Capítulo 7

Contraste de hipótesis

7. Contraste de hipótesis

Ejemplo introductorio
Hace un mes el precio de un bien tenía distribución N(100,10^2) y en la actualidad, se sabe que ha aumentado en θ unidades monetarias, pero se desconoce el valor de θ, por lo que se tomó una m.a. de n comerciantes del bien para registrar sus precios actuales y estimar el parámetro.

a) Halle $\hat{\theta}_{MV}$ y apoyándose en $\hat{\theta}_{MV}$ para generar la variable base, construya un intervalo de confianza al 95% para θ.
b) Si la m.a. dio los precios: 111, 105, 103, 110, 109 y 107, aplique estos datos al I.C. y vea si sería razonable concluir que ***el precio promedio del bien ha crecido en más de 4%.***

Solución:

a) Tenemos que hacer dos cosas, primero estimar θ y luego identificar la distribución del estimador $\hat{\theta}_{MV}$ para tomarla de base en la construcción de un intervalo de confianza.

Cálculo de $\hat{\theta}_{MV}$
Si con X denotamos el precio hace un mes y con Y el precio actual, entonces $Y = X + \theta$, y en particular Y es **función lineal de una variable normal,** de modo que podemos aplicar la **Propiedad Reproductiva** de la distribución Normal: **"Funciones lineales de variables normales, también tienen distribución normal, pero con su respectiva media y varianza"**. Esto es, $X \sim N(100,10^2)$ y $Y = X + \theta \Rightarrow Y \sim N(\mu_Y, \sigma_Y^2)$ donde
$\mu_Y = E(Y) = E(X+\theta) = E(X) + E(\theta) = 100 + \theta$ y $\sigma_Y^2 = V(Y) = E[(Y-\mu_Y)^2] =$
$= E[(X+\theta-100-\theta)^2] = E[(X-100)^2] = V(X) = 10^2$

Escribiendo la función de verosimilitud en términos de la muestra de valores actuales Y_j (pues esos son los datos que tenemos y no los de X_j):

$L(\theta) = \prod_{j=1}^{n} f_{Y_j}(y_j;\theta) = \prod_{j=1}^{n} \frac{e^{-(y_j-\mu_Y)^2/2\sigma_Y^2}}{\sqrt{2\pi}\sigma_Y} = \prod_{j=1}^{n} \frac{e^{-(y_j-100-\theta)^2/2*10^2}}{\sqrt{2\pi}*10} = \frac{e^{-\frac{1}{2*10^2}\sum_{j=1}^{n}(y_j-100-\theta)^2}}{(\sqrt{2\pi}*10)^n}$. Tomando logaritmo neperiano:

$\ln[L(\theta)] = -\frac{1}{2*10^2}\sum_{j=1}^{n}(y_j-100-\theta)^2 - \ln(\sqrt{2\pi}*10)^n$, que es función derivable, así que podemos aplicar cálculo diferencial para maximizar:

$\frac{\partial \ln[L(\theta)]}{\partial \theta} = 0 \Rightarrow -\frac{1}{2*10^2}\sum_{j=1}^{n}2(y_j-100-\theta)(-1) = 0 \Rightarrow \sum_{j=1}^{n}(y_j-100-\theta) = 0 \Rightarrow \sum_{j=1}^{n}y_j - n100 - n\theta = 0$
$\Rightarrow \theta = \bar{Y} - 100$ y verificando el máximo, se llega a que $\hat{\theta}_{MV} = \bar{Y} - 100$. Nótese que:

$E(\hat{\theta}_{MV}) = E(\overline{Y} - 100) = E(\overline{Y}) - 100 = \mu_Y - 100 = (100+\theta) - 100 = \theta$, es decir, $\hat{\theta}_{MV}$ es insesgado. Además $V(\hat{\theta}_{MV}) = V(\overline{Y} - 100) = E[([\overline{Y} - 100] - \theta)^2] = E[(\overline{Y} - (100+\theta))^2] = V(\overline{Y}) = \frac{10^2}{n}$

Intervalo de Confianza para θ

$\hat{\theta}_{MV} = \overline{Y} - 100$ es obviamente función lineal de $\overline{Y}$, que a su vez es función lineal de las Y_j que tienen todas distribución normal $N(\mu_Y = 100+\theta, \sigma_Y^2 = 10^2)$, por tanto (y de nuevo, usando la Propiedad Reproductiva de la distribución normal), podemos escribir: $\hat{\theta}_{MV} \sim N(E[\hat{\theta}_{MV}], V[\hat{\theta}_{MV}])$ y como ya probamos $E[\hat{\theta}_{MV}] = \theta$ y $V[\hat{\theta}_{MV}] = \frac{10^2}{n}$. En conclusión, $\hat{\theta}_{MV} \sim N(\theta, \frac{10^2}{n})$.

- Tomemos como <u>variable base</u> del I.C. a $Z = \frac{(\hat{\theta} - \theta)}{10/\sqrt{n}} \sim N(0,1)$ pues tiene al parámetro θ en su estructura y su distribución no depende de θ .

- Planteando $P(a \le \frac{(\hat{\theta} - \theta)}{10/\sqrt{n}} \le \text{b}) = 0.95$ y tomando a y b simétricos en probabilidad, esto es, $P(Z \le \text{a}) = P(Z \ge b) = 0.05/2 = 0.025 \Rightarrow$ b=1.96 y a=-1.96

- Finalmente, despejando θ del interior del intervalo en la probabilidad:
  $$0..95 = P(-1.96 \le \frac{(\hat{\theta} - \theta)}{10/\sqrt{n}} \le 1.96) = P(-1.96\frac{10}{\sqrt{n}} \le \hat{\theta} - \theta \le 1.96\frac{10}{\sqrt{n}}) = P(\hat{\theta} - 1.96\frac{10}{\sqrt{n}} \le \theta \le \hat{\theta} + 1.96\frac{10}{\sqrt{n}})$$
  $\Rightarrow$ El I.C. para θ es $\hat{\theta} - 1.96\frac{10}{\sqrt{n}} \le \theta \le \hat{\theta} + 1.96\frac{10}{\sqrt{n}}$ con 95% de confianza.

b) De la muestra tenemos: $n = 6$, $\overline{Y} = 107.5 \Rightarrow \hat{\theta}_{MV} = \overline{Y} - 100 = 107.5 - 100 = 7.5$ y reemplazando en la fórmula del I.C.:

$\hat{\theta} - 1.96\frac{10}{\sqrt{n}} \le \theta \le \hat{\theta} + 1.96\frac{10}{\sqrt{n}} \Leftrightarrow 7.5 - 1.96\frac{10}{\sqrt{6}} \le \theta \le 7.5 + 1.96\frac{10}{\sqrt{6}} \Leftrightarrow -0.5 \le \theta \le 15.5$ y como por el contexto del problema $\theta > 0$, tenemos que $0 < \theta \le 15.5$ con 95% de confianza.

Dado que el precio medio inicial era 100, entonces el incremento de precios está entre 0 y 15.5%.

La afirmación ***el precio promedio del bien ha crecido en más de 4%*** equivale a $4 < \theta$ y según el I.C. $0 < \theta \le 15.5$ (con 95% de confianza, claro), por tanto ***no se puede asegurar que el precio promedio ha crecido en más de 4%, pues del I.C. se ve que también podría ser un porcentaje menor.***

El ejemplo muestra la utilidad de los intervalos de confianza para explorar hipótesis generadas a partir de datos. Pero también muestra que a veces los intervalos no pueden ayudarnos a decidir de modo concluyente. Y la razón de ello es que no han sido diseñados para cumplir ese rol. En particular no usan toda la información que hay en la muestra (en el ejemplo, realmente sólo nos interesaba el lado izquierdo del intervalo y la probabilidad a la derecha no se usa) y además no controlan totalmente las probabilidades de error, derivadas del hecho de estar tomando decisiones a partir de muestras. Se necesita una herramienta ad hoc. Esta herramienta se llama Prueba o Contraste de hipótesis.

7.1 El Problema del contraste o prueba de hipótesis

Se tiene una hipótesis de investigación, relativa a una población determinada y se desea estudiar la verdad o falsedad de dicha hipótesis a partir del análisis de una muestra aleatoria tomada de la población.

7.2 Los Elementos

Hipótesis Estadística
Es una afirmación acerca de la distribución poblacional asociada a una variable X. La denotamos mediante H.

Por ejemplo:
(a) H: "X tiene distribución normal con media $\mu = 0$ y $\sigma = 1$"
(b) H: "X es independiente de Y"
(c) H: "La distribución de X es no es normal"

Observaciones:
(1) Si la hipótesis H especifica totalmente la distribución de X, se dice que es "simple" ; en caso contrario, es "compuesta". Así, el ejemplo (a) es de hipótesis simple, y (c) es compuesta.

(2) Toda hipótesis H es una proposición o afirmación, por tanto o es verdadera o es falsa. En consecuencia, H debiera se aceptada o debiera ser rechazada.

(3) La única manera estadística de ver si H es verdadera o falsa, es ver si se cumple o no se cumple en todos y cada uno de los elementos de una Población. Normalmente esto no es posible y tendremos que contentarnos con examinar una muestra y tomar una decisión respecto a H, sabiendo que existe la posibilidad de error. **La ventaja del procedimiento estadístico es que se puede fijar la probabilidad de error.**

En economía, por lo general lo que se tiene es una hipótesis o explicación teórica de algún proceso, y partiendo de la explicación se genera una hipótesis de trabajo (H.T.), que uno desea verificar con evidencia empírica. Si la evidencia proviene de muestras, entonces podemos aprovechar las herramientas estadísticas para evaluar la hipótesis, pero para ello, debemos previamente frasear la H.T. en términos estadísticos.

Por ejemplo, si debido a la demanda de minerales en el mercado mundial y al acceso a mayores cuotas en mercados antes restringidos, observamos una mayor actividad en los sectores minero y agrícola de exportación, es natural esperar un cierto incremento en el empleo. En particular, esto implica que la tasa π de desempleo abierto actual debiera ser menor que la de hace un par de años, en que era de 9%. La hipótesis de trabajo puede ser algo como: "***El desempleo actual en la economía es menor que el de hace dos años"*** ¿Cómo pasar esta hipótesis a un formato estadístico que permita su contraste con datos reales? Una manera podría ser la siguiente:

Supongamos que tomamos al azar a un elemento de la PEA. ¿Cuál es la probabilidad de que se encuentre en condición de desempleo abierto? La respuesta es que esa probabilidad es π . Ahora bien, si tomamos una muestra al azar de n elementos de la PEA, independientemente entre sí, ¿Cuál es la distribución de la v.a. X = número de personas en situación de desempleo abierto en la muestra? Recordando el proceso que genera la distribución binomial, tenemos que $X \sim B(x;n,p=\pi)$, pues se cumplen todos los supuestos del modelo (nótese que aquí la distribución binomial es un <u>modelo de datos</u>, no es un modelo económico)

Ahora bien, notemos que hay dos posibilidades: H.T. es verdadera o es falsa. Y de ahí se deriva que:

H.T. verdadera $\Rightarrow X \sim B(x;n,p<0.09)$, pues bajar el desempleo implica $\pi < 0.09$

H.T. falsa $\Rightarrow X \sim B(x;n,p=0.09)$, ya que aún si no hay reactivación en el empleo, de todos modos sí hay reactivación económica por el contexto de mayor demanda y por tanto mayor producción, lo que si no baja el desempleo, por lo menos no lo aumenta, es decir π no cambiará su valor y seguirá $\pi = 0.09$

Así, podemos someter a prueba la hipótesis estadística $H: X \sim B(x; n, p = 0.09)$ para evaluar nuestra hipótesis de trabajo H.T.

El problema es cómo hacerlo. Esa es la cuestión por resolver. Felizmente, el tener un modelo de datos, provee de material suficiente. Por ejemplo, como ya sabemos que para este modelo, $\mu_X = np = n\pi$, este hecho puede darnos una pista de un método para evaluar nuestra H.T.: simplemente tomemos una muestra de n casos (por ejemplo 100 casos) y si el valor observado cae muy por debajo de lo esperado μ_X (en el caso de $n = 100$, lo esperado es $\mu_X = 9$ desempleados), podemos considerar que nuestra H.T. es cierta. En las líneas que siguen, precisamos mejor estas ideas.

Contraste Estadístico de Hipótesis
Es un sistema de análisis a partir del cual se toma la decisión de aceptar ó rechazar una hipótesis estadística H_0, mediante el análisis de una(s) muestra(s) aleatoria(s).

Observación
Lo que hace el estadístico es tomar una muestra aleatoria y ver si los datos reales coinciden con la hipótesis; si esto último no ocurre, se dice que la hipótesis ha sufrido un "contraste" y se la rechaza. Por eso el nombre de "contraste de hipótesis" que se aplica a esta área de la estadística inferencial. Otra palabra usada es la de "Prueba" o "Test" (del inglés), pero no en el sentido de demostrar, sino de poner a prueba la hipótesis frente a datos reales o "evidencia empírica".

Hipótesis Planteada o Nula
Es aquella que se somete a prueba. La denotaremos H_0

Hipótesis Alterna
Es aquella que nos queda si se rechaza H_0. Esta hipótesis se denotará H_1 .

Observación:
Normalmente H_1 es el complemento de H_0, **pero dentro del contexto de la investigación.**

Ejemplo 1
(1) Si sabemos que $X \sim N(\mu \geq 0, \sigma^2)$ y escribimos H_0: $\mu = 0$, entonces H_1 es H_1: $\mu > 0$
(2) Si sabemos que $X \sim N(\mu, \sigma^2)$ y escribimos H_0: $\mu = 0$, entonces, a diferencia del ejemplo (1), H_1 es dada por H_1: $\mu \neq 0$

Región Crítica
Denotada C, es un evento cuya ocurrencia conduce al rechazo de H_0.

Error Tipo I
Es el error que se produce si se rechaza H_0 injustamente siendo verdadera)

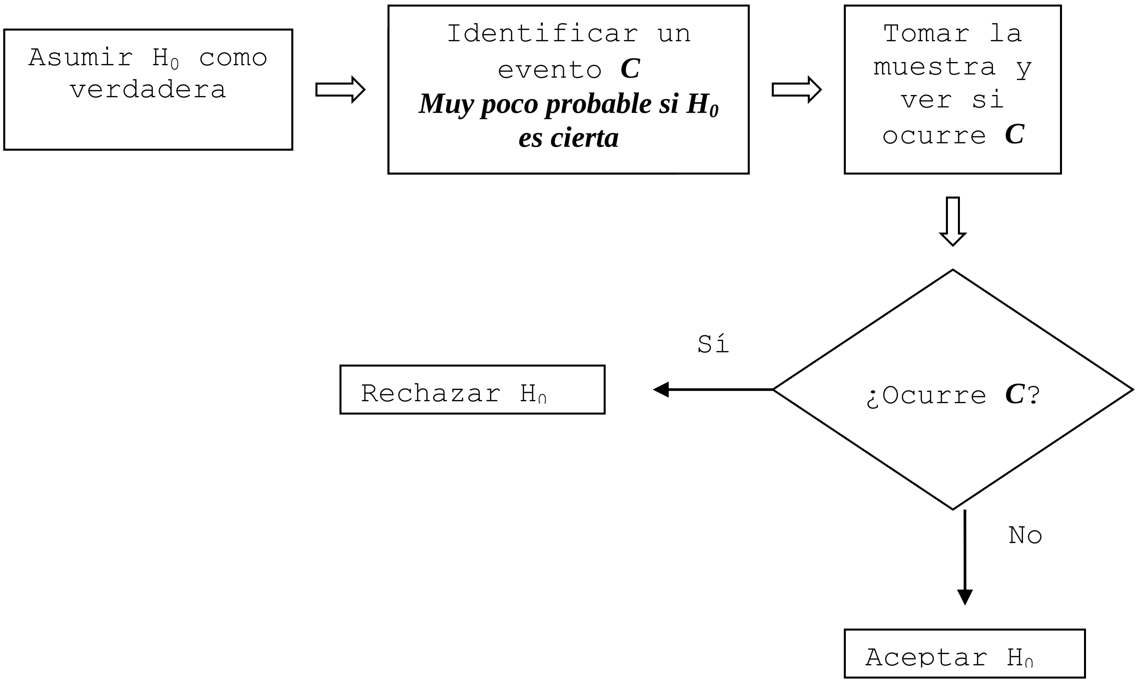
Error Tipo II
Es el error que se produce si se acepta H_0 siendo falsa

Definiciones:
Dada una región crítica C, es posible cometer ó el Error I ó el Error II. La probabilidad del Error I se denota α, y la probabilidad del Error II se denota ß, esto es:

α = P(Error I) = P(C | H_0 es V). La probabilidad α se llama también " Nivel de Significación del contraste" y mide el tamaño probabilístico de la región crítica

ß = P(Error II)= P(C^c | H_0 es F) = P(C^c | H_1 es V).

Metodología general



Observaciones:

a) Internacionalmente se considera que α es apropiado si es menor o igual que 0.05 (o no mayor que 5%)
b) Como ya se dijo, la región C se construye en términos de alguna variable estadística.
c) Lo que hace el estadístico es usar las leyes de probabilidad para encontrar la distribución teórica del estadístico de contraste asumiendo que la hipótesis nula H_0 es cierta; luego se determina un rango C de valores del estadístico de contraste, que tenga una probabilidad conocida α de ocurrir. Este rango C es la región crítica o sea, si al tomar la muestra ocurre C, se rechazará H_0) y su probabilidad α es el Nivel de Significación asociado.
d) Cada test o contraste tiene un estadístico de contraste y una distribución teórica, cuyas probabilidades suelen estar tabuladas.

7.3 El Problema del Contraste Estadístico de Hipótesis.

Dado un juego de hipótesis H_0 vs H_1, se desea hallar la Región Crítica C que minimiza **ß**, para un valor predeterminado de α , y un tamaño de muestra n dado.

Observaciones:
Una región crítica C se dice "óptima" si para un nivel α dado, minimiza la probabilidad β. Las regiones críticas que figuran en los textos y/o paquetes estadísticos son optimas.

Aunque α y β no son complementarios, sí se puede demostrar que dado un tamaño de muestra n, α y β van en sentidos contrarios: O sea, si reducimos α entonces β "crece". *Por eso, si deseamos que ambas probabilidades sean pequeñas, debemos hacer crecer el tamaño de muestra.*

De los dos errores, el Error I se considera "más grave", por eso su probabilidad α es conocida y controlada. **Esta es la razón por la cual se recomienda escribir la hipótesis de investigación como H_1 y su negación como H_0.** La razón tiene que ver con la posición conservadora de la ciencia: Si una hipótesis de trabajo H es falsa y por error se la acepta, entonces se estaría introduciendo un elemento falso en un cuerpo articulado de conocimientos; por esto se debe tener controlada la posibilidad de cometer este error. Si colocamos H como H_1 y la negación de H se escribe como H_0, la probabilidad de rechazar H_0 siendo ésta verdadera, es conocida y baja (es la probabilidad α) o equivalentemente, es conocida y baja la probabilidad de aceptar H_1 (o sea la hipótesis de trabajo H) siendo falsa.

7.4 Hipótesis Simples vs Hipótesis Simples: Teorema de Neyman-Pearson

Sea $X \sim f_X(x;\theta)$ y supongamos que sólo hay dos valores posibles para θ: $\theta=\theta_0$ o $\theta=\theta_1$ y tenemos que decidir entre una de las posibilidades.
Si escribimos $H_0:\theta=\theta_0$ y $H_1:\theta=\theta_1$, estamos ante el caso de H_0 simple vs H_1 simple.

Supongamos ahora que el tamaño de muestra n es dado y que hemos fijado una probabilidad α conocida para el Error I.

Para esta situación existe un teorema que nos proporciona la región crítica C óptima, aquella que minimiza la probabilidad β de Error II

Teorema de Neyman-Pearson
En el contexto anterior, la Región Crítica Optima C es de la forma:

$C=\left\{(X_1,X_2,...,X_n)\mid \dfrac{L(\theta=\theta_1)}{L(\theta=\theta_0)}\geq k\right\}$ donde k satisface la ecuación

$$P\left(\frac{L(\theta=\theta_1)}{L(\theta=\theta_0)}\geq k \mid H_0 \quad es \quad verdadera\right)=\alpha$$

No demostraremos el teorema (la prueba puede encontrarse en el libro de Larson), pero sí mencionaremos que la idea base es simple:
Recordando que la región crítica debe ser "poco probable" si $H_0:\theta=\theta_0$ es verdadera, en este contexto, esperamos que $L(\theta=\theta_1)<L(\theta=\theta_0)$ o $\dfrac{L(\theta=\theta_1)}{L(\theta=\theta_0)}<1$. Por tanto si el cociente $\dfrac{L(\theta=\theta_1)}{L(\theta=\theta_0)}$ es mayor que 1, tenemos razón para sospechar de la verdad de $H_0:\theta=\theta_0$. El asunto es determinar el "punto de corte" k a partir del cual ya tendríamos que rechazar $H_0:\theta=\theta_0$. Como α es una probabilidad que consideramos suficientemente pequeña, lo natural es tomar k tal que $P\left(\dfrac{L(\theta=\theta_1)}{L(\theta=\theta_0)}\geq k \mid H_0 \quad es \quad verdadera\right)=\alpha$ y de ahí se deduce que la regla de decisión será:
"Si al tomar la muestra, ocurre que $\dfrac{L(\theta=\theta_1)}{L(\theta=\theta_0)}\geq k$ *, entonces rechazaremos* $H_0:\theta=\theta_0$ *y aceptaremos su complemento* H_1*"*

Ejemplo 2
Sea $X \sim Exp(x;\theta)$ y se desea contrastar $H_0:\theta=2$ vs $H_1:\theta=6$ a partir de una muestra grande de tamaño n = 36, de modo que la probabilidad α de Error I sea α=0.05. Determine la Región Crítica Optima *C*

Solución:
Tenemos $H_0:\theta=2$ vs $H_1:\theta=6$ y ambas hipótesis son simples, luego apliquemos el Teorema de Neyman-Pearson a este caso específico: *Si* C *es la R.C. Optima* $\Rightarrow$ C *es de la forma*

$C=\left\{(X_1,X_2,...,X_n)\mid \dfrac{L(\theta=6)}{L(\theta=2)}\geq k\right\}$ *donde* k *es tal que* $P(C\geq k\mid\theta=2)=\alpha=0.05$

Entonces:

(1) $X \sim Exp(x;\theta) \Rightarrow$ En general $L(\theta)=\prod_{j=1}^{n} f_{X_j}(x_j;\theta)=\prod_{j=1}^{n}\theta e^{-\theta x_j}=\theta^n e^{-\theta\sum_{j=1}^{n}x_j}=\theta^n e^{-n\theta\overline{X}}$

(2) El “cociente de Neyman-Pearson” es entonces:

$$\frac{L(\theta=6)}{L(\theta=2)}=\frac{6^n e^{-6n\overline{X}}}{2^n e^{-2n\overline{X}}}=3^n e^{-4n\overline{X}}$$

(3) La Desigualdad de Neyman-Pearson toma la forma:

$$\frac{L(\theta=6)}{L(\theta=2)}\geq k \Leftrightarrow 3^n e^{-4n\overline{X}}\geq k \Leftrightarrow n\ln(3)-4n\overline{X}\geq \ln(k)\Leftrightarrow \overline{X}\leq \frac{n\ln(3)-\ln(k)}{4n}\Leftrightarrow \overline{X}\leq k'$$ donde

$$k'=\frac{n\ln 3-\ln k}{4n}$$

(4) La R.C. Optima es de la forma $C=\{(X_1,X_2,...,X_n)\mid \frac{L(\theta=6)}{L(\theta=2)}\geq k\}=\{\overline{X}\mid \overline{X}\leq k'\}$

(5) Deseamos que $P(Error\quad I)=P(C\mid\theta=2)=\alpha=0.05 \Leftrightarrow P[\overline{X}\leq k'\mid\theta=2\}=0.05$

(6) En este caso n=36 permite aplicar el T.L.C. Como sabemos, $\overline{X}\sim N(\mu,\frac{\sigma^2}{36})$ y como X tiene distribución exponencial con $\mu=1/\theta$ y $\sigma^2=1/\theta^2$, entonces $\overline{X}\sim N(\frac{1}{\theta},\frac{1}{36\theta^2})$ y si $H_0:\theta=2$ es verdadera $\Rightarrow$ $\overline{X}\sim N(\frac{1}{2},\frac{1}{144})$ y así $P[\overline{X}\leq k'\mid\theta=2]=0.05\Leftrightarrow$ $P[Z\leq\frac{k'-0.5}{\sqrt{1/144}}]=P[Z\leq 12(k'-0.5)]=0.05$ y de la tabla Z se obtiene 12(k’-0.5)=-1.645 $\Rightarrow$ k’ = 0.363

(7) Finalmente, la R.C Optima es $C=\{\overline{X}\mid\overline{X}\leq 0.363\}$

Observaciones:

- ß=P[Error II]=P[$\overline{X}$ >0.363 | H_0 es F]=P[$\overline{X}$ >0.363 | H_1 es V]=P[$\overline{X}$ >0.363 | θ=6] ; pero si θ=6 $\Rightarrow$ $\overline{X}$ ~N(1/6,1/1296) por tanto ß=P[Error II]=P[X > 0.363 | θ=6] = P[Z >36(0.363-0.167)]=P[Z > 7.056]=0.00
  Es decir, con esta región crítica, la probabilidad de Error II es prácticamente cero.
- Con este procedimiento sólo hemos construido la Región Crítica Optima, o sea sólo tenemos la Regla de Decisión ***: Si al tomar la m.a. resulta una media muestral*** $\overline{X}$ ***menor o igual que 0.363 => Debemos rechazar*** $H_0:\theta=2$
- Para decidir entre H_0 o H_1 **faltaría** tomar la muestra y evaluar $\overline{X}$, verificando si ocurre o no la región crítica C.
- Por ejemplo, si al tomar la m. a. resultó una $\overline{X}$ =1.2, entonces ***no ocurrió la región crítica C y por tanto no rechazamos*** H_0***, esto es, concluimos que*** $\theta=2$

7.5 Hipótesis Simples vs Hipótesis Compuestas: Pruebas Uniformemente Más Poderosas UMP

Sea $X\sim f_X(x;\theta)$ y supongamos que deseamos contrastar $H_0:\theta=\theta_0$ vs $H_1:\theta\in\omega$ donde ω es un conjunto <u>conocido</u> de valores alternativos para θ. Estamos ante el caso de H_0 simple vs H_1 compuesta.

Supongamos ahora que el tamaño de muestra n es dado y que hemos fijado una probabilidad α conocida para el Error I.

Para esta situación no existe un teorema como el de Neyman-Pearson que nos proporcione una región crítica *C* óptima que minimice la probabilidad ß de Error II, pues esto depende de cómo es ω.

Sin embargo, podemos aplicar un artificio y convertir $H_1:\theta\in\omega$ en una hipótesis de apariencia “simple”. En efecto, si escribimos $H_1:\theta\in\omega$ en su forma equivalente $H_1:\theta=\theta_1\wedge\theta_1\in\omega$, aparentemente estamos ante una $H_0:\theta=\theta_0$ simple contra una alterna $H_1:\theta=\theta_1$ ($\theta_1\in\omega$) “pseudo simple” y podemos aplicar el T. de Neyman-Pearson al juego anterior, donde θ_1 es un valor genérico de ω.

Sea $C(\theta_1)$ la región crítica obtenida y supongamos que su construcción es uniformemente válida para **todo** $\boldsymbol{\theta_1\in\omega}$.

En este caso podemos garantizar que C es una región uniformemente óptima (óptima en todas las circunstancias). Este tipo de región crítica se llama Región Crítica Uniformemente Optima o Uniformemente Más Poderosa (UMP)

Definición:
Sea $X \sim f_X(x;\theta)$ y sea el juego de hipótesis $H_0{:}\theta = \theta_0$ vs $H_1{:}\theta \in \omega$ donde θ_0 y ω son conocidos, siendo ω un conjunto conocido de valores alternativos para θ.
Una región crítica C es Uniformemente Optima si dado un Nivel α predeterminado y dado un tamaño de nuestra n, la Región Crítica C satisface el Teorema de Neyman-Pearson para todo $\theta \in \omega$, no dependiendo su forma del valor elegido de $\theta \in \omega$.

Observación:
(1) Si C es UMP entonces P[Error I]$\leq \alpha\ \forall\ \theta \in \omega$ y ß = P[Error II]= Mínima
(2) No siempre existe la R.C. UMP

Ejemplo 3
Si $X \sim Exp(x;\ \theta)$ y se desea contrastar $H_0{:}\theta=\theta_0$ vs $H_1{:}\theta > \theta_0$ a partir de una muestra grande de tamaño n = 36, de modo que la probabilidad α de Error I sea α = 0.05. ¿Existe una R.C. UMP?

Solución:
(1) Transformemos el juego $H_0{:}\theta = \theta_0$ vs $H_1{:}\theta > \theta_0$ en uno de hipótesis simple H_0 vs hipótesis H_1 pseudo-simple: $H_0{:}\theta = \theta_0$ vs $H_1{:}\theta = \theta_1$ ($\theta_1 > \theta_0$). Apliquemos el T. De Neyman-Pearson a este caso específico:

Si C es la R.C. Optima $\Rightarrow$ C *es de la forma* $C = \left\{ (X_1, X_2, ..., X_n) \mid \dfrac{L(\theta_1)}{L(\theta_0)} \geq k \right\}$ donde k es tal que

$$P\left(\frac{L(\theta = \theta_1)}{L(\theta = \theta_0)} \geq k \mid \theta = \theta_0 \right) = \alpha = 0.05$$

Entonces:

(2) Como $X \sim Exp(x;\theta) \Rightarrow$ En general $L(\theta) = \prod_{j=1}^{n} f_{X_j}(x_j;\theta) = \prod_{j=1}^{n} \theta e^{-\theta x_j} = \theta^n e^{-\theta \sum_{j=1}^{n} x_j} = \theta^n e^{-n\theta \overline{X}}$

(3) El cociente de Neyman-Pearson es: $\dfrac{L(\theta_1)}{L(\theta_0)} = \dfrac{\theta_1^n e^{-\theta_1 n \overline{X}}}{\theta_0^n e^{-\theta_0 n \overline{X}}} = \left(\dfrac{\theta_1}{\theta_0} \right)^n e^{-n(\theta_1 - \theta_0)\overline{X}}$

(4) La Desigualdad de Neyman-Pearson toma la forma:

$$\frac{L(\theta_1)}{L(\theta_0)} \geq k \Leftrightarrow \left(\frac{\theta_1}{\theta_0} \right)^n e^{-n(\theta_1 - \theta_0)\overline{X}} \geq k \Leftrightarrow n \ln(\frac{\theta_1}{\theta_0}) - n(\theta_1 - \theta_0)\overline{X} \geq \ln k .$$

Pero por la condición en la hipótesis H_1 se tiene $(\theta_1 - \theta_0) > 0$. Entonces:

$$\frac{L(\theta_1)}{L(\theta_0)} \geq k \Leftrightarrow n \ln(\frac{\theta_1}{\theta_0}) - n(\theta_1 - \theta_0)\overline{X} \geq \ln k \Leftrightarrow \overline{X} \leq \frac{n \ln(\frac{\theta_1}{\theta_0}) - \ln k}{n(\theta_1 - \theta_0)} \Leftrightarrow \overline{X} \leq k' \text{ donde } k' = \frac{n \ln(\frac{\theta_1}{\theta_0}) - \ln k}{n(\theta_1 - \theta_0)}$$

(5) La R.C. Optima es de la forma $C = \{(X_1, X_2, ..., X_n) \mid \dfrac{L(\theta_1)}{L(\theta_0)} \geq k\} = \{\overline{X} \mid \overline{X} \leq k'\} \qquad \forall \theta = \theta_1 > \theta_0$

(6) Deseamos que P[Error I]=P[$C \mid H_0{:}\theta=\theta_0$ es V]=α=0.05 $\Leftrightarrow P[\overline{X} \leq k' \mid \theta = \theta_0] = 0.05$

En este caso n=36 es "grande" y permite aplicar el T.L.C para afirmar que en general $\overline{X} \sim N(\mu, \frac{\sigma^2}{36})$. Como X tiene distribución exponencial, entonces $\mu = 1/\theta$ y $\sigma^2 = 1/\theta^2$, de modo que $\overline{X} \sim N(\frac{1}{\theta}, \frac{1}{36\theta^2})$ **. Ahora bien, si**

***H*₀:*θ*=*θ*₀ es Verdadera** $\Rightarrow \overline{X} \sim N(\frac{1}{\theta_0}, \frac{1}{36\theta_0^2})$ y $P[\overline{X} \le k' \mid \theta = \theta_0] = 0.05 \Leftrightarrow P[Z \le \frac{k' - \frac{1}{\theta_0}}{\frac{1}{6\theta_0}}] = 0.05 \Leftrightarrow$

$P[Z \le 6\theta_0(k' - \frac{1}{\theta_0})] = 0.05$ y de la tabla Z se obtiene que $6\theta_0(k' - \frac{1}{\theta_0}) = -1.645 \Rightarrow k' = \frac{1}{\theta_0} - \frac{1.645}{6\theta_0}$ y como esta R.C. no depende del valor particular $\theta_1 > \theta_0$ que hayamos tomado, podemos decir que hemos encontrado una R.C. UMP que es $C = \{\overline{X} \mid \overline{X} \le \frac{1}{\theta_0} - \frac{1.645}{6\theta_0}\}$

Observaciones:

- En este caso, la Regla de Decisión sería:

Si al tomar la m.a. resulta que $\overline{X} \le \frac{1}{\theta_0} - \frac{1.645}{6\theta_0}$**, entonces debemos rechazar** $H_0: \theta = \theta_0$ **y aceptar** $H_1: \theta > \theta_0$**.**

- Como en el ejemplo anterior, hasta ahora sólo se ha construido la R.C. U.M.P con la regla de decisión asociada. Faltaría tomar la m.a. y obtener $\overline{X}$ para completar el proceso.

Ejemplo 4

Asuma que el número de horas que trabaja un informal es una v.a. X con distribución N(μ,9). Se tenía la hipótesis que $\mu > 8$ y para evaluarla se tomó una m.a. de n=25 ambulantes y se obtuvo una media muestral de 12.5 horas/día. Construya una Región Crítica Optima, aplicando el Teorema de Neyman-Pearson a $H_0: \mu = 8$ vs $H_1: \mu = \mu_1$ donde $\mu_1 > 8$, usando un nivel $\alpha = 0.05$, y úsela para ver si es cierto que $\mu > 8$.

Solución:

Tenemos $H_0: \mu = 8$ vs $H_1: \mu = \mu_1$ donde $\mu_1 > 8$. Con este formato ambas hipótesis son simples (en realidad $H_1: \mu = \mu_1$, $\mu_1 > 8$ es "pseudo simple"), y podemos aplicar el Teorema de Neyman-Pearson a este caso específico:

Si C ***es la R.C. Optima*** $\Rightarrow C$ ***es de la forma*** $C = \left\{(X_1, X_2, \cdots X_{n1}) \mid \frac{L(\mu = \mu_1)}{L(\mu = 8)} \ge k\right\}$ ***donde k es tal que***

$P(C \mid \mu = 8) = \alpha = 0.05$

Entonces:

(1) $X \sim N(\mu, 9) \Rightarrow$ En general $L(\mu) = \prod_{j=1}^{n} f_{X_j}(x_j; \mu) = \prod_{j=1}^{n} \frac{e^{-(x_j - \mu)^2/2\sigma^2}}{(\sqrt{2\pi})\sigma} = \frac{e^{-\sum_{j=1}^{n}(x_j - \mu)^2/2\sigma^2}}{(\sqrt{2\pi})^n \sigma^n}$

(2) El cociente de Neyman-Pearson es entonces:

$$\frac{L(\mu_1)}{L(8)} = \frac{e^{-\sum_{j=1}^{n}(x_j - \mu_1)^2/2\sigma^2}}{e^{-\sum_{j=1}^{n}(x_j - 8)^2/2\sigma^2}} = e^{\sum_{j=1}^{n}\left[(x_j - 8)^2 - (x_j - \mu_1)^2\right]/2\sigma^2} = e^{\sum_{j=1}^{n}\left[(x_j - 8)^2 - (x_j - \mu_1)^2\right]/18}$$, pues $\sigma^2 = 9$

(3) La Desigualdad de Neyman-Pearson toma la forma:

$$\frac{L(\mu_1)}{L(8)} \ge k \Leftrightarrow e^{\sum_{j=1}^{n}\left[(x_j - 8)^2 - (x_j - \mu_1)^2\right]/18} \ge k \Leftrightarrow \sum_{j=1}^{n}\left[(x_j - 8)^2 - (x_j - \mu_1)^2\right]/18 \ge \ln(k) \Leftrightarrow \overline{X} \ge k'$$ pues:

$$\sum_{j=1}^{n}\left[(x_j-8)^2-(x_j-\mu_1)^2\right]\geq 18\ln(k) \Leftrightarrow \sum_{j=1}^{n}\left[(2x_j-8-\mu_1)(\mu_1-8)\right]\geq 18\ln(k) \Leftrightarrow$$

$$(\mu_1-8)\sum_{j=1}^{n}\left[2x_j-8-\mu_1\right]\geq 18\ln(k) \Leftrightarrow (\mu_1-8)\left[2\sum_{j=1}^{n}x_j-8n-n\mu_1\right]\geq 18\ln(k)$$, como según la restricción impuesta en $H_1:\mu=\mu_1$, $\mu_1>8$, entonces $(\mu_1-8)>0$ y al pasar (μ_1-8) al lado izquierdo de la inecuación, ésta no se altera. Luego:

$$(\mu_1-8)\left[2\sum_{j=1}^{n}x_j-8n-n\mu_1\right]\geq 18\ln(k) \Leftrightarrow \left[2\sum_{j=1}^{n}x_j-8n-n\mu_1\right]\geq \frac{18\ln(k)}{(\mu_1-8)} \Leftrightarrow 2n\overline{X}-8n-n\mu_1\geq \frac{18\ln(k)}{(\mu_1-8)} \Leftrightarrow$$

$$2n\overline{X}\geq \frac{18\ln(k)}{(\mu_1-8)}+8n+n\mu_1 \Leftrightarrow 2n\overline{X}\geq \frac{18\ln(k)+n(\mu_1-8)((\mu_1+8)}{(\mu_1-8)} \Leftrightarrow \overline{X}\geq \frac{18\ln(k)+n(\mu_1-8)((\mu_1+8)}{2n(\mu_1-8)}$$

$$\Leftrightarrow \overline{X}\geq k' \text{ donde } k'=\frac{18\ln(k)+n(\mu_1-8)((\mu_1+8)}{2n(\mu_1-8)}$$

(4) Entonces la R.C. Optima es de la forma $C=\{(X_1,X_2,...,X_n)\mid \frac{L(\mu_1)}{L(8)}\geq k\}=\{\overline{X}\mid \overline{X}\geq k'\}$

(5) Deseamos que $P(Error\ I)=P(C\mid H_0:\mu=8\ es\ verdadera)=\alpha=0.05 \Leftrightarrow P\left(\overline{X}\geq k'\mid \mu=8\right)=0.05$.

En este caso, la propiedad reproductiva de la distribución normal permite escribir $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$ y si H_0: $\mu=8$ es Verdadera $\Rightarrow$ $\overline{X}\sim N(8,\frac{9}{25})$ y $P[\overline{X}\geq k'\mid \mu=8]=0.05 \Leftrightarrow P[Z\geq \frac{k'-8}{3/5}]=0.05$.

De la tabla Z se obtiene $\frac{k'-8}{3/5}=1.645 \Rightarrow k'=8+0.987$ y por tanto la R.C Optima es $C=\{\overline{X}\mid \overline{X}\geq 8.987\}$, que no depende del valor $\mu=\mu_1$ que hayamos puesto en la hipótesis $H_1:\mu=\mu_1$, esto quiere decir que la Región Crítica hallada es óptima para todo $\mu_1>8$, o sea que es uniformemente óptima o UMP.

Finalmente, la regla de decisión es:
Si al tomar la m.a. resulta una media muestral $\overline{X}$ superior a 8.987, tendremos que rechazar $H_0:\mu=8$, y al hacer esto tendremos un 5% de probabilidad de equivocarnos.

Ahora bien, como la media muestral resultó $\overline{X}=12.5>8.987$, tenemos evidencia suficiente para rechazar $H_0:\mu=8$. Concluimos entonces que $\mu>8$.

Ejemplo 5
En un modelo de finanzas, se sabe que la rentabilidad de una acción (expresada en porcentaje) es una v.a.c. $X\sim N(0,\sigma^2=4)$. Un analista piensa que la estabilización de la economía ocasionará un menor riesgo y por tanto disminuirá el valor de σ^2. Para examinar su idea, se plantea $H_0:\sigma^2=4$ vs $H_1:\sigma^2<4$. Una muestra de n=8 días dio las siguientes rentabilidades: 0, 2.1, 2.4, 3, 3.4, 1, 3, 2.

a) Construir un I.C. de 95% para σ^2 y usarlo para decidir si $\sigma^2=4$ o si $\sigma^2<4$
b) Halle la R.C. UMP para la hipótesis $H_0:\sigma^2=4$ vs $H_1:\sigma^2<4$ y aplíquela para ver si ha disminuido el riesgo. Use un nivel $\alpha=0.05$

Solución:
a) Un I.C. de 95% para σ^2 requiere una variable base (puede haber más de una) con distribución conocida: Hay varias alternativas, una primera sería

- Basarse en que $X \sim N(0,\sigma^2) \Rightarrow \bar{X} \sim N(0,\frac{\sigma^2}{n})$ y usar como variable base a $Z = \frac{\bar{X}}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$ cuya distribución está tabulada y no depende de σ^2. Se plantea $0.95 = P\left(a \le Z = \frac{\bar{X}}{\sqrt{\frac{\sigma^2}{n}}} \le b\right) \Rightarrow P(Z \le a) = 0.025$ y $P(b < Z) = 0.025 \Rightarrow$ de la tabla normal $a = -1.96$ y $b = 1.96$ de modo que $0.95 = P\left(-1.96 \le \frac{\bar{X}}{\sqrt{\frac{\sigma^2}{n}}} \le 1.96\right)$ y de ahí "despejar" σ^2. Pero este sistema tiene la dificultad adicional de que para despejar σ^2, hay que dividir por $\bar{X}$ y las desigualdades pueden mantenerse o invertirse, dependiendo de si $\bar{X} > 0$ o si $\bar{X} < 0$. **No** es un sistema práctico

Otra alternativa es apoyarse en el hecho conocido que la variable $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(k = n-1 = 7)$.

- Planteamos $0.95 = P\left(a \le W = \frac{(n-1)S^2}{\sigma^2} \le b\right)$ con $\chi^2(k=7)$ y tomando $0.025 = P(W \le a) = P(W \ge b)$, resultan

	Tabla de la distribución Ji cuadrado con k grados de libertad									
	W tiene distribución Ji-cuadrado con k grados del libertad y la tabla da c tal que P(W < c) = p									
	p = Probabilidad acumulada									
k	**0.005**	**0.01**	**0.025**	**0.05**	**0.1**	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955

$a = 1.690$ y $b = 16.013$; entonces $0.95 = P\left(1.69 \le \frac{(n-1)S^2}{\sigma^2} \le 16.013\right)$

- Despejando σ^2
$P\left(\underbrace{\frac{(n-1)S^2}{16.013}}_{L1} \le \sigma^2 \le \underbrace{\frac{(n-1)S^2}{1.69}}_{L2}\right)$ y así el I.C. para σ^2 es: $\underbrace{\frac{(n-1)S^2}{16.013}}_{L1} \le \sigma^2 \le \underbrace{\frac{(n-1)S^2}{1.69}}_{L2}$

- Aplicando el I.C. a la muestra de datos:
$S^2 = \frac{\sum_{i=1}^{8} X_i^2 - n\bar{X}^2}{(n-1)} = \frac{44.73 - 8\times(2.11)^2}{7} = \frac{9.1132}{7} = 1.31$ y $\underbrace{\frac{9.1132}{16.013}}_{L1} \le \sigma^2 \le \underbrace{\frac{9.1132}{1.69}}_{L2}$ o $0.57 \le \sigma^2 \le 5.40$ con 95% de confianza: **No se puede concluir nada acerca de si $\sigma^2 = 4$ o si $\sigma^2 < 4$.**

Una tercera alternativa es apoyarse en que $X \sim N(0,\sigma^2) \Rightarrow X_i \sim N(0,\sigma^2) \Rightarrow Z_i = \frac{X_i}{\sqrt{\sigma^2}} \sim N(0,) \Rightarrow Z_i^2 = \frac{X_i^2}{\sigma^2} \sim \chi^2(k=1) \Rightarrow \sum_{i=1}^{n} \frac{X_i^2}{\sigma^2} \sim \chi^2(k = n = 8)$: La variable base sería $V = \sum_{i=1}^{8} \frac{X_i^2}{\sigma^2}$

- Planteamos $0.95 = P\left(a \le V = \sum_{i=1}^{8} \frac{X_i^2}{\sigma^2} \le b\right)$ con $\chi^2(k=8)$ y tomando $0.025 = P(W \le a) = P(W \ge b)$, resultan $a = 2.180$ y $b = 17.535$

	Tabla de la distribución Ji cuadrado con k grados de libertad									
	W tiene distribución Ji-cuadrado con k grados del libertad y la tabla da c tal que P(W < c) = p									
	p = Probabilidad acumulada									
k	**0.005**	**0.01**	**0.025**	**0.05**	**0.1**	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	**2.180**	2.733	3.490	13.362	15.507	**17.535**	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589

entonces $0.95 = P\left(2.180 \le \sum_{i=1}^{8} \frac{X_i^2}{\sigma^2} \le 17.535\right)$

- Despejando σ^2

  $0.95 = P\left(\underbrace{\frac{\sum_{i=1}^{8} X_i^2}{17.535}}_{L1} \le \sigma^2 \le \underbrace{\frac{\sum_{i=1}^{8} X_i^2}{2.180}}_{L2}\right)$ y así el I.C. para σ^2 es: $\underbrace{\frac{\sum_{i=1}^{8} X_i^2}{17.535}}_{L1} \le \sigma^2 \le \underbrace{\frac{\sum_{i=1}^{8} X_i^2}{2.180}}_{L2}$

- Aplicando el I.C. a la muestra de datos, como $\sum_{i=1}^{8} X_i^2 = 44.73$ entonces: $\underbrace{\frac{44.73}{17.535}}_{L1} \le \sigma^2 \le \underbrace{\frac{44.73}{2.180}}_{L2}$ o

  $2.55 \le \sigma^2 \le 20.52$ con 95% de confianza: **No se puede concluir nada acerca de si $\sigma^2 = 4$ o si $\sigma^2 < 4$.**

Nótese que los dos I.C., construidos con distintas variables base, son también distintos. El preferible es el de menor longitud, el menos ancho.

b) Halle la R.C. UMP para la hipótesis $H_0:\sigma^2 = 4$ vs $H_1:\sigma^2 < 4$ y aplíquela para ver si ha disminuido el riesgo. Use un nivel $\alpha=0.05$

Se desea contrastar la hipótesis $H_0:\sigma^2 = 4$ vs $H_1:\sigma^2 < 4$ con una probabilidad $\alpha = 0.05$ de Error I, pasamos a $H_0:\sigma^2 = 4$ vs $H_1:\sigma^2 = \sigma_1^2\ \ (\sigma_1^2 < 4)$. Con este formato ambas hipótesis son simples (en realidad $H_1:\sigma^2 = \sigma_1^2\ \ (\sigma_1^2 < 4)$ es "pseudo simple"), y podemos aplicar el Teorema de Neyman-Pearson a este caso específico: *Si C es la R.C. Optima $\Rightarrow C$ es de la forma $C = \left\{(X_1, X_2, \cdots X_n) \mid \frac{L(\sigma^2 = \sigma_1^2)}{L(\sigma^2 = 4)} \ge k\right\}$ donde k es tal que $P(C \mid \sigma^2 = 4) = \alpha = 0.05$*

Entonces:

(1) $X \sim N(0,\sigma^2) \Rightarrow$ En general $L(\sigma^2) = \prod_{j=1}^{n} \frac{e^{-x_j^2/2\sigma^2}}{(\sqrt{2\pi})\sigma} = \frac{e^{-\sum_{j=1}^{n} x_j^2/2\sigma^2}}{(\sqrt{2\pi})^n \sigma^n} = \frac{e^{-\sum_{j=1}^{n} x_j^2/2\sigma^2}}{(\sqrt{2\pi})^n (\sigma^2)^{n/2}}$

(2) El cociente de Neyman-Pearson es $\frac{L(\sigma^2 = \sigma_1^2)}{L(\sigma^2 = 4)} = \frac{\dfrac{e^{-\sum_{j=1}^{n} x_j^2/2\sigma_1^2}}{(\sqrt{2\pi})^n (\sigma_1^2)^{n/2}}}{\dfrac{e^{-\sum_{j=1}^{n} x_j^2/2\times 4}}{(\sqrt{2\pi})^n (4)^{n/2}}} = \left(\frac{4}{\sigma_1^2}\right)^{n/2} e^{-\sum_{j=1}^{n} x_j^2\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\times 4}\right)}$

(3) La Desigualdad de Neyman-Pearson toma la forma:

$$\frac{L(\sigma^2=\sigma_1^2)}{L(\sigma^2=4)} \geq k \Leftrightarrow \left(\frac{4}{\sigma_1^2}\right)^{n/2} e^{-\sum_{j=1}^{n} x_j^2\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\times 4}\right)} \geq k \Leftrightarrow \frac{n}{2}\ln\left(\frac{4}{\sigma_1^2}\right) - \sum_{j=1}^{n} x_j^2 \overbrace{\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\times 4}\right)}^{positivo} \geq \ln k$$

pues: $\sigma_1^2 < 4 \Leftrightarrow \frac{1}{\sigma_1^2} > \frac{1}{4} \Leftrightarrow \frac{1}{2\sigma_1^2} - \frac{1}{2\times 4} > 0$, luego

$$\frac{L(\sigma^2=\sigma_1^2)}{L(\sigma^2=4)} \geq k \Leftrightarrow -\sum_{j=1}^{n} x_j^2 \geq \frac{\ln k - \frac{n}{2}\ln\left(\frac{4}{\sigma_1^2}\right)}{\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\times 4}\right)} \Leftrightarrow \sum_{j=1}^{n} x_j^2 \leq \frac{\frac{n}{2}\ln\left(\frac{4}{\sigma_1^2}\right) - \ln k}{\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\times 4}\right)} = k'$$

(4) Entonces la R.C. Optima es de la forma

$$C = \{(X_1, X_2, ..., X_n) \mid \frac{L(\sigma^2=\sigma_1^2)}{L(\sigma^2=4)} \geq k\} = \{(X_1, X_2, ..., X_n) \mid \sum_{j=1}^{n} x_j^2 \leq k'\}$$

(5) De $P(Error\, I) = P(C \mid H_0 : \sigma^2 = 4 \text{ } es\ verdadera) = \alpha = 0.05 \Leftrightarrow P\left(\sum_{j=1}^{n} x_j^2 \leq k' \mid \sigma^2 = 4\right) = 0.05$.

En este caso, de $X_j \sim N(0, \sigma^2)$ y si $H_0 : \sigma^2 = 4$ es verdadera $\Rightarrow Z_j = (\frac{X_j - 0}{\sqrt{4}}) \sim N(0,1)$ y $Z_j^2 = \frac{x_j^2}{4} \sim \chi^2(1)$

$$\Rightarrow \sum_{j=1}^{n} Z_j^2 = \sum_{j=1}^{n} \frac{x_j^2}{4} \sim \chi^2(n) \Rightarrow P\left(\sum_{j=1}^{n} x_j^2 \leq k' \mid \sigma^2 = 4\right) = 0.05 \Leftrightarrow P[\sum_{j=1}^{n} Z_j^2 = \sum_{j=1}^{n} \frac{x_j^2}{4} \leq \frac{k'}{4}] = 0.05.$$

Como $n=8 \Rightarrow \sum_{j=1}^{8} \frac{x_j^2}{4} \sim \chi^2(8)$ y de $P[\sum_{j=1}^{8} \frac{x_j^2}{4} \leq \frac{k'}{4}] = 0.05$ se tiene

	Tabla de la distribución Ji cuadrado con k grados de libertad									
	W tiene distribución Ji-cuadrado con k grados del libertad y la tabla da c tal que P(W < c) = p									
	p = Probabilidad acumulada									
k	**0.005**	**0.01**	**0.025**	**0.05**	**0.1**	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	**1.690**	2.167	[illegible]	[illegible]	14.067	**16.013**	18.475	20.278
8	1.344	1.647	**2.180**	2.733	[illegible]	[illegible]	15.507	**17.535**	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	**19.023**	21.666	23.589

k'/4=2.733

$$\frac{k'}{4} = \chi^2_{0.05}(8) = 2.733 \Rightarrow k' = 10.92$$

La R.C Optima es $C = \{(X_1, X_2, ..., X_n) \mid \sum_{j=1}^{n} x_j^2 \leq k'\} = \{(X_1, X_2, ..., X_n) \mid \sum_{j=1}^{n} x_j^2 \leq 10.92\}$, que no depende del valor σ_1^2 que hayamos puesto y esto quiere decir que la Región Crítica hallada es óptima para todo $\sigma_1^2 < 4$, o sea que es uniformemente óptima o UMP.

Aplicando los datos a la R.C. tenemos $\sum_{j=1}^{n} x_j^2 = 44.73 > 10.92$, por tanto no rechazamos $H_0 : \sigma^2 = 4$, la varianza no ha disminuido.

Nótese como con I.C. no obtuvimos respuesta concreta, e incluso había más de un I.C.; en cambio, con Contraste de Hipótesis, sólo hay una R.C. C óptima.

7.6 Test de Razón De Verosimilitud

Este sistema de generar regiones críticas es bastante general y se aplica cuando no podemos generar una región crítica UMP.

- Sea X ~ f(x;θ) donde tenemos la hipótesis nula H_0: $\theta \in W_0$ vs la hipótesis alterna H_1: $\theta \in W^c_0$ (donde W^c_0 es el complemento de W_0), siendo por lo general ambas compuestas, aunque no es obligatorio que sea así.
- Deseamos contruir una región crítica ***C*** que permita contrastar H_0 vs H_1 de manera razonable.
- Sea ***W*** el espacio de valores posibles para θ ("espacio paramétrico" le llaman). Naturalmente ***W*** = $W_0 \cup W^c_0$ y H_0 nos dice que en verdad, ***W*** se "reduce" a W_0, o en otras palabras, si extraemos una muestra aleatoria $(X_1,X_2,...X_n)$, **un θ en W_0 le dará MAYOR probabilidad a la m.a. que un valor θ en W^c_0**
- Asumamos que H_0 es cierta: Si aplicamos Máxima Verosimilitud para estimar θ, tenemos que "respetar" H_0: $\theta \in W_0$, de modo que tenemos que maximizar **L(θ)** imponiendo la condición adicional (sujetos a la restriccion) $\theta \in W_0$. Sea $\hat{\theta}_0$ este estimador MV de **θ** bajo H_0. Entonces la probabilidad de la m.a. bajo H_0 es, naturalmente

$$L(\hat{\theta}_0) = \prod_{j=1}^{n} f_X(x_j;\hat{\theta}_0)$$

- Por otra parte, si H_0 no es cierta, el correcto estimador MV de **θ** se obtiene con el procedimiento usual de maximizar L(θ) buscando sobre todo el espacio paramétrico ***W*** y ya no sólo sobre W_0. Sea $\hat{\theta}$ este estimador MV de **θ. En este contexto,** la probabilidad de la m.a. es $L(\hat{\theta}) = \prod_{j=1}^{n} f_X(x_j;\hat{\theta})$
- Como W_0 es subconjunto de ***W*** , inevitablemente ocurre que $L(\hat{\theta}_0) \le L(\hat{\theta})$ y por tanto $\frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \le 1$, y como en general **L(θ) es positiva. se tiene que el cociente o "razón de verosimilitud"** $\frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$ satisface $0 \le \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \le 1$. Este cociente se suele denotar $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$ ("lambda")
- Si H_0 es cierta y ***W*** en realidad es W_0, debe ocurrir que $L(\hat{\theta}_0) = L(\hat{\theta})$ y por tanto esperamos que $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = 1$.
- En cambio, si H_0 es falsa esperamos que estrictamente se cumpla que $L(\hat{\theta}_0) < L(\hat{\theta})$ o equivalentemente que $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} < 1$ y cuanto "más falsa" sea H_0 menor debiera ser $L(\hat{\theta}_0)$ en relación a $L(\hat{\theta})$, esto es esperamos que $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$ sea cercano a 0
- Por lo anterior, un valor de $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$ "muy pequeño" **debe ser motivo para rechazar H_0**
- **¿Cuándo considerar pequeño a λ**? : Cuando caiga debajo de cierto punto de corte λ_0, donde λ_0 se escoge de modo que $P(\lambda \le \lambda_0)$ tenga un valor α predeterminado y considerada pequeño (usualmente 0.05).
- **Determinado el valor** λ_0 se tiene la región crítica ***C*** buscada desde el principio, esto es

$$C = \{\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \mid \lambda \le \lambda_0\}$$

- La región crítica ***C*** se llama de Razón de Verosimilitud y el Test correspondiente se llama "Test de Razón de Verosimilitud"
- Se puede desmostrar que si n es "grande" y H_0 es cierta, entonces $-2\ln(\lambda) \sim \chi^2(r)$ donde **r** es el número de parámetros especificado por H_0

Ejemplo 6
Si X ~ Exp(x; θ) y se desea contrastar $H_0:\theta = \theta_0$ vs $H_1:\theta \neq \theta_0$ a partir de una muestra grande de tamaño n = 36, de modo que la probabilidad α de Error I sea α = 0.05. Halle la Región Crítica de Razón de Verosimilitud.

Solución:
Para este problema no existe una R.C. UMP, así que veamos qué sucede si aplicamos la Razón de verosimilitud.

Aquí $H_0:\theta = \theta_0$ vs $H_1:\theta \neq \theta \Leftrightarrow H_0:\theta \in W_0=\{\theta_0\}$ vs $H_1:\theta \in W^C_0=\{\theta|0<\theta\neq\theta_0\}$

Ya sabemos que $\mathbf{L(\theta)} = \prod_{j=1}^{n} \mathbf{f_X(x_j;\theta)} = \prod_{j=1}^{n} \boldsymbol{\theta e^{-\theta x_j}} = \boldsymbol{\theta^n e^{-\theta\sum_{j=1}^{n} x_j}} = \boldsymbol{\theta^n e^{-n\theta\bar{X}}}$ y por tanto:

- Bajo $H_0:\theta \in W_0=\{\theta_0\}$, el máximo de $\mathbf{L(\theta)}$ sobre W_0, es $L(\theta_0) = \theta_0^n e^{-n\theta_0\bar{X}}$ pues aquí no hay nada que estimar, dado que θ_0 ya es conocido (o mejor dicho, ya es especificado por H_0).
- Bajo $H_1:\theta \in W^C_0=\{\theta|0<\theta\neq\theta_0\}$, el máximo de $\mathbf{L(\theta)}$ sobre W^C_0 es el que se obtiene reemplazando el estimador MV de θ $\hat{\theta}_{MV}$ en $L(\theta)$. Ahora bien, si $\theta \neq \theta_0$ entonces, $\hat{\theta}_{MV} = \frac{1}{\bar{X}}$ y
  $L(\hat{\theta}_{MV}) = L(\frac{1}{\bar{X}}) = (\frac{1}{\bar{X}})^{\mathbf{n}} \mathbf{e}^{-\mathbf{n}(\frac{1}{\bar{X}})\bar{X}} = (\frac{1}{\bar{X}})^{\mathbf{n}} \mathbf{e}^{-\mathbf{n}}$
- La razón de verosimilitud es $\lambda = \frac{\mathbf{L(\hat{\theta}_0)}}{\mathbf{L(\hat{\theta})}} = \frac{\theta_0^n e^{-n\theta_0\bar{X}}}{(\frac{1}{\bar{X}})^{\mathbf{n}}\mathbf{e}^{-\mathbf{n}}} = \bar{X}^n \theta_0^n e^{-n(\theta_0\bar{X}-1)}$ y por tanto la Región Crítica de Razón de Verosimilitud es:
  $C = \{\lambda = \frac{\mathbf{L(\hat{\theta}_0)}}{\mathbf{L(\hat{\theta})}} \mid \lambda \leq \lambda_0\} = \{\lambda \mid \lambda = \bar{X}^n \theta_0^n e^{-n(\theta_0\bar{X}-1)} < \lambda_0\} = \{\bar{X} \mid \bar{X}^n \theta_0^n e^{-n(\theta_0\bar{X}-1)} < \lambda_0\}$, donde λ_0 debe calcularse de modo que $P(\lambda < \lambda_0 \mid \theta = \theta_0) = 0.05$
- La distribución de λ no es sencilla, pero como n es grande, podemos usar la propiedad que $-2\ln(\lambda) \sim \chi^2(r=1)$ (pues sólo hay un parámetro). O sea:
  $0.05 = P(\lambda < \lambda_0) = P(\ln\lambda < \ln\lambda_0) = P(-2\ln\lambda > -2\ln\lambda_0) = P(\chi^2(1) > -2\ln\lambda_0) \Rightarrow -2\ln\lambda_0 = 3.84$
- Así la R.C. de Razón de Verosimilitud es
  $C = \{\lambda \mid \lambda < \lambda_0\} = \{\lambda \mid -2\ln\lambda > 3.84\} = \{\bar{X} \mid -2\ln\left(\bar{X}^n \theta_0^n e^{-n(\theta_0\bar{X}-1)}\right) > 3.84\}$ y si al tomar la muestra, calcular $\bar{X}$ y evaluar $-2\ln\left(\bar{X}^n \theta_0^n e^{-n(\theta_0\bar{X}-1)}\right)$, esta última cantidad resulta mayor que 3.84, podemos rechazar $H_0:\theta = \theta_0$

Ejemplo 7
En el ejemplo anterior si las hipótesis fueran $H_0: \theta = \frac{1}{2}$ vs $H_1: \theta \neq \frac{1}{2}$ y en una muestra de $n = 36$ casos ocurrió que $\bar{X} = 1.5$, al nivel $\alpha = 0.05$ de probabilidad de error: ¿Rechaza o no rechaza $H_0: \theta = \frac{1}{2}$?

Solución:
Aplicando la R.C. de Razón de Verosimilitud del ejemplo anterior, justo calculada para un nivel $\alpha = 0.05$

$$C=\{\lambda \mid \lambda<\lambda_0\}=\{\lambda \mid -2\ln\lambda>3.84\}=\{\overline{X} \mid -2\ln\left(\overline{X}^{\,n}\theta_0^n e^{-n(\theta_0\overline{X}-1)}\right)>3.84\}=$$
$$\{\overline{X} \mid -2\ln\left(\overline{X}^{\,36}(0.5)^{36}e^{-36(0.5\overline{X}-1)}\right)>3.84\}$$

Y como $\bar{X}=0.85$, reemplazamos y obtenemos $\lambda=-2\ln\left((1.5)^{36}(0.5)^{36}e^{-36(0.5\times1.5-1)}\right)=2.713<3.84$ este valor de λ **no permite** rechazar $H_0\colon\theta=\frac{1}{2}$, por tanto la aceptamos.