilidad de que una muestra de 1 m^2 contenga al menos un defecto.
b) Encuentre la probabilidad de que una muestra de 3 m^2 contenga al menos un defecto.

-En un aereopuerto se tiene un promedio de 15 llegadas internacionales por hora. En una hora determinada ¿qué probabilidad hay de que lleguen ocho vuelos internacionales o menos?

-Los balleneros noruegos cazan un promedio de 25 ballenas grises al mes. En un mes determinado ¿qué probabilidad hay de que los balleneros maten:
a) entre ocho y veinte ballenas grises?
b) menos de tres ballenas grises?

-El número de colonias de bacterias de cierto tipo en muestras de agua contaminada tiene una media de 2 por cm^3. Hallar la probabilidad de que en una cierta muestra haya menos de 10 colonias por cm^3.

Aproximación de la Binomial a la Poisson

En este punto se estableció que una distribución de Poisson arroja resultados muy semejantes a los que se obtienen de un experimento Binomial en el que se da una cantidad relativamente grande de ensayos, donde para algunos esta consideración de relativamente grande empieza a partir de 15 o 20 ensayos. En ese sentido se planteó también que un experimento Binomial que involucra 30 o más ensayos igualmente puede manejarse como una distribución de Poisson, pero que, como se vería en el siguiente curso, en este caso conviene otro tipo de aproximación, la aproximación a una distribución continua llamada Normal. Así, se estableció que para experimentos de naturaleza Binomial en los que se tuvieran de 15 o 20 (esto tomado a discreción) y hasta 29 ensayos, convenía transformarlos o aproximarlos a una distribución de Poisson a través de la propiedad de que λ = np.

Se presentó un ejemplo y luego, para finalizar el tema, se dio al grupo el siguiente ejercicio de aproximación de la distribución Binomial a la Poisson:

-Una vendedora se da cuenta de que a probabilidad de venta en una entrevista única es aproximadamente 0.03. ¿Cuál es la probabilidad de que ella haga al menos tres ventas al tener 25 compradores posibles?

d) Distribución Geométrica

Se planteó la distribución Geométrica de un modo formal, tras deducir su comportamiento con un ejemplo, estableciendo sus características, su función de distribución y sus propiedades.
-Ejemplo.- Se supone que el 30% de los aspirantes para cierto trabajo industrial tienen un entrenamiento avanzado en programación computacional. Los aspirantes son entrevistados uno tras otro, y son seleccionados al azar del conjunto de aspirantes. Determinar la probabilidad de que se encuentre al primer aspirante con un entrenamiento avanzado en programación:

a) en la quinta entrevista.
b) antes de la tercera.
c) después de la segunda.
d) entre la cuarta y la séptima.

-Características:

1. Consta de una serie de ensayos de Bernoulli iguales, repetidos e independientes.
2. Cada ensayo consta de dos posibles resultados, éxito o fracaso, cuyas probabilidades de ocurrencia, denotadas por p y q respectivamente, permanecen constantes en cada ensayo.

La variable aleatoria de estudio es el número de ensayo en el que ocurre el primer éxito.

Una vez dadas las características se presentó directamente la función de distribución como:

$p(y) = q^{y-1} p$.

Finalmente se presentaron las propiedades.

-Propiedades:

1. $p + q = 1$.
2. $\Sigma\, p(y) = 1$.
3. $\neg\exists\ p(0)$.

de las cuales las dos primeras se habían manejado ya con anterioridad, aclarando solamente que en el caso de la segunda la sumatoria, al igual que en la distribución de Poisson, se realizaba hasta el infinito, ya que el primer éxito podría aparecer tanto en el primer ensayo como hasta el y-ésimo, siendo *y* todo lo grande que se quisiera. En el caso de la tercera propiedad se explicó que, dado que la variable y es el ensayo en el que se presenta el primer éxito, para ese primer éxito se requiere como mínimo de un ensayo, por lo que los valores de y en esta distribución forzosamente estarían en el intervalo $[1, \infty]$.

Posteriormente se presentó el ejemplo correspondiente para terminar con la siguiente serie de ejercicios para el grupo:

-Un explorador de petróleo perforará una serie de pozos en cierta área hasta encontrar un pozo productivo. La probabilidad de que tenga éxito en una prueba es de 0.2. ¿Cuál es la probabilidad de que el primer pozo productivo se encuentre antes de la quinta perforación? ¿Cuál es la probabilidad de que el primer pozo productivo sea el tercer pozo perforado?

-Un contador público titulado ha encontrado que nueve de cada diez auditorías de compañías contienen errores importantes. Si el contador revisa la contabilidad de una serie de compañías ¿cuál es la probabilidad de que la primera contabilidad con errores sustanciales sea:

a) la cuarta revisada?
b) encontrada después de revisar la cuarta?

-Se estima que el 60% de una población de consumidores prefiere una marca particular de pasta de dientes A. ¿Cuál es la probabilidad, al entrevistar a un grupo de consumidores de que se tenga que entrevistar, para encontrar al primer consumidor que prefiere la marca A, a:

a) exactamente cinco personas?
b) a lo más cinco personas?

-La probabilidad que tiene un piloto de Fórmula 1 de accidentarse en una vuelta de carrera es del 15%. En una carrera de 72 vueltas ¿qué probabilidad tiene de accidentarse:
a) después de la séptima vuelta?
b) entre la décima y la décimoquinta vuelta?

-La probabilidad que tiene un reo de conseguir su fuga es del 5%. ¿Qué probabilidad tiene de fugarse en:
a) el primer intento?
b) después del cuarto intento?
c) antes del sexto intento?
d) entre el tercer y séptimo intento?

-La probabilidad que hay de aprobar un examen extraordinario de Estadística es del 7%. ¿Qué probabilidad tiene una persona que presente el extraordinario de aprobarlo:
a) al cuarto intento?
b) del quinto intento en adelante?
c) entre el sexto y el décimo intento?

-Supóngase que la probabilidad de que falle un motor durante cualquier período de una hora es de 0.02. Encuentre la probabilidad de que dicho motor funcione bien durante dos horas.

-La probabilidad de que un cliente acuda al mostrador de una tienda de abarrotes en cualquier período de un segundo es igual a 0.1. Supóngase que los clientes llegan de manera aleatoria y por lo tanto las llegadas en cada intervalo de un segundo son independientes. Encuentre la probabilidad de que la primera llegada ocurra durante:
a) el sexto intervalo de un segundo.
b) del sexto intervalo en adelante.

-Se sabe que siete de cada diez personas que acuden al dentista son mujeres. Encuentre la probabilidad de que en cierto día de consulta la primera mujer atendida sea entre el cuarto y séptimo paciente.

e) Distribución Hipergeométrica

Dada la naturaleza de esta distribución, se inició planteando un problema a guisa de ejemplo que permitió presentar al grupo el hecho de que existen experimentos en los cuales se tiene una población en la que algunos elementos presentan cierta característica –a considerar como éxito- y en los que se busca la probabilidad de tener una cierta cantidad de elementos con dicha característica dentro de una muestra, correspondiendo éstos a experimentos de tipo hipergeométrico, señalando a continuación sus características. El ejemplo, que se tomó directamente de entre los presentes en el grupo, fue: "En este grupo tenemos ahora a 28 personas, ¿quiénes han tenido varicela?; 19 en total. Si seleccionáramos al azar a 10 personas, ¿qué probabilidad habría de que entre ellas hubieran exactamente 5 que hayan tenido varicela?"
-Características:

1. Consta de n ensayos de Bernoulli iguales, repetidos e independientes, y aplicados sobre n elementos tomados de una población de tamaño N.
2. En los N elementos de la población existen r éxitos.
3. La variable aleatoria de estudio es el número de éxitos dentro de los n elementos seleccionados de la población.

Luego se presentó la función como

$$\mathbf{p(y)} = \frac{C_{(r,y)}C_{(N-r,n-y)}}{C_{(N,n)}},$$

y se dieron como sus propiedades las siguientes.

-Propiedades:
1. $\sum p(y) = 1$.
2. $y \leq r$.
3. $n - y \leq N - r$.
4. $n \leq N$.

En donde las tres últimas son en realidad condiciones propias de la operación de Combinación. Y en el caso de la primera propiedad se aclaró que no se ponían límites en la sumatoria por el hecho de que el número de éxitos posibles en el experimento depende del tamaño de r y de n en cada caso particular.

Para reafirmar el manejo de esta función de distribución, ya con la base teórica, se presentó un nuevo ejemplo, al término del cual se comentó que la importancia de esta distribución consistía en que de todas las distribuciones hasta ese momento vistas, la Hipergeométrica resultaba la "más estadística y menos probabilística", ya que en ella se ven elementos de población y selección de muestras, y que en todo momento, excepto en la presentación del resultado, se manejan valores enteros obtenidos de observaciones directas, lo que involucra directamente el concepto de frecuencia absoluta.

Una vez finalizado esto, las soluciones a los siguientes ejercicios corrió a cargo del grupo:

-En un almacén se tienen 10 impresoras de las cuales 4 son defectuosas. Una compañía selecciona 5 de las máquinas al azar, suponiendo que todas funcionan bien. ¿Cuál es la probabilidad de que las 5 máquinas sean:
a) no defectuosas?
b) defectuosas?

-Se desea formar un jurado de 6 personas de un grupo de 20 posibles miembros, de los cuales 8 son negros y 12 blancos. Si el jurado se selecciona aleatoriamente ¿qué probabilidad hay de que en él se tengan a lo más a 2 negros?

-Una urna contiene 10 canicas de las cuales la mitad son verdes. Si se sacan tres canicas al azar ¿qué probabilidad hay de que:
a) las tres sean verdes?
b) ninguna sea verde?

-Para estudiar los hábitos migratorios del lobo plateado un grupo de biólogos capturó, marcó y puso en libertad a 10 lobos de una manada de 25. Al final del período de migración

los biólogos volvieron a capturar a 10 lobos de la misma manada. ¿Qué probabilidad hay de que entre los 10 lobos vueltos a capturar haya:
a) menos de 3 lobos marcados?
b) más de 8 lobos marcados?
c) exactamente la mitad de lobos marcados?

f) Distribución Binomial Negativa

A manera de introducción para esta distribución se mencionó que en realidad se trataba de una generalización de la distribución Geométrica, y se formalizó la distribución Binomial Negativa con sus características y propiedades.

-Características:
1. Consta de una serie de ensayos de Bernoulli iguales, repetidos e independientes.
2. Las probabilidades p y q de éxito y fracaso para cada ensayo de Bernoulli permanecen constantes en todas las repeticiones.
3. La variable aleatoria de estudio es el número de ensayo en el que ocurre el r-ésimo éxito.

La función de distribución queda como:

$p(y) = C_{(y-1,r-1)}\, p^r q^{y-r}$.

-Propiedades:
1. $p + q = 1$.
2. $\Sigma\, p(y) = 1$.
3. $y \geq r$.

Para finalizar la parte teórica se planteó la posibilidad de calcular una función Binomial Negativa para el caso específico de que r = 1 –en sí, sería la búsqueda de la probabilidad de que en el y-ésimo ensayo apareciera el primer éxito-, obteniéndose:
$p(y) = C_{(y-1,1-1)}\, p^1 q^{y-1} = (1)p\, q^{y-1} = q^{y-1} p$, que no es otra sino la distribución Geométrica, con lo que se pudo establecer más firmemente a ésta como un caso particular de la Binomial Negativa, cosa que nos permite calcular no sólo los problemas con las características de una Binomial Negativa sino también aquellos con las características de una Geométrica con la expresión $p(y) = C_{(y-1,r-1)}\, p^r q^{y-r}$.

Una vez más se ilustró la distribución por medio de un ejemplo:

-Ejemplo.- La probabilidad de que una persona tenga un perro se estima en 0.3. Si se pregunta a una serie de personas, una tras otra, si tiene o no un perro, encuentre la probabilidad de que la décima persona entrevistada sea la quinta persona que tenga un perro.

La serie de ejercicios para su resolución, posterior al ejemplo explicativo, fue:

-Un científico inocula varios ratones con un germen de una enfermedad hasta que obtiene dos que la han contraído. Si la probabilidad de contraer la enfermedad es de 1/6, ¿cuál es la probabilidad de que se requieran 8 ratones?

-Suponga que la probabilidad de que una persona determinada crea los reportajes de un programa sensacionalista es de 0.8. ¿Cuál es la probabilidad de que la sexta persona que vea el programa sea la cuarta en darle credibilidad?

-Encuentre la probabilidad de que una persona que lanza una moneda obtenga:

a) la tercera cara en el séptimo lanzamiento.
b) la cuarta cara en el cuarto lanzamiento.

-Un beisbolista tiene un promedio de bateo de 0.300. ¿Qué probabilidad hay de que en un juego pegue un tercer hit en su cuarta ocasión al bat?

-De acuerdo con un grupo de psicólogos, alrededor de las 2/3 partes de consumidores de Valium son mujeres. Suponiendo que esta es una estimación válida, encuentre la probabilidad de que en un determinado día en una farmacia la quinta receta médica por Valium sea:
a) la tercera prescripción para una mujer.
b) a lo más la segunda prescripción para una mujer.

-Considérese el uso de un medicamento que se sabe es efectivo en el 60% de los casos en los cuales se utiliza. Encontrar la probabilidad de que el quinto paciente que presenta recuperación sea el séptimo o el octavo paciente tratado con el medicamento.

Al terminar con este último ejercicio, se puntualizó que en ese momento se estaba finalizando con las funciones de distribución generadas a partir de repeticiones de ensayos de Bernoulli, cuestión conocida también como Proceso de Bernoulli, pero que aún quedaban dos funciones de distribución discretas por ver: la Multinomial y la Hipergeométrica Multivariada, pero que tendrían cierta similitud con lo hecho en la Binomial Negativa, ya que ésta se manejó como una generalización de la Geométrica y aquéllas se verían como generalizaciones de la Binomial y de la Hipergeométrica respectivamente.

g) Distribución Multinomial

Antes de comenzar con la función de distribución Multinomial, se volvió al Cálculo Combinatorio para aprender un arreglo más: la Combinación Múltiple, ya que esta se presenta como parte de la función a describir.

La Combinación Múltiple se planteó como aquel arreglo en el que se asignan n objetos distintos en r grupos diferentes que contienen $n_1, n_2, \ldots, n_r$ respectivamente. El número de formas en las que este arreglo puede llevarse a cabo se denota por $CM_{(n;n_1,n_2,\ldots,n_r)}$ y para calcular el total de formas en las que puede realizarse se dedujo la expresión:

$$\mathbf{CM}_{(n;n_1,n_2,\ldots,n_r)} = \frac{n!}{n_1!n_2!\ldots n_r!},$$

en donde evidentemente debe cumplirse que

$$\sum_{i=1}^{r} n_i = \text{n}.$$

Para la distribución Multinomial se presentaron sus características, la función de distribución y sus propiedades:
-Características:
1. El experimento consiste de n pruebas iguales, repetidas e independientes.
2. El resultado de cada prueba cae en una de r categorías distintas.

3. La probabilidad de que el resultado de una de sola prueba se localice en una categoría i en particular es p_i y permanece constante en todas las pruebas.
4. Las variables aleatorias de estudio son los números de pruebas que caen en cada categoría.

-Función de distribución:

$$p(y_1, y_2, \ldots, y_r) = CM_{(n;y_1,y_2,\ldots,y_r)} p_1^{y_1} p_2^{y_2} \ldots p_r^{y_r}.$$

-Propiedades:
1. $p_1 + p_2 + \ldots + p_r = 1$.
2. $y_1 + y_2 + \ldots + y_r = n$.

Como un caso particular de la Multinomial se presentó la posibilidad de que r = 2, así:

$$p(y_1, y_2) = CM_{(n;y_1,y_2)} p_1^{y_1} p_2^{y_2} = \frac{n!}{y_1!y_2!} p_1^{y_1} p_2^{y_2},$$

pero teniendo en cuenta que $y_1 + y_2 = n$ y que $p_1 + p_2 = 1$, entonces $y_2 = n - y_1$ y $p_2 = 1 - p_1$, lo que hace evidente la forma Binomial de la expresión recién obtenida, de modo que podemos asumir precisamente a la Binomial como un caso particular de la Multinomial, y ocupar así esta última como otra alternativa de solución a problemas de tipo Binomial.

Para ilustrar el manejo de la distribución, se presentó el siguiente problema:

-Ejemplo.- Para cierta materia se sabe que la probabilidad de sacar MB es del 15%, de sacar B es del 20%, de sacar S es del 35% y de sacar NA es del 30%. ¿Qué probabilidad que de un grupo de 25 personas 5 saquen MB, 12 saquen B y los 8 restantes reprueben?

Al terminar la parte explicativa con el ejemplo correspondiente, se dieron los siguientes ejercicios al grupo:

-Se sabe que el 65% de la población tiene la presión arterial alta, el 20% tiene la presión normal y el 15% tiene la presión baja. Si medimos la presión de un grupo de 15 personas, ¿qué probabilidad hay de que 7 de ellas tengan la presión alta, 2 tengan la presión normal y las restantes tengan la presión baja?

-De acuerdo con los datos ajustados del censo de 1990, las proporciones de adultos, clasificados en 5 categorías de edad son como sigue:

Edad	Proporción
18 – 24	0.18
25 – 34	0.23
35 – 44	0.16
45 – 64	0.27
65 - +	0.16

Si se seleccionan al azar 5 adultos de esta población, encuentre la probabilidad de que la muestra contenga a una persona entre las edades 18 – 24, dos entre las edades 25 – 34 y dos entre las edades 45 - 64.

-En cierta carrera se sabe que al finalizar el primer año el 10% de los alumnos han reprobado alguna materia, el 20% ha reprobado alguna materia hasta finalizar el segundo año, el 30% hasta finalizar el tercero y el 40% hasta finalizar el cuarto. Si tomamos al azar a 10 alumnos de esta carrera que ya hayan cursado hasta el cuarto año, ¿qué probabilidad hay de que de ellos 4 hayan reprobado por primera vez en el primer año, 3 en el segundo, 2 en el tercero y 1 en el cuarto?

-En una fábrica se sabe que el 85% de la producción es buena, que 10% es aceptable y que el 5% se rechaza. En un embalaje de 17 piezas de la producción de la fábrica ¿qué probabilidad hay de que aparezcan 10 buenas, 5 aceptables y 2 que deban rechazarse?

-El 35% de las personas usa talla mediana en camisetas, el 50% usa talla chica y el resto usa talla grande. ¿Qué probabilidad hay de que tengan que comprarse 2 camisetas grandes, 5 medianas y 4 chicas para una oncena de jugadores de foot-ball?

-El rendimiento de las principales escuderías de Fórmula 1 en las últimas temporadas se muestra en la siguiente tabla:

ESCUDERIA	RENDIMIENTO
McLaren	0.35
Ferrari	0.30
Jordan	0.20
Williams	0.10
Benetton	0.05

Sobre la base de estos datos ¿qué probabilidad hay de que en la siguiente temporada (17 carreras en total) McLaren gane 7 de ellas, Ferrari gane 5, Jordan 4 y Williams 1?

h) Distribución Hipergeométrica Multivariada

Esta distribución se presentó directamente como la generalización de la distribución Hipergeométrica, en analogía a lo ya visto con la Binomial Negativa con respecto a la Geométrica y a la Multinomial con respecto a la Binomial.

-Características:

1. Consta de n pruebas iguales, repetidas e independientes aplicadas a n elementos tomados de una población de tamaño N.
2. Los N elementos de la población se subdividen en k grupos de r_1, r_2,..., r_k elementos respectivamente.
3. Las variables aleatorias de estudio son el número de elementos que caen dentro de las k categorías dentro del total de los n elementos de la muestra.

La función de distribución se presentó como:

$$p(y_1, y_2,..., y_k) = \frac{C_{(r_1,y_1)}C_{(r_2,y_2)}...C_{(r_k,y_k)}}{C_{(N,n)}}.$$

Finalmente se establecieron las propiedades correspondientes.

-Propiedades:

1. $\sum_{i=1}^{k} y_i = n.$

2. $\sum_{i=1}^{k} r_i = N$.

Con este ejemplo ilustrativo se finalizó con la explicación de la distribución Hipergeométrica Multivariada:
-Ejemplo.- En un congreso de Ciencias hay 10 físicos, 8 químicos, 6 biólogos y 12 matemáticos. Al final del congreso todos fueron a celebrar la clausura a un centro nocturno, y durante una redada 10 de los congresistas fueron detenidos. ¿Qué probabilidad hay de que estos detenidos sean específicamente 5 matemáticos, 2 biólogos, 2 físicos y 1 químico?
El resto fueron los siguientes ejercicios propuestos al grupo:

-De un grupo de 22 personas se sabe que 3 nacieron en primavera, 4 en verano, 10 en otoño y 5 en invierno. Si de este grupo se extrae una muestra de 10 personas, ¿qué probabilidad hay de que una haya nacido en primavera, 2 en verano, 4 en otoño y 3 en invierno? (Este problema se desarrolló en clase con datos tomados directamente del grupo).

-En un gimnasio de Tae Kwon Do hay 25 practicantes, de los cuales 10 son cintas blancas, 7 son amarillas, 3 verdes, 3 rojas y 2 negras. Al acudir a un torneo 15 de los practicantes abordan una camioneta. ¿Qué probabilidad hay de que en la camioneta vayan 5 cintas blancas, 5 amarillas, tres rojas y 2 negras?

-En una dulcera hay 5 chocolates, 7 caramelos, 4 paletas y 9 pastillas. Un niño mete la mano y saca 9 golosinas. ¿Qué probabilidad hay de que haya tomado 3 chocolates, 2 caramelos, 2 pastillas y 2 paletas?

Con esto se terminó no sólo la unidad correspondiente a Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta, sino también todo el primer curso de Estadística.

PROBLEMAS MÁS RECURRENTES QUE SE PRESENTARON EN ESTE CURSO
- En Estadística Descriptiva confusión entre los nombres Media y Mediana, la confusión entre Media y Promedio (aún habiendo establecido que se trata de términos sinónimos), el tomar las frecuencias absolutas como Moda y Mediana, la interpolación, la diferenciación entre Variancia y Desviación Estándar, y los elementos necesarios para el trazo de cada gráfico. Todas estas dificultades se trabajaron explicando de nueva cuenta y de manera individual a aquellos alumnos con problemas. En algunos casos fue posible inducir al razonamiento, y en otros, por cuestiones de tiempo, se tuvo que plantear directamente el hecho en duda como una verdad.
- En Cálculo Combinatorio el problema que se presentó fue el de la identificación del arreglo correspondiente a la solución de un problema por sobre los otros. El cuadro con las características propias de cada arreglo que se presenta dentro del desarrollo de la unidad minimizó mucho esta dificultad, pero aún tuvo que inducirse mucho hacia el razonamiento con alumnos atrasados, utilizando precisamente el cuadro. Igualmente se procuró no trabajar demasiado con la Permutación de Objetos Iguales, dada su naturaleza múltiple con respecto a los demás arreglos. Y esto último fue también razón para tratar la Combinación Múltiple hasta el penúltimo tema de la cuarta unidad.

- En Elementos de Probabilidad no se tuvieron problemas en la resolución de problemas, pero la parte teórica causó desinterés y aburrimiento, y por ende un bajo nivel de comprensión. Esto último procuró combatirse con el planteamiento de ejemplos prácticos de manera frecuente, y con los juegos de dados y de cartas. La capacidad operativa puede atribuirse al conocimiento previo de Teoría de Conjuntos.
- En Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta se presentó de nuevo dificultad para identificar un problema de otro, sobre todo en las primeras funciones, pero la práctica fue minimizando este problema. Así como el presentar que los datos se dan de manera diferente en casi todos los casos, y que en donde llegan a coincidir, el problema pide probabilidades sobre cuestiones diferentes.
-Como problema recurrente a lo largo de todo el curso, se presentó el manejo de la calculadora. En este punto se enseñó a elevar a n-ésimas potencias, calcular factoriales, Combinaciones por proceso o directas si esto era posible, exponenciación, y algunos trucos basados en la factorización y la definición de factoriales para simplificar los cálculos en las distribuciones Geométrica y de Poisson. Así mismo, se proporcionó al grupo al finalizar el curso, y antes de la segunda evaluación parcial, una tabla de Combinaciones desde 0 en 0 hasta 30 en 30 con el fin de optimizar el tiempo de resolución de examen. Esta tabla se proporcionó hasta el final con la intención de dar oportunidad de practica a los alumnos tanto de la operación como de los cálculos en la calculadora, y podrá encontrarse en los anexos al final de este trabajo.
- Finalmente se encaró cierto temor de parte de los alumnos en el proceso de evaluación, en el sentido de que dos exámenes parciales dentro de un programa tan extenso los obligaría a demasiada carga de trabajo, ya que si deseaban exentar el examen final se les pedía ambos parciales aprobados. Finalmente, al considerar dentro de la evaluación total el trabajo desempeñado en clase y el conocimiento adquirido más que la capacidad para dar un resultado, el grupo aceptó el presentar sólo dos exámenes parciales. Modelos de los dos exámenes parciales y de un examen final se encontrarán al término de la presente memoria.

ESTADISTICA II

Unidad I. Conceptos básicos de Muestreo y Estimación

Objetivo:
Que el alumno conozca algunos conceptos básicos de Muestreo y Estimación, tales como estimación puntual, estimación por intervalos, muestreo, parámetro, estimador, como una introducción hacia la Estadística Inferencial.
Que el alumno conozca la necesidad de aplicar cálculo de probabilidades dentro de la Estadística Inferencial.

Como estrategia didáctica el profesor expondrá ante el grupo algunos de los principales elementos propios del Muestreo y la Estimación, haciendo preguntas a los alumnos al final de dicha exposición.

1. Muestreo y Estimación

Conceptos básicos

Esta unidad consistió fundamentalmente en presentar a los alumnos los elementos más usuales en el Muestreo y la Estadística Inferencial, así como en establecer un nexo entre la Estadística Descriptiva y la Estadística Inferencial y dar al mismo tiempo una introducción de lo que es la Estimación en Estadística.

Dado el carácter definitorio de los puntos que acaban de establecerse es de comprender que la metodología se sustentara en la exposición a cargo del profesor. En este caso se procedió así ya que el tema de Distribuciones Muestrales no queda contemplado dentro del programa, puesto que muchos de los conceptos teóricos quedarían fuera del alcance de los conocimientos previos de los alumnos, y adentrarse en ellos no solo implicaría mermar el aprendizaje de los estudiantes sino también dejar de lado el interés de todo el curso, que es el inducir a los alumnos a un manejo práctico de la Estadística Inferencial.

La exposición hecha en la clase fue la siguiente. Se presentó al grupo la alternativa de contemplar la realidad como dos grandes conjuntos, el conjunto de lo que efectivamente es y el conjunto de lo que puede ser, en donde el primero está compuesto de fenómenos determinados y en cierta medida inmutables y que son sujeto de estudio de ciencias como la Física, y en dónde el segundo está compuesto precisamente de fenómenos aleatorios y es sujeto de estudio, obviamente, de la Probabilidad y de la Estadística. Ahora, a la Estadística le interesa lo que puede ser, y se apoya en la Probabilidad para verificar las posibilidades de que sea, en función de llegar a una toma de decisiones; el tener una idea aproximada de cómo será el comportamiento de cierta población con respecto a una determinada situación ayuda a prevenir dificultades que podrían presentarse. Todo esto pone de manifiesto algo sumamente importante: ***la necesidad no sólo de obtener datos, sino también de poder interpretarlos.***

¿De que manera la Estadística puede conseguir esto? Básicamente en lo que podríamos llamar como historia pasada y presente de la situación de interés en la población. Por ejemplo, si nos interesara saber el número de habitantes que habrá en una región en un futuro, podremos recurrir a la cantidad que ha habido en los últimos años y a la que hay actualmente, y así calcular la manera en la que la población ha crecido con lo que podremos tener una idea aproximada de cómo crecerá (o decrecerá, de acuerdo con las condiciones).
Otro ejemplo es el de procurarnos una idea de cuánta gente es analfabeta en el país, y para ello podemos recurrir a qué tan alfabetizada está una parte representativa de la población total –una muestra- con lo que podemos tener una idea aproximada del porcentaje real de analfabetas. Y en ambos casos los resultados nos llevarán a tomar una decisión, específicamente la de cuánto deben incrementarse los servicios o qué tan urgente es una campaña de alfabetización.

Pero el recurrir a esa historia pasada o presente no siempre es fácil. Si queremos estudiar el comportamiento de una población la gran mayoría de las veces encontraremos que esto es imposible, específicamente por tres posibles razones: o la población es demasiado grande como para que podamos analizarla toda (todos los seres humanos, por ejemplo) o la población no existe o la población es infinita (todos los medicamentos que se van a producir si es que éste aprueba las fases experimentales, o bien todas las televisiones de una fábrica incluidas las que ya vendió y que incluso ya se han destruido, las que están terminadas y las que van a fabricar, como ejemplos). Ahora bien, la solución a este problema es relativamente fácil, y consiste en tomar del conjunto que es la población, un subconjunto, que ya se había definido previamente como muestra, que refleje el comportamiento de la población con respecto a la característica que nos interese y sobre el cual pueda realizarse todo un análisis descriptivo, previo al inferencial. Los procesos para seleccionar dicha muestra son precisamente lo que se conoce como Muestreo.

El interés en aplicar los procesos descriptivos a la o las muestras que puedan obtenerse de una población radica en que los elementos descriptivos nos dan una idea clara de cómo se presenta una cierta característica en un grupo de elementos; y si este grupo es un "modelo" de toda una población, entonces podremos tener una descripción aproximada del comportamiento de la población con respecto a alguna característica sobre la que se requiera información. Todo esto funciona de la siguiente forma: una población en sí misma tiene características propias y en cierto sentido inmutables, y los elementos que las describen serán, por ende, inmutables también; esto es, una población tiene por ejemplo una determinada media, que, sin importar cuantas veces se le calcule (asumiendo dicho cálculo como posible) siempre será la misma, pero en una muestra esto no ocurre, ya que para cada población puede existir un sinnúmero de muestras (para mejor ejemplo, la distribución Hipergeométrica) que presentarán lógicamente distintas medias, ó, en general, distintos valores descriptivos en cada una; es decir, los valores descriptivos de una población son constantes, mientras que los de las muestras pueden considerarse variables.

A nivel meramente denominativo, llamemos *Parámetros* a los valores descriptivos de la población y *Estimadores* o *Estadísticos* a los valores descriptivos de la o las muestras, y deduzcamos algunas características:

-Los parámetros son constantes, y que los estimadores son variables; y de hecho, dado que una muestra se toma al azar para evitar errores en la información, los estadísticos son variables cuyo valor está sujeto al azar, es decir, son variables aleatorias.

-Los parámetros son valores desconocidos (de otro modo no tendría sentido aproximarlos), mientras que los estimadores, desde el momento en que somos nosotros quienes los calculamos, son valores perfectamente identificados.

-Como una simbolización genérica, en adelante se representarán los parámetros como θ y los estimadores como $\hat{\theta}$.

Parámetros y elementos de interés dentro del curso

Ahora, para efectos del curso, en específico se tratará de averiguar el comportamiento de los parámetros media, proporción y variancia de una población, y de los parámetros diferencia de medias, diferencia de proporciones y cociente de variancias para efectos de comparar dos poblaciones con respecto a sus medias, proporciones o variancias. Estos parámetros se representarán como μ, p, σ^2, $\mu_1 - \mu_2$, $p_1 - p_2$ y $\frac{\sigma_1^2}{\sigma_2^2}$ respectivamente, siendo sus estimadores simbolizados por $\bar{y}$ (o $\bar{x}$), $\hat{p}$, s^2, $\bar{y}_1 - \bar{y}_2$, $\hat{p}_1 - \hat{p}_2$ y $\frac{s_1^2}{s_2^2}$.

Hasta aquí se ha hablado acerca de estimadores e incluso se les ha simbolizado, pero ¿a qué se refiere en sí el término? Vimos que los parámetros, elementos descriptivos de una población que puede resultar útil conocer, son general e irremediablemente desconocidos, pero que los estimadores nos pueden dar una idea cercana de sus valores reales. Y precisamente en ese punto radica la denominación de estimador, ya que con ellos podemos ***estimar*** los valores de los parámetros, lo que es en sí, a manera un tanto informal, en lo que consiste la ***estimación***.

De hecho, la estimación podría tomarse como la deducción o la inferencia del comportamiento poblacional con respecto a alguna característica en particular, a partir de los valores descriptivos de una o varias muestras obtenidas precisamente de dicha población. Y, para ser más específicos, los estimadores descritos anteriormente se conocen como estimadores puntuales, ya que son en sí excelentes aproximaciones de los correspondientes parámetros valor por valor (o punto por punto); en donde esta aclaración se vuelve pertinente si planteamos la existencia de dos tipos de estimación: puntual y por intervalos. La ***estimación puntual*** consiste en calcular un estimador y asumir, sobre la base de la certeza de haber tomado la muestra adecuadamente, que a ese valor es aproximadamente igual nuestro parámetro; y esto se puede ejemplificar con el juego de tiro al blanco: el centro de la diana es fijo, al igual que el parámetro, y el juego consiste en realizar uno o más disparos -toma de muestras y cálculo de estimadores en comparación- buscando dar justo en el centro -es decir que la estimación sea lo más certera posible-, pero el juego tiene la dificultad de que pegar justo en un blanco pequeño resulta difícil, por lo que puede resultar conveniente ampliar el blanco, y que es justo lo que se hace en la estimación por intervalos. La ***estimación por intervalos*** consiste en generar un intervalo sobre la base de los estimadores, con la idea de encerrar entre esos dos límites al parámetro desconocido. Simbólicamente, la estimación puntual consiste en calcular un estimador $\hat{\theta}$ con la idea de que $\hat{\theta} \approx \theta$, mientras que la estimación por intervalos consiste en construir un intervalo con valores estimados $\hat{\theta}_1$ y $\hat{\theta}_2$ del parámetro con la idea de que $\theta \in [\hat{\theta}_1, \hat{\theta}_2]$.

Finalmente se dijo que el caso de la estimación por intervalos será agotado en la unidad correspondiente a Intervalos de Confianza, mientras que elementos de validación de estimaciones puntuales se verían más a fondo en la unidad correspondiente a Prueba de Hipótesis. Así mismo se dijo que la estimación puede ir con dos propósitos distintos: uno el de conocer la situación actual de una población, y el otro el de hacer predicciones de cómo será la situación a futuro para la población. El primero se toca dentro de lo que es la estimación puntual y por intervalos tal y como acababa de exponerse, mientras que el segundo sería abordado dentro de la unidad de Estimación por Mínimos Cuadrados. Los elementos y cuestiones más específicas de Muestreo se darían en la unidad correspondiente a Consideraciones en el Diseño de Experimentos, dando de este modo a la presente exposición un carácter más: el de servir como una introducción para todo el curso de Estadística II.

Unidad II, Funciones de distribución de probabilidad de variables aleatorias continuas

Objetivos:
Que el alumno recuerde las características de una variable aleatoria discreta y sus diferencias con una variable aleatoria continua.
Que el alumno conozca el concepto de variable aleatoria continua y sus características.
Que el alumno identifique la probabilidad de ocurrencia de valores de variables aleatorias continuas como áreas bajo la curva, a partir de problemas que lleven a una distribución Uniforme.
Que el alumno identifique problemas de cálculo de probabilidades que involucran variables aleatorias continuas.
Que el alumno conozca y aplique la distribución Uniforme como modelo de solución para ciertos problemas de cálculo de probabilidades, y que resuelva dichos problemas.
Que el alumno conozca y aplique la distribución Exponencial como modelo de solución para ciertos problemas de cálculo de probabilidades, y que resuelva dichos problemas.
Que el alumno conozca la distribución Normal y su estandarización a partir de la Regla Empírica.
Que el alumno identifique la distribución Normal Estandarizada como un modelo de solución para ciertos problemas de cálculo de probabilidades de variable aleatoria continua y que resuelva dichos problemas.

Como estrategia didáctica se recordará el concepto de variable aleatoria continua, y por medio del Problema de la Aguja se deducirán las características de las variables aleatorias continuas por medio de lluvia de ideas. Posteriormente se realizará, también a partir de las aportaciones de los estudiantes, la representación gráfica del Problema de la Aguja para deducir la probabilidad de ocurrencia de un valor dentro un intervalo de valores para la variable aleatoria como una fracción de área bajo la curva que describe el ejercicio, y con esto, se deducirán las características de una distribución Uniforme y la manera de resolver de manera sencilla problemas que se ajusten a este modelo, tras lo cual los alumnos resolverán problemas en el aula.
Para el caso de la distribución Exponencial, el profesor presentará directamente el modelo general de solución de problemas de esta naturaleza junto con un ejemplo, manteniendo siempre la idea de la probabilidad como una fracción de área bajo la curva, y posteriormente los alumnos resolverán ejercicios en el aula.
Para el caso de la distribución Normal, el profesor presentará un problema que permita deducir un comportamiento gráfico semejante al de la Campana de Gauss. De aquí se presentará la problemática del cálculo del área bajo la curva Normal, y, recordando la Regla Empírica, se le dará solución, para posteriormente deducir los alumnos la estandarización de la distribución Normal. Para concluir, el profesor mostrará a los alumnos el manejo de las tablas de áreas bajo la Normal Estándar.. Posteriormente los alumnos resolverán ejercicios en el aula.

2. Funciones de distribución de probabilidad de variable aleatoria continua

A manera de introducción se presentó al grupo el llamado Problema de la Aguja. Este experimento, recordando algunos de los elementos manejados en el curso anterior dentro de las unidades de Probabilidad Axiomática y Funciones de Distribución de Variable Aleatoria Discreta, nos remite a un espacio muestral continuo, dado que los resultados de este experimento están dentro de un intervalo de números reales. Ahora, si empezamos a asignar probabilidades de ocurrencia a resultados de esta naturaleza, propiamente estaremos generando una variable aleatoria continua, de la cual se especificaron sus características.

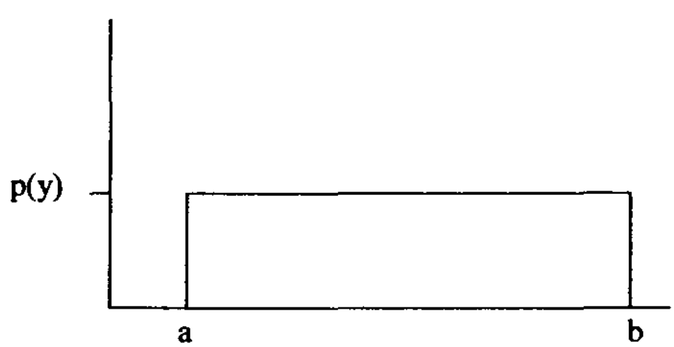
Y aquí podemos ver una característica del cálculo de probabilidades con resultados continuos, y es que las probabilidades puntuales invariablemente son de cero puesto que, en aplicación de la Definición Clásica, se estaría dividiendo una constante sobre infinito. De este modo, el cálculo de probabilidades igualmente puede hacerse sólo sobre intervalos. Generalizando, se debe tener siempre presente que, para variables aleatorias continuas ocurre que:

$P(Y = y) = 0.$

Una vez terminada esta introducción, se procedió al trabajo con específicamente tres funciones de distribución de variable aleatoria continua: la Uniforme, la Normal Estandarizada y la Exponencial. Cabe decir que en ningún momento se habló de funciones de densidad y de distribución, dado el desconocimiento de los alumnos en el Cálculo Integral.

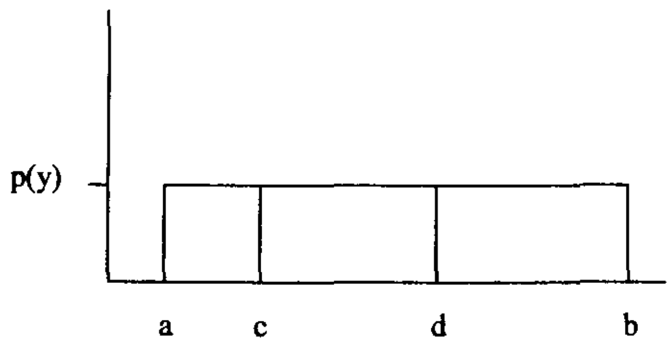
a) Distribución Uniforme

Esta distribución se caracteriza por que todos los posibles resultados del espacio muestral tienen la misma probabilidad de ocurrencia. En ese sentido todos y cada uno de los valores contenidos dentro del intervalo en el que pueden darse los resultados del experimento son equiprobables. De este modo, si graficamos el espacio muestral de esta distribución, asumiendo que el intervalo en el que pueden darse todos los posibles resultados del experimento está entre los valores a y b y con a < b, es decir, el espacio muestral sea el intervalo [a, b], tenemos que la Distribución Uniforme se puede representar como:



En donde p(y) es la probabilidad de ocurrencia de cualquier resultado del espacio muestral. Así, este espacio puede ser seccionado en distintos subintervalos y la probabilidad de obtener un resultado dentro de un subintervalo en particular será simplemente la proporción del rectángulo formado por el subintervalo en particular y p(y) con respecto al rectángulo total mostrado en la gráfica.

Sobre la base de lo anterior, se procedió a establecer el cálculo de la proporción recién mencionada. El problema consiste en lo siguiente: se tiene un espacio muestral continuo conformado por el intervalo [a, b] en el que todos los posibles resultados son igualmente probables, y se desea calcular la probabilidad de obtener específicamente un valor entre los valores digamos c y d, en donde c y d están contenidos en [a, b] y $c < d$; gráficamente tenemos esto:



De esta forma, tenemos que el área del rectángulo mayor representa la probabilidad de ocurrencia del espacio muestral, dado que es el 100% del área representativa del experimento, por lo que, si se desea calcular la proporción del rectángulo pequeño con respecto al mayor, entonces tendremos que calcular ambas áreas. Sabemos que el área de un rectángulo es el producto de la longitud de la base por la longitud de la altura, por lo que el rectángulo mayor, cuya base es $b - a$ y cuya altura es p(y), será $(b - a)p(y)$, y el área del rectángulo menor será $(d - c)p(y)$; así la proporción entre ambas áreas será:

$$\frac{(d-c)p(y)}{(b-a)p(y)} = \frac{d-c}{b-a},$$

con lo que podemos establecer que en una Distribución Uniforme cuyo espacio muestral es el intervalo [a, b], con $a < b$, la probabilidad de obtener un resultado en el intervalo [c, d], contenido en [a, b] y con $c < d$, puede calcularse como:

P(c < la probabilidad de obtener un resultado en el intervalo [c, d], contenido en [a, b] y con $c < d$, puede calcularse como:

$$\mathbf{P(c < Y < d)} = \frac{d-c}{b-a}.$$

Una vez hechas estas deducciones, se presentó la siguiente definición con la finalidad de formalizar un poco el tema:

-Definición.- Cuando una variable aleatoria asume cualquier valor en una escala continua entre dos puntos, de tal forma que ningún valor sea más probable que otro, entonces las probabilidades asociadas con la variable aleatoria se pueden describir mediante la Distribución Uniforme.

Terminada la parte teórica, a la que se recurriría de algún modo en las siguientes distribuciones, se ilustró el uso de esta distribución con un ejemplo.

-Ejemplo.- Supóngase que un vendedor telefonea a las oficinas centrales entre 3 y 4 de la tarde, y según el registro, en ningún momento de este lapso hay más probabilidad de que el vendedor llame. ¿Qué probabilidad hay de que el vendedor llame:

a) entre las 3 y las 3:15?
b) entre las 3:25 y las 3:38?
c) después de las 3:50?
Posteriormente se proporcionó al grupo la siguiente serie de problemas:

-Las ventas de combustible en una gasolinera tienen un máximo de 40,000 litros por día y un mínimo de 30,000 litros por día. Supóngase que estas ventas están distribuidas uniformemente. ¿Qué porcentaje de los días las ventas excederán de 34,000 litros?

-Una compañía maderera corta y vende leños para chimenea, cuya longitud varía uniformemente entre 60 y 90 centímetros. ¿Cuál es la probabilidad de que cualquier leño de la compañía tenga una longitud:
a) mayor que 78 cm?
b) mayor que 90 cm?
c) menor que 75 cm?
d) exactamente de 60 cm?
e) de entre 60 cm y 90 cm?

-Según se ha registrado en cierta región, la temperatura máxima en el mes de enero varía uniformemente de los 0° C a los 6° C. ¿Qué probabilidad hay de que en determinado día de enero en esa región:
a) la temperatura alcance un máximo de 3.5° C?
b) la temperatura no exceda de 1° C?
c) la temperatura sea de entre los 2.2° C y los 4.8° C?

-La cantidad diaria de helado vendido en una fuente de sodas está uniformemente distribuida entre 20 y 50 litros.
a) ¿Qué probabilidad hay de que cierto día las ventas sean de 40 lts o más?
b) ¿Qué probabilidad hay de que las ventas en cierto día sean inferiores a 40 lts?
Si la fuente de sodas obtiene una utilidad de $0.30 por cada litro de helado vendido ¿qué probabilidad hay de que en cierto día la utilidad por venta de helado sea:
c) mayor a $10.20?
d) menor a $7.50?

-Un autobús llega siempre a cierta parada entre las 8 y las 8:10 de la mañana. Hallar la probabilidad de que una persona que llega a la parada a las 8:03, y antes que el autobús, no tenga que esperarlo por más de 5 minutos. Supóngase que las llegadas se distribuyen uniformemente.

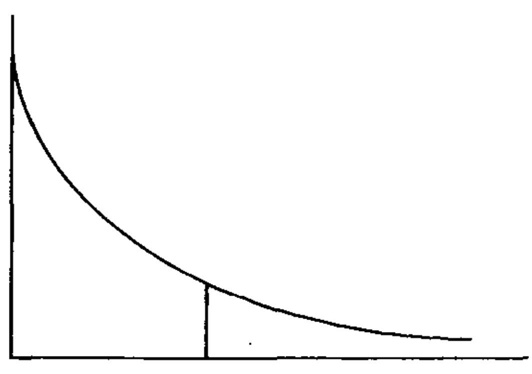
-Las llegadas de los clientes a una caja registradora tienen una distribución uniforme, y se sabe que durante un periodo dado de 30 minutos llegó un cliente a la caja. Encuentre la probabilidad de que el cliente halla llegado:
a) en los últimos cinco minutos del periodo.
b) en los primeros veinte minutos.
c) en los diez minutos intermedios.

-Si un paracaidista cae en un sitio aleatorio de la linea entre los puntos A y B, separados por 20 mts, encuentre la probabilidad de que esté mas cerca de A que de B. Calcular la probabilidad de que la distancia con respecto a A sea más de tres veces la distancia con respecto a B.

-El tiempo de un viaje (ida y vuelta) de los camiones que transportan el concreto hacia una obra en construcción en una carretera está distribuido uniformemente en un intervalo de 50 a 70 minutos. ¿Cuál es la probabilidad de que la duración del viaje sea:
a) mayor a 65 minutos?
b) menor a 55 minutos?
c) de entre 55 y 65 minutos?

b) Distribución Exponencial

Esta distribución se presentó como el "caso continuo" de la distribución de Poisson (con cierta característica de la Distribución Geométrica), y se dijo que su principal aplicación se da dentro de modelos de espacios de tiempo entre los que puede ocurrir un determinado resultado, como por ejemplo los tiempos entre las fallas que pueda presentar un cierto equipo. Y de hecho se tomó como punto de partida la Distribución de Poisson en el sentido de que, al ser λ el número promedio de ocurrencias de éxitos dentro del espacio-tiempo, entonces el espacio entre cada ocurrencia de éxito será de $\frac{1}{\lambda}$. También se dijo que esta distribución, dada la naturaleza del tiempo, es una distribución de variable aleatoria continua, y que, por lo tanto, podría manejarse, al igual que la Distribución Uniforme, como la proporción entre una sección del área bajo una curva que representara al experimento (definida con base en los valores de la variable aleatoria para los que se deseara calcular la probabilidad de ocurrencia) con respecto al área total bajo dicha curva; pero que dada la naturaleza de la curva en específico para esta distribución se presentarían directamente las expresiones necesarias para el cálculo de probabilidades sin pasar por su deducción. Así, se presentó la gráfica para la Distribución Exponencial como:



Aquí se dijo que la las probabilidades para la Distribución exponencial se expresan en términos de tiempo o distancia hasta que un evento no tiene lugar. De esta forma se puede calcular la probabilidad de que el espacio-tiempo antes de que se presente la primera ocurrencia (nótese la característica de primer éxito propia de la Geométrica) sea mayor que un espacio-tiempo de longitud t; o bien se puede calcular la probabilidad de que dicha ocurrencia se dé en un espacio-tiempo de longitud igual o menor a t. Así, se hablaría del porcentaje de área acotado después de t o antes de t respectivamente y de acuerdo con la gráfica presentada. Estas dos probabilidades, están dadas como:

$$P(Y > t) = e^{\frac{-t}{\lambda}},$$

y

$$P(Y \leq t) = 1 - e^{\frac{-t}{\lambda}}.$$

Se hizo notar aquí que en la segunda expresión simplemente se estaba planteando el complemento de la primera expresión.

Se presentaron estos dos ejemplos para dejar claro el funcionamiento de la Distribución Exponencial

-Ejemplo.- Se desea determinar la probabilidad de que un conmutador no reciba llamadas durante un periodo de 2 horas si la razón promedio es de 1.5 llamadas por hora.

-Ejemplo.- Suponga que el tiempo que tardan en recibir su orden en un gran restaurante después de hacerla promedia 10 minutos. Suponga también que ese tiempo se distribuye exponencialmente. Calcular la probabilidad de que el tiempo de espera sea:

a) mayor a 10 minutos.
b) de 10 minutos ó menos.
c) de 3 minutos ó menos.

Una vez resueltos los ejemplos, se dieron los siguientes problemas al grupo:

-Suponga que una máquina falla en promedio cada dos años. Encuentre la probabilidad de que la máquina no falle durante el siguiente año.

-Los defectos en las cuerdas de nylon están distribuidas exponencialmente con una media de un defecto por metro. Obtenga la probabilidad de no encontrar un defecto en los primeros 3 metros de un carrete de este hilo.

-El tiempo de servicio en un almacén de herramientas es aproximado adecuadamente mediante una distribución exponencial, con una media de 4 minutos. ¿Cuál es la probabilidad de que el tiempo de servicio sea:

a) mayor de 4 minutos?
b) menor de 4 minutos?
c) exactamente de 4 minutos?

-Las llamadas de emergencia durante las primeras horas de la mañana del lunes siguen un modelo exponencial con un tiempo medio de 1 hora entre cada llamada.

a) Calcular la probabilidad de tener un periodo de 2 horas sin que se registre ninguna llamada.
b) Obtener la probabilidad de tener periodo de 3 horas sin que halla ninguna llamada.

-Un satélite de comunicación tiene una sola fuente de energía. Determine la probabilidad de que el satélite funcione durante por lo menos 20,000 horas antes de que se dé una falla de energía, si el tiempo promedio entre fallas es:

a) 10,000 horas.
b) 20,000 horas.
c) 40,000 horas.

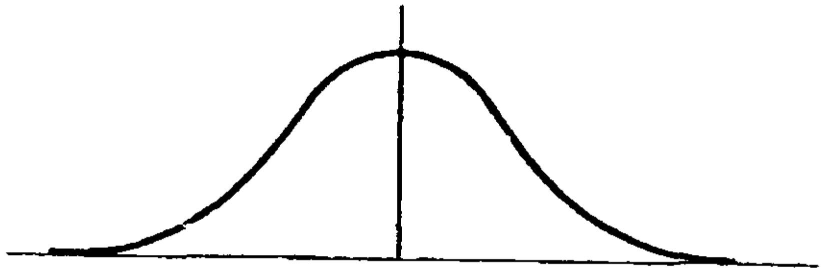
c) Distribución Normal

Al comenzar con este tema se hizo mención que muchos de los fenómenos cuantificables que pueden observarse a través de los elementos de Estadística Descriptiva presentan un comportamiento similar, y que, gráficamente, este comportamiento puede ser representado por medio del Polígono de Frecuencias, gráfico cuya forma se recordó como

por presentar un ejemplo.

Ahora, este gráfico no sólo representa la forma en la que se distribuyen las frecuencias absolutas para una serie de valores de alguna variable, sino también permite una visión clara del comportamiento de la muestra en lo tocante a la Regla Empírica. De este modo, si queremos darnos una idea del comportamiento que pudieran tener uno o varios elementos de una población con respecto a algún elemento descriptivo podríamos recurrir al gráfico como un auxiliar.

El problema aquí es que el Polígono de Frecuencias no puede ser trazado si no es con los datos directamente tomados de la observación, ya que no hay una "fórmula" que lo genere, entre otras cosas por ser único y diferente para cada serie de datos. Y en este caso entonces se recurre a lo que en Matemáticas se conoce como Ajuste de Curvas –tema que será tratado más a profundidad en la unidad correspondiente a Estimación por Mínimos Cuadrados-, y que es un proceso en el que básicamente se busca una función conocida a la que se asemeje la gráfica de interés con la idea de darle a esta última el tratamiento general de dicha función. Así, primero se busca una función que muestre aproximadamente el comportamiento de los polígonos de frecuencias, encontrando como la que más se le parece la curva definida por la función $F(Y)=\frac{e^{\frac{-(y-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$, que es la función de densidad de probabilidad Normal, y después se asume un comportamiento similar para los polígonos de frecuencias que el que se tiene con esta función. Nótese que la función involucra, entre otros elementos, a la media μ y a la desviación estándar σ. La gráfica de la función es



Se explicó aquí que, igual que en las distribuciones Uniforme y Exponencial. en la Distribución Normal igualmente se busca como la probabilidad de ocurrencia de un evento una fracción de área bajo la curva, y que para ello podríamos realizar una de dos cosas: la primera es integrar la función de densidad Normal, cosa que no es sencilla para quien maneja el Cálculo Integral y que es imposible para el que no lo maneja, o la segunda, consistente en tratar de simplificar el problema asumiendo la aplicación de la Regla Empírica a la curva Normal, dado que la podemos tomar como una "representación general" de los polígonos de frecuencias, además de que dicha función involucra los elementos media y desviación estándar, implicados en la Regla Empírica. Esta segunda alternativa puede tomarse de la siguiente forma: la curva Normal es una curva simétrica, y dado que podemos "movernos" sobre el eje horizontal tanto a la derecha como a la izquierda, y que la media está a cero desviaciones estándar de sí misma, entonces tomaremos el eje de simetría de la gráfica como el punto en el que queda situada dicha media, para posteriormente restarle o agregarle desviaciones estándar y recurrir a la regla empírica para verificar las porciones de áreas, y ajustarlas de ser necesario.

Con el fin de dejar más clara esta transformación se plantearon problemas consistentes en encontrar por tanteo el número de desviaciones estándar al que se encontraban de una media ciertas observaciones.

Ahora, de aquí se partió para indicar que, en general, al restarle μ a un cierto valor Y estaríamos midiendo la distancia de Y a μ y que dicha distancia sería una determinada cantidad de veces la desviación estándar. Si llamamos a dicha cantidad de veces σ como z, entonces podemos escribir que $Y-\mu=z\sigma$, pero en realidad nos interesa conocer directamente el valor de z si queremos verificar para cada valor de Y a que distancia de μ en términos de desviaciones estándar nos encontramos, con la finalidad de aplicar la Regla Empírica del mismo modo en que se hizo con el ejemplo anterior, por lo que, despejando, encontramos que $z=\frac{Y-\mu}{\sigma}$, expresión que se conoce como *Normal Estandarizada.*

Finalmente queda por resolver un problema: la Regla Empírica nos dice qué ocurre al restar o sumar al promedio una, dos, tres y tres y media veces la desviación estándar, pero no habla por ejemplo de qué fracción de área está en el intervalo obtenido al restar y sumar de la media 1.72 veces la desviación estándar, por citar un ejemplo. Pero este problema queda sorteado en términos de que para valores de z desde 0 hasta 3.5 existen tablas que especifican qué fracción de área corresponde a valores mayores a dicha z (o menores a ella, o entre z y cero, o entre z y –z). Dichas tablas, tomadas del libro de *Estadística Matemática con Aplicaciones* de William Mendenhall y que pueden encontrarse al final del presente trabajo como *Areas bajo la curva Normal*, fueron puestas a disposición del grupo, y cuyo manejo fue explicado por medio de un ejemplo, y una vez terminado éste se especificó al grupo que a esta distribución se le denomina Normal en función de que la gran mayoría de los fenómenos aleatorios continuos tienden a comportarse precisamente de esta manera. El ejemplo fue el siguiente:

-Ejemplo.- Los resultados de un examen en un colegio tienen una Distribución Normal con una media de 75 y una desviación estándar de 10. ¿Qué fracción de los resultados cae:
a) arriba de 85?
b) abajo de 60?
c) arriba de 55?
d) entre 80 y 90?
e) entre 55 y 95?

Después se dieron los siguientes problemas para ser resueltos por los alumnos:

-Se observó durante cierto periodo que la cantidad semanal gastada en el mantenimiento y en las reparaciones en cierta fábrica tiene aproximadamente una Distribución Normal con una media de $400.- y una desviación estándar de $20.-. Si el presupuesto para la próxima semana es de $450.-, ¿cuál es la probabilidad de que los costos reales sean mayores que la cantidad presupuestada?

-Una empresa produce rodamientos con diámetros que tienen una Distribución Normal con una media de 3.0005 cm. y una desviación estándar de 0.001 cm. Las especificaciones requieren que los diámetros estén en el intervalo [3 $\pm$ 0.002] cm. Se rechazan los cojinetes que queden fuera del intervalo y deben volverse a maquinar. Con la maquinaria actual, ¿qué fracción de la producción total será rechazada?

-Los promedios de las calificaciones de una gran población de estudiantes en un colegio tienen aproximadamente una Distribución Normal con una media de 2.4 y una desviación estándar de 0.8. ¿Qué fracción de los estudiantes tendrá un promedio superior a 3?

-Refiérase al ejercicio anterior. Si se expulsan del colegio a los estudiantes que tienen un promedio igual o menor a 1.9, ¿qué porcentaje de los estudiantes será expulsado?

-Los alambres que se usan en cierta computadora deben tener una resistencia de entre 0.12 y 0.14 ohms. Las resistencias reales de los alambres producidos por la compañía A tienen una Distribución Normal con una media de 0.13 ohms y una desviación estándar de 0.005 ohms. ¿Cuál es la probabilidad de que un alambre seleccionado al azar de la producción de la compañía A satisfaga las especificaciones para la computadora?

-Un método para hacer predicciones económicas es la "aproximación por consenso". Se obtiene el pronóstico de cada uno de un gran número de analistas, y el promedio de los pronósticos individuales es el pronóstico general. Suponga que los pronósticos individuales de enero de 1997 con respecto a la tasa de interés mínima de todos los analistas económicos tienen aproximadamente una Distribución Normal con una media igual a 14% y una

desviación estándar de 2.6%. Si se selecciona al azar a un solo analista de este grupo, ¿cuál es la probabilidad de que el pronóstico de la tasa de interés mínima del analista sea:
a) mayor que 18%?
b) menor que 16%?

-El diámetro de los pernos de una fábrica tiene una Distribución Normal con una media de 950 mm. y una desviación estándar de 10 mm. ¿Cuál es la probabilidad de que un perno escogido al azar tenga un diámetro entre 947 mm y 958 mm?

-Se supone que los resultados de una prueba de pesos tienen una Distribución Normal con una media de 78 kg y una variancia de 36. ¿Cuál es la probabilidad de que al pesar a una persona ésta tenga más de 72 kg?

Al finalizar con esta serie de problemas se planteó otra forma de trabajar con la Distribución Normal consistente en hallar z a partir del valor de probabilidad –o lo que es igual, de la fracción de área bajo la curva-, esto con el fin de introducir al grupo en la forma en que la tabla se manejará en los temas de Intervalos de Confianza y de Prueba de Hipótesis.

Con este fin se presento un problema al grupo consistente en el cálculo de z a partir de la probabilidad. De este modo, se trabajaría con las tablas buscando entre los valores de probabilidad el que resultara igual o más cercano a la probabilidad planteada, para luego ubicar el valor de z a partir del renglón y a la columna de centésimas en las que se encontrara dicha probabilidad. Igualmente se sugirió hacer el boceto de curva Normal que representara el problema.

Una vez entendida la mecánica para la obtención de z a partir de las probabilidades, se resolvió un ejemplo, tomado de un ejercicio resuelto por el grupo.

-Ejemplo.- En el problema de los costos de mantenimiento en la fábrica, ¿de cuánto tendría que ser el presupuesto para que la cantidad presupuestada solamente sea rebasada con una probabilidad de 0.1?

Para finalizar esta parte, se presentó al grupo el siguiente problema para su resolución:

-En el problema de los rodamientos, ¿de cuánto deberá ser el diámetro máximo para que la fracción de rodamientos rechazados a causa de presentar un diámetro demasiado grande sea del 0.5%?

Aproximación de la Binomial a la Normal

Este subtema se presentó con la finalidad de dar una primera aproximación a las expresiones que se manejarán en Intervalos de Confianza y Prueba de Hipótesis para la proporción, además de permitir el manejo de problemas de naturaleza Binomial en los que el número de ensayos es demasiado grande, recordando aquí que una primera alternativa ya había sido manejada en el subtema Aproximación de la Binomial a la Poisson en el curso de Estadística I.

Se presentó al grupo la alternativa de un problema Binomial con muchos ensayos, y se comentó que nos encontraríamos con serias dificultades no sólo en el cálculo de las combinaciones sino también el cálculo de las potencias, sobre todo si p y q resultaran ser una muy grande y otra muy pequeña. Además, utilizar la Distribución de Poisson también resultaría impráctico ya que tendrían que calcularse y sumarse demasiadas probabilidades. De este modo, en problemas de esta índole resulta conveniente otro tipo de solución: la Distribución Normal, ya que la Binomial puede manejarse aquélla para los casos en los que el número n de ensayos es demasiado grande, específicamente mayor o igual a 30, a través de la expresión:

$$z = \frac{\frac{y}{n} - p}{\sqrt{\frac{pq}{n}}}.$$

Los elementos involucrados son, en el caso del lado derecho de la igualdad, todos aquellos involucrados en los problemas de tipo Binomial: n como número de ensayos, p como probabilidad de éxito, q como probabilidad de fracaso y y como número de éxitos sobre cuya ocurrencia se busca calcular la probabilidad, y aquí cabe decir que est número invariablemente estará planteado en estos casos dentro de una desigualdad, ya que sería ocioso presentarlo dentro de una igualdad estricta desde el momento en que para las funciones de probabilidad de variable aleatoria continua, como lo es la Normal, las probabilidades puntuales, como ya se manejó, son iguales a cero. Y en el caso del lado izquierdo, tenemos a la misma z ya vista en la Distribución Normal y que es el número de desviaciones estándar a la media a las que se sitúa cierto valor, y que se utiliza para efectos de lectura de la tabla.

Para efectos de verificar la manera de realizar esta aproximación, nuevamente se resolvió un ejemplo demostrativo.

-Ejemplo.- Un candidato considera que puede ganar una elección en una ciudad si obtiene al menos el 55% de los votos en el Distrito 1. Además supone que alrededor del 50% de los votantes en la ciudad están a su favor. Si 100 votantes acuden a votar en el Distrito 1, ¿cuál es la probabilidad de que el candidato reciba al menos 55% de los votos?

Para finalizar con el subtema y con la unidad, se dieron al grupo los siguientes problemas:

-Del total de focas nacidas en el Artico, 35% son muertas por cazadores de pieles cuando aún son bebés. De una generación de 100 focas, ¿qué probabilidad hay de que al menos 55 de ellas no sean victimadas por los cazadores?

-¿Qué probabilidad hay de que de 100 ratones de laboratorio, más de 90 se condicionen exitosamente, si el porcentaje de condicionamiento es del 85%?

-En un experimento Binomial con p = 0.001 y 100 repeticiones, ¿qué probabilidad hay de que se obtengan:
a) 10 o menos éxitos?
b) entre 20 y 30 éxitos?

-La probabilidad de hacer una tirada ganadora en un juego de casino es 0.22. Si se realizan 120 tiradas, ¿qué probabilidad hay de ganar más de 34 de ellas?

-Si la probabilidad de lluvia en cierta región de la Tierra es de 0.03 por día, ¿qué probabilidad hay de que haya más de dos días de lluvia en los siguientes 100 días?

-La probabilidad de que una persona tenga cierto tipo de sangre es 0.075. ¿Qué probabilidad hay de que haya más de 12 personas con dicho tipo de sangre de entre 250 donadores?

En esta unidad la dificultad más seria que hubo que enfrentar fue al momento de buscar z a partir de la fracción de área. Usualmente los alumnos tomaron dicha fracción como una z y encontraron su probabilidad de ocurrencia. Se requirió de trabajar individualmente con todos aquellos que cometieron el error y en ocasiones sobre más de un ejemplo para lograr la comprensión. Más adelante, en las dos unidades siguientes, se pudo observar que con la práctica la gran mayoría consiguieron revertir el proceso de lectura de tablas que se enseñó en un principio, y fueron capaces de encontrar z a partir del área asociada a ella.

Unidad III, Intervalos de confianza

Objetivos:
Que el alumno conozca el concepto de estimación por intervalos y de los elementos que lo componen.
Que el alumno conozca los modelos de estimación por intervalos que utilizan las distribuciones muestrales Normal, t de Student Ji-Cuadrada y F de Snedecor y resuelva problemas de estimación de los parámetros media, proporción, diferencia de medias, diferencia de proporciones, variancia y cociente de variancias con el modelo correspondiente e interpretando las soluciones.

Como estrategia didáctica se discutirá en el aula la inconveniencia de la estimación puntual y las ventajas de la estimación por intervalo, con la idea de generar la necesidad de este tipo de estimación y enfatizando la importancia de la interpretación de los resultados que de este tipo de estimación se puedan obtener. Después, y a partir de utilizar la distribución Normal Estandarizada, se llegará a una expresión general de estimación por intervalos utilizando el modelo Normal. Posteriormente el profesor presentará los otros tres modelos de estimación por intervalos contemplados para la unidad por medio de ejemplos y mostrando el manejo de las tablas correspondientes. Al final de las deducciones y/o explicaciones propias de cada modelo, los alumnos resolverán ejercicios en el aula.

3. Intervalos de confianza

Para esta unidad se recordaron algunas de las consideraciones hechas al inicio del curso, en la unidad de Muestreo y Estimación, en dónde se comentó que la estimación por intervalos consiste en generar un intervalo con valores obtenidos de una muestra dentro del que se espera esté algún parámetro en particular. Ahora, se sabe que una muestra es tomada de manera aleatoria, por lo que deberá de incluirse en la estimación un elemento de probabilidad, y este elemento precisamente nos permitirá dar cierto grado de certidumbre a nuestras estimaciones; esto es, si sabemos como se comporta la población de interés, o al menos la muestra con la que se trabaje, en términos de distribución de probabilidad, entonces podemos decidir el porcentaje de seguridad para nuestra estimación, es decir, podemos establecer de antemano la probabilidad de que, al construir el intervalo de confianza, efectivamente consigamos que dentro de éste se encuentre el parámetro a estimar.

a) Intervalo de confianza para la media, la proporción, la diferencia de medias y la diferencia de proporciones con muestras grandes ($n \geq 30$)

En el caso de la estimación por intervalos existen distintas posibilidades de trabajar, dependiendo tanto del parámetro a estimar como de algunas características de la muestra con la que se trabaje. En ese sentido se especificó que en esta unidad se verían cuatro modelos de estimación por intervalos, siendo el primero de ellos el correspondiente a la estimación para los parámetros media, proporción, diferencia de medias y diferencia de proporciones a partir de una muestra grande, entendiendo como tal a aquella muestra conformada de 30 o más elementos. Y aquí se hizo el comentario de que, para estos parámetros en particular, toda muestra grande se distribuye de manera normal – específicamente la referencia corresponde al Teorema del Límite Central, pero dado que no está contemplado el tema de Distribuciones Muestrales, este Teorema no se planteó de manera formal-, por lo que en este subtema se trabajaría directamente con esta distribución.

En este punto entonces se presentó de nuevo la expresión correspondiente a la Distribución Normal Estándar $z = \frac{y-\mu}{\sigma}$, pero se recordó que en esta distribución se medía la probabilidad sobre una sola observación, mientras que aquí se tiene el caso de trabajar con 30 o más de ellas, es decir, con una muestra, por lo que se tienen que realizar algunos ajustes sobre la expresión, consistentes, uno, en generalizar para cualquiera de los parámetros y estimadores de interés, y otro, en manejar la desviación estándar correspondiente a cada estimador. Para esto, primero se dijo que se utilizaría la notación planteada en la unidad de Muestreo y Estimación dejando z como $z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$ dónde θ es el parámetro a estimar, $\hat{\theta}$ es el estimador puntual correspondiente, y $\sigma_{\hat{\theta}}$ es la desviación estándar del estimador, y segundo, planteando el siguiente cuadro en el que se especifican cada uno de los tres elementos establecidos dentro de la nueva expresión de z:

Cuadro de Estimadores Puntuales.

PARAMETRO θ	ESTIMADOR $\hat{\theta}$	TAMAÑO DE LAS MUESTRAS	DESVIACION ESTANDAR DEL ESTIMADOR $\sigma_{\hat{\theta}}$
μ	$\bar{y}$	n	$\sqrt{\frac{\sigma^2}{n}}$
p	$\hat{p}$	n	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	n_1 y n_2	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	n_1 y n_2	$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

Ahora bien, este cuadro tendrá la función de "ajustar" la expresión genérica de z para cada parámetro en específico que presente en cada problema; pero antes se requiere encontrar la expresión para calcular los límites del intervalo de confianza, cuestión que se desarrolló como sigue.

Primero se retomó el hecho de que podemos establecer libremente la probabilidad de que nuestra estimación sea acertada, de modo que denotaremos dicha probabilidad como $1 - \alpha$ y la llamaremos de aquí en adelante como Coeficiente de Confianza, por ser precisamente el grado de confianza que estableceremos para nuestras estimaciones. De esta forma podemos decir que la probabilidad de que la estimación sea fallida es de α, pero existen dos maneras de fallar la estimación: una consistente en que el parámetro sea en realidad menor al límite inferior del intervalo, y la otra consistente en que el parámetro sea en realidad mayor al límite superior del intervalo; así tendremos que repartir α para cada una de las opciones como $\alpha/2$, dado que se tiene una Distribución Normal Estándar para la muestra, distribución que sabemos es simétrica. Ahora, es evidente que se buscará que α sea lo más pequeña posible, por lo que en la curva Normal, estaríamos tomando fracciones de área bajo la curva de $\alpha/2$ sobre las partes más extremas de las colas, áreas que serían limitadas por valores de –z y de z, y que simbolizaremos como $-z_{\frac{\alpha}{2}}$ y $z_{\frac{\alpha}{2}}$, al ser valores que se obtendrán a partir de $\alpha/2$. Así mismo, dentro de los límites de estos valores z, se tendrá precisamente el intervalo de confianza; y dado que la probabilidad de "atrapar" entre estos dos valores al parámetro de interés se establece como $1 - \alpha$, entonces podemos establecer que:

$$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha.$$

Si sustituimos z por la expresión planteada en un principio y realizamos una serie de despejes tendremos:

$$P(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}) = 1 - \alpha$$

$$P(-z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}} < \theta - \hat{\theta} < z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}}) = 1 - \alpha$$

$$P(-\hat{\theta} - z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}} < -\theta < -\hat{\theta} + z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}}) = 1 - \alpha$$

$$P(\hat{\theta} + z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}} > \theta > \hat{\theta} - z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}}) = 1 - \alpha$$

$$P(\hat{\theta} - z_{\alpha/2}\,\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\frac{\alpha}{2}}\,\sigma_{\hat{\theta}}) = 1 - \alpha$$

con lo que hemos "cercado" el parámetro θ con dos límites y cumpliendo las condiciones preestablecidas de probabilidad, por lo que dichos límites serán precisamente los del intervalo de confianza, por lo que podemos expresar el cálculo de dicho intervalo como: $\hat{\theta} \pm z_{\alpha/2}\,\sigma_{\hat{\theta}}$.

Una vez que hemos encontrado la expresión para el cálculo de los límites del intervalo, se planteó al grupo un punto importante en la Estadística Inferencial: no es suficiente el cálculo numérico de un resultado, sino que también debemos saber interpretarlo, y para ello deben tenerse en cuenta básicamente el parámetro con el que se está trabajando y el contexto en el que se presente el problema. Y en este sentido se especificó al grupo que en todo problema de estimación por intervalo invariablemente deberá darse una interpretación del intervalo.

Una vez mostrado el funcionamiento práctico por medio de un ejemplo.

-Ejemplo.- Se registraron los tiempos utilizados en la compra para 64 clientes seleccionados al azar en un supermercado. La media y la variancia de los 64 clientes fueron 33 minutos y 256 respectivamente. Estime el valor del tiempo medio realizado por los clientes en la compra con un coeficiente de confianza del 90%. Interprete el resultado.

Para finalizar, se le dio al grupo la siguiente serie de problemas para su resolución, aprovechando su resolución para explicar que en los casos de la diferencia de medias y de la diferencia de proporciones lo que se busca es comparar las medias o las proporciones, por lo que si los dos límites del intervalo son positivos eso implicará que la primera población tiene una media (o proporción) mayor que la segunda, si los dos límites son negativos eso implicara que la segunda población tiene una media (o una proporción) mayor que la primera, y si el límite inferior es negativo y el superior positivo eso implicará que no se tiene la suficiente información para inferir una diferencia en el parámetro a favor de alguna de las dos poblaciones:

-Dos marcas de refrigeradores, A y B, tienen ambas una garantía de un año. En una muestra aleatoria de 50 refrigeradores de la marca A, 12 se descompusieron antes de terminar el periodo de garantía. Una muestra aleatoria de 60 refrigeradores de la marca B reveló también 12 descomposturas durante el periodo de garantía. Estime la diferencia real entre las proporciones de fallas durante el periodo de garantía con un coeficiente de confianza del 98% e interprete el resultado.

-Cierto periódico reportó que los estudiantes de Derecho se oponen a la pena de muerte. Se hizo esa declaración basándose en una encuesta para la cual se escogieron al azar y entrevistaron a 86 estudiantes de Derecho. El 52% de los entrevistados declararon que se oponían a la pena de muerte. A partir de esta información obtenga un intervalo de confianza del 95% para la proporción real de los estudiantes de Derecho que se oponen a la pena de muerte. ¿Se justifica la declaración del periódico?

-La Dirección Médica de una clínica desea estimar el número promedio de días necesarios para el tratamiento de pacientes con edades entre 25 y 34 años. Una muestra aleatoria de 500 pacientes de la clínica con esas edades proporcionó una media y una desviación estándar de 5.4 y 3.1 respectivamente. Obtener e interpretar el intervalo de confianza del 95% para el promedio de tiempo de estancia de la población de pacientes de la cual se obtuvo la muestra.

-Se publicaron los resultados de un estudio sobre la relación entre la práctica deportiva y la destreza manual. De una muestra aleatoria de 37 niños que practicaron deporte se obtuvo una calificación media de destreza manual de 32.19 y una desviación estándar de 4.34. De una muestra aleatoria independiente de 37 niños que no practicaron deporte se calculó una calificación media de destreza manual de 31.68 y una desviación estándar de 4.56. Estime la diferencia entre los promedios reales de los resultados para los dos grupos con un intervalo de confianza del 90%. ¿Parecería que la calificación promedio de destreza manual para quienes practican deporte difiere de la calificación promedio de quienes no lo practican?

-Según un reporte ecológico, la lluvia ácida causada por la reacción de ciertos contaminantes en el aire con el agua de la lluvia es un problema creciente que afecta al suelo y corroe las superficies metálicas expuestas. La lluvia pura que se precipita a través del aire limpio tiene un pH de 5.7 (el pH es una medida para la acidez: el 0 es ácido y el 14 es alcalino). Supóngase que se analizan 40 lluvias con respecto a su pH y que la media y la desviación estándar son iguales a 3.7 y a 0.5 respectivamente. Determinar el intervalo de confianza del 99% para la media del pH en las lluvias e interpretarlo.

-Para comparar las proporciones en los artículos defectuosos producidos por dos líneas de producción, se seleccionan muestras aleatorias independientes de 100 artículos de cada línea. La línea A produjo 18 artículos defectuosos en la muestra, y la línea B 12. Obtener un intervalo de confianza del 98% para la diferencia real entre las proporciones de artículos defectuosos para las dos líneas e interpretarlo.

-Una encuesta incluye las entrevistas por teléfono de 497 adultos seleccionados al azar. Al preguntárseles si consideraban decisivo el sexo de un candidato político, el 62% de los hombres y el 49% de las mujeres opinaron que el sexo del candidato no importaba. Si se entrevistaron a 241 hombres y 256 mujeres, construya un intervalo de confianza del 99% para la diferencia entre las proporciones de hombres y mujeres para los cuales el sexo del candidato no importa. Interpretar el resultado.

-Una encuesta realizada con respecto a la política de jubilaciones reveló que una alta proporción de trabajadores es muy pesimista con respecto a sus perspectivas cuando lleguen a jubilarse. Al preguntárseles si consideran que su jubilación será suficiente, 62.9% de los 6100 trabajadores entrevistados indicaron que su ingreso al jubilarse definitivamente no sería suficiente. Calcular un intervalo de confianza del 95% para la proporción de todos los trabajadores que consideran que al jubilarse su ingreso por pensión no será suficiente. Interpretar el intervalo.

b) Intervalo de confianza para la media y la diferencia de medias con muestras chicas (n < 30)

Para este subtema se planteó que al trabajar con muestras chicas, la distribución de probabilidad de la muestra deja de tener un comportamiento Normal en forma general, y pasa a tomar la misma distribución que tenga la población de la que fue tomada. Ahora, se ofrece aquí una alternativa para la estimación de la media y de la diferencia de medias a través de la Distribución t de Student, que nos da las bases probabilísticas para la estimación, siempre y cuando la o las muestras con las que se trabaje hayan sido seleccionadas de una población Normal, por lo que incluiremos este supuesto en todos los problemas de estimación para estos dos parámetros que se sustenten en información obtenida a partir de muestras chicas; y de la misma manera deberemos incluir el supuesto de igualdad en las variancias poblacionales para el caso de estimar la diferencia entre dos medias. Igualmente se dijo que en este punto ya no se trabajaría con pruebas para proporciones.

Para poder realizar las estimaciones en estas condiciones, se proporcionó al grupo la tabla de *Puntos porcentuales de las distribuciones t* tomada del libro de Estadística Matemática con Aplicaciones de William Mendenhall, y que puede encontrarse en el anexo al final de este trabajo. Se dijo que la lectura de la tabla se explicaría con el ejemplo de estimación y que si bien su manejo es sumamente sencillo, se requiere de un nuevo elemento para la lectura, conocido como grados de libertad. Respecto a este punto, se dijo de los grados de libertad que cuando se tiene una muestra de n elementos, y de ellos se conoce un parámetro, como lo puede ser la media muestral, entonces podemos asignar valores de forma un tanto arbitraria o libre a n-1 de los elementos de la muestra, pero que el n-ésimo elemento quedaría determinado tanto por el estimador como por los valores que asignáramos, y que precisamente como grados de libertad entenderíamos el número de elementos de la muestra a los que podríamos dar valores, y a partir de conocer algún estimador de dicha muestra (esta idea se tomó de la página 338 del libro de Taro Yamane referido en la bibliografía). Y para redondear esta explicación inicial se presentó el siguiente cuadro en el que se establecen directamente las fórmulas para el cálculo de los límites del intervalo de confianza para cada uno de los dos parámetros con los que se va a trabajar:

PARAMETRO	ESTIMADOR	INTERVALO	GRADOS DE LIBERTAD (g. l.)
μ	$\bar{y}$	$\bar{y} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}$	v = n - 1
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	$(\bar{y}_1 - \bar{y}_2) \pm t_{\frac{\alpha}{2}} S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ donde $S = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$	v = $n_1 + n_2$ - 2

Se dijo aquí que el elemento $t_{\frac{\alpha}{2}}$ era el valor obtenido de las nuevas tablas así como $z_{\frac{\alpha}{2}}$ fue el valor obtenido de las tablas de la Normal.

De este modo se explicó la parte práctica con un ejemplo, en el cual se mostró también el uso de la tabla correspondiente.

-Ejemplo.- Los resultados de una muestra aleatoria de 16 personas que presentaron un examen de idioma tuvieron una media de 540 y una desviación estándar de 50. Obtener un intervalo de confianza del 95% para la media poblacional de los resultados del examen, suponiendo que estos tienen una Distribución Normal. Interpretar el resultado.

Al final se dieron los siguientes problemas a resolver al grupo:

-Un fabricante de pólvora desarrolló una nueva fórmula que se probó con 8 granadas. Las velocidades iniciales resultantes, en metros por segundo, fueron las siguientes: 3005, 2925, 2935, 2965, 2995, 3005, 2937 y 2905. Encuentre un intervalo de confianza del 95% para la media real en las velocidades para granadas de este tipo e interprételo. Suponga que las velocidades iniciales se distribuyen de manera Normal.

-La tasa de consumo de oxígeno es una medida importante de la actividad fisiológica de los corredores. Se publicó el informe respecto a las diferencias en las tasas de consumo para atletas entrenados con dos métodos diferentes, uno que utiliza el entrenamiento continuo durante cierto lapso cada día, y otro que utiliza un entrenamiento intermitente de una duración total igual. Las medias, las desviaciones estándar y los tamaños de las muestras se indican en la tabla, en mililitros por kilogramo minuto. Si se supone que las mediciones proceden de poblaciones con distribuciones normales de variancias iguales, estime la diferencia en las medias poblacionales con un coeficiente del 95%. Interprete el resultado.

Entrenamiento continuo	Entrenamiento intermitente
$n_1 = 9$	$n_2 = 7$
$\bar{y}_1 = 43.71$	$\bar{y}_2 = 39.63$
$s_1 = 5.88$	$s_2 = 7.68$

-Se seleccionan 15 resistencias de la producción de un proceso que produce supuestamente resistencias de 10 omhs. Las 15 resistencias tuvieron una media muestral de 9.8 ohms y una desviación estándar de la muestra de 0.5 ohms. Obtener un intervalo de confianza del 95% para la verdadera media de los valores de las resistencias producidas por este proceso e interpretarlo. Suponga que las mediciones tienen aproximadamente una distribución normal.

-Una operación de montaje en una fábrica requiere un periodo de entrenamiento para que un nuevo empleado alcance la máxima eficiencia. Se sugirió un nuevo método para el entrenamiento y se realizó una prueba para comparar el método nuevo con el procedimiento estándar. Se entrenaron dos grupos de 9 empleados nuevos con los dos métodos diferentes y se midió el tiempo en minutos que necesitó cada empleado para montar el dispositivo. Las mediciones se muestran en el cuadro. Estime la diferencia real de las medias con un coeficiente de confianza del 98%. Suponga que los tiempos de montaje tienen aproximadamente una Distribución Normal y que las variancias son iguales para los dos métodos e interprete el resultado.

Estándar	32	37	35	28	41	44	35	31	34
Nuevo	35	31	29	25	34	40	27	32	31

-Se administraron dos nuevos medicamentos a pacientes con un padecimiento cardiaco. El primer medicamento bajó la presión sanguínea de 16 pacientes en un promedio de 11 puntos, con una desviación estándar de 6 puntos. El segundo medicamento bajó la presión sanguínea de otros 20 pacientes en un promedio de 12 puntos, con una desviación estándar de 8 puntos. Determinar un intervalo de confianza del 90% para la diferencia en la reducción media de la presión sanguínea suponiendo que las mediciones tienen distribuciones normales con variancias iguales e interpretar el intervalo.

-Una agencia para la protección ambiental reunió información con respecto a mediciones de CL50 (concentración letal que mata al 50% de los animales de experimentación) para ciertos productos químicos que se pueden encontrar probablemente en ríos y lagos de agua dulce. Para cierta especie de peces las mediciones de CL50 para DDT en 12 experimentos fueron: 16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4 y 9 (las mediciones se indican en partes por millón). Estimar el verdadero promedio de CL50 para DDT con un coeficiente de confianza del 90% e interpretar. Suponga que las mediciones tienen una Distribución Normal.

-En referencia con el ejercicio anterior, otro insecticida común, diazinón, dio las siguientes mediciones de CL50 en tres experimentos: 7.8, 1.6 y 1.3.
a) Estimar la media de CL50 para diazinón por medio de un intervalo de confianza del 90%.
b) Estimar la diferencia en los promedios de CL50 para DDT y para diazinón mediante un intervalo de confianza del 90%.
Supóngase que ambas poblaciones son normales y de variancias iguales e interprete ambos resultados.

-Una fábrica de artículos deportivos realizó un estudio con respecto a los impulsos repartidos a la pelota con raquetas de tenis de diversos materiales. Se le dieron tres impulsos para cada tipo de raqueta, y para una raqueta de madera la media fue de 2.41 y la desviación estándar de 0.02, mientras que para una raqueta de grafito la media fue de 2.22 y la desviación estándar de 0.07. Estimar la diferencia real en impulsos promedio para las dos raquetas con un coeficiente de confianza del 98%, suponiendo que las poblaciones son normales y de variancias iguales, e interpretar el resultado.

-En cierto laboratorio se mide la porosidad del cobre que se produce al sintetizar un polvo bajo ciertas condiciones. Una muestra de cuatro mediciones independientes de la porosidad tuvo una media de 0.22 y una variancia de 0.001. Un segundo laboratorio repite el mismo proceso con un polvo idéntico y obtiene cinco mediciones independientes de la porosidad con una media de 0.17 y una variancia de 0.002. Estimar la diferencia real entre las medias poblacionales de las mediciones de estos dos laboratorios con un coeficiente de confianza del 99% e interpretar el intervalo. Suponga que las poblaciones se distribuyen normalmente y con variancias iguales.

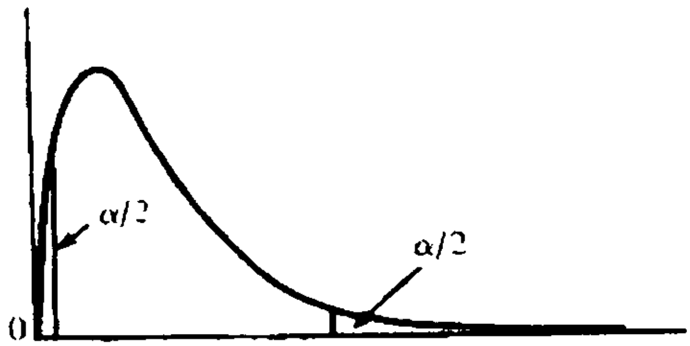
-Se registraron las áreas en hectáreas que ocupan los caimanes según las estaciones en cierto lago. Cinco caimanes registrados en la primavera mostraron áreas de 8.0, 12.1, 8.1, 18.2 y 31.7. Cuatro caimanes diferentes registrados en el verano mostraron áreas de 102.0, 81.7, 54.7 y 50.7. Estimar la diferencia en los promedios de las áreas en la primavera y en el verano por medio de un intervalo de confianza del 95%, suponiendo que se tienen poblaciones normales con variancias iguales, e interpretar el resultado.

c) Intervalo de confianza para la variancia

Para este subtema se planteó que ahora se realizaría la estimación para un solo parámetro: la variancia, y que esta vez las cuestiones de probabilidad se sustentarían sobre una distribución llamada Ji-cuadrada y simbolizada por χ^2, cuya tabla se dio al grupo. También se planteó que, al igual que en la estimación para la media y la diferencia de medias con muestras chicas, aquí se requerirá nuevamente del supuesto de normalidad en la población de la que se tome la muestra. Y finalmente se dijo que para la estimación de la variancia no hay diferencia si se tiene una muestra grande o chica, y se estableció el siguiente cuadro en el que se tienen las fórmulas para los límites del intervalo. En cuanto al manejo de la tabla, se dijo que, del mismo modo en que se hizo en el subtema anterior, éste se vería dentro de la resolución del ejemplo pero que de nueva cuenta se vería el elemento de grados de libertad.

PARAMETRO	ESTIMADOR	INTERVALO	GRADOS DE LIBERTAD (g. l.)
σ^2	s^2	$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}\right]$	v = n - 1

La razón de que las áreas para la búsqueda en tablas se manejen diferente a los dos casos anteriores se explicó en términos de que la Normal Estandarizada y la t de Student son distribuciones simétricas con respecto a 0, pero que este no es el caso para la χ^2, que es una distribución asimétrica y para valores sólo positivos. El siguiente es un boceto de gráfica de la χ^2, en donde se manejan los valores sobre la horizontal basándose en el área de la cola (derecha):



Como es usual, se presentó un problema como ejemplo. en el que se explicaron el cálculo del intervalo, el manejo de la tabla y la interpretación.

-Ejemplo.- Un experimentador quiere verificar la variabilidad de un equipo diseñado para medir el volumen de una fuente de audiofrecuencia. Tres mediciones independientes registradas con este equipo fueron 4.1, 5.2 y 10.2. Estime la variancia con un coeficiente de confianza del 90% suponiendo una población normalmente distribuida e interprete el intervalo.

Para concluir, la serie de ejercicios correspondientes al subtema fue:

-Se selecciona una muestra aleatoria de 21 ingenieros de un grupo que labora para un fabricante de equipo electrónico. La desviación estándar de la muestra de las horas de trabajo por semana fue de 7 horas. Determinar un intervalo de confianza del 90% para la variancia de la población de las horas de trabajo por semana para todos los ingenieros que laboran para el fabricante, suponiendo que estas mediciones tienen una distribución normal. Interpretar el intervalo.

-En el trabajo de un laboratorio es deseable verificar cuidadosamente la variabilidad de las lecturas obtenidas en muestras estándar. En un estudio de la concentración de calcio en agua potable, como parte de la valoración de la calidad del agua, se pasó el mismo patrón de medida 6 veces por el laboratorio en intervalos aleatorios. Las lecturas, en partes por millón, fueron: 9.54, 9.61, 9.32, 9.48, 9.70 y 9.26. Estimar la variancia de la población para las lecturas obtenidas del patrón con un intervalo de confianza del 98%, suponiendo que la población de las mediciones se distribuye normalmente. Interpretar.

-Las edades de 5 profesores universitarios en una muestra aleatoria son 39, 54, 61, 72 y 52 años. Obtener un intervalo de confianza para la variancia poblacional de las edades de todos los profesores de la universidad, suponiendo que estas se distribuyen normalmente e interpretar el resultado.

-Un instrumento de precisión tiene como garantía el leer con un error máximo de dos unidades. Una muestra de cuatro lecturas del mismo objeto dio como mediciones 353, 351, 351 y 355. Calcular el intervalo de confianza del 95% para la variancia de la población, suponiendo que esta se distribuye normalmente. ¿Es adecuada la garantía?

d) Intervalo de confianza para el cociente de dos variancias

Para el último modelo de estimación por intervalos se dijo que la idea principal es la de comparar las variancias de dos poblaciones diferentes, y que para esto se recurre a la estimación del cociente de éstas, expresado como $\frac{\sigma_1^2}{\sigma_2^2}$, y en donde la comparación –a diferencia de lo visto en las diferencias de medias y de proporciones, en la que esta se hace contra cero- se realiza contra 1 de la siguiente forma: si el cociente es mayor a uno entonces la variancia del numerador es mayor a la del denominador, si el cociente es menor a uno entonces la variancia del denominador es mayor a la del numerador, y si el cociente es igual a 1 entonces ambas variancias serán iguales. Así, para efectos de interpretación de intervalos, si los dos límites son mayores a uno se tendrá invariablemente el primero de los tres casos, si los dos límites son menores a uno se tendrá invariablemente el segundo caso, y si el límite inferior es menor a uno y el superior es mayor entonces podría suceder cualquiera de los tres casos por lo que no se tendría información suficiente para inferir una diferencia.

También se dijo que para este modelo de estimación se trabajaría con las tablas de la Distribución F de Snedecor, tablas que fueron puestas a disposición del grupo con la aclaración de que su manejo sería mostrado sobre el ejemplo con el antecedente de que nuevamente se manejaría el elemento grados de libertad, pero esta vez requiriendo de dos valores diferentes para dicho elemento en lugar de uno, como había sido en los dos modelos anteriores. Las tablas mencionadas son las marcadas como Puntos porcentuales de las distribuciones F que se anexan al final de este trabajo, y que igualmente se tomaron del libro Estadística Matemática con Aplicaciones de William Mendenhall.

Para efectos de realizar la estimación se presentó el siguiente cuadro en el que se resume el proceso de estimación, haciendo la aclaración que, al igual que en los dos modelos anteriores, también en este se requiere del supuesto de normalidad en las poblaciones para poder llevarlo a cabo:

PARAMETRO	ESTIMADOR	INTERVALO	GRADOS DE LIBERTAD (g. l.)
$\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{s_1^2}{s_2^2}$	$\left[\frac{s_1^2}{s_2^2 F^{v_1}_{v_2\,\alpha_2}}, \frac{s_1^2 F^{v_2}_{v_1\,\alpha_2}}{s_2^2}\right]$	$v_1 = n_1 - 1$ $v_2 = n_2 - 1$

Antes de pasar al ejemplo y con el fin de hacer más claro el proceso de enseñanza de manejo de tablas, se explicó la notación de los valores F estableciendo el modelo genérico como $F^{a}_{b\,\alpha/2}$, en donde $\alpha/2$ sigue calculándose a partir del coeficiente de confianza $1 - \alpha$, a se denomina como grados de libertad del numerador y b como grados de libertad del denominador. Así mismo se hizo notar que en cada una de las fórmulas para cada límite de intervalo no sólo cambiaban los valores de F de denominador a numerador, sino que también se invierten los grados de libertad, esto como una manera de salvar la asimetría característica de esta distribución, que en forma es muy semejante a la distribución χ^2.

El ejemplo de trabajo fue tomado de uno de los ejercicios vistos en la estimación con la Distribución t de Student e igualmente permitió la explicación del cálculo, la interpretación y el manejo de las tablas.
-Ejemplo.- En el experimento en el cual se comparó la cantidad de CL50 para DDT y para diazinón en el organismo de ciertos peces, para DDT se obtuvieron 12 muestras con una variancia de 37.8, y para diazinón en una muestra de 3 elementos se obtuvo una variancia de 8.9. Estimar el cociente de las dos variancias poblacionales por medio de un intervalo de confianza del 90% e interpretar. Suponga que las poblaciones se distribuyen normalmente.
Y terminado el ejemplo y aclaradas las dudas respecto al manejo de las tablas (esto individualmente) y a la interpretación, se dieron al grupo los siguientes problemas:

-Se tienen dos muestras, una de tamaño 15 y otra de tamaño 12, con desviaciones estándar de 3.07 y 0.8 respectivamente. Calcular un intervalo de confianza del 98% para el cociente de las dos variancias poblacionales e interpretarlo. Suponga poblaciones normales.

-Determine un intervalo de confianza del 95% para el cociente de las variancias en el rendimiento de gasolina para vehículos VW y Toyota, si 12 VW tuvieron una desviación estándar del rendimiento de 1 km/lt, y 10 Toyota tuvieron una desviación estándar del rendimiento de 1.8 km/lt. Suponga poblaciones normales e interprete el resultado.

-Una compañía transportista desea comprar neumáticos y contempla dos opciones. Realizó un experimento con 13 neumáticos de cada marca rodándolos hasta el desgaste. La primera marca tuvo una desviación estándar de 5000 km y la segunda de 6100 km. Calcule un intervalo de confianza para el cociente de las variancias del 90% e interprételo. Suponga que las poblaciones se distribuyen normalmente.

-Se tomó una muestra de cinco películas de una compañía y se obtuvo una variancia de 61.04 para la duración. En una muestra de siete películas de otra compañía la variancia de la duración fue de 887.92. Calcular un intervalo de confianza del 99% para el cociente de las variancias suponiendo que la duración de las películas se distribuye normalmente. Interpretar el resultado.

Con esto se finalizó con la unidad correspondiente a Intervalos de Confianza, estableciéndose como la dificultad más grande en los alumnos la interpretación de los resultados. Incluso las quejas y las frustraciones al respecto fueron manifestadas de manera abierta y por casi la totalidad de los alumnos. Se explicó que todo el problema se reduce a explicar con palabras el resultado numérico dentro del contexto que el mismo problema plantea. A la larga y con la experiencia adquirida el problema tiende a reducirse, mientras que a lo largo de la unidad se advirtió que en el examen correspondiente la interpretación tendría un valor del 50% del total de cada problema, por lo que la presión consiguió de los alumnos un poco más del esfuerzo requerido. Finalmente, en dicho examen efectivamente se le dio un valor del 50% del problema a la interpretación, 50% que se obtuvo con el hecho de *intentar* la interpretación (salvo en los casos de que ésta fuera muy irreal o sobre un parámetro distinto al estimado), lo que a la larga generaría en los alumnos la confianza para realizar futuras inferencias en los temas de Estadística Inferencial del resto del curso.

Unidad IV, Prueba de hipótesis

Objetivos:
Que el alumno conozca el concepto de Prueba de Hipótesis y de los elementos que la componen.
Que el alumno aplique los modelos de prueba de hipótesis que utilizan las distribuciones muestrales Normal, t de Student Ji-Cuadrada y F de Snedecor para resolver problemas de prueba de hipótesis de los parámetros media, proporción, diferencia de medias, diferencia de proporciones, variancia y cociente de variancias con el modelo correspondiente e interpretando las soluciones.

Como estrategia didáctica se discutirá en el aula la necesidad de probar estadísticamente supuestos que pudieran hacerse con respecto al comportamiento de algún parámetro, para ir deduciendo, por medio de lluvia de ideas, los elementos de una prueba de hipótesis, con la conciencia de que el profesor deberá redondear y en ocasiones presentar directamente y de forma clara, estos elementos. Posteriormente se planteará el modelo de prueba de hipótesis con la distribución Normal Estandarizada y el profesor resolverá un ejemplo de su aplicación, para finalizar con la resolución de ejercicios en el aula por parte de los alumnos. Posteriormente, y para cada uno de los tres modelos restantes, se presentará el modelo y se resolverá un ejemplo por parte del profesor para posteriormente dejar la solución de ejercicios a los alumnos dentro del aula.

4. Pruebas de hipótesis

Esta unidad se presentó como una nueva alternativa de la Estadística Inferencial. al ser un elemento que permite la validación de estimaciones puntuales. La mecánica consiste en hacer suposiciones con respecto a algún parámetro poblacional a partir de la información obtenida a partir del correspondiente estimador. Ahora bien, aunque las suposiciones parezcan aceptables en un principio, éstas deberán de validarse con métodos más estrictos que la simple observación de un estimador; y estos métodos corresponden precisamente a la Prueba de Hipótesis.

En primera instancia se tiene una suposición inicial realizada a partir de los resultados muestrales y que es la que se busca demostrar como válida. En segundo lugar, y bajo la premisa de que existe la posibilidad de que la suposición inicial en realidad esté equivocada, se tendrá que plantear aquella opción que quedará como válida si se concluye que el supuesto inicial es errado. Estos elementos son conocidos como *Hipótesis Nula* e *Hipótesis Alternativa*, y se simbolizan como H_0 y H_a respectivamente. En tercer lugar tenemos el manejo probabilístico de la prueba, dado que los resultados muestrales son aleatorios y dado el hecho de que la suposición inicial sea válida o no es aleatorio también, y dicho manejo probabilístico dependerá de la distribución de probabilidad a la que obedezcan la población y la muestra. En ese sentido partiremos de lo más elemental: o se asume que la población es Normal o se toma una muestra grande, a sabiendas que una muestra grande se distribuye normalmente, siendo el segundo punto con el que se trabajará inicialmente. Así, al trabajar con una muestra Normal se asume que los estimadores con los que se trabaje podrán ser estandarizados a valores z a través de la forma genérica de la Distribución Normal Estándar; y en dónde la expresión específica de estandarización se denominará en adelante como *Estadístico de Prueba*, siendo éste el tercer elemento dentro de una Prueba de Hipótesis. Para el cuarto elemento debemos establecer previamente una cuestión más de probabilidad que tiene que ver con las opciones de plantear una hipótesis que puede ser verdadera o falsa (eso no lo sabemos) con las acciones de aceptarla o rechazarla como válida, con la consecuencia de que algunas de las combinaciones de estado natural con la decisión de aceptar o rechazar la validez de la suposición sobre la base de las observaciones en un experimento derivarán en una acción acertada, mientras que otras devendrán en un error. Así si planteamos una hipótesis y basándonos en el experimento la aceptamos cuando en realidad es válida, o si planteamos la contraparte como hipótesis y sobre la base del experimento aceptamos esa contraparte cuando en realidad es válida, estaremos tomando una decisión acertada, mientras que si planteamos una hipótesis y con la base del experimento la aceptamos cuando en realidad es falsa, o si planteamos la contraparte como hipótesis y sobre la base del experimento la aceptamos cuando en realidad es verdadera la otra hipótesis, estaremos tomando una decisión errónea. Estas opciones se presentan en la siguiente tabla, en las que catalogaremos los errores como Tipo I y Tipo II:

	VERDADERO	FALSO
ACEPTAR	Acierto.	Error Tipo II.
RECHAZAR	Error Tipo I.	Acierto.

En términos un poco más formales definiremos estos dos errores de la siguiente manera:
- Error Tipo I.- Es el que se comete al rechazar la hipótesis nula siendo que es verdadera.
- Error Tipo II.- Es el que se comete al aceptar la hipótesis nula siendo que es falsa.

Ahora, el que se cometa cualquiera de estos errores es una cuestión meramente aleatoria, así que denotaremos la probabilidad de cometer el Error Tipo I como α y la probabilidad de cometer el Error Tipo II como β. Y en adelante trabajaremos dentro de los planteamientos probabilísticos con el riesgo de rechazar H_0 siendo que es verdadera, es decir con α, valor al que denominaremos *Nivel de Significación.*

Con este valor podemos establecer el cuarto elemento de una prueba de hipótesis. Supongamos que tenemos una hipótesis nula que es efectivamente válida, pero si obtenemos ciertas observaciones en el experimento corremos el riesgo de rechazar erróneamente la hipótesis de legalidad. Es decir, para ciertos valores de nuestro estadístico se pensará en rechazar H_0. Estos valores del estadístico de prueba para los cuales erróneamente rechazaríamos una hipótesis nula que es en realidad verdadera se le denominan como Región de Rechazo y se denotan como R. R.

Finalmente, se establece que la región de rechazo dependerá del valor que queramos asignarle a α. Y así, si en principio se trabaja en términos de una Distribución Normal Estándar, entonces no sólo se estandarizarán los valores muestrales, sino también se establecerán como región de rechazo valores z dependientes de α. Específicamente, podemos establecer la región de rechazo de tres formas: para valores menores a nuestro parámetro, lo que implica un área hacia el lado izquierdo de la distribución Normal; para valores mayores a nuestro parámetro, lo que implica un área hacia el lado derecho de la distribución Normal; y para valores simplemente diferentes a nuestro parámetro, lo que implica un área hacia ambos lados de la distribución Normal. Dichas alternativas de R. R. se les conocen respectivamente como *región de rechazo de cola izquierda*, *región de rechazo de cola derecha* y *región de rechazo de dos colas*.

Con esto hemos establecido los cuatro elementos de una Prueba de Hipótesis como Hipótesis Nula (H_0), Hipótesis Alternativa (H_a), Estadístico de Prueba y Región de Rechazo (R. R.), además de definir el Nivel de Significación (α), elementos que se utilizarán en todas las pruebas que se presentarán en esta unidad. Y en ese tenor, se dijo que nuevamente se trabajaría sobre los mismos parámetros y bajo las mismas condiciones vistos en Intervalos de Confianza; es decir, se realizarían pruebas de hipótesis sobre media, proporción, diferencia de medias y diferencia de proporciones a partir de muestras grandes, sobre media y diferencia de medias a partir de muestras chicas, sobre la variancia, y sobre el cociente de dos variancias, aplicando las correspondientes distribuciones de probabilidad Normal, t de Student, Ji-cuadrada y F de Snedecor.

a) Prueba de Hipótesis para la media, la proporción, la diferencia de medias y la diferencia de proporciones con muestras grandes ($n \geq 30$)

Este primer modelo será manejado a través de la Distribución Normal, y de nueva cuenta requeriremos de tener disponible el cuadro de estimadores puntuales para los parámetros de interés, así como recordar que s^2 funciona como estimador puntual de σ^2.

Cuadro de Estimadores Puntuales

PARAMETRO θ	ESTIMADOR $\hat{\theta}$	TAMAÑO DE LAS MUESTRAS	DESVIACION ESTANDAR DEL ESTIMADOR $\sigma_{\hat{\theta}}$
μ	$\bar{y}$	n	$\sqrt{\frac{\sigma^2}{n}}$
p	$\hat{p}$	n	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	n_1 y n_2	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	n_1 y n_2	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Ahora, para efectos de presentar de manera más directa el procedimiento de la Prueba de Hipótesis para los cuatro parámetros listados arriba y bajo la condición de tamaño de muestra $n \geq 30$, se dio al grupo el siguiente cuadro en el que se establecen los cuatro elementos de la Prueba de Hipótesis correspondientes al caso:

H_0	H_a	Estadístico de Prueba	R. R.
$\theta \leq \theta_0$	$\theta > \theta_0$		$z > z_\alpha$
$\theta \geq \theta_0$	$\theta < \theta_0$	$z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$	$z < -z_\alpha$
$\theta = \theta_0$	$\theta \neq \theta_0$		$\lvert z \rvert > z_{\alpha/2}$

en donde θ_0 corresponde a una constante contra la cual comparamos nuestro parámetro y que sustituirá a $\hat{\theta}$ en el Estadístico de Prueba, y donde z_α y $z_{\alpha/2}$ corresponden a los valores z a partir de los cuales comienza el área bajo la curva de valor α o $\alpha/2$ y en donde se sitúa R. R. En este último punto se dijo que las tres regiones de rechazo listadas y que dependen de la manera en la que se establezca H_0 se denominan respectivamente como Región de Cola Derecha (por estar situada al lado derecho de la curva), Región de Cola Izquierda (por estar situada la lado izquierdo de la curva) y Región de Dos Colas (por estar situada a ambos lados de la curva, lo que además provoca que α sea repartida equitativamente hacia ambos lados de la distribución); e igualmente se explicó que en esta última región, la desigualdad con la operación valor absoluto implica el comparar simultáneamente ambas colas de una sola vez. El resto de los elementos son familiares desde la unidad anterior.

Como aclaración final se hizo notar que todas las hipótesis nulas ofrecen la opción de igualdad contra la constante mientras que las alternativas sólo se plantean como desigualdades estrictas, y que esta característica debería de asumirse siempre; así, al extraer de algún problema las desigualdades para el planteamiento de las hipótesis, siempre deberá de tomarse para H_0 aquella opción que presente la posibilidad de la igualdad del parámetro con la constante, lo que permitirá además establecer R. R. de manera directa con la ayuda del cuadro.

Para ilustrar el procedimiento, se presento y resolvió un ejemplo ante el grupo, tras el cual, y para terminar con las explicaciones, se dijo que ya que se estaba trabajando solo con la probabilidad de cometer el error tipo I y no con la de cometer el error tipo II todas las inferencias se harían en términos de H_0 y nunca en términos de H_a.

-Ejemplo.- Se tiene que reparar una máquina en cierta fábrica si produce más del 10% de artículos defectuosos en la producción de un día. Una muestra aleatoria de 100 artículos de la producción diaria contiene 15 defectuosos y el capataz decide que debe repararse la máquina. ¿La evidencia de la muestra apoya su decisión? Utilice un nivel de significación del 1%.

Por último de dieron a los alumnos los siguientes problemas para su solución:

-El vicepresidente de ventas de una corporación afirma que los vendedores tienen un promedio no mayor de 15 prospectos de venta por semana. Se seleccionan 36 vendedores al azar para verificar su afirmación, y se registra el número de contactos en una sola semana seleccionada en forma aleatoria, La muestra tiene una media de 17 prospectos y una variancia de 9. ¿Contradicen los hechos la afirmación del vicepresidente? Utilice un nivel de significación del 5%.

-Se realizó un estudio psicológico para comparar los tiempos de reacción de hombres y de mujeres con respecto a cierto estímulo. Se utilizaron en el experimento muestras aleatorias independientes de 50 hombres y 50 mujeres. Los resultados se muestran en la tabla. ¿Presentan los datos suficiente evidencia para sugerir una diferencia entre los promedios verdaderos de los tiempos de reacción para hombres y mujeres? Utilice un nivel de significación del 5%.

Hombres	Mujeres
$n_1 = 50$	$n_2 = 50$
$\bar{y}_1 = 3.6$ seg.	$\bar{y}_2 = 3.8$ seg.
$s_1^2 = 0.18$	$s_2^2 = 0.14$

-Los salarios diarios en una industria en particular presentan una Distribución Normal, con una media de \$13.20 y una desviación de \$2.50. Si en esta industria una compañía que emplea a 40 trabajadores les paga en promedio \$12.20, ¿puede acusarse a esta compañía de pagar salarios inferiores? Utilice un nivel de significación del 1%.

-El índice de dureza Rockwell para acero se determina al rayar el acero con un punzón de diamante y medir la profundidad de la penetración. 50 muestras de cierto tipo de acero sometidas al índice de dureza Rockwell tuvieron una media de 62 con una desviación estándar de 8. El productor afirma que este acero tiene un índice de dureza promedio de por lo menos 64. ¿Hay suficiente evidencia para refutar la afirmación del productor a un nivel de significación del 1%?

-Mediciones respecto del esfuerzo cortante obtenidas a partir de pruebas de compresión independientes para dos tipos de suelos dieron los resultados de la tabla (mediciones en tn/m^2). ¿Difieren los dos suelos con respecto al esfuerzo cortante promedio a un nivel de significación del 1%?

Suelo I	Suelo II
$n_1 = 30$	$n_2 = 35$
$\bar{y}_1 = 1.65$	$\bar{y}_2 = 1.43$
$s_1^2 = 0.26$	$s_2^2 = 0.22$

-Un productor afirma que al menos el 20% del público prefiere su producto. Se toma una muestra de 100 personas para verificar su afirmación, y se encuentra que 25 de ellas prefieren el producto. ¿Es válida la afirmación del productor a un nivel de significación del 5%?

-En una encuesta llevada a cabo en los Estados Unidos se entrevistó a 871 adultos. Con respecto a una pregunta, el 53% de los entrevistados estuvieron a favor de un apoyo militar para Israel. ¿Concluiría que la mayoría de los adultos en Estados Unidos está a favor del apoyo militar a Israel? Utilice un nivel de significación del 5%.

-Un investigador político afirma que la fracción de republicanos es mayor que la fracción de demócratas a favor de la pena de muerte. Obtuvo muestras aleatorias independientes de 200 republicanos y 200 demócratas, y encontró que 46 republicanos y 34 demócratas estuvieron a favor de la pena de muerte. ¿Fundamenta esta evidencia un apoyo estadístico para la opinión del investigador? Utilice un nivel de significación del 5%.

-En un ejercicio se afirma que de una muestra al azar de 500 mediciones acerca del tiempo de hospitalización en clínicas se obtuvieron una media muestral de 5.4 días y una desviación estándar muestral de 3.1 días. Una agencia de reglamentación supone que el promedio del tiempo de hospitalización es mayor que 5 días. ¿Apoyan los datos esta hipótesis a un nivel de significación del 5%?

b) Prueba de hipótesis para la media y la diferencia de medias con muestras chicas ($n < 30$)

Con esta prueba se dijo que de nuevo se validarían supuestos referentes a la media de una población y a la diferencia de medias de dos poblaciones diferentes, y que de nueva cuenta, al igual que en el tema de Intervalos de Confianza, dependeríamos de estimadores obtenidos de muestras chicas, por lo que se trabajaría de nueva cuenta con la Distribución t de Student. Para los efectos de seguir el proceso de la Prueba de Hipótesis se dio al grupo el siguiente cuadro en el que se resumen los elementos de la prueba involucrados en el proceso, aclarando que a diferencia de la prueba anterior (Normal) aquí no se establecería un modelo genérico sino el específico para cada uno de los dos parámetros a validar:

H_0	H_a	Estadístico de prueba	R. R.	Grados de libertad (g. l.)
$\mu \leq \theta_0$	$\mu > \theta_0$		$t > t_\alpha$	
$\mu \geq \theta_0$	$\mu < \theta_0$	$t = \dfrac{\bar{y} - \mu}{s / \sqrt{n}}$	$t < -t_\alpha$	$v = n - 1$
$\mu = \theta_0$	$\mu \neq \theta_0$		$\lvert t \rvert > t_{\alpha/2}$	
-------------	-------------	---------------------	---------	--------------
$\mu_1 - \mu_2 \leq \theta_0$	$\mu_1 - \mu_2 > \theta_0$		$t > t_\alpha$	
$\mu_1 - \mu_2 \geq \theta_0$	$\mu_1 - \mu_2 < \theta_0$	$t = \dfrac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ donde $S = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$	$t < -t_\alpha$	$v = n_1 + n_2 - 2$
$\mu_1 - \mu_2 = \theta_0$			$\lvert t \rvert > t_{\alpha/2}$	

Asimismo, se recordó que para poder utilizar la prueba t de Student se requiere del supuesto de población Normal en el caso de la prueba para una media, y del supuesto de poblaciones normales con variancias iguales en el caso de la prueba para la diferencia de medias.

Para redondear la explicación referente al subtema, se presentó al grupo un ejemplo tomado de uno de los ejercicios resueltos por los alumnos en el subtema de Intervalos de confianza para la media y la diferencia de medias con muestras chicas.

-Ejemplo.- En el ejercicio del fabricante de pólvora, la media muestral y la desviación estándar de las velocidades iniciales para 8 granadas que se probaron con un nuevo tipo de pólvora fueron 2959 m/s y 39.1 m/s respectivamente. El productor afirma que la nueva pólvora produce una velocidad promedio no menor a 3000 m/s. ¿Aportan los datos de la muestra suficiente evidencia para contradecir la afirmación del productor a un nivel de significación del 2.5%? Supóngase que las velocidades iniciales se distribuyen de manera Normal.

Para finalizar el subtema, los alumnos resolvieron los siguientes ejercicios:

-Una máquina expendedora de refrescos se diseñó para servir en promedio 7 onzas de refresco por vaso. Con el objeto de verificar lo anterior se eligieron 10 vasos llenos de bebida y se midieron los contenidos. La media y la desviación estándar de las 10 mediciones fueron 7.1 onzas y 0.12 onzas respectivamente. ¿Presentan estos datos suficiente evidencia para indicar que la descarga media difiere de 7 onzas a un nivel de significación del 10%? Suponga que la población de las mediciones es Normal.

-Una encuesta reveló que los sueldos del personal de investigación es en promedio de \$13,221.-. Si una muestra aleatoria de sueldos de 20 físicos tuvo una media de \$33,120.- y una desviación estándar de \$2,140, ¿concluiría que se paga más en promedio a los físicos con respecto a la media global? Utilice un nivel de significación del 5%, asumiendo que los sueldos se distribuyen normalmente.

-Se aplicaron dos nuevos métodos para enseñar a leer a dos grupos de niños y se hizo una comparación basada en una prueba de comprensión de lectura al final del período de enseñanza. En la tabla se ofrecen los resultados de las pruebas. ¿Presentan los datos suficiente evidencia que indique una diferencia en los resultados promedio para las poblaciones asociadas a los dos métodos de enseñanza a un nivel de significación del 5%? Suponga que ambas poblaciones son normales y con variancias iguales.

Método I	Método II
$n_1 = 11$	$n_2 = 14$
$\bar{y}_1 = 64$	$\bar{y}_2 = 69$
$s_1^2 = 52$	$s_2^2 = 71$

-Se efectuó un estudio para estimar las cantidades de residuos químicos encontrados en los tejidos cerebrales de pelícanos cafés. En una prueba sobre DDT, muestras aleatorias de pelícanos jóvenes y polluelos dieron los resultados indicados en la tabla (las mediciones están dadas en partes por millón): Pruebe la hipótesis de que no existe diferencia en las cantidades promedio de DDT encontradas en pelícanos jóvenes y polluelos a un nivel de significación del 5%. Supóngase que ambas poblaciones son normales y de variancias iguales.

Jóvenes.	Polluelos
$n_1 = 10$	$n_2 = 13$
$\bar{y}_1 = 0.041$	$\bar{y}_2 = 0.026$
$s_1 = 0.017$	$s_2 = 0.006$

-En referencia con el ejercicio anterior, ¿existe evidencia de que el promedio en pelícanos jóvenes es mayor que en los polluelos en más de 0.01 partes por millón al mismo nivel de significación y bajo los mismos supuestos?

c) Prueba de hipótesis para la variancia

Se estableció que esta prueba tiene la finalidad de validar supuestos sobre el parámetro variancia, y que en el fondo seguiríamos trabajando sobre la misma idea genérica de la Prueba de Hipótesis, en este caso para el parámetro mencionado. Y, tal y como se dio en Intervalos de Confianza, para la prueba sobre variancia de nueva cuenta recurrimos a la Distribución Ji-cuadrada como base probabilística, sustentada esta siempre con el supuesto de normalidad en la población asociada a cada problema, sin olvidar también que para el trabajo con este parámetro ya no hay ninguna diferencia al realizar la prueba si se tiene una muestra grande o chica.

El cuadro de resumen de los elementos de la prueba es el siguiente:

H_0	H_a	Estadístico de prueba	R. R.	Grados de libertad (g. l.)
$\sigma^2 \leq \theta_0$	$\sigma^2 > \theta_0$		$\chi^2 > \chi^2_{\alpha}$	
$\sigma^2 \geq \theta_0$	$\sigma^2 < \theta_0$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$	$\chi^2 < \chi^2_{1-\alpha}$	v = n - 1
$\sigma^2 = \theta_0$	$\sigma^2 \neq \theta_0$		$\chi^2 > \chi^2_{\alpha_2}$ ó $\chi^2 < \chi^2_{1-\alpha_2}$	

Como se puede ver, independientemente del parámetro, el modo de plantear las hipótesis nula y alternativa es idéntico que para las dos pruebas anteriores, por lo que se puede inferir del igualmente que se tienen las tres mismas regiones de rechazo para cada planteamiento: de cola derecha, de cola izquierda y de dos colas, siguiendo el orden en el cuadro; y donde también sabemos que en realidad la prueba Ji-cuadrada, dada su asimetría, debe manejarse de manera especial al trabajar con la cola izquierda (que en el sentido estricto no sería una cola, ya que la distribución en realidad presenta sólo una cola, y esta está situada hacia loa derecha), lo que justifica las diferencias con respecto a uno involucradas para el nivel de significación.

Para mostrar el proceso de esta prueba se realizó el siguiente ejemplo ante el grupo:

-Ejemplo.- Se afirma que una de las piezas de un motor producida por una compañía tiene una variancia del diámetro no mayor que 0.0002. Una muestra aleatoria de 10 partes reveló una variancia muestral de 0.0003. ¿Apoya esto la afirmación referente a la pieza a un nivel de significación del 5%? Suponga que los diámetros están distribuidos normalmente.

Para finalizar el subtema se dieron al grupo los siguientes problemas para su solución:

-Un investigador está convencido de que su equipo de medición tiene una variabilidad que se traduce en una desviación estándar de 2. 16 mediciones tuvieron como resultado una variancia de 6.1. ¿Están los datos en desacuerdo con su afirmación a un nivel de significación del 5%? Suponga que la población de las mediciones obedecen a una Distribución Normal.

-Un fabricante de máquinas para empacar detergentes afirma que su máquina podría llenar con un peso dado las cajas. La media y la variancia de una muestra de 8 cajas de 3 libras fueron iguales a 3.1 lb y 0.018 respectivamente. Pruebe la hipótesis de que la variancia de la población de las mediciones de los pesos es igual a 0.01 con un nivel de significación del 5% y suponiendo que dichos pesos están distribuidos normalmente.

-Se garantiza que un instrumento de medición es exacto con un margen de dos unidades. Una muestra de cuatro lecturas del mismo objeto dio como mediciones 353, 351, 351 y

355. Pruebe la hipótesis de que la desviación estándar no es mayor a 2 a un nivel de significación del 1%. Suponga que la población de las mediciones realizadas con el instrumento son Normales.

-Las pruebas de aptitud deberían de producir resultados con una gran cantidad de variación de tal manera que un administrador pueda diferenciar entre personas con una aptitud baja y personas con una aptitud alta. La prueba estándar que utiliza cierta industria ha estado produciendo resultados con una desviación estándar de 10 puntos. Se aplica una nueva prueba a 20 aspirantes a empleados de esa industria y se obtiene una desviación estándar muestral de 12 puntos. ¿Son os resultados de la nueva prueba significativamente más variables que los resultados de la prueba estándar? Utilice un nivel de significación del 1% y suponga una población de resultados distribuida de manera Normal.

-Un fabricante de fusibles asegura que la duración de sus fusibles tiene una desviación estándar de 0.9 años. Si una muestra aleatoria de 10 de estos fusibles tiene una desviación estándar de 1.2 años, ¿podría afirmarse que la desviación estándar es mayor que 0.9 años a un nivel de significación del 5%? Supóngase que la vida útil de los fusibles obedece a una Distribución Normal.

d) Prueba de hipótesis para el cociente de dos variancias

Para terminar con el tema de Prueba de Hipótesis se presentó la opción de validar supuestos acerca de la relación entre las variancias de dos distintas poblaciones. Para ello se recordó que en la estimación por intervalos se había recurrido al cociente de las dos variancias como parámetro para verificar la relación de igualdad o desigualdad que pudiera existir entre ambas, por lo que podríamos acudir al mismo parámetro dentro de la Prueba de Hipótesis. Previamente al establecimiento de los elementos de la prueba se dijo que en esta ocasión el planteamiento sería mucho más accesible que en los tres casos anteriores, ya que se establecería sólo una manera de plantear la hipótesis nula, aunque ello implicaría manejar los valores muestrales bajo una cierta condición que se establecería más adelante. De esa manera, el cuadro correspondiente al subtema resultó ser el siguiente:

H_0	H_a	Estadístico de prueba	R. R.	Grados de libertad (g. l.)
$\sigma_1^2=\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	$F=\frac{s_1^2}{s_2^2}$	$F>F^{v_1}_{v_2\,\alpha}$	$v_1=n_1-1$ para el numerador $v_2=n_2-1$ para el denominador

y en el que se dijo que H_0 y H_a invariablemente se plantearían del modo establecido como único y que la región de rechazo curiosamente sería siempre de cola derecha y no de dos colas como podría sugerirlo en principio la hipótesis nula, con lo que el razonamiento del problema se reduce casi a sólo identificar que se busca probar una suposición que implique comparar dos variancias poblacionales. Pero también se dijo que para poder tener esta comodidad debe cuidarse que siempre que se vacíen los datos de las muestras invariablemente hay que elegir como la variancia del numerador a aquella que de las dos que se nos presenten sea la mayor, es decir, siempre debe darse que $s_1^2 > s_2^2$, lo que dará como consecuencia que n_1 sea la muestra de la que se obtenga la mayor variancia y que v_1, los grados de libertad para el numerador, se obtenga directamente de esa n_1, considerando este orden importante si se recuerda el manejo de la tabla de Puntos Porcentuales de las Distribuciones F que son con las que se trabajará en esta prueba. Asimismo se hizo notar que las hipótesis en esta prueba no tenían la estructura de las tres pruebas anteriores en las que se compara el parámetro con una constante, sino que en este caso se están comparando directamente dos parámetros de dos poblaciones. Como exigencia final se estableció que, al igual que en el tema de Intervalos de confianza, se requiere del supuesto de poblaciones Normales.

Para clarificar toda la teoría, se planteó y resolvió un problema como ejemplo.

-Ejemplo.- Supóngase que dos compañías producen una cierta refección y que se desea comparar la variación en los diámetros para las piezas de cada compañía. La variancia muestral de la primera compañía basada en 10 diámetros fue de 0.0003, mientras que la variancia de las mediciones de los diámetros para una muestra de 20 de las partes de la segunda compañía fue 0.0001. ¿Presentan los datos suficiente información para inferir una menor variancia en los diámetros de la segunda compañía? Realice la prueba con un nivel de significación del 5%, suponiendo que las poblaciones de los diámetros de las piezas de las dos compañías son Normales.

Una vez concluidas las explicaciones correspondientes al subtema, se dieron al grupo los siguientes problemas para su resolución:

-Un experimento para estudiar los umbrales del dolor producidos por descargas eléctricas en hombres y mujeres reveló los datos resumidos en la tabla. ¿Muestran los datos evidencia suficiente para indicar una evidencia significativa en la variabilidad en los umbrales del dolor para hombres y mujeres? Utilice un nivel de significación del 10%, suponiendo poblaciones Normales.

	Hombres	Mujeres
n	14	10
$\bar{y}$	16.2	14.9
s^2	12.7	26.4

-Se registraron los precios al cierre de las operaciones de dos acciones comunes durante un período de 16 días. Las medias y las variancias se muestran en tabla. ¿Presentan estos datos suficiente evidencia para indicar una diferencia en variabilidad para los precios al cierre de las operaciones de las dos acciones a un nivel de significación del 2%? Suponga que el comportamiento de los precios al cierre para los dos tipos de acciones están distribuidos normalmente.

-La tabla presenta los resultados de una prueba para detectar DDT en el tejido cerebral de pelícanos cafés. ¿Hay suficiente evidencia a un nivel de significación del 5% para concluir que la variancia en las mediciones de los niveles de DDT es mayor para los pelícanos jóvenes que para los polluelos? Suponga que ambas poblaciones son Normales.

Jóvenes	Polluelos
$n = 10$	$n = 13$
$\bar{y}_1 = 0.041$	$\bar{y}_2 = 0.026$
$s_1 = 0.017$	$s_2 = 0.006$

-En la tabla se muestran los resultados obtenidos al aplicar dos métodos para enseñar a leer a los niños. ¿Podría pensarse que las dos variancias poblacionales son iguales a un nivel de significación del 1%, suponiendo que los resultados para los dos métodos tienen un a Distribución Normal?

	Método I	Método II
n	11	14
$\bar{y}$	64	69
s^2	52	71

-Se aplicaron dos diferentes métodos de secado a dos muestras de concreto para verificar su resistencia. Los resultados se muestran en la tabla. ¿Parecería que con el primer método se obtiene una variabilidad mayor en la resistencia que con el segundo método? Utilice un nivel de significación del 5%, suponiendo que las poblaciones asociadas a las resistencias obtenidas por ambos métodos son Normales.

	Método I	Método II
n	7	10
$\bar{y}$	3250	3240
s	210	190

-En el problema de estimación de consumo de oxígeno para dos diferentes métodos de entrenamiento, cuyos resultados se muestran en la tabla, se supuso que las variancias poblacionales eran iguales. ¿Podría afirmarse que esto es cierto a un nivel de significación del 0.5%? Recuérdese que se asumió que las poblaciones asociadas a los dos métodos son Normales.

Entrenamiento continuo	Entrenamiento intermitente
$n = 9$	$n = 7$
$\bar{y} = 43.71$	$\bar{y} = 39.63$
$s = 5.88$	$s = 7.68$

-Dos laboratorios realizaron una prueba sobre la porosidad del cobre sintetizado y obtuvieron los resultados de la tabla. Al hacer una estimación de la diferencia en los promedios se supuso que las variancias poblacionales eran iguales. ¿Podría afirmarse tal cosa a un nivel de significación del 10%? Supóngase que las poblaciones de las mediciones son distribuidas normalmente.

	Laboratorio I	Laboratorio II
n	4	5
$\bar{y}$	0.22	0.17
s^2	0.001	0.002

Con esta serie de problemas se dio por finalizado el tema de Prueba de Hipótesis. En esta unidad en realidad el único problema significativo que presentaron los alumnos fue el de poder plantear la desigualdad que como supuesto para hipótesis presente un problema; es decir, aún existen dificultades dentro del lenguaje matemático en lo tocante a expresiones como "al menos", "no mayor", etcétera, aunque con el antecedente de la unidad correspondiente a las Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta estas dificultades ya no fueron tan recurrentes, y tendieron a reducirse a medida que se avanzó en la unidad, o equivalentemente, con práctica. Un problema mucho menos serio, dado que se presentó muy esporádicamente, y no por desconocimiento o falta de comprensión en los alumnos, fue el de confundir la variancia y la desviación estándar en algunos problemas, pero esto como resultado, en casi la totalidad de los casos, de falta de concentración.

Unidad V, Estimación por mínimos cuadrados

Objetivos:
Que el alumno establezca una relación lineal, parabólica, exponencial o logarítmica entre dos variables aleatorias aparejadas involucradas en un problema estadístico. cuando esto sea posible, con el fin de resolver problemas de estimación por medio de regresión lineal o regresión no lineal, según sea el caso.
Que el alumno conozca y utilice el método de mínimos cuadrados y los elementos que lo conforman (diagramas de dispersión, coeficiente de correlación) para predecir el comportamiento de variables aleatorias aparejadas.

Como estrategia didáctica se hablará en clase de la posibilidad de observar variables aparejadas dentro de diversos fenómenos y se ilustrarán algunos de ellos, para posteriormente crear un problema que implique tomar una muestra con variables aparejadas y que se comporten de forma más o menos lineal, muestra que será tomada de entre los mismos estudiantes. Posteriormente se presentarán la necesidad de la estimación del valor de una de las variables a partir de un valor dado de la otra con la consecuente necesidad de establecer una función que las relacione. Se le pedirá al grupo que grafique los pares obtenidos en la muestra para deducir que el mejor ajuste sería a una recta y se recordará la forma algebraica de la recta; se mostrarán al grupo pasos sencillos que permitan generar el sistema de ecuaciones para obtener la recta de mejor ajuste y se resolverá el ejemplo en clase con la participación tanto del profesor como de los alumnos. Posteriormente se hará alguna predicción para mostrar la utilidad del ajuste, y se terminará la explicación presentando el concepto de coeficiente de correlación, su fórmula, su utilidad, sus características y el cálculo del mismo para el ejemplo desarrollado. Posteriormente se resolverán ejercicios en el aula. Para los siguientes modelos se seguirá la estrategia de presentar datos aparejados que se ajusten a cualquiera de los modelos deseados y se pedirá a los alumnos que encuentren el sistema propio de cada uno de acuerdo a lo visto en el modelo lineal, para terminar con la solución de ejercicios en clase.

5. Estimación por mínimos cuadrados

En esta unidad se presentó a los alumnos un problema común: el asumir el comportamiento futuro de alguna situación aleatoria basándonos exclusivamente en su comportamiento presente y/o pasado. Y se dijo que de alguna manera este problema se había trabajado ya con las interpretaciones de los resultados de los intervalos de confianza y de las pruebas de hipótesis, pero que existen otros tipos de estructuras que podrían presentarse en la Estadística y en las que no se tendría la facilidad de recurrir a los métodos inferenciales que hasta ahora se habían manejado. Uno de estos casos es aquel en el que se tienen observaciones aparejadas, es decir dos series de resultados en los que puede decirse que una de la serie depende de alguna manera de la otra, como pueden ser fricción y calor, oferta y demanda, distancia recorrida hasta la escuela y gasto de transporte, estatura y peso, etcétera, y en donde, al involucrar dos variables, el problema tiende a ser el poder predecir el valor que se espera tome una de las variables bajo el supuesto de que la otra efectivamente tome un valor específico y previamente decidido por nosotros. Para hacer esto mas claro se retomó uno de los ejemplos recién mencionados, el de estaturas y pesos en metros y kilogramos respectivamente, y se tomaron datos con algunas de las personas en el grupo anotándose en una tabla sobre la cual se establecieron dos cuestiones: una que puede suponerse la existencia de una cierta relación entre ambas variables -estatura y peso-, y la otra que la presentación en tabulación de la estatura de cada persona aparejada con su respectivo peso tiene la forma de las tabulaciones con las que en cursos anteriores de Matemáticas se realizaban las gráficas de las funciones algebraicas, con lo que podría pensarse en la existencia de una función con la cual poder obtener, al menos de una manera aproximada, el paso de una persona a partir de conocer su estatura.

Ahora, para poder realizar esto podemos recurrir a dos alternativas: o desarrollamos una función nueva y específica para el problema, cuestión que de entrada se presenta demasiado compleja, o buscamos entre las funciones elementales ya existentes aquella a la que se asemeje más el comportamiento de la relación que nos presenta el problema. Obviamente puede verse que la segunda alternativa resulta ser más práctica, y para ello se recordó que previamente ya se había realizado algo así al momento de comenzar a trabajar con la Distribución Normal, al partir del Polígono de Frecuencias y de la función Normal como aquella curva que mejor lo representaba de manera genérica. En este caso la forma en que esto se realiza empieza por graficar en el plano cartesiano toda la serie de pares que el problema presenta como datos, para después identificar el tipo de función al que más se asemeja dicho gráfico, al cual llamaremos *Gráfico de Dispersión*; y una vez realizado lo anterior buscar entonces la función en específico a la que mejor se aproxima o ajusta el comportamiento de nuestros datos, cuestión que le da al proceso el nombre de *Ajuste de Curvas*. En particular en este tema se tocarán cuatro distintos modelos de ajuste: el lineal, de la forma $y = a + bx$; el cuadrático o parabólico, de la forma $y = a + bx + cx^2$; el exponencial, de la forma $y = ae^{bx}$; y el logarítmico, de la forma $y = a + b\ln(x)$. De esta manera, y como se puede ver, todo se reduce a buscar las constantes a, b y en su caso c para definir la función del modelo como única para cada problema; y así no sólo se tiene el comportamiento aparejado de nuestras variables expresado en forma de función, sino también la alternativa de realizar predicciones o estimaciones de alguna de las dos variables a partir de asignar algún valor de interés a la otra variable.

A lo largo de este tema se verán los procesos para hallar las constantes en cada uno de los modelos recién presentados.

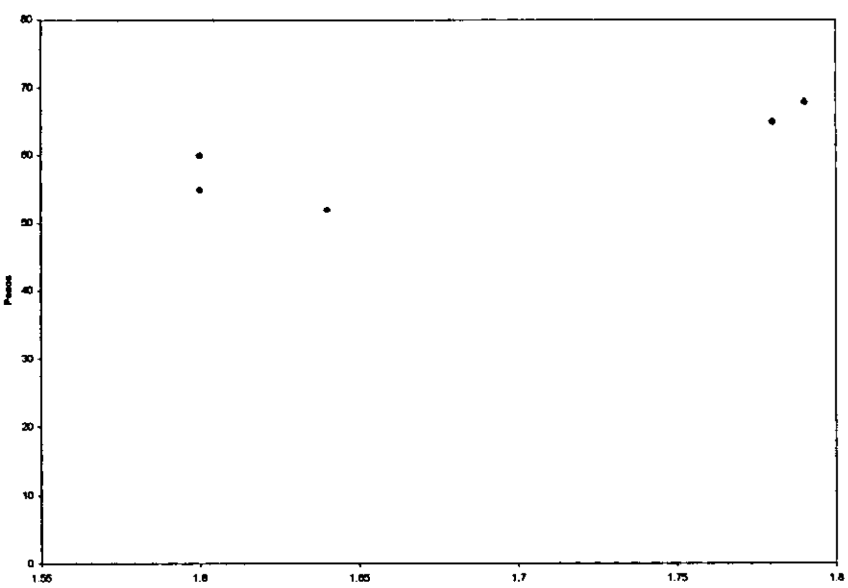
a) Recta de mínimos cuadrados

En este punto veremos el proceso para calcular a y b para el modelo lineal $y = a + bx$, y que igualmente podremos establecer para los otros tres modelos a manejar más adelante. Para ello retomamos el ejemplo de los datos de peso y estatura de cinco diferentes personas, y si el diagrama de dispersión indica un comportamiento más o menos lineal, entonces procederemos a establecer los pasos para hallar a y b, y una vez obtenidos, realizaremos la estimación del peso para una persona de 1.92 de estatura.

Los datos fueron

Estatura	Peso
1.79	68
1.60	55
1.60	60
1.78	65
1.64	52

y el correspondiente diagrama de dispersión fue



en el que podemos ver que el comportamiento efectivamente se asemeja a una línea recta, con las lógicas desviaciones de lo que podríamos llamar la "recta ideal", y que sería aquella que mejor contuviera a los cinco puntos. De esta forma efectivamente tenemos que buscar a y b para obtener dicha "recta ideal", y el proceso es simple: primero tomamos el modelo hacia el cual ajustaremos los datos, en este caso en particular la recta $y = a + bx$; segundo, sustituimos las variables por las sumatorias de dichas variables al tiempo en que la constante independiente la multiplicamos por el número de observaciones hechas n, obteniendo

$$\sum y = na + b\sum x\,;$$

tercero, multiplicamos los elementos de esta ecuación por x y sustituimos n por la suma de x, resultando

$$\sum xy = a\sum x + b\sum x^2\,;$$

cuarto, creamos con estas dos últimas ecuaciones el sistema

$$\begin{cases}\sum y = na + b\sum x \\ \sum xy = a\sum x + b\sum x^2\end{cases}$$

de dos ecuaciones con dos incógnitas, que son precisamente a y b, ya que las sumas son fácilmente realizables, y el valor de n conocido. En nuestro ejemplo bástese con agregar las columnas con los valores xy resultantes de multiplicar cada x por su pareja y con los valores x^2 resultantes de elevar cada estatura al cuadrado, y realizar las sumas:

Estatura x	Peso y	xy	x^2
1.79	68	121.72	3.2041
1.60	55	88	2.56
1.60	60	96	2.56
1.78	65	115.7	3.1684
1.64	52	85.28	2.6896
8.41	*300*	*506.70*	*14.1821*

con lo que el sistema correspondiente al problema, sabiendo que n = 5, es

$$\begin{cases}300 = 5a + 8.41b \\ 506.70 = 8.41a + 14.1821b\end{cases}$$

y el cual puede ser resuelto por cualquiera de los métodos ya conocidos. En este caso recurrimos al método de suma y resta:

$$\begin{cases}(300 = 5a + 8.41b)(8.41) \\ (506.70 = 8.41a + 14.1821b)(5)\end{cases}$$

$$\begin{cases}2523 = 42.05a + 70.7281b \\ 2533.50 = 42.05a + 70.9105b\end{cases}$$

$$-10.5 = \qquad -0.1824b$$

$$\frac{-10.5}{--0.1824} = b$$

b = 57.566

300 = 5a + (8.41)(57.566)

300 = 5a + 484.128

300 – 484.128 = 5a

-184.128 = 5a

$$\frac{-184.128}{5} = a$$

a = -36.826
aproximadamente.

Así, hemos encontrado que a = -36.826 y b = 57.566 aproximadamente, con lo que la recta a la que mejor se ajustan nuestros datos es
y = -36.826 + 57.566x
y a partir de la cual podemos realizar la estimación pedida en un principio -el peso (y) de alguno de 1.92 (x) de estatura- sustituyendo 1.92 en x en nuestra función:
y = -36.826 + (57.566)(1.92) = 73.701
con lo cual podemos decir que se esperaría que alguno de 1.92 mt de estatura pese 73.701 kg .

De la misma manera podríamos asumir un valor arbitrario para el peso y a partir de él estimar la estatura. Sólo sustituiríamos dicho valor en y, y hallaríamos x (la estatura) a partir de simples despejes.

Coeficiente de correlación

Existe un elemento más a tratar en el modelo lineal, y corresponde a un valor con el cual podemos decidir sobre una base un poco más teórica si efectivamente el ajuste es o no recomendable para nuestros datos, y este elemento es el llamado *Coeficiente de Correlación*, que es un valor que denotamos por r y que se calcula con la fórmula

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

Tiene la característica de ser un valor dentro del intervalo [-1, 1], y a medida que r tome un valor cercano a -1 o 1, entonces nuestros datos tenderán a comportarse más como una recta, mientras que si r es cercano a 0 entonces nuestras observaciones aparejadas guardarán una mínima o nula relación entre sí y el ajuste así como las estimaciones carecerán de exactitud. En cuanto a qué tan cercano a 1 o a –1 tiene que ser r para que pueda establecerse que efectivamente existe una relación lineal entre las dos variables involucradas dentro de algún problema de estimación lineal, se dijo que ello dependía tanto del criterio del investigador como de la naturaleza del problema que se estuviera resolviendo, y que finalmente podría apoyarse esta decisión un poco en el diagrama de dispersión. Cabe decir que el *coeficiente de determinación*, r^2, que nos permite saber qué tanto porcentaje de error con respecto a la recta ideal se ha eliminado en el ajuste (dese 0 hasta 1, en términos de su valor aritmético), no es un elemento que se contemple dentro del curso. Para calcular r en nuestro ejemplo, agregamos una columna con los cuadrados de los pesos (es decir, con valores y^2) y calculamos su correspondiente suma:

Estatura x	Peso y	xy	x^2	y^2
1.79	68	121.72	3.2041	4624
1.60	55	88	2.56	3025
1.60	60	96	2.56	3600
1.78	65	115.7	3.1684	4225
1.64	52	85.28	2.6896	2704
8.41	*300*	*506.70*	*14.1821*	*18178*

así

$$r = \frac{5(506.70)-(8.41)(300)}{\sqrt{(5(14.1821)-(8.41)^2)(5(18178)-(300)^2)}} = 0.824103727979$$

aproximadamente, y que es un valor cercano a 1, con lo que podemos asumir que el ajuste lineal es adecuado y las estimaciones confiables.

Una vez finalizado este ejemplo junto con los correspondientes planteamientos teóricos, se plantearon al grupo los siguientes problemas para su resolución:

-Ajustar los siguientes datos a una recta de mínimos cuadrados, trazar el diagrama de dispersión, hallar el coeficiente de correlación y estimar el valor de y si x = 10:

x	y
-2	0
-1	0
0	1
1	1
2	3

-Los auditores deben comparar muchas veces el valor revisado de un artículo del catálogo de un inventario con el valor en los libros (o nominal). Si una compañía tiene su inventario y sus libros al día debe existir una gran relación entre los valores revisados y los valores nominales. Una muestra de 10 artículos del catálogo de cierta compañía dio los datos siguientes acerca de los valores revisados y los valores nominales. Ajuste el modelo y = a + bx a través de los datos y calcule el coeficiente de correlación. Si el valor nominal es igual a 100, ¿cuál será el valor revisado estimado?

Valor nominal x	Valor revisado y
10	9
12	14
9	7
27	29
47	45
112	109
36	40
241	238
59	60
167	170

-Los promedios de los precios de venta de casas nuevas unifamiliares durante un período de 8 años se indican en la tabla siguiente. Ajuste el modelo $y = a + bx$, calcule el coeficiente de correlación y haga una estimación del valor promedio de las casas unifamiliares para 1997.

Año	Precio promedio de venta (en miles)
1989	27.6
1990	32.5
1991	35.9
1992	39.3
1993	44.2
1994	48.8
1995	55.7
1996	62.9

Con este problema, y previamente a que el grupo comenzara la resolución, se dijo que en muchas ocasiones el método de estimación por mínimos cuadrados se aplica a series de tiempo como la del ejercicio, ya que hay muchas cuestiones (inflación, crecimiento demográfico, esperanza de vida) que varían en función del tiempo. Pero eso implica un problema: el tomar esta serie de periodos de tiempo tal y como se listan genera valores muy grandes y poco manejables en los cuadrados de x y en los productos xy, por lo que se recomienda contabilizar estos periodos a partir de uno, cuidando solamente que sean iguales (todos anuales, o mensuales, o quinquenales, etcétera). De esta forma la tabla quedó como

Año	Precio promedio de venta (en miles)
1	27.6
2	32.5
3	35.9
4	39.3
5	44.2
6	48.8
7	55.7
8	62.9

y la estimación se realizará sobre $x = 9$.

-Sobre la base de la tabla siguiente ajuste el modelo $y = a + bx$, calcule el coeficiente de correlación y obtenga una estimación para $x = 5$.

x	y
-2	3
-1	2
0	1
1	1
2	0.5

-En la siguiente tabla se muestra la clasificación combinada del número de millas por galón y el volumen del motor para 9 automóviles subcompactos, con transmisión automática, de cuatro cilindros y que utilizan gasolina. El tamaño del motor se da en pulgadas cúbicas totales de cilindrada. Utilice la recta de mínimos cuadrados para estimar el promedio de millas por galón para un automóvil subcompacto que tiene un volumen del motor de 125 in^3. Calcule el coeficiente de correlación.

VEHICULO	CILINDRADA x	MPG (COMBINADO) y
VW Rabbit	97	24
Datsun 210	85	29
Chevette	98	26
Dodge Omni	105	24
Mazda 626	120	24
Oldsmobil Starfire	151	22
Mercury Capri	140	23
Toyota Celica	134	23
Datsun 810	136	21

b) Parábola de mínimos cuadrados

El segundo modelo a trabajar es aquel en el que el diagrama de dispersión presenta un comportamiento de los pares ordenados similar al de una parábola. En este caso buscaremos ajustar nuestros datos a una curva de la forma $y = a + bx + cx^2$, y para ello seguiremos un proceso similar al planteado en el modelo lineal, aunque un paso más extenso. En el modelo lineal desarrollamos un sistema de dos ecuaciones dado que el problema es hallar dos incógnitas -a y b-, pero como en el modelo cuadrático se tienen tres incógnitas -a, b y c- entonces se requerirá de un sistema de tres ecuaciones. Este sistema lo desarrollamos a partir primero de sustituir las variables por las sumatorias de dichas variables al tiempo en que la constante independiente la multiplicamos por el número de observaciones hechas n, obteniendo

$$\sum y = na + b\sum x + c\sum x^2 \;;$$

luego multiplicamos los elementos de esta ecuación por x y sustituimos n por la suma de x, resultando

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \;;$$

después multiplicamos de nueva cuenta esta segunda expresión, término a término, por x, logrando

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \,,$$

para finalmente generar nuestro sistema de tres ecuaciones con tres incógnitas a partir de estas tres últimas igualdades, de esta manera:

$$\begin{cases} \sum y = na + b\sum x + c\sum x^2 \\ \sum xy = a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \end{cases}$$

reduciéndose todo a la solución del sistema, la sustitución en el modelo, y la sustitución del valor de variable a partir del cual se busca hacer la estimación.

Cabe decir que para este modelo ya no manejaremos coeficiente de correlación. Como ejemplo para este modelo de ajuste se presentó el siguiente ejercicio:

-Ejemplo.- Ajuste los siguientes datos a una parábola de mínimos cuadrados, verificando la validez de dicho ajuste con el diagrama de dispersión. Estime el valor de y para x = 10.

x	y
-2	5
-1	3
0	1
1	0
2	-1
3	0
4	1
5	2

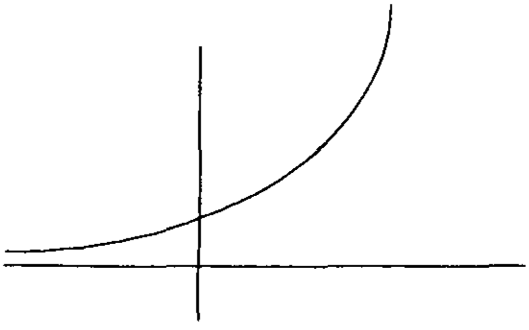
Terminado el ejemplo explicativo que siguió al planteamiento teórico, el grupo resolvió el siguiente ejercicio:

-Trace el diagrama de dispersión para los siguientes datos, ajústelos por medio de una parábola de mínimos cuadrados si es esto adecuado y estime el valor de y para x = 4.

x	y
-2	1
-1	0
0	2
1	0
2	2

c) Exponencial de mínimos cuadrados

De este modelo se dijo que se recurriría a él en caso de que el diagrama de dispersión mostrara un comportamiento semejante al de la función y = e^x, función que ya había sido manejado con las funciones de distribución de probabilidad de Poisson y Exponencial, y cuya gráfica tiene la forma



En este caso, la curva a la que se ajustarán los datos que muestren un comportamiento similar al de esta gráfica será $y = ae^{bx}$, en dónde de nueva cuenta se requiere encontrar dos constantes -a y b-, pero con la dificultad de que desde este modelo no pueden establecerse directamente las dos ecuaciones para el sistema como se hizo en el modelo lineal, ya que en primera instancia a es un factor, y en segunda que b es parte de una potencia. Para salvar este problema recurrimos a aplica a ambas partes la función *logaritmo natural*, que es la función inversa de la función exponencial, por lo que:
$\ln(y) = \ln(ae^{bx})$.

Ahora, aplicamos una propiedad de la función logaritmo natural[1] que dice que el logaritmo natural de un producto es igual a la suma de los logaritmos naturales de cada uno de los factores, con lo que tenemos
$\ln(y) = \ln(a) + \ln(e^{bx})$,
pero al ser la función exponencial y la función logaritmo natural una inversa de la otra, operativamente se cancelan y nos queda
$\ln(y) = \ln(a) + bx$
en donde ln(a) es una constante, con lo cual nuestro modelo tiene ahora una estructura similar a la del modelo lineal, a partir de la cual podemos desarrollar nuestro sistema de dos ecuaciones con dos incógnitas de la misma manera en que lo hicimos anteriormente, obteniendo el sistema

$$\begin{cases} \sum \ln(y) = n\ln(a) + b\sum x \\ \sum x\ln(y) = \ln(a)\sum x + b\sum x^2 \end{cases}$$

Aquí deberá de tenerse en cuenta que al resolver el sistema obtendremos directamente el valor de b, pero no el de a, sino el de ln(a), por lo que una vez calculado ln(a), a dicho valor tendremos que aplicarle la función inversa de logaritmo natural -la función exponencial- para obtener a y poder plantear nuestro modelo para cada problema. Finalmente se estableció que tampoco aquí aplicaríamos el coeficiente de correlación, por lo que la selección del modelo se sustentaría exclusivamente en el diagrama de dispersión.

Para aclarar todo esto un poco más se realizó un ejemplo demostrativo en el cual también se enseñó al grupo a calcular el logaritmo natural de un número utilizando la calculadora, aunque no se dio ninguna explicación teórica, salvo algo muy elemental y de manera personal a aquellos alumnos que tuvieron la curiosidad de preguntar el significado matemático de la función.

-Ejemplo.- Ajustar una exponencial de mínimos cuadrados a los datos siguientes, si es que el diagrama de dispersión muestra un comportamiento acorde con el modelo. Estimar para x = 15.

[1] Estas propiedades se presentaron como un hecho, dado que el tema del Logaritmos y las correspondientes demostraciones de sus propiedades quedan fuera del alcance de este curso.

x	y
3	202
4	1490
5	11013
6	81377
7	601302
8	4443055
9	32829985

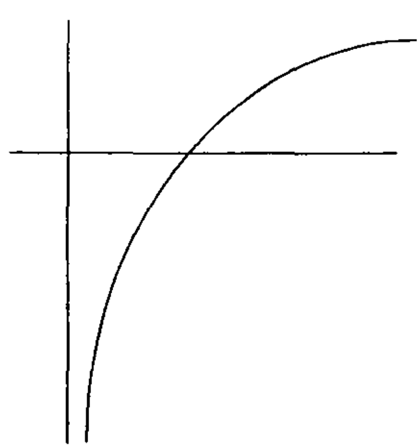
Una vez terminado dicho ejemplo, la resolución del siguiente ejercicio corrió a cargo del grupo:

-Trazar el diagrama de dispersión de los siguientes datos, y ajustarlos por medio de una exponencial de mínimos cuadrados si esto parece factible. Estimar el valor de y si x = 7.

x	y
-1	3
0	4
1	5
2	6
4	10
10	48

d) Logarítmica de mínimos cuadrados

Este modelo se presentó como el requerido cuando nuestro diagrama de dispersión presente aproximadamente el trazo de la gráfica de la función logaritmo natural $y = \ln(x)$. y que es



y en donde se buscará ajustar los datos a la curva $y = a + b\ln(x)$, cuya estructura es semejante a la del modelo lineal, y a partir de la cual obtendremos nuestro sistema de dos por dos del mismo modo que en dicho modelo lineal. El sistema será

$$\begin{cases} \sum y = na + b\sum \ln(x) \\ \sum xy = a\sum x + b\sum x\ln(x) \end{cases}$$

Y para verificar la solución a través de este modelo se realizó un ejemplo, aclarando que de nueva cuenta el coeficiente de correlación no aparecería en este tipo de ajustes.

-Ejemplo.- Ajustar una logarítmica de mínimos cuadrados con los datos siguientes, verificando antes por medio del diagrama de dispersión si este ajuste es adecuado.

Estimar y para x = 500.

x	y
0.5	2.6
1	3
3	3.5
10	4.1
15	4.3
125	5.4

Para finalizar con el modelo, se dio al grupo el siguiente problema:

-Ajustar una logarítmica de mínimos cuadrados a los siguientes datos si el diagrama de dispersión muestra que dicho ajuste es adecuado, y estime y para x = 10.

x	y
2	11
7	14
17	16
200	21
1500	25

Para concluir con el tema, se dijo al grupo que de los cuatro modelos vistos, es factible encontrar situaciones o variables aparejadas que efectivamente se ajusten a cualquiera de ellos, pero que, primero, un modelo cuadrático no es algo muy común, y segundo, que en el caso de los modelos logarítmico y exponencial, salvo que se tengan demasiados datos y estos presenten de manera muy evidente cualquiera de los dos comportamientos, estos dos modelos pueden ser fácilmente sustituidos por el modelo lineal, ya que los diagramas de dispersión de estas curvas igualmente se asemejan a una recta (y para ejemplificarlo se remitió a los alumnos a los diagramas de dispersión trazados para ambos modelos).

En esta unidad no se presentaron problemas didácticos más allá de los correspondientes al manejo de la calculadora. Incluso en algunos casos -cuando la calculadora de algunos alumnos lo permitía- se explicó cómo obtener las sumatorias con solo cargar las variables. El resto de la comprensión fue fácil dados los antecedentes de Estadística Descriptiva, de Funciones de Distribución de Probabilidad de Variable Aleatoria Continua y de Estadística Inferencial.

Unidad VI, Algunas consideraciones en el Diseño de Experimentos

Objetivo:
Que el alumno conozca algunos elementos básicos a considerar dentro del diseño de un experimento, a partir de retomar los elementos básicos de Muestreo y Estimación y relacionarlos con los conocimientos adquiridos en el curso.

Como estrategia didáctica el profesor expondrá ante el grupo algunos de los principales elementos propios del Diseño de Experimentos, haciendo preguntas a los alumnos al final de dicha exposición.

6. Algunas consideraciones en el diseño de experimentos

Esta unidad se presentó como un resumen de todo el Proceso Estadístico visto a lo largo de los dos cursos, e igualmente como una serie de cuestiones a tomar en cuenta al tratar de obtener información a través de la Estadística.

Se sabe que con la Estadística se busca obtener información acerca de alguna característica específica de una población. Ahora, hay algunos elementos que deben de tomarse en cuenta antes de comenzar a buscar dicha información con el fin de optimar los resultados. Con esto podríamos pensar que entonces esta unidad debería de abordarse antes de la unidad correspondiente a Estadística Descriptiva en el primer curso, pero dado que entonces el conocimiento de los alumnos acerca de la materia era muy limitado, mucho de lo que aquí se verá habría resultado incomprensible.

Para entrar en materia, primero debemos de tener presente que la información que podamos obtener, sobre todo por la cuestión de la aleatoriedad, no será exacta, pero siempre buscaremos que sea lo más precisa posible. De esta forma es conveniente identificar que elementos afectan la información que se obtenga de una muestra. Para explicar mejor esto se puso como ejemplo la información que una persona pueda transmitir a otra a través del teléfono, o la que un orador pueda transmitir a un auditorio; en ambos casos si se desea que la información sea transmitida lo mejor posible se requiere de que no exista ruido que interfiera con el mensaje, o de otra manera quien escuche tendrá que inferir parte del mensaje; y si existe interferencia también ello obliga a quien habla a aumentar la fuerza en la voz para poder transmitir la información. En Estadística sucede algo semejante: si se busca mayor información deberá o reducirse la interferencia o aumentar la intensidad, que retirando la analogía implica buscar que la variabilidad en los resultados sea mínima o que la muestra sea lo más representativa posible, cuestión que puede conseguirse si el tamaño de dicha muestra tiende a ser grande. Pero ambas cuestiones, la reducción de la interferencia y el aumento en la intensidad, conllevan, como se verá, un cierto grado de complejidad.

Previo al análisis de los problemas de disminución de interferencia y aumento de intensidad, se vieron algunas definiciones más de elementos involucrados dentro de la Estadística:

-Definición.- Los objetos sobre los cuales se hacen mediciones se denominan *unidades experimentales.*

Ahora, para obtener información de las variables experimentales, debe hacerse algo con ellas que las haga diferentes del resto de los individuos. Por ejemplo, si se desea saber qué tanta corriente eléctrica pasa a través de una resistencia bajo distintas temperaturas, entonces tendríamos que variar la temperatura de las resistencias tomadas como unidades experimentales y luego hacer cruzar por ellas una cierta cantidad de corriente eléctrica; o bien podemos querer comprobar en cuál estación del año los caimanes ocupan mayor área en cierta región, con lo cual deberemos hacer mediciones en primavera, verano, otoño e invierno. Estas cuestiones que generan información de las unidades experimentales se denominan *tratamientos*, aunque veremos una definición formal más adelante, dado que requeriremos de otra serie de definiciones previas.

-Definición.- Las variables experimentales independientes se denominan *factores.*

-Definición.- Un *factor cuantitativo* es un factor que puede tomar valores correspondientes a los puntos de una recta real. Los factores que no son cuantitativos se denominan *cualitativos.*
-Definición.- Al grado de intensidad de un factor se le llama *nivel.*
-Definición.- Un tratamiento es una combinación específica de niveles de un factor o de factores.

Con lo anterior, se procedió a presentar los pasos aplicados en el diseño de un experimento, ejemplificando con alguno de los problemas resueltos a lo largo del curso:

Pasos aplicados en le diseño de un experimento

1. Se seleccionan los factores que deben incluirse en el experimento y se especifica el o los parámetros de interés.
2. Se decide cuanta información conviene utilizar acerca de los parámetros de interés.
3. Se seleccionan los tratamientos que deben utilizarse en el experimento y se decide el número de unidades experimentales que deben asignarse a cada uno. En este tercer punto se tiene la alternativa de aumentar la intensidad en la información.
4. Se decide como deben aplicarse los tratamientos a las unidades experimentales. En este cuarto punto se tiene la alternativa de disminuir la interferencia

Otro punto a tratar, previamente a la disminución de la interferencia y al aumento en la intensidad, es lo tocante al Muestreo. Hasta ahora la única referencia que se ha manejado corresponde al *muestreo aleatorio simple*, y que es aquel que se caracteriza por el hecho de que justo antes de tomar la muestra debe cumplirse que todos y cada uno de los elementos de la población tengan exactamente la misma probabilidad de ser seleccionados para dicha muestra. Este tipo de muestreo tiene dos propósitos: uno el evitar sesgos que pudieran darse por hacer una selección no aleatoria (como el realizar una encuesta de edades sólo a la salida de una primaria dentro de un estudio sobre las edades en una cierta ciudad), y otro el de suministrar la base probabilística en la que se sustenta el trabajo de la Estadística Inferencial (al saber como se distribuye una población, o en su defecto la muestra, podemos establecer la manera de realizar una estimación por intervalos o una prueba de hipótesis).

Ahora, podemos llevar el muestreo aleatorio simple un paso más adelante, específicamente para experimentos en los que se busque comparar dos o más poblaciones. Pensemos en el ejemplo de querer comparar la duración para cinco distintas marcas de llantas; evidentemente no seleccionaríamos por ejemplo 100 llantas para luego verificar la marca, ni asignaríamos cada quinta parte como si fuera una de las marcas, sino que en realidad se tomarían cinco muestras aleatorias de cada marca y luego procederíamos a las comparaciones. Un diseño semejante al de este ejemplo se le conoce como *diseño completamente aleatorizado*, y que podemos definir de la siguiente forma:
-Definición.- A la selección de muestras aleatorias independientes de k poblaciones se la denomina *diseño completamente aleatorizado.*

El siguiente elemento dentro de esta unidad corresponde ya al diseño de experimentos en los que se aumenta la intensidad. Aquí se explicó que en este punto aparece uno de los problemas más fuertes dentro del muestreo: la selección del tamaño de la muestra. En principio se planteó la posibilidad de que, si se desea realizar una estimación por intervalo o una prueba de hipótesis cuyas fórmulas involucran precisamente n, podía realizarse primero una pequeña prueba con una muestra de 30 o 40 elementos, (ya que así se tendrá una muestra distribuida normalmente) y después con los valores a obtenidos de los estimadores y la decisión del coeficiente de confianza o del nivel de significación, despejar n de la fórmula y obtener así una idea aproximada de que tan grande debe ser. También se mencionó la existencia de algunas expresiones para dicho cálculo que resultan útiles sólo bajo ciertas condiciones. Y también se dijo que en muchas ocasiones el investigador decide el valor de n basado exclusivamente en su experiencia. De cualquier forma, se hizo énfasis en el hecho de que el valor de n es un elemento sumamente importante, ya que una n muy chica no arroja información muy confiable, pero que una n demasiado grande puede ser, para efectos prácticos de recopilación de información, poco manejable además de muy costosa.

A nivel un poco más práctico, se comentó que algunos métodos propios del Cálculo Diferencial pueden ser útiles al abordar este problema; y que podrían plantearse un par de sugerencias: una, que al realizar un experimento en el que se tenga una comparación entre dos medias, proporciones o variancias, se busque que las dos muestras (o con las que se trabaje) sean del mismo tamaño; y otra, que al tener el diseño de un experimento en el que se vaya a realizar una estimación por medio de una recta de mínimos cuadrados, es conveniente construir un intervalo de valores x, y tomar la mitad de las observaciones a realizar en el límite inferior del intervalo, la otra mitad en el límite superior, y dos o tres puntos más a la mitad de estos dos límites, lo que reducirá las variaciones de las observaciones con respecto a la "recta ideal" a la cual se desea ajustar los datos observados.

Finalmente, se explicó una forma de reducir la interferencia. Esta interferencia usualmente se da cuando hay datos demasiado heterogéneos, con lo cual podremos tener menos interferencia (o alternativa de identificarla más fácilmente) cuando tenemos observaciones más homogéneas. Un experimento que reduzca la heterogeneidad puede realizarse a través de un diseño que se conoce como diseño aleatorizado de bloques, el cual se define como:

-Definición.- Un diseño aleatorizado de bloques que contiene b bloques y p tratamientos, consiste en b bloques de p unidades experimentales cada uno. Se asignan aleatoriamente los tratamientos a las unidades en cada bloque, y cada tratamiento aparece exactamente una sola vez en cada bloque.

Para concluir con este punto se dijo que existen otras formas de generar muestras de modo tal que se optimen los resultados, pero que no se vería más allá de los puntos ya tocados, por no ser este un curso de Muestreo y porque precisamente un curso de dicha naturaleza requeriría de mayores antecedentes escolares que aquellos con los que el grupo contaba hasta ese momento. De este modo se dio por finalizada la última unidad del curso.

En este tema no se presentó ningún problema de comprensión, dado el antecedente de los dos cursos y el carácter casi nada práctico con el que se presentó la unidad al grupo. La única dificultad con la que se tuvo que lidiar en la sesión y media que consumió el tema fue el del aburrimiento en los alumnos por el carácter discursivo que se dio a la unidad.

CONCLUSIONES

Conclusiones

De los cursos que se han descrito en el presente trabajo se puede decir que en principio partían de la finalidad de que los alumnos que los llevaran aprendieran algunas cuestiones elementales de la Estadística. Si esto se conseguía entonces en consecuencia se cumplirían también los objetivos con los que se inició dicha descripción: dotar de una herramienta práctica para la investigación, sin perder por completo las bases teóricas, y reducir los índices de reprobación en la materia.

A lo largo de las unidades que componen los dos cursos se han especificado los problemas más recurrentes en el aprendizaje, y escuetamente la forma en que dichas complicaciones fueron, en la medida de lo posible, resueltas. En casi todas las veces en que se presentaron dificultades, estas forzaron al profesor a trabajar directamente con cada uno de los estudiantes y buscar una forma específica de explicar adecuada a cada alumno; fueron mínimas las ocasiones en las que se repitió algo completo para todo el grupo como resultado de una total incomprensión. Este trato personal generó una mayor confianza entre los alumnos, tanto para cuestionar al profesor como para decidirse a adentrarse sin miedo en la materia y a arriesgarse a intentar los procedimientos. Además esta confianza tendió a reforzarse con los métodos de evaluación, en donde se dio más importancia al trabajo hecho en clase y al esfuerzo por aprender la materia y demostrar este aprendizaje clase con clase, que a la tradicional evaluación numérica sustentada básicamente en exámenes, con la ventaja para el profesor de que el trato personal con aquellos alumnos que presentaron problemas de comprensión permitió una evaluación del aprendizaje mucho más directa y objetiva precisamente con quienes más la requerían.

Otra ventaja que se tuvo dentro del diseño de los cursos fue el hecho de que la materia no resultó tediosa en términos generales, ya que en casi todas las clases se mantuvo a los alumnos en actividad constante con la resolución de los problemas, o con la atención en las explicaciones desde el momento en que comprendieron que esa era la única vía hacia el entendimiento de los temas y de las dinámicas que se les exigiría dominar más adelante. Igualmente este diseño de trabajo, en la que el grupo resolvió problemas que luego fueron realizados por el profesor a fin de verificar los resultados y aclarar dudas, permitió abarcar un programa que en apariencia resultaría demasiado extenso para dos semestres, así como desarrollar un dominio, si bien no total, al menos adecuado de los procesos estadísticos y probabilísticos tratados.

Todo lo anterior efectivamente tuvo como consecuencia una disminución notable en los índices de reprobación usuales en la materia, aunque la forma de trabajo (continuo y de mucha carga) provocó altos índices de deserción de alumnos no dispuestos a tal ritmo en clases. Igualmente pudo optimizarse el proceso de enseñanza-aprendizaje en la mayoría de los temas vistos (las mayores dificultades se dieron en términos de la interpretación de resultados dentro de la Estadística Inferencial), así como en los tiempos establecidos en cada semestre. Dicha optimización pudo observarse al mostrar los estudiantes una cierta capacidad de abstracción e interpretación de resultados, un mayor dominio del lenguaje estadístico y una manera más ordenada de abordar los problemas que a lo largo de los dos semestres se les fueron presentando.

En lo tocante a otro de los objetivos perseguidos en el curso, debe reconocerse que una medición real de la utilidad que como herramientas comprendidas y aprendidas tendrán los temas abordados entre los alumnos no será posible de realizar mas que por cada uno de dichos estudiantes y por los resultados que ellos arrojen dentro de su desarrollo como estudiantes de nivel licenciatura o como egresados de alguna carrera y dentro del ámbito laboral y profesional, aunque existe la confianza de que buena parte de ellos lograrán aplicar la Estadística en el corto, mediano y largo plazos.

Por lo que respecta a los objetivos perseguidos con el presente trabajo puede decirse que efectivamente se ha conseguido crear un documento de apoyo a los cursos de Estadística I y II creado y probado dentro del contexto del CCH Naucalpan, en el cual pueden encontrarse ejercicios, actividades, conceptos, cuadros y tablas, fórmulas, elementos de evaluación y algunas consideraciones didácticas que pueden ser de utilidad para quien imparta la materia. Lo que aún no ha podido medirse es que este material efectivamente sea utilizado por todos aquellos profesores interesados en la enseñanza de la Estadística en el CCH Naucalpan, aunque partes del presente documento ya han sido tomadas y utilizadas en cursos por algunos maestros del plantel, por lo que podría pensarse que en un futuro este trabajo pueda tener completa aceptación entre aquellos a quienes pueda resultarles de utilidad.

Finalmente cabe decir que la factibilidad de un curso de esta naturaleza parte del compromiso del profesor quien lo aplique a dar lo mismo que pide a los alumnos: trabajo. Y a dar también un 100% de asistencia, ya que de su presencia depende buena parte del funcionamiento de cada clase, puesto que muchos de los conceptos y desarrollos prácticos deben de explicarse al alumno, así como por el requerimiento de una medición constante de la comprensión de lo enseñado en clase, y dado también que ambos programas son demasiado ambiciosos en cuanto a cobertura de temas.

BIBLIOGRAFIA

Bibliografía

Probabilidad y Estadística. Aplicaciones y métodos.- Canavos, George C. McGraw-Hill. México, 1988.

Estadística.- Hayslett Jr., H. T. Compañía General de Ediciones, S. A. México, 1978.

Estadística Elemental.- Johnson, Robert. Grupo Editorial Iberoamérica. México, 1990.

Estadística Matemática con aplicaciones.- Mendenhall; Scheaffer; Wackerly. Grupo Editorial Iberoamérica. México, 1986.

Técnicas de Conteo.- Olivera Salazar; Zúñiga Barrera. Noriega Editores. México, 1987.

Estadística.- Spiegel, Murray R. McGraw-Hill. Colombia, 1969.

Estadística para Administración y Economía.- Stevenson, William. Editorial Harla. México, 1994.

¿De cuántas formas?.- Vilenkin. Editorial Mir. URSS, 1972.

Estadística y Probabilidad.- Walpole; Myers. McGraw-Hill.

Razonamiento Estadístico.- Williams, Frederick. Editorial Nuevo Mundo.

Estadística.- Yamane, Taro. Editorial Harla. México, 1979.

ANEXOS

Anexos

A continuación se presentan modelos de los exámenes parciales y finales de cada uno de los cursos descritos en este trabajo. Igualmente se presentan modelos de exámenes extraordinarios para ambos cursos. Cabe señalar que estos diseños se realizaron sobre un temario distinto al manejado en los cursos ordinarios, bajo la idea de incluir los temas que impartieran la generalidad de los profesores, y dar así alternativas más reales de aprobación a aquellos alumnos que tuviesen que presentar cualquiera de las dos asignaturas en instancia extraordinaria. Mas adelante se hará referencia a estos temarios para examen extraordinario, igualmente diseñados por mí.

ESTADÍSTICA I. PRIMER PARCIAL

1. HALLAR:
A) MEDIA Y MODA
B) DESVIACIÓN ESTÁNDAR
C) 3ER. DECIL
D) COEFICIENTE DE APUNTALAMIENTO (α_4)
E) POLÍGONO DE FRECUENCIAS

CLASES	F
10 – 19	2
20 – 29	10
30 – 39	12
40 – 49	9
50 – 59	7

(5 PUNTOS).

2. ¿DE CUÁNTAS FORMAS DIFERENTES PUEDEN SALIR CINCO SELECCIONADOS DE UN TORNEO SELECTIVO DE 18 ATLETAS? (2.5 PUNTOS).

3. ¿DE CUÁNTAS FORMAS PUEDE CONTESTARSE UN EXAMEN DE 10 PREGUNTAS DE OPCIÓN MÚLTIPLE, SI CADA PREGUNTA TIENE TRES DISTINTAS ALTERNATIVAS? (2.5 PUNTOS).

ESTADÍSTICA I. SEGUNDO PARCIAL

1. CADA AÑO, EN CIERTA REGIÓN, UN PROMEDIO DE 30 CRÍAS DE FOCAS ES CAZADO PARA HACER ABRIGOS. HALLAR LA PROBABILIDAD DE QUE EN EL PROXIMO AÑO SEAN CAZADAS EN LA REGIÓN ENTRE 25 Y 27 CRÍAS DE FOCA. (2 PUNTOS).

2. DE LA POBLACIÓN ASIÁTICA DE TIGRES, SOLO EL 10% SON TIGRES DE BENGALA. ¿QUÉ PROBABILIDAD HAY DE QUE AL CAPTURAR A 10 TIGRES, MÁS DE 2 SEAN DE BENGALA? (2 PUNTOS).

3. EN LA CACERÍA DE BALLENAS EL 30% DE LAS BALLENAS MUERTAS SON ORCAS, 35% SON BALLENAS JOROBADAS, EL 10% SON BALLENAS AZULES Y LAS RESTANTES SON BALLENAS GRISES. AL ATRAPAR UN BALLENERO JAPONÉS A 12 BALLENAS ¿QUÉ PROBABILIDAD HAY DE QUE LA DISTRIBUCIÓN SEA DE 6 ORCAS, 3 JOROBADAS, 2 AZULES Y UNA GRIS? (1 PUNTO).

4. EN UN ZOOLÓGICO SE TIENEN EN LA MISMA ÁREA A TODOS LOS LOBOS: 5 LOBOS PLATEADOS, 6 LOBOS NEGROS, 8 LOBOS GRISES, 7 LOBOS EUROPEOS Y 4 LOBOS MEXICANOS. SI SE ESCAPAN 19 LOBOS DEL ZOOLÓGICO ¿QUÉ PROBABILIDAD HAY DE QUE SE HAYAN ESCAPADO 3 LOBOS PLATEADOS, 4 LOBOS NEGROS, 5 LOBOS GRISES, 5 LOBOS EUROPEOS Y 2 LOBOS MEXICANOS? (1 PUNTO).

5. LA PROBABILIDAD DE QUE UN PUMA SOBREVIVA A UNA PARTIDA DE CAZADORES ES DE 15%. ¿QUÉ PROBABILIDAD HAY DE QUE UN PUMA PUEDA ESCAPAR A 4 PARTIDAS DE CAZADORES ANTES DE SER MUERTO POR ELLOS? ¿QUÉ PROBABILIDAD HAY DE QUE EL CUARTO PUMA QUE ESCAPE LO HAGA DE LA SÉPTIMA PARTIDA? (2 PUNTOS).

6. EN UNA MANADA DE 20 RENOS HAY 7 QUE HAN SIDO HERIDOS POR LAS TRAMPAS. SI UN GRUPO DE VETERINARIOS CAPTURA A 10 DE ESTOS RENOS, ¿QUÉ PROBABILIDAD HAY DE QUE ENTRE ELLOS HAYA MAS DE 5 RENOS HERIDOS? (2 PUNTOS).

EXAMEN FINAL Y DE RECUPERACIÓN. ESTADÍSTICA I

INSTRUCCIONES.- SI PRESENTAS RECUPERACIÓN DEL PRIMER EXAMEN. RESUELVE LAS PREGUNTAS 1 Y 2. SI PRESENTAS RECUPERACIÓN DEL SEGUNDO EXAMEN, RESUELVE LAS PREGUNTAS 3 Y 4. SI PRESENTAS EXAMEN FINAL, RESUELVE LAS CUATRO PREGUNTAS.

1. CON

CLASES	F
14 – 16	15
16 – 18	10
18 – 20	7
20 – 22	5
22 – 24	3

HALLAR MEDIA, MODA, VARIANCIA, HISTOGRAMA Y POLÍGONO DE FRECUENCIAS.

2. ¿DE CUÁNTAS FORMAS DIFERENTES PUEDEN ORGANIZARSE LOS RECORRIDOS POR LAS 7 SALAS MÁS IMPORTANTES DEL MUSEO NACIONAL DE ANTROPOLOGÍA?

3. CIERTO MUSEO QUE CUENTA CON UNA BUENA CANTIDAD DE SALAS, TIENE UN PROMEDIO DE 28 PIEZAS EN EXHIBICIÓN POR SALA. ¿QUÉ PROBABILIDAD HAY DE QUE SI ENTRAMOS EN UNA DE SUS SALAS AL AZAR HAYA EN ELLA ENTRE 25 Y 28 PIEZAS EN EXHIBICIÓN?

4. EL 88% DE LAS PERSONAS QUE VAN AL MUSEO DE HISTORIA NATURAL ACUDEN OBLIGADOS POR LAS ESCUELAS. ¿QUÉ PROBABILIDAD HAY DE QUE DE 15 PERSONAS QUE ESTAN VIENDO LOS HUESOS DEL BRONTOSAURIO MÁS DE 12 ESTEN POR VOLUNTAD PROPIA Y NO POR OBLIGACIÓN?

ESTADÍSTICA II. PRIMER PARCIAL

1.UNA MUESTRA DE 12 SANDÍAS PESÓ EN PROMEDIO 9.8 LB, CON UNA DESVIACIÓN ESTÁNDAR DE 2.8 LB. ESTIMAR EL PROMEDIO REAL DE PESO DE LAS SANDÍAS CON UN INTERVALO DE CONFIANZA DEL 90%, SUPONIENDO QUE LOS PESOS SE DISTRIBUYEN NORMALMENTE. INTERPRETAR (2 PUNTOS).

2. SE MIDIÓ EL TIEMPO DE ESPERA EN LA FILA DE UN BANCO PARA 50 CLIENTES Y SE OBTUVIERON UNA MEDIA DE 3.7 MIN. Y UNA DESVIACIÓN ESTÁNDAR DE 1.4 MIN. ESTIMAR LA MEDIA REAL DE TIEMPO DE ESPERA EN LA FILA DEL BANCO CON UN COEFICIENTE DE CONFIANZA DEL 99%. INTERPRETAR (2 PUNTOS).

3. EN UNA MUESTRA DE 5 DÍAS SE ANALIZÓ LA PRODUCCIÓN DIARIA DE UN PROCESO QUÍMICO CON RESPECTO A LA CANTIDAD OBTENIDA DE CIERTO PRODUCTO. ESTE ANÁLISIS ARROJÓ UNA VARIANCIA MUESTRAL DE 47.2. DETERMINAR UN INTERVALO DE CONFIANZA DEL 98% PARA LA VARIANCIA DE LA PRODUCCIÓN DIARIA, SUPONIENDO QUE DICHA PRODUCCION SE DISTRIBUYE NORMALMENTE. INTERPRETAR (2 PUNTOS).

4. EN 4 MEDICIONES RESPECTO A LA POROSIDAD DEL COBRE SINTETIZADO, UN LABORATORIO HALLÓ UNA VARIANCIA DE 0.001. UN SEGUNDO LABORATORIO REALIZÓ 5 MEDICIONES SOBRE EL MISMO EXPERIMENTO Y OBTUVO UNA VARIANCIA DE 0.002. ESTIMAR EL COCIENTE DE LAS DOS VARIANCIAS CON UN INTERVALO DEL 95% E INTERPRETAR. SUPONGA POBLACIÓNES NORMALES (2 PUNTOS).

5. CALCULAR LAS PROBABILIDADES QUE SE PIDEN:
A) EL DIÁMETRO DE LAS MONEDAS DE $10.- ESTÁ UNIFORMEMENTE DISTRIBUIDO DE 2.82 CM. A 3.15 CM. ¿QUÉ PROBABILIDAD HAY DE QUE UNA CIERTA MONEDA DE 10$.- TENGA UN DIÁMETRO NO MAYOR A 3 CM?

B) CIERTOS TUBOS FABRICADOS POR UNA COMPAÑIA TIENEN UNA DURACIÓN NORMALMENTE DISTRIBUIDA CON UNA MEDIA 800 HORAS Y UNA DESVIACIÓN ESTÁNDAR DE 40 HORAS. HALLAR LA PROBABILIDAD DE QUE UNO DE ESTOS TUBOS TENGA UNA DURACIÓN MEDIA DE MAS DE 860 HORAS.

C) SE SABE QUE EL TIEMPO ENTE LAS FALLAS QUE SE PRESENTAN EN LOS SATÉLITES DE COMUNICACIONES ESTÁ EXPONENCIALMENTE DISTRIBUIDA CON UN PROMEDIO DE 4 AÑOS. HALLAR LA PROBABILIDAD DE QUE UN SATÉLITE RECIÉN PUESTO EN ORBITA NO FALLE SINO HASTA DENTRO DE 6 AÑOS (2 PUNTOS TODO).

ESTADÍSTICA II. SEGUNDO PARCIAL

1. AJUSTAR LOS DATOS DE LA SIGUIENTE TABLA POR MEDIO DE UNA RECTA DE MÍNIMOS CUADRADOS Y ESTIMAR EL VALOR DE Y SI X = 15. CALCULAR EL COEFICIENTE DE CORRELACIÓN Y TRAZAR EL DIAGRAMA DE DISPERSIÓN (2 PUNTOS).

X	Y
2	1
3	3
5	7
7	11

2. SE MIDIÓ EL TIEMPO QUE 31 ESTUDIANTES OCUPAN EN TRANSPORTE EN UNA ZONA DE LA CIUDAD, Y SE REPITIO EL EXPERIMENTO CON OTROS 31 ESTUDIANTES DE OTRA ZONA. LOS RESULTADOS FUERON

/	ZONA I	ZONA II
n	31	31
$\bar{y}$	81	76
s	5.2	3.4

PRUEBE LA HIPÓTESIS DE QUE EN REALIDAD NO HAY DIFERENCIA EN LOS TIEMPOS PROMEDIO DE TRANSPORTE PARA LAS DOS ZONAS A UN NIVEL DE SIGNIFICACIÓN DEL 5%, Y SUPONIENDO VARIANCIAS POBLACIÓNALES IGUALES (2 PUNTOS).

3. DE ACUERDO CON LOS DATOS DEL PROBLEMA ANTERIOR, ¿ES VÁLIDO EL SUPUESTO DE QUE LAS VARIANCIAS POBLACIÓNALES SON IGUALES A UN NIVEL DE SIGNIFICACIÓN DEL 5%? (2 PUNTOS).

4. EL ADMINISTRADOR DE UN CINE SUPONE QUE SE VENDEN EN PROMEDIO 300 BOLETOS O MÁS AL DIA. SI UNA MUESTRA DE 15 DIAS ARROJÓ UN PROMEDIO DE 270 BOLETOS VENDIDOS POR DIA, CON UNA DESVIACIÓN ESTÁNDAR DE 16, ¿PODRÍA DECIRSE QUE SU APRECIACIÓN ES INCORRECTA? UTILICE UN NIVEL DE SIGNIFICACIÓN DEL 10%, Y SUPONGA QUE LA CANTIDAD DIARIA DE BOLETOS VENDIDOS ES NORMALMENTE DISTRIBUIDA (2 PUNTOS).

5. PARA LOS DATOS DEL PROBLEMA ANTERIOR, ¿PODRÍA AFIRMARSE QUE LA DESVIACIÓN ESTÁNDAR POBLACIÓNAL ES MAYOR A 12 A UN NIVEL DE SIGNIFICACIÓN DEL 1%? RECUERDE QUE SE SUPONE UNA POBLACIÓN NORMAL (2 PUNTOS).

EXAMEN FINAL Y DE RECUPERACIÓN. ESTADÍSTICA II

INSTRUCCIONES.- SI PRESENTAS RECUPERACIÓN DEL PRIMER EXAMEN, RESUELVE LAS PREGUNTAS 1 Y 2. SI PRESENTAS RECUPERACIÓN DEL SEGUNDO EXAMEN, RESUELVE LAS PREGUNTAS 3 Y 4. SI PRESENTAS EXAMEN FINAL, RESUELVE LAS CUATRO PREGUNTAS.

1. LA CANTIDAD DE PALOMITAS VENDIDAS POR FUNCIÓN EN UN CINE ESTA UNIFORMEMENTE DISTRIBUIDA ENTRE 150 Y 500 GRAMOS POR PERSONA. ¿QUÉ PROBABILIDAD HAY DE QUE UNA PERSONA EN LA ULTIMA FUNCIÓN COMPRE MAS DE 350 GRAMOS DE PALOMITAS?

2. LA VARIABILIDAD EN LOS NIVELES DE SONIDO DE 16 PELICULAS DE UNA MISMA PRODUCTORA SE TRADUJO EN UNA DESVIACIÓN ESTÁNDAR DE 10 DECIBELES. ESTIME LA VARIANCIA REAL DE LOS NIVELES DE SONIDO POR MEDIO DE UN INTERVALO DEL 98% E INTERPRETAR. SUPONGA UNA POBLACIÓN NORMAL.

3. ANTERIORMENTE UNA PLANTA QUÍMICA INDUSTRIAL HA PRODUCIDO UN PROMEDIO DE 1100 LIBRAS DE UN PRODUCTO QUÍMICO POR DIA. LOS REGISTROS PARA EL AÑO PASADO, BASADOS EN 260 DÍAS DE TRABAJO, MUESTRAN UN PROMEDIO DE 1060 LIBRAS DIARIAS CON UNA DESVIACIÓN ESTÁNDAR DE 340 LIBRAS DIARIAS. ¿PRESENTAN LOS DATOS SUFICIENTE EVIDENCIA QUE INDIQUE UNA BAJA EN LA PRODUCCION DIARIA PROMEDIO A UN NIVEL DE SIGNIFICACIÓN DEL 5%?

4.SE REALIZÓ UN ESTUDIO PARA DETERMINAR SI EXISTE UNA RELACIÓN LINEAL ENTRE EL PESO ESPECIFICO DE LA MADERA Y LA RESISTENCIA A LA RUPTURA DE VIGAS HECHAS DE MADERA. DIEZ VIGAS SELECCIONADAS AL AZAR, DE LAS MISMAS DIMENSIONES EN LA SECCIÓN TRANSVERSAL, SE SOMETIERON A UN ESFUERZO HASTA LA RUPTURA. EN LA TABLA SE MUESTRAN LA DENSIDAD DE LA MADERA Y LA RESISTENCIA A LA RUPTURA PARA CADA VIGA. A TRAVÉS DEL COEFICIENTE DE CORRELACION DETERMINE SI EL MODELO LINEAL PRESENTA EL AJUSTE MAS ADECUADO, Y, DE SER ASI, REALICE DICHO AJUSTE Y ESTIME LA RESISTENCIA A LA RUPTURA PARA UNA VIGA DE DENSIDAD IGUAL 0.500.

PESO ESPECÍFICO	RESISTENCIA
0.499	11.14
0.558	12.74
0.604	13.13
0.441	11.51
0.550	12.38
0.528	12.60
0.418	11.13
0.480	11.70
0.406	11.02
0.467	11.41

Temario para examen extraordinario de Estadística I

I. Estadística Descriptiva
a) Medidas de tendencia central.- Media, median y moda.
b) Medidas de dispersión.- Variancia y desviación estándar.
c) Gráficos.- Histograma, polígono de frecuencias y ojiva porcentual.
d) Medidas de posición.- Cuantiles.
II. Cálculo Combinatorio
a) Combinación, permutación y ordenación.
III. Probabilidad Axiomática
a) Unión e intersección de eventos. Evento complemento.
b) Probabilidad marginal, conjunta y condicional.
c) Eventos mutuamente excluyentes y eventos independientes.
IV. Función de probabilidad de variable aleatoria discreta
a) Función de distribución de probabilidad de variable aleatoria discreta.
b) Valor esperado (media) y variancia de una variable aleatoria discreta.
V. Distribución Binomial.

EXAMEN EXTRAORDINARIO DE ESTADÍSTICA I

1. HALLAR
A) MEDIA Y MEDIANA. B) VARIANCIA Y DESVIACIÓN ESTÁNDAR.
C) HISTOGRAMA Y POLÍGONO DE FRECUENCIAS. D) PRIMER CUARTIL.

CLASES	F
0 – 10	50
10 – 20	27
20 – 30	62
30 – 40	80
40 – 50	11

2. PARA LOS EVENTOS A Y B TALES QUE P(A) = 0.2, P(B) = 0.4 Y $P(A \cup B) = 0.5$, HALLAR
A) $P(A \cap B)$ B) $P\langle A|B\rangle$ C) $P\langle B|A\rangle$

3. PARA QUE LA SIGUIENTE TABLA REPRESENTE UNA DISTRIBUCIÓN DE PROBABILIDAD DE VARIABLE ALEATORIA, HALLAR k. HALLAR EL VALOR ESPERADO Y LA VARIANCIA.

y	0.1	0.3	0.5	0.7
p(y)	k	10k	k	8k

4. ¿DE CUÁNTAS MANERAS PUEDEN COLOCARSE 8 TORRES EN UN TABLERO DE AJEDREZ DE FORMA TAL QUE NO PUEDAN COMERSE UNAS A OTRAS?

5. LA PROPORCIÓN DE ZURDOS ENTRE LAS PERSONAS SE SUPONE DEL 25%. SI SELECCIONAMOS A 15 PERSONAS AL AZAR, ¿QUÉ PROBABILIDAD HAY DE QUE MENOS DE 6 SEAN ZURDOS?

Temario para examen extraordinario de Estadística II

I. Función de probabilidad de variable aleatoria discreta
a) Binomial.
b) Poisson.
c) Geométrica.
II. Función de distribución de variable aleatoria continua
a) Normal.
b) Aproximación de la Binomial a la Normal.

EXAMEN EXTRAORDINARIO DE ESTADÍSTICA II

1. LA PROBABILIDAD DE CONDICIONAR EXITOSAMENTE A UN RATON DE LABORATORIO ES DEL 85%. SI SE TRABAJA EN CIERTO LABORATORIO CON 16 RATONES, ¿QUÉ PROBABILIDAD HAY DE EXACTAMENTE LA MITAD SE CONDICIONEN CON ÉXITO?

2. EN REFERENCIA AL PROBLEMA ANTERIOR, SI SE CONDICIONA UNA SERIE DE RATONES UNO TRAS OTRO, ¿QUÉ PROBABILIDAD HAY DE QUE EL PRIMER CONDICIONAMIENTO EXITOSO SE DÉ EN EL TERCER RATÓN?

3. ¿QUÉ PROBABILIDAD HAY DE QUE DE 100 RATONES, MAS DE 90 SE CONDICIONEN EXITOSAMENTE, SABIENDO QUE EL PORCENTAJE DE EXITO ES DEL 85%?

4. SI EL PROMEDIO DE CONDICIONAMIENTO EXITOSO ES DE 6 RATONES POR LABORATORIO, ¿QUÉ PROBABILIDAD HAY DE QUE EN UN LABORATORIO EN PARTICULAR SE DEN MÁS DE TRES CONDICIONAMIENTOS EXITOSOS?

5. EN EL HISTORIAL DE CIERTO LABORATORIO SE PUEDE VER QUE DE TODOS LOS EXPERIMENTOS DE CONDICIONAMIENTO REALIZADOS, SE OBTIENE UN PROMEDIO DE 8 ÉXITOS CON UNA DESVIACIÓN ESTÁNDAR DE 4. ¿QUÉ PROBABILIDAD HAY DE QUE EN EL PROXIMO EXPERIMENTO QUE REALICEN HAYA 3 O MENOS CONDICIONAMIENTOS EXITOSOS?

A continuación se presenta una tabla diseñada para facilitar el trabajo con las distribuciones Binomial, Binomial Negativa, Hipergeométrica e Hipergeométrica Multivariada, en la que se enumeran las combinaciones desde 0 en 0 hasta 30 en 30, bajo la idea de que no todos los alumnos cuentan con calculadora que realice directamente la operación, y en consecuencia agotan tiempo realizando los productos y la división. Esta tabla se dio a los alumnos justo antes del segundo examen parcial del primer curso, con la finalidad de que en el curso de Calculo Combinatorio y de Funciones de Distribución de Probabilidad de Variable Aleatoria Discreta trabajaran de manera más directa con esta operación, y desarrollaran la costumbre de depender directamente de resultados ya establecidos.

				TABLAS DE COMBINACIONES								
n	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
r												
0	1	1	1	1	1	1	1	1	1	1	1	1
1		1	2	3	4	5	6	7	8	9	10	11
2			1	3	6	10	15	21	28	36	45	55
3				1	4	10	20	35	56	84	120	165
4					1	5	15	35	70	126	210	330
5						1	6	21	56	126	252	462
6							1	7	28	84	210	462
7								1	8	36	120	330
8									1	9	45	165
9										1	10	55
10											1	11
11												1
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
r												
n	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***

TABLAS DE COMBINACIONES

12	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
1	1	1	1	1	1	1	1	1	1	1	1	1
12	13	14	15	16	17	18	19	20	21	22	23	24
66	78	91	105	120	136	153	171	190	210	231	253	276
220	286	364	455	560	680	816	969	1140	1330	1540	1771	2024
495	715	1001	1365	1820	2380	3060	3876	4845	5985	7315	8855	10626
792	1287	2002	3003	4368	6188	8568	11628	15504	20349	26334	33649	42504
924	1716	3003	5005	8008	12376	18564	27132	38760	54264	74613	100947	134596
792	1716	3432	6435	11440	19448	31824	50388	77520	116280	170544	245157	346104
495	1287	3003	6435	12870	24310	43758	75582	125970	203490	319770	490314	735471
220	715	2002	5005	11440	24310	48620	92378	167960	293930	497420	817190	1307504
66	286	1001	3003	8008	19448	43758	92378	184756	352716	646646	1144066	1961256
12	78	364	1365	4368	12376	31824	75582	167960	352716	705432	1352078	2496144
1	13	91	455	1820	6188	18564	50388	125970	293930	646646	1352078	2704156
	1	14	105	560	2380	8568	27132	77520	203490	497420	1144066	2496144
		1	15	120	680	3060	11628	38760	116280	319770	817190	1961256
			1	16	136	816	3876	15504	54264	170544	490314	1307504
				1	17	153	969	4845	20349	74613	245157	735471
					1	18	171	1140	5985	26334	100947	346104
						1	19	190	1330	7315	33649	134596
							1	20	210	1540	8855	42504
								1	21	231	1771	10626
									1	22	253	2024
										1	23	276
											1	24
												1
12	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*

	TABLAS DE COMBINACIONES					
25	***26***	***27***	***28***	***29***	***30***	***n***
						r
1	1	1	1	1	1	***0***
25	26	27	28	29	30	***1***
300	325	351	378	406	435	***2***
2300	2600	2925	3276	3654	4060	***3***
12650	14950	17550	20475	23751	27405	***4***
53130	65780	80730	98280	118755	142506	***5***
177100	230230	296010	376740	475020	593775	***6***
480700	657800	888030	1184040	1560780	2035800	***7***
1081575	1562275	2220075	3108105	4292145	5852925	***8***
2042975	3124550	4686825	6906900	10015005	14307150	***9***
3268760	5311735	8436285	13123110	20030010	30045015	***10***
4457400	7726160	13037895	21474180	34597290	54627300	***11***
5200300	9657700	17383860	30421755	51895935	86493225	***12***
5200300	10400600	20058300	37442160	67863915	119759850	***13***
4457400	9657700	20058300	40116600	77558760	145422675	***14***
3268760	7726160	17383860	37442160	77558760	155117520	***15***
2042975	5311735	13037895	30421755	67863915	145422675	***16***
1081575	3124550	8436285	21474180	51895935	119759850	***17***
480700	1562275	4686825	13123110	34597290	86493225	***18***
177100	657800	2220075	6906900	20030010	54627300	***19***
53130	230230	888030	3108105	10015005	30045015	***20***
12650	65780	296010	1184040	4292145	14307150	***21***
2300	14950	80730	376740	1560780	5852925	***22***
300	2600	17550	98280	475020	2035800	***23***
25	325	2925	20475	118755	593775	***24***
1	26	351	3276	23751	142506	***25***
	1	27	378	3654	27405	***26***
		1	28	406	4060	***27***
			1	29	435	***28***
				1	30	***29***
					1	***30***
						r
25	***26***	***27***	***28***	***29***	***30***	***n***

Para finalizar con los anexos, se dan a continuación copias de las tablas de las distribuciones Normal estandarizada, t de Student, Ji-cuadrada y F de Snedecor empleadas en los temas Intervalos de Confianza y Pruebas de Hipótesis del curso de Estadística II.

TABLA 4 Áreas bajo la curva normal
Probabilidad normal estándar de cola superior
(para valores negativos de z, las áreas se obtienen por simetría)

	Segundo decimal de z									
z	*.00*	*.01*	*.02*	*.03*	*.04*	*.05*	*.06*	*.07*	*.08*	*.09*
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

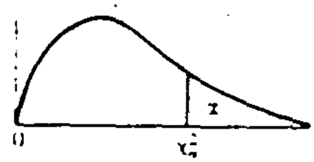
TABLA 5 Puntos porcentuales de las distribuciones t

t_α

$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	g.l.
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15
1.337	1.746	2.120	2.583	2.921	16
1.333	1.740	2.110	2.567	2.898	17
1.330	1.734	2.101	2.552	2.878	18
1.328	1.729	2.093	2.539	[illegible]	19
1.325	1.725	2.086	2.528	2.84[illegible]	20
1.323	1.721	2.080	2.518	2.831	21
1.321	1.717	2.074	2.508	2.819	22
1.319	1.714	2.069	2.500	2.807	23
1.318	1.711	2.064	2.492	2.797	24
1.316	1.708	2.060	2.485	2.787	25
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	1.645	1.960	2.326	2.576	inf.

TABLA 6

Puntos porcentuales de las distribuciones χ^2

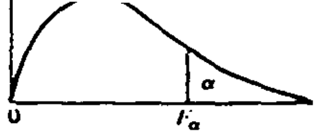


g.l.	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

TABLA 6 (Continuación)

$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$	g.l.
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8454	31.9999	34.2672	16
24.7690	27.5871	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
51.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1539	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100

Puntos porcentuales de las distribuciones F



g.l. del denominador	*α*	*g.l. del numerador* 1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
	.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
	.010	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
	.005	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	[illegible]	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	[illegible]	7.19	6.99	6.84	6.72
	.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51

g.l. del numerador											g.l. del denominador
10	12	15	20	24	30	40	60	120	∞	α	
60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33	.100	1
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	.050	
968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018	.025	
6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	.010	
24224	24426	24630	24836	24940	25044	25148	25253	25359	25465	.005	
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	.100	2
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	.050	
39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50	.025	
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	.010	
199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	.005	
5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13	.100	3
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	.050	
14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	.025	
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	.010	
43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83	.005	
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	.100	4
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	.050	
8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	.025	
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	.010	
20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32	.005	
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10	.100	5
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	.050	
6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	.025	
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	.010	
13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14	.005	
2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	.100	6
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	.050	
5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	.025	
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	.010	
10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88	.005	
2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	.100	7
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	.050	
4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	.025	
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	.010	
8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08	.005	

F_α

g.l. del denominador		*g.l. del numerador*								
	α	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72

F_α

g.l. del numerador											g.l. del denominador
10	12	15	20	24	30	40	60	120	∞	α	
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	.100	8
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	.050	
4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	.025	
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	.010	
7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95	.005	
2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	.100	9
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	.050	
3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	.025	
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	.010	
6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19	.005	
2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06	.100	10
2.98	2.91	2.85	2.74	2.77	2.70	2.66	2.62	2.58	2.54	.050	
3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	.025	
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	.010	
5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64	.005	
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97	.100	11
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	.050	
3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	.025	
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	.010	
5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23	.005	
2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90	.100	12
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	.050	
3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	.025	
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	.010	
5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90	.005	
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85	.100	13
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	.050	
3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	.025	
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	.010	
4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65	.005	
2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80	.100	14
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	.050	
3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	.025	
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	.010	
4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44	.005	

F_α

g.l. del denominador	g.l. del numerador									
	α	1	2	3	4	5	6	7	8	9
15	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88

TABLA 7 (Continuación)

F_{α}

g.l. del numerador: 10	12	15	20	24	30	40	60	120	∞	α	g.l. del denominador
2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76	.100	15
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	.050	
3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.45	2.40	.025	
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	.010	
4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26	.005	
2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72	.100	16
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	.050	
2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	.025	
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	.010	
4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11	.005	
2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69	.100	17
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	.050	
2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	.025	
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	.010	
4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98	.005	
1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66	.100	18
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	.050	
2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	.025	
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	.010	
4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87	.005	
1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63	.100	19
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	.050	
2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	.025	
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	.010	
3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78	.005	
1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61	.100	20
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	.050	
2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	.025	
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	.010	
3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69	.005	
1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59	.100	21
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	.050	
2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04	.025	
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	.010	
3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61	.005	

TABLA 7 (Continuación)

F,

g.l. del denominador	α	1	2	3	4	5	6	7	8	9
22	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.005	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.005	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.005	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.005	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	.005	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52

F_α

g.l. del numerador											g.l. del denominador
10	*12*	*15*	*20*	*24*	*30*	*40*	*60*	*120*	∞	α	
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57	.100	22
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	.050	
2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00	.025	
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	.010	
3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55	.005	
1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55	.100	23
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	.050	
2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97	.025	
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	.010	
3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48	.005	
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53	.100	24
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	.050	
2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	.025	
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	.010	
3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43	.005	
1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	.100	25
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	.050	
2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	.025	
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	.010	
3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38	.005	
1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	.100	26
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	.050	
2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	.025	
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	.010	
3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33	.005	
1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	.100	27
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	.050	
2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	.025	
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	.010	
3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29	.005	
1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	.100	28
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	.050	
2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	.025	
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	.010	
3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25	.005	

TABLA 7 (Continuación)

F_{α}

g.l. del denominador	*g.l. del numerador*									
	α	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
29	.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	.005	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.005	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	.100	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	.050	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
	.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	.100	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

TABLA 7 (Continuación)

F_a

g.l. del numerador											g.l. del denominador
10	12	15	20	24	30	40	60	120	∞	α	
1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	.100	29
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	.050	
2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	.025	
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	.010	
3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21	.005	
1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	.100	30
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	.050	
2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	.025	
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	.010	
3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18	.005	
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	.100	40
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	.050	
2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	.025	
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	.010	
3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93	.005	
1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	.100	60
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	.050	
2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	.025	
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	.010	
2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69	.005	
1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19	.100	120
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	.050	
2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	.025	
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	.010	
2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43	.005	
1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00	.100	∞
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	.050	
2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	.025	
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	.010	
2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00	.005	