



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO.

FACULTAD DE ESTUDIOS SUPERIORES
ACATLÁN.

UNA BREVE INTRODUCCIÓN A LA TEORÍA DE PRUEBAS DE
HIPÓTESIS.

TESINA.

QUE PARA OBTENER EL TÍTULO DE
ACTUARIO.

PRESENTA

JORGE LUIS CRUZ FLORES.

ASESOR: MAHIL HERRERA MALDONADO.

Agosto 2011.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mi madre, mi familia y amigos,
muchas gracias.

Jorge Luis Cruz Flores.

Indice

Introducción.....	1
Capítulo I.- Pruebas paramétricas.....	3
1.1 Elementos de una prueba estadística.....	4
1.2 Prueba más potente.....	10
1.3 Prueba uniformemente más potente.....	14
1.4 Pruebas insegadas.....	22
1.5 Prueba de razón de verosimilitud.....	27
1.6 Pruebas de hipótesis e intervalos de confianza.....	40
Capítulo II.- Pruebas libres de distribución.....	42
2.1 Prueba de razón de verosimilitud y la prueba de bondad de ajuste de Pearson.....	43
2.2 Prueba de Kolmogorov-Smirnov.....	48
2.3 Tablas de contingencia.....	54
2.4 Prueba Binomial.....	60
2.5 Prueba de rangos de Wilcoxon.....	63
Capítulo III.- Regresión lineal simple.....	68
3.1 Definición.....	69
3.2 Método de los minimos cuadrados ordinarios.....	70
3.3 Estimadores lineales insegados de varianza mínima.....	72
3.4 Regresión lineal simple bajo normalidad.....	77
3.5 Inferencia en el modelo de regresión lineal simple.....	86
3.6 Coeficiente de determinación R^2	91
3.7 Predicción y estimación del valor medio de la variable de respuesta.....	93
Apéndice de tablas.....	96
Conclusiones.....	105
Bibliografía.....	107

Introducción.

En el campo de la estadística uno de los problemas principales es el de hacer inferencia acerca de una población basada en una muestra. Se busca inferir sobre el modelo generador de los datos, tal inferencia se centra en un conjunto de medidas que identifican el modelo denominadas parámetros. El problema de hacer inferencia estadística se divide de la siguiente manera:

- Estimación de los parámetros (puntual y por intervalos).
- Pruebas de Hipótesis acerca de los valores de los parámetros.

En la licenciatura en Actuaría estos temas son vistos dentro de las materias de Estadística I y Estadística II respectivamente. El objetivo del presente trabajo es analizar de forma teórica las pruebas de hipótesis estadísticas, para que sirva de apoyo a la materia de Estadística II de la carrera de Actuaría de la FES Acatlán.

Cuando se realizan pruebas de hipótesis sobre los parámetros del modelo generador de los datos de manera que con ellos el modelo queda completamente especificado, la prueba se denomina prueba paramétrica, por ejemplo, los datos siguen una distribución normal con media cero y varianza uno, este tipo de pruebas son analizadas en el capítulo I. En caso contrario, es decir cuando el modelo no queda especificado completamente se denomina prueba libre de distribución o prueba no paramétrica, este tipo de pruebas se analizan en el capítulo II. Finalmente en el capítulo III se analiza el modelo de regresión lineal simple con objeto de cubrir el temario correspondiente a la materia de Estadística II.

Capítulo I.

Pruebas paramétricas

Las técnicas estadísticas que comunmente se emplean en la estadística inferencial y descriptiva fueron desarrolladas por un pequeño grupo de hombres ingleses durante la segunda mitad del siglo XIX y la primera mitad del siglo XX. A Sir Francis Galton (1822-1911) se le atribuye el mérito de introducir los conceptos estadísticos de correlación y regresión a finales del siglo XIX, Galton hizo a la ciencia de la estadística como Biometría, debido a que la observó como una disciplina que empleaba las matemáticas para estudiar problemas tanto en ciencias sociales como biológicas. Karl Pearson (1857-1936) subsecuentemente desarrolló el procedimiento matemático para calcular el coeficiente de correlación, junto con Sir Ronald Fisher, Pearson es visto como el mayor contribuyente a las bases de la estadística moderna, adicionalmente descubrió la distribución ji-cuadrada y la prueba de bondad de ajuste ji-cuadrada. William S. Gosset (1876-1937) quien estudió un breve periodo con Pearson fué quien descubrió la distribución *t*-Student (seudónimo bajo el cual publicó su trabajo) y desarrolló la prueba *t*. Sir Ronald Fisher (1890-1962) desarrolló una mayor cantidad de conceptos fundamentales en el campo de la estadística inferencial que cualquier otro estadístico de la era moderna, las contribuciones más importantes de Fisher fueron: su desarrollo del análisis de la varianza y sus ideas en los temas de diseños experimentales y pruebas de hipótesis, en éste último tema introdujo el concepto de hipótesis nula y nivel de significancia. El término significancia implica que cada uno determina cuando la presencia o ausencia de una diferencia obtenida en un experimento es debido a la posibilidad de la presencia genuina de un efecto experimental. A través de su carrera profesional Fisher y Karl Pearson fueron férreos adversarios, como profesor de Eugenesia en la University College en Londres, Karl Pearson dirigió el laboratorio Galton. Entre los que estudiaron o trabajaron bajo la protección de Pearson se encuentran su hijo Egon y el estadístico polaco Jerzy Neyman. Cuando Karl Pearson se retiró en 1933 Fisher lo sucedió como profesor de Eugenesia, debido a que Fisher no pretendía congeniar con Karl Pearson o sus asociados y con objeto de apaciguar el retiro de Pearson la escuela estableció un departamento de estadística separado eligiendo a Egon Pearson como su dirigente, fue precisamente la colaboración de Egon Pearson con Jerzy Neyman lo que dio como resultado lo que se conoce como el modelo de pruebas de hipótesis de Neyman-Pearson, siendo éste fuente de controversia con la visión de Fisher de las pruebas de hipótesis.

Fisher (1925) establecía que la hipótesis nula debería ser evaluada de la siguiente manera: empleando los datos muestrales, un investigador determina cuando la información contenida en los datos muestrales se desvía o no lo suficiente del valor presentado en la hipótesis nula de modo que la hipótesis nula no sea probable. Fisher no vió necesidad de tener una hipótesis alternativa, el

argumento de Fisher era que el propósito de una prueba estadística era determinar cuando existía o no suficiente evidencia para rechazar la hipótesis nula, dentro de éste contexto introdujo el valor de 0.05 como un nivel de significancia estándar para rechazar una hipótesis nula y sugirió el valor 0.01 como un valor alternativo mas estricto; aunque posteriormente estableció tras una revisión de sus observaciones que los investigadores deberían publicar el nivel de significancia exacto calculado para un conjunto de datos, finalmente estableció que la aceptación de una hipótesis nula no indicaba que esta haya sido "probada" y por lo tanto aceptada, tomó la postura de que cualquier conclusión alcanzada dentro del marco una prueba de hipótesis debería ser sujeta a una reevaluación basada en los resultados de futuras investigaciones. Fisher por lo tanto, rechazó el uso de un conjunto de reglas mecanizadas las cuales obligaran al investigador a tomar una decisión final acerca del estatus de la hipótesis nula basada en el análisis de un solo experimento. En contraste, a finales de 1920 y principios de 1930 Neyman y Pearson propusieron que el investigador no sólo presentara una hipótesis nula sino que además debía especificar una o más hipótesis alternativas contra las cuales la hipótesis nula debería ser evaluada. Dentro de éste contexto introdujeron los conceptos de: error de tipo I, error de tipo II y potencia de una prueba estadística. Rechazaron el uso de niveles de significancia estándar debido a que ellos percibían más esencial que un investigador alcanzara un balance entre los errores tipo I y tipo II.

La teoría de pruebas de hipótesis moderna es resultado de la combinación del trabajo y vision de Sir Ronald Fisher, Jerzy Neyman y Egon Pearson.

1.1 Elementos de una prueba estadística.

El concepto más básico en la teoría de pruebas de hipótesis es la *hipótesis*, dentro del marco de la inferencia estadística una hipótesis puede definirse como una predicción acerca de una sola población o acerca de la relación entre dos o mas poblaciones. En contraste una *prueba de hipótesis* es un procedimiento en el cual se emplea un conjunto de datos muestrales para evaluar una hipótesis. Al emplear el término hipótesis se hace una distinción entre una hipótesis de investigación y una hipótesis estadística. Una hipótesis de investigación es un enunciado general acerca de lo que predice un investigador, por ejemplo: pacientes diagnosticados con depresión que toman un nuevo medicamento antidepresivo durante un periodo de seis meses presentarán menor depresión que pacientes con depresión que toman un placebo durante el mismo periodo. Con el fin de evaluar una hipótesis de investigación, ésta debe describirse dentro del marco de dos hipótesis estadísticas. A través del uso de un formato simbólico, las hipótesis estadísticas resumen la hipótesis de investigación haciendo referencia al parámetro o parámetros poblacionales bajo estudio.

Las dos hipótesis estadísticas son la *hipótesis nula* representada por H_0 y la *hipótesis alternativa* representada por H_1 . La hipótesis nula representa la no

diferencia o el no efecto en el estudio, debido a que la hipótesis de investigación generalmente predice un efecto o una diferencia con respecto a lo que se está estudiando, generalmente es la hipótesis que el investigador espera rechazar. Por otro lado, la hipótesis alternativa representa un enunciado estadístico que indica la presencia de un efecto o una diferencia, por lo que generalmente el investigador espera que ésta sea respaldada por la información de la muestra.

El siguiente paso en un procedimiento de prueba de hipótesis es: una vez recolectados los datos para el estudio, evaluarlos a través de la utilización de una apropiada *prueba estadística*, una prueba estadística es esencialmente una ecuación que describe un conjunto de operaciones matemáticas a realizarse en los datos obtenidos del estudio, el resultado de llevarlas a cabo da como resultado una *estadística de prueba*, este valor es evaluado en referencia a una *distribución muestral*, la cual es una distribución de probabilidad teórica de todos los posibles valores que la estadística de prueba puede asumir suponiendo que la hipótesis nula es correcta y se llevarán a cabo una infinidad de estudios empleando el mismo número de observaciones (tamaño de muestra) utilizadas en el estudio.

Se comenzará definiendo de manera formal cada uno de los conceptos anteriores, recordando antes algunos conceptos adicionales. Una muestra aleatoria representada por X_1, \dots, X_n es un conjunto de variables aleatorias independientes con la misma distribución de probabilidad, se emplea la notación x_1, \dots, x_n para representar una realización de la muestra aleatoria, al conjunto de todas las posibles realizaciones de la muestra aleatoria se le denomina espacio muestral generalmente denotado por Ω . La media muestral representada por \bar{X} y la varianza muestral representada por S^2 se encuentran definidas por

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{y} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

donde n representa el tamaño de la muestra.

Definición 1.1.1 Hipótesis estadística. Una hipótesis estadística (regularmente representada por H) es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias. Si la hipótesis estadística específica completamente la distribución es llamada simple, en caso contrario es llamada compuesta.

Definición 1.1.2 Prueba de una hipótesis estadística. Una prueba de una hipótesis estadística es una regla o procedimiento para decidir cuando rechazar o no la hipótesis estadística con motivo de que el investigador vea respaldada o no su conjetura por los datos observados.

Ejemplo 1.1.3 Sea X_1, \dots, X_n una muestra aleatoria de una distribución normal con media μ y varianza σ^2 conocida. Considere la hipótesis $H : \mu = \mu_0$. Sería razonable rechazar H si el valor de \bar{X} estuviera muy alejado de μ_0 , por lo

que una prueba puede formularse como sigue: rechazar H si $|\bar{X} - \mu_0| > k$, con $k > 0$.

Definición 1.1.4 Prueba no aleatorizada y región crítica. Considere una prueba de una hipótesis estadística H definida de la siguiente manera: rechazar H si y sólo si $(x_1, \dots, x_n) \in C$, donde C es un subconjunto del espacio muestral (es decir, $C \subset \Omega$). Entonces la prueba es llamada no aleatorizada y C es la región crítica de la prueba.

Obsérvese que una prueba no aleatorizada de una hipótesis estadística H es una descomposición del espacio muestral Ω en dos subconjuntos C y C^c , donde si $(x_1, \dots, x_n) \in C$ entonces H es rechazada. Por lo que una prueba no aleatorizada esta completamente especificada por su correspondiente región crítica

Ejemplo 1.1.5 La prueba expuesta en (1.1.3) es un ejemplo de una prueba no aleatorizada cuya región crítica esta dada por

$$C_k := \{(x_1, \dots, x_n) : |\bar{x} - \mu_0| > k, k > 0\}$$

Definición 1.1.6 Prueba aleatorizada. Una prueba de una hipótesis estadística H es una prueba aleatorizada si la decisión de rechazar o no la hipótesis estadística esta basada en una función $\varphi(\cdot, \dots, \cdot) : \Omega \rightarrow [0, 1]$ definida por:

$$\varphi(x_1, \dots, x_n) = P[\text{Rechazar } H \mid \text{la observación } (x_1, \dots, x_n)]$$

Es decir, la probabilidad condicional de rechazar la hipótesis estadística dado que la observación de la muestra fue x_1, \dots, x_n . La función $\varphi(\cdot, \dots, \cdot)$ es llamada la función crítica de la prueba.

Ejemplo 1.1.7 Sea X_1, \dots, X_{2n} una muestra aleatoria de $f(x; \theta) = \theta^x (1 - \theta)^{1-x}$ $x = 0, 1$. Suponga que se desea probar la hipótesis $H : \theta < \frac{1}{2}$. Una prueba se plantea como sigue: rechazar H si $\sum_{i=0}^{2n} x_i > n$, no rechazar si $\sum_{i=0}^{2n} x_i < n$ y para los puntos muestrales tales que $\sum_{i=0}^{2n} x_i = n$ asignar a cualquiera de las acciones una probabilidad de $\frac{1}{2}$ (la decisión entre rechazar o no H podría entonces apoyarse en el lanzamiento de una moneda legal). Esta prueba particiona el espacio muestral en tres conjuntos y la función crítica para esta prueba estaría dada por:

$$\varphi(x_1, \dots, x_{2n}) := \begin{cases} 1 & \text{si } \sum_{i=0}^{2n} x_i > n \\ \frac{1}{2} & \text{si } \sum_{i=0}^{2n} x_i = n \\ 0 & \text{si } \sum_{i=0}^{2n} x_i < n \end{cases}$$

Se ha mencionado que una prueba no aleatorizada queda especificada por su región crítica. De igual manera una prueba aleatorizada queda completamente especificada por su función crítica. De hecho, cualquier función $\varphi(\cdot, \dots, \cdot)$ definida en el espacio muestral Ω y codominio el intervalo $[0, 1]$ es una posible función crítica y define una prueba aleatorizada.

Se observa que si una prueba tiene una función crítica definida por:

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } (x_1, \dots, x_n) \in C \\ 0 & \text{en otro caso} \end{cases}$$

Donde C es un subconjunto del espacio muestral Ω , entonces se trata de una prueba no aleatorizada con región crítica C . Es decir, la función crítica de una prueba no aleatorizada es igual a la función indicadora de la región crítica.

En muchos problemas de pruebas de hipótesis, dos hipótesis son discutidas: la primera, la hipótesis a probar es llamada la hipótesis nula denotada por H_0 y la segunda es llamada la hipótesis alternativa denotada por H_1 . La postura en el contexto de las pruebas de hipótesis es: si la hipótesis nula es falsa entonces la hipótesis alternativa se acepta como cierta y viceversa.

Definición 1.1.8 Tipos de error y tamaño del error. Rechazar la hipótesis nula H_0 cuando en realidad es verdadera es llamado *error de tipo I* y aceptar H_0 cuando en realidad es falsa es llamado *error de tipo II*. El tamaño del error de tipo I denotado por α se define como la probabilidad de cometer un error de tipo I, de manera similar el tamaño del error de tipo II denotado por β se define como la probabilidad de cometer un error de tipo II.

De acuerdo a la definición (1.1.8) la probabilidad de aceptar una hipótesis nula verdadera es $1 - \alpha$ y la probabilidad de rechazar una hipótesis nula falsa es $1 - \beta$. Lo anterior puede resumirse en el siguiente cuadro:

	Decisión	
	Aceptar H_0	Rechazar H_0
H_0 verdadera	Acción correcta ($1 - \alpha$)	Error de tipo I (α)
H_0 falsa	Error de tipo II (β)	Acción correcta ($1 - \beta$)

Si la distribución de la cual fue obtenida la muestra tiene parámetro θ , donde $\theta \in \Theta$ (espacio parametral). Entonces asociamos con cualquier prueba su función de potencia cuya definición se presenta a continuación.

Definición 1.1.9 Función de potencia. Considere una prueba de la hipótesis nula H_0 . La función de potencia de la prueba denotada por $\pi(\theta)$ se define como la probabilidad de que H_0 sea rechazada cuando la distribución verdadera de donde se obtuvo la muestra tiene parámetro θ .

La definición de la función de potencia nos dice: $\pi(\theta) = P_\theta[\text{rechazar } H_0]$, donde θ es el valor verdadero del parámetro. Si se considera una prueba no aleatorizada entonces $\pi(\theta) = P_\theta[(X_1, \dots, X_n) \in C]$, donde C es la región crítica asociada con la prueba. Si se considera una prueba aleatorizada con función crítica $\varphi(\cdot, \dots, \cdot)$ entonces:

$$\begin{aligned}\pi(\theta) &= P_\theta[\text{rechazar } H_0] \\ &= E_\theta[\varphi(X_1, \dots, X_n)]\end{aligned}$$

La función de potencia juega el mismo papel en pruebas de hipótesis que el que juega el error cuadrático medio en la estimación. Usualmente será nuestra regla para valorar que tan buena es una prueba o para realizar una comparación entre dos pruebas.

Definición 1.1.10 Tamaño de la prueba. Considere una prueba de la hipótesis $H_0 : \theta \in \Theta_0$; donde $\Theta_0 \subset \Theta$. El tamaño de la prueba de H_0 se define como $\sup_{\theta \in \Theta_0} \pi(\theta)$. El tamaño de una prueba no aleatorizada se refiere al tamaño de la región crítica C .

Ejemplo 1.1.11 Sea X_1, \dots, X_n una muestra aleatoria de una distribución uniforme en $[0, \theta]$ y $Y_{(n)} := \max\{X_1, \dots, X_n\}$. Para el contraste de hipótesis $H_0 : \theta \geq 2$ vs $H_1 : \theta < 2$ considérese la prueba que rechaza H_0 si $Y_{(n)} \leq 1.5$. La función de potencia de esta prueba es:

$$\pi(\theta) = P_\theta(Y_{(n)} \leq 1.5)$$

Nótese que si $\theta \leq 1.5$ entonces $\pi(\theta) = 1$ mientras que si $\theta > 1.5$, $\pi(\theta) = \left(\frac{1.5}{\theta}\right)^n$. Se observa que la función de potencia es decreciente, por lo que el tamaño de la prueba se obtiene de

$$\sup_{\theta \in \Theta_0} \pi(\theta) = \sup_{\theta \geq 2} \pi(\theta) = \pi(2) = \left(\frac{1.5}{2}\right)^n$$

Para $n = 10$ la gráfica de la función de potencia se presenta en la figura (1)

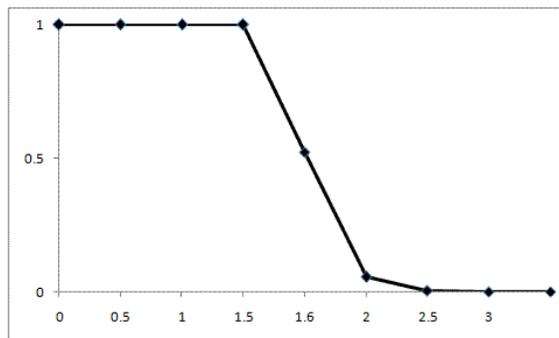


figura (1). $\pi(\theta)=P_{\theta}(Y_{(10)}\leq 1.5)$

En el estudio de la estimación puntual se tiene que bajo ciertas consideraciones podemos restringir nuestra atención a estimadores que sean función de estadísticas suficientes. Podemos hacer la misma consideración para las pruebas de hipótesis cuando la función de potencia es usada como base para comparar pruebas, como lo muestra el siguiente resultado.

Teorema 1.1.12 Si X_1, \dots, X_n es una muestra aleatoria de $f(x; \theta)$ donde $\theta \in \Theta$ y $S_1 = s_1(x_1, \dots, x_n), \dots, S_r = s_r(x_1, \dots, x_n)$ es un conjunto de estadísticas suficientes. Entonces para toda prueba con función crítica $\varphi(\cdot)$, existe una prueba con función crítica ψ que depende solamente del conjunto de las estadísticas suficientes que satisface:

$$\pi_{\psi}(\theta) = \pi_{\varphi}(\theta) \quad \text{para todo } \theta \in \Theta.$$

Demostración: Definimos

$$\psi(x_1, \dots, x_n) = E(\varphi(x_1, \dots, x_n) \mid S_1 = s_1, \dots, S_r = s_r)$$

entonces ψ es una función crítica. Por otro lado

$$\begin{aligned} \pi_{\psi}(\theta) &= E_{\theta}[\psi(x_1, \dots, x_n)] \\ &= E_{\theta}[E(\varphi(x_1, \dots, x_n) \mid S_1 = s_1, \dots, S_r = s_r)] \\ &= E_{\theta}[\varphi(x_1, \dots, x_n)] = \pi_{\varphi}(\theta) \quad \blacksquare \end{aligned}$$

El teorema muestra que dada cualquier prueba es posible encontrar otra prueba basada en un conjunto de estadísticas suficientes de tal manera que la función de potencia de ambas es idéntica.

El problema de las pruebas de hipótesis es semejante al de la estimación en el sentido de que primero es necesario encontrar un método para obtener pruebas y posteriormente establecer un criterio de comparación entre pruebas.

1.2 Prueba más potente.

Primero se considera el problema de probar una hipótesis nula simple contra una hipótesis alternativa simple. Este caso no tiene realmente muchas aplicaciones estadísticas pero sirve como introducción a la teoría de las pruebas de hipótesis. Asumimos que tenemos una muestra aleatoria X_1, \dots, X_n que proviene de una de dos densidades específicas $f_0(x)$ o $f_1(x)$ y se pretende realizar el contraste $H_0 : X \sim f_0(\cdot)$ vs $H_1 : X \sim f_1(\cdot)$, el cual suele expresarse por $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Una buena prueba es aquella para la cual $\pi(\theta_0) = P[\text{rechazar } H_0 \mid H_0 \text{ es verdadera}]$ es pequeña, muy próxima a cero y $\pi(\theta_1) = P[\text{rechazar } H_0 \mid H_0 \text{ es falsa}]$ llamada la **potencia de la prueba** sea grande (lo cual minimiza el error tipo II), muy próxima a 1. Podríamos usar razonablemente los valores de $\pi(\theta_0) = \alpha$ y $1 - \pi(\theta_1) = \beta$ a fin de obtener un criterio para definir la mejor prueba. El método más frecuente se basa en fijar el tamaño del error de tipo I y minimizar entonces el error de tipo II. La justificación para fijar α (usualmente pequeño: 0.05, 0.01 o 0.001) parece surgir de aquellas situaciones donde se formulan las dos hipótesis de tal manera que un tipo de error es más serio que el otro. Las hipótesis se establecen considerando más grave el error de tipo I, es por ello que se busca que éste sea pequeño.

Definición 1.2.1 Prueba más potente. Una prueba de $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ se dice que es la prueba más potente de tamaño α ($0 < \alpha < 1$) si y sólo si:

- i) $\pi^*(\theta_0) = \alpha$.
- ii) $\pi^*(\theta_1) \geq \pi(\theta_1)$ para cualquier otra prueba tal que $\pi(\theta_0) \leq \alpha$.

La definición anterior indica que una prueba es la más potente de tamaño α si el tamaño del error de tipo I es igual a α y la potencia de la prueba es mayor que la de cualquier otra prueba con tamaño menor o igual a α .

Con respecto a encontrar la prueba más potente se presenta el siguiente resultado.

TEOREMA 1.2.2 Lema fundamental de Neyman-Pearson.

Sean F_0 y F_1 funciones de distribución con funciones de densidad f_0 y f_1 respectivamente con respecto a una medida μ .

a) Para probar $H_0 : X \sim f_0(\cdot)$ vs $H_1 : X \sim f_1(\cdot)$ existe una prueba φ y una constante k tal que

$$E_0(\varphi(X)) = \alpha \quad 0 \leq \alpha \leq 1 \quad (1.1)$$

y

$$\varphi(x) := \begin{cases} 1 & \text{si } f_1(x) > kf_0(x) \\ 0 & \text{si } f_1(x) < kf_0(x) \end{cases} \quad (1.2)$$

b) Si existe una prueba que satisface (1.1) y (1.2) para alguna k entonces es la prueba más potente de tamaño α para probar H_0 vs H_1 .

c) Si φ es la prueba más potente de tamaño α para probar H_0 vs H_1 . Entonces para alguna k se satisfacen (1.2) casi en cualquier parte con respecto a μ y (1.1) a menos que exista una prueba de tamaño menor que α y potencia 1.

Demostración: Para $\alpha = 0$ y $\alpha = 1$ puede considerarse $k = \infty$ y $k = 0$ respectivamente en (1.2) y el producto $0 \cdot \infty$ es interpretado como 0. Por lo que la demostración considera $0 < \alpha < 1$.

a) Sea $\alpha(c) = P_0\{f_1(X) > cf_0(X)\}$. Debido a que la probabilidad es calculada bajo f_0 la desigualdad anterior puede ser considerada solo sobre el conjunto donde $f_0(x) > 0$. Por lo tanto $\alpha(c)$ representa la probabilidad de que la v.a. $\frac{f_1(X)}{f_0(X)}$ sea mayor que c , $\alpha(c)$ es no creciente y continua por la derecha, $\alpha(-\infty) = 1$, $\alpha(\infty) = 0$ y $\alpha(c-0) - \alpha(c) = P_0\left\{\frac{f_1(X)}{f_0(X)} = c\right\}$. Dada $0 < \alpha < 1$ sea k tal que $\alpha(k) \leq \alpha \leq \alpha(k-0)$ y considere la prueba definida por

$$\varphi(x) = \begin{cases} 1 & \text{si } f_1(x) > kf_0(x) \\ \frac{\alpha - \alpha(k)}{\alpha(k-0) - \alpha(k)} & \text{si } f_1(x) = kf_0(x) \\ 0 & \text{si } f_1(x) < kf_0(x) \end{cases}$$

donde el término de en medio no es significativo si $\alpha(k) = \alpha(k-0)$, ya que $P_0\{f_1(X) = cf_0(X)\} = 0$.

El tamaño de esta prueba es

$$E_0(\varphi(X)) = P_0\left\{\frac{f_1(X)}{f_0(X)} > k\right\} + \frac{\alpha - \alpha(k)}{\alpha(k-0) - \alpha(k)} P_0\left\{\frac{f_1(X)}{f_0(X)} = k\right\} = \alpha$$

b) Suponemos que φ es una prueba que satisface (1.1) y (1.2) y que φ^* es cualquier otra prueba con $E_0(\varphi^*(X)) \leq \alpha$. Denotemos por S^+ y S^- al

conjunto de puntos muestrales donde $\varphi(x) - \varphi^*(x) > 0$ y $\varphi(x) - \varphi^*(x) < 0$ respectivamente.

Si $x \in S^+$, $\varphi(x)$ debe ser positivo y $f_1(x) \geq kf_0(x)$. De igual manera $f_1(x) \leq kf_0(x)$ si $x \in S^-$. Con lo que se tiene

$$\int (\varphi - \varphi^*)(f_1(x) - kf_0(x)) d\mu = \int_{S^+ \cup S^-} (\varphi - \varphi^*)(f_1(x) - kf_0(x)) d\mu \geq 0$$

Por lo tanto la diferencia entre la potencia de φ y φ^* satisface

$$\int (\varphi - \varphi^*) f_1(x) d\mu \geq k \int (\varphi - \varphi^*) f_0(x) d\mu \geq 0$$

por lo que φ es entonces la prueba más potente de tamaño α .

c) Sea φ^* la prueba más potente de tamaño α para probar H_0 vs H_1 y sea que φ satisfaga (1.1) y (1.2). Denotemos por S la intersección del conjunto $S^+ \cup S^-$ (donde φ y φ^* difieren) con el conjunto $\{x : f_1(x) \neq kf_0(x)\}$, suponemos que $\mu(S) > 0$. Como $(\varphi - \varphi^*)(f_1(x) - kf_0(x))$ es positivo en S se tiene

$$\int_{S^+ \cup S^-} (\varphi - \varphi^*)(f_1(x) - kf_0(x)) d\mu = \int_S (\varphi - \varphi^*)(f_1(x) - kf_0(x)) d\mu > 0$$

lo cual contradice que φ^* sea la prueba más potente de tamaño α , por lo que $\mu(S) = 0$.

Si φ^* fuera de tamaño menor a α y potencia menor que 1, es posible incluir en la región de rechazo puntos adicionales o proporciones de puntos adicionales para incrementar la potencia de la prueba hasta que la potencia sea igual a 1 o el tamaño sea igual a α . ■

La demostración anterior considera f una función de densidad de probabilidad. El caso en que f es una función masa de probabilidad puede tratarse intercambiando el símbolo de integral por el de suma considerando una pequeña dificultad que puede presentarse. Puede ocurrir que no sea posible obtener para cada $\alpha \in (0, 1)$ que

$$P_0(\text{rechazar } H_0) = \sum_{x \in C} f_0(x) = \alpha$$

debido a que si un cierto punto es incluido en C el valor de α no es obtenido (Ejemplo 1.2.4).

La demostración de la parte c) muestra que la prueba más potente esta determinada de manera única por (1.1) y (1.2) excepto en el conjunto donde $f_1(x) = kf_0(x)$. En este conjunto la prueba puede definirse de manera arbitraria.

En la práctica la prueba expresada por (1.2) puede ser manipulada para obtener una desigualdad equivalente mas sencilla y la prueba es entonces expresada en términos de la nueva desigualdad. Para ayudar a clarificar lo anterior se presentan los siguientes ejemplos:

Ejemplo 1.2.3 Sea X_1, \dots, X_n una muestra aleatoria de

$$f(x : \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$$

donde $\theta = \theta_0$ o $\theta = \theta_1$. θ_0 y θ_1 son valores conocidos con $\theta_1 > \theta_0 > 0$. Se desea probar el contraste $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, de acuerdo al lema de N-P la prueba más potente tiene la forma: Rechazar H_0 si $\frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} > k$ o si $\left(\frac{\theta_1}{\theta_0}\right)^n \exp\left[-(\theta_1 - \theta_0) \sum_{i=1}^n x_i\right] > k$ lo cual es equivalente a

$$\sum_{i=1}^n x_i < \frac{1}{\theta_0 - \theta_1} \ln\left(\left(\frac{\theta_0}{\theta_1}\right)^n k\right) = k^*$$

donde k^* es precisamente una constante. La desigualdad $\frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} > k$ se ha simplificado y expresado por la desigualdad equivalente $\sum_{i=1}^n x_i < k^*$. Por otro lado $\sum_{i=1}^n X_i$ tiene una distribución gamma con parámetros n y θ , por lo que k^* puede determinarse de la condición (1.1)

$$\begin{aligned} \alpha &= P_{\theta_0}(\text{rechazar } H_0) = P_{\theta_0}\left(\sum_{i=1}^n x_i < k^*\right) \\ &= \int_0^{k^*} \frac{1}{\Gamma(n)} \theta_0^n x^{n-1} e^{-x\theta_0} dx \end{aligned}$$

Ejemplo 1.2.4 Sea X_1, \dots, X_n una muestra aleatoria de

$$f(x : \theta) = \theta^x (1 - \theta)^{1-x} I_{\{0,1\}}(x)$$

donde $\theta = \theta_0$ o $\theta = \theta_1$, $0 < \theta_0 < \theta_1 < 1$. Se desea probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Ahora $\frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} > k$ si y sólo si

$$\frac{\theta_1^{\sum x_i} (1 - \theta_1)^{n - \sum x_i}}{\theta_0^{\sum x_i} (1 - \theta_0)^{n - \sum x_i}} > k \Leftrightarrow \left[\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}\right]^{\sum x_i} > k \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^n = k'$$

que es equivalente a

$$\sum x_i > \frac{k'}{\ln\left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}\right)} = k^*$$

(Obsérvese que $\ln\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right) > 0$). Por lo que la prueba más potente rechaza H_0 para valores grandes de $\sum_{i=1}^n x_i$. Considere $\theta_0 = \frac{1}{3}$, $\theta_1 = \frac{2}{3}$ y $n = 10$. De la condición (1.1) k^* es tal que

$$\alpha = P_{\theta_0} \left(\sum_{i=1}^{10} X_i > k^* \right) = \sum_{i=k^*+1}^{10} \binom{10}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^i$$

Si $\alpha = 0.0197$ entonces $k^* = 6$ mientras que si $\alpha = 0.0766$ $k^* = 5$. Con lo anterior se observa que para $\alpha = 0.05$ no existe una prueba no aleatorizada y una constante k^* de la forma dada en el lema de Neyman-Pearson. En estos casos es usual cambiar el tamaño de la prueba a algún valor α para el cual pueda determinarse una prueba de la forma dada por el lemma de N-P. Sin embargo, observando la demostración del lema de N-P debe notarse que una prueba más potente de tamaño α existe (en este caso será una prueba aleatorizada). En el ejemplo considerando $\alpha = 0.05$ la prueba más potente esta dada por la función crítica

$$\varphi(x_1, \dots, x_n) := \begin{cases} 1 & \text{si } \sum_{i=1}^{10} x_i > 6 \\ \frac{0.05-0.0197}{0.0766-0.0197} & \text{si } \sum_{i=1}^{10} x_i = 6 \\ 0 & \text{si } \sum_{i=1}^{10} x_i \leq 5 \end{cases}$$

Corolario 1.2.5 Sea que $\pi(\theta_1)$ denote la potencia de la prueba más potente de tamaño $\alpha \in (0, 1)$ para probar $H_0 : X \sim f_0(\cdot)$ vs $H_1 : X \sim f_1(\cdot)$. Entonces $\alpha < \pi(\theta_1)$ a menos que $F_0 = F_1$.

Demostración: Sea φ la prueba que rechaza H_0 con probabilidad α para cualquier observación. Por lo tanto se tiene

$$E_0(\varphi) = \alpha = E_1(\varphi) \quad \alpha \leq \pi(\theta_1)$$

Si $\alpha = \pi(\theta_1)$, entonces φ es la prueba más potente y debe satisfacer (1.1) y (1.2) del lema de N-P lo que sería una contradicción a menos que $F_0 = F_1$. ■

1.3 Prueba uniformemente más potente.

Ahora se considera el problema de probar hipótesis más generales las cuales son llamadas hipótesis compuestas. Asumimos que tenemos una muestra aleatoria de $f(x; \theta)$, $\theta \in \Theta$ y queremos probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, donde $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ con Θ_0 y Θ_1 disjuntos, regularmente $\Theta_1 = \Theta - \Theta_0$.

Definición 1.3.1 Prueba uniformemente más potente. Una prueba de $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta - \Theta_0$ se dice que es la prueba uniformemente más potente (UMP) de tamaño α si y sólo si:

- i) $\sup_{\theta \in \Theta_0} [\pi^*(\theta)] = \alpha$
- ii) $\pi^*(\theta) \geq \pi(\theta)$ para todo $\theta \in \Theta - \Theta_0$ y para toda prueba con tamaño menor o igual a α .

Una prueba es uniformemente más potente de tamaño α si tiene tamaño α y si entre todas las pruebas de tamaño menor o igual a α su potencia es mayor o igual para todos los posibles valores alternativos de θ . El adjetivo "uniformemente" hace referencia a que la prueba es más potente sobre todos los valores alternativos de θ . Cabe señalar que no siempre existe una prueba uniformemente más potente. Sin embargo existe un conjunto de familias de distribuciones que poseen la característica de poder determinar una prueba UMP para un contraste de hipótesis compuestas. Antes de exponer el resultado que muestra como obtener esta prueba se presenta la siguiente definición.

Definición 1.3.2 Razón de verosimilitud monótona.

Sea $\theta \in \mathbb{R}$ y $\{f(x; \theta) : \theta \in \Theta\}$ una familia de densidades o masas de probabilidad, se dice que $\{f(x; \theta) : \theta \in \Theta\}$ tiene una razón de verosimilitud monótona si existe una estadística $T = t(\mathbf{X})$ tal que para cada $\theta_0 < \theta_1$ las distribuciones F_0 y F_1 son distintas y la razón:

$$r(x) := \frac{f(x; \theta_1)}{f(x; \theta_0)}$$

es una función no decreciente (no creciente) de T .

Teorema 1.3.3 Sean $\theta \in \mathbb{R}$ y $\{f(x; \theta) : \theta \in \Theta\}$ una familia de densidades con una razón de verosimilitud monótona en la estadística $T = t(\mathbf{X})$. Entonces para probar $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ existe una prueba uniformemente más potente dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } t(\mathbf{x}) > k \\ \gamma & \text{si } t(\mathbf{x}) = k \\ 0 & \text{si } t(\mathbf{x}) < k \end{cases}$$

donde k y γ se determinan por

$$\alpha = E_{\theta_0}(\varphi(\mathbf{X})) \tag{1.3}$$

con función de potencia

$$\pi(\theta) = E_{\theta}(\varphi(\mathbf{X})) \tag{1.4}$$

la cual es una función estrictamente creciente de θ para todos los valores tales que $0 < \pi(\theta) < 1$ y para cualquier $\theta < \theta_0$ la prueba minimiza $\pi(\theta)$ (la

probabilidad de un error de tipo I) de entre todas las pruebas que satisfacen (1.3).

Demostración: Consideramos primero un contraste de hipótesis simples $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1 > \theta_0$. Los principales puntos muestrales que hacen rechazar H_0 son aquellos para los que

$$r(\mathbf{x}) := \frac{f(\mathbf{x}; \theta_1)}{f(\mathbf{x}; \theta_0)}$$

es suficientemente grande. Si $t(\mathbf{x}) < t(\mathbf{x}')$ entonces $r(\mathbf{x}) \leq r(\mathbf{x}')$ y \mathbf{x}' es tan deseable como \mathbf{x} . Por lo que la prueba que rechaza H_0 para valores grandes de $t(\mathbf{x})$ es más potente. De la misma manera que en el teorema 1.2.2 existen k y γ que satisfacen (1.3) y (1.4), la prueba resultante es además la prueba más potente de tamaño $\alpha' = \pi(\theta')$ para probar

$$H_0 : \theta = \theta' \text{ vs } H_1 : \theta = \theta'' > \theta'$$

Además $\alpha' < \pi(\theta'')$ por el corolario 1.2.5, por lo que para toda $\theta' < \theta''$ se tiene $\pi(\theta') < \pi(\theta'')$ y $\pi(\theta)$ es estrictamente creciente para todo θ tal que $0 < \pi(\theta) < 1$.

Debido a que $\pi(\theta)$ es por tanto no decreciente la prueba satisface

$$E_\theta(\varphi(\mathbf{X})) \leq \alpha \quad \text{para } \theta \leq \theta_0 \quad (1.5)$$

La clase de las pruebas que satisfacen (1.5) esta contenida en la clase de las que satisfacen $E_\theta(\varphi(X)) \leq \alpha$. Dado que la prueba dada maximiza $\pi(\theta_1)$ dentro de la clase más extensa, también lo hace sujeto a (1.5), debido a que ésto es independiente de la elección particular de $\theta_1 > \theta_0$, esta es la prueba uniformemente más potente contra la alternativa H_1 .

La demostración de que para cualquier $\theta < \theta_0$ la prueba minimiza $\pi(\theta)$ para un contraste de hipótesis simples se obtiene aplicando el teorema 1.2.2 con las desigualdades en el sentido inverso.

Se observa que si $r(\mathbf{x})$ es no creciente en $t(\mathbf{x})$ las desigualdades en la prueba son en el sentido inverso

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } t(\mathbf{x}) < k \\ \gamma & \text{si } t(\mathbf{x}) = k \\ 0 & \text{si } t(\mathbf{x}) > k \end{cases} \blacksquare$$

Intercambiando las desigualdades presentes en el teorema anterior se obtiene de manera sencilla la solución para la prueba uniformemente más potente del contraste de hipótesis $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$.

Esto es, si $r(\mathbf{x})$ es no decreciente en $t(\mathbf{x})$ la prueba uniformemente más potente para $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$ esta dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } t(\mathbf{x}) < k \\ \gamma & \text{si } t(\mathbf{x}) = k \\ 0 & \text{si } t(\mathbf{x}) > k \end{cases}$$

donde k y γ se determinan por

$$\alpha = E_{\theta_0}(\varphi(\mathbf{X}))$$

Considerando en este caso que para un contraste de hipótesis simples $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1 < \theta_0$, $r(\mathbf{x}) \leq r(\mathbf{x}')$ si $t(\mathbf{x}) > t(\mathbf{x}')$, por lo que la prueba que rechaza H_0 para valores pequeños de $t(\mathbf{x})$ es más potente. Si $r(\mathbf{x})$ es no creciente en $t(\mathbf{x})$ las desigualdades en la prueba se invierten.

El tipo de contraste de hipótesis expuesto en el teorema 1.3.3 se denomina unilateral. Un contraste de hipótesis de la forma $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ se denomina contraste bilateral y será analizado más adelante.

Corolario 1.3.4 Sea $\theta \in \mathbb{R}$ y X variable aleatoria con función de densidad

$$f(x; \theta) = a(\theta) e^{c(\theta)d(x)} b(x)$$

donde $c(\theta)$ es estrictamente monótona. Entonces existe una prueba uniformemente más potente para $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Si $c(\theta)$ es creciente la prueba esta dada por:

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } d(\mathbf{x}) > k \\ \gamma & \text{si } d(\mathbf{x}) = k \\ 0 & \text{si } d(\mathbf{x}) < k \end{cases}$$

donde k y γ se determinan por $E_{\theta_0}(\varphi(\mathbf{X})) = \alpha$. Si $c(\theta)$ es decreciente las desigualdades son en el sentido inverso.

Demostración: Para $\theta' < \theta''$ se tiene

$$r(\mathbf{x}) = \frac{f(\mathbf{x}; \theta'')}{f(\mathbf{x}; \theta')} = \frac{a(\theta'')}{a(\theta')} e^{d(\mathbf{x})[c(\theta'') - c(\theta')]}$$

Por lo que si $d(\mathbf{x}) < d(\mathbf{x}')$ entonces $r(\mathbf{x}) \leq r(\mathbf{x}')$ o $r(\mathbf{x}) \geq r(\mathbf{x}')$ dependiendo de si $c(\theta)$ es creciente o bien es decreciente respectivamente. Con lo que se cumple la propiedad de la razón de verosimilitud monótona y el resultado se sigue del teorema 1.3.3 ■

Ejemplo 1.3.5 Sea X_1, \dots, X_n una muestra aleatoria de una distribución Bernoulli con parámetro θ ($0 < \theta < 1$) desconocido. Se observa que para cualesquiera dos valores θ' y θ'' con $0 < \theta' < \theta'' < 1$ se tiene

$$\frac{f(\mathbf{x}; \theta'')}{f(\mathbf{x}; \theta')} = \left[\frac{\theta''(1-\theta')}{\theta'(1-\theta'')} \right]^{\sum x_i} \left(\frac{1-\theta''}{1-\theta'} \right)^n$$

Esta ecuación muestra que la razón $\frac{f(\mathbf{x}; \theta'')}{f(\mathbf{x}; \theta')}$ depende de la muestra sólo a través de la estadística $T = \sum_{i=1}^n x_i$ y que en función de T la razón es creciente.

Por lo que la familia de funciones masa de probabilidad Bernoulli tiene una razón de verosimilitud monótona en la estadística $T = \sum_{i=1}^n x_i$. Ahora si se desea probar el contraste de hipótesis $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ el teorema 1.3.3 muestra que la prueba UMP viene dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } \sum_{i=1}^n x_i > k \\ \gamma & \text{si } \sum_{i=1}^n x_i = k \\ 0 & \text{si } \sum_{i=1}^n x_i < k \end{cases} \quad (1.6)$$

donde k y γ se determinan por $\alpha = E_{\theta_0}(\varphi(\mathbf{X}))$

Concentrándonos en una prueba no aleatorizada la región crítica de tamaño α tiene la forma

$$C := \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n x_i > k \right\}$$

donde k se determina por la condición $\alpha = P_{\theta_0} \left(T = \sum_{i=1}^n x_i > k \right)$. Cuando $\theta = \theta_0$, $T \sim \text{Bin}(n, \theta_0)$, por lo que T toma solo un número finito de valores diferentes, en consecuencia sólo hay un número finito de valores posibles para α . Si el valor de n es suficientemente grande con ayuda del teorema central del límite y la función de potencia

$$\begin{aligned} \pi(\theta) &= P_{\theta} \left(\sum_{i=1}^n x_i > k \right) = P_{\theta} \left(\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i > k^* \right) \\ &= P_{\theta} \left(\frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} > \frac{k^* - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \right) = P_{\theta} \left(z > \frac{k^* - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \right) \\ &= 1 - P_{\theta} \left(z \leq \frac{k^* - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \right) = 1 - \Phi \left(\frac{k^* - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \right) \end{aligned}$$

donde k^* puede determinarse de

$$\alpha = \pi(\theta_0) = 1 - \Phi \left(\frac{k^* - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \right)$$

De donde se obtiene

$$\frac{k^* - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} = Z_{1-\alpha}$$

donde $Z_{1-\alpha}$ es el $(1-\alpha)$ -percentil de la distribución normal estándar.

Por lo que

$$k^* = \theta_0 + Z_{1-\alpha} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}$$

con lo que la región crítica puede expresarse

$$C := \left\{ (x_1, \dots, x_n) : \hat{\theta} > \theta_0 + Z_{1-\alpha} \sqrt{\frac{\theta_0(1-\theta_0)}{n}} \right\}$$

Si se desea probar el contraste $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$ la prueba uniformemente más potente de tamaño α queda determinada por la expresión (1.6) con las desigualdades en el sentido inverso, considerando que la prueba sea no aleatorizada la región crítica queda

$$C := \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n x_i < k \right\}$$

donde k se determina por la condición $\alpha = P_{\theta_0} \left(T = \sum_{i=1}^n x_i < k \right)$. Si n es suficientemente grande y se emplea la aproximación normal

$$\begin{aligned} C &= \left\{ (x_1, \dots, x_n) : \hat{\theta} < \theta_0 + Z_{\alpha} \sqrt{\frac{\theta_0(1-\theta_0)}{n}} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \hat{\theta} < \theta_0 - Z_{1-\alpha} \sqrt{\frac{\theta_0(1-\theta_0)}{n}} \right\} \end{aligned}$$

Ejemplo 1.3.6 Sea X_1, \dots, X_n una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 conocida. Entonces para cualesquiera $\mu' < \mu''$

$$\begin{aligned} \frac{f(\mathbf{x}; \mu'')}{f(\mathbf{x}; \mu')} &= \frac{\frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu'')^2 \right) \right\}}{\frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu')^2 \right) \right\}} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu'')^2 - \sum_{i=1}^n (x_i - \mu')^2 \right) \right\} \\ &= \exp \left\{ \frac{n(\mu'' - \mu')}{\sigma^2} \left(\bar{x} - \frac{\mu'' + \mu'}{2} \right) \right\} \end{aligned}$$

Por lo que la razón $\frac{f(\mathbf{x}; \mu'')}{f(\mathbf{x}; \mu')}$ depende de la muestra sólo a través de la media muestral \bar{x} y es creciente en función de \bar{x} .

Entonces la familia de densidades normales con media desconocida y varianza conocida tiene una razón de verosimilitud monótona en la estadística \bar{x} . Si se desea probar el contraste de hipótesis $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ del teorema 1.3.3 se sigue que la prueba UMP de tamaño α es

$$\varphi(x) := \begin{cases} 1 & \text{si } \bar{x} > k \\ \gamma & \text{si } \bar{x} = k \\ 0 & \text{si } \bar{x} < k \end{cases}$$

donde k y γ se determinan por

$$\alpha = E_{\theta_0}(\varphi(X))$$

En este caso \bar{X} tiene una distribución continua ($\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$) por lo que $P(\bar{X} = k) = 0$ y la prueba UMP tiene región crítica

$$C := \{(x_1, \dots, x_n) : \bar{x} > k\}$$

donde k se determina por la condición $\alpha = P_{\mu_0}(\bar{x} > k)$. Bajo H_0 se tiene que k es el $(1 - \alpha)$ -percentil de la distribución normal con media μ_0 y varianza $\frac{\sigma^2}{n}$. Para determinar la función de potencia

$$\begin{aligned} \pi(\mu) &= P_{\mu}(\bar{x} > k) = 1 - P_{\mu}(\bar{x} \leq k) \\ &= 1 - P_{\mu}\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{k - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \Phi\left(\frac{k - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \end{aligned}$$

La condición $\alpha = P_{\mu_0}(\bar{x} > k)$ se emplea para determinar el valor de

$$k = \mu_0 + Z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

Por lo que la región crítica puede reescribirse como

$$C := \left\{ (x_1, \dots, x_n) : \bar{x} > \mu_0 + Z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

Si el contraste que se desea probar es $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$ la región crítica de la prueba uniformemente más potente de tamaño α tiene la forma

$$C := \{(x_1, \dots, x_n) : \bar{x} < k\}$$

con $\alpha = P_{\mu_0}(\bar{x} < k)$ por lo que k es el α -percentil de la distribución normal con media μ_0 y varianza $\frac{\sigma^2}{n}$. De manera alternativa

$$\begin{aligned} C &= \left\{ (x_1, \dots, x_n) : \bar{x} > \mu_0 + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \bar{x} > \mu_0 - Z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\} \end{aligned}$$

Ejemplo 1.3.7 Sea X_1, \dots, X_n una muestra aleatoria de una distribución normal con media μ conocida y varianza σ^2 desconocida. Para $\sigma_1^2 < \sigma_2^2$ la razón

$$\frac{f(\mathbf{x}; \sigma_1^2)}{f(\mathbf{x}; \sigma_2^2)} = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{\frac{n}{2}} \exp \left\{ \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) \right\}$$

es una función no decreciente de $\sum_{i=1}^n (x_i - \mu)^2$, por lo que la familia de densidades normales con media conocida y varianza desconocida tiene una razón de verosimilitud monótona en la estadística $T = \sum_{i=1}^n (x_i - \mu)^2$. Del teorema 1.3.3 la prueba UMP para el contraste $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ esta dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 > k \\ \gamma & \text{si } \sum_{i=1}^n (x_i - \mu)^2 = k \\ 0 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 < k \end{cases}$$

donde k y γ se determinan por $\alpha = E_{\theta_0}(\varphi(\mathbf{X}))$. Bajo H_0 se tiene que $\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2}$ tiene una distribución ji-cuadrada con n grados de libertad, por lo que $P_{\sigma_0^2} \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} = k \right) = 0$. Con esto la prueba puede reescribirse en términos de la región crítica

$$C := \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \mu)^2 > k \right\}$$

donde k se determina por $\alpha = P_{\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu)^2 > k \right)$. Empleando que bajo H_0 se tiene $\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} \sim \chi_{(n)}^2$

$$\alpha = P_{\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu)^2 > k \right) = P_{\sigma_0^2} \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} > \frac{k}{\sigma_0^2} \right)$$

de donde $k = \sigma_0^2 \chi_{(n, 1-\alpha)}^2$ y

$$C := \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} > \chi_{(n, 1-\alpha)}^2 \right\}$$

Empleando un argumento similar se obtiene que la prueba UMP de tamaño α para el contraste $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ tiene región crítica

$$C := \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} < \chi_{(n, 1-\alpha)}^2 \right\}$$

1.4 Pruebas insesgadas.

Ahora se analizará el problema de determinar si existe o no una prueba UMP para un contraste de hipótesis de la forma $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. El teorema 1.3.3 afirma que si la familia de densidades o masas de probabilidad tiene una razón de verosimilitud monótona en la estadística $T = t(\mathbf{X})$ siempre existe una prueba φ que es UMP para el contraste $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$ y para $\theta < \theta_0$, $\pi(\theta)$ es menor que el tamaño de la prueba. Derivado del mismo teorema se tiene que existe siempre una prueba φ^* que es UMP para el contraste $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$ y para $\theta > \theta_0$, $\pi(\theta)$ es menor que el tamaño de la prueba. Sin embargo, estas pruebas son muy distintas en su forma ya que una rechaza H_0 para valores grandes de T y la otra rechaza H_0 para valores pequeños de T . Por lo que una prueba UMP para el contraste bilateral $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ no existe.

En estos casos uno puede en algunas ocasiones encontrar una prueba UMP en una subclase de entre todas las pruebas posibles imponiendo ciertas restricciones adicionales. Una restricción que podría parecer obvia para una prueba φ del contraste de hipótesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ sería que

$$\begin{aligned} E_{\theta_0}(\varphi(X)) &= \alpha \\ E_{\theta}(\varphi(X)) &= \pi(\theta) \geq \alpha \quad \text{para } \theta \neq \theta_0 \end{aligned} \tag{1.7}$$

Esta condición nos dice que la función potencia de la prueba tiene solamente un mínimo en $\theta = \theta_0$. Una prueba φ que satisface la condición (1.7) es llamada prueba insesgada de tamaño α para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Resulta obvio que la clase de las pruebas insesgadas de tamaño α es una subclase de la clase de las pruebas de tamaño α para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Cuando es posible encontrar una prueba UMP dentro de la clase de las pruebas insesgadas de tamaño α para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, se dice que es la prueba insesgada uniformemente más potente (IUMP) de tamaño α para el contraste bilateral $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Debido a las limitaciones matemáticas y alcance del presente trabajo solamente se expone como poder determinar la prueba IUMP dentro de la familia exponencial uniparamétrica.

*Teorema 1.4.1*¹ Sea $\theta \in \mathbb{R}$ y X variable aleatoria con función de densidad

$$f(x; \theta) = a(\theta) e^{c(\theta)d(x)} b(x)$$

donde $c(\theta)$ es estrictamente monótona. Entonces existe una prueba insesgada uniformemente más potente para $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Si $c(\theta)$ es creciente la prueba esta dada por:

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } d(\mathbf{x}) < k_1 \text{ o } d(\mathbf{x}) > k_2 \\ \gamma_1 & \text{si } d(\mathbf{x}) = k_1 \\ \gamma_2 & \text{si } d(\mathbf{x}) = k_2 \\ 0 & \text{si } k_1 < d(\mathbf{x}) < k_2 \end{cases}$$

¹Para una demostración de este teorema el lector puede referirse a Lehmann.

donde k_1 , k_2 , γ_1 y γ_2 se determinan por:

$$E_{\theta_0}(\varphi(\mathbf{X})) = \alpha \text{ y } E_{\theta_0}(\varphi(\mathbf{X})d(\mathbf{X})) = \alpha E_{\theta_0}(d(\mathbf{X})).$$

Ejemplo 1.4.2 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 conocida. Se considera el problema de probar $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. Ahora

$$\begin{aligned} f(\mathbf{x}; \mu) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ \frac{n\mu\bar{x}}{\sigma^2} \right\} \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\} \end{aligned}$$

de donde tomando

$$\begin{aligned} a(\mu) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\}, \quad b(\mathbf{x}) = \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\} \\ d(\mathbf{x}) &= \bar{x} \quad \text{y} \quad c(\mu) = \frac{n\mu}{\sigma^2} \end{aligned}$$

y observando que $c(\mu)$ es creciente se sigue del teorema 1.4.1 que la prueba IUMP de tamaño α para este contraste esta dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } \bar{x} < k_1 \text{ o } \bar{x} > k_2 \\ \gamma_1 & \text{si } \bar{x} = k_1 \\ \gamma_2 & \text{si } \bar{x} = k_2 \\ 0 & \text{si } k_1 < \bar{x} < k_2 \end{cases} \quad (1.8)$$

donde (con $T = \bar{X}$)

$$\int_{-\infty}^{k_1} f_T(t; \mu_0) dt + \int_{k_2}^{\infty} f_T(t; \mu_0) dt + \gamma_1 P_{\mu_0}(\bar{X} = k_1) + \gamma_2 P_{\mu_0}(\bar{X} = k_2) = \alpha \quad (1.9)$$

$$\int_{-\infty}^{k_1} t f_T(t; \mu_0) dt + \int_{k_2}^{\infty} t f_T(t; \mu_0) dt + \gamma_1 P_{\mu_0}(\bar{x} = k_1) + \gamma_2 P_{\mu_0}(\bar{x} = k_2) = \alpha E_{\mu_0}(T) \quad (1.10)$$

Cuando H_0 es verdadero $T = \bar{X}$ se distribuye como una normal con media μ_0 y varianza $\frac{\sigma^2}{n}$. Entonces $P_{\mu_0}(\bar{X} = k_1) = 0 = P_{\mu_0}(\bar{X} = k_2)$ y

$$P_{\mu_0} \left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -u \right) = P_{\mu_0} \left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > u \right) \quad \text{para toda } u \in \mathbb{R} \quad (1.11)$$

Por lo que la ecuación (1.9) se reduce a

$$\int_{-\infty}^{k_1} f_T(t; \mu_0) dt + \int_{k_2}^{\infty} f_T(t; \mu_0) dt = \alpha \quad (1.12)$$

Utilizando (1.10) se tiene

$$\int_{-\infty}^{k_1} (t - \mu_0) f_T(t; \mu_0) dt + \int_{k_2}^{\infty} (t - \mu_0) f_T(t; \mu_0) dt = 0$$

Sea $Y = \frac{(T - \mu_0)\sqrt{n}}{\sigma}$ entonces la ecuación anterior se reduce a

$$\int_{-\infty}^{\frac{(k_1 - \mu_0)\sqrt{n}}{\sigma}} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \int_{\frac{(k_2 - \mu_0)\sqrt{n}}{\sigma}}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 0$$

de donde se concluye que $-\frac{(k_1 - \mu_0)\sqrt{n}}{\sigma} = \frac{(k_2 - \mu_0)\sqrt{n}}{\sigma}$. De la ecuación (1.12) se tiene

$$\begin{aligned} \alpha &= \int_{-\infty}^{\frac{(k_1 - \mu_0)\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \int_{\frac{(k_2 - \mu_0)\sqrt{n}}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= 2 \int_{\frac{(k_2 - \mu_0)\sqrt{n}}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \end{aligned}$$

Por lo tanto (1.8) queda expresada como una prueba de colas iguales por

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{si } \frac{|\bar{x} - \mu_0|\sqrt{n}}{\sigma} > k_2 \\ 0 & \text{si } \frac{|\bar{x} - \mu_0|\sqrt{n}}{\sigma} < k_2 \end{cases}$$

donde k_2 se determina por $\alpha = P_{\mu_0} \left(\frac{|\bar{x} - \mu_0|\sqrt{n}}{\sigma} > k_2 \right) = P_{\mu_0} \left(\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} < -k_2 \right) + P_{\mu_0} \left(\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > k_2 \right)$.

El término de colas iguales expresa que el tamaño de la prueba se reparte equitativamente entre las colas de la distribución.

Empleando el hecho de que $\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \sim N(0, 1)$ y la ecuación (1.11) para determinar el valor de k_2 la prueba IUMP de tamaño α para probar

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0$$

expresada en términos de su región crítica es

$$\begin{aligned}
C &= \left\{ (x_1, \dots, x_n) : \frac{|\bar{x} - \mu_0| \sqrt{n}}{\sigma} > Z_{1-\alpha} \right\} \\
&= \left\{ (x_1, \dots, x_n) : \frac{(\bar{x} - \mu_0) \sqrt{n}}{\sigma} > Z_{1-\frac{\alpha}{2}} \right\} \cup \\
&\quad \cup \left\{ (x_1, \dots, x_n) : \frac{(\bar{x} - \mu_0) \sqrt{n}}{\sigma} < -Z_{1-\frac{\alpha}{2}} \right\} \\
&= \left\{ (x_1, \dots, x_n) : \bar{x} > \mu_0 + \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}} \right\} \cup \left\{ (x_1, \dots, x_n) : \bar{x} < \mu_0 - \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}} \right\}
\end{aligned}$$

Ejemplo 1.4.3 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ conocida y varianza σ^2 desconocida. Se considera el problema de probar $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$. Ahora

$$f(\mathbf{x}; \mu) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \right\}$$

de donde tomando

$$\begin{aligned}
a(\sigma^2) &= \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\}, b(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{n}{2}} \\
d(\mathbf{x}) &= \sum_{i=1}^n (x_i - \mu)^2 \text{ y } c(\sigma^2) = -\frac{1}{2\sigma^2}
\end{aligned}$$

la cual es creciente se sigue del teorema 1.4.1 que la prueba IUMP de tamaño α para este contraste esta dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 < k_1 \text{ o } \sum_{i=1}^n (x_i - \mu)^2 > k_2 \\ \gamma_1 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 = k_1 \\ \gamma_2 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 = k_2 \\ 0 & \text{si } k_1 < \sum_{i=1}^n (x_i - \mu)^2 < k_2 \end{cases}$$

donde k_1 , k_2 , γ_1 y γ_2 se determinan por

$$\begin{aligned}
E_{\sigma_0}(\varphi(\mathbf{X})) &= \alpha \\
E_{\sigma_0}(\varphi(\mathbf{X})d(\mathbf{X})) &= \alpha E_{\sigma_0}(d(\mathbf{X})).
\end{aligned} \tag{1.13}$$

Como las X_i 's son independientes e idénticamente distribuidas, cuando H_0 es verdadera $Z = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2}$ tiene una distribución ji-cuadrada con n grados de

libertad, por lo que

$$P_{\sigma_0^2} \left(Z = \frac{k_i}{\sigma_0^2} \right) = 0, \quad i = 1, 2 \quad \text{y} \quad E_{\sigma_0^2}(Z) = n$$

Entonces las condiciones enunciadas en (1.13) se reducen a

$$\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} f_Z(z; \sigma_0^2) dz = 1 - \alpha \quad (1.14)$$

y

$$n - \int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} z f_Z(z; \sigma_0^2) dz = \alpha n$$

o

$$\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} \frac{1}{\Gamma\left(\frac{n+2}{2}\right)} \frac{1}{2^{\frac{n+2}{2}}} z^{\left(\frac{n+2}{2}\right)-1} e^{-\frac{z}{2}} dz = 1 - \alpha \quad (1.15)$$

donde el integrando anterior es la función de densidad de una distribución ji-cuadrada con $n + 2$ grados de libertad. Integrando por partes (1.15) se tiene

$$1 - \alpha = -\frac{1}{\Gamma\left(\frac{n+2}{2}\right)} \frac{1}{2^{\frac{n+2}{2}}} e^{-\frac{z}{2}} z^{\left(\frac{n+2}{2}\right)-1} \Big|_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} + \int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{2^{\frac{n}{2}}} z^{\left(\frac{n}{2}\right)-1} e^{-\frac{z}{2}} dz \quad (1.16)$$

Combinando (1.15) y (1.16) se tiene

$$k_1^{\frac{n}{2}} e^{-\frac{k_1}{2\sigma_0^2}} = k_2^{\frac{n}{2}} e^{-\frac{k_2}{2\sigma_0^2}} \quad (1.17)$$

Por lo tanto la prueba IUMP para probar $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ esta dada por

$$\varphi(\mathbf{x}) := \begin{cases} 1 & \text{si } \sum_{i=1}^n (x_i - \mu)^2 < k_1 \text{ o } \sum_{i=1}^n (x_i - \mu)^2 > k_2 \\ 0 & \text{si } k_1 < \sum_{i=1}^n (x_i - \mu)^2 < k_2 \end{cases} \quad (1.18)$$

donde las constantes k_1 y k_2 se determinan por las ecuaciones (1.14) y (1.17). Considerando el caso en que n es suficientemente grande tal que bajo H_0 verdadera $T = \frac{Z-n}{\sqrt{2n}}$ tenga aproximadamente una distribución normal estándar y

empleando la propiedad de simetría con respecto a la media de la distribución normal, podemos como en el ejemplo anterior expresar la región crítica de la prueba de colas iguales dada por (1.18) donde k_1 y k_2 son tales que

$$\int_0^{\frac{k_1}{\sigma_0^2}} f_Z(z; \sigma_0^2) dz = \int_{\frac{k_2}{\sigma_0^2}}^{\infty} f_Z(z; \sigma_0^2) dz = \frac{\alpha}{2}$$

mediante

$$C = \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \mu)^2 < \sigma_0^2 \chi_{(n, \frac{\alpha}{2})}^2 \right\} \cup \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \mu)^2 > \sigma_0^2 \chi_{(n, 1 - \frac{\alpha}{2})}^2 \right\}$$

1.5 Prueba de razón de verosimilitud.

Anteriormente se ha expuesto (teorema 1.3.3, corolario 1.3.4) como obtener pruebas UMP para contrastes de hipótesis unilaterales de la forma $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ o $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$ cuando la familia de densidades o masas de probabilidad tienen una razón de verosimilitud monótona o pertenecen a la familia exponencial uniparamétrica, además se ha mostrado (teorema 1.4.1) que para la familia exponencial uniparamétrica existe una prueba IUMP para el contraste bilateral $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. A continuación se discutirá un método general de construir una prueba en el que θ puede ser un vector.

Para una muestra aleatoria X_1, \dots, X_n de una densidad $f(x; \theta)$, $\theta \in \Theta$, buscamos una prueba para $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$.

Definición 1.5.1 Sea $L(\theta; \mathbf{x})$ la función de verosimilitud de la muestra \mathbf{X} con función de densidad conjunta $f(\mathbf{x}; \theta)$ donde $\theta \in \Theta$. La estadística de prueba de razón de verosimilitud para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$ esta definida por:

$$\lambda = \lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

Una prueba de razón de verosimilitud es cualquier prueba con una región de rechazo de la forma $C := \{(x_1, \dots, x_n) : \lambda(\mathbf{x}) \leq c\}$ con $0 \leq c \leq 1$.

Se debe mencionar que en este caso λ es una función de las observaciones x_1, \dots, x_n . Cuando las observaciones son reemplazadas por sus correspondientes variables aleatorias X_1, \dots, X_n escribimos Λ en lugar de λ . Además $\lambda \geq 0$ debido a que es el cociente de dos cantidades no negativas y $\lambda \leq 1$ debido a que el supremo del denominador se obtiene de un conjunto más grande que el del numerador, por lo que el denominador no puede ser más pequeño que el numerador.

El principio de prueba de razón de verosimilitud es puramente intuitivo, bajo la idea de que λ tenderá a ser pequeño cuando H_0 no sea cierta, debido a que el denominador tenderá a ser más grande que el numerador. En general será una buena prueba, en muchos de los problemas clásicos ha mostrado poseer propiedades óptimas cuando n tiende a infinito, aunque existen ejemplos donde la prueba de razón de verosimilitud establece una exposición pobre comparada con otras pruebas.

Entre las posibles desventajas de la prueba es que algunas veces es difícil obtener el $\sup L(\theta; \mathbf{x})$ y en algunas ocasiones es difícil encontrar la distribución de Λ bajo la hipótesis nula H_0 necesaria para determinar la región crítica, si H_0 es compuesta la distribución de Λ puede no ser única por lo cual no pueda determinarse c de manera única. Nos concentraremos en problemas donde c pueda determinarse de manera única. Adicionalmente se debe mencionar que en los casos donde c no puede determinarse de manera única, una solución muy satisfactoria puede ser obtenida tomando el tamaño de la muestra n suficientemente grande. Cuando n tiende a infinito bajo H_0 , $-2 \ln(\Lambda)$ tiene aproximadamente una distribución ji-cuadrada con $k - r$ grados de libertad² donde los valores de k y r representan las dimensiones de Θ y Θ_0 respectivamente. Debido a que $-2 \ln(\lambda)$ crece cuando λ decrece la región crítica de la hipótesis nula H_0 en términos de $-2 \ln(\lambda)$ estará dada por

$$C := \{(x_1, \dots, x_n) : -2 \ln(\lambda(\mathbf{x})) \geq c\}$$

donde la constante c se determina por $P(\chi_{k-r}^2 \geq c) = \alpha$.

Si pensamos en maximizar sobre ambos; el espacio parametral completo (maximizar sin restricción) y un subconjunto del espacio parametral (maximización restringida) entonces la correspondencia entre las pruebas de razón de verosimilitud y los estimadores máximo verosímiles comienza a ser más clara. Supongase que existe $\hat{\theta}$ un estimador máximo verosímil de θ , $\hat{\theta}$ es obtenido maximizando sin restricción $L(\theta; \mathbf{x})$. Podemos además considerar el estimador máximo verosímil de θ llamémosle $\hat{\theta}_0$ obtenido maximizando $L(\theta; \mathbf{x})$ asumiendo que Θ_0 es el espacio de parámetros, es decir, $\hat{\theta}_0 = \hat{\theta}_0(\mathbf{x})$ es el valor de $\theta \in \Theta_0$

²Para una demostración de este resultado el lector puede referirse a Kendall and Stuarts o Lehmann

que maximiza $L(\theta; \mathbf{x})$. Entonces:

$$\lambda = \frac{\hat{L}(\hat{\theta}_0; \mathbf{x})}{\hat{L}(\hat{\theta}; \mathbf{x})}$$

Ejemplo 1.5.2 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 ambos parámetros desconocidos. Considere el problema de probar $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

La función de verosimilitud esta dada por

$$L(\theta; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \right\}$$

donde $\theta = (\mu, \sigma^2)$. Tenemos

$$\begin{aligned} \Theta_0 &= \{ \theta = (\mu_0, \sigma^2) : \sigma^2 > 0 \} \\ \Theta &= \{ \theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \} \end{aligned}$$

Para $\theta \in \Theta_0$ se tiene

$$L(\theta; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right\}$$

Para obtener el máximo se observa

$$\frac{\delta \ln(L(\theta; \mathbf{x}))}{\delta \sigma^2} = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu_0)^2}{(\sigma^2)^2} = 0$$

si $\sigma^2 = \sum_{i=1}^n \frac{(x_i - \mu_0)^2}{n}$. Por lo que

$$L(\hat{\theta}_0; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{n}{2}} \left[\frac{n}{\sum_{i=1}^n (x_i - \mu_0)^2} \right]^{\frac{n}{2}} e^{-\frac{n}{2}}$$

De manera similar para $\theta \in \Theta$ el máximo de $L(\theta; \mathbf{x})$ se alcanza en

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$$

con

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2 = \frac{(n-1)s^2}{n} \quad (1.19)$$

son los estimadores máximo verosímiles de μ y σ^2 respectivamente, con

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

Entonces

$$\lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} = \left[\frac{1}{1 + \left(\frac{y^2}{n-1}\right)} \right]^{\frac{n}{2}}$$

donde

$$y = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Cuando H_0 es verdadera $Y = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$ donde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{y} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

tiene una distribución t central con $n-1$ grados de libertad; es decir $Y \sim t_{n-1}$.

Debido a que λ es monótona decreciente en función de y^2 , la condición $\lambda \leq c$ puede representarse de manera equivalente por $y^2 \geq c$ con $0 \leq c \leq 1$. Entonces la región crítica de la prueba de razón de verosimilitud de tamaño α esta dada por

$$C := \{(x_1, \dots, x_n) : y \leq -c \text{ o } y \geq c\}$$

donde c es tal que $P(Y \leq -c) + P(Y \geq c) = \alpha$. Como $Y \sim t_{n-1}$ se cumple $P(Y \leq -c) = P(Y \geq c)$, por lo que la constante c puede ser determinada por $P(Y \geq c) = \frac{\alpha}{2}$. Y la región crítica puede reescribirse

$$\begin{aligned} C &= \left\{ (x_1, \dots, x_n) : \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \leq t_{(\frac{\alpha}{2}, n-1)} \text{ o } \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \geq t_{(1-\frac{\alpha}{2}, n-1)} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \bar{x} \leq \mu_0 + \frac{s}{\sqrt{n}} t_{(\frac{\alpha}{2}, n-1)} \right\} \cup \\ &\quad \cup \left\{ (x_1, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} t_{(1-\frac{\alpha}{2}, n-1)} \right\} \end{aligned}$$

Ejemplo 1.5.3 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 ambos parámetros desconocidos. Considere el problema de probar $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$. La función de verosimilitud es

$$L(\theta; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \right\}$$

donde $\theta = (\mu, \sigma^2)$. Tenemos

$$\Theta_0 = \{\theta = (\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}$$

$$\Theta = \{\theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

Para $\theta \in \Theta$ el máximo de $L(\theta; \mathbf{x})$ se alcanza en $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$ donde $\hat{\mu}$ y $\hat{\sigma}^2$ están dados por (1.19).

Por otro lado para $\theta \in \Theta_0$ se observa que si $\hat{\mu} \leq \mu_0$ se concluye que el máximo sobre el conjunto restringido es el mismo que sobre el conjunto completo, mientras que si $\hat{\mu} > \mu_0$ el máximo restringido está en $\hat{\theta}_0 = (\mu_0, \hat{\sigma}_0^2)$ donde

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2. \text{ Por lo tanto}$$

$$\lambda = \begin{cases} 1 & \text{si } \bar{x} = \hat{\mu} \leq \mu_0 \\ \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} & \text{si } \bar{x} = \hat{\mu} > \mu_0 \end{cases} \quad (1.20)$$

Por lo que la prueba de razón de verosimilitud puede enunciarse: rechazar H_0 si $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \left[\frac{1}{1 + \left(\frac{y^2}{n-1}\right)} \right]^{\frac{n}{2}} \leq c$, donde $y = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$ y $0 \leq c < 1$. Cuando H_0 es verdadera $Y = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}$ donde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{y} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Como en el ejemplo (1.5.2) la condición $\lambda \leq c$ puede representarse por $y^2 \geq c$ y de (1.20) la región crítica de tamaño α de esta prueba toma la forma

$$\begin{aligned} C &= \{(x_1, \dots, x_n) : y \geq c\} = \left\{ (x_1, \dots, x_n) : \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \geq t_{(1-\alpha, n-1)} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} t_{(1-\alpha, n-1)} \right\} \end{aligned}$$

Con un argumento similar se obtiene que la región crítica de tamaño α de la prueba de razón de verosimilitud para probar $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$ la cual está dada por

$$\begin{aligned} C &= \left\{ (x_1, \dots, x_n) : \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \leq t_{(\alpha, n-1)} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \bar{x} \leq \mu_0 + \frac{s}{\sqrt{n}} t_{(\alpha, n-1)} \right\} \end{aligned}$$

Ejemplo 1.5.4 Sea X_1, \dots, X_m una muestra aleatoria de una distribución normal con media μ_1 y varianza σ_1^2 ambos parámetros desconocidos y sea Y_1, \dots, Y_n

una muestra aleatoria (independiente de X_1, \dots, X_m) de otra distribución normal con media μ_2 y varianza σ_2^2 ambos parámetros también desconocidos. Considere el problema de probar $H_0 : \mu_1 = \mu_2 = \mu$ vs $H_1 : \mu_1 \neq \mu_2$. Para este problema se tiene

$$\Theta = \{ \theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : \mu_i \in \mathbb{R}, \sigma_i^2 > 0, i = 1, 2 \}$$

$$\Theta_0 = \{ \theta = (\mu, \sigma_1^2, \sigma_2^2) : \mu \in \mathbb{R}, \sigma_i^2 > 0, i = 1, 2 \}$$

La función de verosimilitud de las observaciones muestrales es

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{1}{\sigma_1^2} \right)^{\frac{m}{2}} \left(\frac{1}{\sigma_2^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^m \frac{(x_i - \mu_1)^2}{\sigma_1^2} + \sum_{i=1}^n \frac{(y_i - \mu_2)^2}{\sigma_2^2} \right) \right\}$$

Empleando los estimadores máximo verosímiles se tiene que para $\theta \in \bar{\Theta}$ el máximo de $L(\theta; \mathbf{x}, \mathbf{y})$ es

$$L(\hat{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{m}{(m-1)s_1^2} \right)^{\frac{m}{2}} \left(\frac{n}{(n-1)s_2^2} \right)^{\frac{n}{2}} e^{-\frac{m+n}{2}}$$

Mientras que para $\theta \in \Theta_0$, $L(\theta; \mathbf{x}, \mathbf{y})$ tiene un máximo cuando $\mu, \sigma_1^2, \sigma_2^2$ satisfacen

$$\bar{x} - \mu + \frac{n\sigma_1^2}{m\sigma_2^2}(\bar{y} - \mu) = 0, \quad \sigma_1^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2, \quad \sigma_2^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2$$

Donde se encontrará que el estimador máximo verosímil de μ esta dado por la raíz de una ecuación cúbica y será una función muy compleja de las observaciones. Obviamente en este caso la prueba de razón de verosimilitud no es muy apropiada para usos prácticos. Se consideran algunos casos especiales que presentan soluciones simples.

Caso 1: Varianzas iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Redefinimos el espacio parametral Θ como sigue

$$\Theta = \{ \theta = (\mu_1, \mu_2, \sigma^2) : \mu_i \in \mathbb{R}, \sigma^2 > 0, \}$$

$$\Theta_0 = \{ \theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \}$$

Para $\theta \in \Theta$ se tiene

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{m+n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^m (x_i - \mu_1)^2 + \sum_{i=1}^n (y_i - \mu_2)^2 \right) \right\}$$

Donde el máximo de $L(\theta)$ ocurre en

$$L(\hat{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{m+n}{(m-1)s_1^2 + (n-1)s_2^2} \right)^{\frac{m+n}{2}} e^{-\frac{m+n}{2}} \quad (1.21)$$

donde

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2 \quad \text{y} \quad s_2^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Para $\theta \in \Theta_0$ se tiene

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{m+n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^m (x_i - \mu)^2 + \sum_{i=1}^n (y_i - \mu)^2 \right) \right\}$$

por lo que los estimadores máximo verosímiles de μ y σ^2 estan dados por

$$\begin{aligned} \hat{\mu} &= \frac{m\bar{x} + n\bar{y}}{m+n} & (1.22) \\ \hat{\sigma}^2 &= \frac{1}{m+n} \left[\sum_{i=1}^m (x_i - \hat{\mu})^2 + \sum_{i=1}^n (y_i - \hat{\mu})^2 \right] \\ &= \frac{1}{m+n} \left[(m-1)s_1^2 + (n-1)s_2^2 + \frac{mn}{m+n} (\bar{x} - \bar{y})^2 \right] \end{aligned}$$

Por tanto

$$L(\hat{\theta}_0) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left[\frac{m+n}{(m-1)s_1^2 + (n-1)s_2^2 + \frac{mn}{m+n} (\bar{x} - \bar{y})^2} \right]^{\frac{m+n}{2}} e^{-\frac{m+n}{2}} \quad (1.23)$$

De (1.21) y (1.23), la estadística de razón de verosimilitud esta dada por

$$\lambda = \left[\frac{1}{1 + \frac{(\frac{mn}{m+n})(\bar{x} - \bar{y})^2}{(m-1)s_1^2 + (n-1)s_2^2}} \right]^{\frac{m+n}{2}} = \left(\frac{1}{1 + \frac{z^2}{m+n-2}} \right)^{\frac{m+n}{2}} \quad (1.24)$$

donde

$$z^2 = \frac{(\bar{x} - \bar{y})^2 / \left(\frac{1}{m} + \frac{1}{n} \right)}{\left((m-1)s_1^2 + (n-1)s_2^2 \right) / (m+n-2)}$$

Sea

$$\begin{aligned} Z &= \frac{(\bar{X} - \bar{Y}) / \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}}}{\left[\left((m-1)S_1^2 + (n-1)S_2^2 \right) / (m+n-2) \right]^{\frac{1}{2}}} \\ &= \frac{(\bar{X} - \bar{Y}) / \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}} \sigma}{\left[\left((m-1)S_1^2 + (n-1)S_2^2 \right) / (m+n-2) \right]^{\frac{1}{2}} \sigma} \end{aligned}$$

Cuando la hipótesis nula es verdadera $\bar{X} - \bar{Y} \sim N \left(0, \left(\frac{1}{m} + \frac{1}{n} \right) \sigma^2 \right)$.

Además $\frac{(m-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2}$ son variables aleatorias independientes con distribución ji-cuadrada con $m-1$ y $n-1$ grados de libertad respectivamente e

independientes de $\bar{X} - \bar{Y}$. Entonces Z tiene una distribución t con $m + n - 2$ grados de libertad. De la ecuación (1.24) se sigue que la región crítica para probar $H_0 : \mu_1 = \mu_2 = \mu$ vs $H_1 : \mu_1 \neq \mu_2$ esta dada por

$$C := \{(x_1, \dots, x_n), (y_1, \dots, y_n) : z \leq -c \text{ o } z \geq c\}$$

Donde c se elige de tal manera que bajo H_0 ,

$$P(Z \leq -c) + P(Z \geq c) = \alpha \quad (1.25)$$

Como $Z \sim t_{m+n-2}$, $P(Z \leq -c) = P(Z \geq c)$, por lo que la constante c puede ser determinada por $P(Z \geq c) = \frac{\alpha}{2}$. Y la región crítica puede reescribirse

$$C = \left\{ (x_1, \dots, x_n), (y_1, \dots, y_n) : \bar{x} - \bar{y} \leq s^* \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}} t_{(\frac{\alpha}{2}, m+n-2)} \right\} \cup \\ \cup \left\{ (x_1, \dots, x_n), (y_1, \dots, y_n) : \bar{x} - \bar{y} \geq s^* \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}} t_{(1-\frac{\alpha}{2}, m+n-2)} \right\}$$

donde

$$s^* = \left(\frac{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}{m + n - 2} \right)^{\frac{1}{2}} \quad (1.26)$$

Si el contraste de hipótesis a considerar es de la forma $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$, se observa que el máximo sobre Θ esta dado por (1.21). Si $\bar{x} \leq \bar{y}$ el máximo sobre Θ_0 también es igual a (1.21), por lo que $\lambda = 1$. Si $\bar{x} > \bar{y}$ puede verse que $\mu_1 = \mu_2$ es necesario con objeto de obtener el máximo, en este caso el máximo es expresado por (1.22) y (1.23), entonces el valor de λ es el mismo que en (1.24). Con lo anterior la prueba puede expresarse como rechazar H_0 si $\lambda \leq c$ con $0 \leq c < 1$. De la ecuación (1.24) se sigue que la región crítica para probar $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$ esta dada por

$$C := \{(x_1, \dots, x_n), (y_1, \dots, y_n) : y \geq c\}$$

Donde c es tal que bajo H_0 satisface (1.25), Como $Z \sim t_{m+n-2}$, c se elige tal que $P(Z \geq c) = \frac{\alpha}{2}$, entonces la región crítica de la prueba es

$$C := \left\{ (x_1, \dots, x_n), (y_1, \dots, y_n) : \bar{x} - \bar{y} \geq s^* \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}} t_{(1-\alpha, m+n-2)} \right\}$$

con s^* igual que en (1.26).

De manera análoga se obtiene que la región crítica para el contraste de hipótesis $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$ es la siguiente

$$C := \left\{ (x_1, \dots, x_n), (y_1, \dots, y_n) : \bar{x} - \bar{y} \leq s^* \left(\frac{1}{m} + \frac{1}{n} \right)^{\frac{1}{2}} t_{(\alpha, m+n-2)} \right\}$$

Caso 2: $\sigma_1^2 \neq \sigma_2^2$ y $m = n$.

Una solución apropiada para este caso se alcanza utilizando el siguiente artificio de emparejamiento. Se define $Z_i = X_i - Y_i$, $i = 1, \dots, m$. Entonces Z_1, \dots, Z_m constituye una muestra aleatoria de una distribución normal con media $\mu_1 - \mu_2 = v$ y varianza $\sigma_1^2 + \sigma_2^2 = \sigma^2$. En términos de v y σ^2 , el problema se reduce a probar $H_0 : v = 0$ vs $H_1 : v \neq 0$, cuando σ^2 es desconocida. Del ejemplo (1.5.2) se tiene que la región crítica para este contraste es

$$C := \{(x_1, \dots, x_n) : y \leq -c \text{ o } y \geq c\} \quad (1.27)$$

$$\text{donde } y = \frac{\sqrt{m}\bar{z}}{s}, \quad s^2 = \sum_{i=1}^m \frac{(z_i - \bar{z})^2}{m-1}.$$

Bajo H_0 , $Y = \frac{\sqrt{m}\bar{Z}}{S}$ tiene una distribución t -Student con $m-1$ grados de libertad y la constante c se elige de tal manera que $P(Y \geq c) = \frac{\alpha}{2}$ cuando H_0 es verdadera.

Cuando el contraste de hipótesis es de la forma $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$ o $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$, el mismo artificio de emparejamiento reduce los contrastes a $H_0 : v \leq 0$ vs $H_1 : v > 0$ y $H_0 : v \geq 0$ vs $H_1 : v < 0$ respectivamente cuando σ^2 es desconocida. Entonces las regiones críticas tienen la forma expuesta en el ejemplo (1.5.3)

Caso 3: Solución de Scheffe para $\sigma_1^2 \neq \sigma_2^2$ y $m \neq n$.

Asumiendo que $m < n$ se define $Z_i = X_i - \left(\frac{m}{n}\right)^{\frac{1}{2}} Y_i + \frac{1}{\sqrt{mn}} \sum_{i=1}^m Y_i - \frac{1}{n} \sum_{i=1}^n Y_i$, para $i = 1, \dots, n$. Entonces $E(Z_i) = \mu_1 - \mu_2$, $Var(Z_i) = \sigma_1^2 + \left(\frac{m}{n}\right) \sigma_2^2$ y $Cov(Z_i, Z_j) = 0$, $i \neq j$. Por lo que Z_1, \dots, Z_m constituye una muestra aleatoria de una distribución normal con media $v = \mu_1 - \mu_2$ y varianza $\sigma_1^2 + \left(\frac{m}{n}\right) \sigma_2^2$. De donde se sigue que en base a las observaciones z_1, \dots, z_m la región crítica de la prueba de razón de verosimilitud de $H_0 : v = 0$ vs $H_1 : v \neq 0$ esta dada por (1.27).

Ejemplo 1.5.5 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 ambos parámetros desconocidos. Considere el problema de probar $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$.

La función de verosimilitud esta dada por

$$L(\theta; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2\right)\right\}$$

donde $\theta = (\mu, \sigma^2)$. Tenemos

$$\begin{aligned} \Theta_0 &= \{\theta = (\mu, \sigma_0^2) : \mu \in \mathbb{R}\} \\ \Theta &= \{\theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} \end{aligned}$$

Después de algunos cálculos se obtiene

$$L(\hat{\theta}) = \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n}{2}} \left[\frac{n}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^{\frac{n}{2}} e^{-\frac{n}{2}}$$

$$L(\hat{\theta}_0) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\right\}$$

Sea $y = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma_0^2}$. Entonces la estadística de prueba para probar H_0 vs H_1 esta dada por

$$\lambda = \left(\frac{y}{n}\right)^{\frac{n}{2}} e^{-\frac{(y-n)}{2}}$$

En base a la gráfica de λ en función de y se hace aparente (figura 2) que la desigualdad $\lambda \leq c$ es equivalente a $y \leq k_1$ o $y \geq k_2$. Donde k_1 y k_2 son valores de y tal que $\lambda = c$, esto es

$$k_1^{n/2} e^{-\frac{k_1}{2}} = k_2^{n/2} e^{-\frac{k_2}{2}} \quad (1.28)$$

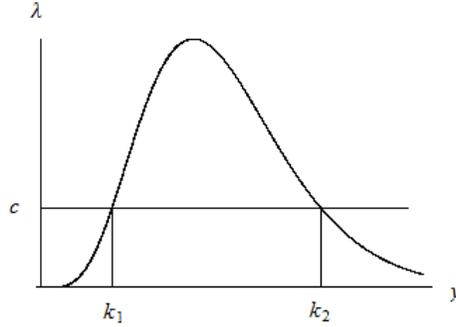


figura 2

Sea $Y = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma_0^2}$, entonces $Y \sim \chi_{(n-1)}^2$ cuando H_0 es verdadera. Por lo tanto la región crítica para probar $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ es equivalente a

$$C := \{(x_1, \dots, x_n) : y \leq k_1 \text{ o } y \geq k_2\}$$

donde k_1 y k_2 satisfacen (1.28). Como en el ejemplo (1.4.3) si n es suficientemente grande k_1 y k_2 pueden elegirse de tal manera que

$$\int_0^{k_1} f_Y(y; \sigma_0^2) dz = \int_{k_2}^{\infty} f_Y(y; \sigma_0^2) dz = \frac{\alpha}{2}$$

por lo tanto

$$\begin{aligned}
C &= \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \bar{x})^2 < \sigma_0^2 \chi_{(n-1, \frac{\alpha}{2})}^2 \right\} \cup \\
&\cup \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \bar{x})^2 > \sigma_0^2 \chi_{(n-1, 1-\frac{\alpha}{2})}^2 \right\} \\
&= \left\{ (x_1, \dots, x_n) : s^2 < \frac{\sigma_0^2}{n-1} \chi_{(n-1, \frac{\alpha}{2})}^2 \right\} \cup \\
&\cup \left\{ (x_1, \dots, x_n) : s^2 > \frac{\sigma_0^2}{n-1} \chi_{(n-1, 1-\frac{\alpha}{2})}^2 \right\}
\end{aligned}$$

$$\text{con } s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}.$$

Ejemplo 1.5.6 Sea X_1, \dots, X_m muestra aleatoria de una distribución normal con media μ_1 y varianza σ_1^2 ambos parámetros desconocidos y sea Y_1, \dots, Y_n muestra aleatoria (independiente de X_1, \dots, X_m) de otra distribución normal con media μ_2 y varianza σ_2^2 ambos parámetros también desconocidos. Considere el problema de probar $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Para este problema se tiene

$$\Theta = \{ \theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : \mu_i \in \mathbb{R}, \sigma_i^2 > 0, i = 1, 2 \}$$

$$\Theta_0 = \{ \theta = (\mu_1, \mu_2, \sigma^2) : \mu_i \in \mathbb{R}, \sigma^2 > 0, i = 1, 2 \}$$

La función de verosimilitud de las observaciones muestrales es

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{1}{\sigma_1^2} \right)^{\frac{m}{2}} \left(\frac{1}{\sigma_2^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^m \frac{(x_i - \mu_1)^2}{\sigma_1^2} + \sum_{i=1}^n \frac{(y_i - \mu_2)^2}{\sigma_2^2} \right) \right\}$$

Como en el ejemplo (1.5.4)

$$L(\hat{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{m}{(m-1)s_1^2} \right)^{\frac{m}{2}} \left(\frac{n}{(n-1)s_2^2} \right)^{\frac{n}{2}} e^{-\frac{m+n}{2}}$$

Además para $\theta \in \Theta_0$ el máximo de $L(\theta; \mathbf{x}, \mathbf{y})$ se alcanza cuando

$$\mu_1 = \bar{x}, \quad \mu_2 = \bar{y} \quad \text{y} \quad \sigma^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n}$$

Por lo que

$$L(\hat{\theta}_0) = \left(\frac{1}{\sqrt{2\pi}} \right)^{\frac{m+n}{2}} \left(\frac{m+n}{(m-1)s_1^2 + (n-1)s_2^2} \right)^{\frac{m+n}{2}} e^{-\frac{m+n}{2}}$$

Con lo que se obtiene

$$\lambda = \frac{(m+n)^{\frac{m+n}{2}} \left[\left(\frac{m-1}{n-1} \right) f \right]^{\frac{m}{2}}}{m^{\frac{n}{2}} n^{\frac{n}{2}} \left[1 + \left(\frac{m-1}{n-1} \right) f \right]^{\frac{m+n}{2}}}$$

donde $f = \frac{s_1^2}{s_2^2}$. Se define $F = \frac{S_1^2}{S_2^2}$, bajo H_0 , $\frac{S_1^2}{\sigma^2}$ y $\frac{S_2^2}{\sigma^2}$ son variables aleatorias independientes con distribución ji-cuadrada con $m-1$ y $n-1$ grados de libertad respectivamente. De donde se sigue que F tiene una distribución F -Fisher central con $m-1, n-1$ grados de libertad.

Como en el ejemplo anterior una gráfica de λ en función de y puede ser útil para establecer que la desigualdad $\lambda \leq c$ es equivalente a $f \leq k_1$ o $f \geq k_2$. Donde k_1 y k_2 son valores de y tal que $\lambda = c$, esto es

$$\left(\frac{m-1}{n-1} k_1 \right)^{\frac{m}{2}} \left(1 + \frac{m-1}{n-1} k_2 \right)^{\frac{m+n}{2}} = \left(\frac{m-1}{n-1} k_2 \right)^{\frac{n}{2}} \left(1 + \frac{m-1}{n-1} k_1 \right)^{\frac{m+n}{2}} \quad (1.29)$$

De donde se obtiene que la prueba de razón de verosimilitud tiene por región crítica

$$C := \{(x_1, \dots, x_n), (y_1, \dots, y_n) : f \leq k_1 \text{ o } f \geq k_2\}$$

donde k_1 y k_2 son constantes que satisfacen la ecuación (1.29) y

$$\int_0^{k_1} f_F(f) df + \int_{k_2}^{\infty} f_F(f) df = \alpha$$

Usualmente se eligen k_1 y k_2 tal que

$$\int_0^{k_1} f_F(f) df = \int_{k_2}^{\infty} f_F(f) df = \frac{\alpha}{2}$$

debido a que existen tablas estadísticas de la distribución F que hacen fácil determinar el valor de k_1 y k_2 .

después de realizar una prueba de hipótesis, las conclusiones deben ser reportadas de alguna manera. Se ha mencionado anteriormente que un método para reportar los resultados de una prueba es reportando el tamaño α de la prueba empleada para decidir entre aceptar o rechazar H_0 . El tamaño de la prueba proporciona información muy importante: si α es pequeño la decisión de rechazar H_0 es muy confiable, pero si α es grande la decisión de rechazar H_0 no es muy confiable debido a que existe una probabilidad grande de tomar una decisión incorrecta (error de tipo I). Otra manera de reportar los resultados de una prueba es reportar el valor de cierto tipo de estadística llamada *p-value*.

Definición 1.5.7 Un p -value $p(\mathbf{X})$ es una estadística de prueba que satisface $0 \leq p(\mathbf{x}) \leq 1$ para todos los puntos muestrales. Valores pequeños de $p(\mathbf{X})$ proporcionan evidencia de la veracidad de H_1 . Un p -value es válido si para todo $\theta \in \Theta_0$ y para cada $\alpha \in [0, 1]$

$$P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha \quad (1.30)$$

Si $p(\mathbf{X})$ es un p -value válido, resulta sencillo construir una prueba de nivel α basada en $p(\mathbf{X})$. La prueba que rechaza H_0 si y sólo si $p(\mathbf{X}) \leq \alpha$ es una prueba de nivel α por (1.30). Una ventaja de reportar el resultado de una prueba empleando el p -value es que cada lector puede elegir que valor de α considera apropiado y entonces comparar ese valor con el del p -value para conocer cuando los datos caen en la región de rechazo o aceptación de H_0 . Además mientras más pequeño sea el valor del p -value, la evidencia para rechazar H_0 será más fuerte. El valor del p -value puede interpretarse como el nivel más pequeño para el que H_0 debería ser rechazada dada la observación muestral. Por lo que un p -value reporta los resultados de una prueba en una escala continua en lugar de sólo la decisión dicotómica: "Rechazar H_0 " o "No rechazar H_0 ".

Para las pruebas que son de la forma "Rechazar H_0 cuando $T(\mathbf{X}) \geq c$ " existe una manera directa de calcular p -values³, el siguiente teorema expone esta manera de obtener p -values válidos. La demostración del teorema hace uso de los siguientes resultados: si X tiene una función de distribución continua $F_X(x)$, entonces la variable aleatoria definida por $Y = F_X(X)$ tiene una distribución uniforme en $(0, 1)$. Si X es una variable aleatoria discreta con función de distribución $F_X(x)$ entonces $Y = F_X(X)$ es estocásticamente mayor que $U(0, 1)$; esto es $P(Y > y) \geq P(U > y) = 1 - y$ para todo $y \in (0, 1)$ y $P(Y > y) > P(U > y) = 1 - y$ para algún $y \in (0, 1)$

Teorema 1.5.8 Sea $T(X_1, \dots, X_n)$ una estadística de prueba para la que valores grandes de T proporcionan evidencia de la veracidad de H_1 . Para cada punto muestral (x_1, \dots, x_n) se define

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta(T(\mathbf{X}) \geq T(\mathbf{x})) \quad (1.31)$$

Entonces $p(\mathbf{X})$ es un p -value válido.

Demostración: Sea $\theta \in \Theta_0$ un valor fijo y sea que $F_\theta(t)$ denote la función de distribución de $-T(\mathbf{X})$. Definimos

$$\begin{aligned} p_\theta(\mathbf{x}) &= P_\theta(T(\mathbf{X}) \geq T(\mathbf{x})) \\ &= P_\theta(-T(\mathbf{X}) \leq -T(\mathbf{x})) = F(-T(\mathbf{x})) \end{aligned}$$

³Los software estadísticos actuales proporcionan el valor del p -value para las pruebas estadísticas disponibles en ellos.

Entonces la variable aleatoria $p_\theta(\mathbf{X})$ es igual a $F_\theta(-T(\mathbf{X}))$, por lo que la distribución de $p_\theta(\mathbf{X})$ es estocásticamente mayor o igual a $U(0, 1)$ lo que significa que para cada $\alpha \in [0, 1]$ $P_\theta(p_\theta(\mathbf{X}) \leq \alpha) \leq \alpha$. Debido a que

$$p(\mathbf{x}) = \sup_{\theta' \in \Theta_0} p_{\theta'}(\mathbf{x}) \geq p_\theta(\mathbf{x}) \quad \text{para todo } (x_1, \dots, x_n)$$

$$P_\theta(p(\mathbf{X}) \leq \alpha) \leq P_\theta(p_\theta(\mathbf{X}) \leq \alpha) \leq \alpha$$

lo anterior es válido para todo $\theta \in \Theta_0$ y cada $\alpha \in [0, 1]$, por lo tanto $p(\mathbf{X})$ es un p -value válido. ■

Ejemplo 1.5.9 Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 ambos parámetros desconocidos. Considere el problema de probar $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. En el ejemplo (1.5.2) se observa que la prueba de razón de verosimilitud rechaza H_0 para valores grandes de $T(\mathbf{X}) = \frac{\sqrt{n}|\bar{X} - \mu_0|}{s}$, cuando $\mu = \mu_0$ independientemente del valor de σ , $\frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$ tiene una distribución t de Student con $n - 1$ grados de libertad. Por lo que para calcular (1.31) se observa que la probabilidad es la misma para todos los valores de $\theta \in \Theta_0$, empleando la simetría de la distribución t se obtiene que el p -value para esta prueba esta dado por

$$\begin{aligned} p(\mathbf{x}) &= P\left(T \leq -\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right) + P\left(T \geq \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right) \\ &= 2P\left(T \geq \frac{\sqrt{n}|\bar{x} - \mu_0|}{s}\right) \end{aligned}$$

donde T tiene una distribución t de Student con $n - 1$ grados de libertad.

1.6 Pruebas de hipótesis e intervalos de confianza.

Existe una relación muy estrecha entre las pruebas de hipótesis y los intervalos (conjuntos) de confianza, esta relación se expone en el siguiente resultado.

Teorema 1.6.1 Para cada $\theta \in \Theta$ sea $A(\theta_0)$ la región de aceptación de una prueba de tamaño α de $H_0 : \theta = \theta_0$. Para cada $\mathbf{x} \in \mathbf{X}$ se define un conjunto en el espacio parametral mediante

$$I(\mathbf{x}) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\}$$

Entonces el conjunto aleatorio $I(\mathbf{X})$ es el $(1 - \alpha)$ conjunto de confianza. Recíprocamente, sea $I(\mathbf{X})$ el $(1 - \alpha)$ conjunto de confianza. Para cada $\theta_0 \in \Theta$ se define

$$A(\theta_0) = \{\mathbf{x} : \theta_0 \in I(\mathbf{X})\}$$

Entonces $A(\theta_0)$ es la región de aceptación de una prueba de tamaño α de $H_0 : \theta = \theta_0$ ⁴

Se examinará la aplicación por medio de un ejemplo.

Ejemplo 1.6.2. En el ejemplo (1.4.2) se encontró que si X_1, \dots, X_n es una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 conocida, la prueba (IUMP) de tamaño α de $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ tiene región crítica

$$C = \left\{ (x_1, \dots, x_n) : \frac{|\bar{x} - \mu_0| \sqrt{n}}{\sigma} > Z_{1-\alpha/2} \right\}$$

Por lo que H_0 es aceptada si $|\bar{x} - \mu_0| \leq \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$ o de manera equivalente si

$$\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \leq \mu_0 \leq \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \quad (1.32)$$

Como la prueba tiene tamaño α se tiene

$$P(\text{Rechazar } H_0 \mid \mu = \mu_0) = \alpha$$

Por lo que $P(\text{Aceptar } H_0 \mid \mu = \mu_0) = 1 - \alpha$. Combinando este hecho con (1.32)

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \leq \mu_0 \leq \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right) = 1 - \alpha$$

Sin embargo, la ecuación anterior es cierta para todo valor de μ_0 , por lo tanto se tiene

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right) = 1 - \alpha$$

Por lo tanto el intervalo $\left[\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right]$ obtenido de invertir la región de aceptación de la prueba de tamaño α es el $100(1 - \alpha)\%$ intervalo de confianza para la media cuando la varianza σ^2 es conocida.

⁴Para una demostración de este teorema el lector puede referirse a Casella and Berger seg. ed.

Capítulo II.

Pruebas libres de distribución.

Aún cuando Savage (1953) designó 1936 como el inicio de la estadística no paramétrica, marcada por la publicación del artículo en correlación de rangos de Hotteling and Pabst (1936). La mayor investigación en estadística no paramétrica se realizó en la década de 1960, siendo la principal contribución la realizada por Hodges y Lehmann (1963) ellos mostraron como deducir estimadores en pruebas de rangos y establecieron que estos estimadores poseían propiedades óptimas. Su trabajo encaminó la utilización de la aproximación no paramétrica para obtener estimadores en el entorno de diseño de experimentos y pruebas no paramétricas y estimación en regresión.

Dentro de las ventajas de los métodos no paramétricos (libres de distribución) se pueden enunciar las siguientes:

1.- Requieren pocos supuestos acerca de la cual se obtiene los datos, en particular renuncian al supuesto tradicional de distribución normal de la población.

2.- Permiten al experimentador obtener valores exactos para los P-values, probabilidades de cobertura exactos para intervalos y bandas de confianza sin la necesidad de suponer distribución normal de la población.

3.- Frecuentemente (aunque no siempre) son más fáciles de aplicar y entender que sus contrapartes basadas en el supuesto de normalidad.

4.- Usualmente los métodos no paramétricos son ligeramente menos eficientes que sus contrapartes basadas en el supuesto de normalidad cuando la población sigue este comportamiento, mientras que son más eficientes cuando el supuesto de normalidad sobre la población no se cumple.

5.- Son relativamente no sensitivos a observaciones extremas.

6.- Son aplicables en muchas situaciones donde la teoría normal no puede ser utilizada. Muchos de los métodos libres de distribución requieren sólo de los rangos de las observaciones antes que sus valores, mientras que los métodos paramétricos requieren los valores de las observaciones.

Considere x_1, \dots, x_n observaciones independientes de una variable aleatoria con función de distribución $F(x)$ desconocida.

Suponga que se desea probar la hipótesis

$$H_0 : F(x) = F_0(x) \quad (2.1)$$

donde $F_0(x)$ es una función de distribución particular, la cual puede ser continua o discreta.

El problema de probar (2.1) se denomina prueba de bondad de ajuste. Si $F_0(x)$ esta completamente especificada H_0 es una hipótesis simple, en caso de que H_0 no este completamente especificada H_0 será una hipótesis compuesta, por ejemplo:

a) H_0 simple:

i) Las n observaciones provienen de una distribución normal con media y varianza especificadas.

b) H_0 compuesta:

i) Las n observaciones provienen de una distribución normal cuya media se especifica no así su varianza.

ii) Las n observaciones provienen de una distribución normal de la cual no se especifican los parámetros, frecuentemente llamada "Prueba de Normalidad".

Dos métodos muy conocidos de pruebas de bondad de ajuste dependen de un artificio simple.

2.1 Prueba de razón de verosimilitud y la prueba de bondad de ajuste de Pearson.

Se considera inicialmente el caso en que $F_0(x)$ esta completamente especificada.

Supongase que el rango de la variable aleatoria X se divide en k clases mutuamente excluyentes (en la practica usualmente se toman intervalos sucesivos en el rango de X). Debido a que $F_0(x)$ esta completamente especificada puede calcularse la probabilidad de que una observación pertenesca a cada una de las k clases, denotaremos esta probabilidad mediante $p_{0,i}$, $i = 1, \dots, k$ y las frecuencias observadas dentro de las k clases por n_i , $i = 1, \dots, k$. Entonces las frecuencias observadas siguen una distribución multinomial, es decir,

$$f_0(n_1, \dots, n_k) = n! \prod_{i=1}^k \frac{p_{0,i}^{n_i}}{n_i!}$$

donde

$$\sum_{i=1}^k p_{0,i} = 1 \quad \text{y} \quad \sum_{i=1}^k n_i = n$$

Por otro lado si la función de distribución verdadera es $F_1(x)$ la cual puede ser cualquier otra función de distribución, denotaremos las probabilidades para las k clases por $p_{1,i}$, $i = 1, \dots, k$.

Con lo anterior puede formarse la prueba de razón de verosimilitud de la hipótesis (2.1) con la hipótesis alternativa compuesta $H_1 : F(x) = F_1(x)$.

La función de verosimilitud bajo $F_1(x)$ queda

$$L_1 = L_1(n_1, \dots, n_k; p_{1,1}, \dots, p_{1,k}) = n! \prod_{i=1}^k \frac{p_{1,i}^{n_i}}{n_i!}$$

de donde se tiene

$$\ln(L_1) = \ln(n!) + \sum_{i=1}^k n_i \ln(p_{1,i}) + \sum_{i=1}^k \ln(n_i)$$

para obtener los estimadores máximo verosímiles se emplea el método de los multiplicadores de Lagrange con la restricción $\sum_{i=1}^k p_{1,i} = 1$, derivando con respecto a cada una de las probabilidades $p_{1,i}$ e igualando a cero se obtiene

$$\frac{n_i}{p_{1,i}} - \lambda = 0 \quad \text{para } i : 1, \dots, k$$

sumando todas las ecuaciones anteriores y del hecho que $\sum_{i=1}^k n_i = n$ se obtiene $\lambda = n$. Por lo tanto los estimadores máximo verosímiles son

$$\hat{p}_{1,i} = \frac{n_i}{n}$$

Por lo que la estadística de razón de verosimilitud para probar H_0 vs H_1 toma la forma:

$$\lambda = \frac{L_0}{L_1} = n^n \prod_{i=1}^k \left(\frac{p_{0,i}}{n_i} \right)^{n_i}$$

de donde la prueba rechaza H_0 para valores suficientemente pequeños de λ .

La distribución exacta de la estadística λ es desconocida; sin embargo, se ha mencionado (sección 1.5) que cuando H_0 es verdadero y $n \rightarrow \infty$,

$$-2 \ln(\lambda) = 2 \sum_{i=1}^k n_i \ln \left(\frac{n_i}{np_{0,i}} \right) \quad (2.2)$$

sigue asintóticamente una distribución ji-cuadrada con $k - 1$ grados de libertad.

La estadística de prueba expuesta por Karl Pearson (1900) es quizá la más conocida y utilizada para esta situación.

La cuál puede expresarse bajo nuestra notación actual por

$$Q_n = \sum_{i=1}^k \frac{(n_i - np_{0,i})^2}{np_{0,i}} \quad (2.3)$$

que bajo H_0 sigue asintóticamente una distribución ji-cuadrada con $k - 1$ grados de libertad⁵.

En el uso de esta aproximación frecuentemente se construyen las clases de modo que la frecuencia esperada en cada uno de ellas sea de por lo menos 5, para las clases con frecuencias esperadas menores a 5 el procedimiento usual consiste en combinar clases adyacentes en la distribución de frecuencias hasta que la restricción sea satisfecha.

Debido a que en el numerador de la estadística de prueba de Pearson Q_n se puede observar la diferencia entre las frecuencias observadas y las frecuencias esperadas cuando la distribución verdadera de las observaciones es $F_0(x)$, la prueba rechaza H_0 para valores grandes de Q_n , más aún la prueba rechaza H_0 cuando $Q_n \geq c$ donde c es el $(1 - \alpha)$ -percentil de la distribución χ^2 con $k - 1$ grados de libertad.

La esencia de las dos pruebas anteriores es reducir el problema de probar 2.1 a uno que concierne a la distribución multinomial. La necesidad de agrupar los datos dentro de clases envuelve el sacrificio de cierta cantidad de información, especialmente si la variable es continua. Sin embargo, este defecto viene acompañado de una ventaja, no es necesario conocer los valores de las observaciones individuales solo se necesita tener k clases para las cuales las probabilidades hipotéticas $p_{0,i}$ puedan ser calculadas; por lo que podemos utilizar cualquiera de las pruebas anteriores aún si los datos originales se refieren a una clasificación no numérica.

Ejemplo 2.1.1

En algunos experimentos clásicos de germinación de guisantes Mendel observó las frecuencias de diferentes tipos de semillas de cruces entre plantas con semillas amarillas redondas y plantas con semillas verdes rugosas.

Las frecuencias observadas se presentan a continuación junto con las probabilidades teóricas bajo la teoría mendeliana de herencia.

Semillas	Frecuencias observadas (n_i)	Probabilidad teórica ($p_{0,i}$)
Amarillas redondas	315	$\frac{9}{16}$
Amarillas rugosas	101	$\frac{3}{16}$
Verdes redondas	108	$\frac{3}{16}$
Verdes rugosas	32	$\frac{1}{16}$

⁵Las estadísticas expresadas en 2.2 y 2.3 son asintóticamente equivalentes, una demostración de este hecho puede encontrarse en Kendall and Stuart volumen 2.

De donde se obtiene el valor de la estadística de Pearson

$$\begin{aligned}
 Q_n &= \frac{(315 - (556) \frac{9}{16})^2}{(556) \frac{9}{16}} + \frac{(101 - (556) \frac{3}{16})^2}{(556) \frac{3}{16}} + \\
 &\quad + \frac{(108 - (556) \frac{3}{16})^2}{(556) \frac{3}{16}} + \frac{(32 - (556) \frac{1}{16})^2}{(556) \frac{1}{16}} \\
 &\approx 0.47
 \end{aligned}$$

Por lo que de la tabla A.3 la prueba de tamaño $\alpha = 0.05$ de la hipótesis nula

$$H_0 : p_1 = \frac{9}{16}, p_2 = \frac{3}{16} = p_3, p_4 = \frac{1}{16}$$

rechaza H_0 si Q_n excede a $\chi^2_{1-\alpha}(k-1) = \chi^2_{(0.95,3)} = 7.815$.

Como $Q_n \approx 0.47 < 7.815 = \chi^2_{(0.95,3)}$. Las observaciones se ajustan con la hipótesis nula, es decir, existe un buen ajuste con el modelo Mendeliano.

Ejemplo 2.1.2 Prueba de uniformidad. Sea X_1, \dots, X_n una muestra aleatoria de una función de distribución continua F , considere el problema de probar la hipótesis nula simple $H_0 : F = F_0$ donde $F_0 \sim U(0, 1)$ contra cualquier otra distribución alternativa.

Para este problema se considera dividir el intervalo $(0,1)$ en k subintervalos de longitud $\frac{1}{k}$; para $i = 1, \dots, k$ sea n_i el número de observaciones X_i que caen en el intervalo $[\frac{i-1}{k}, \frac{i}{k})$

Entonces bajo H_0 la distribución conjunta de (n_1, \dots, n_k) es una multinomial basada en n ensayos y k categorías de igual probabilidad $\frac{1}{k}$. Por lo que para probar H_0 puede emplearse la prueba χ^2 que rechaza la hipótesis nula empleando la estadística de Pearson que en este problema toma la forma particular

$$Q_n = \sum_{i=1}^k \frac{(n_i - \frac{n}{k})^2}{\frac{n}{k}}$$

Este ejemplo es de interés particular debido a que dada X variable aleatoria continua con función de distribución $F_X(\cdot)$, si se emplea la transformación

$$Y = F_X(x)$$

entonces Y tiene una distribución uniforme en el intervalo $(0, 1)$.

Hipótesis compuestas

En este caso bajo la hipótesis nula H_0 la prueba de razón de verosimilitud permanece asintóticamente equivalente a la prueba ji-cuadrada de Pearson, por lo que limitaremos nuestra atención a ésta última.

Supongase que $F_0(x)$ no esta completamente especificada, es decir algunos o quizás todos los parámetros de la distribución no han sido especificados. Entonces en la formulación del problema para la distribución multinomial se tiene que las probabilidades hipotéticas $p_{0,i}$ no pueden ser calculadas de manera inmediata, debido a que estan en función de s (asumimos que $s < k-1$) parámetros no especificados $\theta := (\theta_1, \dots, \theta_s)$.

Con el propósito de calcular las probabilidades hipotéticas podríamos emplear algun vector $\hat{\theta}$ de estimadores de θ con lo que la estadística de Pearson (2.3) se reescribe

$$Q_n(\theta) = \sum_{i=1}^k \frac{(n_i - np_{0,i}(\theta))^2}{np_{0,i}(\theta)} \quad (2.4)$$

Fisher (1928) sugirió basar la prueba en la estadística $Q_n(\hat{\theta})$ donde $\hat{\theta}$ es el estimador máximo verosímil (o cualquier estimador eficiente) de θ bajo la hipótesis nula en el problema multinomial, posteriormente Neyman (1949) recomendó basar la prueba en $Q_n(\tilde{\theta})$ donde $\tilde{\theta}$ se elige de manera que minimice $Q_n(\theta)$, el caso típico es que si la hipótesis nula es verdadera

$$Q_n(\hat{\theta}) - Q_n(\tilde{\theta}) \xrightarrow{P} 0$$

Como en el caso de la hipótesis nula simple, si la hipótesis nula es verdadera entonces:

$$-2 \ln(\lambda) - Q_n(\hat{\theta}) \xrightarrow{P} 0$$

donde λ es la estadística de razón de verosimilitud.

Por lo que $Q_n(\hat{\theta})$ tendrá una distribución asintótica χ^2 con $(k-1) - s$ grados de libertad.

Ejemplo 2.1.3

Considere una muestra aleatoria X_1, \dots, X_n tomada de una distribución continua desconocida, se desea probar la hipótesis nula H_0 de que la muestra proviene de una distribución normal contra la alternativa H_1 de que la distribución no es normal.

Si se aplica la prueba χ^2 de bondad de ajuste a este problema debe dividirse la recta real en k subintervalos y contar el número de observaciones n_i muestrales que caen en el i -ésimo subintervalo con puntos terminales a_i y b_i $i = 1, \dots, k$.

Si H_0 es verdadera y μ, σ^2 representan la media y varianza de la distribución normal respectivamente ($\theta = (\mu, \sigma^2)$), entonces las probabilidades hipotéticas para el i -ésimo estan dadas por

$$p_{0,i}(\theta) = \int_{a_i}^{b_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \quad (2.5)$$

Con el objeto de calcular la estadística $Q_n(\theta)$ de (2.4) deben ser encontrados los estimadores máximo verosímiles $\hat{\mu}$ y $\hat{\sigma}^2$ usando los números n_1, \dots, n_k de observaciones en los diferentes subintervalos, es decir, $\hat{\mu}$ y $\hat{\sigma}^2$ serán los valores de μ y σ^2 que maximizan la función de verosimilitud

$$L(\theta) = \prod_{i=1}^k [p_{0,i}(\theta)]^{n_i} \quad (2.6)$$

debido a la naturaleza complicada de $p_{0,i}(\theta)$ en (2.5) usualmente requerirá un largo calculo numérico para determinar los valores de μ y σ^2 que maximizan $L(\theta)$. Por otro lado, se conoce que los estimadores máximo verosímiles de μ y σ^2 basados en las n observaciones muestrales X_1, \dots, X_n son la media \bar{X} y varianza muestral $\frac{S^2}{n}$.

Además si los estimadores empleados para calcular $Q_n(\theta)$ son aquellos que maximizan (2.6) entonces si H_0 es verdadera $Q_n(\theta)$ tendrá una distribución χ^2 con $k - 3$ grados de libertad.

Chernoff y Lehmann establecieron en 1954 que si se emplea en el cálculo de la estadística $Q_n(\theta)$ los estimadores máximo verosímiles basados en las observaciones muestrales, entonces $Q_n(\theta)$ no tendría una distribución asintótica χ^2 , sin embargo, mostraron que su distribución se encuentra acotada entre las distribuciones χ^2 con $k - 3$ grados de libertad y χ^2 con $k - 1$ grados de libertad⁶.

2.2 Prueba de Kolmogorov-Smirnov.

Suponga que X_1, \dots, X_n es una muestra aleatoria de alguna distribución continua con función de distribución $F(x)$ desconocida. Se define

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}$$

llamada la función de distribución empírica de los datos, $F_n(x)$ denota la proporción de X_i 's que son menores o iguales que x . Por otro lado $F(x)$ representa la probabilidad de que cualquier observación X_i sea menor o igual que x . Por tanto se sigue de la ley fuerte de los grandes números que

$$F_n(x) \xrightarrow{P} F(x) \quad x \in \mathbb{R}$$

⁶El resultado general afirma que cuando se emplean los estimadores máximo verosímiles basadas en las n observaciones muestrales la distribución de $Q_n(\theta)$ esta acotada entre las distribuciones χ^2 con $(k - 1) - s$ y $k - 1$ grados de libertad.

Otro fuerte resultado conocido como el teorema de Glivenko-Cantelli asegura en base a la distancia de Kolmogorov-Smirnov⁷ que $F_n(x)$ convergerá uniformemente a $F(x)$ para todo x y forma las bases para la prueba Kolmogorov-Smirnov de bondad de ajuste.

Considérese el problema de probar la hipótesis simple $H_0 : F = F_0$ contra $H_1 : F \neq F_0$, donde F_0 es una función de distribución continua específica. La estadística de prueba de Kolmogorov-Smirnov se define

$$D_n := \sup_x |F_n(x) - F_0(x)|$$

En otras palabras D_n es la máxima diferencia entre la función de distribución empírica y la función de distribución hipotética. Cuando H_0 es verdadera la distribución de probabilidad de D_n es independiente de la función de distribución particular F_0 ⁸, es decir la distribución de D_n es la misma para todas las posibles funciones de distribución continuas F_0 , por lo que se denomina libre de distribución. Se sigue del teorema de Glivenko-Cantelli que el valor de D_n tenderá a ser pequeño si H_0 es verdadera y D_n tenderá a ser grande si la función de distribución real F es diferente de F_0 . Por lo tanto un procedimiento razonable para este contraste será rechazar H_0 si D_n excede una constante c apropiada.

Es conveniente expresar el procedimiento en términos de $T_n = n^{\frac{1}{2}} D_n$ en lugar de D_n . Por lo que la prueba de Kolmogorov-Smirnov rechaza H_0 si $T_n > s_{n,1-\alpha}$ donde $s_{n,1-\alpha}$ es el $(1 - \alpha)$ -percentil de la distribución de T_n bajo la hipótesis nula y F_0 es la distribución uniforme sobre $(0, 1)$, sin embargo la forma exacta de esta distribución es difícil de expresar⁹. Kolmogorov (1933) mostró que si F_0 es continua T_n tiene una distribución límite continua y creciente en $(0, \infty)$. De manera más precisa, si H_0 es verdadera entonces para cada $t > 0$

$$\lim_{n \rightarrow \infty} P(T_n \leq t) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 t^2}$$

La tabla A.5 en el apéndice proporciona las probabilidades de T_n bajo la hipótesis nula cuando F_0 es continua.

⁷ Si F y G son dos funciones de distribución continuas la distancia de Kolmogorov-Smirnov entre F y G esta definida por $d_k(F, G) := \sup_x |F(x) - G(x)|$. Notese que $d_k(F, G) = 0$ si y solo si $F = G$.

⁸ Lehmann

⁹ Algunas aproximaciones para obtener esta distribución son discutidas en Durbin(1973) y la sección 4.3 de Gibbons y Chakraborti(1992). Valores de $s_{n,1-\alpha}$ han sido tabulados en Birnbaum(1952)

Ejemplo 2.2.1 Considere probar la hipótesis de que los siguientes valores constituyen una muestra aleatoria de una distribución uniforme en el intervalo $[0, 1]$.

0.42	0.06	0.88	0.40	0.90
0.38	0.78	0.71	0.57	0.66
0.48	0.35	0.16	0.22	0.08
0.11	0.29	0.79	0.75	0.82
0.30	0.23	0.01	0.41	0.09

Las 25 observaciones ordenadas de menor a mayor están representadas por y_1, \dots, y_{25} y son listadas en la siguiente tabla. Además se muestran los valores correspondientes a la función de distribución empírica $F_n(y_i)$ de la muestra y los de la distribución uniforme en $[0, 1]$.

i	y_i	$F_n(y_i)$	$F_0(y_i)$
1	0.01	0.04	0.01
2	0.06	0.08	0.06
3	0.08	0.12	0.08
4	0.09	0.16	0.09
5	0.11	0.20	0.11
6	0.16	0.24	0.16
7	0.22	0.28	0.22
8	0.23	0.32	0.23
9	0.29	0.36	0.29
10	0.30	0.40	0.30
11	0.35	0.44	0.35
12	0.38	0.48	0.38
13	0.40	0.52	0.40
14	0.41	0.56	0.41
15	0.42	0.60	0.42
16	0.48	0.64	0.48
17	0.57	0.68	0.57
18	0.66	0.72	0.66
19	0.71	0.76	0.71
20	0.75	0.80	0.75
21	0.78	0.84	0.78
22	0.79	0.88	0.79
23	0.82	0.92	0.82
24	0.88	0.96	0.88
25	0.90	1	0.90

Con ayuda de la tabla anterior encontramos que la mayor diferencia entre $F_n(x)$ y $F_0(x)$ ocurre en la observación 15 de donde se observa que $D_n = 0.60 - 0.42 = 0.18$. El ejemplo consta de 25 observaciones por lo que $T_n = n^{\frac{1}{2}} D_n = 0.9$.

De la tabla A.5 encontramos que $F(T_n) = F(0.9) = 0.60727$, de aquí que el área de la cola correspondiente al valor observado de T_n es 0.39273, por lo que no deberíamos rechazar la hipótesis nula a niveles α menores a 0.39273.

Es importante enfatizar que cuando el tamaño de la muestra es grande, hasta un pequeño valor del área de la cola correspondiente al valor observado de T_n no indica necesariamente que la verdadera función de distribución $F(x)$ sea muy diferente de la función de distribución hipotética $F_0(x)$, debido a que cuando n es grande hasta una pequeña diferencia entre $F(x)$ y $F_0(x)$ será suficiente para generar un valor grande de T_n , por lo tanto, antes de rechazar una hipótesis nula debería asegurarse que existe una función de distribución alternativa plausible con la cual la función de distribución empírica $F_n(x)$ proporcione un mejor ajuste.

Prueba de Kolmogorov Smirnov para dos muestras.

Considere un problema en el cual una muestra aleatoria de m observaciones X_1, \dots, X_m es tomada de una distribución $F(x)$ la cual es desconocida y una muestra aleatoria independiente de n observaciones Y_1, \dots, Y_n es tomada de otra distribución $G(x)$ también desconocida. Asumiremos que ambas distribuciones $F(x)$ y $G(x)$ son continuas y se desea probar la hipótesis de que estas funciones son idénticas sin necesidad de especificar su forma, es decir, se desea probar

$$\begin{aligned} H_0 &: F(x) = G(x) \text{ para toda } x \in \mathbb{R} \\ H_1 &: H_0 \text{ es falsa} \end{aligned}$$

Sea $F_m(x)$ la función de distribución empírica calculada a partir de las observaciones X_1, \dots, X_m y $G_n(x)$ la función de distribución empírica calculada a partir de las observaciones Y_1, \dots, Y_n , además se considera la siguiente estadística

$$D_{mn} := \sup_x |F_m(x) - G_n(x)|$$

Cuando la hipótesis nula es verdadera y $F(x)$ y $G(x)$ son funciones idénticas las funciones de distribución empírica $F_m(x)$ y $G_n(x)$ tenderán a estar muy cerca una de la otra. De hecho cuando H_0 es verdadera se sigue del teorema de Glivenko-Cantelli que

$$D_{mn} \xrightarrow{P} 0 \text{ cuando ambos } m \rightarrow \infty \text{ y } n \rightarrow \infty$$

Por lo tanto parece razonable utilizar un procedimiento de prueba que especifique rechazar H_0 cuando

$$\left(\frac{mn}{m+n} \right)^{\frac{1}{2}} D_{mn} \geq c$$

donde c es una constante apropiada.

La conveniencia de trabajar con la estadística $\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} D_{mn}$ en lugar D_{mn} radica en el siguiente resultado: para cada $t > 0$, si H_0 es verdadera entonces

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} P\left(\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} D_{mn} \leq t\right) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 t^2}$$

Ejemplo 2.2.2 Supóngase que 25 observaciones son seleccionadas al azar de una distribución $F(x)$ la cual es desconocida, los valores obtenidos se encuentran en la siguiente tabla

0.61	0.29	0.06	0.59	-1.73
-0.74	0.51	-0.56	-0.39	1.64
0.05	-0.06	0.64	-0.82	0.31
1.77	1.09	-1.28	2.36	1.31
1.05	-0.32	-0.40	1.06	-2.47

Supóngase además que 20 observaciones son seleccionadas al azar de otra distribución $G(x)$ también desconocida, los valores obtenidos se presentan en la siguiente tabla

2.20	1.66	1.38	0.20
0.36	0.00	0.96	1.56
0.44	1.50	-0.30	0.66
2.31	3.29	-0.27	-0.37
0.38	0.70	0.52	-0.71

Se desea probar la hipótesis de que $F(x)$ y $G(x)$ son funciones idénticas.

Las observaciones ordenadas y las funciones de distribución empíricas de las observaciones se muestran en la siguiente tabla:

i	x_i	$F_m(x_i)$	y_i	$G_n(y_i)$
1	-2.47	0.04	-0.71	0.05
2	-1.73	0.08	-0.37	0.10
3	-1.28	0.12	-0.30	0.15
4	-0.82	0.16	-0.27	0.20
5	-0.74	0.20	0.00	0.25
6	-0.56	0.24	0.20	0.30
7	-0.40	0.28	0.36	0.35
8	-0.39	0.32	0.38	0.40
9	-0.32	0.36	0.44	0.45
10	-0.06	0.40	0.52	0.50
11	0.05	0.44	0.66	0.55
12	0.06	0.48	0.70	0.60
13	0.29	0.52	0.96	0.65
14	0.31	0.56	1.38	0.70
15	0.51	0.60	1.50	0.75
16	0.59	0.64	1.56	0.80
17	0.62	0.68	1.66	0.85
18	0.64	0.72	2.20	0.90
19	1.05	0.76	2.31	0.95
20	1.06	0.80	3.29	1
21	1.09	0.84		
22	1.31	0.88		
23	1.64	0.92		
24	1.77	0.96		
25	2.36	1		

Calculando los valores absolutos de las diferencias entre las funciones de distribución empíricas se encuentra que $D_{mn} = 0.27$, de donde se tiene que $\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} D_{mn} = 0.9$. Como en el ejemplo 2.2.1 el área de la cola correspondiente al valor observado de la estadística $\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} D_{mn}$ es 0.39273, por lo que no debería rechazarse la hipótesis nula a niveles α menores a 0.39273.

La base de la prueba de Kolmogorov-Smirnov es una medida de discrepancia entre la función de distribución hipotética $F_0(x)$ y la función de distribución empírica $F_n(x)$. Cualquier estadística de este tipo se denomina estadística de función de distribución empírica. En particular si d es una métrica en el espacio de las funciones de distribución, cualquier estadística de la forma $d(F_n, F_0)$ es una estadística de función de distribución empírica, la elección $d = D_n$ lleva a la prueba Kolmogorov-Smirnov de bondad de ajuste.

Otra clase de estadísticas de función de distribución empíricas es proporcionada por la familia de estadísticas de Cramer-Von Mises

$$V_n = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 \psi(x) dF_0(x)$$

Haciendo $\psi(x) = 1$ se obtiene la estadística de Cramer-Von Mises, mientras que si

$$\psi(x) = [F_0(x)[1 - F_0(x)]]^{-1}$$

se obtiene la estadística de Anderson-Darling.

2.3 Tablas de contingencia.

Se entiende por datos categóricos aquellos datos que se presenten en forma de frecuencias que "caen" dentro de ciertas categorías o clases. Nuestra atención se limitará a problemas asociados con las relaciones estadísticas entre 2 o más variables expresadas en forma de categorías.

Una variable categórica puede ser simplemente una clasificación conveniente dentro de grupos de una variable medible. En otras palabras no siempre puede ser expresada en términos de una variable medible básica. Por ejemplo, podemos clasificar hombres por:

- a) Peso.
- b) Color de cabello.
- c) Su actriz favorita.

La clasificación propuesta en a) es un ejemplo de *variable medible* categorizada, ya que se basa en una unidad de medida básica (kilogramos o libras), lo cual no ocurre con las clasificaciones propuestas en b) y c). Cabe mencionar una diferencia adicional, el color de cabello puede ser expresado en orden de escala acorde a la pigmentación, de claro a oscuro, lo que no es posible en c).

Si bien podemos imponer varios tipos de clasificaciones a los nombres de las actrices, las actrices no están intrínsecamente ordenadas de ninguna manera. Se hace referencia al ejemplo b) como una categorización ordenada mientras que a c) como una no ordenada.

Existe un punto adicional a tener en consideración, ocurre en ocasiones que las dos variables a ser investigadas pueden ser simplemente la misma variable observada en diferentes ocasiones (por ejemplo antes y después de un evento) o dos muestras relacionadas (por ejemplo padres e hijos, marido y mujer, etc.) en cuyo caso se refiere a una categorización idéntica.

Considerese en primer lugar una población clasificada de acuerdo a la presencia o ausencia de un atributo. El problema más simple de interdependencia surge

cuando hay dos atributos A_1 y A_2 , por lo que hay cuatro posibles subgrupos en los que las observaciones pueden ser clasificadas

	A_2	A_2^c	total	(2.7)
A_1	A_1A_2	$A_1A_2^c$	n_{A_1}	
A_1^c	$A_1^cA_2$	$A_1^cA_2^c$	$n_{A_1^c}$	
total	n_{A_2}	$n_{A_2^c}$	n	

Ahora, considerense las frecuencias observadas de una muestra aleatoria X_1, \dots, X_n y supongase que bajo la distribución poblacional las probabilidades de dichas frecuencias son p_{11}, p_{12}, p_{21} y p_{22} respectivamente, dichas probabilidades se representan en la siguiente tabla

p_{11}	p_{12}	$p_{1\cdot}$
p_{21}	p_{22}	$p_{2\cdot}$
$p_{\cdot 1}$	$p_{\cdot 2}$	1

donde para $i = 1, 2$,

$$p_{i\cdot} = p_{i1} + p_{i2}, \quad p_{\cdot i} = p_{1i} + p_{2i} \quad \text{y} \quad p_{1\cdot} + p_{2\cdot} = 1 = p_{\cdot 1} + p_{\cdot 2}.$$

La tabla (2.7) será representará por

n_{11}	p_{12}	$n_{1\cdot}$
n_{21}	p_{22}	$n_{2\cdot}$
$n_{\cdot 1}$	$n_{\cdot 2}$	n

donde n_{ij} representa las frecuencias observadas en el (i, j) -ésimo grupo y

$$\sum_{i=1}^2 \sum_{j=1}^2 n_{ij} = n = \sum_{i=1}^2 n_{i\cdot} = \sum_{j=1}^2 n_{\cdot j}.$$

La distribución de las frecuencias muestrales esta dada por la distribución multinomial con función de verosimilitud

$$L = L(p_{11}, p_{12}, p_{21}, p_{22}) = \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!} p_{11}^{n_{11}} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}}$$

de donde los estimadores máximo verosímiles de las proporciones p_{ij} estan dados por

$$\hat{p}_{ij} = \frac{n_{ij}}{n} \tag{2.8}$$

Supongase que se desea probar la hipótesis de independencia en la tabla de 2x2, es decir, se desea probar

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j} \quad i, j = 1, 2 \tag{2.9}$$

o de manera equivalente

$$H_0 : p_{11}p_{22} = p_{12}p_{21}$$

En este caso H_0 es una hipótesis compuesta e impone una restricción, además tiene dos grados de libertad debido a que los valores de p_{21} y p_{22} pueden ser obtenidos a partir de los de p_{11} y p_{12} por medio de las expresiones

$$\begin{aligned} p_{21} &= \frac{p_{11}(1 - p_{11} - p_{12})}{p_{11} + p_{12}} \\ p_{22} &= \frac{p_{12}(1 - p_{11} - p_{12})}{p_{11} + p_{12}} \end{aligned} \quad (2.10)$$

Por lo que el logaritmo de la función de verosimilitud es proporcional a

$$\begin{aligned} \ln(L) &\propto n_{11} \ln(p_{11}) + n_{12} \ln(p_{12}) + n_{21} \ln(p_{21}) + n_{22} \ln(p_{22}) \\ &= n_{11} \ln(p_{11}) + n_{12} \ln(p_{12}) + \\ &\quad + n_{21} [\ln(p_{11}) + \ln(1 - p_{11} - p_{12}) - \ln(p_{11} + p_{12})] \\ &\quad + n_{22} [\ln(p_{12}) + \ln(1 - p_{11} - p_{12}) - \ln(p_{11} + p_{12})] \\ &= n_{.1} \ln(p_{11}) + n_{.2} \ln(p_{12}) + n_{.2} [\ln(1 - p_{1.}) - \ln(p_{1.})] \end{aligned}$$

Para estimar los parámetros se forman las ecuaciones

$$\begin{aligned} 0 &= \frac{\partial \ln(L)}{\partial p_{11}} = \frac{n_{.1}}{p_{11}} - n_{.2} \left[\frac{1}{1 - p_{1.}} + \frac{1}{p_{1.}} \right] = \frac{n_{.1}}{p_{11}} - \frac{n_{.2}}{p_{1.}(1 - p_{1.})} \\ 0 &= \frac{\partial \ln(L)}{\partial p_{12}} = \frac{n_{.2}}{p_{12}} - \frac{n_{.2}}{p_{1.}(1 - p_{1.})} \end{aligned}$$

Que proporcionan los estimadores máximo verosímiles bajo H_0

$$\hat{p}_{11} = \left(\frac{n_{1.}}{n} \right) \left(\frac{n_{.1}}{n} \right), \quad \hat{p}_{12} = \left(\frac{n_{1.}}{n} \right) \left(\frac{n_{.2}}{n} \right)$$

Empleando las ecuaciones en (2.10) se obtienen expresiones similares para p_{21} y p_{22}

$$\hat{p}_{21} = \left(\frac{n_{2.}}{n} \right) \left(\frac{n_{.1}}{n} \right), \quad \hat{p}_{22} = \left(\frac{n_{2.}}{n} \right) \left(\frac{n_{.2}}{n} \right)$$

Por lo que estimamos las probabilidades en cada celda por medio del producto de las frecuencias marginales proporcionales.

Sustituyendo los estimadores máximo verosímiles en la verosimilitud se tiene bajo H_0

$$L_0 \propto \frac{(n_{1.}n_{.1})^{n_{11}} (n_{1.}n_{.2})^{n_{12}} (n_{2.}n_{.1})^{n_{21}} (n_{2.}n_{.2})^{n_{22}}}{n^{2n}}$$

mientras que para el máximo no condicionado se emplean los estimadores en (2.8)

$$L_1 \propto \frac{(n_{11})^{n_{11}} (n_{12})^{n_{12}} (n_{21})^{n_{21}} (n_{22})^{n_{22}}}{n^n}$$

para formar la estadística de razón de verosimilitud

$$\lambda = \frac{L_0}{L_1} = \left(\frac{n_{1.}n_{.1}}{nn_{11}} \right)^{n_{11}} \left(\frac{n_{1.}n_{.2}}{nn_{12}} \right)^{n_{12}} \left(\frac{n_{2.}n_{.1}}{nn_{21}} \right)^{n_{21}} \left(\frac{n_{2.}n_{.2}}{nn_{22}} \right)^{n_{22}}$$

escribiendo

$$n\hat{p}_{ij} = \frac{n_{i.}n_{.j}}{n} = e_{ij} \quad (2.11)$$

se tiene

$$\lambda = \left(\frac{e_{11}}{n_{11}} \right)^{n_{11}} \left(\frac{e_{12}}{n_{12}} \right)^{n_{12}} \left(\frac{e_{21}}{n_{21}} \right)^{n_{21}} \left(\frac{e_{22}}{n_{22}} \right)^{n_{22}}$$

Se ha mencionado anteriormente (sección 1.5) que si la hipótesis nula es verdadera $-2 \ln(\lambda)$ se distribuye asintóticamente como una χ^2 con un grado de libertad.

Otra prueba de (2.9) puede obtenerse observando que la hipótesis nula H_0 implica un conjunto de frecuencias hipotéticas e_{ij} y que la prueba de independencia es equivalente a una prueba de bondad de ajuste entre las observaciones y estas frecuencias hipotéticas, por tanto podemos emplear la estadística de Pearson (2.3) en la forma

$$Q_n = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

la cual tiene también una distribución asintótica χ^2 con un grado de libertad.

Recordando que los valores esperados en las celdas de la tabla de 2x2 bajo la hipótesis de independencia de las dos variables categoricas estan dados por

$$e_{ij} = \frac{n_{i.}n_{.j}}{n} \quad i, j = 1, 2$$

dependen sólo de las cuatro frecuencias marginales y el tamaño de la muestra n , debería tomarse en cuenta la manera en la que fue formada la tabla y en particular la forma en que se generan las frecuencias marginales. Con n fijo se tienen tres posibilidades con respecto a las frecuencias marginales. Ambos conjuntos de frecuencias marginales pueden ser variables aleatorias o ambos conjuntos pueden ser fijados previamente (este caso es más raro en la práctica que los otros dos).

De manera alternativa un conjunto de frecuencias marginales puede ser fijo, debido a que una clasificación es únicamente una manera de etiquetar dos muestras (por ejemplo hombres y mujeres) los cuales son comparados respecto a otra clasificación (por ejemplo número de infectados y no infectados por cierta enfermedad). Si el número en las dos muestras se fija con anterioridad tenemos un conjunto de frecuencias marginales fijo y un conjunto variable. Cuando estamos comparando dos (o más) muestras con respecto a una característica, frecuentemente se refiere a esta prueba como una prueba de **homogeneidad** de dos (o k) muestras.

TABLAS DE $r \times c$.

Si se considera el caso más general en el cual dos variables son clasificadas dentro de dos o más categorías, se puede extender la notación para escribir la tabla de $r \times c$ en la forma

$$\begin{array}{cccc|c}
 n_{11} & n_{12} & \cdots & n_{1c} & n_{1\cdot} \\
 n_{21} & n_{22} & \cdots & n_{2c} & n_{2\cdot} \\
 \vdots & \vdots & \ddots & \vdots & \vdots \\
 n_{r1} & n_{r2} & \cdots & n_{rc} & n_{r\cdot} \\
 \hline
 n_{\cdot 1} & n_{\cdot 2} & \cdots & n_{\cdot c} & n
 \end{array} \tag{2.12}$$

La tabla (2.12) es llamada **tabla de contingencia** de $r \times c$. Los resultados de la sección previa pueden generalizarse para ser empleados en este problema.

En este problema se desea probar la hipótesis

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j} \quad \text{para } i = 1, \dots, r \text{ y } j = 1, \dots, c$$

en el caso de una prueba de independencia. Donde

$$p_{i\cdot} = \sum_{j=1}^c p_{ij}, \quad p_{\cdot j} = \sum_{i=1}^r p_{ij} \quad \text{y} \quad \sum_{i=1}^r \sum_{j=1}^c p_{ij} = 1$$

Y en el caso de una prueba de homogeneidad

$$H_0 : p_{1j} = p_{2j} = \cdots = p_{rj} \quad \text{para } j = 1, \dots, c$$

En ambos planteamientos se puede emplear la estadística de Pearson para probar la hipótesis nula, la estadística toma la forma

$$Q_n = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Ahora el número de parámetros a estimar cuando H_0 es verdadero serán $(r-1) + (c-1) = r + c - 2$, además debido a que la tabla de contingencia tiene rc celdas se tiene que cuando H_0 es verdadera y $n \rightarrow \infty$ la función de distribución de Q_n converge a una χ^2 con grados de libertad igual a

$$rc - 1 - (r + c - 2) = (r - 1)(c - 1).$$

Por lo que la prueba de tamaño α rechaza la hipótesis nula si

$$Q_n \geq \chi_{1-\alpha}^2 (r - 1)(c - 1).$$

Ejemplo 2.3.1 Chase y Dummer (1992) estudiaron las actitudes de los niños en edad escolar en Michigan. Entre la diversa información que se obtuvo acerca

de cada niño, se les cuestionó acerca de que era más importante para ellos: buenas calificaciones, habilidades atléticas o popularidad. La siguiente tabla muestra los resultados de las respuestas de 478 niños clasificados por sexo y su respuesta a la pregunta. Se desea probar la hipótesis de que la respuesta de los niños es independiente del sexo.

	Buenas calificaciones	Habilidad atlética	Popularidad	Total
niños	117	60	50	227
niñas	130	30	91	251
Total	247	90	141	478

De las observaciones se obtienen bajo la hipótesis nula de independencia los valores esperados

	Buenas calificaciones	Habilidad atlética	Popularidad
niños	117.299	42.741	66.960
niñas	129.701	47.259	74.040

Calculando el valor de la estadística de prueba se obtiene $Q_n \approx 21.46$ y de la tabla A.3 $\chi^2_{(0.995,2)} = 10.597$, por lo que se rechaza la hipótesis nula al nivel $\alpha = 0.005$. Con ayuda de algún paquete estadístico puede encontrarse que el área de la cola correspondiente es 0.000022. Por lo tanto sólo rechazaríamos H_0 para valores de $\alpha \geq 0.000022$.

Ejemplo 2.3.1 El estudio de Chase y Dummer (1992) planteado en el ejemplo anterior fue realmente obtenido de una de tres subpoblaciones de acuerdo a la zona de localización de las escuelas: rural, suburbana y urbana. La tabla siguiente muestra las respuestas a la pregunta de estudio clasificadas por la localización de la escuela. Se desea probar la hipótesis de que la distribución de las respuestas es la misma para los tres tipos de zona en que se localiza la escuela.

	Buenas calificaciones	Habilidad atlética	Popularidad	Total
Rural	57	42	50	149
Suburbana	87	22	42	151
Urbana	103	26	49	178
Total	247	90	141	478

Nuevamente de las observaciones se obtienen bajo la hipótesis nula de independencia los correspondientes valores esperados

	Buenas calificaciones	Habilidad atlética	Popularidad	Total
Rural	76.99	28.05	43.95	149
Suburbana	78.03	28.43	44.54	151
Urbana	91.98	33.52	52.51	178
Total	247	90	141	478

Calculando el valor de la estadística de prueba se obtiene $Q_n \approx 18.83$ y de la tabla A.3 $\chi^2_{(0.995,4)} = 14.860$, por lo que se rechaza la hipótesis nula

al nivel $\alpha = 0.005$. Como en el ejemplo anterior puede emplearse un paquete estadístico para determinar que el área de la cola correspondiente es 0.00085. Por lo tanto sólo rechazaríamos la hipótesis nula de homogeneidad para valores de $\alpha \geq 0.00085$.

2.4 Prueba Binomial.

Sea X_1, \dots, X_n una muestra aleatoria obtenida a partir de la repetición de ensayos Bernoulli independientes con probabilidad de éxito p .

Considérese el problema de probar

$$H_0 : p = p_0$$

donde $0 < p_0 < 1$ es algún valor específico.

Se define la estadística T mediante

$$T = \sum_{i=1}^n \phi_i \quad (2.13)$$

donde

$$\phi_i = \begin{cases} 1 & \text{si ocurre un éxito en el } i\text{-ésimo ensayo} \\ 0 & \text{si no ocurre un éxito en el } i\text{-ésimo ensayo} \end{cases}$$

En otras palabras la estadística T se define como el número de éxitos en los n ensayos Bernoulli, por lo que si H_0 es verdadera T tendrá una distribución binomial con parámetros n y p_0 .

Además $\frac{T}{n}$ es un estimador insesgado del valor verdadero del parámetro p . Por lo que si $p > p_0$ entonces $\frac{T}{n}$ tenderá a ser mayor que p_0 . De lo anterior se deduce una prueba unilateral para el contraste

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p > p_0 \quad (2.14)$$

la cual rechaza H_0 si $T \geq c_{(1-\alpha, n)}$, donde $c_{(1-\alpha, n)}$ es el $(1 - \alpha)$ -percentil de la distribución binomial con parámetros n y p_0 .

Aproximación para muestras grandes.

Si definimos

$$T^* = \frac{T - E_{p_0}(T)}{\sqrt{Var_{p_0}(T)}} = \frac{T - np_0}{\sqrt{np_0(1-p_0)}}$$

donde $E_{p_0}(T) = np_0$ y $Var_{p_0}(T) = np_0(1-p_0)$ son la media y la varianza de T respectivamente cuando $p = p_0$. Si $n \rightarrow \infty$ y H_0 es verdadera entonces T^* tendrá una distribución asintóticamente normal con media cero y varianza uno,

por lo que para muestras grandes puede emplearse la teoría de aproximación normal y basar la prueba de (2.14) en la estadística T^* , entonces la prueba rechaza H_0 si $T^* \geq Z_{1-\alpha}$ donde $Z_{1-\alpha}$ es el $(1 - \alpha)$ -percentil de la distribución normal estándar.

Ejemplo 2.4.1 Pruebas de diferencia sensitiva son usadas frecuentemente en control y evaluación de calidad. La prueba triangular (Bradley (1963)) es una prueba de diferencia sensitiva que proporciona una útil aplicación del modelo binomial. En su forma simple consiste en que a cada uno de los n panelistas se les presentan tres muestras de prueba en un orden aleatorio, se conoce de antemano que dos de las muestras son idénticas mientras que la restante es diferente. El panelista debe elegir supuestamente la muestra que es diferente en base a un atributo sensitivo específico.

Si los panelistas son expertos o jueces capacitados homogéneos, el experimento puede interpretarse como n ensayos Bernoulli independientes donde un "éxito" corresponde a una correcta identificación de la muestra diferente (si no fueran homogéneos la probabilidad de elegir la muestra correcta no sería la misma entre los panelistas, por lo que podría cuestionarse la validez de aplicar la prueba binomial).

Bajo la hipótesis de que no existen bases para discriminar entre las muestras la probabilidad p de éxito es $\frac{1}{3}$; mientras que valores de $p > \frac{1}{3}$ exhiben bases para la discriminación.

Byer Y Abrams (1953) consideraron pruebas triangulares de amargura en las que cada catador recibía tres copas, dos conteniendo la misma solución de quinina y la tercera una solución de quinina diferente.

En su primer prueba de amargura las soluciones contenían 0.0075% y 0.0050% de sulfato de quinina respectivamente. Las seis posibles presentaciones BAA, ABA, AAB, ABB, BAB y BBA donde A representa la concentración alta y B la concentración baja fueron aleatoriamente distribuidas entre los catadores. En 25 ensayos se encontraron 25 selecciones correctas y 25 incorrectas.

Considerando la prueba binomial para el contraste

$$H_0 : p = \frac{1}{3} \quad \text{vs} \quad H_1 : p > \frac{1}{3}$$

y empleando la aproximación para muestras grandes se observa que

$$T = 25, \quad E_{p_0}(T) = \frac{50}{3} = Var_{p_0}(T)$$

Por lo que $T^* \approx 2.042$. Dado $\alpha = 0.05$ se tiene de la tabla A.1 que $Z_{0.95} = 1.645 \leq T^*$, entonces la prueba rechaza H_0 al nivel $\alpha = 0.05$. Más aún, se determina que $Z_{0.97932} = 2.04$, entonces el nivel más pequeño al cual debe rechazarse H_0 en base a la muestra es $\alpha = 0.02068$.

El ejemplo (2.4.1) es una muestra de la aplicación de la prueba binomial a un experimento que no es precisamente una serie de repeticiones de ensayos Bernoulli, lo necesario para aplicar esta prueba puede enunciarse de la siguiente manera:

- a) El resultado de cada ensayo puede ser clasificado como "éxito" o "fracaso".
- b) La probabilidad de "éxito" denotada por p permanece constante de ensayo a ensayo.
- c) Los n ensayos son independientes.

Por lo que la distribución de la estadística T queda determinada por los supuestos a), b), c) y el valor específico $p = p_0$ más allá del supuesto respecto a la distribución de la cual se desprenden los datos, es por ello que se denomina libre de distribución.

De manera similar puede deducirse una prueba unilateral para el contraste

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p < p_0$$

basada en la estadística T , la cual rechaza H_0 si $T \leq c_{(\alpha, n)}$, donde $c_{(\alpha, n)}$ es el α -percentil de la distribución binomial con parámetros n y p_0 .

Y una prueba bilateral para el contraste

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0$$

que rechaza H_0 si $T \geq c_{(1-\alpha_1, n)}$ o $T \leq c_{(\alpha_2, n)}$, con $\alpha_1 + \alpha_2 = \alpha$.

Un caso particularmente importante en el que puede aplicarse la prueba binomial ocurre cuando se desea probar hipótesis acerca de la mediana θ desconocida de una población. La aplicación de la prueba binomial a este problema establece como estadística de prueba T al número de observaciones muestrales mayores que un valor hipotético específico digamos $\theta = \theta_0$. En este caso particular, se refiere a la estadística T como estadística del signo y a la prueba asociada como **prueba del signo**.

De manera alternativa puede definirse T^- como el número de "fallas" en los n ensayos Bernoulli, es decir T^- puede representarse mediante (2.13) reemplazando ϕ_i por $(1 - \phi_i)$, $i = 1, \dots, n$

Debido a que $T^- = n - T$ las pruebas expuestas en esta sección pueden basarse en T^- observando que bajo H_0 , $T^- \sim Bin(n, 1 - p_0)$.

2.5 Prueba de rangos de Wilcoxon.

Este procedimiento está diseñado para análisis estadístico en el que el interés primario se centra en el parámetro de localización (mediana) de una población. Pueden encontrarse dos tipos de datos para los cuales un análisis como éste es importante, el primero de ellos conocido como pares replicados representa parejas de observaciones de pre y postratamiento, en el cual se está interesado en una traslación en la localización hecha por el "tratamiento". El segundo tipo de dato consiste en observaciones provenientes de una población particular a cuyo parámetro de localización se desea hacer inferencia.

PARES REPLICADOS.

Supongase que se obtienen $2n$ observaciones, 2 observaciones en cada uno de los n sujetos.

sujeto i	x_i	y_i
1	x_1	y_1
2	x_2	y_2
\vdots	\vdots	\vdots
n	x_n	y_n

A continuación se enuncian los supuestos para la prueba de rangos de Wilcoxon.

1) Se define $Z_i = Y_i - X_i$ y se forma el modelo

$$Z_i = \theta + e_i \quad i = 1, \dots, n$$

donde los e_i 's son variables aleatorias no observables y el parámetro de interés θ es el efecto desconocido del "tratamiento".

2) Los e_i 's son mutuamente independientes.

3) Cada e_i proviene de una distribución continua (no necesariamente la misma para cada uno) simétrica respecto a cero.

Para probar la hipótesis nula

$$H_0 : \theta = 0$$

es decir, no existe efecto por parte del "tratamiento". Se forman las diferencias absolutas $|Z_1|, \dots, |Z_n|$. Se denota mediante R_i al rango de $|Z_i|$, $i = 1, \dots, n$ en el rango conjunto, del menor al mayor de $|Z_1|, \dots, |Z_n|$.

Se definen las variables indicadoras ϕ_i mediante

$$\phi_i = \begin{cases} 1 & \text{si } Z_i > 0 \\ 0 & \text{si } Z_i < 0 \end{cases}$$

Con los productos $R_1\phi_1, \dots, R_n\phi_n$ se define la estadística

$$T^+ = \sum_{i=1}^n R_i \phi_i \tag{2.15}$$

El producto $R_i\phi_i$ es conocido como el rango positivo de Z_i , el cual toma el valor de 0 si Z_i es negativo y es igual al rango de Z_i si es positivo. La estadística T^+ es la suma de los rangos positivos.

Sea N el número de Z 's positivas y sea que $r_1 < r_2 < \dots < r_N$ denoten los rangos de los valores absolutos de estos Z 's positivas. Entonces, para $\theta > 0$ estos rangos tenderán a ser más grandes que los correspondientes rangos de los valores absolutos de las Z 's negativas. Lo anterior sugiere rechazar H_0 en favor de $\theta > 0$ para valores grandes de $\sum_{i=1}^N r_i$, se observa que $T^+ = \sum_{i=1}^N r_i$ por lo que la prueba unilateral para el contraste de hipótesis

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta > 0 \quad (2.16)$$

rechaza H_0 si $T^+ \geq c_{(\alpha, n)}$ donde c es el α -percentil de la distribución de la estadística T^+ bajo H_0 .

La distribución de la estadística T^+ bajo la hipótesis nula puede obtenerse observando que cada una de los 2^n posibles resultados para (r_1, \dots, r_N) ocurre con probabilidad $(\frac{1}{2})^n$. La derivación de la distribución de T^+ bajo H_0 no depende de la clase de distribuciones a la que pertenecen los e 's, lo único que se requiere es que dichas distribuciones sean continuas y simétricas respecto a cero, por ello el procedimiento basado en la estadística T^+ de Wilcoxon es llamado libre de distribución.

Aproximación para muestras grandes.

Cuando la hipótesis nula H_0 es verdadera la estadística

$$\begin{aligned} T^* &= \frac{T^+ - E_{p_0}(T^+)}{\sqrt{Var_{p_0}(T^+)}} \\ &= \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \end{aligned}$$

tiene una distribución asintótica normal estándar cuando $n \rightarrow \infty$. Por lo que la prueba basada en la distribución asintótica de T^* rechaza H_0 si $T^* \geq Z_\alpha$.

Si existen valores para los que $Z_i = 0$ estos valores se descartan y se redefine n como el número de Z_i 's distintas de cero.

Si existen empates entre los $|Z|$'s para los que las Z_i 's son distintas de cero, el rango asignado a cada uno de las Z_i 's para las que existe el empate es el promedio de sus rangos, con objeto de calcular la estadística T^+ . Para la aproximación en muestras grandes se calcula T^* empleando promedios de rangos en los empates y remplazando la varianza por

$$Var_0(T^*) = \frac{1}{24} \left[n(n+1)2n+1 - \frac{1}{2} \sum_{j=1}^g t_j(t_j-1)(t_j+1) \right]$$

donde g es el número de grupos empatados y t_j es el tamaño del j -ésimo grupo empatado.

De manera análoga si $\theta < 0$ los rangos $r_1 < r_2 < \dots < r_N$ de los valores absolutos de las Z 's positivas tenderán a ser más pequeños que los rangos de los valores absolutos de las Z 's negativas, por lo que la prueba unilateral para el contraste de hipótesis

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta < 0 \quad (2.17)$$

rechaza H_0 si $T^+ \leq c_{(\alpha,n)}$ donde c es el α -percentil de la distribución de la estadística T^+ bajo H_0 .

Cuando H_0 es verdadera la distribución de T^+ es simétrica con respecto a su media $\frac{n(n+1)}{4}$, esto implica que

$$P(T^+ \geq x) = P\left(T^+ \leq \frac{n(n+1)}{2} - x\right)$$

para $x = 0, 1, \dots, \frac{n(n+1)}{2}$.

Entonces la prueba rechaza H_0 si

$$T^+ \leq \frac{n(n+1)}{2} - c_{(\alpha,n)}$$

Y la prueba bilateral del contraste

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \neq 0 \quad (2.18)$$

rechaza H_0 si

$$T^+ \geq c_{(\alpha_1,n)} \quad \text{o} \quad T^+ \leq \frac{n(n+1)}{2} - c_{(\alpha_2,n)}$$

con $\alpha_1 + \alpha_2 = \alpha$.

Ejemplo 2.5.1 Los datos mostrados en la siguiente tabla fueron obtenidos por Salsburg(1970)

Paciente i	X_i	Y_i
1	1.83	0.878
2	0.50	0.647
3	1.62	0.598
4	2.48	2.05
5	1.68	1.06
6	1.88	1.29
7	1.55	1.06
8	3.06	3.14
9	1.30	1.29

Valores del factor IV en la escala de depresión Hamilton

Los datos basados en nueve pacientes que recibieron el tranquilizante T, fueron tomados de un ensayo clínico doblemente ciego que envolvía dos tranquilizantes. La medida utilizada fue el factor IV (factor suicida) en la escala de depresión Hamilton (1960). El valor de X fue tomado en la primer visita del paciente después de iniciar la terapia, el valor de Y fue tomado en la segunda visita del paciente después de iniciar la terapia, los pacientes habían sido diagnosticados con una mezcla de ansiedad y depresión. En este ejemplo que envuelve tranquilizantes una mejora correspondería a una disminución en los valores del factor. Se llevará acabo la prueba unilateral expuesta en (2.16), los cálculos necesarios para formar la estadística T^+ en (2.15) se muestran a continuación

i	Z_i	$ Z_i $	R_i	ϕ_i	$R_i\phi_i$
1	-0.952	0.952	8	0	0
2	0.147	0.147	3	1	3
3	-1.022	1.022	9	0	0
4	-0.430	0.430	4	0	0
5	-0.620	0.620	7	0	0
6	-0.590	0.590	6	0	0
7	-0.490	0.490	5	0	0
8	0.080	0.080	2	1	2
9	-0.010	0.010	1	0	0

Por lo que se tiene

$$T^+ = \sum_{i=1}^9 R_i\phi_i = 5$$

De la tabla A.6 se encuentra para $\alpha = 0.049$

$$T^+ = 5 < 8 = 45 - 37 = \left(\frac{9(10)}{2}\right) - c_{(0.049,9)}$$

Por lo que se rechaza H_0 en favor de $\theta < 0$ al nivel $\alpha = 0.049$. Se observa que el nivel más bajo al cual debería rechazarse H_0 usando la prueba unilateral es $\alpha = 0.020$ debido a que

$$\left(\frac{9(10)}{2}\right) - c_{(0.020,9)} = 45 - 40 = 5$$

Si se emplea la aproximación para muestras grandes se tiene

$$T^* = \frac{5 - \left(\frac{9(10)}{4}\right)}{\sqrt{\frac{9(10)(19)}{24}}} = \frac{-17.5}{8.44} = -2.07$$

Por lo que de la tabla A.1 se encuentra que el nivel más bajo al cual se rechaza H_0 con la aproximación normal es $\alpha = 0.01923$. De aquí que las pruebas indican una fuerte evidencia de que el tranquilizante lleva a una mejora en el nivel del factor IV en la escala Hamilton.

La estadística T^+ de Wilcoxon es la suma de los rangos positivos de las Z_i 's positivas. Si se define T^- como la suma de los rangos de los valores absolutos de las Z_i 's negativas, se observa que T^- puede definirse mediante (2.15) reemplazando ϕ_i por $(1 - \phi_i)$ $i = 1, \dots, n$. Debido a que

$$T^- = \frac{n(n+1)}{2} - T^+$$

Las pruebas expuestas en esta sección pueden expresarse en términos de T^- .

Finalmente, para probar la hipótesis $H_0 : \theta = \theta_0$ donde $\theta_0 \neq 0$, se obtienen las observaciones modificadas

$$Z'_i = Z_i - \theta_0 \quad i = 1, \dots, n$$

y se calcula el valor de la estadística T^+ emplando entonces las observaciones modificadas.

Capítulo III.

Regresión lineal simple.

El término *regresión* tiene una historia interesante que data del trabajo de Sir Francis Galton en el siglo XIX. Galton, nieto de Erasmus Darwin y primo de Charles Darwin estudió medicina en Cambridge desempeñándose de manera competente y sin gran distinción. Después de recibir su herencia abandonó la carrera de medicina y se dedicó a observar el mundo, proceso durante el cual desarrolló una gran curiosidad. En 1853 recibió la medalla dorada de la Royal Geographical Society en reconocimiento a sus logros tras haber explorado África durante el periodo de 1850-1852.

Durante la década de 1860 emergieron los primeros signos de su habilidad e intereses en el campo de la estadística. Empleando los registros de todas las estaciones del clima de Europa de Diciembre de 1861 construyó intrincados mapas climáticos y de manera visual fue capaz de descubrir la existencia de anticiclones, sus mapas empleaban símbolos ingeniosos que mostraban de manera simultánea la dirección del viento, temperatura y presión barométrica de cada estación, fue un notable triunfo temprano del uso de métodos gráficos para el análisis de datos multivariados. Los resultados fueron publicados en *Meteorographica* (1863). Las publicaciones de Galton se concentraron posteriormente en materia de psicología, antropología, sociología y educación, aunque el tema dominante en su trabajo de 1865 fue el estudio de la herencia. La característica distintiva del estudio de Galton en herencia fue su naturaleza estadística.

En su autobiografía Galton nos cuenta de su primer encuentro con el método de Quetelet sobre el ajuste de curvas normales. Se encontró fascinado por la curiosa aparición de la ley teórica "*desviación a la media*" (o como se renombró posteriormente "regresión a la media") en situaciones tales como la relación entre las alturas de padres e hijos, en medidas de habilidades o de clasificación. Siguiendo a Quetelet, proponía que la semejanza de la apariencia de los datos a esta curva característica era una prueba de lo apropiado de clasificar los datos en un sólo grupo, mientras que, la no semejanza a esta curva indicaba que los datos no deberían ser tratados juntos (dentro de un mismo grupo). En su publicación *Hereditarius Genius* (1869) ilustró la forma de la curva mediante un diagrama de puntos en el que cada punto representaba la altura de una persona. Esta idea fue recurrente en el trabajo de Galton, aunque para 1875 planteó una manera diferente de mostrar los datos, ordenó los datos en forma creciente y graficó los valores de los datos contra sus rangos, Galton llamó a esta curva "*ojiva*" (curva que actualmente conocemos como la función inversa de la función de distribución acumulativa normal). El desarrollo del concepto de regresión y sus bases probabilísticas alcanzaron su plenitud en su libro *Natural Inheritance* (1889).

3.1 Definición.

La palabra *regresión* se emplea en estadística para representar una relación entre variables.

En particular una regresión lineal simple satisface una relación de la forma

$$Y_i = \alpha + \beta x_i + e_i \quad (3.1)$$

donde Y_i es una variable aleatoria y x_i es otra variable observable. Las cantidades α y β el intercepto y la pendiente de la regresión respectivamente se suponen parámetros fijos desconocidos y e_i es necesariamente variable aleatoria. Es práctica común suponer que $E(e_i) = 0$ (en caso contrario podría asignarse el exceso al valor del intercepto α), por lo tanto de (3.1) se tiene

$$E(Y_i) = \alpha + \beta x_i \quad (3.2)$$

En general la función que proporciona el valor esperado de Y en función de x es llamada la **función de regresión poblacional**, entonces la ecuación (3.2) define la función de regresión poblacional para el modelo lineal simple.

Uno de los principales propósitos de la regresión es hacer predicciones de los valores de Y_i a partir del conocimiento de x_i usando una relación como la enunciada en (3.2), debido a este conocimiento para hacer inferencia acerca de la relación entre Y_i y x_i (3.2) puede reescribirse

$$E(Y_i | x_i) = \alpha + \beta x_i \quad (3.3)$$

para reforzar el aspecto condicional.

En la práctica comunmente se interpreta diciendo que Y_i depende de x_i , por lo que es común referirse a Y_i como la variable "dependiente" y a x_i como la variable "independiente". Cabe mencionar que esta terminología resulta un poco confusa debido a que la palabra independiente tiene un significado diferente en la teoría estadística (las x_i 's no son necesariamente variables aleatorias, por lo que no pueden ser estocásticamente independientes). Por ello se emplea una terminología alternativa más descriptiva refiriéndose a Y_i como la variable de respuesta y a x_i como la variable predictora.

Cuando se hace referencia a una regresión que es lineal, puede entenderse que la esperanza condicional de Y dado $X = x$ es una función lineal de x .

El término regresión lineal se refiere a la linealidad de los parámetros, por ejemplo

$$E(Y_i | x_i) = \alpha + \beta e^{x_i} \quad \text{y} \quad E(\ln(Y_i) | x_i) = \alpha + \frac{\beta}{x_i}$$

especifican regresiones lineales, mientras que

$$E(Y_i | x_i) = \alpha + \beta^2 x_i$$

no especifica una regresión lineal debido a que el parámetro para la pendiente es un término cuadrático.

Cuando se emplea una ecuación como (3.3) se está suponiendo implícitamente que la regresión de Y en X es lineal, este supuesto puede no estar justificado debido a la falta de teoría subyacente que soporte dicha relación lineal; sin embargo, como una relación lineal es muy conveniente para usos prácticos estaríamos muy interesados en suponer que la regresión de Y en X puede ser aproximada de manera adecuada por una función lineal. Es decir, esperaríamos que

$$E(Y_i | x_i) \approx \alpha + \beta x_i$$

Cuando se realiza un análisis de regresión existen dos pasos en el análisis. El primer paso se encuentra totalmente orientado a los datos cuyo objetivo es resumir los datos observados. Este paso siempre se realiza dado que generalmente se calculan medias, varianzas o alguna otra estadística. Se debe tener en cuenta que el primer paso que es el "ajuste de datos" no envuelve inferencia estadística, puesto que no se realiza ningún supuesto acerca de la población, sólo nos interesan los datos disponibles.

El segundo paso en el análisis de regresión es la parte estadística en la cual la atención se concentra en hacer inferencia sobre la función de regresión poblacional, para realizarlo es necesario realizar supuestos acerca de la población.

En un problema de regresión lineal simple se tienen n pares de observaciones $(x_1, y_1), \dots, (x_n, y_n)$ los cuales no estarán exactamente sobre una línea recta, pero estaremos interesados en resumir la información muestral mediante el "ajuste" a una recta.

3.2 Método de los mínimos cuadrados ordinarios.

Es uno de los métodos más usados para ajustar una recta a un conjunto de puntos $(x_1, y_1), \dots, (x_n, y_n)$. Este método no considera ningún supuesto acerca de las observaciones, simplemente son consideradas como n pares ordenados en el plano.

Dada la recta $y = \alpha + \beta x$ se define la suma de cuadrados residual o suma de cuadrados del error (SSE) mediante

$$SSE = \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \quad (3.4)$$

La suma de cuadrados residual mide la distancia vertical de cada punto (x_i, y_i) a la recta $y = \alpha + \beta x$ y suma los cuadrados de estas distancias.

El método de mínimos cuadrados ordinarios consiste en minimizar la suma de cuadrados residual, aquellos valores $\hat{\alpha}$ y $\hat{\beta}$ que minimizan (3.4) reciben el nombre de estimadores mínimo cuadráticos de α y β respectivamente.

Derivando parcialmente SSE con respecto a α y β e igualando a cero se obtienen las siguientes ecuaciones

$$\begin{aligned}\frac{\partial SSE}{\partial \alpha} &= -2 \sum_{i=1}^n (y_i - (\alpha + \beta x_i)) = 0 \\ \frac{\partial SSE}{\partial \beta} &= -2 \sum_{i=1}^n (y_i - (\alpha + \beta x_i)) x_i = 0\end{aligned}$$

simplificando ambas ecuaciones se obtienen las denominadas *ecuaciones normales*

$$\begin{aligned}n\alpha + \beta \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ \alpha \sum_{i=1}^n x_i + \beta \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i\end{aligned}$$

Resolviendo las ecuaciones normales y definiendo las siguientes cantidades:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

como las *sumas de cuadrados* y

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

como la *suma de productos cruzados*, los estimadores mínimo cuadráticos $\hat{\alpha}$ y $\hat{\beta}$ están dados por

$$\hat{\beta} = \frac{s_{xy}}{S_{xx}}, \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad (3.5)$$

Debido a que no hay suposición alguna acerca del modelo probabilístico que genera las observaciones el método de mínimos cuadrados ordinarios debe ser considerado sólo como un método para ajustar una recta a un conjunto de datos, por lo que los valores en (3.5) pueden denominarse soluciones mínimo cuadráticas.

Ejemplo 3.2.1 Se desea establecer una función de gastos de materiales para una rama de la economía que cuenta con nueve empresas, los datos se presentan

en la siguiente tabla donde Y representa gastos en materiales y X la producción.

Empresa	Y (miles de pesos)	X (miles de pesos)
1	117.22	390.00
2	92.94	336.52
3	147.36	463.23
4	80.60	320.78
5	83.38	326.94
6	65.92	280.40
7	58.97	338.27
8	64.69	269.06
9	41.73	138.2

Con los datos anteriores se obtienen las estadísticas

$$\begin{aligned}\bar{x} &= 318.1556, & \bar{y} &= 83.6456 \\ S_{xx} &= \sum_{i=1}^9 (x_i - \bar{x})^2 = 63253.9740 \\ S_{xy} &= \sum_{i=1}^9 (x_i - \bar{x})(y_i - \bar{y}) = 20462.3103 \\ S_{yy} &= \sum_{i=1}^9 (y_i - \bar{y})^2 = 8321.8116\end{aligned}$$

De donde los estimadores mínimo cuadráticos son:

$$\hat{\beta} = \frac{20462.3103}{63253.9740} = 0.3235 \quad \text{y} \quad \hat{\alpha} = 83.6456 - (0.3235) 318.1556 = -19.276$$

Por lo que la recta ajustada es $\hat{y} = -19.276 + 0.3235x$

3.3 Estimadores lineales insesgados de varianza mínima.

Con objeto de encontrar estimadores con algunas propiedades óptimas se impondrán algunos supuestos al modelo de regresión lineal simple

- 1) Los valores y_1, \dots, y_n son observaciones provenientes de variables aleatorias no correlacionadas Y_1, \dots, Y_n .
- 2) Cada una de las variables aleatorias Y_1, \dots, Y_n tiene la misma varianza σ^2 .

Tales supuestos pueden expresarse de la siguiente manera. Supóngase que

$$Y_i = \alpha + \beta x_i + e_i \quad i = 1, \dots, n$$

donde e_1, \dots, e_n son variables aleatorias no correlacionadas tales que para $i = 1, \dots, n$

$$E(e_i) = 0 \quad \text{y} \quad \text{Var}(e_i) = \sigma^2$$

Los e_1, \dots, e_n son llamados errores aleatorios. Debido a que x_1, \dots, x_n son valores fijos conocidos se verifica

$$\begin{aligned} E(Y_i) &= \alpha + \beta x_i, & \text{Var}(Y_i) &= \text{Var}(e_i) = \sigma^2 \quad i = 1, \dots, n \\ \text{Cov}(Y_i, Y_j) &= 0 \quad \text{para } i \neq j \end{aligned} \quad (3.6)$$

Para estimar los parámetros α y β nos restringiremos a la clase de los estimadores lineales; es decir, estimadores de la forma

$$\sum_{i=1}^n d_i Y_i$$

donde d_1, \dots, d_n son constantes conocidas. De (3.6) se deriva que la varianza de cualquier estimador lineal esta dada por

$$\text{Var}\left(\sum_{i=1}^n d_i Y_i\right) = \sum_{i=1}^n d_i^2 \text{Var}(Y_i) = \sigma^2 \sum_{i=1}^n d_i^2 \quad (3.7)$$

Adicionalmente dentro de la clase de los estimadores lineales nos restringiremos a aquellos estimadores que sean insesgados, lo cual limita los posibles valores para d_1, \dots, d_n .

Comenzaremos obteniendo un estimador lineal insesgado para β el cual debe satisfacer

$$\begin{aligned} \beta &= E\left(\sum_{i=1}^n d_i Y_i\right) = E\left(\sum_{i=1}^n d_i (\alpha + \beta x_i + e_i)\right) \\ &= \alpha \sum_{i=1}^n d_i + \beta \sum_{i=1}^n d_i x_i \end{aligned}$$

igualdad que es cierta si

$$\sum_{i=1}^n d_i = 0 \quad \text{y} \quad \sum_{i=1}^n d_i x_i = 1 \quad (3.8)$$

Recordando que un estimador insesgado es el "mejor" si tiene la varianza más pequeña de entre todos los estimadores insesgados, puede pensarse en buscar el mejor estimador lineal insesgado para β ; es decir, un estimador lineal que cumpla las restricciones en (3.8) y que minimice (3.7). Para obtener tal estimador debe minimizarse $\sum_{i=1}^n d_i^2$ sujeto a las restricciones en (3.8). Empleando el método de los multiplicadores de Lagrange se debe minimizar

$$L = \sum_{i=1}^n d_i^2 - \lambda_1 \left(\sum_{i=1}^n d_i x_i - 1\right) - \lambda_2 \sum_{i=1}^n d_i$$

derivando parcialmente e igualando a cero se obtienen las ecuaciones

$$\begin{aligned}\frac{\partial L}{\partial d_i} &= 2d_i - \lambda_1 x_i - \lambda_2 = 0, \quad i = 1, \dots, n \\ \frac{\partial L}{\partial \lambda_1} &= -\sum_{i=1}^n d_i x_i + 1 = 0 \\ \frac{\partial L}{\partial \lambda_2} &= -\sum_{i=1}^n d_i = 0\end{aligned}\tag{3.9}$$

Sumando las primeras n ecuaciones y empleando que $\sum_{i=1}^n d_i = 0$ se obtiene

$$\lambda_1 \sum_{i=1}^n x_i + n\lambda_2 = 0$$

Entonces

$$\lambda_1 = -\frac{\lambda_2}{\bar{x}}\tag{3.10}$$

Y multiplicando la i -ésima ecuación en (3.9) por x_i y sumando las ecuaciones se obtiene

$$2\sum_{i=1}^n d_i x_i = \lambda_1 \sum_{i=1}^n x_i^2 + \lambda_2 \sum_{i=1}^n x_i$$

Dado que $\sum_{i=1}^n d_i x_i = 1$

$$2 = \lambda_1 \sum_{i=1}^n x_i^2 + \lambda_2 \sum_{i=1}^n x_i$$

Sustituyendo (3.10) en la ecuación anterior se encuentra que

$$\lambda_2 = \frac{2\bar{x}}{n\bar{x}^2 - \sum_{i=1}^n x_i^2} = -\frac{2\bar{x}}{S_{xx}}$$

valor que sustituido en (3.10) proporciona

$$\lambda_1 = \frac{2}{S_{xx}}$$

Por lo que, sustituyendo los valores obtenidos para λ_1 y λ_2 en (3.9) y resolviendo para d_i , $i = 1, \dots, n$ se obtiene

$$d_i = \frac{x_i - \bar{x}}{S_{xx}}$$

Por lo tanto

$$\hat{\beta} = \sum_{j=1}^n d_j y_j = \sum_{j=1}^n y_j \left(\frac{x_j - \bar{x}}{S_{xx}} \right) = \frac{\sum_{j=1}^n y_j x_j - n\bar{x}\bar{y}}{S_{xx}} = \frac{S_{xy}}{S_{xx}}$$

es el estimador lineal insesgado de varianza mínima del parámetro β , el cual coincide con el estimador mínimo cuadrático de β . El valor de la varianza de $\hat{\beta}$ es

$$Var(\hat{\beta}) = \sigma^2 \sum_{i=1}^n d_i^2 = \frac{\sigma^2}{S_{xx}}$$

De manera similar, el estimador lineal insesgado de varianza mínima de α debe satisfacer

$$\alpha = E\left(\sum_{i=1}^n d_i Y_i\right) = \alpha \sum_{i=1}^n d_i + \beta \sum_{i=1}^n d_i x_i$$

igualdad que es cierta si

$$\sum_{i=1}^n d_i = 1 \quad \text{y} \quad \sum_{i=1}^n d_i x_i = 0$$

y minimizar (3.7) sujeto a las restricciones anteriores. Empleando nuevamente el método de los multiplicadores de Lagrange se debe minimizar

$$L = \sum_{i=1}^n d_i^2 - \lambda_1 \left(\sum_{i=1}^n d_i - 1 \right) - \lambda_2 \sum_{i=1}^n d_i x_i$$

donde derivando parcialmente e igualando a cero se obtienen las ecuaciones

$$\begin{aligned} \frac{\partial L}{\partial d_i} &= 2d_i - \lambda_1 - \lambda_2 x_i = 0, \quad i = 1, \dots, n & (3.11) \\ \frac{\partial L}{\partial \lambda_1} &= -\sum_{i=1}^n d_i + 1 = 0 \\ \frac{\partial L}{\partial \lambda_2} &= -\sum_{i=1}^n d_i x_i = 0 \end{aligned}$$

Sumando las primeras n ecuaciones y empleando que $\sum_{i=1}^n d_i = 1$ se obtiene

$$n\lambda_1 + \lambda_2 \sum_{i=1}^n x_i = 2 \quad (3.12)$$

Multiplicando la i -ésima ecuación en (3.11) por x_i y sumando las ecuaciones se obtiene

$$2 \sum_{i=1}^n d_i x_i = \lambda_1 \sum_{i=1}^n x_i + \lambda_2 \sum_{i=1}^n x_i^2$$

Como $\sum_{i=1}^n d_i x_i = 0$, se obtiene

$$\lambda_1 = -\frac{\lambda_2 \sum_{i=1}^n x_i^2}{n\bar{x}}$$

y sustituyendolo en (3.12) proporciona

$$\lambda_2 = \frac{-2\bar{x}}{S_{xx}}$$

Al sustituir los valores obtenidos para λ_1 y λ_2 en (3.11) y resolviendo para d_i , $i = 1, \dots, n$ se obtiene

$$d_i = \frac{1}{nS_{xx}} \left(\sum_{i=1}^n x_i^2 - n\bar{x}x_i \right)$$

Por lo tanto

$$\hat{\alpha} = \sum_{j=1}^n d_j y_j = \bar{y} - \hat{\beta}\bar{x}$$

es el estimador lineal insesgado de varianza mínima del parámetro α , que también coincide con el estimador mínimo cuadrático de α . Para calcular la varianza de $\hat{\alpha}$ se observa que d_i puede reescribirse de la siguiente manera

$$d_i = \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}}$$

Entonces

$$Var(\hat{\alpha}) = \sigma^2 \sum_{i=1}^n d_i^2 = \sigma^2 \sum_{i=1}^n \left[\frac{1}{n^2} - \frac{2\bar{x}(x_i - \bar{x})}{S_{xx}} + \frac{\bar{x}^2(x_i - \bar{x})^2}{(S_{xx})^2} \right]$$

el segundo término en la suma se cancela y finalmente se obtiene

$$Var(\hat{\alpha}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2 S_{xx}}{(S_{xx})^2} \right] = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] = \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2$$

El resultado de que los estimadores obtenidos por el método de mínimos cuadrados ordinarios sean efectivamente los estimadores lineales insesgados de varianza mínima es conocido como el *teorema de Gauss-Markov*.

Hasta este momento sólo se han hecho supuestos acerca de los dos primeros momentos (media, varianza y covarianza) de los datos y se obtuvieron estimadores con propiedades estadísticas importantes (mínima varianza e insesgades) aunque sólo dentro de la clase de los estimadores lineales insesgados. A pesar de ello no pueden construirse pruebas de hipótesis o intervalos de confianza debido a que no se ha especificado completamente un modelo de probabilidad para los datos. A continuación se presenta el modelo de regresión lineal simple más común.

3.4 Regresión lineal simple bajo normalidad.

Los datos observados son los n pares $(x_1, y_1), \dots, (x_n, y_n)$ donde los valores de la variable predictora se consideran fijos y conocidos y los valores de la variable de respuesta y_1, \dots, y_n son observaciones de las variables aleatorias Y_1, \dots, Y_n las cuales se suponen independientes, más aún la distribución de las Y_i 's se supone normal, específicamente

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2) \quad i = 1, \dots, n$$

Por lo que la función de regresión poblacional es una función lineal de x , es decir, $E(Y | x) = \alpha + \beta x$ y todas las Y_i 's tienen la misma varianza.

Tales supuestos pueden expresarse de la siguiente manera

$$Y_i = \alpha + \beta x_i + e_i \quad i = 1, \dots, n \quad (3.13)$$

donde e_1, \dots, e_n llamados errores aleatorios son variables aleatorias independientes e idénticamente distribuidas con

$$e_i \sim N(0, \sigma^2) \quad i = 1, \dots, n$$

Este modelo es un caso especial del analizado en la sección anterior, que los errores aleatorios no estén correlacionados se sigue del supuesto de independencia, además la distribución de Y_i depende solamente de e_i por lo que $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$.

La función de densidad conjunta de Y_1, \dots, Y_n es el producto de las funciones de densidad marginales

$$\begin{aligned} f(\mathbf{Y}; \alpha, \beta, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2}{2\sigma^2}\right] \end{aligned}$$

Para obtener los estimadores máximo verosímiles de α, β y σ^2 se emplea el logaritmo de la función de verosimilitud

$$\ln(L(\alpha, \beta, \sigma^2; \mathbf{x}, \mathbf{y})) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2$$

Derivando parcialmente con respecto a cada uno de los parámetros e igualando a cero las ecuaciones resultantes se encuentra que los estimadores máximo

verosímiles de α , β y σ^2 deben satisfacer las tres ecuaciones

$$\begin{aligned}\sum_{i=1}^n (y_i - (\alpha + \beta x_i)) &= 0 \\ \sum_{i=1}^n (y_i - (\alpha + \beta x_i)) x_i &= 0 \\ \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 &= n\sigma^2\end{aligned}\tag{3.14}$$

Se observa que las dos primeras ecuaciones en (3.14) son las ecuaciones normales obtenidas bajo el método de mínimos cuadrados ordinarios, por lo tanto los estimadores mínimo cuadráticos

$$\hat{\beta} = \frac{s_{xy}}{S_{xx}} \quad \text{y} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

son también los estimadores máximo verosímiles de β y α bajo el supuesto de normalidad.

Finalmente, el estimador máximo verosímil de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2$$

la suma de cuadrados residual (SSE) evaluada en la recta de mínimos cuadrados dividido por el tamaño de la muestra.

En la sección anterior se mostró que $\hat{\alpha}$ y $\hat{\beta}$ son estimadores lineales insesgados de α y β ; sin embargo el estimador máximo verosímil $\hat{\sigma}^2$ no es un estimador insesgado de σ^2 . Para obtener el sesgo de $\hat{\sigma}^2$ se emplearán los siguientes resultados.

Teorema 3.4.1 Dadas X_1, \dots, X_n variables aleatorias. Entonces

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var} (X_i) + 2 \sum_{i < j} \text{Cov} (X_i, X_j)$$

Demostración:

$$\begin{aligned}
\text{Var} \left(\sum_{i=1}^n X_i \right) &= E \left[\left(\sum_{i=1}^n X_i - E \left(\sum_{i=1}^n X_i \right) \right)^2 \right] \\
&= E \left[\left(\sum_{i=1}^n (X_i - E(X_i)) \right)^2 \right] \\
&= E \left[\sum_{i=1}^n \sum_{j=1}^n (X_i - E(X_i)) (X_j - E(X_j)) \right] \\
&= \sum_{i=1}^n \sum_{j=1}^n E [(X_i - E(X_i)) (X_j - E(X_j))] \\
&= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \quad \blacksquare
\end{aligned}$$

Teorema 3.4.2 Sean X_1, \dots, X_n y Y_1, \dots, Y_m dos conjuntos de variables aleatorias y sean a_1, \dots, a_n y b_1, \dots, b_m dos conjuntos de constantes. Entonces

$$\text{Cov} \left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$$

Demostración: Sea $Z = \sum_{j=1}^m b_j Y_j$, entonces

$$\begin{aligned}
\text{Cov} \left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) &= \text{Cov} \left(\sum_{i=1}^n a_i X_i, Z \right) \\
&= E \left(\left(\sum_{i=1}^n a_i X_i \right) Z \right) - E \left(\sum_{i=1}^n a_i X_i \right) E(Z) \\
&= E \left(\sum_{i=1}^n a_i (X_i Z) \right) - \left(\sum_{i=1}^n a_i E(X_i) \right) E(Z) \\
&= \sum_{i=1}^n a_i [E(X_i Z) - E(X_i) E(Z)] = \sum_{i=1}^n a_i \text{Cov}(X_i, Z)
\end{aligned}$$

Ahora

$$\begin{aligned}
Cov(X_i, Z) &= Cov\left(X_i, \sum_{j=1}^m b_j Y_j\right) \\
&= E\left(X_i \left(\sum_{j=1}^m b_j Y_j\right)\right) - E(X_i) E\left(\sum_{j=1}^m b_j Y_j\right) \\
&= \sum_{j=1}^m b_j [E(X_i, Y_j) - E(X_i) E(Y_j)] = \sum_{j=1}^m b_j Cov(X_i, Y_j)
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
Cov\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) &= \sum_{i=1}^n a_i Cov(X_i, Z) \\
&= \sum_{i=1}^n a_i \sum_{j=1}^m b_j Cov(X_i, Y_j) \\
&= \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j) \quad \blacksquare
\end{aligned}$$

Corolario 3.4.3 Si Y_1, \dots, Y_n son variables aleatorias no correlacionadas con $Var(Y_i) = \sigma^2$, $i = 1, \dots, n$ y a_1, \dots, a_n y b_1, \dots, b_n son dos conjuntos de constantes, entonces

$$Cov\left(\sum_{i=1}^n a_i Y_i, \sum_{i=1}^n b_i Y_i\right) = \sum_{i=1}^n a_i b_i Var(Y_i)$$

De (3.13) se tiene

$$e_i = Y_i - \alpha - \beta x_i \quad i = 1, \dots, n$$

Se define el *residual de la regresión* mediante

$$\hat{e}_i = Y_i - \hat{\alpha} - \hat{\beta} x_i \quad i = 1, \dots, n \quad (3.15)$$

Por lo que

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta} x_i)\right)^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2$$

Además

$$\begin{aligned}
E(\hat{e}_i) &= E(Y_i - \hat{\alpha} - \hat{\beta} x_i) = E(Y_i) - E(\hat{\alpha}) - E(\hat{\beta}) x_i \\
&= (\alpha + \beta x_i) - \alpha - \beta x_i = 0
\end{aligned}$$

Empleando el teorema (3.4.1)

$$\begin{aligned}
Var(\hat{e}_i) &= E(\hat{e}_i^2) = Var(Y_i - \hat{\alpha} - \hat{\beta}x_i) \\
&= Var(Y_i) + Var(\hat{\alpha}) + x_i^2 Var(\hat{\beta}) \\
&\quad - 2Cov(Y_i, \hat{\alpha}) - 2x_i Cov(Y_i, \hat{\beta}) + 2x_i Cov(\hat{\alpha}, \hat{\beta})
\end{aligned} \tag{3.16}$$

Empleando el corolario (3.4.3) y el hecho de que $Cov(Y_i, Y_j) = 0$ para $i \neq j$

$$Cov(Y_i, \hat{\alpha}) = Cov\left(Y_i, \sum_{j=1}^n \left(\frac{1}{n} - \frac{\bar{x}(x_j - \bar{x})}{S_{xx}}\right) Y_j\right) \tag{3.17}$$

$$= \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}}\right) \sigma^2$$

$$Cov(Y_i, \hat{\beta}) = Cov\left(Y_i, \sum_{j=1}^n \left(\frac{x_j - \bar{x}}{S_{xx}}\right) Y_j\right)$$

$$= \left(\frac{x_i - \bar{x}}{S_{xx}}\right) \sigma^2$$

$$Cov(\hat{\alpha}, \hat{\beta}) = Cov\left(\sum_{j=1}^n \left(\frac{1}{n} - \frac{\bar{x}(x_j - \bar{x})}{S_{xx}}\right) Y_j, \sum_{j=1}^n \left(\frac{x_j - \bar{x}}{S_{xx}}\right) Y_j\right)$$

$$= \sum_{j=1}^n \left(\frac{1}{n} - \frac{\bar{x}(x_j - \bar{x})}{S_{xx}}\right) \left(\frac{x_j - \bar{x}}{S_{xx}}\right) \sigma^2$$

$$= \sigma^2 \left[\sum_{j=1}^n \left(\frac{x_j - \bar{x}}{nS_{xx}}\right) - \sum_{j=1}^n \frac{\bar{x}(x_j - \bar{x})^2}{(S_{xx})^2} \right] = -\frac{\bar{x}}{S_{xx}} \sigma^2$$

Sustituyendo los resultados anteriores en (3.16) se obtiene

$$\begin{aligned}
Var(\hat{e}_i) &= E(\hat{e}_i^2) = \sigma^2 + \frac{\sigma^2}{nS_{xx}} \sum_{j=1}^n x_j^2 + \frac{x_i^2}{S_{xx}} \sigma^2 - 2\sigma^2 \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}}\right) \\
&\quad - 2\sigma^2 \frac{x_i(x_i - \bar{x})}{S_{xx}} - 2\sigma^2 \frac{x_i \bar{x}}{S_{xx}} \\
&= \sigma^2 \left[\frac{n-2}{n} + \frac{1}{S_{xx}} \left(\frac{1}{n} \sum_{j=1}^n x_j^2 + x_i^2 - 2(x_i - \bar{x})^2 - 2x_i \bar{x} \right) \right]
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{1}{n} \sum_{i=1}^n E(\hat{e}_i^2) \\
&= \frac{1}{n} \sum_{i=1}^n \left[\frac{n-2}{n} + \frac{1}{S_{xx}} \left(\frac{1}{n} \sum_{j=1}^n x_j^2 + x_i^2 - 2(x_i - \bar{x})^2 - 2x_i \bar{x} \right) \right] \sigma^2 \\
&= \left[\frac{n-2}{n} + \frac{1}{nS_{xx}} \left(\sum_{j=1}^n x_j^2 + \sum_{i=1}^n x_i^2 - 2S_{xx} - 2n\bar{x}^2 \right) \right] \sigma^2 \\
&= \frac{n-2}{n} \sigma^2
\end{aligned}$$

debido a que $\sum x_i^2 - n\bar{x}^2 = S_{xx}$.

Por lo tanto el estimador insesgado de σ^2 comúnmente más empleado es

$$S^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{e}_i^2$$

Con el fin de desarrollar estimaciones y pruebas de hipótesis basadas en los estimadores obtenidos deben conocerse sus distribuciones muestrales, para ello se presentan algunos resultados sin demostración que serán empleados y una breve discusión de la representación matricial del modelo de regresión lineal simple.

Lema 3.4.4 Sean X_1, \dots, X_n variables aleatorias independientes tales que $X_j \sim N(\mu_j, \sigma_j^2)$ $j = 1, \dots, n$. Si para constantes a_{ij} y b_{rj} ($j = 1, \dots, n; i = 1, \dots, k; r = 1, \dots, m$) donde $k + m \leq n$ se definen

$$\begin{aligned}
U_i &= \sum_{j=1}^n a_{ij} X_j, \quad i = 1, \dots, k \\
V_i &= \sum_{j=1}^n b_{rj} X_j, \quad r = 1, \dots, m
\end{aligned}$$

Entonces las variables U_i y V_i son independientes si y sólo si $Cov(U_i, V_i) = 0$.

En la teoría de modelos lineales se encuentra el siguiente resultado.

Corolario 3.4.5 Si $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ entonces $\mathbf{x}^t \mathbf{A} \mathbf{x} \sim \chi_r^2$ si y sólo si \mathbf{A} es idempotente y de rango r .

Es decir, si \mathbf{x} es un vector aleatorio normal con vector de medias $\boldsymbol{\mu} = \mathbf{0}$ y matriz de covarianzas igual a la matriz identidad, entonces la forma cuadrática

$\mathbf{x}^t \mathbf{A} \mathbf{x}$ tendrá una distribución ji-cuadrada si la matriz \mathbf{A} es idempotente (es decir $\mathbf{A} \mathbf{A} = \mathbf{A}$) y el rango de la matriz \mathbf{A} determina los grados de libertad de la distribución ji-cuadrada.

El modelo considerado en (3.13) puede ser representado matricialmente mediante

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.18)$$

donde $\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ y $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ con $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Mediante el método de mínimos cuadrados ordinarios se obtiene el vector de estimadores

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

donde $(\mathbf{X}^t \mathbf{X})^{-1} = \frac{1}{nS_{xx}} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}$

que proporciona los estimadores mínimo cuadráticos obtenidos en (3.5)

$$\hat{\beta} = \frac{s_{xy}}{S_{xx}}, \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}.$$

El vector de los residuos de la regresión puede obtenerse de (3.18)

$$\begin{aligned} \mathbf{e} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} &= \mathbf{y} - \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} \\ &= (\mathbf{I} - \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t) \mathbf{y} = \mathbf{M} \mathbf{y} \end{aligned}$$

donde

$$\mathbf{M}_{ij} = \delta_{ij} - \frac{1}{nS_{xx}} \left[\sum x^2 - (x_i + x_j) \sum x + nx_i x_j \right]$$

$$\text{con } \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Puede mostrarse que la matriz \mathbf{M} es simétrica e idempotente y que $\mathbf{M} \mathbf{X} = \mathbf{0}$. Además empleando el hecho de que el rango de una matriz idempotente es igual

a su traza

$$\begin{aligned}
 \text{rango}(\mathbf{M}) &= \text{traza}(\mathbf{M}) = \sum_{i=1}^n \mathbf{M}_{ii} \\
 &= \sum_{i=1}^n \left[\delta_{ii} - \frac{1}{nS_{xx}} \left[\sum x^2 - 2x_i \sum x + nx_i^2 \right] \right] \\
 &= n - \frac{1}{nS_{xx}} \left[n \sum x^2 - 2 \sum x \sum_{i=1}^n x_i + n \sum_{i=1}^n x_i^2 \right] \\
 &= n - \frac{2nS_{xx}}{nS_{xx}} = n - 2
 \end{aligned}$$

Teorema 3.4.6 Considerando el modelo de regresión lineal simple (3.13) bajo el supuesto de normalidad, los estimadores $\hat{\alpha}$, $\hat{\beta}$ y S^2 tienen las siguientes distribuciones muestrales

$$\hat{\alpha} \sim N \left(\alpha, \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2 \right), \quad \hat{\beta} \sim N \left(\beta, \frac{\sigma^2}{S_{xx}} \right)$$

con

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{x}}{S_{xx}} \sigma^2$$

Además $(\hat{\alpha}, \hat{\beta})$ y S^2 son independientes y

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$$

Demostración: En la sección anterior se mostró que $\hat{\alpha}$ y $\hat{\beta}$ son estimadores lineales de α y β respectivamente, por lo que son funciones lineales de las Y_i 's las cuales tienen una distribución normal, por lo tanto $\hat{\alpha}$ y $\hat{\beta}$ siguen una distribución normal. Además se demostró que

$$\begin{aligned}
 E(\hat{\alpha}) &= \alpha, & \text{Var}(\hat{\alpha}) &= \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2 \\
 E(\hat{\beta}) &= \beta, & \text{Var}(\hat{\beta}) &= \frac{\sigma^2}{S_{xx}}
 \end{aligned}$$

Por lo que $\hat{\alpha}$ y $\hat{\beta}$ tienen las distribuciones enunciadas. Y de (3.17)

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{x}}{S_{xx}} \sigma^2$$

Con el fin de demostrar que $(\hat{\alpha}, \hat{\beta})$ y S^2 son independientes se observa que el residual \hat{e}_i de la regresión definido en (3.15) puede reescribirse como

$$\hat{e}_i = \sum_{j=1}^n [\delta_{ij} - (c_j + d_j x_i)] Y_j$$

con

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \quad c_j = \frac{1}{n} - \frac{\bar{x}(x_j - \bar{x})}{S_{xx}}, \quad \text{y} \quad d_j = \frac{(x_j - \bar{x})}{S_{xx}}$$

Recordando que $\hat{\alpha} = \sum_{j=1}^n c_j Y_j$ se tiene

$$\begin{aligned} Cov(\hat{\alpha}, \hat{e}_i) &= \sum_{j=1}^n c_j [\delta_{ij} - (c_j + d_j x_i)] Var(Y_j) \quad (\text{Por el corolario 3.4.3}) \\ &= \sigma^2 \left[c_i - \sum_{j=1}^n (c_j^2 + c_j d_j x_i) \right] \\ &= \sigma^2 \left[c_i - \left(\frac{1}{n} + \left(\frac{x_i - 2\bar{x}}{nS_{xx}} \right) \sum_{j=1}^n (x_j - \bar{x}) - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}} \right) \right] \\ &= \sigma^2 (c_i - c_i) = 0 \end{aligned}$$

ya que $\sum_{j=1}^n (x_j - \bar{x}) = 0$

Por otro lado $\hat{\beta} = \sum_{j=1}^n d_j Y_j$, entonces

$$\begin{aligned} Cov(\hat{\beta}, \hat{e}_i) &= \sum_{j=1}^n d_j [\delta_{ij} - (c_j + d_j x_i)] Var(Y_j) \quad (\text{Corolario 3.4.3}) \\ &= \sigma^2 \left[d_i - \sum_{j=1}^n (d_j c_j + d_j^2 x_i) \right] \\ &= \sigma^2 \left[d_i - \left(\frac{1}{nS_{xx}} \sum_{j=1}^n (x_j - \bar{x}) + \frac{(x_i - \bar{x})}{S_{xx}} \right) \right] \\ &= \sigma^2 (d_i - d_i) = 0 \end{aligned}$$

Por lo que se sigue del lema (3.4.4) que $\hat{\alpha}$ y $\hat{\beta}$ son independientes de S^2 .

El estimador máximo verosímil de σ^2 esta basado en la suma de los cuadrados de los residuos

$$\begin{aligned} \mathbf{e}^t \mathbf{e} &= (\mathbf{M}\mathbf{y})^t \mathbf{M}\mathbf{y} = [\mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})]^t \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\varepsilon}^t \mathbf{M}\boldsymbol{\varepsilon} \end{aligned}$$

Por lo que del corolario (3.4.5).

$$\frac{(n-2)S^2}{\sigma^2} = \frac{\mathbf{e}^t \mathbf{e}}{\sigma^2} = \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)^t \mathbf{M} \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)$$

tiene una distribución χ_{n-2}^2 debido a que $\text{rango}(\mathbf{M}) = n - 2$. ■

3.5 Inferencia en el modelo de regresión lineal simple.

Del teorema (3.4.6) se deduce que

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2}} \quad \text{y} \quad \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{S_{xx}}}}$$

tienen una distribución normal estándar y que $\frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$, por lo que las estadísticas

$$\frac{\hat{\alpha} - \alpha}{S \sqrt{\frac{1}{nS_{xx}} \sum_{i=1}^n x_i^2}} \tag{3.19}$$

y

$$\frac{\hat{\beta} - \beta}{S \sqrt{\frac{1}{S_{xx}}}} \tag{3.20}$$

tienen una distribución t de Student con $n - 2$ grados de libertad. Estas estadísticas son las que comúnmente se utilizan para hacer inferencia acerca de los parámetros α y β .

Usualmente existe mayor interés en el parámetro β que en α . El parámetro α es el valor esperado de la variable de respuesta Y cuando la variable predictora toma el valor de cero, es decir, $E(Y | x = 0) = \alpha$, dependiendo del problema esta cantidad puede no ser de interés, en particular puede que cero no sea un valor razonable para la variable predictora. Sin embargo, β es la razón de cambio de $E(Y | x)$ como función de x ; esto es, β expresa la cantidad que cambia $E(Y | x)$ cuando x varía una unidad. El valor $\beta = 0$ es de un interés particular debido a que expresa lo no existencia de una relación lineal entre la variable de respuesta y la variable predictora.

Para el contraste de hipótesis

$$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0 \tag{3.21}$$

la prueba de tamaño α que emplea la estadística en (3.20) rechaza H_0 si

$$\left| \frac{\widehat{\beta}}{S\sqrt{\frac{1}{S_{xx}}}} \right| > t_{(1-\frac{\alpha}{2}, n-2)}$$

Si se considera una prueba de razón de verosimilitud para el contraste de hipótesis en (3.21) se tiene $\theta = (\alpha, \beta, \sigma^2)$, la función de verosimilitud es

$$L(\theta; \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \right] \quad (3.22)$$

Y

$$\begin{aligned} \Theta_0 &= \{ \theta = (\alpha, \beta, \sigma^2) : \alpha \in \mathbb{R}, \beta = 0, \sigma^2 > 0 \} \\ \Theta &= \{ \theta = (\alpha, \beta, \sigma^2) : \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \sigma^2 > 0 \} \end{aligned}$$

Para encontrar $\sup_{\theta \in \Theta_0} L(\theta; \mathbf{y})$ sustituimos $\beta = 0$ en (3.22) con lo que se obtiene la función de verosimilitud

$$L(\alpha, \sigma^2; \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha)^2 \right]$$

que es la función de verosimilitud de una muestra de tamaño n de una distribución normal con media α y varianza σ^2 , por lo tanto los estimadores máximo verosímiles son

$$\tilde{\alpha} = \bar{y} \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\text{Entonces } \sup_{\theta \in \Theta_0} L(\theta; \mathbf{y}) = (2\pi\tilde{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}.$$

Por otro lado en la sección anterior se encontraron los estimadores máximo verosímiles de (3.22)

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x}, \quad \widehat{\beta} = \frac{S_{xy}}{S_{xx}}, \quad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - (\widehat{\alpha} + \widehat{\beta}x_i) \right)^2$$

$$\text{Entonces } \sup_{\theta \in \Theta} L(\theta; \mathbf{y}) = (2\pi\widehat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}.$$

Por lo tanto la estadística de razón de verosimilitud esta dada por

$$\lambda = \left(\frac{\widehat{\sigma}^2}{\tilde{\sigma}^2} \right)^{\frac{n}{2}}$$

En lugar de examinar λ se examinará $(n-2) \left(\lambda^{-\frac{2}{n}} - 1 \right)$ que es una función monótona de la estadística λ , por lo que proporcionará una prueba equivalente.

Se observa que

$$\begin{aligned} \lambda^{-\frac{2}{n}} - 1 &= \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} = \frac{S_{yy} - \sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2}{\sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2} \\ &= \frac{S_{yy} - \sum_{i=1}^n \left[(y_i - \bar{y}) - \hat{\beta} (x_i - \bar{x}) \right]^2}{\sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2} \\ &= \frac{2\hat{\beta}S_{xy} - \hat{\beta}^2 S_{xx}}{\sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2} = \frac{\hat{\beta}^2 S_{xx}}{\sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2} \end{aligned}$$

Por lo tanto

$$(n-2) \left(\lambda^{-\frac{2}{n}} - 1 \right) = \frac{\hat{\beta}^2 S_{xx}}{S^2} = \frac{\left(\hat{\beta}^2 S_{xx} \right)}{\frac{S^2}{\sigma^2}}$$

la cual bajo H_0 es la razón de dos variables aleatorias independientes con distribución ji-cuadrada divididas por sus respectivos grados de libertad, 1 en el caso del numerador y $n-2$ para el denominador. Entonces bajo H_0 $(n-2) \left(\lambda^{-\frac{2}{n}} - 1 \right)$ tiene una distribución F de Fisher con 1 y $n-2$ grados de libertad. La prueba de razón de verosimilitud rechaza H_0 si y sólo si $\lambda \leq c$ o de manera equivalente si

$$(n-2) \left(\lambda^{-\frac{2}{n}} - 1 \right) \geq (n-2) \left(c^{-\frac{2}{n}} - 1 \right) = c^*$$

Por lo tanto la prueba rechaza H_0 si

$$\frac{\hat{\beta}^2 S_{xx}}{S^2} \geq F_{(\alpha, 1, n-2)}$$

Nótese que $(n-2) \left(\lambda^{-\frac{2}{n}} - 1 \right)$ es el cuadrado de

$$\frac{\hat{\beta}}{S \sqrt{\frac{1}{S_{xx}}}}$$

De manera similar si el contraste de hipótesis es de la forma

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0$$

la prueba de tamaño α que emplea la estadística en (3.20) rechaza H_0 si

$$\left| \frac{\widehat{\beta} - \beta_0}{S \sqrt{\frac{1}{S_{xx}}}} \right| > t_{(1-\frac{\alpha}{2}, n-2)}$$

Esta prueba puede emplearse para obtener un intervalo de confianza para β

$$P \left(-t_{(1-\frac{\alpha}{2}, n-2)} < \frac{\widehat{\beta} - \beta_0}{S \sqrt{\frac{1}{S_{xx}}}} < t_{(1-\frac{\alpha}{2}, n-2)} \right) = 1 - \alpha$$

Después de algunas simplificaciones se obtiene

$$P \left(\widehat{\beta} - t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{\frac{S^2}{S_{xx}}} < \beta < \widehat{\beta} + t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{\frac{S^2}{S_{xx}}} \right) = 1 - \alpha$$

Por lo tanto $\left(\widehat{\beta} - t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{\frac{S^2}{S_{xx}}}, \widehat{\beta} + t_{(1-\frac{\alpha}{2}, n-2)} \sqrt{\frac{S^2}{S_{xx}}} \right)$ es un $100(1 - \alpha)\%$ intervalo de confianza para β .

Por otro lado, para el contraste de hipótesis

$$H_0 : \alpha = \alpha_0 \quad \text{vs} \quad H_1 : \alpha \neq \alpha_0$$

la prueba de tamaño α que emplea la estadística en (3.19) rechaza H_0 si

$$\left| \frac{\widehat{\alpha} - \alpha_0}{S \sqrt{\frac{1}{nS_{xx}} \sum_{i=1}^n x_i^2}} \right| > t_{(1-\frac{\alpha}{2}, n-2)}$$

Y un $100(1 - \alpha)\%$ intervalo de confianza para α basado en esta prueba es

$$\left(\widehat{\alpha} - t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{\frac{1}{nS_{xx}} \sum_{i=1}^n x_i^2}, \widehat{\alpha} + t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{\frac{1}{nS_{xx}} \sum_{i=1}^n x_i^2} \right)$$

Finalmente, recordando que $U = \frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$ independientemente del valor de σ^2 se sigue que U es una pivotal, por lo que para obtener un intervalo de confianza de σ^2

$$P \left(\chi_{\left(\frac{\alpha}{2}, n-2\right)}^2 \leq U \leq \chi_{\left(1-\frac{\alpha}{2}, n-2\right)}^2 \right) = 1 - \alpha$$

Simplificando

$$P \left(\frac{(n-2) S^2}{\chi^2_{(1-\frac{\alpha}{2}, n-2)}} \leq \sigma^2 \leq \frac{(n-2) S^2}{\chi^2_{(\frac{\alpha}{2}, n-2)}} \right) = 1 - \alpha$$

Por lo que $\left(\frac{(n-2)S^2}{\chi^2_{(1-\frac{\alpha}{2}, n-2)}}, \frac{(n-2)S^2}{\chi^2_{(\frac{\alpha}{2}, n-2)}} \right)$ es un $100(1 - \alpha)\%$ intervalo de confianza para σ^2 .

Ejemplo 3.5.1 Considerando los datos del ejemplo 3.2.1 se analizará el contraste de hipótesis

$$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0$$

y se determinará los intervalos de confianza para α, β y σ^2 .

Para el contraste de hipótesis de que la pendiente de la recta de regresión es cero se tiene

$$\begin{aligned} S^2 &= \frac{1}{n-2} \sum_{i=1}^n \left(y_i - (\hat{\alpha} + \hat{\beta}x_i) \right)^2 \\ &= \frac{1702.3675}{7} = 243.1954 \end{aligned}$$

Por lo que

$$\frac{\hat{\beta}}{S \sqrt{\frac{1}{S_{xx}}}} = \frac{0.3235}{15.5947 \sqrt{0.00001581}} = 5.217$$

De la tabla de la distribución t (A.2) se observa que al nivel $\alpha = 0.05$, $t_{(0.975, 7)} = 2.365 < 5.217$, por lo que la prueba rechaza la hipótesis nula al nivel $\alpha = 0.05$.

Los intervalos de confianza al 95% para α, β y σ^2 son respectivamente:

$$\begin{aligned} \left(-19.2760 - 36.882\sqrt{2.2003}, -19.2760 + 36.882\sqrt{2.2003} \right) &= (-73.985, 35.316) \\ \left(0.3235 - 2.365\sqrt{0.0038}, 0.3235 + 2.365\sqrt{0.0038} \right) &= (0.1777, 0.4693) \\ \left(\frac{7(243.1954)}{16.013}, \frac{7(243.1954)}{1.69} \right) &= (106.31, 1007.3) \end{aligned}$$

3.6 Coeficiente de determinación R^2 .

Para obtener una medida de la bondad de ajuste del modelo de regresión lineal simple se descompone la variación total de Y en dos partes: una debida a la relación lineal de Y con x y otra debida a causas no controlables.

Se observa que la desviación de un valor observado y_i con respecto a la media \bar{y} puede escribirse como

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

Elevando al cuadrado y sumando sobre todas las observaciones se obtiene

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i) (\hat{y}_i - \bar{y})$$

Por otro lado

$$\sum_{i=1}^n (y_i - \hat{y}_i) (\hat{y}_i - \bar{y}) = \sum_{i=1}^n (y_i - \hat{y}_i) \hat{y}_i - \bar{y} \sum_{i=1}^n (y_i - \hat{y}_i)$$

De la definición de los residuales de la regresión en (3.15) se observa que

$$\begin{aligned} \sum_{i=1}^n \hat{e}_i &= \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) \\ &= \sum_{i=1}^n y_i - n\hat{\alpha} - \hat{\beta} \sum_{i=1}^n x_i = 0 \end{aligned}$$

por la primera ecuación normal. Además

$$\begin{aligned} \sum_{i=1}^n \hat{e}_i x_i &= \sum_{i=1}^n (y_i - \hat{y}_i) x_i = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) x_i \\ &= \sum_{i=1}^n x_i y_i - \hat{\alpha} \sum_{i=1}^n x_i - \hat{\beta} \sum_{i=1}^n x_i^2 = 0 \end{aligned}$$

por la segunda ecuación normal. Finalmente

$$\begin{aligned} \sum_{i=1}^n \hat{e}_i \hat{y}_i &= \sum_{i=1}^n (y_i - \hat{y}_i) \hat{y}_i = \sum_{i=1}^n \hat{e}_i (\hat{\alpha} + \hat{\beta}x_i) \\ &= \hat{\alpha} \sum_{i=1}^n \hat{e}_i - \hat{\beta} \sum_{i=1}^n \hat{e}_i x_i = 0 \end{aligned}$$

Por lo tanto

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$SST = \sum_{i=1}^n (y_i - \bar{y})^2$ representa la variación total de las y 's y se denomina suma de cuadrados total.

$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ se había definido previamente como la suma de cuadrados residual o suma de cuadrados del error y representa la variación de las y 's debida a causas no controlables.

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ representa la variación de las y 's que es explicada por su relación lineal con x y se denomina suma de cuadrados debida a la regresión.

Dado que $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i = \bar{y} + \hat{\beta}(x_i - \bar{x})$ se tiene que

$$SSR = \hat{\beta}^2 S_{xx}$$

Se define el coeficiente de determinación R^2 mediante

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

El análisis de varianza en regresión consiste en descomponer la variación total de la variable de respuesta en varias partes llamadas fuentes de variación. Se ha mencionado que en el modelo de regresión lineal simple solo hay dos fuentes: una debida a la regresión y otra debida a causas no controlables. Cada variación es cuantificada por una suma de cuadrados y cada una de las sumas de cuadrados tiene una distribución ji-cuadrada ($SST \sim \chi_{n-1}^2$, $SSE \sim \chi_{n-2}^2$ y $SSR \sim \chi_1^2$).

La división de las sumas de cuadrados por sus grados de libertad se denomina *cuadrado medio*.

$\frac{SSR}{1}$: Cuadrado medio de regresión (MSR).

$\frac{SSE}{n-2}$: Cuadrado medio del error (MSE).

$\frac{SST}{n-1}$: Cuadrado medio del total (MST).

Se observa que si la hipótesis $H_0 : \beta = 0$ es verdadera $F = \frac{MSR}{MSE} = \frac{\hat{\beta}^2 S_{xx}}{S^2} \sim F_{(1, n-2)}$. La información anterior se exhibe en la tabla de análisis de varianza

Fuente de variación	Grados de libertad	Sumas de cuadrados	Cuadrados medios	F
Regresión	1	SSR	MSR	$\frac{MSR}{MSE}$
Error	$n - 2$	SSE	MSE	
Total	$n - 1$	SST		

Ejemplo 3.6.1 Tomando en cuenta los datos del ejemplo 3.2.1 se determina el coeficiente de determinación del modelo de regresión y se construye la tabla de análisis de varianza.

$$\begin{aligned}
 R^2 &= \frac{S_{xy}^2}{S_{xx}S_{yy}} \\
 &= \frac{(20462.3103)^2}{(63253.9740)(8321.8116)} \\
 &= 0.7954
 \end{aligned}$$

Por lo que el porcentaje de la variación total explicada por la regresión es del 79%.

La tabla de análisis de varianza correspondiente es

Fuente de variación	Grados de libertad	Sumas de cuadrados	Cuadrados medios	F
Regresión	1	6619.444	6619.444	27.219
Error	7	1702.368	243.195	
Total	8	8321.812		

Observando en la tabla A.4 que $F_{(0.95,1,7)} = 5.59 < 27.219$, se rechaza la hipótesis nula $H_0 : \beta = 0$ al nivel $\alpha = 0.05$

3.7 Predicción y estimación del valor medio de la variable de respuesta.

Supóngase que $(x_1, y_1), \dots, (x_n, y_n)$ satisfacen la condición de normalidad en el modelo de regresión lineal simple y que en base a estas n observaciones se obtienen los estimadores $\hat{\alpha}, \hat{\beta}$ y S^2 . Sea x_0 un valor específico de la variable predictora. Se considera en primera instancia estimar el valor medio de Y asociado a x_0 , es decir, se desea estimar $E(Y | x_0) = \alpha + \beta x_0$. Una elección que parece obvia para un estimador puntual es $\hat{\alpha} + \hat{\beta} x_0$. Este es un estimador insesgado debido a que

$$E(\hat{\alpha} + \hat{\beta} x_0) = E(\hat{\alpha}) + E(\hat{\beta}) x_0 = \alpha + \beta x_0$$

Empleando los resultados del teorema (3.4.6) se obtiene

$$\begin{aligned}
\text{Var}(\widehat{\alpha} + \widehat{\beta}x_0) &= \text{Var}(\widehat{\alpha}) + x_0^2 \text{Var}(\widehat{\beta}) + 2x_0 \text{Cov}(\widehat{\alpha}, \widehat{\beta}) \\
&= \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n x_i^2 + \frac{x_0^2 \sigma^2}{S_{xx}} - \frac{2x_0 \bar{x} \sigma^2}{S_{xx}} \\
&= \frac{\sigma^2}{S_{xx}} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 + \bar{x}^2 - 2x_0 \bar{x} + x_0^2 \right) \\
&= \frac{\sigma^2}{S_{xx}} \left(\frac{S_{xx}}{n} + (x_0 - \bar{x})^2 \right) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)
\end{aligned}$$

Debido a que $\widehat{\alpha}$ y $\widehat{\beta}$ son funciones lineales de Y_1, \dots, Y_n también lo es $\widehat{\alpha} + \widehat{\beta}x_0$. Por lo tanto

$$\widehat{\alpha} + \widehat{\beta}x_0 \sim N \left(\alpha + \beta x_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \right) \quad (3.23)$$

Del teorema (3.4.6) se tiene que $(\widehat{\alpha}, \widehat{\beta})$ y S^2 son independientes, por lo que S^2 también es independiente de $\widehat{\alpha} + \widehat{\beta}x_0$ y la estadística

$$\frac{\widehat{\alpha} + \widehat{\beta}x_0 - (\alpha + \beta x_0)}{S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t_{n-2}$$

es una pivotal mediante la cual se obtiene el $100(1 - \alpha)\%$ intervalo de confianza para $\alpha + \beta x_0$ dado por

$$\left(\widehat{\alpha} + \widehat{\beta}x_0 - t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \widehat{\alpha} + \widehat{\beta}x_0 + t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

la longitud del intervalo de confianza para $\alpha + \beta x_0$ depende de los valores x_1, \dots, x_n a través del valor $\frac{(x_0 - \bar{x})^2}{S_{xx}}$. Puede observarse que la longitud del intervalo será más corta si x_0 está próximo a \bar{x} y es mínima cuando $x_0 = \bar{x}$. Por lo que en el diseño de un experimento deberían elegirse valores x_1, \dots, x_n de tal manera que el valor x_0 (del cual se desea estimar el valor promedio) este próximo a \bar{x} .

Frecuentemente uno esta interesado en hacer inferencia en un solo valor no observado de la variable aleatoria Y y no un promedio de valores, tal inferencia es llamada *predicción*. Para ello supóngase que la nueva observación Y_0 tomada en $x = x_0$ tiene una distribución normal con media $\alpha + \beta x_0$ y varianza σ^2 ,

independiente de los datos previos $(x_1, y_1), \dots, (x_n, y_n)$. Como los estimadores $\hat{\alpha}, \hat{\beta}$ y S^2 se obtienen a partir de los datos previos serán independientes de Y_0 . Empleando (3.23) se observa que $Y_0 - (\hat{\alpha} + \hat{\beta}x_0)$ tiene una distribución normal con media

$$E\left(Y_0 - (\hat{\alpha} + \hat{\beta}x_0)\right) = \alpha + \beta x_0 - (\alpha + \beta x_0) = 0$$

y varianza

$$\begin{aligned} \text{Var}\left(Y_0 - (\hat{\alpha} + \hat{\beta}x_0)\right) &= \text{Var}(Y_0) - \text{Var}(\hat{\alpha} + \hat{\beta}x_0) \\ &= \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \end{aligned}$$

Debido a que S^2 es independiente de $Y_0 - (\hat{\alpha} + \hat{\beta}x_0)$ se observa que

$$\frac{Y_0 - (\hat{\alpha} + \hat{\beta}x_0)}{S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t_{n-2}$$

lo cual puede emplearse para obtener el $100(1 - \alpha)\%$ intervalo de predicción para la nueva observación Y_0

$$\left(\hat{\alpha} + \hat{\beta}x_0 - t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{\alpha} + \hat{\beta}x_0 + t_{(1-\frac{\alpha}{2}, n-2)} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

Ejemplo 3.7.1 Considerando nuevamente los datos del ejemplo 3.2.1, suponiendo que la producción es igual a 500 se calcula un intervalo de confianza para el valor medio y uno de predicción al 95%. Empleando la recta de regresión ajustada $\hat{y} = -19.276 + 0.3235x$ se estima el gasto de material en

$$Y_0 = -19.276 + (0.3235) 500 = 142.47$$

Por lo que el intervalo de confianza al 95% para el valor medio es

$$(142.47 - 36.882\sqrt{0.6656}, 142.47 + 36.882\sqrt{0.6656}) = (112.38, 172.56)$$

Y el intervalo de predicción al 95%

$$(142.47 - 36.882\sqrt{1.6656}, 142.47 + 36.882\sqrt{1.6656}) = (94.871, 190.07)$$

Apéndice de tablas.

A.1 Tabla de la distribución normal estándar.

$$P(Z \leq z)$$

z	0	1	2	3	4	5	6	7	8	9
-3	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
-1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
-0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
-0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
-0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
-0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
-0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414

A.1 Tabla de la distribución normal estándar (continuación).

		$P(Z \leq z)$									
z	0	1	2	3	4	5	6	7	8	9	
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	
1	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900	

A.2 Tabla de la distribución t de Student
 Proporciona el valor x tal que $P(X \leq x) = t(\alpha, n)$

grados de libertad	t(0.55)	t(0.60)	t(0.65)	t(0.70)	t(0.75)	t(0.80)	t(0.85)	t(0.90)	t(0.95)	t(0.975)	t(0.99)	t(0.995)	t(0.9995)
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646

A.3 Tabla de la distribución χ^2 .
 Proporciona el valor \mathbf{x} tal que $P(X \leq \mathbf{x}) = \chi^2(\alpha, n)$

grados de libertad	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	0.064	0.148	0.275	0.455	0.708	1.074	1.642	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.446	0.713	1.022	1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.005	1.424	1.869	2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.649	2.195	2.753	3.357	4.045	4.878	5.989	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.343	3.000	3.655	4.351	5.132	6.064	7.289	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.070	3.828	4.570	5.348	6.211	7.231	8.558	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	3.822	4.671	5.493	6.346	7.283	8.383	9.803	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	4.594	5.527	6.423	7.344	8.351	9.524	11.030	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.380	6.393	7.357	8.343	9.414	10.656	12.242	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.179	7.267	8.295	9.342	10.473	11.781	13.442	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	6.989	8.148	9.237	10.341	11.530	12.899	14.631	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	7.807	9.034	10.182	11.340	12.584	14.011	15.812	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	8.634	9.926	11.129	12.340	13.636	15.119	16.985	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	9.467	10.821	12.078	13.339	14.685	16.222	18.151	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	10.307	11.721	13.030	14.339	15.733	17.322	19.311	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.152	12.624	13.983	15.338	16.780	18.418	20.465	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.002	13.531	14.937	16.338	17.824	19.511	21.615	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	12.857	14.440	15.893	17.338	18.868	20.601	22.760	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	13.716	15.352	16.850	18.338	19.910	21.689	23.900	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	14.578	16.266	17.809	19.337	20.951	22.775	25.038	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	15.445	17.182	18.768	20.337	21.991	23.858	26.171	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	16.314	18.101	19.729	21.337	23.031	24.939	27.301	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	17.187	19.021	20.690	22.337	24.069	26.018	28.429	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	18.062	19.943	21.652	23.337	25.106	27.096	29.553	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	18.940	20.867	22.616	24.337	26.143	28.172	30.675	34.382	37.652	40.646	44.314	46.928
30	13.787	14.953	16.791	18.493	20.599	23.364	25.508	27.442	29.336	31.316	33.530	36.250	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	32.345	34.872	37.134	39.335	41.622	44.165	47.269	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	41.449	44.313	46.864	49.335	51.892	54.723	58.164	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	50.641	53.809	56.620	59.335	62.135	65.227	68.972	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	59.898	63.346	66.396	69.334	72.358	75.689	79.715	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	69.207	72.915	76.188	79.334	82.566	86.120	90.405	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	78.558	82.511	85.993	89.334	92.761	96.524	101.054	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	87.945	92.129	95.808	99.334	102.946	106.906	111.667	118.498	124.342	129.561	135.807	140.169

A.4 Tabla de la distribución F con $\alpha = 0.05$.

$$P(X \leq \mathbf{x}) \leq F_{(0.95, n, m)}$$

Grados de libertad del numerador (n).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	30	40	60	120
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95	248.01	250.10	251.14	252.20	253.25
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.45	19.46	19.47	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.66	8.62	8.59	8.57	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.80	5.75	5.72	5.69	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.56	4.50	4.46	4.43	4.40
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.87	3.81	3.77	3.74	3.70
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.44	3.38	3.34	3.30	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.15	3.08	3.04	3.01	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.94	2.86	2.83	2.79	2.75
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.77	2.70	2.66	2.62	2.58
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.65	2.57	2.53	2.49	2.45
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.54	2.47	2.43	2.38	2.34
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.46	2.38	2.34	2.30	2.25
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.39	2.31	2.27	2.22	2.18
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.33	2.25	2.20	2.16	2.11
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.28	2.19	2.15	2.11	2.06
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.23	2.15	2.10	2.06	2.01
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.19	2.11	2.06	2.02	1.97
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.16	2.07	2.03	1.98	1.93
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.12	2.04	1.99	1.95	1.90
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.10	2.01	1.96	1.92	1.87
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.07	1.98	1.94	1.89	1.84
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.05	1.96	1.91	1.86	1.81
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.03	1.94	1.89	1.84	1.79
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.01	1.92	1.87	1.82	1.77
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.93	1.84	1.79	1.74	1.68
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.84	1.74	1.69	1.64	1.58
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.75	1.65	1.59	1.53	1.47
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80	1.78	1.75	1.66	1.55	1.50	1.43	1.35

A.5 Tabla de T_n para la prueba de Kolmogorv Smirnov cuando F_0 es continua.

$$\lim_{n \rightarrow \infty} P(T_n \leq t) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 t^2}$$

t	F(t)	t	F(t)	t	F(t)	t	F(t)
0.30	0.00000	0.74	0.35598	1.18	0.87655	1.62	0.98949
0.31	0.00002	0.75	0.37283	1.19	0.88226	1.63	0.99015
0.32	0.00004	0.76	0.38964	1.20	0.88775	1.64	0.99078
0.33	0.00009	0.77	0.40637	1.21	0.89303	1.65	0.99136
0.34	0.00017	0.78	0.42300	1.22	0.89810	1.66	0.99192
0.35	0.00030	0.79	0.43951	1.23	0.90297	1.67	0.99244
0.36	0.00051	0.80	0.45586	1.24	0.90765	1.68	0.99293
0.37	0.00083	0.81	0.47204	1.25	0.91213	1.69	0.99339
0.38	0.00128	0.82	0.48803	1.26	0.91644	1.70	0.99382
0.39	0.00193	0.83	0.50381	1.27	0.92056	1.71	0.99423
0.40	0.00281	0.84	0.51937	1.28	0.92451	1.72	0.99461
0.41	0.00397	0.85	0.53468	1.29	0.92829	1.73	0.99497
0.42	0.00548	0.86	0.54975	1.30	0.93191	1.74	0.99531
0.43	0.00738	0.87	0.56455	1.31	0.93537	1.75	0.99563
0.44	0.00973	0.88	0.57907	1.32	0.93868	1.76	0.99592
0.45	0.01259	0.89	0.59332	1.33	0.94185	1.77	0.99620
0.46	0.01600	0.90	0.60727	1.34	0.94487	1.78	0.99646
0.47	0.02000	0.91	0.62093	1.35	0.94776	1.79	0.99670
0.48	0.02468	0.92	0.63429	1.36	0.95051	1.80	0.99693
0.49	0.03002	0.93	0.64734	1.37	0.95314	1.81	0.99715
0.50	0.03605	0.94	0.66008	1.38	0.95565	1.82	0.99735
0.51	0.04281	0.95	0.67252	1.39	0.95804	1.83	0.99753
0.52	0.05031	0.96	0.68464	1.40	0.96032	1.84	0.99771
0.53	0.05853	0.97	0.69645	1.41	0.96249	1.85	0.99787
0.54	0.06750	0.98	0.70794	1.42	0.96455	1.86	0.99802
0.55	0.07718	0.99	0.71913	1.43	0.96652	1.87	0.99817
0.56	0.08758	1.00	0.73000	1.44	0.96838	1.88	0.99830
0.57	0.09866	1.01	0.74057	1.45	0.97016	1.89	0.99842
0.58	0.11039	1.02	0.75083	1.46	0.97185	1.90	0.99854
0.59	0.12276	1.03	0.76078	1.47	0.97345	1.91	0.99864
0.60	0.13572	1.04	0.77044	1.48	0.97497	1.92	0.99874
0.61	0.14923	1.05	0.77979	1.49	0.97641	1.93	0.99884
0.62	0.16326	1.06	0.78886	1.50	0.97778	1.94	0.99892
0.63	0.17775	1.07	0.79764	1.51	0.97908	1.95	0.99900
0.64	0.19268	1.08	0.80613	1.52	0.98031	2.00	0.99933
0.65	0.20799	1.09	0.81434	1.53	0.98148	2.05	0.99955
0.66	0.22364	1.10	0.82228	1.54	0.98258	2.10	0.99971
0.67	0.23958	1.11	0.82995	1.55	0.98362	2.15	0.99981
0.68	0.25578	1.12	0.83736	1.56	0.98461	2.20	0.99988
0.69	0.27219	1.13	0.84450	1.57	0.98554	2.30	0.99995
0.70	0.28877	1.14	0.85140	1.58	0.98643	2.40	0.99998
0.71	0.30547	1.15	0.85804	1.59	0.98726	2.50	0.99999
0.72	0.32227	1.16	0.86444	1.60	0.98805	2.54	1.00000
0.73	0.33911	1.17	0.87061	1.61	0.98879		

A.6 Tabla de T^+ para la prueba de rangos de Wilcoxon.

Para una n específica, la entrada en la tabla para el valor \mathbf{x} es $P_0(T^+ \geq \mathbf{x})$.
 Bajo estas condiciones, si \mathbf{x} es tal que $P_0(T^+ \geq \mathbf{x}) = \alpha$, entonces $c_{(\alpha,n)} = \mathbf{x}$

x	n									
	3	4	5	6	7	8	9	10	11	12
3	0.625									
4	0.375									
5	0.250	0.563								
6	0.125	0.438								
7		0.313								
8		0.188	0.500							
9		0.125	0.406							
10		0.063	0.313							
11			0.219							
12			0.156	0.500						
13			0.094	0.422						
14			0.063	0.344						
15			0.031	0.281	0.531					
16				0.219	0.469					
17				0.156	0.406					
18				0.109	0.344					
19				0.078	0.289	0.527				
20				0.047	0.234	0.473				
21				0.031	0.188	0.422				
22				0.016	0.148	0.371				
23					0.109	0.320	0.500			
24					0.078	0.273	0.455			
25					0.055	0.230	0.410			
26					0.039	0.191	0.367			
27					0.023	0.156	0.326			
28					0.016	0.125	0.285	0.500		
29						0.008	0.098	0.248	0.461	
30							0.074	0.213	0.423	
31							0.055	0.180	0.385	
32							0.039	0.150	0.348	
33							0.027	0.125	0.312	
34							0.020	0.102	0.278	
35							0.012	0.082	0.246	
36							0.008	0.064	0.216	
37							0.004	0.049	0.188	
38								0.037	0.161	
39								0.027	0.138	
40								0.020	0.116	
41								0.014	0.097	
42								0.010	0.080	
43								0.006	0.065	
44								0.004	0.053	
45								0.002	0.042	
46									0.032	
47									0.024	
48									0.019	
49									0.014	
50									0.010	
51									0.007	
52									0.005	
53									0.003	
54									0.002	
55									0.001	

A.6 Tabla de T^+ para la prueba de rangos de Wilcoxon (continuación).
 Para una n específica, la entrada en la tabla para el valor \mathbf{x} es $P_0(T^+ \geq \mathbf{x})$.
 Bajo estas condiciones, si \mathbf{x} es tal que $P_0(T^+ \geq \mathbf{x}) = \alpha$, entonces $c_{(\alpha, n)} = \mathbf{x}$

		n									
x	11	12	13	14	15	x	11	12	13	14	15
33	0.517					77	0.011	0.013	0.068	0.180	
34	0.483					78	0.000	0.011	0.059	0.165	
35	0.449					79		0.009	0.052	0.151	
36	0.416					80		0.007	0.045	0.138	
37	0.382					81		0.005	0.039	0.126	
38	0.350	0.526				82		0.004	0.034	0.115	
39	0.319	0.496				83		0.003	0.029	0.104	
40	0.289	0.466				84		0.002	0.025	0.094	
41	0.260	0.436				85		0.002	0.021	0.084	
42	0.232	0.407				86		0.001	0.018	0.076	
43	0.207	0.378				87		0.001	0.015	0.068	
44	0.183	0.350				88		0.001	0.012	0.060	
45	0.160	0.322				89		0.000	0.010	0.053	
46	0.139	0.296	0.500			90		0.000	0.008	0.047	
47	0.120	0.270	0.473			91		0.000	0.007	0.042	
48	0.103	0.246	0.446			92			0.005	0.036	
49	0.087	0.223	0.420			93			0.004	0.032	
50	0.074	0.201	0.393			94			0.003	0.028	
51	0.062	0.181	0.368			95			0.003	0.024	
52	0.051	0.162	0.342			96			0.002	0.021	
53	0.042	0.144	0.318	0.500		97			0.002	0.018	
54	0.034	0.128	0.294	0.476		98			0.001	0.015	
55	0.027	0.113	0.271	0.452		99			0.001	0.013	
56	0.021	0.099	0.249	0.428		100			0.001	0.011	
57	0.016	0.087	0.227	0.404		101			0.000	0.009	
58	0.012	0.076	0.207	0.380		102			0.000	0.008	
59	0.009	0.066	0.188	0.357		103			0.000	0.006	
60	0.007	0.057	0.170	0.335	0.511	104			0.000	0.005	
61	0.005	0.050	0.153	0.313	0.489	105				0.004	
62	0.003	0.043	0.137	0.292	0.467	106				0.003	
63	0.002	0.037	0.122	0.271	0.445	107				0.003	
64	0.001	0.032	0.108	0.251	0.423	108				0.002	
65	0.001	0.028	0.095	0.232	0.402	109				0.002	
66	0.000	0.024	0.084	0.213	0.381	110				0.001	
67		0.021	0.073	0.196	0.360	111				0.001	
68		0.019	0.064	0.179	0.339	112				0.001	
69		0.017	0.055	0.163	0.319	113				0.001	
70		0.016	0.047	0.148	0.300	114				0.000	
71		0.014	0.040	0.134	0.281	115				0.000	
72		0.013	0.034	0.121	0.262	116				0.000	
73		0.013	0.029	0.108	0.244	117				0.000	
74		0.012	0.024	0.097	0.227	118				0.000	
75		0.012	0.020	0.086	0.211	119				0.000	
76		0.011	0.016	0.077	0.195	120				0.000	

Adaptada de la tabla A.4 de Nonparametric Statistical Methods, Hollander Myles (1973).

Conclusiones.

Aún cuando el número de pruebas expuestas en el presente trabajo es muy limitado comparado con la gran cantidad que se puede encontrar en la literatura, pienso que los alumnos de Actuaría, para quienes va dirigido el presente trabajo, así como todo aquel interesado en la teoría de pruebas de hipótesis, encontrarán en el texto una herramienta básica que les ayude a comprender las bases de dicha teoría.

Para la exposición de las pruebas, se hizo un mayor énfasis en la parte teórica por lo que sólo se presentaron algunos ejemplos numéricos con el fin ilustrar las pruebas expuestas dejando que el lector sea el que desarrolle los ejemplos que considere necesarios para afianzar sus conocimientos.

No puede dejar de mencionarse que existen diversos paquetes estadísticos (SPSS, STATISTICA, SAS, SPLUS, R,...) en los que se implementan, además de otras, las pruebas expuestas en este trabajo, los cuales son una herramienta indispensable para el desarrollo profesional de los futuros actuarios, de manera que durante la lectura del texto puede ser de gran utilidad que el lector se interese en la utilización de alguno de estos paquetes.

BIBLIOGRAFIA.

CALERO VINELO ARÍSTIDES (1998), Estadística Tomo II. IPN. México.

DEGROOT MORRIS H. and SCHERVISH MARK J. (2002), Probability and Statistics. Third edition. Addison Wesley.

G. CASELLA; R.L. BERGER (1996), Statistical Inference. Wadsworth & Brooks. USA.

GEISSER SEYMOUR (2006), Modes of Parametric Statistical Inference. Wiley series in probability and statistics. John Wiley & Sons.

GIRI NARAYAN C. (1993), Introducción to Probability and Statistics. Second edition. Statistics, textbooks and monographs. Marcel Dekker Inc. USA.

GREENE WILLIAM H. (1999), Análisis Econométrico. 3^a edición. Prentice Hall. España.

HOLLANDER MYLES (1973), Nonparametric Statistical Methods. John Wiley & Sons.

KENDALL MAURICE and STUART ALAN (1979), The Advanced Theory of Statistics volume 2. Fourth edition. Charles Griffin & Company Limited. London.

LEHMANN E.L. and ROMANO JOSEPH P. (2005), Testing Statistical Hypotheses. Third edition. Springer. USA.

L.R. MAYA P. y F.J. MARTÍN P. (2002), Fundamentos de Inferencia Estadística. AC Thopson. España.

MENDENHALL W., SCHEAFFER R., WACKERLY D. (2002), Estadística Matemática con aplicaciones. 6^a edición. Thopson. México.

MOOD A., GRAYBILL F. and BOES D. (1995), Introduction to Theory of Statistics. McGraw-Hill. USA.

NEAVE HENRY R. and WORTHINGTON PETER L. (1988), Distribution Free Tests. London UNWIN HYMAN.