Universidad Nacional Agraria La Molina
Departamento De Estadística E Informática
Profesor: Clodomiro Fernando Miranda Villagómez

PRUEBAS DE HIPÓTESIS

NOTA: Empezar esta Guía con el Script R hasta terminar el ejemplo de proporciones con datos pareados.

Tabla del ejemplo de prueba de hipótesis para proporciones con datos pareados

		Segundo Año	
	¿Café?	No	Si
Primer Año	No	174	21
	Si	5	50

Hipótesis Estadística
Es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias. El objetivo es rechazar la hipótesis nula.

Hipótesis Nula y Alternativa
Estamos interesados en un parámetro desconocido θ, el cual está dentro del espacio paramétrico Θ . $\theta \in \Theta$. Θ se puede particionar en dos sub conjuntos disjuntos, Θ_0 y Θ_1, tales que, $\Theta = \Theta_0 \cup \Theta_1$. Debemos decidir si $\theta \in \Theta_0$ o $\theta \in \Theta_1$. Sea H_0 la hipótesis $\theta \in \Theta_0$ y H_1 la hipótesis $\theta \in \Theta_1$, entonces $H_0 : \theta \in \Theta_0$ es la hipótesis nula y $H_1 : \theta \in \Theta_1$ la hipótesis alternativa.

Hipótesis Simple y Compuesta
Una hipótesis es simple cuando se especifica por completo la distribución.
Ejemplo: H: una v.a tiene una distribución N(μ=40, σ^2=64).
Una hipótesis es compuesta cuando no especifica por completo la distribución.
Ejemplo: H: una v.a tiene una distribución Poisson con λ>1.32.
Si $\Theta_0\left(\Theta_1\right)$ contiene un punto, $\Theta_0\left(\Theta_1\right)$ es llamada hipótesis simple, de otro modo $\Theta_0\left(\Theta_1\right)$ se llama hipótesis compuesta.

Datos
Asumimos que $X_i \sim f\left(x_i / \theta\right)$, $i = 1, 2, \ldots, n$, $\theta \in \Theta$. $\underline{x} = \left(x_1, \ldots, x_n\right)^t \in S$, donde S se llama espacio muestral.

Región Crítica o de Rechazo
Sea $C_1 \subseteq S$ tal que cuando $\left(x_1, \ldots, x_n\right)^t \in C_1$ se rechaza la H_0 y se acepta la H_1. Sea $C_0 \subseteq S$, $C_0 \cap C_1 = \phi$, y cuando $\left(x_1, \ldots, x_n\right)^t \in C_0$ no se rechaza la H_0. Sea $C_2 \subseteq S$, $C_0 \cap C_2 = C_1 \cap C_2 = \phi$ y cuando $\left(x_1, \ldots, x_n\right)^t \in C_2$ decidimos aleatoriamente rechazar H_0. $C_0 \cup C_1 \cup C_2 = S$. Notar que C_2 usualmente es el ϕ.

Función Prueba φ**:** $\varphi(x_1,\ldots,x_n)\in[0,1]$

Función Prueba No Aleatoria

$$\varphi(x_1,\ldots,x_n)=\begin{cases}1, & (x_1,\ldots,x_n)\in C_1\\ 0, & (x_1,\ldots,x_n)\in C_0\end{cases}$$

Se rechaza la H_0 si $\varphi=1$. No se rechaza la H_0 si $\varphi=0$. Notar que $C_2=\phi$ $(\phi$ es el vacío$)$.

Función Prueba Aleatoria

$$\varphi(x_1,\ldots,x_n)=\begin{cases}1, & (x_1,\ldots,x_n)\in C_1\\ \gamma, & (x_1,\ldots,x_n)\in C_2\\ 0, & (x_1,\ldots,x_n)\in C_0\end{cases}$$

Se rechaza la H_0 si $\varphi=1$. No se rechaza la H_0 si $\varphi=0$. Se rechaza la H_0 con probabilidad γ si $\varphi=\gamma$.

Ejemplo 1: Sea $X_1,\ldots,X_n$ una m.a de $N(\mu,\sigma^2)$ y sea S el espacio muestral de observaciones. Considerando Ho: $\mu\le 15$ y la prueba de hipótesis: rechazar Ho si y solo si $\overline{X}\ge 15+1.4\dfrac{\sigma}{\sqrt{n}}$, entonces la región de rechazo o región crítica es:

$$C_1=\left\{\underline{x}=(x_1,\ldots,x_n)\in S\;/\;\overline{X}\ge 15+1.4\frac{\sigma}{\sqrt{n}}\right\}$$

Ejemplo 2: Para el ejemplo anterior en que Ho: $\mu\le 15$:

$$\varphi(\underline{x})=\begin{cases}1 & \text{,si } \bar{x}\ge 15+1.4\sigma/\sqrt{n}\\ 0 & \text{,si } \bar{x}< 15+1.4\sigma/\sqrt{n}\end{cases}$$

Errores

Error Tipo I: Aceptar la H_1 cuando la H_0 es verdadera.

Error Tipo II: No rechazar la H_0 cuando la H_1 es verdadera

Sea $\theta_0\in\Theta_0$.

$$\begin{aligned}E_{\theta_0}\left[\varphi(x_1,\ldots,x_n)\right]&=1\times P_{\theta_0}\left[(x_1,\ldots,x_n)^t\in C_1\right]+\gamma\times P_{\theta_0}\left[(x_1,\ldots,x_n)^t\in C_2\right]\\&=P(\text{Error Tipo I})\end{aligned}$$

Sea $\theta_1\in\Theta_1$.

$$\begin{aligned}E_{\theta_1}\left[1-\varphi(x_1,\ldots,x_n)\right]&=(1-\gamma)\times P_{\theta_1}\left[(x_1,\ldots,x_n)^t\in C_2\right]+1\times P_{\theta_1}\left[(x_1,\ldots,x_n)^t\in C_0\right]\\&=P(\text{Error Tipo II})=\beta_\varphi(\theta_1)=\beta(\theta_1)=\beta\end{aligned}$$

	En Realidad

		H_o Verdadera	H_1 Verdadera $\left(H_o \text{ Falsa}\right)$
Decisión	No Rechazar H_o	Especificidad Verdadero Negativo $1-\alpha$	Error Tipo II Falso Negativo β
	Rechazar H_o	Error Tipo I Falso Positivo α	Potencia Sensibilidad Verdadero Positivo $1-\beta$

Nivel de Significación o Tamaño de la Prueba

$\alpha = \underset{\theta \in \Omega_0}{\text{Sup}}\left[E_\theta\left(\varphi\right)\right]$. Este es el límite superior mínimo de todas las probabilidades de error tipo I.

Potencia o Poder de la Prueba

Representa la probabilidad de rechazar correctamente la H_0 para un $\theta_1 \in \Theta_1$ dado.

$\text{Potencia} = 1-\beta$

Ejemplo 3: Se sabe que la longitud, en cm, de un artículo tiene distribución normal con $\sigma^2 = 64$. Se desea probar $H_0 : \mu = 120 \quad vs \quad H_1 : \mu < 120$, usando una m.a de tamaño 16 y se define la prueba estadística de la siguiente manera, rechazar la hipótesis nula si $\overline{X} < 115$ cm .

a. Halle la probabilidad de cometer error tipo I.

Solución

X = Longitud, $X \sim N\left(\mu, 64\right) \rightarrow \overline{X} \sim N\left(\mu, \frac{64}{16}\right) \sim N\left(\mu, 4\right)$

$$\alpha = P\left(\text{Error tipo } I\right) = P_{\mu=120}\left(\overline{X} < 115\right) = P\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} < \frac{115-120}{2}\right) = P\left(Z \langle < -2.5\right) = 0.0062$$

Utilizando el software R se concluye que a medida que aumenta el tamaño de muestra disminuye el nivel de significación. Se debe mencionar que el nivel de significación lo fija y controla el investigador.

```
n1=14:20
pnorm((115-120)/(64/n1)^.5)
```

Ver el Script R.

```
#0.009679734 0.007747147 0.006209665 0.004983948 0.004004971
#0.003221782 0.002594304
```

b. Determine el tamaño de muestra que se debe considerar, si se desea tener una probabilidad de cometer error tipo II de 0.0197, cuando en verdad $\mu = 112$.

Solución

$$\beta = P_{\mu=112}\left(\bar{X} \geq 115\right) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \geq \frac{115-112}{8/\sqrt{n}}\right) = P\left(Z \geq \frac{3}{8}\sqrt{n}\right) \quad \to P\left(Z < \frac{3}{8}\sqrt{n}\right) = 0.9803$$

$$\to \frac{3}{8}\sqrt{n} = Z_{(0.9803)} = 2.06 \to n = 30.17 \cong 31$$

Utilizando el software R se concluye que a medida que aumenta el tamaño de muestra disminuye la probabilidad de cometer el error tipo II. Se debe mencionar que la probabilidad de cometer el error tipo II no la controla el investigador pero con muestras grandes se reduce dicho error.

```
n2=25:35
pnorm((115-112)/(64/n2)^.5,lower.tail = F)

#0.03039636 0.02792961 0.02567417 0.02361045 0.02172092 0.01998980
#0.01840290 0.01694743 0.01561181 0.01438561 0.01325936
```

Ejemplo 4: Suponga que el fabricante de un nuevo medicamento para la gastritis afirma que el 90% de los pacientes a quienes se les da el medicamento se recuperan. Un investigador sospecha que menos del 90% de pacientes se recuperan. El investigador selecciona una muestra de 25 pacientes y manifiesta que, si menos de 20 de ellos se recuperan, se quejará ante la INDECOPI. Calcule la probabilidad de cometer error tipo I y luego la probabilidad de cometer error tipo II (asumiendo que el 60% se recupera).

$H_0 : \theta = 0.9$ vs $H_1 : \theta < 0.9$

Sea $X_1, \ldots, X_{25}$ una muestra aleatoria de tamaño 25 de una distribución $\text{Binomial}\left(1, \theta\right)$ entonces $T = \sum_{i=1}^{25} X_i \sim \text{Binomial}\left(25, \theta\right)$

$$\alpha = P\left(\text{Error tipo I}\right) = P\left(T < 20 / \theta = 0.9\right) = \sum_{t=0}^{19} \binom{25}{t} \left(0.9\right)^t \left(0.1\right)^{25-t} = 0.0334$$

Hay una probabilidad de 0.0334 de que el investigador se queje innecesariamente ante la INDECOPI.

$$\beta = P\left(\text{Error tipo II}\right) = P\left(T \geq 20 / \theta = 0.6\right) = \sum_{t=20}^{25} \binom{25}{t} \left(0.6\right)^t \left(0.4\right)^{25-t} = 0.0294$$

Hay una probabilidad de 0.0294 de que el investigador no se queje ante la INDECOPI, pero debería hacerlo.

Se ha comparado una hipótesis simple vs otra hipótesis simple. La prueba estadística no aleatoria es:

$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } T < 20 \\ 0 & ,\text{si } T \geq 20 \end{cases}$$

Por ejemplo, si T=18 entonces se rechaza la hipótesis nula y el pvalor será $\sum_{t=0}^{18} \binom{25}{t} (0.9)^t (0.1)^{25-t} = 0.009476 < \alpha$.

Ejemplo 5: La edad de los estudiantes universitarios de sexto ciclo de cierta universidad tiene distribución $N(\mu, \sigma^2 = 1 \text{ año}^2)$. Se desea probar $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$ con$(\mu_1 > \mu_0)$.

a. Halle k tal que $\bar{X} > k$ proporcione una región crítica de tamaño $\alpha = 0.05$ para una muestra aleatoria de tamaño n.

$$\alpha = P(\bar{X} > k / \mu_0) = P\left[\frac{(\bar{X} - \mu_0)}{\sigma}\sqrt{n} > \frac{(k - \mu_0)}{1}\sqrt{n}\right] = P\left[Z > (k - \mu_0)\sqrt{n}\right] = 0.05$$

$$\rightarrow (k - \mu_0)\sqrt{n} = 1.64 \rightarrow k = \mu_0 + \frac{1.64}{\sqrt{n}}$$

Se ha comparado una hipótesis simple vs otra hipótesis simple. La prueba estadística no aleatoria es:

$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } \bar{X} > k \\ 0 & ,\text{si } \bar{X} \leq k \end{cases}$$

b. Halle el mínimo tamaño de muestra necesario para probar $H_0 : \mu = 20$ vs $H_1 : \mu = 21$, considere una probabilidad de cometer error tipo II de a lo más 0.05.

$$\beta = P(\text{Error tipo II}) = P\left(\bar{X} < \mu_0 + \frac{1.64}{\sqrt{n}} / \mu_1\right) = P\left(\bar{X} < 20 + \frac{1.64}{\sqrt{n}} / \mu = 21\right) =$$

$$\beta = P\left[\frac{(\bar{X} - \mu_1)}{\sigma}\sqrt{n} < \frac{\left(20 + \frac{1.64}{\sqrt{n}} - 21\right)}{1}\sqrt{n}\right] = P\left[Z < -\sqrt{n} + 1.64\right] = 0.05$$

$$\rightarrow -\sqrt{n} + 1.64 = -1.64 \rightarrow n = 10.8 \approx 11$$

Ejercicio: Un lote de 6 lavadoras contiene un número A desconocido de lavadoras defectuosas. Para probar $H_0: A=2$ vs $H_1: A>2$ se tomará una muestra aleatoria simple (sin reemplazo) de 2 lavadoras y se rechazará la hipótesis nula si ambas están defectuosas

a. Halle el nivel de significación de la prueba.
Sea $X =$ Número de lavadoras defectuosas en una muestra aleatoria simple de tamaño 2,
$X \sim \text{Hipergeométrica}\left(N=6, n=2, A=\text{Desconocido}\right)$

$$\alpha = P\left[X = 2 / A=2\right] = \frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}} = \frac{1}{15}$$

b. Halle el poder de la prueba si en realidad A=4.

$$\text{Potencia de la prueba} = P\left[X = 2 / A=4\right] = \frac{\binom{4}{2}\binom{2}{0}}{\binom{6}{2}} = \frac{6}{15}$$

c. Halle el poder de la prueba si en realidad A=5.
d. Halle el poder de la prueba si en realidad A=6.
e. Resolver a, b y c si la selección es con reemplazo y considerando el orden de extracción.
f. Resolver a, b y c si la selección es con reemplazo y sin considerar el orden de extracción.

Ejemplo 6: La muestra aleatoria $X_1,\ldots,X_{10}$ de tamaño 10 es extraída de una distribución $\text{Poisson}(\lambda)$ para probar $H_0: \lambda = 0.12$ vs $H_1: \lambda > 0.12$. Se sabe que $Y = \sum_{i=1}^{10} X_i \sim \text{Poisson}\left(10\times 0.12 = 1.2\right)$ y asumiendo que se rechaza la hipótesis nula con valores altos de Y.

a. Halle la prueba aleatoria apropiada con $\alpha = 0.01$.

$$\varphi(Y) = \begin{cases} 1 & \text{,si } Y > k \quad \left(Y \geq k+1\right) \\ \gamma & \text{, si } Y = k \\ 0 & \text{,si } Y < k \end{cases}$$

Se prueba con diferentes valores y se encuentra que en $Y = 4$ está el "límite" porque $\underbrace{P\left(Y \geq 4\right) = 0.0338}_{\text{como pvalor}} > \alpha = 0.01$ mientras que $\underbrace{P\left(Y \geq 5\right) = 0.008}_{\text{como pvalor}} < \alpha = 0.01$.

Cálculo de γ considerando que $\alpha = E\left[\varphi\left(Y\right)\right] = 0.01$:

$P(\text{Rechazar } H_0 / H_0 \text{ verdadero}) = 1 \times P(Y \geq 5) + \gamma \times P(Y = 4) + 0 \times P(Y < 4) =$

$P(\text{Rechazar } H_0 / H_0 \text{ verdadero}) = 0.008 + \gamma \times 0.026 + 0 = \alpha = 0.01 \rightarrow \gamma = \frac{1}{13}$

Por lo tanto, la prueba randomizada será:

$$\varphi(Y) = \begin{cases} 1 & \text{,si } Y > 4 \quad (Y \geq 5) \\ \gamma = \frac{1}{13} & \text{, si } Y = 4 \\ 0 & \text{,si } Y < 4 \end{cases}$$

NOTA: Ver el Script R para hallar el valor de k=4 con el R

b. Con la muestra 0, 0, 1, 0, 0, 0, 2, 0, 1, 0 ¿cuál es la conclusión?

Como $Y = \sum_{i=1}^{10} X_i = 4$ entonces se rechaza la hipótesis nula con probabilidad $\frac{1}{13}$.

Ejemplo 7: Una muestra de tamaño 1 es extraída de una distribución $\text{Uniforme}(a,b)$ para probar las hipótesis simples $H_0 : a = 2 \ , \ b = 4$ vs $H_1 : a = 3 \ , \ b = 7$.

a. Obtenga la prueba aleatoria apropiada con $\alpha = 0.05$.

Para la regla de decisión se debe tener en cuenta lo siguiente:

Si $x \in (2,3)$ la H_0 no se rechazaría, porque H_1 no puede ser cierta.
Si $x \in (4,7)$ la H_0 si se rechazaría, porque H_1 si puede ser cierta.
Si $x \in (3,4)$ la H_0 a veces se rechazaría y a veces no, porque H_0 y H_1 pueden ser ciertas.

$$\varphi(x) = \begin{cases} 1 & \text{,si } x \in (4,7) \\ \gamma & \text{, si } x \in [3,4] \\ 0 & \text{,si } x \in (2,3) \end{cases}$$

Cálculo de p considerando que $\alpha = E[\varphi(Y)] = 0.05$:

$P(\text{Rechazar } H_0 / H_0 \text{ verdadero}) = 1 \times P[x \in (4,7)] + \gamma \times P[x \in [3,4]] + 0 \times P[x \in (2,3)] =$

$P(\text{Rechazar } H_0 / H_0 \text{ verdadero}) = 1 \times 0 + \gamma \times \int_3^4 \frac{1}{2} dx + 0 = \alpha = 0.05 \rightarrow \gamma = 0.1$

Por lo tanto, la prueba randomizada será:

$$\varphi(x) = \begin{cases} 1 & ,\text{si } x \in (4,7) \\ \gamma = 0.1 & ,\text{ si } x \in [3,4] \\ 0 & ,\text{si } x \in (2,3) \end{cases}$$

b. Halle el poder de la prueba.

$$P(\text{Rechazar } H_0 / H_1 \text{ verdadero}) = 1 \times P[x \in (4,7)] + 0.1 \times P[x \in [3,4]] + 0 \times P[x \in (2,3)] =$$

$$P(\text{Rechazar } H_0 / H_1 \text{ verdadero}) = 1 \times \int_4^7 \frac{1}{4} dx + 0.1 \times \int_3^4 \frac{1}{4} dx + 0 = \frac{31}{40} = 0.775$$

Ejercicio: La muestra aleatoria $X_1, \ldots, X_{10}$ de tamaño 10 es extraída de una distribución $\text{Binomial}(1,\theta)$ para probar $H_0 : \theta = 0.2$ vs $H_1 : \theta > 0.2$. Se sabe que $Y = \sum_{i=1}^{10} X_i \sim \text{Binomial}(10,\theta)$ y asumiendo que se rechaza la hipótesis nula con valores altos de Y. Halle la prueba aleatoria apropiada con $\alpha = 0.05$.

Respuesta:

$$\varphi(Y) = \begin{cases} 1 & ,\text{si } Y > 4 \quad (Y \geq 5) \\ \gamma = 0.1952 & ,\text{ si } Y = 4 \\ 0 & ,\text{si } Y < 4 \end{cases}$$

FUNCIÓN POTENCIA O PODER DE LA PRUEBA

Se usa cuando al menos una de las hipótesis es compuesta.
Sea φ una prueba de hipótesis nula Ho.

$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } \underline{x} \in C \\ 0 & ,\text{si } \underline{x} \notin C \end{cases}$$

Para un sistema de hipótesis, puede haber más de una regla de decisión. En este caso, necesitamos comparar estas reglas de decisión en términos de probabilidad de cometer los dos tipos de errores. La función potencia de la prueba φ (o regla de decision), se define como una función del parámetro θ definida como:

$\Pi_\varphi(\theta) = P_\theta$ (rechazar Ho)

Por lo tanto $1 - \Pi_\varphi(\theta) = P(\text{No rechazar } H_0)$, y se puede asociar la función potencia con los dos errores que se pueden cometer de la siguiente manera:

- Si H_0 es verdadera, o sea si $\theta \in \Theta_0$, entonces al rechazar H_0, se comete error tipo I, y la máxima probabilidad de cometer este error es α. De tal manera que $\Pi_\varphi(\theta) \leq \alpha$.

- Si H_0 es falsa, o sea si $\theta \in \Theta_0^c$ $\left(\text{no necesariamente } \theta \in \Theta_1\right)$, entonces al no rechazar H_0, se comete error tipo II. De tal manera que $1-\Pi_\varphi(\theta)=\beta$.

Por lo tanto, la función potencia se expresa de la siguiente manera:

$$\Pi_\varphi(\theta)=\begin{cases}\text{Probabilidad de cometer error tipo I} \leq \alpha & \text{si } \theta \in \Theta_0 \\ 1-\beta & \text{si } \theta \in \Theta_0^c \ \left(\text{no necesariamente } \theta \in \Theta_1\right)\end{cases}$$

Ver el Script R.

Prueba Uniformemente más Poderosa (UMP)

Sea φ una prueba de tamaño α. Sea φ_i otra prueba de tamaño al menos α. Si $\beta_\varphi(\theta) \leq \beta_{\varphi_i}(\theta)$, $\forall i$ y $\theta \in \Omega_1$, entonces φ es una prueba UMP.

$$\beta_\varphi(\theta) \leq \beta_{\varphi_i}(\theta) \Rightarrow 1-\beta_\varphi(\theta) \geq 1-\beta_{\varphi_i}(\theta)$$
$$\Rightarrow \Pi(\theta/\varphi) \geq \Pi(\theta/\varphi_i)$$

Ejemplo: Suponga que asumiendo normalidad se tiene el siguiente sistema de hipótesis: $H_0: \mu=\mu_0$ vs $H_1: \mu \neq \mu_0$ con σ conocida.

$$\Pi(\mu)=P(\text{Rechazar } H_0)=P\left(\bar{X}>\mu_0+z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}} \ \text{ ó } \ \bar{X}<\mu_0-z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right)=$$
$$=P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}>\frac{\mu_0-\mu}{\sigma/\sqrt{n}}+z_{(1-\alpha/2)}\right)+P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{\mu_0-\mu}{\sigma/\sqrt{n}}-z_{(1-\alpha/2)}\right)$$

Ver el Script R

Ejemplo: Suponga que asumiendo normalidad se tiene el siguiente sistema de hipótesis: $H_0: \sigma^2=\sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$ con μ conocido.

$$\Pi(\sigma^2)=P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2}>\chi^2_{(n,1-\alpha/2)}\right)+P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2}<\chi^2_{(n,\alpha/2)}\right)=$$
$$=P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}>\frac{\sigma_0^2}{\sigma^2}\chi^2_{(n,1-\alpha/2)}\right)+P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}<\frac{\sigma_0^2}{\sigma^2}\chi^2_{(n,\alpha/2)}\right)=$$

VER EL SCRIPT R

Ejemplo: Suponga que asumiendo normalidad se tiene el siguiente sistema de hipótesis: $H_0: \mu=\mu_0$ vs $H_1: \mu \neq \mu_0$ con σ desconocida.

$$\Pi(\mu) = P(\text{Rechazar } H_0) = P\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > t_{(n-1,1-\alpha/2)}\right) + P\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} < -t_{(n-1,1-\alpha/2)}\right) =$$

Pero ahora, $\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right) \rightarrow \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N\left(\dfrac{\mu_0-\mu}{\sigma/\sqrt{n}}, 1\right)$ y $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

Como $\bar{X}$ y $S^2 \rightarrow \dfrac{\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{S^2}{\sigma^2}}} = \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t^{\text{nc}}_{n-1,\delta}$ con $\delta = \dfrac{\mu-\mu_0}{\sigma/\sqrt{n}} \rightarrow$

$$\Pi(\mu) = 1 - Ft^{\text{nc}}_{n-1,\delta}\left(t_{n-1,1-\alpha/2}\right) + Ft^{\text{nc}}_{n-1,\delta}\left(t_{n-1,\alpha/2}\right)$$

Donde $Ft^{\text{nc}}_{n-1,\delta}$ es la fda de la distribución $t^{\text{nc}}_{n-1,\delta}$.

VER EL SCRIPT R

Ejemplo 8: Suponga que el fabricante de un nuevo medicamento para la gastritis desea probar:

Ho: $\theta \geq 0.9$ (Por lo menos el 90% de los pacientes se recuperan)
Ha: $\theta < 0.9$ (Menos del 90% de los pacientes se recuperan)

Si X= número de pacientes que se recuperan en un grupo de 25 pacientes; considere la prueba de hipótesis nula

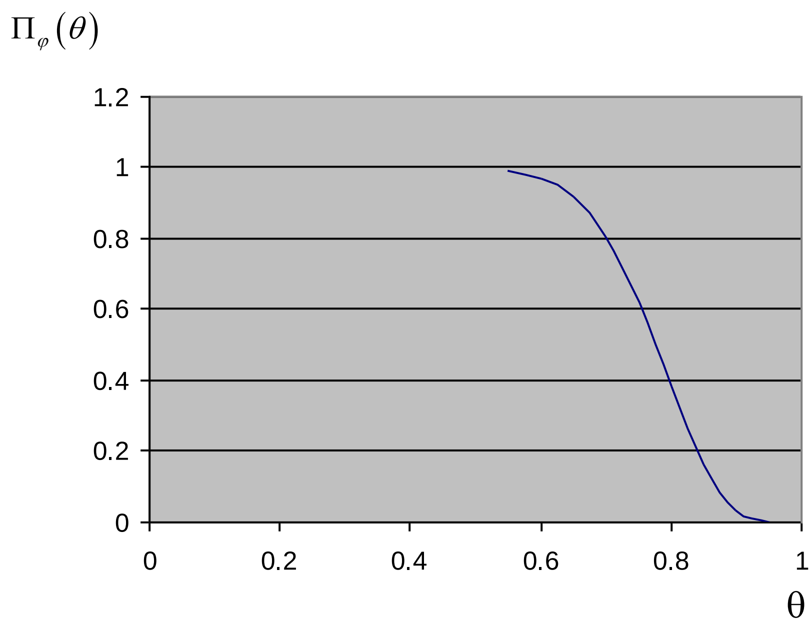
$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } \; x < 20 \\ 0 & ,\text{si } \; x \geq 20 \end{cases}$$

trazar la gráfica de la función potencia de φ o sea $\Pi_\varphi(\theta)$.

Solución:

X ~Binomial (25, θ)

θ	P(X<20)	β =P(X $\geq$ 20: θ)	$\Pi_\varphi(\theta)$ (P. de rechazar Ho)
0.95	0.00121296		0.00121296
0.9	0.03339994		0.03339994
0.85		0.8384846	0.1615154
0.8		0.61668941	0.38331059
0.75		0.37827851	0.62172149
0.7		0.19348844	0.80651156
0.65		0.0826247	0.9173753
0.6		0.0293622	0.9706378
0.55		0.00859907	0.99140093



TAMAÑO DE LA PRUEBA

Sea φ una prueba de Ho: $\theta \in \Theta_o$, donde $\Theta_o \subset \Theta$. El tamaño de la prueba φ o nivel de significación para probar Ho vs. H_1 es: $\alpha = \sup_{\theta\in\Theta}\left\{\Pi_{\varphi}(\theta)\right\}$ ó $\alpha = \Pi_{\varphi}(\theta_o)$.

En el ejemplo anterior Ho: $\theta \geq 0.9$

$\Rightarrow \Theta_o = \{\theta / \theta \geq 0.9\} \Rightarrow \alpha = \sup_{\theta\in\{\theta/\theta\geq 0.9\}}\left\{\Pi_{\varphi}(\theta)\right\} = \Pi_{\varphi}(0.9)$ = P(X<20; θ =0.9) =

$$= \sum_{X=0}^{19}\binom{25}{x}0.9^{x}0.1^{25-x} = 0.0334$$

Ejemplo 9: Sea $X_1,...,X_9$ una m.a. de N (θ,9) donde ; Ho: $\theta \leq 15$ vs. H_1: $\theta > 15$

Considere la prueba de Hipótesis nula:

$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } \bar{x} > 17 \\ 0 & ,\text{si } \bar{x} \leq 17 \end{cases}$$

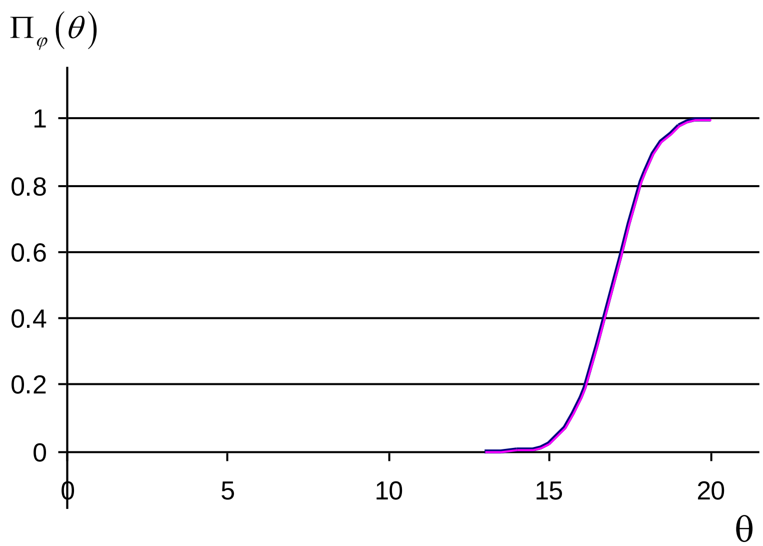
Trazar la grafica de la función potencia de φ o sea $\Pi_{\varphi}(\theta)$

Solución:

$\Pi_{\varphi}(\theta) = P_{\theta}$ (rechazar Ho) $= P_{\theta}(\underline{x} \in C) = P_{\theta}(\bar{x} > 17) = 1 - P_{\theta}(\bar{x} \leq 17) =$

$= 1 - P_{\theta}\left(\frac{\bar{x}-\theta}{3/\sqrt{n}} \leq \frac{17-\theta}{1}\right) = 1 - P(Z \leq 17-\theta)$; n = 9

θ	β	$\Pi_{\varphi}(\theta)$
20	0.0014	0.9986
19	0.0228	0.9772
18	0.1586	0.8414
17	0.5	0.5
16	0.8413	0.1587
15	0.9773	0.0227
14	0.9987	0.0013
13	0.9999	0.0001



$$\alpha = \sup_{\theta\in\{\theta/\theta\le 15\}}\{\Pi_{\varphi}(\theta)\} = \Pi_{\varphi}(15) = 0.0227$$

PRUEBAS ÓPTIMAS, MÁS PODEROSAS O POTENTES PARA HIPÓTESIS SIMPLES

Prueba de Razón de Verosimilitud Simple:

Sea $X_1,\ldots,X_n$ una m.a. de $f(x;\theta)$ donde $\theta \in \Theta = \{\theta_0,\theta_1\}$. Se dice que la función crítica φ es una prueba de razón de verosimilitud simple para probar las hipótesis simples; Ho: $\theta = \theta_0$ Vs. H1: $\theta = \theta_1$, si la regla de correspondencia de φ es dada por:

$$\varphi(\underline{x}) = \begin{cases} 1 & ,\text{si } \lambda(\underline{x}) \le k \\ 0 & ,\text{si } \lambda(\underline{x}) > k \end{cases}$$

donde $\lambda(\underline{x})$ es la razón de la verosimilitud simple dada por:

$$\lambda(\underline{x}) = \lambda(x_1,\ldots,x_n) = \frac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})} = \frac{\prod_{i=1}^{n} f(x_i,\theta_0)}{\prod_{i=1}^{n} f(x_i,\theta_1)} \quad \text{y} \quad k > 0$$

Prueba Más Poderosa o Potente:

Se dice que φ^* es una prueba más poderosa de tamaño α (0< α <1) sii:

1. $\prod_{\varphi^*}(\theta_0) = \alpha$
2. $\prod_{\varphi^*}(\theta_1) \geq \prod_{\varphi}(\theta_1)$ para cualquier otra prueba φ tal que $\prod_{\varphi}(\theta_0) \leq \alpha$

Construcción de Pruebas Más Poderosas de tamaño α para contrastar Hipótesis Simple vs Simple

TEOREMA: Lema de Neyman-Pearson

Sea $X_1,\ldots,X_n$ una m.a de $f(x;\theta)$ donde $\theta \in \Theta = \{\theta_0,\theta_1\}$ y sea α un número fijo. Sea k una constante positiva y $C^* \subset \mathbb{X}$ tal que:

1. $C^* = \left\{ \underline{X} = (x_1,\ldots,x_n) / \lambda(\underline{x}) = \dfrac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})} < k \right\}$

2. $P_{\theta_0}\left(\underline{X} = (x_1,\ldots,x_n) \in C^*\right) = \alpha = \prod_{\varphi^*}(\theta_0)$

Entonces: $\varphi^*(\underline{x}) = \begin{cases} 1 \quad , \quad \text{si } \dfrac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})} < k \\ \gamma \quad , \quad \text{si } \dfrac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})} = k \\ 0 \quad , \quad \text{si } \dfrac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})} > k \end{cases}$

Es una prueba más poderosa o potente, de tamaño α para probar la hipótesis: Ho: $\theta = \theta_0$ vs. H1: $\theta = \theta_1$.

La prueba de Neyman-Pearson es una función de las estadísticas suficientes.

Nota: bajo el contexto de hipótesis simples y cuando se aplican las condiciones del Lema de Neyman Pearson la prueba más poderosa es única pero en el contexto de hipótesis compuestas no siempre hay una prueba más poderosa que sea única.

NOTA: Si $X_1,\ldots,X_n$ es una muestra aleatoria de una distribución discreta entonces la prueba es randomizada cuando se considera $0<\gamma<1$, y sería no randomizada si se asume que $\gamma=0$. Pero si la muestra aleatoria proviene de una distribución continua se cumple que siempre $\gamma=0$, y la prueba es no randomizada.

Ejemplo 10: El número de fallas de un sistema por mes tiene distribución $\text{Poisson}(\lambda)$. Se tomó una muestra aleatoria de 12 meses para probar: $H_0:\lambda=2$ vs $H_1:\lambda=3$.

a. Halle la prueba no randomizada más poderosa de tamaño $\alpha=0.0678$ para probar: $H_0:\lambda=2$ vs $H_1:\lambda=3$.

Utilizando el Lema de Neyman-Pearson:
$\lambda_0=2$ y $\lambda_1=3$

1. $C^*=\left\{(x_1,...,x_n)\,/\,\lambda(\underline{x})=\dfrac{L(\lambda_0,\underline{x})}{L(\lambda_1,\underline{x})}\le k\right\}$

$$\lambda(\underline{x})=\frac{L(\lambda_0=2)}{L(\lambda_1=3)}=\frac{\prod_{i=1}^{12}\frac{e^{-2}2^{x_i}}{x_i!}}{\prod_{i=1}^{12}\frac{e^{-3}3^{x_i}}{x_i!}}=\frac{e^{-12(2)}2^{\sum_{i=1}^{12}x_i}}{e^{-12(3)}3^{\sum_{i=1}^{12}x_i}}=e^{12}\left(\frac{2}{3}\right)^{\sum_{i=1}^{12}x_i}\le k$$

Tomando logaritmos en la desigualdad:

$$12+\left(\sum_{i=1}^{12}x_i\right)\underbrace{(\ln 2-\ln 3)}_{\text{Negativo}}\le \ln k$$

$$\sum_{i=1}^{12}x_i\ge\frac{\ln k-12}{\ln 2-\ln 3}$$

$\sum_{i=1}^{12}x_i\ge k_1$, la $\sum_{i=1}^{12}x_i$ es una estadística suficiente.

La prueba no randomizada más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{12}x_i\ge k_1\\ 0 & , \text{ o.m}\end{cases}$$

Cálculo de k_1:
De acuerdo al enunciado se tiene que:

$$Y=\sum_{i=1}^{12}x_i\sim\text{Poisson}(n\lambda)\sim\text{Poisson}(12\times 2)\sim\text{Poisson}(24)$$

$$\alpha = 0.0678 = P_{\lambda=2}\left(Y \geq k_1\right) = \sum_{y=k_1}^{\infty} \frac{e^{-24}24^y}{y!} \rightarrow$$

$$1-\alpha = 0.9322 = 1-\sum_{y=k_1}^{\infty} \frac{e^{-24}24^y}{y!} = \sum_{y=0}^{k_1-1} \frac{e^{-24}24^y}{y!} \rightarrow k_1 - 1 = 31 \rightarrow k_1 = 32$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , \quad \text{si } \sum_{i=1}^{12} x_i \geq 32 \\ 0 \quad , \quad \text{o.m} \end{cases}$$

Hay que tener en cuenta que $\sum_{i=1}^{12} x_i$ es una estadística suficiente.

2. $P_{\lambda_0=2}\left[\left(x_1,...,x_n\right) \in C^*\right] = \sum_{y=32}^{\infty} \frac{e^{-24}24^y}{y!} = 1-\sum_{y=0}^{31} \frac{e^{-24}24^y}{y!} =$

$$P_{\lambda_0=2}\left[\left(x_1,...,x_n\right) \in C^*\right] = 1-0.9322 = \alpha = 0.0678 = \prod_{\varphi^*}\left(\lambda_0\right)$$

b. Si se tiene la muestra de tamaño doce: 1, 3, 5, 2, 4, 3, 4, 2, 1, 5, 4, 3. ¿Cuál es su conclusión?

Si la hipótesis nula es cierta se tiene que:

$$Y = \sum_{i=1}^{12} x_i \sim \text{Poisson}\left(n\lambda\right) \sim \text{Poisson}\left(12 \times 2\right) \sim \text{Poisson}\left(24\right)$$

Con los datos de la muestra $\sum_{i=1}^{12} x_i = 37 > 32 \rightarrow$ se rechaza la hipótesis nula.

$$\text{pvalor} = P_{\lambda=2}\left(\sum_{i=1}^{12} x_i \geq 37\right) = 1 - P_{\lambda=2}\left(\sum_{i=1}^{12} x_i < 37\right) = 1-0.9918 = 0.0082 < \alpha$$

c. Calcule la potencia de la prueba.

La potencia de la prueba consiste en la probabilidad de rechazar una H_0 falsa $\rightarrow$

Si la hipótesis nula es falsa se tiene que:

$$Y = \sum_{i=1}^{12} x_i \sim \text{Poisson}\left(n\lambda\right) \sim \text{Poisson}\left(12 \times 3\right) \sim \text{Poisson}\left(36\right)$$

$$\prod_{\varphi^*}\left(\lambda\right) = P_{\lambda=3}\left(Y \geq 32\right) = 1 - P_{\lambda=3}\left(Y < 32\right) = 1 - \sum_{y=0}^{31} \frac{e^{-36}36^y}{y!} =$$

$$\prod_{\varphi^*}\left(\lambda\right) = 1 - 0.23055 = 0.76974$$

d. Obtenga la prueba randomizada más poderosa para probar $H_0 : \lambda = 2$ vs $H_1 : \lambda = 3$. Use $\alpha = 0.05$.

Utilizando el Lema de Neyman-Pearson:

$\lambda_0 = 2$ y $\lambda_1 = 3$

1. $C^* = \left\{ \left(x_1, ..., x_n \right) / \lambda\left(\underline{x}\right) = \dfrac{L\left(\lambda_0, \underline{x}\right)}{L\left(\lambda_1, \underline{x}\right)} \le k \right\}$

$$\lambda\left(\underline{x}\right) = \frac{L\left(\lambda_0 = 2\right)}{L\left(\lambda_1 = 3\right)} = \frac{\prod_{i=1}^{12} \frac{e^{-2} 2^{x_i}}{x_i !}}{\prod_{i=1}^{12} \frac{e^{-3} 3^{x_i}}{x_i !}} = \frac{e^{-12(2)} 2^{\sum_{i=1}^{12} x_i}}{e^{-12(3)} 3^{\sum_{i=1}^{12} x_i}} = e^{12} \left(\frac{2}{3} \right)^{\sum_{i=1}^{12} x_i} < k$$

Tomando logaritmos en la desigualdad:

$$12 + \left(\sum_{i=1}^{12} x_i \right) \underbrace{\left(\ln 2 - \ln 3 \right)}_{\text{Negativo}} < \ln k$$

$$\sum_{i=1}^{12} x_i > \frac{\ln k - 12}{\ln 2 - \ln 3}$$

$$\sum_{i=1}^{12} x_i > k_1$$

La prueba más poderosa será:

$$\varphi^*\left(x_1, \ldots, x_n\right) = \begin{cases} 1 \quad , \quad \text{si} \quad \dfrac{L\left(\theta_0, \underline{x}\right)}{L\left(\theta_1, \underline{x}\right)} < k \rightarrow \sum_{i=1}^{12} x_i > k_1 \\ \gamma \quad , \quad \text{si} \quad \dfrac{L\left(\theta_0, \underline{x}\right)}{L\left(\theta_1, \underline{x}\right)} = k \rightarrow \sum_{i=1}^{12} x_i = k_1 \\ 0 \quad , \quad \text{si} \quad \dfrac{L\left(\theta_0, \underline{x}\right)}{L\left(\theta_1, \underline{x}\right)} > k \rightarrow \sum_{i=1}^{12} x_i < k_1 \end{cases}$$

Cálculo de k_1:

De acuerdo al enunciado se tiene que:

$$Y = \sum_{i=1}^{12} x_i \sim \text{Poisson}\left(n\lambda\right) \sim \text{Poisson}\left(12 \times 2\right) \sim \text{Poisson}\left(24\right)$$

$$\alpha = 0.05 = P_{\lambda=2}\left(Y > k_1\right) = \sum_{y=k_1+1}^{\infty} \frac{e^{-24}24^y}{y!} \to k_1 + 1 = 33 \to k_1 = 32$$

$$P_{\lambda=2}\left(Y > 32\right) = \sum_{y=33}^{\infty} \frac{e^{-24}24^y}{y!} = 0.0467$$

$$P_{\lambda=2}\left(Y \geq 32\right) = \sum_{y=33}^{\infty} \frac{e^{-24}24^y}{y!} = 0.0678$$

$$P_{\lambda=2}\left(Y = 32\right) = 0.0211$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , & \text{si } \sum_{i=1}^{12} x_i > 32 \\ \gamma \quad , & \text{si } \sum_{i=1}^{12} x_i = 32 \\ 0 \quad , & \text{si } \sum_{i=1}^{12} x_i < 32 \end{cases}$$

Cálculo de γ considerando que $\alpha = E\left[\varphi\left(Y\right)\right] = 0.05$:

$$P\left(\text{Rechazar } H_0 \,/\, H_0 \text{ verdadero}\right) = 1 \times P\left(Y > 32\right) + \gamma \times P\left(Y = 32\right) + 0 \times P\left(Y < 32\right) =$$

$$P\left(\text{Rechazar } H_0 \,/\, H_0 \text{ verdadero}\right) = 0.0467 + \gamma \times 0.0211 + 0 = \alpha = 0.05 \to \gamma = 0.1564$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , & \text{si } \sum_{i=1}^{12} x_i > 32 \\ 0.1564, & \text{si } \sum_{i=1}^{12} x_i = 32 \\ 0 \quad , & \text{si } \sum_{i=1}^{12} x_i < 32 \end{cases}$$

Ejemplo 11: El $\theta \times 100\%$ de los estudiantes de cierta universidad son hijos de padres separados. Sea $X_1,\ldots,X_9$ una muestra aleatoria de 9 estudiantes.

a. Encuentre la prueba no randomizada más poderosa para probar $H_0 : \theta = 0.6$ vs $H_1 : \theta = 0.8$. Use $\alpha = 0.07054$.

$X_1,\ldots,X_9$ es una muestra aleatoria de una distribución $\text{Binomial}\left(1,\theta\right)$

Utilizando el Lema de Neyman-Pearson:

$\theta_0 = 0.6$ y $\theta_1 = 0.8$

1. $C^{*}=\left\{\left(x_1,...,x_n\right)/\lambda\left(\underline{x}\right)=\dfrac{L\left(\theta_0,\underline{x}\right)}{L\left(\theta_1,\underline{x}\right)}\leq k\right\}$

$$\lambda\left(\underline{x}\right)=\frac{L\left(\theta_0=0.6\right)}{L\left(\theta_1=0.8\right)}=\frac{\prod_{i=1}^{9}0.6^{x_i}\left(1-0.6\right)^{1-x_i}}{\prod_{i=1}^{9}0.8^{x_i}\left(1-0.8\right)^{1-x_i}}=\frac{0.6^{\sum_{i=1}^{9}x_i}\times 0.4^{9-\sum_{i=1}^{9}x_i}}{0.8^{\sum_{i=1}^{9}x_i}\times 0.2^{9-\sum_{i=1}^{9}x_i}}=\left(\frac{3}{4}\right)^{\sum_{i=1}^{9}x_i}\times 2^{9-\sum_{i=1}^{9}x_i}\leq k$$

Tomando logaritmos en la desigualdad:

$$\left(\sum_{i=1}^{9}x_i\right)\ln 0.75+9\ln 2-\left(\sum_{i=1}^{9}x_i\right)\ln 2\leq \ln k$$

$$\left(\sum_{i=1}^{9}x_i\right)\underbrace{\left(\ln 0.75-\ln 2\right)}_{\text{Negativo}}+9\ln 2\leq \ln k$$

$$\sum_{i=1}^{9}x_i\geq\frac{\ln k-9\ln 2}{\ln 0.75-\ln 2}$$

$\sum_{i=1}^{9}x_i\geq k_1$, la $\sum_{i=1}^{9}x_i$ es una estadística suficiente.

La prueba más poderosa será:

$$\varphi^{*}\left(x_1,\ldots,x_n\right)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{9}x_i\geq k_1\\ 0 & , \text{ o.m}\end{cases}$$

Cálculo de k_1:

De acuerdo al enunciado se tiene que:

$$Y=\sum_{i=1}^{9}x_i\sim\text{Binomial}\left(9,\theta\right)$$

$$\alpha=0.07054=P_{\theta=0.6}\left(Y\geq k_1\right)=\sum_{y=k_1}^{9}\binom{9}{y}0.6^{y}\times 0.4^{9-y}\rightarrow k_1=8$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^{*}\left(x_1,\ldots,x_n\right)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{9}x_i\geq 8\\ 0 & , \text{ o.m}\end{cases}$$

Hay que tener en cuenta que $\sum_{i=1}^{9}x_i$ es una estadística suficiente.

2. $P_{\theta_0=0.6}\left[\left(x_1,...,x_n\right)\in C^{*}\right]=\sum_{y=8}^{9}\binom{9}{y}0.6^{y}\times 0.4^{9-y}=\alpha=0.07054=\prod_{\varphi^{*}}\left(\theta_0\right)$

b. Concluya con la muestra 1, 0, 0, 1, 1, 1, 1, 1, 0.

Respuesta: No se rechaza la hipótesis nula.

Ejemplo 12: Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución $N\left(\mu,\sigma^2=4\right)$. Halle la prueba óptima de tamaño α para probar $H_0:\mu=\mu_0$ vs $H_1:\mu=\mu_1$ considerando $\mu_1>\mu_0$.

Utilizando el Lema de Neyman-Pearson:
Teniendo en cuenta que $\mu_1>\mu_0$.

1. $C^*=\left\{\left(x_1,...,x_n\right)/\,\lambda\left(\underline{x}\right)=\dfrac{L\left(\mu_0,\underline{x}\right)}{L\left(\mu_1,\underline{x}\right)}\le k\right\}$

$$\lambda\left(\underline{x}\right)=\frac{L\left(\mu_0\right)}{L\left(\mu_1\right)}=\frac{\prod_{i=1}^{n}\left(2\pi\right)^{-\frac{1}{2}}\times 2^{-1}\exp\left[-\frac{1}{2\times 4}\sum_{i=1}^{n}(x_i-\mu_0)^2\right]}{\prod_{i=1}^{n}\left(2\pi\right)^{-\frac{1}{2}}\times 2^{-1}\exp\left[-\frac{1}{2\times 4}\sum_{i=1}^{n}(x_i-\mu_1)^2\right]}=$$

$$\lambda\left(\underline{x}\right)=\frac{\left(2\pi\right)^{-\frac{n}{2}}\times 2^{-n}\exp\left[-\frac{1}{8}\left(\sum_{i=1}^{n}x_i^2-2\mu_0\sum_{i=1}^{n}x_i+n\mu_0^2\right)\right]}{\left(2\pi\right)^{-\frac{n}{2}}\times 2^{-n}\exp\left[-\frac{1}{8}\left(\sum_{i=1}^{n}x_i^2-2\mu_1\sum_{i=1}^{n}x_i+n\mu_1^2\right)\right]}$$

$$\lambda\left(\underline{x}\right)=\exp\left[-\frac{1}{4}\left(\mu_1-\mu_0\right)\sum_{i=1}^{n}x_i-\frac{n}{8}\left(\mu_0^2-\mu_1^2\right)\right]\le k$$

Tomando logaritmos en la desigualdad y operando:

$$\sum_{i=1}^{n}x_i\ge\frac{4}{\mu_1-\mu_0}\left[\frac{n}{8}\left(\mu_0^2-\mu_1^2\right)-\ln k\right]$$

$\sum_{i=1}^{n}x_i\ge k_1$, la $\sum_{i=1}^{9}x_i$ es una estadística suficiente.

La prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right)=\begin{cases}1 & ,\ \text{si}\ \sum_{i=1}^{n}x_i\ge k_1\\ 0 & ,\ \text{o.m}\end{cases}$$

$$\alpha = P_{\mu_0}\left(\sum_{i=1}^{n} x_i \geq k_1\right) = P_{\mu_0}\left(\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{k_1}{n}\right) = P_{\mu_0}\left(\bar{x} \geq k_2\right) =$$

$$\alpha = P_{\mu_0}\left(Z \geq \frac{\left(k_2 - \mu_0\right)}{2}\sqrt{n}\right) \rightarrow \frac{\left(k_2 - \mu_0\right)}{2}\sqrt{n} = Z\left(1-\alpha\right) \rightarrow$$

$$k_2 = \mu_0 + \frac{2}{\sqrt{n}} Z\left(1-\alpha\right)$$

La prueba más poderosa en función de $\bar{x}$ (que también es suficiente) será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , \quad \text{si } \bar{x} \geq \mu_0 + \frac{2}{\sqrt{n}} Z\left(1-\alpha\right) \\ 0 \quad , \quad \text{o.m} \end{cases}$$

Ejemplo 13: La siguiente distribución corresponde a la proporción de vitamina B en porciones de 100g de zanahoria: $f\left(x\right) = \theta x^{\theta-1} I_{(0,1)}\left(x\right)$ con $\theta > 0$.

a. Con una muestra de tamaño 1 halle la prueba más poderosa de tamaño α para probar $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$, con $\theta_1 > \theta_0$.

Utilizando el Lema de Neyman-Pearson:
$\theta_1 > \theta_0$

1. $C^* = \left\{\left(x_1,...,x_n\right) / \lambda\left(\underline{x}\right) = \frac{L\left(\theta_0, \underline{x}\right)}{L\left(\theta_1, \underline{x}\right)} \leq k\right\}$

$$\lambda\left(\underline{x}\right) = \frac{L\left(\theta_0\right)}{L\left(\theta_1\right)} = \frac{\theta_0 x^{\theta_0 - 1}}{\theta_1 \times x^{\theta_1 - 1}} = \frac{\theta_0}{\theta_1} x^{\theta_0 - \theta_1} \leq k$$

$$\rightarrow \left(\theta_0 - \theta_1\right)\log\left(x\right) \leq c \rightarrow \log\left(x\right) \geq c_1$$

Se demuestra que $Y = -\log\left(X\right) \sim$ Exponencial(Tasa $= \theta$), $-\log\left(X\right)$ es suficiente.

La prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , \quad \text{si } -\log\left(x\right) \leq k_1 \\ 0 \quad , \quad \text{o.m} \end{cases}$$

Cálculo de k_1:

$$\alpha = P_{\theta=\theta_0}\left(-\log\left(X\right) \leq k_1\right) = P_{\theta=\theta_0}\left(Y \leq k_1\right) = \int_0^{k_1} \theta_0 \times e^{-\theta_0 y} dx = 1 - e^{-\theta_0 k_1} \rightarrow k_1 = -\frac{\log\left(1-\alpha\right)}{\theta_0}$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^*(x)=\begin{cases}1 \quad , \quad \text{si} \quad -\log(x)\le -\dfrac{\log(1-\alpha)}{\theta_0}\\ 0 \quad , \quad \text{o.m}\end{cases}$$

b. Calcule el poder de la prueba.

La potencia de la prueba consiste en la probabilidad de rechazar una H_0 falsa $\rightarrow$
Si la hipótesis nula es falsa se tiene que:

$$\prod_{\varphi^*}(\theta)=P_{\theta=\theta_1}(Y\le k_1)=\int_0^{k_1}\theta_1\times e^{-\theta_1 y}dy=1-e^{-\theta_1 k_1}$$

c. Con una muestra aleatoria de tamaño n halle la mejor región crítica de tamaño α para probar $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1$, con $\theta_1>\theta_0$.

Utilizando el Lema de Neyman-Pearson:
$\theta_1>\theta_0$

1. $C^*=\left\{(x_1,...,x_n)\,/\,\lambda(\underline{x})=\dfrac{L(\theta_0,\underline{x})}{L(\theta_1,\underline{x})}\le k\right\}$

$$\lambda(\underline{x})=\frac{L(\theta_0)}{L(\theta_1)}=\frac{\prod_{i=1}^{n}\theta_0\times x_i^{\theta_0-1}}{\prod_{i=1}^{n}\theta_1\times x_i^{\theta_1-1}}=\left(\frac{\theta_0}{\theta_1}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{\theta_0-\theta_1}\le k$$

$$\rightarrow\left(\prod_{i=1}^{n}x_i\right)^{\theta_0-\theta_1}\le c\rightarrow\log\left(\prod_{i=1}^{n}x_i\right)\ge c_1\rightarrow\sum_{i=1}^{n}\log(X_i)\ge c_1$$

Se sabe que: $\ln\prod_{i=1}^{n}x_i=\sum_{i=1}^{n}\ln x_i$

Se puede verificar que: $-\ln x\sim\text{Exponencial}\left(\text{con media }\dfrac{1}{\theta}\right)\rightarrow$

$$-\sum_{i=1}^{n}\ln x_i\sim\text{Gamma}(n,\theta)\rightarrow 2\left(-\sum_{i=1}^{n}\ln x_i\right)\theta\sim\chi^2(2n),\ -\sum_{i=1}^{n}\ln x_i \text{ es suficiente.}$$

Al considerarse $-\sum_{i=1}^{n}\ln x_i$ cambia el sentido de la desigualdad en Neyman-Pearson.

La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 \quad , \quad \text{si} \quad 2\left(-\sum_{i=1}^{n}\ln x_i\right)\theta\le k_2\\ 0 \quad , \quad \text{o.m}\end{cases}$$

Planteamiento del cálculo de k_2:

$$\alpha = P_{\theta=\theta_0}\left[2\left(-\sum_{i=1}^{n}\ln x_i\right)\theta_0 \le k_2\right] = P_{\theta=\theta_0}\left[\chi^2(2n) \le k_2\right]$$

Ejemplo 14: Suponga que el contenido de β-Caroteno en 100 g de papaya tiene distribución $\text{Laplace}(\mu = 833,\ \lambda)$. Con la muestra aleatoria 833.4, 834.2, 832.8, 832.1 y 834.4 verifique: $H_0: \lambda = 0.5$ vs $H_1: \lambda = 0.6$. Use α=0.05.

Utilizando el Lema de Neyman-Pearson:

$\lambda_0 = 0.5$ y $\lambda_1 = 0.6$

1. $C^* = \left\{(x_1,...,x_n) / \lambda(\underline{x}) = \dfrac{L(\lambda_0,\underline{x})}{L(\lambda_1,\underline{x})} \le k\right\}$

$$\lambda(\underline{x}) = \frac{L(\lambda_0 = 0.5)}{L(\lambda_1 = 0.6)} = \frac{\prod_{i=1}^{n}\dfrac{\exp\left\{-\dfrac{|x_i-\mu|}{0.5}\right\}}{1}}{\prod_{i=1}^{n}\dfrac{\exp\left\{-\dfrac{|x_i-\mu|}{0.6}\right\}}{1}} = \exp\left[-\frac{1}{3}\sum_{i=1}^{n}|x_i-\mu|\right] \le k$$

$\rightarrow \sum_{i=1}^{n}|x_i-\mu| \ge c,\ \sum_{i=1}^{n}|x_i-\mu|$ es suficiente.

Hallando la densidad de $|x-\mu|$

$$y = |x-\mu| = \begin{cases} -x+\mu, & x<\mu \\ x-\mu, & x>\mu \end{cases}$$

$x < \mu$

$y = -x+\mu \rightarrow x = \mu - y \rightarrow J_1 = -1$

$$f_Y(y) = f_X(\mu-y) = \frac{\exp\left(\dfrac{-|\mu-y-\mu|}{\lambda}\right)}{2\lambda} = \frac{\exp\left(\dfrac{-|-y|}{\lambda}\right)}{2\lambda} I_{(0,\infty)}(y) = \frac{\exp\left(\dfrac{-y}{\lambda}\right)}{2\lambda} I_{(0,\infty)}(y)$$

$x > \mu$

$y = x-\mu \rightarrow x = \mu + y \rightarrow J_1 = 1$

$$f_Y(y) = f_X(\mu+y) = \frac{\exp\left(\dfrac{-|\mu+y-\mu|}{\lambda}\right)}{2\lambda} = \frac{\exp\left(\dfrac{-|y|}{\lambda}\right)}{2\lambda} I_{(0,\infty)}(y) = \frac{\exp\left(\dfrac{-y}{\lambda}\right)}{2\lambda} I_{(0,\infty)}(y)$$

$$\rightarrow f_Y(y) = \frac{\exp\left(\dfrac{-y}{\lambda}\right)}{\lambda} I_{(0,\infty)}(y) \rightarrow Y = |x-\mu| \sim \text{Exponencial}(\text{con media } \lambda)$$

$$T = \sum_{i=1}^{n}|x_i-\mu| \sim \text{Gamma}\left(n=5, \frac{1}{\lambda}\right) \rightarrow T_c = \sum_{i=1}^{5}|x_i-\mu| = 4.1$$

$$Q = 2\times\sum_{i=1}^{n}\left|x_i-\mu\right|\times\frac{1}{\lambda} \sim \chi^2_{(2n)} \sim \chi^2_{(10)} \to Q_c = 2\times 4.1\times\frac{1}{0.5} = 16.4$$

$$\text{pvalor} = P\left[\chi^2_{(10)} > 16.4\right] = 0.08 \to \text{NR}H_0$$

Ejemplo 15: Con la variable aleatoria X definida como la proporción de estudiantes universitarios que son hijos únicos por universidad se definen las hipótesis $H_0: f_1(x)=4(x-0.25)I_{(0,1)}(x)$ vs $H_1: f_2(x)=\frac{24}{25}\left(5x^2-\frac{5}{4}x\right)I_{(0,1)}(x)$. Con la muestra unitaria: 0.4745, concluya utilizando $\alpha = 0.05$.

Utilizando el Lema de Neyman-Pearson:

$$C^* = \left\{x / \lambda(x) = \frac{L(H_0)}{L(H_1)} \le k\right\}$$

1.
$$\lambda(x) = \frac{L(H_0)}{L(H_1)} = \frac{4\left(x-\frac{1}{4}\right)}{\frac{24}{25}\left(5x^2-\frac{5}{4}x\right)} = \frac{4\left(x-\frac{1}{4}\right)}{\frac{24}{5}x\left(x-\frac{1}{4}\right)} = \frac{20}{24x} \le k$$

$$\to x \ge \frac{20}{24k} \to x \ge k_1$$

La prueba más poderosa será:

$$\varphi^*(x) = \begin{cases} 1 & , \text{ si } x \ge k_1 \\ 0 & , \text{ o.m} \end{cases}$$

Cálculo de k_1:

$$\alpha = 0.05 = P_{H_0}(x \ge k_1) = \int_{k_1}^{1} 4\left(x-\frac{1}{4}\right)dx = 1-2k_1^2+k_1 = 0.05$$

$$\to 2k_1^2 + k_1 + 0.95 = 0 \to k_1 = 0.9831$$

Por lo tanto, la prueba más poderosa será:

$$\varphi^*(x) = \begin{cases} 1 & , \text{ si } x \ge 0.9831 \\ 0 & , \text{ o.m} \end{cases}$$

Con la muestra 0.4745 se concluye que no se rechaza la hipótesis nula.

Ejemplo 16: Con la variable aleatoria X definida como el ingreso mensual en um de recién egresados de Matemáticas se definen las hipótesis $H_0: f_1(x)=\frac{5}{22}\left(x+\frac{1}{5}\right)I_{(1,3)}(x)$ vs

$H_1: f_2(x)=\frac{3}{64}\left(2x^2+x\right)I_{(1,3)}(x)$. Con la muestra unitaria: 1.84822, concluya utilizando $\alpha=0.05$.

Utilizando el Lema de Neyman-Pearson:

1. $C^*=\left\{x/\lambda(x)=\frac{L(H_0)}{L(H_1)}\le k_1\right\}$

$$\lambda(x)=\frac{L(H_0)}{L(H_1)}=\frac{\frac{5}{22}\left(x+\frac{1}{5}\right)}{\frac{3}{64}\left(2x^2+x\right)}=\frac{160}{33}\frac{\left(x+\frac{1}{5}\right)}{\left(2x^2+x\right)}\le k_1$$

$$\rightarrow\frac{\left(x+\frac{1}{5}\right)}{\left(2x^2+x\right)}\le k\rightarrow 2kx^2+(k-1)x-\frac{1}{5}\ge 0\rightarrow$$

$$(x-k_2)(x-k_3)\ge 0$$

$$\underbrace{(x-k_2)\le 0 \text{ y } (x-k_3)\le 0}_{x\le r}\quad \text{o}\quad \underbrace{(x-k_2)\ge 0 \text{ y } (x-k_3)\ge 0}_{x\ge t}$$

La prueba más poderosa será:

$$\varphi^*(x)=\begin{cases}1 & , \text{ si } x\le r \;\cup\; x\ge t\\ 0 & , \text{ o.m}\end{cases}$$

Cálculo de r y t:

$$\frac{\alpha}{2}=0.025=P_{H_0}(x\le r)=\int_1^r\frac{5}{22}\left(x+\frac{1}{5}\right)dx\rightarrow\frac{5}{44}r^2+\frac{r}{22}-\frac{81}{440}=0$$

$\rightarrow r=1.0884$

$$\frac{\alpha}{2}=0.025=P_{H_0}(x\ge t)=\int_t^3\frac{5}{22}\left(x+\frac{1}{5}\right)dx\rightarrow-\frac{5\left(t+\frac{1}{5}\right)^2}{44}+\frac{64}{55}=0.025$$

$\rightarrow t=2.96544$

Por lo tanto, la prueba más poderosa será:

La prueba más poderosa será:

$$\varphi^*(x)=\begin{cases}1 & , \text{ si } x\le 1.0884098 \;\cup\; x\ge 2.96544\\ 0 & , \text{ o.m}\end{cases}$$

Con la muestra 1.84822 se concluye que no se rechaza la hipótesis nula.

Pruebas Óptimas para Hipótesis Simples vs Compuestas

En el ejemplo 13 de simples vs simples se comparaban las hipótesis $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1$, con $\theta_1>\theta_0$. Esto es equivalente a comparar una hipótesis simple vs compuesta $H_0:\theta=\theta_0$ vs $H_1:\theta>\theta_0$ y el procedimiento sería similar a lo hecho en el ejemplo 13.

Pruebas Óptimas para Hipótesis nula Compuesta

En este caso se sigue aplicando el Lema de Neyman Pearson, porque $\alpha=\max\limits_{\theta\in\Theta_0} P\left(\text{Rechazar } H_0;\theta\right)$ y esto ocurre cuando $\theta=\theta_0$.

Sea $X_1,\ldots,X_n$ una m.a. de $f(x;\theta)$ donde $\theta\in\Theta\subset\Re$ y suponga que se desea probar las hipótesis compuestas: Ho: $\theta\in\Theta_0$ vs. H1: $\theta\in\Theta_1$;donde:

$\Theta_0\cap\Theta_1=\phi$ y $\Theta_0\cup\Theta_1=\Theta$.

Prueba Uniformemente Más Poderosa (U.M.P.)

Se dice que φ^* es una prueba UMP de tamaño α para probar las hipótesis compuestas: Ho: $\theta\in\Theta_0$ vs. H1: $\theta\in\Theta_1$ si y solo si:

1. $\sup\limits_{\theta\in\Theta_0}\left\{\Pi_{\varphi^*}(\theta)\right\}=\alpha$
2. $\prod_{\varphi^*}(\theta_0)\geq\Pi_{\varphi}(\theta)$ $\forall\theta\in\Theta_1$ y para cualquier otra prueba φ tal que:

   $\sup\limits_{\theta\in\Theta_0}\left\{\Pi_{\varphi}(\theta)\right\}\leq\alpha$

Recordar que con hipótesis compuestas no siempre se encuentra una prueba uniformemente más poderosa, incluso aplicando las condiciones del Lema de Neyman Pearson.

TEOREMA

Si $X_1,\ldots,X_n$ una m.a. de una densidad que pertenece a la familia exponencial, entonces $f(x;\theta)=a(\theta)b(x)\exp\left\{c(\theta)d(x)\right\}$ $\theta\in\Theta\subset\Re$. Si T($\underline{x}$) = $\sum_{i=1}^{n}d(x_i)$ y se desea probar:

1) Si $c(\theta)$ es monótona creciente en θ y si existe $k>0$.

Si se quiere probar $H_0:\theta\leq\theta_0$ vs $H_1:\theta>\theta_0$

$$\varphi^*(\underline{x})=\begin{cases}1 & , \text{ si } T(\underline{x})>k\\ \gamma & , \text{ si } T(\underline{x})=k\\ 0 & , \text{ si } T(\underline{x})<k\end{cases}$$

entonces $\varphi^*(\underline{x})$ es una prueba UMP.

Donde k y γ $\left(k>0 \;\;,\;\; 0<\gamma<1\right)$ son hallados con

$E_{\theta_0}\left[\varphi^*\left(\underline{x}\right)\right]=P_{\theta_0}\left[T>k\right]+\gamma P_{\theta_0}\left[T=k\right]=\alpha$.

El poder de la prueba está dado por:

$\pi_{\varphi^*}\left(\theta\right)=P_{\theta}\left[T>k\right]+\gamma P_{\theta}\left[T=k\right]$ con $\theta>\theta_0$

Si se quiere probar $H_0:\theta\geq\theta_0$ vs $H_1:\theta<\theta_0$

$$\varphi^*\left(\underline{x}\right)=\begin{cases}1 & ; \quad T(\underline{x})<k\\ \gamma & ; \quad T(\underline{x})=k\\ 0 & ; \quad T(\underline{x})>k\end{cases}$$

entonces $\varphi^*\left(\underline{x}\right)$ es una prueba UMP.

Donde k y γ $\left(k>0 \;\;,\;\; 0<\gamma<1\right)$ son hallados con

$E_{\theta_0}\left[\varphi^*\left(\underline{x}\right)\right]=P_{\theta_0}\left[T<k\right]+\gamma P_{\theta_0}\left[T=k\right]=\alpha$.

El poder de la prueba está dado por:

$\pi_{\varphi^*}\left(\theta\right)=P_{\theta}\left[T<k\right]+\gamma P_{\theta}\left[T=k\right]$ con $\theta<\theta_0$

2) si $c(\theta)$ es monótona decreciente en θ y si existe $k>0$.

Si se quiere probar $H_0:\theta\leq\theta_0$ vs $H_0:\theta>\theta_0$

$$\varphi^*\left(\underline{x}\right)=\begin{cases}1 & ; \quad T(\underline{x})<k\\ \gamma & ; \quad T(\underline{x})=k\\ 0 & ; \quad T(\underline{x})>k\end{cases}$$

entonces $\varphi^*\left(\underline{x}\right)$ es una prueba UMP.

Donde k y γ $\left(k>0 \;\;,\;\; 0<\gamma<1\right)$ son hallados con

$E_{\theta_0}\left[\varphi^*\left(\underline{x}\right)\right]=P_{\theta_0}\left[T<k\right]+\gamma P_{\theta_0}\left[T=k\right]=\alpha$.

El poder de la prueba está dado por:

$\pi_{\varphi^*}\left(\theta\right)=P_{\theta}\left[T<k\right]+\gamma P_{\theta}\left[T=k\right]$ con $\theta<\theta_0$

Si se quiere probar $H_0:\theta\geq\theta_0$ vs $H_0:\theta<\theta_0$

$$\varphi^*\left(\underline{x}\right) = \begin{cases}1 & , \quad \text{si } T(\underline{x})>k\\ \gamma & , \quad \text{si } T(\underline{x})=k\\ 0 & , \quad \text{si } T(\underline{x})<k\end{cases}$$

entonces $\varphi^*\left(\underline{x}\right)$ es una prueba UMP.

Donde k y γ $(k>0\ ,\ 0<\gamma<1)$ son hallados con
$E_{\theta_0}\left[\varphi^*(\underline{x})\right]=P_{\theta_0}\left[T>k\right]+\gamma P_{\theta_0}\left[T=k\right]=\alpha$.

El poder de la prueba está dado por:
$\pi_{\varphi^*}(\theta)=P_\theta\left[T>k\right]+\gamma P_\theta\left[T=k\right]$ con $\theta>\theta_0$

Ejemplo 17: La siguiente muestra: 14.2, 9.1, 2.2, 13.3, 6.2 corresponde a los tiempos, en segundos, de respuesta en cierto terminal de computadora en línea (es el tiempo entre el fin de la consulta del usuario y el principio de la respuesta del sistema a esa consulta). Asumiendo que la muestra se extrajo de una distribución exponencial con media $\frac{1}{\text{tasa}}$ $(\text{o tasa } \theta)$.

a. Halle una prueba uniformemente más poderosa para probar $H_0:\theta\geq\frac{1}{5}$ vs $H_1:\theta<\frac{1}{5}$. Use $\alpha=0.05$.

$$f(x)=\theta\times 1\times\exp(-\theta\times x)$$
$$a(\theta)=\theta\ ,\ b(x)=1\ ,\ c(\theta)=-\theta\ ,\ d(x)=x$$
$$c(\theta)=-\theta \text{ es monótona decreciente}\rightarrow d(x)=x\rightarrow T=\sum_{i=1}^{5}x_i$$
$\rightarrow$ La prueba más poderosa será:
$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & ,\ \text{si } \sum_{i=1}^{5}x_i\geq k\\ 0 & ,\ \text{o.m}\end{cases}$$
$$\sum_{i=1}^{n}X_i\sim \text{Gamma}(n,\theta)$$
$$\sum_{i=1}^{5}x_i=45 \text{ y } n=5$$
$$\text{pvalor}=P_{\theta_0}\left(\sum_{i=1}^{5}x_i\geq 45\right)=P_{\theta_0}\left[2\left(\sum_{i=1}^{5}x_i\right)\theta_0\geq 2(45)\left(\frac{1}{5}\right)\right]=P_{\theta_0}\left[\chi^2(10)\geq 18\right]=$$
$$\text{pvalor}=0.0549>\alpha=0.05\rightarrow \text{No se rechaza la } H_0$$
$$\text{También: } 18<\chi^2(0.95,10)=18.307$$

Este problema se puede resolver utilizando el teorema de Neyman-Pearson. Cosa que se hará a continuación:
$$\theta_0=\frac{1}{5} \quad \text{y} \quad \theta<\frac{1}{5}$$

1. $C^* = \left\{ \left(x_1,...,x_n\right) / \lambda\left(\underline{x}\right) = \dfrac{L\left(\theta_0, \underline{x}\right)}{L\left(\theta, \underline{x}\right)} \le k \right\}$

$$\lambda\left(\underline{x}\right) = \frac{L\left(\theta_0 = \dfrac{1}{5}\right)}{L\left(\theta < \dfrac{1}{5}\right)} = \frac{\prod_{i=1}^{5} \theta_0 \exp\left(-\theta_0 x_i\right)}{\prod_{i=1}^{5} \theta \exp\left(-\theta x_i\right)} = \frac{\left(\theta_0\right)^n \exp\left(-\theta_0 \sum_{i=1}^{5} x_i\right)}{\left(\theta\right)^n \exp\left(-\theta \sum_{i=1}^{5} x_i\right)} =$$

$$\lambda\left(\underline{x}\right) = \left(\frac{\theta}{\theta_0}\right)^n \exp\left[\left(\theta - \theta_0\right)\sum_{i=1}^{5} x_i\right] \le k$$

$$\exp\left[\left(\theta - \theta_0\right)\sum_{i=1}^{5} x_i\right] \le k_1$$

Tomando logaritmos en la desigualdad:

$$\left(\theta - \theta_0\right)\sum_{i=1}^{5} x_i \le k_3 \rightarrow \sum_{i=1}^{5} x_i \ge c \ \text{ porque } \left(\theta - \theta_0\right) < 0$$

La $\sum_{i=1}^{n} x_i$ es una estadística suficiente.

Entonces la prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \ , \ \text{si } \sum_{i=1}^{5} x_i \ge c \\ 0 \ , \ \text{o.m} \end{cases}$$

Lo que sigue es igual a lo que ya se hizo anteriormente.

b. Grafique el poder de la prueba. Incluya en el gráfico el poder cuando $\theta = \dfrac{1}{10.2}$.

$c = \text{qgamma}\left(0.05,\ n,\ \theta,\ \text{lower.tail = F}\right)$

$\theta_1 = \dfrac{1}{10.2}$

$\text{Poder1} = P\left(\sum_{i=1}^{n} X_i > c\right) = P\left[\text{Gamma}\left(n, \theta_1\right) > c\right]$

$\text{Poder} = P\left(\sum_{i=1}^{n} X_i > c\right) = P\left[\text{Gamma}\left(n, \theta\right) > c\right]$

Ver el Script R.

c. Halle la prueba uniformemente más poderosa para probar $H_0 : \theta \le \dfrac{1}{5}$ vs $H_1 : \theta > \dfrac{1}{5}$. Use $\alpha = 0.05$.

$f(x)=\theta\times 1\times \exp(-\theta\times x)$

$a(\theta)=\theta$, $b(x)=1$, $c(\theta)=-\theta$, $d(x)=x$

$c(\theta)=-\theta$ es monótona decreciente $\rightarrow d(x)=x\rightarrow T=\sum_{i=1}^{5}x_i$

$\rightarrow$ La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 \ , \ \text{si} \ \sum_{i=1}^{5}x_i\le k\\ 0 \ , \ \text{o.m}\end{cases}$$

$\sum_{i=1}^{5}x_i=45$ y $n=5$

$$\text{pvalor}=P_{\theta_0}\left(\sum_{i=1}^{5}x_i\le 45\right)=P_{\theta_0}\left[2\left(\sum_{i=1}^{5}x_i\right)\theta_0\le 2(45)\left(\frac{1}{5}\right)\right]=P_{\theta_0}\left[\chi^2(10)\le 18\right]=0.9450364$$

$\text{pvalor}=0.9450364>\alpha=0.05\rightarrow$ No se rechaza la H_0

También: $18>\chi^2(0.05,10)=3.9402991$

Este problema se puede resolver utilizando el teorema de Neyman-Pearson. Cosa que se hará a continuación:

$\theta_0=\frac{1}{5}$ y $\theta>\frac{1}{5}$

1. $C^*=\left\{(x_1,...,x_n)/\lambda(\underline{x})=\frac{L(\theta_0,\underline{x})}{L(\theta,\underline{x})}\le k\right\}$

$$\lambda(\underline{x})=\frac{L\left(\theta_0=\frac{1}{5}\right)}{L\left(\theta>\frac{1}{5}\right)}=\frac{\prod_{i=1}^{5}\theta_0\exp(-\theta_0 x_i)}{\prod_{i=1}^{5}\theta\exp(-\theta x_i)}=\frac{(\theta_0)^n\exp\left(-\theta_0\sum_{i=1}^{5}x_i\right)}{(\theta)^n\exp\left(-\theta\sum_{i=1}^{5}x_i\right)}=$$

$$\lambda(\underline{x})=\left(\frac{\theta}{\theta_0}\right)^n\exp\left[(\theta-\theta_0)\sum_{i=1}^{5}x_i\right]\le k$$

$$\exp\left[(\theta-\theta_0)\sum_{i=1}^{5}x_i\right]\le k_1$$

Tomando logaritmos en la desigualdad:

$$(\theta-\theta_0)\sum_{i=1}^{5}x_i\le k_3\rightarrow\sum_{i=1}^{5}x_i\le c \ \ \text{porque} \ (\theta-\theta_0)>0$$

Entonces la prueba más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right)=\begin{cases}1 \quad , \quad \text{si } \sum_{i=1}^{5} x_i \le c \\ 0 \quad , \quad \text{o.m}\end{cases}$$

Lo que sigue es igual a lo que ya se hizo anteriormente.

d. Halle la prueba uniformemente más poderosa para probar $H_0:\theta=\frac{1}{5}$ vs $H_1:\theta\neq\frac{1}{5}$. Use $\alpha=0.05$.

En este caso la hipótesis planteada puede ser $H_0^1:\theta\ge\frac{1}{5}$ o $H_0^2:\theta\le\frac{1}{5}$. Considerando los resultados de las subpreguntas a y c se obtendrá la siguiente prueba más poderosa:

$$\varphi^*\left(x_1,\ldots,x_n\right)=\begin{cases}1 \quad , \quad \text{si } \sum_{i=1}^{5} x_i \le k_1 \ \text{ o } \ \sum_{i=1}^{5} x_i \ge k_2 \\ 0 \quad , \quad \text{o.m}\end{cases}$$

$$\sum_{i=1}^{5} x_i = 45 \ \text{ y } \ n=5$$

Como $\sum_{i=1}^{5} X_i \sim \text{Gamma}\left(r=5,\lambda=\frac{1}{5}\right)$

$$P\left(\sum_{i=1}^{5} x_i \le k_1 \ \text{ o } \ \sum_{i=1}^{5} x_i \ge k_2\right)=P\left(\sum_{i=1}^{5} x_i \le k_1\right)+P\left(\sum_{i=1}^{5} x_i \ge k_2\right)=$$

$$=P\left(2\times\sum_{i=1}^{5} x_i\times\theta_0 \le 2k_1\times\frac{1}{5}\right)+P\left(2\times\sum_{i=1}^{5} x_i\times\theta_0 \ge 2k_2\times\frac{1}{5}\right)=$$

$$=P\left(\chi^2(10)\le 2k_1\times\frac{1}{5}\right)+P\left(\chi^2(10)\ge 2k_2\times\frac{1}{5}\right)=\frac{0.05}{2}+\frac{0.05}{2}=0.025+0.025$$

$$\rightarrow 2k_1\times\frac{1}{5}=3.2469728 \qquad \text{y} \qquad 2k_2\times\frac{1}{5}=20.483177$$

$$\rightarrow k_1=8.117432 \qquad \text{y} \qquad k_2=51.2079425$$

Entonces:

$$\varphi^*\left(x_1,\ldots,x_n\right)=\begin{cases}1 \quad , \quad \text{si } \sum_{i=1}^{5} x_i \le 8.117432 \ \text{ o } \ \sum_{i=1}^{5} x_i \ge 51.2079425 \\ 0 \quad , \quad \text{o.m}\end{cases}$$

$$\sum_{i=1}^{5} x_i = 45 \rightarrow \text{No se rechaza la } H_0$$

e. Con $Y_1 = \min(X_1, \ldots, X_5)$ derive una prueba de hipótesis de tamaño $\alpha = 0.05$ para verificar $H_0 : \theta \geq \frac{1}{5}$ vs $H_1 : \theta < \frac{1}{5}$. Asuma que se rechaza la H_0, en favor de H_1, si $Y_1 > c$ para algún c determinado. Nota: $Y_1 \sim \text{Exponencial}(\text{tasa} = n\theta)$

Hallando c:

$$\alpha = \max P(\text{Error tipo I}) = \max P(\text{Rechazar } H_0 \text{ verdadera}) = \max P(Y_1 > c \ ; \ \theta) =$$

$$\alpha = \max_{\lambda \leq \lambda_0}\left[1 - \left(1 - e^{-n\theta c}\right)\right] = e^{-n\theta_0 c} \rightarrow c = -\frac{1}{n\theta_0}\ln(\alpha)$$

$$\beta = \max P(\text{Error tipo II}) = \max P(\text{No rechazar } H_0 \text{ cuando es falsa}) = \max_{\theta > \theta_0} P(Y_1 \leq c \ ; \ \theta) =$$

$$\beta = \max_{\theta > \theta_0} P\left(Y_1 \leq -\frac{1}{n\theta_0}\ln(\alpha) \ ; \ \theta\right) = \max_{\theta > \theta_0}\left[P\left(Y_1 \leq -\frac{1}{n\theta_0}\ln(\alpha) \ ; \ \theta\right)\right] =$$

$$\beta = \max_{\theta > \theta_0}\left(1 - e^{-n\theta\left(-\frac{1}{n\theta_0}\ln(\alpha)\right)}\right) = \max_{\theta > \theta_0}\left[1 - e^{\ln\left[(\alpha)^{\frac{\theta}{\theta_0}}\right]}\right] = \max_{\theta > \theta_0}\left[1 - (\alpha)^{\frac{\theta}{\theta_0}}\right] = 1 - (\alpha)^{\frac{\theta_0}{\theta_0}} = (1 - \alpha)$$

$$\Pi(\theta) = P(\text{Rechazar } H_0 \ ; \ \theta) = P\left(Y_1 \geq -\frac{1}{n\theta_0}\ln(\alpha) \ ; \ \theta\right) = e^{-n\theta\left(-\frac{1}{n\theta_0}\ln(\alpha)\right)} =$$

$$\Pi(\theta) = (\alpha)^{\frac{\theta}{\theta_0}}$$

En el Script R ver la comparación gráfica de las funciones Potencia de la preguntas a y e.

Ejemplo 18: En cierta ciudad la proporción de microbuseros que tienen por lo menos cuarenta infracciones de tránsito es p. En una muestra aleatoria de 18 microbuseros 13 tienen por lo menos cuarenta infracciones.

a. Con una prueba no aleatorizada pruebe $H_0 : p \geq 0.84$ vs $H_1 : p < 0.84$. Use $\alpha = 0.10$.

$$X_1, \ldots, X_{18} \rightarrow X_i \sim \text{Binomial}(1, p) \rightarrow$$

$$f(x) = p^x (1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = (1-p) \times 1 \times \exp\left\{\left[\ln\left(\frac{p}{1-p}\right)\right] \times x\right\}$$

$$a(p) = 1 - p \ \ , \ b(x) = 1 \ \ , \ c(p) = \ln\left(\frac{p}{1-p}\right) \ \ , \ d(x) = x$$

$$c(p) = \ln\left(\frac{p}{1-p}\right) \text{ es monótona creciente} \rightarrow d(x) = x \rightarrow T = \sum_{i=1}^{18} d(x_i) = \sum_{i=1}^{18} X_i$$

La $\sum_{i=1}^{18} X_i$ es una estadística suficiente.

$\rightarrow$ La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{18} x_i \le k \\ 0 & , \text{ o.m}\end{cases}$$

Nota: Con variables discretas, en general, el tamaño de la prueba es diferente al nivel de significación

$$\sum_{i=1}^{18} x_i = 13 \ , \ \sum_{i=1}^{18} x_i \sim \text{Binomial}(18, p)$$

$$\text{pvalor} = P_{p=0.84}\left(\sum_{i=1}^{18} x_i \le 13\right) = \sum_{y=0}^{13}\binom{18}{y} 0.84^y \times 0.16^{18-y} = 0.1482$$

$$\text{pvalor} = 0.1482 > \alpha = 0.10 \rightarrow \text{No se rechaza la } H_0 .$$

b. Con una prueba no aleatorizada pruebe $H_0: p \le 0.84$ vs $H_1: p > 0.84$. Use $\alpha = 0.10$.

$$X_1,\ldots,X_{18} \rightarrow X_i \sim \text{Binomial}(1, p) \rightarrow$$

$$f(x) = p^x(1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = (1-p)\times 1 \times \exp\left\{\left[\ln\left(\frac{p}{1-p}\right)\right]\times x\right\}$$

$$a(p) = 1-p \ , \ b(x) = 1 \ , \ c(p) = \ln\left(\frac{p}{1-p}\right) \ , \ d(x) = x$$

$$c(p) = \ln\left(\frac{p}{1-p}\right) \text{ es monótona creciente} \rightarrow d(x) = x \rightarrow T = \sum_{i=1}^{18} d(x_i) = \sum_{i=1}^{18} X_i$$

$\rightarrow$ La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{18} x_i \ge k \\ 0 & , \text{ o.m}\end{cases}$$

Nota: Con variables discretas, en general, el tamaño de la prueba es diferente al nivel de significación

$$\sum_{i=1}^{18} x_i = 13 \ , \ \sum_{i=1}^{18} x_i \sim \text{Binomial}(18, p)$$

$$\text{pvalor} = P_{p=0.84}\left(\sum_{i=1}^{18} x_i \ge 13\right) = \sum_{y=13}^{18}\binom{18}{y} 0.84^y \times 0.16^{18-y} = 0.9449$$

$$\text{pvalor} = 0.9449 > \alpha = 0.10 \rightarrow \text{No se rechaza la } H_0 .$$

c. Pruebe $H_0: p = 0.84$ vs $H_1: p \ne 0.84$. Use $\alpha = 0.10$.

En este caso la hipótesis planteada puede ser $H_0^1: p \ge 0.84$ o $H_0^2: p \le 0.84$.

Considerando los resultados de las subpreguntas a y b se obtendrá la siguiente prueba más poderosa:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{18} x_i \le k_1 \text{ o } \sum_{i=1}^{18} x_i \ge k_2 \\ 0 & , \text{ o.m}\end{cases}$$

$$\sum_{i=1}^{18} x_i = 13 \;,\; \sum_{i=1}^{18} x_i \sim \text{Binomial}(18, p)$$

$$0.10 = P_{p=0.84}\left(\sum_{i=1}^{18} x_i \le k_1 \text{ o } \sum_{i=1}^{18} x_i \ge k_2\right) = P_{p=0.84}\left(\sum_{i=1}^{18} x_i \le k_1\right) + P_{p=0.84}\left(\sum_{i=1}^{18} x_i \ge k_2\right) =$$

$$= \frac{0.10}{2} + \frac{0.10}{2} = 0.05 + 0.05 \to \text{En forma aproximada: } k_1 = 12 \text{ y } k_2 = 17$$

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{18} x_i \le 12 \text{ o } \sum_{i=1}^{18} x_i \ge 17 \\ 0 & , \text{ o.m}\end{cases}.$$

Como $\sum_{i=1}^{18} x_i = 13 \to$ No se rechaza la H_0

d. Con una prueba randomizada pruebe $H_0 : p \ge 0.84$ vs $H_1 : p < 0.84$. Use $\alpha = 0.10$.

$$X_1,\ldots,X_{18} \to X_i \sim \text{Binomial}(1, p) \to$$

$$f(x) = p^x(1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = (1-p)\times 1 \times \exp\left\{\left[\ln\left(\frac{p}{1-p}\right)\right]\times x\right\}$$

$$a(p) = 1-p \;,\; b(x) = 1 \;,\; c(p) = \ln\left(\frac{p}{1-p}\right) \;,\; d(x) = x$$

$$c(p) = \ln\left(\frac{p}{1-p}\right) \text{ es monótona creciente} \to d(x) = x \to T = \sum_{i=1}^{18} d(x_i) = \sum_{i=1}^{18} X_i$$

Cálculo del valor de k :
De acuerdo al enunciado se tiene que:

$$Y = \sum_{i=1}^{18} x_i \sim \text{Binomial}(n = 18, p = 0.84)$$

$$\alpha = 0.10 = P_{p=0.84}\left(Y < k\right) = \sum_{y=0}^{k-1}\binom{18}{y}0.84^{y}\times 0.16^{18-y} \rightarrow k-1 = 13 \rightarrow k = 14$$

$$\left[\text{qbinom}\left(0.10,18,0.84\right) = 13\right]$$

$$P_{p=0.84}\left(Y < 14\right) = P_{p=0.84}\left(Y \leq 13\right) = \sum_{y=0}^{13}\binom{18}{y}0.84^{y}\times 0.16^{18-y} = 0.1482$$

El anterior será:

$$P_{p=0.84}\left(Y < 13\right) = P_{p=0.84}\left(Y \leq 12\right) = 0.0551$$

$$P_{p=0.84}\left(Y = 13\right) = 0.0931$$

$\rightarrow$ La prueba randomizada más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , \quad \text{si } \sum_{i=1}^{18} x_i < 13 \\ \gamma \quad , \quad \text{si } \sum_{i=1}^{18} x_i = 13 \\ 0 \quad , \quad \text{si } \sum_{i=1}^{18} x_i > 13 \end{cases}$$

Cálculo del valor de γ.

$$E_{\theta_0}\left[\varphi^*\left(\underline{x}\right)\right] = P_{\theta_0}\left[T < k\right] + \gamma P_{\theta_0}\left[T = k\right] = \alpha$$

$$E_{p_0}\left[\varphi^*\left(\underline{x}\right)\right] = P_{p_0=0.84}\left[\sum_{i=1}^{18} x_i < 13\right] + \gamma P_{p_0=0.84}\left[\sum_{i=1}^{18} X_i = 13\right] = 0.10.$$

$$E_{p_0}\left[\varphi^*\left(\underline{x}\right)\right] = 0.0551 + \gamma\left(0.0931\right) = 0.10 \rightarrow \gamma = 0.4823$$

$\rightarrow$ La prueba randomizada más poderosa será:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 \quad , \quad \text{si } \sum_{i=1}^{18} x_i < 13 \\ 0.4823 \quad , \quad \text{si } \sum_{i=1}^{18} x_i = 13 \\ 0 \quad , \quad \text{si } \sum_{i=1}^{18} x_i > 13 \end{cases}$$

Con $\sum_{i=1}^{18} x_i = 13$, existe una probabilidad de 0.4823 de rechazo o no rechazo de la $H_0 : p \geq 0.84$.

El poder de la prueba está dado por:

$$\pi_{\varphi^*}(p) = P_p\left[\sum_{i=1}^{18} X_i < 13\right] + 0.4823 \times P_p\left[\sum_{i=1}^{18} X_i = 13\right] \text{ con } p < 0.84$$

NOTA: Ver en el Script R la forma de hallar k=13

Ejemplo 19: El caudal máximo de un rio pequeño tiene densidad: $f(x) = \alpha \exp\left\{-\alpha(x-\mu) - \exp\left[-\alpha(x-\mu)\right]\right\} I_{(-\infty,\infty)}(x)$, con $\alpha = 0.0256$. Si se toma una muestra aleatoria de 50 años y se obtiene $\sum_{i=1}^{50} \exp(-\alpha x_i) = 8.32$. Verifique: $H_0 : \mu \geq 79$ pies^3 vs $H_1 : \mu < 79$ pies^3. Use el nivel de significación de 0.05.

$$f(x) = \alpha \exp\left\{-\alpha(x-\mu) - \exp\left[-\alpha(x-\mu)\right]\right\} =$$

$$f(x) = \alpha \exp(\alpha\mu) \times \exp(-\alpha x) \times \exp\left\{\left[-\exp(\alpha\mu)\right] \times \left[\exp(-\alpha x)\right]\right\}$$

$$a(\mu) = \alpha \exp(\alpha\mu), b(x) = \exp(-\alpha x), c(\mu) = -\exp(\alpha\mu), d(x) = \exp(-\alpha x)$$

$c(\mu) = -\exp(\alpha\mu)$ es monótona decreciente en μ.

Se demuestra que $d(x) = \exp(-\alpha x) \sim$ Exponencial(con media: $\exp(-\alpha\mu)$)

$$\rightarrow \sum_{i=1}^{50} \exp(-\alpha x_i) \sim \text{Gamma}\left(n = 50, \exp(\alpha\mu)\right) \rightarrow$$

$$2\left[\sum_{i=1}^{50} \exp(-\alpha x_i)\right] \exp(\alpha\mu) \sim \chi^2(2n = 100)$$

$\rightarrow$ La prueba más poderosa será:

$$\varphi^*(x_1, \ldots, x_n) = \begin{cases} 1 \quad , \quad \text{si } \sum_{i=1}^{50} \exp(-\alpha x_i) \geq k \rightarrow 2\left[\sum_{i=1}^{50} \exp(-\alpha x_i)\right] \exp(\alpha\mu) \geq c \\ 0 \quad , \quad \text{o.m} \end{cases}$$

$$\chi_c^2 = 2[8.32] \exp(0.0256 \times 79) = 125.74$$

$$\text{pvalor} = P_{\mu=79}\left[\chi^2(100) \geq 125.74\right] = 0.0418 < 0.05 \rightarrow \text{Se rechaza la } H_0$$

Ejemplo 20: Se hace el siguiente experimento: Considere que un “intento” consiste en determinar el número de vicuñas analizadas con un contenido de colesterol menor de 78 mg/DL antes de encontrar la tercera vicuña con un contenido de colesterol de por lo menos 78 mg/DL. En una muestra aleatoria de 5 “intentos” se encontraron los siguientes valores 2, 1, 3, 1 y 2. Suponga que en el $\pi \times 100\%$ de las vicuñas tienen un contenido de colesterol de por lo menos 78 mg/DL. Haga una prueba aleatoria para verificar $H_0 : \pi = 0.4$ vs $H_1 : \pi > 0.4$ con un nivel de significación de 1.3%.

Solución

$X_1,\ldots,X_5 \to X_i \sim \text{Pascal}(\text{k}=3,\pi) \to$

$f(x) = \binom{3+x-1}{x} \pi^3 (1-\pi)^x I_{\{0,1,2,\ldots\}}(x) = \pi^3 \times \binom{3+x-1}{x} \times \exp\{[\ln(1-\pi)]\times x\}$

$a(\pi)=\pi^3 \;\; , \;\; b(x)=\binom{3+x-1}{x} \; , \;\; c(\pi)=\ln(1-\pi) \;\; , \;\; d(x)=x$

$c(\pi)=\ln(1-\pi)$ es monótona decreciente $\to d(x)=x \to T=\sum_{i=1}^{5} d(x_i)=\sum_{i=1}^{5} X_i$

La $\sum_{i=1}^{5} X_i$ es una estadística suficiente.

$\to$ La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases} 1 \;\; , \;\; \text{si} \; \sum_{i=1}^{5} x_i < k \\ \gamma \;\; , \;\; \text{si} \; \sum_{i=1}^{5} x_i = k \\ 0 \;\; , \;\; \text{si} \; \sum_{i=1}^{5} x_i > k \end{cases}$$

Cálculo del valor de k :

De acuerdo al enunciado se tiene que:

$Y=\sum_{i=1}^{5} x_i \sim \text{Pascal}(nk=15,\pi=0.4) \to f_Y(y)=\binom{15+y-1}{\text{y}} \pi^{15}(1-\pi)^y I_{\{0,1,2,\ldots\}}(y)$

$\alpha=0.013=P_{\pi=0.4}(Y<k)=\sum_{y=0}^{k-1}\binom{15+y-1}{\text{y}}\pi^{15}(1-\pi)^y \to k-1=9 \to k=10$

$P_{\pi=0.4}(Y<10)=P_{\pi=0.4}(Y\le 9)=\sum_{y=0}^{9}\binom{15+y-1}{\text{y}}\pi^{15}(1-\pi)^y=0.0216$

$P_{\pi=0.4}(Y\le 8)=\sum_{y=0}^{9}\binom{15+y-1}{\text{y}}\pi^{15}(1-\pi)^y=0.01281$

$P_{\pi=0.4}(Y=9)=0.008842685251$

Cálculo del valor de γ .

$E_{\pi_0}\left[\varphi^*(\underline{x})\right]=P_{\pi_0}[T<k]+\gamma P_{\pi_0}[T=k]=\alpha$

$E_{\pi_0}\left[\varphi^*(\underline{x})\right]=P_{\pi=0.4}\left[\sum_{i=1}^{5} x_i<9\right]+\gamma P_{\pi=0.4}\left[\sum_{i=1}^{5} X_i=9\right]=0.013$.

$E_{\pi_0}\left[\varphi^*(\underline{x})\right]=0.01281429234+\gamma(0.008842685251)=0.013 \to \gamma=0.02100127447$

$\rightarrow$ La prueba más poderosa será:

$$\varphi^*(x_1,\ldots,x_n)=\begin{cases}1 & , \text{ si } \sum_{i=1}^{5} x_i < 9 \\ 0.02100127447 & , \text{ si } \sum_{i=1}^{5} x_i = 9 \\ 0 & , \text{ si } \sum_{i=1}^{5} x_i > 9\end{cases}$$

Como el valor calculado en la muestra es $\sum_{i=1}^{5} x_i = 9$, entonces a veces no se rechaza y a veces se rechaza la hipótesis planteada con probabilidad 0.02100127447.

<u>Familia de Densidades con la Propiedad de Razón de Verosimilitud Monótona (RVM)</u>

Una familia de densidades $f(x;\theta)$; $\theta \in \Theta \subset \Re$ tiene la propiedad de RVM si existe T(x) tal que:

$$\Psi = \frac{L(\theta_1,\underline{x})}{L(\theta_2,\underline{x})} = \frac{\prod_{i=1}^{n} f(x_i,\theta_1)}{\prod_{i=1}^{n} f(x_i,\theta_2)}$$

Es una función creciente o decreciente en T; $\forall \theta_1 < \theta_2$.

TEOREMA

Sea $X_1,\ldots,X_n$ una m.a. de $f(x;\theta)$ donde $\theta \in \Theta \subset \Re$. suponga que $f(x;\theta)$ pertenece a una familia de densidades con propiedad de RVM en T(x). Si se desea probar:

1. $H_0:\theta \le \theta_0$ vs $H_1:\theta > \theta_0$

Si la RVM es creciente en T(x) entonces.

$$\varphi^*(\underline{x}) = \begin{cases}1 & , \text{ si } T(\underline{x}) \le k^* \\ 0 & , \text{ o.m.}\end{cases}$$

$\varphi^*(\underline{x})$ es una prueba Uniformemente Más Poderosa (UMP)

Si la RVM es decreciente en T(x) entonces.

$$\varphi^*\left(\underline{x}\right) = \begin{cases} 1 & , \quad \text{si } T(\underline{x}) \geq k^* \\ 0 & , \quad \text{o.m.} \end{cases}$$

$\varphi^*\left(\underline{x}\right)$ es una prueba UMP.

2. $H_0 : \theta \geq \theta_0$ vs $\theta < \theta_0$

Si la RVM es creciente en T(x) entonces.

$$\varphi^*\left(\underline{x}\right) = \begin{cases} 1 & , \quad \text{si } T(\underline{x}) \geq k^* \\ 0 & , \quad \text{o.m.} \end{cases}$$

$\varphi^*\left(\underline{x}\right)$ es una prueba UMP

Si la RVM es decreciente en T(x) entonces.

$$\varphi^*\left(\underline{x}\right) = \begin{cases} 1 & , \quad \text{si } T(\underline{x}) \leq k^* \\ 0 & , \quad \text{o.m.} \end{cases}$$

$\varphi^*\left(\underline{x}\right)$ es una prueba UMP.

Ejemplo 21: En cierta Universidad el tiempo X, en minutos, que transcurre entre la hora de inicio programada y la hora en que empiezan las clases tiene distribución $\text{Uniforme}\left(0,\theta\right)$.

a. Si se toma una muestra al azar de 10 clases y se obtiene: 0.81, 0.32, 1.25, 1.92, 1.12, 0.45, 1.99, 0.00, 0.00, 1.25. Pruebe $H_0 : \theta \leq 2 \quad vs \quad H_1 : \theta > 2$. Use $\alpha = 0.05$.

$f\left(x\right) = \dfrac{1}{\theta} \notin$ a la familia exponencial.

Para $\theta_1 < \theta_2$

$$\Psi = \frac{L\left(\theta_1\right)}{L\left(\theta_2\right)} = \frac{\left(\dfrac{1}{\theta_1}\right)^n I_{(0,\theta_1)}\left(Y_n\right)}{\left(\dfrac{1}{\theta_2}\right)^n I_{(0,\theta_2)}\left(Y_n\right)} = \left(\frac{\theta_2}{\theta_1}\right)^n \frac{I_{(0,\theta_1)}\left(Y_n\right)}{I_{(0,\theta_2)}\left(Y_n\right)} = \Psi\left(Y_n\right)$$

$$\Psi\left(Y_n\right) = \begin{cases} \left(\dfrac{\theta_2}{\theta_1}\right)^n, & 0 < Y_n < \theta_1 \\ 0 \quad , & \theta_1 < Y_n < \theta_2 \end{cases}$$

$\Psi\left(Y_n\right)$ es decreciente en Y_n

$$\varphi\left(\underset{\sim}{x}\right) = \begin{cases} 1, & Y_n \geq k \\ 0, & \text{de otro modo} \end{cases}$$

$$\alpha = P_{\theta=2}\left(Y_n \geq k\right) = 1 - P_{\theta=2}\left(Y_n < k\right) = 0.05 \rightarrow P_{\theta=2}\left(Y_n < k\right) = 0.95 = \int_0^k f\left(Y_n\right) dY_n$$

teniendo en cuenta que $n = 10$:

$$0.95 = \int_0^k f\left(Y_n\right) dY_n = \int_0^k \frac{10Y_n^9}{2^{10}} dY_n = \left(\frac{k}{2}\right)^{10} = 0.95 \rightarrow k = 1.989768$$

$$\varphi\left(\underset{\sim}{x}\right) = \begin{cases} 1, & Y_n \geq 1.989768 \\ 0, & \text{de otro modo} \end{cases}$$

Como en la muestra $Y_n = 1.99 \rightarrow$ se rechaza la H_0

Otra forma de concluir:

$$\text{pvalor} = P_{\theta=2}\left(Y_n \geq 1.99\right) = \int_{1.99}^{2} \frac{10Y_n^9}{2^n} dY_n = 0.048 < \alpha = 0.05 \rightarrow \text{Se rechaza } H_0.$$

b. Si se toma una muestra al azar de 10 clases y se obtiene: 0.81, 0.32, 1.25, 1.92, 1.12, 0.45, 1.99, 0.00, 0.00, 1.25. Pruebe $H_0 : \theta \geq 2 \quad vs \quad H_1 : \theta < 2$. Use $\alpha = 0.05$.

$f\left(x\right) = \dfrac{1}{\theta} \notin$ a la familia exponencial.

Para $\theta_1 < \theta_2$

$$\Psi = \frac{L\left(\theta_1\right)}{L\left(\theta_2\right)} = \frac{\left(\frac{1}{\theta_1}\right)^n I_{\left(0,\theta_1\right)}\left(Y_n\right)}{\left(\frac{1}{\theta_2}\right)^n I_{\left(0,\theta_2\right)}\left(Y_n\right)} = \left(\frac{\theta_2}{\theta_1}\right)^n \frac{I_{\left(0,\theta_1\right)}\left(Y_n\right)}{I_{\left(0,\theta_2\right)}\left(Y_n\right)} = \Psi\left(Y_n\right)$$

$$\Psi\left(Y_n\right) = \begin{cases} \left(\dfrac{\theta_2}{\theta_1}\right)^n, & 0 < Y_n < \theta_1 \\ 0 \quad , & \theta_1 < Y_n < \theta_2 \end{cases}$$

$\Psi\left(Y_n\right)$ es decreciente en Y_n

$$\varphi\left(\underset{\sim}{x}\right) = \begin{cases} 1, & Y_n \leq k \\ 0, & \text{de otro modo} \end{cases}$$

$$\alpha = P_{\theta=2}\left(Y_n \leq k\right) = 0.05 \rightarrow 0.05 = \int_0^k f\left(Y_n\right) dY_n$$

teniendo en cuenta que $n = 10$:

$$0.05 = \int_0^k f\left(Y_n\right) dY_n = \int_0^k \frac{10Y_n^9}{2^{10}} dY_n = \left(\frac{k}{2}\right)^{10} = 0.05 \rightarrow k = 1.482269$$

$$\varphi\left(\underset{\sim}{x}\right) = \begin{cases} 1, & Y_n \leq 1.482269 \\ 0, & \text{de otro modo} \end{cases}$$

Como en la muestra $Y_n = 1.99 \rightarrow$ No se rechaza la H_0

Otra forma de concluir:

$$\text{pvalor} = P_{\theta=2}\left(Y_n \le 1.99\right) = \int_0^{1.99} \frac{10Y_n^9}{2^n}\, dY_n = 0.951110 > \alpha = 0.05 \rightarrow \text{No se rechaza la } H_0.$$

c. Si se toma una muestra al azar de 10 clases y se obtiene: 0.81, 0.32, 1.25, 1.92, 1.12, 0.45, 1.99, 0.00, 0.00, 1.25. Pruebe $H_0 : \theta = 2 \quad vs \quad H_1 : \theta \neq 2$. Use $\alpha = 0.10$.

En este caso la hipótesis planteada puede ser $H_0^1 : \theta \ge 2$ o $H_0^2 : \theta \le 2$. Considerando los resultados de las subpreguntas a y b se obtendrá la siguiente prueba más poderosa:

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 & , \text{ si } Y_n \le k_1 \text{ o } Y_n \ge k_2 \\ 0 & , \text{ o.m} \end{cases}$$

$$Y_n = 1.99 \;\; , \; f_{Y_n}\left(y\right) = \frac{10y^9}{\theta^{10}} I_{(0,\theta)}\left(y\right)$$

$$0.10 = P_{\theta=2}\left(Y_n \le k_1 \text{ o } Y_n \ge k_2\right) = P_{\theta=2}\left(Y_n \le k_1\right) + P_{\theta=2}\left(Y_n \ge k_2\right) =$$

$$= \frac{0.10}{2} + \frac{0.10}{2} = 0.05 + 0.05 \rightarrow \quad k_1 = 1.482269 \quad \text{y} \quad k_2 = 1.989768$$

$$\varphi^*\left(x_1,\ldots,x_n\right) = \begin{cases} 1 & , \text{ si } Y_n \le 1.482269 \text{ o } Y_n \ge 1.989768 \\ 0 & , \text{ o.m} \end{cases}.$$

Como $Y_n = 1.99 \rightarrow$ Se rechaza la H_0

Ejemplo 22: Sea $X_1,\ldots,X_n$ una m.a. de $f\left(x;\theta\right) = e^{-\frac{1}{8}\left(x-\frac{1}{\theta}\right)} I_{\left(\frac{1}{\theta},\infty\right)}\left(x\right)$; $\theta \in \mathbb{R}$. Por lo tanto, la distribución es exponencial con media 8 pero truncada en $\frac{1}{\theta}$.

a. Halle una prueba UMP de tamaño α para probar $H_0 : \frac{1}{\theta} \le \frac{1}{\theta_0}$ vs. $H_1 : \frac{1}{\theta} > \frac{1}{\theta_0}$.

$f\left(x;\theta\right) = e^{-\frac{1}{8}\left(x-\frac{1}{\theta}\right)} \notin$ a la familia exponencial (usar la propiedad de RVM).

Para $\frac{1}{\theta_1} < \frac{1}{\theta_2}$

$$\Psi=\frac{L(\theta_1)}{L(\theta_2)}=\frac{\prod_{i=1}^{n} e^{-\frac{1}{8}\left(x-\frac{1}{\theta_1}\right)}}{\prod_{i=1}^{n} e^{-\frac{1}{8}\left(x-\frac{1}{\theta_2}\right)}}=\frac{e^{\frac{1}{8}\left(\frac{n}{\theta_1}-\sum_{i=1}^{n}x_i\right)} I_{\left(\frac{1}{\theta_1},\infty\right)}(Y_1)}{e^{\frac{1}{8}\left(\frac{n}{\theta_2}-\sum_{i=1}^{n}x_i\right)} I_{\left(\frac{1}{\theta_2},\infty\right)}(Y_1)}=e^{\frac{n}{8}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)}\frac{I_{\left(\frac{1}{\theta_1},\infty\right)}(Y_1)}{I_{\left(\frac{1}{\theta_2},\infty\right)}(Y_1)}=\Psi(Y_1)$$

$$\Psi(Y_1)=\begin{cases}\infty & , \quad \frac{1}{\theta_1}<Y_1<\frac{1}{\theta_2}\\ e^{\frac{n}{8}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)} & , \quad Y_1>\frac{1}{\theta_2}\end{cases}$$

$\Psi(Y_1)$ es decreciente en Y_1

$$\varphi(\underset{\sim}{x})=\begin{cases}1, & Y_1\geq k\\ 0, & \text{de otro modo}\end{cases}$$

$$\alpha=P_{\theta_0}(Y_1\geq k)=\int_k^{\infty} f(Y_1)\,dY_1=\int_k^{\infty}\frac{n}{8}e^{-\frac{n}{8}\left(Y_1-\frac{1}{\theta_0}\right)}dY_1=-e^{-\frac{n}{8}\left(Y_1-\frac{1}{\theta_0}\right)}\Big|_k^{\infty}=e^{-\frac{n}{8}\left(k-\frac{1}{\theta_0}\right)}=\alpha\rightarrow$$

$$\rightarrow k=\frac{1}{\theta_0}-\frac{8\ln(\alpha)}{n}$$

$$\varphi(\underset{\sim}{x})=\begin{cases}1, & Y_1\geq k=\frac{1}{\theta_0}-\frac{8\ln(\alpha)}{n}\\ 0, & \text{de otro modo}\end{cases}$$

Ver el Script R. En ese Script se considera $H_0:\theta=\frac{1}{10}$ vs $H_1:\theta>\frac{1}{10}$

NOTA: Las pruebas de hipótesis con densidades que pertenecen a la familia exponencial se pueden desarrollar con la RVM, pero las pruebas de hipótesis con densidades que no pertenecen a la familia exponencial no se pueden trabajar como si fueran de la familia exponencial.

Ejemplo 23: En cierta ciudad la proporción de microbuseros que tienen por lo menos cuarenta infracciones de tránsito es p. En una muestra aleatoria de 18 microbuseros 13 tienen por lo menos cuarenta infracciones. Proceda con el método de la RVM.

a. Con una prueba no randomizada pruebe $H_0:p\geq 0.84$ vs $H_1:p<0.84$. Use $\alpha=0.10$.

$$X_1,\ldots,X_{18}\rightarrow X_i\sim\text{Binomial}(1,p)\rightarrow$$

$$f(x)=p^x(1-p)^{1-x}=(1-p)\left(\frac{p}{1-p}\right)^x=(1-p)\times 1\times\exp\left\{\left[\ln\left(\frac{p}{1-p}\right)\right]\times x\right\}$$

$f(x) = p^x (1-p)^{1-x} \in$ a la familia exponencial.

Para $p_1 < p_2$

$$\Psi = \frac{L(p_1)}{L(p_2)} = \frac{\prod_{i=1}^{18} p_1^{x_i}(1-p_1)^{1-x_i}}{\prod_{i=1}^{18} p_2^{x_i}(1-p_2)^{1-x_i}} = \left(\frac{p_1}{p_2}\right)^{\sum_{i=1}^{18} x_i} \left(\frac{1-p_1}{1-p_2}\right)^{18-\sum_{i=1}^{18} x_i} = \Psi\left(\sum_{i=1}^{18} x_i\right)$$

Como $p_1 < p_2 \rightarrow$ considerando por ejemplo $0.5 < 0.9$ y $t = \sum_{i=1}^{18} x_i$, se tendría:

$\Psi = \left(\frac{5}{9}\right)^t \left(\frac{5}{1}\right)^{18-t}$ la cual es decreciente en $t = \sum_{i=1}^{18} x_i$ entonces:

$$\varphi(\underset{\sim}{x}) = \begin{cases} 1, & \sum_{i=1}^{18} X_i \leq k \\ 0, & \text{de otro modo} \end{cases}$$

$\sum_{i=1}^{18} x_i = 13$, $\sum_{i=1}^{18} x_i \sim \text{Binomial}(18, p)$

$$\text{pvalor} = P_{p=0.84}\left(\sum_{i=1}^{18} x_i \leq 13\right) = \sum_{y=1}^{13} \binom{18}{y} 0.84^y \times 0.16^{18-y} = 0.1482$$

$\text{pvalor} = 0.1482 > \alpha = 0.10 \rightarrow$ No se rechaza la H_0.

Prueba de Razón de Verosimilitud General (PRVG)

Sea $X_1, \ldots, X_n$ una m.a. de $f(x;\theta)$; $\theta \in \Theta \subset \Re$. Sean las hipótesis: Ho: $\theta \in \Theta_0$ vs. H1: $\theta \in \Theta_1$; con $\Theta_0 \cap \Theta_1 = \phi$ y $\Theta_0 \cup \Theta_1 = \Theta$.

La R. V.G., se define por:

$$\Lambda(\underline{x}) = \frac{\underset{\theta\in\Theta_0}{Sup}\, L(\underline{x},\theta)}{\underset{\theta\in\Theta}{Sup}\, L(\underline{x},\theta)} = \frac{\max L(\Theta_0)}{\max L(\Theta)}$$

Principio de la PRVG

Se dice que φ es una PRVG si: $\varphi(\underline{x}) = \begin{cases} 1 & ,si\, \Lambda(\underline{x}) \leq k \\ 0 & ,si\, \Lambda(\underline{x}) > k \end{cases}$

OBSERVACIONES

1. $\Lambda(\underline{x})$ es una estadística que no depende del parámetro desconocido
2. El parámetro θ puede ser un vector.
3. $0 \leq \Lambda(\underline{x}) \leq 1$

4. Para usar esta prueba se debe obtener la distribución exacta o aproximada de $\Lambda(\underline{x})$ pero si no es así $\Lambda(\underline{x})$ debe estar en función de otro estadístico cuya distribución se conozca.

Ejemplo 26: Si $X_1,\ldots,X_n$ es una muestra aleatoria de una distribución $\text{Binomial}(1,\pi)$. Para contrastar $H_0:\pi\le\pi_0$ vs $H_1:\pi>\pi_0$, mediante el test de RVG se tiene:

$$\Lambda(\underline{x})=\frac{\sup\limits_{\pi\le\pi_0}\pi^{\sum_{i=1}^{n}x_i}(1-\pi)^{n-\sum_{i=1}^{n}x_i}}{\sup\limits_{\pi\in[0,1]}\pi^{\sum_{i=1}^{n}x_i}(1-\pi)^{n-\sum_{i=1}^{n}x_i}}$$

El supremo del denominador se obtiene en $\pi=\bar{X}$ y para el numerador si $\bar{x}\le\pi_0$ es $\pi_0=\pi_0$, también al ser el numerador una función creciente en π, el supremo se obtiene en π_0, por lo tanto:

$$\max L(\Theta_0)=\prod_{i=1}^{n}\pi_0^{x_i}(1-\pi_0)^{1-x_i}=\pi_0^{\sum_{i=1}^{n}x_i}(1-\pi_0)^{n-\sum_{i=1}^{n}x_i}$$

$$\max L(\Theta)=\prod_{i=1}^{n}\bar{X}^{x_i}(1-\bar{X})^{1-x_i}=\bar{X}^{\sum_{i=1}^{n}x_i}(1-\bar{X})^{n-\sum_{i=1}^{n}x_i}$$

$$\Lambda(\underline{x})=\frac{\pi_0^{\sum_{i=1}^{n}x_i}(1-\pi_0)^{n-\sum_{i=1}^{n}x_i}}{\bar{X}^{\sum_{i=1}^{n}x_i}(1-\bar{X})^{n-\sum_{i=1}^{n}x_i}}=\left(\frac{\pi_0}{\bar{X}}\right)^{\sum_{i=1}^{n}x_i}\left(\frac{1-\pi_0}{1-\bar{X}}\right)^{n-\sum_{i=1}^{n}x_i}$$

$$\Lambda(\underline{x})=\begin{cases}1 & \text{si } \bar{X}>\pi_0\\ \left(\dfrac{\pi_0}{\bar{X}}\right)^{\sum_{i=1}^{n}x_i}\left(\dfrac{1-\pi_0}{1-\bar{X}}\right)^{n-\sum_{i=1}^{n}x_i} & \text{si } \bar{X}\le\pi_0\end{cases}$$

Como la función $\Lambda(\underline{x})$ es decreciente en $\bar{x}$, la región crítica es $\bar{x}>c$. Con un nivel de significación conocido, para hallar la constante hay que utilizar el hecho de que $\sum_{i=1}^{n}X_i\sim\text{Binomial}(n,\pi)$

Ejemplo 27: Determine la región crítica de la prueba de razón de verosimilitud para verificar $H_0:\mu=\mu_0$ vs $H_1:\mu\ne\mu_0$ considerando una muestra aleatoria de tamaño n de una distribución normal con σ^2 conocida.

Según la hipótesis planteada: En Θ_0 se tiene que el EMV restringido de μ es μ_0, entonces:

$$\max L(\Theta_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu_0)^2\right] = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2\right]$$

Sin restricción: En Θ se tiene:
El estimador máximo verosímil de μ es $\overline{X}$, entonces:

$$\max L(\Theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x_i - \overline{x}\right)^2\right] = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]$$

Por lo tanto:

$$\Lambda(\underline{x}) = \frac{\max L(\Theta_0)}{\max L(\Theta)} = \exp\left[-\frac{n}{2\sigma^2}\left(\overline{x} - \mu_0\right)^2\right] \le k$$

Tomando logaritmos y despejando:

$$\frac{\left(\overline{x} - \mu_0\right)^2}{\frac{\sigma^2}{n}} \ge \underbrace{(-2)\left(\underset{\text{Negativo}}{\ln k}\right)}_{\text{Positivo}} = k_1 \text{ , porque } 0 < k < 1$$

Primera opción
Lo anterior se puede poner así:

$$\left|\frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right| = |Z| \ge Z_{\left(1-\frac{\alpha}{2}\right)} \text{ que viene a ser la región crítica.}$$

Segunda opción

$$\left(\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}\right)^2 \ge \chi^2(\alpha, 1) \text{ es otra forma de representar la region crítica.}$$

NOTA: $\frac{n}{\sigma^2}\left(\overline{x} - \mu_0\right)^2 = \left(\frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)^2 = Z^2 \sim \chi^2(1)$

Ejemplo: Determine la región crítica de la prueba de razón de verosimilitud para verificar $H_0: \mu \le \mu_0$ vs $H_1: \mu > \mu_0$ considerando una muestra aleatoria de tamaño n de una distribución normal con σ^2 desconocida.

El EMV sin restricción es $\hat{\mu} = \overline{X}$.

El EMV restringido es $\hat{\mu}_0 = \begin{cases} \overline{X} & \text{si } \overline{X} > \mu_0 \\ \mu_0 & \text{si } \overline{X} \le \mu_0 \end{cases}$

Entonces

$$\Lambda\left(\underline{x}\right)=\frac{L\left(\hat{\mu}_0\right)}{L\left(\hat{\mu}\right)}=\begin{cases}1 & \text{si} \quad \bar{X}>\mu_0\\ \dfrac{L\left(\mu_0\right)}{L\left(\bar{X}\right)} & \text{si} \quad \bar{X}\le\mu_0\end{cases}$$

$$\Lambda\left(\underline{x}\right)=\begin{cases}1 & \text{si} \quad \bar{X}>\mu_0\\ \exp\left[-\dfrac{1}{2\sigma^2}\left(-2\mu_0 n\bar{X}+n\mu_0^2+n\bar{X}^2\right)\right] & \text{si} \quad \bar{X}\le\mu_0\end{cases}$$

Ejemplo 28: Determine la región crítica de la prueba de razón de verosimilitud para verificar $H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$ considerando una muestra aleatoria de tamaño n de una distribución normal con σ^2 desconocida.

Según la hipótesis planteada: En Θ_0 se tiene que los EMVs restringidos de μ y σ^2 son μ_0 y $\hat{\sigma}_0^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$ respectivamente:

$$\max L\left(\Theta_0\right)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sqrt{\hat{\sigma}_0^2}}\exp\left[-\frac{1}{2\hat{\sigma}_0^2}\left(x_i-\mu_0\right)^2\right]=\left(\frac{1}{\sqrt{2\pi}\sqrt{\hat{\sigma}_0^2}}\right)^n\exp\left[-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{n}\left(x_i-\mu_0\right)^2\right]$$

Sin restricción: En Θ se tiene que los EMVs de μ y σ^2 son $\bar{X}$ y $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ respectivamente, entonces:

$$\max L\left(\Theta\right)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sqrt{\hat{\sigma}^2}}\exp\left[-\frac{1}{2\hat{\sigma}^2}\left(x_i-\bar{x}\right)^2\right]=\left(\frac{1}{\sqrt{2\pi}\sqrt{\hat{\sigma}^2}}\right)^n\exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right]$$

Por lo tanto:

$$\Lambda\left(\underline{x}\right)=\frac{\max L\left(\Theta_0\right)}{\max L\left(\Theta\right)}=\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}}\le k$$

Pero:

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\mu_0)^2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sum_{i=1}^{n}\left[(X_i-\bar{X})+(\bar{X}-\mu_0)\right]^2}=\left[1+\frac{n\left(\bar{X}-\mu_0\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]^{-1}=$$

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}=\left\{1+\frac{1}{n-1}\left[\frac{n\left(\bar{X}-\mu_0\right)^2}{S^2}\right]\right\}^{-1}=\left[1+\frac{t_{(n-1)}^2}{n-1}\right]^{-1}$$

$$\Lambda\left(\underline{x}\right)=\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}}\le k\rightarrow\left[1+\frac{t_{(n-1)}^2}{n-1}\right]^{-\frac{n}{2}}\le k$$

Tomando logaritmos y despejando:

$$-\frac{n}{2}\ln\left[1+\frac{t_{(n-1)}^2}{n-1}\right]\le\left(\underset{\text{Negativo}}{\ln k}\right)\text{, porque } 0<k<1$$

$$\ln\left[1+\frac{t_{(n-1)}^2}{n-1}\right]\ge\underbrace{\left(-\frac{2}{n}\right)\left(\underset{\text{Negativo}}{\ln k}\right)}_{\text{Positivo}}=k_1$$

$$\left[1+\frac{t_{(n-1)}^2}{n-1}\right]\ge e^{k_1}=k_2\rightarrow t_{(n-1)}^2\ge c$$

Como $t_{(n-1)}\sim t\left(n-1\right)$ y $t_{(n-1)}^2\sim F\left(1,n-1\right)$ se puede hallar la región crítica.

Teorema de Wilks (Distribución asintótica de la razón de verosimilitud)
Este Teorema requiere que se cumplan las condiciones de regularidad.

Para n grande, $-2\ln\left(\Lambda\left(\underline{x}\right)\right)$ converge en distribución a $\chi_{(r)}^2$; donde:

r= número de parámetros desconocidos bajo Θ menos el Nº de parámetros desconocidos bajo Θ_0, y la prueba es unilateral a la derecha.

NOTA: Como la PRVG no siempre genera una estadística con distribución conocida entonces el teorema anterior ayuda con n grande (con el TLC $n\ge 30$, pero con el Teorema de Wilks se sugiere tamaños de muestras más grandes).

Ejemplo 29
En cierto distrito la proporción de jóvenes menores de 18 años que consumen drogas es p. Si se seleccionaron aleatoriamente a 150 de estos jóvenes y se encontró que 100 consumen drogas pruebe $H_0: p=0.72$ vs $H_1: p\ne 0.72$. Use $\alpha=0.05$.

$$X_i\sim\text{Binomial}\left(1,p\right),\ i=1,2,\ldots,150\rightarrow f\left(x_i\right)=p^{x_i}\left(1-p\right)^{1-x_i}I_{\{0,1\}}\left(x_i\right)$$

$$n=150,\quad\sum_{i=1}^{150}X_i=100$$

$$L=\prod_{i=1}^{150} f\left(x_i\right)=\prod_{i=1}^{150} p^{x_i}\left(1-p\right)^{1-x_i} = p^{\sum_{i=1}^{150} x_i}\left(1-p\right)^{150-\sum_{i=1}^{150} x_i} \rightarrow \text{El EMVde } p \text{ es: } \frac{\sum_{i=1}^{150} x_i}{150}=\frac{100}{150}$$

$$\max\left[L\left(\Theta_0\right)\right]=0.72^{100}\times 0.28^{150-100} = 0.72^{100}\times 0.28^{50}$$

$$\max\left[L\left(\Theta\right)\right]=\left(\frac{100}{150}\right)^{100}\times\left(1-\frac{100}{150}\right)^{150-100}=\left(\frac{100}{150}\right)^{100}\times\left(\frac{50}{150}\right)^{50}$$

$$\Lambda=\frac{\max\left[L\left(\Theta_0\right)\right]}{\max\left[L\left(\Theta\right)\right]}=\frac{150^{150}}{50^{50}\times 100^{100}}0.72^{100}\times 0.28^{50}$$

$$-2\ln\left(\Lambda\right)\sim\chi^2_{(1-0)}\sim\chi^2_{(1)}\rightarrow -2\ln\left(\Lambda\right)=2.043\rightarrow \text{pvalor}=P\left[\chi^2_{(1)}>2.043\right]=0.1529>\alpha=0.05$$

$\rightarrow$ No se rechaza la H_0

Ejemplo 30. Multinomial con una muestra

El Psicólogo Jorge Luis Borges, utilizando un método estadístico multivariado, ha clasificado a los pacientes de acuerdo con su edad, peso, sexo, estado civil, distracciones, estatura, ingreso mensual familiar, y otros, en cinco grupos. Él está interesado en estudiar la proporción de alcohólicos peruanos que pertenecen a cada grupo, y para esto selecciona al azar con reemplazo y considerando el orden de extracción una muestra de 100 pacientes y encuentra que al primer grupo pertenecen 18 pacientes, al segundo 22, al tercero 17, al cuarto 23 y al quinto 20.

a. Verifique la hipótesis nula $H_0: p_i=\frac{1}{5}$, $i=1,2,\ldots,5$. Utilice $\alpha=0.05$

Solución

$$L=\frac{n!}{x_1!\ldots x_k!x_{k+1}!}p_1^{x_1}\ldots p_k^{x_k}\left(1-p_1-\cdots-p_k\right)^{x_{k+1}}$$

$$\max\left[L\left(\Theta_0\right)\right]=\frac{n!}{x_1!\ldots x_k!x_{k+1}!}p_{1,0}^{x_1}\ldots p_{k,0}^{x_k}p_{k+1,0}^{x_{k+1}} \quad \text{donde } p_{\text{i},0}=\frac{1}{5}, i=1,2,\ldots,5$$

El EMV de p_{i} es $\hat{p}_i=\frac{x_i}{n}, i=1,2,\ldots,5\rightarrow$

$$\max\left[L\left(\Theta\right)\right]=\frac{n!}{x_1!\ldots x_k!x_{k+1}!}\left(\frac{x_1}{n}\right)^{x_1}\ldots\left(\frac{x_{k+1}}{n}\right)^{x_{k+1}}=n^{-n}\frac{n!}{x_1!\ldots x_k!x_{k+1}!}\left(x_1\right)^{x_1}\ldots\left(x_{k+1}\right)^{x_{k+1}}$$

$$\rightarrow\Lambda=\frac{\max\left[L\left(\Theta_0\right)\right]}{\max\left[L\left(\Theta\right)\right]}=n^n\left(\frac{p_{1,0}}{x_1}\right)^{x_1}\ldots\left(\frac{p_{\text{k}+1,0}}{x_{k+1}}\right)^{x_{k+1}}$$

$$\Lambda = 100^{100}\left(\frac{1}{5\times18}\right)^{18}\left(\frac{1}{5\times22}\right)^{22}\left(\frac{1}{5\times17}\right)^{17}\left(\frac{1}{5\times23}\right)^{23}\left(\frac{1}{5\times20}\right)^{20}$$

$-2\ln(\Lambda)\sim\chi^2(4-0)\sim\chi^2(4)$,se consideran 4 parámetros desconocidos en todo el espacio paramétrico porque $p_{k+1}=p_5=1-p_1-\cdots-p_4$ ya se conoce, dado que la suma de las probabilidades es uno.

$-2\ln(\Lambda)=1.30408\rightarrow \text{pvalor}=P\left[\chi^2(4)>1.30408\right]=0.8606829.$

$\therefore$ No se rechaza la hipótesis planteada.

b. Suponga que se cree que el porcentaje de alcohólicos peruanos que pertenece al grupo 1 y al grupo 2 son aproximadamente iguales. En este caso el sistema de hipótesis es el siguiente: $H_0: p_1=p_2$ vs $H_1: p_1\neq p_2$.

En este caso el numerador de Λ bajo $H_0: p_1=p_2$ considerará a la función de verosimilitud como: $\max\left[L(\Theta_0)\right]=\dfrac{n!}{x_1!\ldots x_4!x_5!}p_{1,0}^{x_1}p_{2,0}^{x_2}p_{3,0}^{x_3}p_{4,0}^{x_4}p_{5,0}^{x_5}=\dfrac{n!}{x_1!\ldots x_4!x_5!}p_{1,0}^{x_1+x_2}p_{3,0}^{x_3}p_{4,0}^{x_4}p_{5,0}^{x_5}$

La restricción $p_1+p_2+p_3+p_4+p_5=1$ ahora toma la forma $2p_1+p_3+p_4+p_5=1$

Como se tiene un problema de maximización con restricciones se recurre a la técnica del multiplicador de Lagrange:

$$\max\left[L(\Theta_0)\right]=\frac{n!}{x_1!\ldots x_4!x_5!}p_{1,0}^{x_1}p_{2,0}^{x_2}p_{3,0}^{x_3}p_{4,0}^{x_4}p_{5,0}^{x_5}=\frac{n!}{x_1!\ldots x_4!x_5!}p_{1,0}^{x_1+x_2}p_{3,0}^{x_3}p_{4,0}^{x_4}p_{5,0}^{x_5}$$

En esta parte Λ hace las veces de la función de Lagrange.

$$\Lambda=\ln(n!)-\ln(x_1!)-\cdots-\ln(x_5!)+(x_1+x_2)\ln\left(p_{1,0}\right)+x_3\ln\left(p_{3,0}\right)+x_4\ln\left(p_{4,0}\right)+x_5\ln\left(p_{5,0}\right)+$$
$$+\lambda\left(2p_{1,0}+p_{3,0}\cdots+p_{5,0}-1\right)$$

$$\frac{d\Lambda}{dp_{1,0}}=\frac{x_1+x_2}{p_{1,0}}+2\lambda=0\rightarrow p_{1,0}=-\frac{x_1+x_2}{2\lambda}$$

$$\frac{d\Lambda}{dp_{i,0}}=\frac{x_i}{p_{i,0}}+\lambda=0\rightarrow p_{i,0}=-\frac{x_i}{\lambda}\ ,\ i=3,4,5$$

$$\frac{d\Lambda}{d\lambda}=2p_{1,0}+p_{3,0}\cdots+p_{5,0}-1=0\rightarrow 2p_{1,0}+p_{3,0}\cdots+p_{5,0}=1$$

$$\rightarrow 2p_{1,0}+p_{3,0}\cdots+p_{5,0}=1\rightarrow 2\left(-\frac{x_1+x_2}{2\lambda}\right)+\left(-\sum_{i=3}^{5}\frac{x_i}{\lambda}\right)=1\rightarrow\lambda=-(x_1+x_2)-\sum_{i=3}^{5}x_i\rightarrow\lambda=-n$$

$$p_{1,0}=-\frac{x_1+x_2}{(-2n)}=\frac{x_1+x_2}{2(x_1+x_2+x_3+x_4+x_5)}=\frac{x_1+x_2}{2n}$$

$$\rightarrow p_{i0}=-\frac{x_i}{\lambda}=-\frac{x_i}{-(x_1+x_2+x_3+x_4+x_5)}=-\frac{x_i}{(-n)}=\frac{x_i}{n}=\overline{X}_i\ ,\ i=3,4,5$$

De la pregunta anterior se tiene:

$$\max\left[L(\Theta)\right]=\frac{n!}{x_1!\ldots x_4!x_5!}\left(\frac{x_1}{n}\right)^{x_1}\ldots\left(\frac{x_5}{n}\right)^{x_5}=n^{-n}\frac{n!}{x_1!\ldots x_4!x_5!}(x_1)^{x_1}\ldots(x_5)^{x_5}$$

En esta parte Λ es la razón de verosimilitud general.

$$\rightarrow\Lambda=\frac{\max\left[L(\Theta_0)\right]}{\max\left[L(\Theta)\right]}=\left(\frac{x_1+x_2}{2n}\right)^{x_1+x_2}\left(\frac{1}{\frac{x_1}{n}}\right)^{x_1}\left(\frac{1}{\frac{x_2}{n}}\right)^{x_2}\left(\frac{\frac{x_3}{n}}{\frac{x_3}{n}}\right)^{x_3}\left(\frac{\frac{x_4}{n}}{\frac{x_4}{n}}\right)^{x_4}\left(\frac{\frac{x_5}{n}}{\frac{x_5}{n}}\right)^{x_5}=$$

$$\Lambda=\left(\frac{x_1+x_2}{2n}\right)^{x_1+x_2}\left(\frac{1}{\frac{x_1}{n}}\right)^{x_1}\left(\frac{1}{\frac{x_2}{n}}\right)^{x_2}$$

$$\Lambda=\left(\frac{x_1+x_2}{2n}\right)^{x_1+x_2}\left(\frac{1}{\frac{x_1}{n}}\right)^{x_1}\left(\frac{1}{\frac{x_2}{n}}\right)^{x_2}$$

$$-2\ln(\Lambda)=-2\times\left\{(18+22)\ln\left[(18+22)/(2\times100)\right]+18\ln(100/18)+22\ln(100/22)\right\}=$$

$$-2\ln(\Lambda)=0.4006693477$$

En este problema el número de parámetros desconocidos bajo el espacio paramétrico Θ es 5-1=4 y el número de parámetros desconocidos bajo la hipótesis nula es 5-2=3 porque una vez que se conocen p_3, p_4 y $p_5 \rightarrow 2p_1+p_3+p_4+p_5=1 \rightarrow p_1=p_2=\frac{1-(p_3+p_4+p_5)}{2}$ Por lo tanto:

$$-2\ln(\Lambda)\sim\chi^2(4-3)\sim\chi^2(1)\rightarrow \text{pvalor}=P\left[\chi^2(1)>0.4006693477\right]=0.5267438$$

En conclusion no se rechaza la hipótesis nula.

c. Verifique la hipótesis nula $H_0: p_1=0.19, p_2=0.23, p_3=0.19, p_4=0.32, p_5=0.07$. Use $\alpha=0.05$. (Queda como ejercicio)

Ejemplo 31. Multinomial con dos muestras

Suponga que se desea conocer la opinión que tienen los consumidores respecto a un nuevo producto que una empresa lanzó hace tres meses al mercado. Se encuestó a un grupo de personas preguntándoles por su opinión respecto al producto, y la respuesta puede ser Bueno, Regular o Malo. Se entrevistó a 800 personas, de las cuales 450 son hombres y 350 mujeres. Interesaba saber si hay diferencia significativa de los sexos respecto a la percepción que tienen del producto. Si se concluye que la percepción es baja en alguno de los sexos, se plantearán estrategias para atraerlos y así la empresa tendrá mayores ganancias.

	Bueno	Regular	Malo	Total
Hombre	220	180	50	450
Mujer	190	130	30	350

$H_0:(p_{11},\ldots,p_{k1})=(p_{12},\ldots,p_{k2})$ vs $H_1:(p_{11},\ldots,p_{k1})\neq(p_{12},\ldots,p_{k2})$

H_0: No hay diferencia entre los sexos respecto
a la percepción del nuevo producto

Sean dos vectores aleatorios que tienen la misma dimensión $(X_1,\ldots,X_k)$ y $(Y_1,\ldots,Y_k)$ que se distribuyen independientemente según las distribuciones Multinomial$(n_X, p_{11},\ldots,p_{k1})$ y Multinomial$(n_Y, p_{12},\ldots,p_{k2})$. La función de verosimilitud conjunta es:

$$L=\frac{n_X!}{x_1!\ldots x_k!}p_{11}^{x_1}\cdots p_{k1}^{x_k}\frac{n_Y!}{y_1!\ldots y_k!}p_{12}^{y_1}\cdots p_{k2}^{y_k}$$

Se utiliza el multiplicador de Lagrange para hallar los estimadores máximo verosímiles. Esto se hace maximizando $\ln L$.

$$\ln L=x_1\ln(p_{11})+\ldots+x_k\ln(p_{k1})+y_1\ln(p_{12})+\ldots+y_k\ln(p_{k2})+\ln\left(\frac{n_X!}{x_1!\ldots x_k!}\right)+\ln\left(\frac{n_Y!}{y_1!\ldots y_k!}\right)$$

Sujeto a las restricciones: $p_{11}+\ldots+p_{k1}=1$ y $p_{12}+\ldots+p_{k2}=1$, de tal manera que los estimadores máximo verosímiles son:

$$p_{i1}=\frac{X_i}{n_X}=\overline{X}_i \quad \text{y} \quad p_{i2}=\frac{Y_i}{n_Y}=\overline{Y}_i, \quad i=1,2,\ldots,k$$

Utilizando la prueba de razón de verosimilitud general para el sistema de hipótesis anterior y considerando que $\Theta_0\cup\Theta_1=\Theta$ es el espacio paramétrico de los vectores $(p_{11},\ldots,p_{k1})$ y $(p_{12},\ldots,p_{k2})$ entonces bajo todo el espacio paramétrico la función de verosimilitud $L=\frac{n_X!}{x_1!\ldots x_k!}p_{11}^{x_1}\cdots p_{k1}^{x_k}\frac{n_Y!}{y_1!\ldots y_k!}p_{12}^{y_1}\cdots p_{k2}^{y_k}$ será evaluada con los estimadores $p_{i1}=\frac{X_i}{n_X}=\overline{X}_i$ y $p_{i2}=\frac{Y_i}{n_Y}=\overline{Y}_i$, $i=1,2,\ldots,k$.

De otro lado, bajo H_0 se tiene que $p_{i1}=p_{i2}=p_i$ para $i=1,2,\ldots,k$. Por lo tanto, la función de verosimilitud se convierte en:

$$L=\frac{n_X!}{x_1!\ldots x_k!}p_1^{x_1}\cdots p_k^{x_k}\frac{n_Y!}{y_1!\ldots y_k!}p_1^{y_1}\cdots p_k^{y_k}=p_1^{x_1+y_1}\cdots p_k^{x_k+y_k}\frac{n_X!}{x_1!\ldots x_k!}\frac{n_Y!}{y_1!\ldots y_k!}$$

Maximizando $\ln(L)$ considerando la restricción $p_1+\ldots+p_k=1$ se tiene el siguiente estimador de p_i bajo H_0: $p_i=\frac{x_i+y_i}{n_X+n_Y}$ para $i=1,2,\ldots,k$. Por lo tanto:

$$-2\ln(\Lambda)=$$

$$=-2\times\left[(x_1+y_1)\ln\frac{x_1+y_1}{n_X+n_Y}+(x_k+y_k)\ln\frac{x_k+y_k}{n_X+n_Y}-x_1\ln\overline{x}_1-\ldots-x_k\ln\overline{x}_k-y_1\ln\overline{y}_1-\ldots-y_k\ln\overline{y}_k\right]$$

Bajo todo el espacio paramétrico la restricción es $(p_{11},\ldots,p_{k1})$ y $(p_{12},\ldots,p_{k2})$ entonces hay $2(k-1)$ parámetros desconocidos (grados de libertad) mientras que bajo la hipótesis nula los parámetros de las dos muestras son iguales, entonces el número de parámetros desconocidos (grados de libertad) es $(k-1)$, en consecuencia $-2\ln(\Lambda) \sim \text{asym } \chi^2(k-1)$.

En el problema:

$-2\ln(\Lambda) \sim \text{asym } \chi^2(k-1=2) \rightarrow -2\ln(\Lambda) = 2.817609$

$\text{pvalor} = P\left[\chi^2(2) > 2.817609\right] = 0.2444353$

No hay diferencia significativa entre los hombres y las mujeres respecto a la percepción del nuevo producto.

NOTA: Lo anterior se puede aplicar con m muestras (m>2). En este caso $-2\ln(\Lambda) \sim \text{asym } \chi^2(k-1)\times(m-1)$

Ejercicio

Se estudia el perfil de los diferentes estratos respecto a la frecuencia que practican deportes. Se hizo una encuesta a persona de tres estratos respecto a la frecuencia de práctica deportiva.

	Estrato Bajo	Estrato Medio	Estrato Alto
Más de 3 veces semanales	92	72	84
Entre 1 y 3 veces a la semana	100	89	111
De vez en cuando	53	94	17
No practica deportes	28	52	10
Total de Encuestados	273	307	222

H_0 : No hay diferencia entre los estratos respecto a la frecuencia de practicas deportivas

Ver el Script R.

Ejemplo 32

Se tienen tres marcas de TV (I, II y III). Se quiere estudiar la fracción de TV con defectos que son enviados a repararse durante el periodo de garantía. 10 de 85 TV de la marca I, 4 de 95 TV de la marca II y 4 de 98 TV de la marca III tuvieron defectos y se enviaron a reparación durante el periodo de garantía. Se puede considerar que el número de TV con defectos que fueron enviados a repararse durante el periodo de garantía de las tres marcas son v.as Binomiales independientes.

a. ¿Hay evidencia estadística que nos permita afirmar que la fracción de TV con defectos que fueron enviados a repararse durante el periodo de garantía es la misma para las tres marcas? Use $\alpha = 0.05$

$X_i \sim \text{Binomial}(1,\pi_1)$, $i = 1,2,\ldots,85$, $Y_i \sim \text{Binomial}(1,\pi_2)$, $i = 1,2,\ldots,95$,
$Z_i \sim \text{Binomial}(1,\pi_3)$, $i = 1,2,\ldots,98$

$$n_1 = 85, \quad \sum_{i=1}^{85} X_i = 10; \; n_2 = 95, \quad \sum_{i=1}^{95} Y_i = 4; \; n_3 = 98, \quad \sum_{i=1}^{98} Z_i = 4$$

$$L = \prod_{i=1}^{n_1} f(x_i, \pi_1) \prod_{i=1}^{n_2} f(y_i, \pi_2) \prod_{i=1}^{n_3} f(z_i, \pi_3) =$$

$$= \prod_{i=1}^{n_1} \pi_1^{x_i} (1-\pi_1)^{1-x_i} \prod_{i=1}^{n_2} \pi_2^{y_i} (1-\pi_2)^{1-y_i} \prod_{i=1}^{n_3} \pi_3^{z_i} (1-\pi_3)^{1-z_i}$$

$$= \left[\pi_1^{\sum_{i=1}^{n_1} x_i} (1-\pi_1)^{n_1 - \sum_{i=1}^{n_1} x_i} \right] \left[\pi_2^{\sum_{i=1}^{n_2} y_i} (1-\pi_2)^{n_2 - \sum_{i=1}^{n_2} y_i} \right] \left[\pi_3^{\sum_{i=1}^{n_3} z_i} (1-\pi_3)^{n_3 - \sum_{i=1}^{n_3} z_i} \right]$$

$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi$ vs H_1 : Al menos un π_i es diferente.

$\Theta_0 = \{\pi_1 = \pi_2 = \pi_3 = \pi\} \rightarrow$

$$L = \pi^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i} (1-\pi)^{n_1+n_2+n_3 - \sum_{i=1}^{n_1} x_i - \sum_{i=1}^{n_2} y_i - \sum_{i=1}^{n_3} z_i} \rightarrow \text{El EMVde } \pi \text{ es: } \hat{\pi} = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i}{n_1 + n_2 + n_3}$$

Se necesitan estimadores MV de π_1, π_2 y π_3 en Θ:

$$\hat{\pi}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1}, \; \hat{\pi}_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2}, \; \hat{\pi}_3 = \frac{\sum_{i=1}^{n_3} z_i}{n_3}$$

$$\Lambda = \frac{\max\left[L(\Theta_0)\right]}{\max\left[L(\Theta)\right]} = \frac{\hat{\pi}^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i} (1-\hat{\pi})^{n_1+n_2+n_3 - \sum_{i=1}^{n_1} x_i - \sum_{i=1}^{n_2} y_i - \sum_{i=1}^{n_3} z_i}}{\left[\hat{\pi}_1^{\sum_{i=1}^{n_1} x_i} (1-\hat{\pi}_1)^{n_1 - \sum_{i=1}^{n_1} x_i} \right] \left[\hat{\pi}_2^{\sum_{i=1}^{n_2} y_i} (1-\hat{\pi}_2)^{n_2 - \sum_{i=1}^{n_2} y_i} \right] \left[\hat{\pi}_3^{\sum_{i=1}^{n_3} z_i} (1-\hat{\pi}_3)^{n_3 - \sum_{i=1}^{n_3} z_i} \right]}$$

Primero se reemplazan los siguientes valores en Λ:

$$\hat{\pi} = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i}{n_1 + n_2 + n_3} = \frac{10+4+4}{85+95+98} = \frac{18}{278}$$

$$\hat{\pi}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{10}{85}, \; \hat{\pi}_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2} = \frac{4}{95}, \; \hat{\pi}_3 = \frac{\sum_{i=1}^{n_3} z_i}{n_3} = \frac{4}{98}$$

Luego se obtiene:

$$-2\ln(\Lambda) \sim \chi^2_{(3-1)} \sim \chi^2_{(2)} \rightarrow -2\ln(\Lambda) = 5.18 \rightarrow \text{pvalor} = P\left[\chi^2_{(2)} > 5.18\right] = 0.0752 > \alpha = 0.05$$

$\rightarrow$ No se rechaza la H_0

b. Verifique si: $H_0 : \pi_1 = \pi_2 = \pi_3 = 0.09$ vs H_1 : Al menos un π_i es diferente de 0.09. Use $\alpha = 0.10$.

$X_i \sim \text{Binomial}(1,\pi_1)$, $i=1,2,\ldots,85$, $Y_i \sim \text{Binomial}(1,\pi_2)$, $i=1,2,\ldots,95$,

$Z_i \sim \text{Binomial}(1,\pi_3)$, $i=1,2,\ldots,98$

$$n_1=85,\quad \sum_{i=1}^{85} X_i=10;\ n_2=95,\quad \sum_{i=1}^{95} Y_i=4;\ n_3=98,\quad \sum_{i=1}^{98} Z_i=4$$

$$L=\prod_{i=1}^{n_1} f(x_i,\pi_1)\prod_{i=1}^{n_2} f(y_i,\pi_2)\prod_{i=1}^{n_3} f(z_i,\pi_3)=\prod_{i=1}^{n_1}\pi_1^{x_i}(1-\pi_1)^{1-x_i}\prod_{i=1}^{n_2}\pi_2^{y_i}(1-\pi_2)^{1-y_i}\prod_{i=1}^{n_3}\pi_3^{z_i}(1-\pi_3)^{1-z_i}$$

$$=\left[\pi_1^{\sum_{i=1}^{n_1}x_i}(1-\pi_1)^{n_1-\sum_{i=1}^{n_1}x_i}\right]\left[\pi_2^{\sum_{i=1}^{n_2}y_i}(1-\pi_2)^{n_2-\sum_{i=1}^{n_2}y_i}\right]\left[\pi_3^{\sum_{i=1}^{n_3}z_i}(1-\pi_3)^{n_3-\sum_{i=1}^{n_3}z_i}\right]$$

$\Theta_0=\{\pi_1=\pi_2=\pi_3=\pi_0\}, \pi_0=0.09 \rightarrow$

$$L=\pi_0^{\sum_{i=1}^{n_1}x_i++\sum_{i=1}^{n_2}y_i+\sum_{i=1}^{n_3}z_i}(1-\pi_0)^{n_1+n_2+n_3-\sum_{i=1}^{n_1}x_i-\sum_{i=1}^{n_2}y_i-\sum_{i=1}^{n_3}z_i}$$

Se necesitan estimadores MV de π_1,π_2 y π_3 en Θ:

$$\widehat{\pi}_1=\frac{\sum_{i=1}^{n_1}x_i}{n_1}=\frac{10}{85},\ \widehat{\pi}_2=\frac{\sum_{i=1}^{n_2}y_i}{n_2}=\frac{4}{95},\ \widehat{\pi}_3=\frac{\sum_{i=1}^{n_3}z_i}{n_3}=\frac{4}{98}$$

$$\Lambda=\frac{\max\left[L(\Theta_0)\right]}{\max\left[L(\Theta)\right]}=\frac{\pi_0^{\sum_{i=1}^{n_1}x_i+\sum_{i=1}^{n_2}y_i+\sum_{i=1}^{n_3}z_i}(1-\pi_0)^{n_1+n_2+n_3-\sum_{i=1}^{n_1}x_i-\sum_{i=1}^{n_2}y_i-\sum_{i=1}^{n_3}z_i}}{\left[\widehat{\pi}_1^{\sum_{i=1}^{n_1}x_i}(1-\widehat{\pi}_1)^{n_1-\sum_{i=1}^{n_1}x_i}\right]\left[\widehat{\pi}_2^{\sum_{i=1}^{n_2}y_i}(1-\widehat{\pi}_2)^{n_2-\sum_{i=1}^{n_2}y_i}\right]\left[\widehat{\pi}_3^{\sum_{i=1}^{n_3}z_i}(1-\widehat{\pi}_3)^{n_3-\sum_{i=1}^{n_3}z_i}\right]}$$

Primero se reemplazan los siguientes valores en Λ:

$$\pi_0=0.09, \widehat{\pi}_1=\frac{\sum_{i=1}^{n_1}x_i}{n_1}=\frac{10}{85},\ \widehat{\pi}_2=\frac{\sum_{i=1}^{n_2}y_i}{n_2}=\frac{4}{95},\ \widehat{\pi}_3=\frac{\sum_{i=1}^{n_3}z_i}{n_3}=\frac{4}{98}$$

Luego se obtiene:

$$-2\ln(\Lambda)\sim\chi^2_{(3-0)}\sim\chi^2_{(3)}\rightarrow -2\ln(\Lambda)=7.558\rightarrow \text{pvalor}=P\left[\chi^2_{(3)}>7.558\right]=0.056<\alpha=0.10$$

$\rightarrow$ Se rechaza la H_0

c. Verifique si: $H_0:\pi_1=\pi_3=0.08$ vs H_1: Al menos un π_i es diferente de 0.08. Use $\alpha=0.05$.

De la subpregunta b:

$$\Lambda = \frac{\max\left[L\left(\Theta_0\right)\right]}{\max\left[L\left(\Theta\right)\right]} = \frac{\pi_0^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_3} z_i}\left(1-\pi_0\right)^{n_1+n_3-\sum_{i=1}^{n_1} x_i - \sum_{i=1}^{n_3} z_i}}{\left[\widehat{\pi}_1^{\sum_{i=1}^{n_1} x_i}\left(1-\widehat{\pi}_1\right)^{n_1-\sum_{i=1}^{n_1} x_i}\right]\left[\widehat{\pi}_3^{\sum_{i=1}^{n_3} z_i}\left(1-\widehat{\pi}_3\right)^{n_3-\sum_{i=1}^{n_3} z_i}\right]}$$

Primero se reemplazan los siguientes valores en Λ:

$$\pi_0 = 0.08, \widehat{\pi}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{10}{85}, \ \widehat{\pi}_3 = \frac{\sum_{i=1}^{n_3} z_i}{n_3} = \frac{4}{98}$$

Luego se obtiene:

$$-2\ln\left(\Lambda\right) \sim \chi^2_{(2-0)} \sim \chi^2_{(2)} \rightarrow -2\ln\left(\Lambda\right) = 3.904 \rightarrow \text{pvalor} = P\left[\chi^2_{(2)} > 3.904\right] = 0.142 > \alpha = 0.05$$

$\rightarrow$ No se rechaza la H_0

Ejemplo 32

La Sicóloga Leonor Respeto hizo el estudio del número de agresiones hechas por las barras bravas en las afueras de los estadios cuando hay un día de clásico. Se considera que el número de agresiones, por unidad de tiempo, en los países tiene distribución Poisson. Ella quiso comparar 3 países (Perú, Ecuador y Colombia) e hizo el registro del total de clásicos jugados por país en cierto periodo: 58, 62 y 74. El número total de agresiones que se registró en cada muestra de estos países fueron 1470, 1478 y 1892. ¿Hay evidencias estadísticas para establecer que, en promedio, el número de agresiones por día de clásico no difiere en los tres países? Use $\alpha = 0.05$.

$X_i \sim \text{Poisson}\left(\lambda_1\right), i = 1,2,\ldots,58, Y_i \sim \text{Poisson}\left(\lambda_2\right), i = 1,2,\ldots,62, Z_i \sim \text{Poisson}\left(\lambda_3\right)$
$, i = 1,2,\ldots,74$

$$n_1 = 58, \quad \sum_{i=1}^{58} X_i = 1470; \ n_2 = 62, \quad \sum_{i=1}^{62} Y_i = 1478; \ n_3 = 74, \quad \sum_{i=1}^{74} Z_i = 1892$$

$$L = \prod_{i=1}^{n_1} f\left(x_i\right) \prod_{i=1}^{n_2} f\left(y_i\right) \prod_{i=1}^{n_3} f\left(z_i\right) =$$

$$= \prod_{i=1}^{n_1} \frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!} \prod_{i=1}^{n_2} \frac{e^{-\lambda_2}\lambda_2^{y_i}}{y_i!} \prod_{i=1}^{n_3} \frac{e^{-\lambda_3}\lambda_3^{z_i}}{z_i!}$$

$$= \left(\frac{1}{\prod_{i=1}^{n_1} x_i! \prod_{i=1}^{n_2} y_i! \prod_{i=1}^{n_3} z_i!}\right)\left[e^{-n_1\lambda_1}\lambda_1^{\sum_{i=1}^{n_1} x_i}\right]\left[e^{-n_2\lambda_2}\lambda_2^{\sum_{i=1}^{n_2} y_i}\right]\left[e^{-n_3\lambda_3}\lambda_3^{\sum_{i=1}^{n_3} z_i}\right]$$

$H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda$ vs H_1: Al menos un λ_i es diferente.

$\Theta_0 = \{\lambda_1 = \lambda_2 = \lambda_3 = \lambda\} \rightarrow$

$$L = \left(\frac{1}{\prod_{i=1}^{n_1} x_i! \prod_{i=1}^{n_2} y_i! \prod_{i=1}^{n_3} z_i!} \right) e^{-(n_1+n_2+n_3)\lambda} \lambda^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i} \rightarrow \text{El EMVde } \pi \text{ es: } \lambda = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i}{n_1 + n_2 + n_3}$$

Se necesitan estimadores MV de λ_1, λ_2 y λ_3 en Θ:

$$\lambda_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1}, \widehat{\lambda_2} = \frac{\sum_{i=1}^{n_2} y_i}{n_2}, \widehat{\lambda_3} = \frac{\sum_{i=1}^{n_3} z_i}{n_3}$$

$$\Lambda = \frac{\max\left[L(\Theta_0)\right]}{\max\left[L(\Theta)\right]} = \frac{e^{-(n_1+n_2+n_3)\hat{\lambda}} \lambda^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i}}{e^{-n_1\hat{\lambda}_1} e^{-n_2\hat{\lambda}_2} e^{-n_3\hat{\lambda}_3} \left[\lambda_1^{\sum_{i=1}^{n_1} x_i}\right]\left[\lambda_2^{\sum_{i=1}^{n_2} y_i}\right]\left[\lambda_3^{\sum_{i=1}^{n_3} z_i}\right]}$$

Primero se reemplazan los siguientes valores en Λ:

$$\lambda = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i + \sum_{i=1}^{n_3} z_i}{n_1 + n_2 + n_3} = \frac{1470 + 1478 + 1892}{58 + 62 + 74} = \frac{2420}{97}$$

$$\lambda_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{1470}{58}, \lambda_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2} = \frac{1478}{62}, \lambda_3 = \frac{\sum_{i=1}^{n_3} z_i}{n_3} = \frac{1892}{74}$$

Luego se obtiene:

$-2\ln(\Lambda) \sim \chi^2_{(3-1)} \sim \chi^2_{(2)} \rightarrow -2\ln(\Lambda) = 4.59789 \rightarrow$

$\text{pvalor} = P\left[\chi^2_{(2)} > 4.59789\right] = 0.10036 > \alpha = 0.05$

$\rightarrow$ No se rechaza la H_0

Ver el Script R.

Ejercicios

1. Se hace el siguiente experimento: Considere que un "intento" consiste en determinar el número de días de acopio de pollo en pie donde la demanda fue menor que 175 mil kg antes de encontrar el día donde la demanda fue de por lo menos 175 mil kg. En una muestra aleatoria de 5 "intentos" se encontraron los siguientes valores 4, 6, 5, 7 y 3. Suponga que en el $\theta \times 100\%$ de los días la demanda es de por lo menos 175 mil kg. Haga una prueba aleatoria para verificar $H_0: \theta = 0.3$ vs $H_1: \theta = 0.2$ con un nivel de significación de 5%.

2. Sea 8.9, 12.4, 12.1, 10, 9.2, 13.7, 13.9, 9.1, 8.8, 6.3, 12.1, 10.9, 12.5, 9.5, 8.2, 10.2 una muestra aleatoria de pesos de niños atendidos en una clínica. La distribución de los pesos es $N(\mu,\sigma^2=9)$. Usando el Lema de Neyman-Pearson halle la prueba óptima de tamaño $\alpha=0.05$ para probar $H_0:\mu=10$ vs $H_1:\mu=11$. Cuál es su conclusión.

3. Suponga que se toma una muestra aleatoria de cuatro observaciones de la densidad Weibull: $f(x;\theta)=\frac{5x^4}{\theta}\exp\left(-\frac{x^5}{\theta}\right)I_{(0,\infty)}(x)$. Determine la prueba estadística más potente de H_0: $\theta = 2$ vs H_1: $\theta = 4$. Usar α=0.1

4. Se estudió el rendimiento X en toneladas que se tiene con unidades experimentales de 15 toneladas de materia prima de harina de la variedad B de yuca que se procesaron en tolvas para obtener pastas de yuca. Se desea probar las hipótesis $H_0: f_1(x)=\frac{2(x+1)}{301}I_{(17,24)}(x)$ vs $H_1: f_2(x)=\frac{3(x^2+1)}{8932}I_{(17,24)}(x)$. Si una unidad experimental registró un rendimiento de 22 toneladas ¿cuál es su conclusión? Utilice un nivel de significación de 5%.

5. Suponga que X~ N(0,1) bajo H_0 y $g(x)=\frac{1}{2}\exp\{-|x|\}I_{(-\infty,\infty)}(x)$ bajo H_1. Halle la prueba más poderosa para probar H_0 vs H_1, basados en una observación de X. ¿Cuál es su conclusión si x = 2.2 ? Use α = 0.05.

6. La Ingeniera María Anna Mozart estudió la proporción de mediciones de suelo X, en cierta región, donde la temperatura es mayor que 19°C y concluyó que X tiene distribución $\text{Uniforme}(0,1)$. En cambio la Ingeniera Chabuca Granda ha establecido que la densidad de X es una $\text{Beta}(1,2)$. El Ingeniero Antonio Salieri tomó una muestra de tamaño 1 (100 mediciones con 4 mayores de 19°C) para contrastar la hipótesis planteada de María Anna Mozart vs la alternativa de Chabuca Granda. Con un nivel de significación de 5%, cuál será la conclusión de Antonio Salieri. Calcule la potencia de la prueba.

7. Se quiere hacer una encuesta a una pareja que espera su primer hijo en cada uno de 8 distritos. Se estudia la variable aleatoria X definida como el número de parejas que son consultadas antes de encontrar la pareja que acceda a responder la encuesta en el distrito i. Sea 1, 0, 0, 3, 1, 1, 0, 0 una muestra de los 8 distritos. Halle la prueba aleatoria apropiada y de su conclusión para $H_0:\theta=0.4$ vs $H_1:\theta>0.4$, si se rechaza la hipótesis planteada con valores bajos de la estadística suficiente. Utilice $\alpha=0.05$.

8. De los camales de Lima Metropolitana se tiene la información del número de cabezas y los volúmenes de ganado vacuno en miles de kilos, beneficiados. Suponga que el número de cabezas de ganado vacuno beneficiado en 5 minutos tiene distribución de $\text{Poisson}(\lambda)$. Si 5, 4, 14, 9 y 12 es una muestra aleatoria de 5 intervalos de 5 minutos, primero con una prueba aleatoria verifique si $H_0:\lambda=10$ vs $H_1:\lambda=8$, y luego calcule la potencia de la prueba.

9. Si la proporción de escolares con padres separados, en los colegios de cierta ciudad, tiene distribución $f(x)=(1+\theta)x^{\theta}I_{(0,1)}(x)$ con $\theta>-1$.

a. Pruebe $H_0:\theta=-0.25$ vs $H_1:\theta=-0.23$, si 0.61, 0.59, 0.64, 0.58, 0.63 es una muestra obtenida en 5 colegios. Utilice $\alpha=0.05$.

b. Verifique $H_0:\theta=-0.20$ vs $H_1:\theta>-0.20$ con la muestra de la pregunta a). Emplee $\alpha=0.05$.

10. De los camales de Lima Metropolitana se tiene la información del número de cabezas y los volúmenes de ganado vacuno en miles de kilos, beneficiados. La siguiente muestra aleatoria de 6 días: 52, 84, 404, 308, 224 y 388 corresponde a los volúmenes, en miles de kilos, de ganado vacuno. Asumiendo que la muestra se extrajo de una distribución exponencial con media θ. Pruebe $H_0:\theta=200$ vs $H_1:\theta\neq 200$.

11. Se considera que el rendimiento mínimo mensual X (en toneladas que se tiene con unidades experimentales de 15 toneladas de materia prima de harina de la variedad C de yuca que se procesaron en tolvas para obtener pastas de yuca) tiene distribución del valor mínimo tipo II de Fréchet cuya densidad es $f(x)=\frac{\alpha}{\mu}\left(\frac{x}{\mu}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\mu}\right)^{\alpha}\right]I_{(0,\infty)}(x)$. Considere que el parámetro $\alpha=5$ y que una m.a de tamaño 4 de $f(x)$ es 35, 28, 32, 34 toneladas y pruebe $H_0:\mu=36$ vs $H_1:\mu\neq 36$. Utilice un nivel de significación de 0.05.

12. De los camales de Lima Metropolitana se tiene la información del número de cabezas y los volúmenes de ganado porcino en miles de kilos, beneficiados. La siguiente muestra aleatoria de 6 días: 224, 48, 128, 68, 74 y 104 corresponde a los volúmenes, en miles de kilos, de ganado porcino. Asumiendo que la muestra se extrajo de una distribución $\text{Pareto}(\alpha,\beta=2.25)$. Pruebe $H_0:\alpha=50$ vs $H_1:\alpha\neq 50$.

13. Sea X = Nº de "fallas" de una pareja (que desea tener dos niñas) hasta que la pareja tenga dos niñas . Pruebe si el porcentaje de niñas que nacen es a lo más 45%. Considere a: 1, 0, 0, 2, 0 como las observaciones de una muestra aleatoria de tamaño 5. use $\alpha = 0.05$.

14. La siguiente distribución corresponde a la proporción de población gestante atendida por personal capacitado en los países de las Américas: $f\left(x\right) = \theta x^{\theta-1} I_{(0,1)}\left(x\right)$ con $\theta > 0$. Con la muestra aleatoria 0.28, 0.84, 0.99, 0.78 y 1.00 verifique lo siguiente $H_0 : E\left(X\right) = 0.8$ vs $H_1 : E\left(X\right) > 0.8$. Utilice un nivel de significación de 5%.

15. En la Tesis del Ingeniero Mariano Melgar se encontró que la cantidad de nitrógeno X, en porcentaje, en el aire del suelo de cierta región A tiene densidad $f\left(x\right) = \frac{1}{5} e^{-\frac{1}{5}(x-\theta)} I_{(\theta;\infty)}(x)$. El Ingeniero Melgar tomó una muestra de mediciones con valores 78.45, 78.42, 78.89, 78.94 y 78.54. Con esa muestra pruebe las hipótesis $H_0 : \theta = 78\%$ vs $H_1 : \theta > 78\%$ usando un nivel de significación de 0.05. Diga su conclusión.

16. Si 10.5, 15.6, 10.4, 18.7 y 17.3 son las observaciones de una m.a de tamaño 5 de $f\left(x\right) = \frac{4x^3}{\theta^4} I_{(0,\theta)}\left(x\right)$. Utilice la prueba uniformemente más poderosa de tamaño $\alpha = 0.05$ para verificar $H_0 : \theta \geq 22$ vs $H_1 : \theta < 22$.

17. La Licenciada Leonor Respeto hizo un estudio del costo en unidades monetarias de las agresiones hechas por las barras bravas en las afueras de los estadios cuando hay un día de clásico. El costo por atender a cada agredido, en decenas de soles, tiene densidad $f\left(x\right) = \frac{3\sigma^3}{x^4} I_{(\sigma,\infty)}\left(x\right)$. Con la muestra aleatoria: 25, 22, 102 y 44 considere la prueba uniformemente más poderosa de tamaño $\alpha = 0.05$ para probar $H_0 : \sigma \leq 16$ vs $H_1 : \sigma > 16$ ¿Cuál es su conclusión?

18. El tiempo, en minutos, que demora un ratón de laboratorio en recorrer un laberinto experimental tiene densidad $f\left(x\right) = \frac{1}{\theta x^{\frac{1}{\theta}+1}} I_{(1,\infty)}\left(x\right)$ con $0 < \theta < 1$. Si 1.11, 1.15, 1.14 y 1.18 son las observaciones de una muestra aleatoria de esa distribución. Con un nivel de significación de 0.05 verifique si en promedio los ratones demoran menos de 2 minutos en recorrer el laberinto.

19. La siguiente muestra: 10.1, 11.3, 5.2, 13.4, 6.3, 4.5 corresponde a la cantidad de metros de alambre que una máquina produce hasta encontrar una falla y proviene de

la siguiente distribución $f(x)=\frac{1}{\lambda}e^{-\frac{1}{\lambda}(x-\theta)}I_{(\theta,\infty)}$, $\lambda>0$ con $\theta=2$. Halle la prueba uniformemente más poderosa para probar $H_0:\lambda\geq 13$ vs $H_1:\lambda<13$.Concluya con $\alpha=0.05$.

20. Con la muestra: 2.0, 2.3, 2.2, 2.1 de la densidad $f\left(x\right)=\frac{8\,x^7\,\exp\left(-\frac{x^8}{\theta}\right)}{\theta}I_{(0,\infty)}\left(x\right)$. Halle la prueba uniformemente más poderosa para $H_0:\theta=2$ *vs* $H_1:\theta>2$ y concluya con $\alpha=0.05$.

21. La velocidad máxima anual del viento ($millas/h$) en la ciudad de Chiclayo tiene como densidad: $f\left(x\right)=\frac{k}{\mu}\left(\frac{\mu}{x}\right)^{k+1}\exp\left[-\left(\frac{\mu}{x}\right)^{k}\right]I_{[0,\infty)}\left(x\right)$ con k conocida. Considere en esta pregunta que el parámetro $k=6.5$ y una m.a de tamaño 4 de $f\left(x\right)$ es: 53, 55, 58, 56 $millas/h$ y pruebe $H_0:\mu=62$ vs $H_1:\mu<62$. Utilice un nivel de significación de 0.05.

22. Sea $X_1,\ldots,X_{10}$ una m.a de una $N\left(\mu,\sigma^2\right)$ con $\mu\,y\,\sigma^2$ desconocidas. Halle la prueba de razón de verosimilitud generalizada para $H_0:\mu=4$ vs $H_1:\mu\neq 4$. Use $\alpha=0.05$ y concluya con $\sum_{i=1}^{10}x_i=50$.

23. Según el Ingeniero Manuel Scorza, la proporción de nitrógeno en el aire del suelo en una región B tiene densidad $f\left(x\right)=\theta x^{\theta-1}I_{(0,1)}\left(x\right)$. Con una muestra de tamaño 1, primero halle la prueba más poderosa de tamaño $\alpha=0.10$ para probar $H_0:\theta=1$ vs $H_1:\theta\neq 1$y luego diga si es una prueba insesgada.

24. Un Ingeniero hace el siguiente experimento: Considerando que un “intento” consiste en determinar el número de días de acopio de porcino donde el ingreso no fue de por lo menos 220 mil kg antes de encontrar el día donde el ingreso fue de por lo menos 220 mil kg. él seleccionó una muestra aleatoria de 500 intentos y obtuvo un total de 1800 días de acopio de porcino donde el ingreso no fue de por lo menos 220 mil kg antes de encontrar el día donde el ingreso fue de por lo menos 220 mil kg. Con un nivel de significación $\alpha=0.05$ haga una prueba no aleatoria para verificar $H_0:\theta=0.20$ vs $H_1:\theta\neq 0.20$.

25. Se considera que el número de autos robados, por unidad de tiempo, en tres ciudades (A, B y C) tiene distribución Poisson. Fueron seleccionados 200, 250 y

240 días de cada ciudad y el número total de autos robados que se registró en cada muestra fue 900, 1000 y 1008. ¿Hay evidencias estadísticas para establecer que en promedio, el número de autos robados por día no difiere en las tres ciudades? Use $\alpha = 0.05$.

26. En tres ciudades se hizo un estudio del número de compromisos (enamoramientos) que tuvo una persona adulta antes de casarse. Se tomaron muestras de 205, 208 y 203 adultos de cada ciudad y se encontró que: $\sum_{i=1}^{205} X_{iA} = 109, \sum_{i=1}^{208} X_{iB} = 115 \ y \sum_{i=1}^{203} X_{iC} = 98$. Con un nivel de significación de 5% verifique si el porcentaje de adultos casados en las tres ciudades es igual a 62%.

27. Se dice que $X \sim \text{Laplace}(\mu, \lambda)$ si $f_X(x) = \dfrac{\exp\left[-\dfrac{|x-\mu|}{\lambda}\right]}{2\lambda}$. En Ecuador, Perú y Argentina la tasa de crecimiento anual, en %, de las empresas farmacéuticas tienen distribución $\text{Laplace}(\mu = 0, \lambda_i)$ para $i = 1, 2, 3$. Se tomaron muestras de 205, 218 y 225 empresas de cada país y se encontró $\sum_{i=1}^{205} |X_{i1}| = 307.5$, $\sum_{i=1}^{218} |X_{i2}| = 414.2$ y $\sum_{i=1}^{225} |X_{i3}| = 442.8$. Pruebe si:

a. $H_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda$ con $\alpha = 0.05$.

b. $H_0 : \lambda_1 = \lambda_2 = \lambda_3 = 1.698$ con $\alpha = 0.05$.

28. Bacterias Psicrófilas: Son aquellas capaces de vivir a temperaturas por debajo de los 5 °C. Para las preguntas de abajo se utilizan porciones de carne de 75 gr. Se mide el número de bacterias psicrófilas en la carne después de 9 días de almacenamiento a 4ºC. La variable respuesta o medición fue el logaritmo del número de bacterias por cm^2. (Estas bacterias se encuentran en la superficie de la carne y aparecen cuando la carne se echó a perder). En el laboratorio A, el logaritmo del número de bacterias por cm^2 mínimo diario X tiene distribución del valor mínimo tipo II de Fréchet entonces $f(x) = \dfrac{\alpha}{\mu}\left(\dfrac{x}{\mu}\right)^{\alpha-1} \exp\left[-\left(\dfrac{x}{\mu}\right)^{\alpha}\right] I_{(0,\infty)}(x)$.

a. Considere en esta pregunta que el parámetro $\alpha = 4$ y que una muestra aleatoria de 4 porciones de 75 gr de carne se hallaron los siguientes logaritmos 7.4, 8.2, 6.8 y 7.6. Pruebe $H_0 : \mu = 7$ vs $H_1 : \mu \neq 7$. Utilice un nivel de significación de 0.05.

b. Suponga en esta pregunta que el parámetro $\alpha = 4$ y que en una porción de 75 gr de carne se obtuvo un logaritmo igual a 6.9644. Verifique $H_0: \mu \leq 6$ ó $\mu \geq 8$ vs $H_1: 6 < \mu < 8$. Emplee un nivel de significación de 0.05.