Capítulo 10 Pruebas de Hipótesis

10.1 Introducción

Recuerde que uno de los objetivos de la estadística es hacer inferencia acerca de parámetros poblaciones desconocidos con base a información contenida en datos muestrales. Estas inferencias se interpretan de dos formas, como estimaciones de los parámetros respectivos o como pruebas de hipótesis acerca de sus valores.

En este capítulo se revisarán las pruebas de hipótesis. En muchos aspectos el procedimientos formal para pruebas de hipótesis es similar al método científico. Este observa la naturaleza, formula una teoría y la confronta con lo observado. En nuestro contexto el científico plantea una hipótesis respecto a uno o más parámetros poblacionales. En seguida toma una muestra de la población y compara sus observaciones con la hipótesis. Si las observaciones no concuerdan con la hipótesis, las rechaza. De lo contrario, concluye que la hipótesis es verdadera o que la muestra no detectó la diferencia entre los valores real e hipotético de los parámetros poblacionales.

10.2 Elementos de una prueba estadística

Los elementos de una prueba estadística

1. Hipótesis nula, H_0
2. Hipótesis alternativa, H_a
3. Estadístico de prueba
4. Región de rechazo

Como se menciono brevemente en la sección anterior, muchas veces, el objetivo de una prueba estadística es probar una hipótesis concerniente a los valores de uno o más parámetros poblacionales. Por lo general se tiene una teoría, es decir una *hipótesis de investigación*, acerca del o los parámetros que se desea apoyar. Para realizar una prueba de esta hipótesis primero se propone una hipótesis nula de como se cree que se comporta el o los parámetros en cuestión. Esta sería la hipótesis a ser probada o rechazada. De ser rechazada la hipótesis nula se necesita proponer una hipótesis alternativa (o *investigación*). Esta es la hipótesis a ser aceptada en caso de que H_0 sea rechazada. En muchas ocasiones la hipótesis alternativa es la que queremos comprobar con base en la información contenida en la muestra. Esto se le llama una prueba por contradicción.

Las partes esenciales de una prueba estadística son el estadístico de prueba y la región de rechazo asociada. El *estadístico de prueba* (al igual que un estimador) es una función de las mediciones muestrales en las que la decisión estadística estará basada. La *región de rechazo*, denotada por RR, especifica los valores del estadístico de prueba para el cual la hipótesis nula ha de ser *rechazada* a favor de la hipótesis alternativa. En otras palabras, si para una muestra particular, el valor calculado del estadístico de prueba cae en la región de rechazo RR entonces se rechaza la hipótesis nula H_0y aceptamos la hipótesis al ternativa H_a. Si el valor del estadístico de prueba no cae en la RR, aceptamos H_0. Para cualquier región de rechazo fija (determinada por un valor particular de k), dos ditpos de errores se pueden cometer al llegar a una decisión. Estos se muestran en la siguiente definición.

Definición 10.1

Se comete un *error tipo I* si H_0 es rechazada cuando H_0 es verdadera. La *probabilidad de un erro tipo I* está denotada por α. EL valor de α se denomina *nivel* de la prueba.
Se comete un *error tipo II* si H_0 es aceptada cuando H_a es verdadera. La *probabilidad de un error tipo II* está denotada por β-

Entonces, α y β, las probabilidades de cometer estos dos tipos de error, miden los riesgos relacionados con las dos posibles decisiones erróneas que podrían resultar de una prueba estadística. Estos valores se ven afectados por la región de rechazo. Si cambiamos la región de rechazo para aumentar α, entonces β disminuirá. Del mimso modo, si el cambio en la región de rechazo resulta en una disminución en α, entonces β aumentará. Por lo tanto, α y β están relacionados de manera inversa.

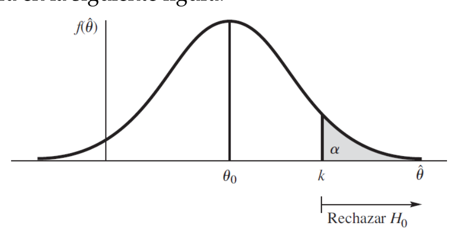
10.3 Pruebas comunes con muestras grandes

Suponga que deseamos probar un conjunto de hipótesis respecto a un parámetro θ con base en una muestra aleatoria $Y_1, Y_2, \dots, Y_n$. Suponemos que el estimador $\hat{\theta}$ tiene distribución normal (aproximadamente) con media θ y error estándar $\theta_{\hat{\theta}}$. Hay tres posibles pruebas que se pueden obtener. Se ejemplificara el desarrollo de una.

Si θ_0 es un valor especifico de θ, podemos probar $H_0 = \theta = \theta_0$ contra $H_a : \theta > \theta_0$. Intuitivamente se puede ver que valores grandes de $\hat{\theta}$ favorecen el rechazo de H_0 y aceptación de la hipótesis alternativa. Esto se puede expresar de la siguiente forma.

$H_0 : \theta = \theta_0$.
$H_a : \theta > \theta_0$.
Estadístico de prueba $\hat{\theta}$.
Región de rechazo: RR $= \{\hat{\theta} > k\}$ para alguna seleción de k.

El valor real de k en la región de rechazo RR se determina al fijar la probabilidad α de error tipo I y escoger k de conformidad. Como se muestra en la siguiente figura.



Entonces si H_0 es verdadera, $\hat{\theta}$ tiene una distribución aproximadamente normal con media θ_0 y error estándar $\sigma_{\hat{\theta}}$ y además se desea una prueba de nivel α se tiene que,

$$k = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$$

es la selección apropiada para k [si Z tiene una distribución normal estándar, entonces z_α es tal que $P(Z > z_\alpha) = \alpha$]. Como

$$\text{RR} = \{\hat{\theta} : \hat{\theta} > \theta_0 + z_\alpha \sigma_{\hat{\theta}}\} = \left\{\hat{\theta} : \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha\right\}$$

si $Z = (\hat{\theta} - \theta_0)/\sigma_{\hat{\theta}}$ se usa como estadístico de prueba, la región de rechazo se puede escribir como RR $= \{z > z_\alpha\}$. Z mide el número de errores estándar entre el estimador para θ y θ_0. Una forma alternativa de la prueba de hipótesis con nivel α, es la siguiente:

$H_0 : \theta = \theta_0$.
$H_a : \theta > \theta_0$.
Estadístico de prueba $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$.
Región de rechazo: RR $= \{z > z_\alpha\}$.

H_0 es rechazada si Z cae suficientemente alejada en la cola superior de la distribución normal estándar. La hipótesis alternativa $H_a; \theta > \theta_0$ se denomina alternativa de *cola superior* y RR $= \{z > z_\alpha\}$ se conoce como *región de rechazo de cola superior*. En el siguiente cuadro se muestran dos pruebas alternativas.

Pruebas de hipótesis de nivel α para muestras grandes

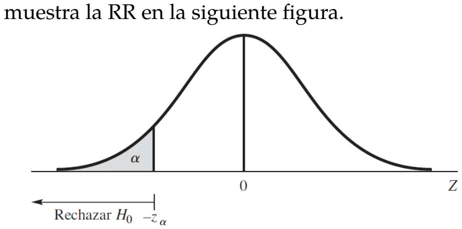
$H_0 : \theta = \theta_0$.

$$H_a : \begin{cases} \theta > \theta_0 & \text{(alternativa de cola superior).} \\ \theta < \theta_0 & \text{(alternativa de cola inferior).} \\ \theta \neq \theta_0 & \text{(alternativa de dos colas).} \end{cases}$$

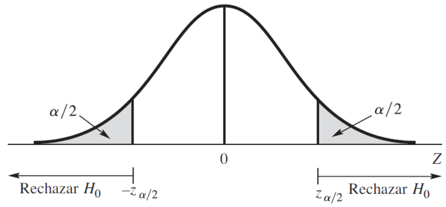
Estadístico de prueba: $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$.

$$\text{Región de rechazo: } \begin{cases} \{z > z_\alpha\} & \text{(RR de cola superior)} \\ \{z < -z_\alpha\} & \text{(RR de cola inferior)} \\ \{|z| > z_{\alpha/2}\} & \text{(RR de dos colas)} \end{cases}$$

La prueba con región de rechazo de la cola interior se puede demostrar de manera análoga al de la cola superior. En este caso se rechaza la H_0 cuando z cae lo suficientemente lejos en la cola inferior. Se muestra la RR en la siguiente figura.



En la prueba de dos colas se rechaza la H_0 si $\hat{\theta}$ es mucho menor o mucho mayor que θ_0. En otras palabras se rechaza H_0 si $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$. Las regiones de rechazo se muestra en la siguiente figura.



10.4 Cálculo de las probabilidades del error tipo II y determinación del tamaño muestral para la prueba Z

El cálculo de β puede ser muy difícil para algunas pruebas estadísticas, pero es fácil hacerlo para las pruebas desarrolladas en la sección anterior. Entonces se puede usar la prueba Z para demostrar el cálculo de β.

Para la prueba $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$, podemos calcular probabilidades de un error tipo II sólo para valores específicos de θ en H_a. Suponga que el experimentador tiene en mente una alternativa específica, por ejemplo $\theta = \theta_a$ (donde $\theta_a > \theta_0$). Como la región de rechazo es de la forma

$$\text{RR} = \{\hat{\theta} : \hat{\theta} > k\}$$

la probabilidad β de un error tipo II es

$$\begin{aligned}\beta &= P(\hat{\theta} \text{ no está en RR cuando } H_a \text{ es verdadera})\\ &= P(\hat{\theta} \leq k \text{ cuando } \theta = \theta_a)\\ &= P\left(\frac{\hat{\theta}-\theta_a}{\sigma_{\hat{\theta}}} \leq \frac{k-\theta_a}{\sigma_{\hat{\theta}}} \text{ cuando } \theta = \theta_a\right)\end{aligned}$$

Si θ_a es el verdadero valor de θ, entonces $(\hat{\theta} - \theta_a)/\sigma_{\hat{\theta}}$ tiene una distribución normal estándar aproximadamente. Esto significa que β puede ser determinada (en forma aproximada) al hallar un área correspondiente bajo una curva normal estándar.

Ejemplo

Suponga que se desea probar lo siguiente, $H_0 : \mu = 15$ contra $H_a : \mu = 16$. Se tiene que $n = 36$, $\overline{y} = 17$ y $s^2 = 9$. Suponemos que la muestra es lo suficientemente grande para aplicar lo visto en la sección anterior. Entonces la región de rechazo de nivel 0.05 esta dada por

$$z = \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}} > 1.645$$

que es equivalente a

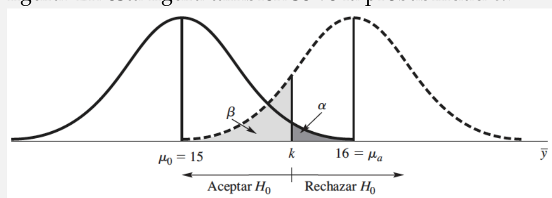
$$\overline{y}-\mu_0 > 1.645\left(\frac{\sigma}{\sqrt{n}}\right) \quad \Rightarrow \quad \overline{y} > \mu_0 + 1.645\left(\frac{\sigma}{\sqrt{n}}\right)$$

Sustituyendo $\mu_0 = 15$, $n = 36$ y usando s para aproximar σ, se tiene que la región de rechazo es

Ejemplo

$$\overline{y} > 15 + 1.645\left(\frac{3}{\sqrt{9}}\right) = 15.8225$$

Por definición, $\beta = P(\overline{Y} \leq 15.8225 \text{ cuando } \mu = 16)$ está dada por el área sombreada bajo la curva de línea interrumpida a la izquierda de $k = 15.8225$ en la siguiente figura. En esta figura también se ve la probabilidad α.



Por lo tanto, para $\mu_a = 16$,

$$\begin{aligned}\beta &= P\left(\frac{\overline{Y}-\mu_a}{\sigma/\sqrt{n}} \leq \frac{15.8225-16}{3/\sqrt{36}}\right) = P(Z \leq -0.36)\\ &= 0.3594\end{aligned}$$

De esto se puede ver que para una muestra fija de tamaño n, el tamaño de β depende de la distancia entre θ_a y θ_0. Si θ_a está cerca de θ_0, el verdadero valor de θ (y sea θ_0 o θ_a) es difícil de detectar y la probabilidad de aceptar H_0 cuando H_a es verdadera tiende a ser grande. Si θ_a está lejos de θ_0, el verdadero valor es relativamente fácil de detectar y β es mucho menor. Sin embargo este valor se puede disminuir con un tamaño muestral lo suficientemente grande

Tamaño muestral para una prueba de cola superior de nivel α

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2}$$

Demostración

Supongamos que se desea probar $H_0 : \mu = \mu_0$ contra $H_a : \mu > \mu_0$. Si se especifican los valores deseados de α y β (donde β se evalúa cuando $\mu = \mu_a$ y $\mu_a > \mu_0$) cualquier ajuste posterior de la prueba debe comprender dos cantidades restantes: el tamaño muestral n y el punto en el que empieza la región de rechazo, k. Como α y β se pueden escribir como probabilidades que contengan n y k, tenemos dos ecuaciones con dos incógnitas, de las que se puede despejar n simultáneamente. Entonces,

Demostración

$$\begin{aligned}\alpha &= P(\overline{Y} > k \text{ cuando } \mu = \mu_0)\\ &= P\left(\frac{\overline{Y}-\mu_0}{\sigma/\sqrt{n}} > \frac{k-\mu_0}{\sigma/\sqrt{n}} \text{cuando } \mu = \mu_0\right)\\ &= P(Z > z_\alpha)\\ \beta &= P(\overline{Y} \leq k \text{ cuando } \mu = \mu_a)\\ &= \left(\frac{\overline{Y}-\mu_a}{\sigma/\sqrt{n}} \leq \frac{k-\mu_a}{\sigma/\sqrt{n}} \text{cuando } \mu = \mu_a\right)\\ &= P(Z \leq -z_\beta)\end{aligned}$$

De estas ecuaciones para α y β, se tiene

$$\frac{k-\mu_0}{\sigma/\sqrt{n}} = z_\alpha \qquad \frac{k-\mu_a}{\sigma/\sqrt{n}} = -z_\beta$$

Despejando k de ambas ecuaciones tendremos

$$k = \mu_0 + z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - z_\beta\left(\frac{\sigma}{\sqrt{n}}\right)$$

Por lo tanto,

$$(z_\alpha + z_\beta)\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - \mu_0 \quad \Rightarrow \quad \sqrt{n} = \frac{(z_\alpha+z_\beta)\sigma}{(\mu_a-\mu_0)}$$

10.5 Relaciones entre los procedimientos de pruebas de hipótesis e intervalos de confianza

De la sección 8.6, observamos que si $\hat{\theta}$ es un estimador de θ con una distribución muestral aproximadamente normal, un intervalos de confianza bilateral para θ con coeficiente de confianza $1-\alpha$ está dado por

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

En esta expresión $\sigma_{\hat{\theta}}$ es el error estándar del estimador $\hat{\theta}$ y $z_{\alpha/2}$ es un número obtenido usando la tabla normal estándar tal que $P(Z > z_{\alpha/2}) = \alpha/2$. De la sección 10.3 se obtuvo que la región de rechazo de una prueba de dos colas es $\{|z| > z_{\alpha/2}\}$, el complemento de la región de rechazo es conocida como *región de aceptación* para la prueba. Esta región de aceptación está dada por $\overline{\text{RR}} = \{-z_{\alpha/2} \leq z \leq z_{\alpha/2}\}$. Esto es, no rechazamos la $H_0 : \theta = \theta_0$ a favor de la alternativa de dos colas si

$$-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2} \quad \Rightarrow \quad \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta_0 \leq \hat{\theta} + z_{\alpha/2}$$

Entonces, existe una dualidad entre nuestros procedimientos de muestra grande para construir un intervalo de confianza $100(1-\alpha)$ de dos lados y para implantar una prueba de hipótesis de dos lados con nivel α. No se rechaza la H_0 a favor de la H_a si el valor de θ_0 se encuentra dentro de un intervalo de confianza $100(1-\alpha)$ para θ. Rechazamos la H_0 si esta fuera de este intervalo. De manera equivalente, un intervalo de confianza $100(1-\alpha)$ de dos lados se interpreta como el conjunto de valores de θ_0 para los cuales $H_0 : \theta = \theta_0$ es aceptable en nivel α.

Observe que *cualquier* valor dentro del intervalo de confianza es un valor aceptable del parámetro. No hay *un valor aceptable* para el parámetro sino muchos. Por esta razón, por lo general no *aceptamos* la hipótesis nula de que $\theta = \theta_0$. Decimos que no hay suficiente evidencia para rechazar la hipótesis nula por que muchos valores de θ son aceptables por lo que no es razonable aceptar que un valor individual de θ como *el* valor verdadero.

10.6 Otra forma de presentar los resultados de una prueba estadística: niveles de significancia alcanzados o valores p

Como se ha indicado anteriormente, es frecuente que la probabilidad α de un error tipo I reciba el nombre de *nivel de significancia* o bien, *nivel* de la prueba. Aun cuando se recomienden pequeños valores de α, el valor real para usar en un análisis es un tanto arbitrario. Por esto razón es común utilizar lo que se conoce como valor p.

Definición 10.2

Si W es un estadístico de prueba, el *valor* p, o *nivel de significancia alcanzado*, es el nivel más pequeño de significancia α para el cual la información observada indica que la hipótesis nula debe ser rechazada.

Cuanto más pequeño sea el valor de p, es más fuerte la evidencia de que la hipótesis nula debe ser rechazada. Además, si el valor deseado de α es mayor o igual que el valor p, la hipótesis nula es rechazada para ese valor de α.

Para calcular esto valores se necesita hacer lo siguiente. Si fuéramos a rechazar H_0 en favor de H_a para valores pequeños de un estadístico de prueba W, por ejemplo RR : $\{w \leq k\}$, el valor p relacionado con un valor observado w_0 de W está dado por

$$\text{valor p} = P(W \leq w_0, \text{cuando } H_0 \text{ es verdadera})$$

Análogamente, si fuéramos a rechazar H_0 a favor de H_a para valores grandes de W, por ejemplo RR : $\{w \geq k\}$, el valor p relacionado con el valor observado w_0 es

$$\text{valor p} = P(W \geq w_0, \text{cuando } H_0 \text{ es verdadera})$$

Para el calculo de una prueba de ambas colas es de manera similar a los métodos anteriores. Como se esta tratando con una distribución normal entonces solo se necesita multiplicar por dos el valor p de una prueba para una cola.

10.8 Prueba de hipótesis con muestras pequeñas para μ y $\mu_1 - \mu_2$

Al igual que los intervalos de confianza con muestras pequeñas para medias y comparaciones de medias en el caso de pruebas de hipótesis se ocupa la distribución t para definir la región de rechazo y el estadístico de prueba. Como la distribución t es simétrica y en forma de campana, la región de rechazo se determina de forma semejante a la empleada con el estadístico Z de una muestra grande. Con esto se tiene lo siguiente,

Prueba de muestra pequeña para μ

Suposiciones: $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con $E(Y_i) = \mu$

$H_0 : \mu = \mu_0$

$$H_a : \begin{cases} \mu > \mu_0 & \text{(alternativa de cola superior)} \\ \mu < \mu_0 & \text{(alternativa de cola inferior)} \\ \mu \neq \mu_0 & \text{(alternativa de dos colas)} \end{cases}$$

Estadístico de prueba: $T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

$$\text{Región de rechazo: } \begin{cases} t > t_\alpha & \text{(RR de cola superior)} \\ t < -t_\alpha & \text{(RR de cola inferior)} \\ |t| > t_{\alpha/2} & \text{(RR de dos colas)} \end{cases}$$

donde t_α es tal que $P\{T > t_\alpha\} = \alpha$ para una distribución t con $n-1$ grados de libertad.

Para una prueba de hipótesis acerca de la diferencia de dos medias con muestra chica se puede hacer de manera similar al caso de encontrar los intervalos de confianza. Se utiliza el mismo estadístico de prueba y la varianza ponderada. De tal manera se tiene que para dos muestras aleatorias $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ y $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ donde la media y la varianza de la i-ésima población son μ_i y σ^2 para $i = 1, 2$,

Pruebas con muestras pequeñas para comparar dos medias poblacionales

Suposiciones: muestras independientes de distribuciones normales con $\sigma_1 = \sigma_2$

$H_0 : \mu_1 - \mu_2 = D_0$

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0 & \text{(alternativa de cola superior)} \\ \mu_1 - \mu_2 < D_0 & \text{(alternativa de cola inferior)} \\ \mu_1 - \mu_2 \neq D_0 & \text{(alternativa de dos colas)} \end{cases}$$

Estadístico de prueba: $T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ donde

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

Pruebas con muestras pequeñas para comparar dos medias poblacionales

$$\text{Región de rechazo: } \begin{cases} t > t_\alpha & \text{(RR de cola superior)} \\ t < -t_\alpha & \text{(RR de cola inferior)} \\ |t| > t_{\alpha/2} & \text{(RR de dos colas)} \end{cases}$$

donde t_α es tal que $P\{T > t_\alpha\} = \alpha$ para una distribución t con $v = n_1 + n_2 - 2$ grados de libertad.

Tanto la prueba t para una sola media y la prueba t para comprar dos medias poblacionales (a veces llamada *prueba t de dos muestras*) son robustas con respecto a la suposición de normalidad, al igual que con respecto a la suposición de que $\sigma_1 = \sigma_2$ cuando n_1 y n_2 son iguales. Esto significa que son insensibles a las violaciones de los supuestos formales.

10.9 Pruebas de hipótesis referentes a varianzas

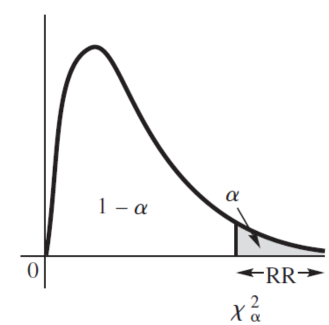
Al igual que en la sección pasada, para las pruebas de hipótesis referentes a varianzas nos aprovecharemos de la relación que existe entre los intervalos de confianza y las pruebas de hipótesis. Entonces para una $H_0 : \sigma^2 = \sigma_0^2$ se puede ocupar el siguiente estadístico de prueba,

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Este es el mismo que se ocupo para un intervalo de confianza de la varianza poblacional. Este estadístico tiene una distribución χ^2 con $n-1$ grados de libertad. Entonces para desarrollar las regiones de rechazo posibles se basa en esta distribución. Observe que S^2 es grande con respecto a σ_0^2 si y sólo si $\chi^2 = (n-1)S^2/\sigma_0^2$ es grande. Entonces si se desea una prueba de cola superior con probabilidad de error tipo I de α se tiene que la región de rechazo es

$$\text{RR} = \{\chi^2 > \chi^2_\alpha\}$$

Esta es el área sombreada de la siguiente figura.



Para la prueba de cola inferior y prueba de dos colas se tiene lo siguiente.

Pruebas de hipótesis referentes a una varianza poblacional

Suposiciones: $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$

$H_0 : \sigma^2 = \sigma_0^2$

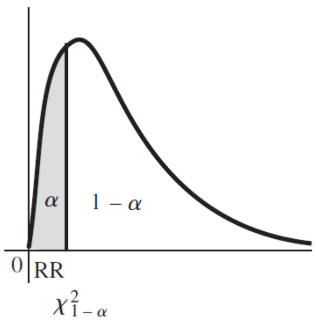
$$H_a : \begin{cases} \sigma^2 > \sigma_0^2 & \text{(alternativa de cola superior)} \\ \sigma^2 < \sigma_0^2 & \text{(alternativa de cola inferior)} \\ \sigma^2 \neq \sigma_0^2 & \text{(alternativa de dos colas)} \end{cases}$$

Estadístico de prueba: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$.

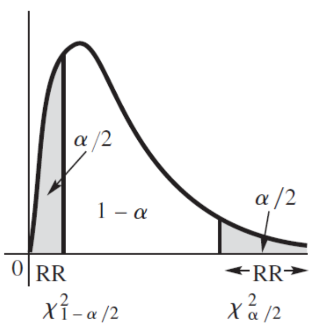
$$\text{Región de rechazo:} \begin{cases} \chi^2 > \chi^2_{\alpha}, \text{(RR de cola superior)} \\ \chi^2 < \chi^2_{1-\alpha}, \text{(RR de cola inferior)} \\ \chi^2 > \chi^2_{\alpha/2} \quad \text{o} \\ \chi^2 < \chi^2_{1-\alpha/2}, \text{(RR de dos colas)} \end{cases}$$

Observe que χ^2_α se elige de modo que, para $v = n - 1$ grados de libertad, $P(\chi^2 > \chi^2_\alpha) = \alpha$

La región de rechazo de la prueba de la cola inferior se muestra en la siguiente figura.



La región de rechazo de la prueba de dos colas es el área subrayada de la siguiente figura.



Al igual que en el caso de medias poblacionales, se puede desarrollar una prueba para la comparación de las varianzas de dos distribuciones normales. Particularmente se puede desarrollar una prueba para determinar si son iguales.

Suponemos que $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ y $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ son muestras aleatorias independientes de distribuciones normales con medias desconocidas y que $V(Y_{1i}) = \sigma_1^2$ y $V(Y_{2i}) = \sigma_2^2$, donde σ_1^2 y σ_2^2 son incógnitas. Se desea probar la hipótesis nula $H_0 : \sigma_1^2 = \sigma_2^2$ contra la alternativa $H_a : \sigma_1^2 > \sigma_2^2$.

Como las varianzas muestrales S_1^2 y S_2^2 estiman las respectivas varianzas poblacionales, rechazamos H_0 a favor de H_a si S_1^2 es mucho mayor que S_2^2. Estos significa que la región de rechazo tiene la forma

$$\text{RR} = \left\{ \frac{S_1^2}{S_2^2} > k \right\}$$

donde k se elige de modo que la probabilidad de cometer un erro tipo I sea α. El valor apropiado de k depende de la distribución de probabilidad del estadístico S_1^2/S_2^2. De capítulos anteriores se sabe que $(n_1-1)S_1^2/\sigma_1^2$ y $(n_2-1)S_2^2/\sigma_2^2$ son variables aleatorias independientes χ^2. De la definición 7.3 se tiene que

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

tiene una distribución F con $(n_1 - 1)$ grados de libertad en el numerador y $(n_2 - 1)$ grados de libertad en el denominador. Con esto se desarrolla lo siguiente.

Prueba de hipótesis $\sigma_1^2 = \sigma_2^2$

Suposiciones: muestras independientes de poblaciones normales.

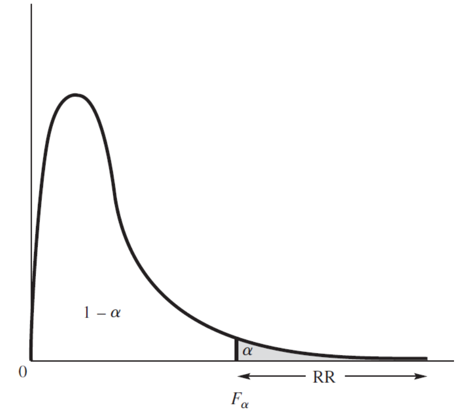
$H_0 : \sigma_1^2 = \sigma_2^2$.
$H_a : \sigma_1^2 > \sigma_2^2$.

Estadístico de prueba: $F = \frac{S_1^2}{S_2^2}$.

Región de rechazo $F > F_\alpha$

Donde F_α se elige para que $P(F > F_\alpha) = \alpha$ donde F tiene $v_1 = n_1 - 1$ grados de libertad en el numerador y $v_2 = n_2 - 1$ grados de libertad en el denominador.

Esta región de rechazo se muestra en la siguiente figura.



No es necesario establecer la prueba alternativa de $\sigma_1^2 < \sigma_2^2$ ya que se puede elegir la población que se desea para la varianza del numerador. Para la prueba alternativa de $H_a : \sigma_1^2 \neq \sigma_2^2$ se hace de manera similar a la prueba de una sola varianza. Se necesitan encontrar la cola superior de la distribución F.

Observemos que $F = S_1^2/S_1^2$ y $F^{-1} = S_1^2/S_2^2$ tienen distribuciones F, pero los grados de libertad se intercambian. Denotemos con F_b^a una variable aleatoria con una distribución F con a y b grados de libertad en numerador y denominador, respectivamente y sea $F^a_{b,\alpha/2}$ tal que

$$P(F_b^a > F^a_{b,\alpha/2}) = \alpha/2$$

Entonces

$$P((F_b^a)^{-1} > (F^a_{b,\alpha/2})^{-1}) = \alpha/2$$

por lo tanto,

$$P[F_b^a < (F^a_{b,\alpha/2})^{-1}] = \alpha/2$$

Esto significa que, el valor que corta un área de cola inferior de $\alpha/2$ para una distribución F_a^b se puede hallar al invertir $F^a_{b,\alpha/2}$. Entonces para la prueba alternativa $H_a : \sigma_1^2 \neq \sigma_2^2$ la región de rechazo es

$$\text{RR} : \left\{ F > F^{n_1-1}_{n_2-1,\alpha/2} \text{ o } F < \left(F^{n_2-1}_{n_1-1,\alpha/2}\right)^{-1} \right\}$$

Una prueba equivalente se obtiene denotando n_L y n_S a los tamaños muestrales relacionados con las varianzas muestrales mayor y menor, respectivamente. Se pone la varianza muestral mayor en el numerador y la varianza muestral menor en el denominador del estadístico F y rechazamos $H_0 : \sigma_1^2 = \sigma_2^2$ a favor de $H_a : \sigma_1^2 \neq \sigma_2^2$ si $F > F_{\alpha/2}$, donde $F_{\alpha/2}$ está determinada por $v_1 = n_L - 1$ y $v_2 = n_S - 1$ grados de libertad en el numerador y denominador respectivamente. Las pruebas basadas en la χ^2 y F no son pruebas robustas. Esto es, son muy sensibles a desviaciones respecto a la suposición de normalidad.

10.10 Potencia de las prueba y el lema de Neyman-Pearson

La bondad de una prueba es medida por α y β, las probabilidades de errores tipo I y II. Por lo general, el valor de α se elige de antermano y determina la ubicación de la región de rechazo. Un concepto relacionado pero muy útil es el siguiente.

Definición 10.3

Suponga que W es el estadístico de prueba y RR es la región de rechazo para una prueba de una hipótesis que involucra el valor de un parámetro θ. Entonces la *potencia* de la prueba, denotada por potencia(θ), es la probabilidad de que la prueba lleve al rechazo de H_0 cuando el valor real del parámetro es θ. Esto es,

$$\text{potencia}(\theta) = P(W \text{ en RR cuando el valor del parámetro es } \theta)$$

Suponga que deseamos probar la hipótesis nula $H_0 : \theta = \theta_0$ y que θ_a es un valor particular para θ elegido de H_a. La potencia de una prueba en $\theta = \theta_0$, potencia(θ_0), es igual a la probabilidad de rechazar H_0 cuando H_0 es verdadera. Esto es, potencia$(\theta_0) = \alpha$, la probabilidad de cometer un error tipo I.

Para cualquier valor de θ a partir de H_a, la potencia de una prueba mide la capacidad de la prueba para detectar que la hipótesis nula es falsa. Esto es, para $\theta = \theta_a$,

$$\text{potencia}(\theta_a) = P(\text{rechazar } H_0 \text{ cuando } \theta = \theta_a)$$

Si expresamos la probabilidad β de cometer un error tipo II cuando $\theta = \theta_a$ como $\beta(\theta_a)$, entonces

$$\beta(\theta_a) = P(\text{aceptar } H_0 \text{ cuando } \theta = \theta_a)$$

Se deduce que la potencia de la prueba en θ_a y la probabilidad de cometer un error tipo II están relacionadas de la siguiente manera.

Relación entre potencia y β
Si θ_a es un valor de θ en la hipótesis alternativa H_a, entonces

$$\text{potencia}(\theta_a) = 1 - \beta(\theta_a)$$

Idealmente, una prueba detectaría una desviación desde $H_0 : \theta = \theta_0$ con certeza; en otras palabras, potencia(θ_a) sería 1 para toda θ_a en H_a. Porque para un tamaño muestral fijo, α y β no pueden hacerse arbitrariamente pequeños, esto no es posible. Entonces, para un tamaño muestral n, se adopta el procedimientos de seleccionar un valor (pequeño) para α y hallar una región de rechazo RR para *minimizar* $\beta(\theta_a)$ en cada θ_a en H_a. En otras palabras se selecciona la RR para *maximizar* la potencia(θ) para θ en H_a. Para hacer esto primero se necesita la siguiente defnición.

Definición 10.4

Si se toma una muestra aleatoria de una distribución con parámetro θ, se dice que una hipótesis es *simple* si *especifica de manera única* la distribución de la población de la cual se toma la muestra. Cualquier hipótesis que no sea simple se denomina *hipótesis compuesta*.

Supongamos se desea probar una hipótesis nula *simple* $H_0 : \theta = \theta_0$ contra una hipótesis alternativa *simple* $H_a : \theta = \theta_a$. Como estamos interesados sólo en dos valores particulares de θ (θ_0 y θ_a). Nos gustaría escoger una región de rechazo RR para que $\alpha =$ potencia(θ_0) sea un valor fijo y potencia(θ_a) sea tan grande como sea posible. Esto es, buscamos la *más potente* prueba de nivel α. En el siguiente teorema se proporciona la metodología para obtener la más potente prueba para probar H_0 simple contra H_a simple.

Teorema 10.1

El lema de Neyman-Pearson Suponga que deseamos probar la hipótesis nula simple $H_0 : \theta = \theta_0$ contra la hipótesis alternativa simple $H_a : \theta = \theta_a$, con base en una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de una distribución con parámetro θ. Sea $L(\theta)$ la verosimilitud de la muestra cuando el valor del parámetro es θ. Entonces, para una α dada, la prueba que maximiza la potencia en θ_a tiene una región de rechazo, RR, determinada por

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

El valor de k se escoge de modo que la prueba tenga el valor deseado para α. Esta prueba es la más potente en el nivel α para H_0 contra H_a.

Suponga que hacemos un muestreo de una población cuya distribución está especificada por completo, excepto para el valor de un solo parámetro θ. SI deseamos probar $H_0 : \theta = \theta_0$ (simple) contra $H_a : \theta > \theta_0$ (compuesto). El teorema 10.1 se puede aplicar para obtener una más potente prueba para $H_0 : \theta = \theta_0$ contra $H_a : \theta = \theta_a$ para cualquier valor individual θ_a donde $\theta_a > \theta_0$. En muchas situaciones, la región de rechazo real para la más potente prueba depende sólo del valor de θ_0 (y no depende de la selección particular de θ_a). Cuando una prueba obtenidar por el Toerema 10.1 en realidad maximiza la potencia para todo valor de θ mayor que θ_0, se dice que es una prueba *uniformemente más potente* para $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$. Hay una ligera desventaja, el hecho de que si hay más de un parámetro desconocido entonces no se puede aplicar este teorema.

10.11 Pruebas de razón de probabilidad

En esta sección se vera un procedimiento que función para hipótesis simples o compuestas ya sea que estén presentes o no otro parámetros con valores desconocidos. Suponga que una muestra aleatoria se selecciona de una distribución y que la función de probabilidad $L(y_1, y_2, \ldots, y_n | \theta_1, \theta_2, \ldots, \theta_n)$ es una función de k parámetros, $\theta_1, \theta_2, \ldots, \theta_k$. Para simplificar la notación, denotemos con Θ el vector de todos los parámetros k, es decir, $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$ y escribimos la función de probabilidad como $L(\Theta)$.

Si no nos interesan todos los parámetros entonces se denomina a los parámetros de no interés como *parámetro de ruido* y la función de probabilidad sería una función con parámetros de ruido desconocidos y un parámetro de interés.

Ahora, supongamos que la hipótesis nula especifica que Θ (puede ser un vector) se encuentra en un conjunto particular de posibles valores, por ejemplo Ω_0 y que la hipótesis alternativa especifica que θ está en otro conjunto de posibles valores Ω_a, que no se traslapa con Ω_0. Cualquiera de las dos hipótesis H_0 y H_a, o ambas, pueden ser compuestas porque podrían contener valores múltiples del parámetro de interés o porque puedan estar presentes otros parámetros desconocidos.

Denotemos con $L(\widehat{\Omega}_0)$ el máximo de la función de verosimilitud para toda $\Theta \in \Omega_0$. Esto es, $L(\widehat{\Omega}_0) = \max_{\Theta \in \Omega_0} L(\Theta)$. Observe que $L(\widehat{\Omega}_0)$ representa la mejor explicación para los datos observados para toda $\Theta \in \Omega_0$ y puede hallarse mediante métodos similares en la Sección 9.7. Del mismo modo, $L(\widehat{\Omega}) = \max_{\Theta \in \Omega} L(\Theta)$ representa la mejor explicación para los datos observados para toda $\Theta \in \Omega = \Omega_0 \cup \Omega_a$.

Si $L(\widehat{\Omega}_0) = L(\widehat{\Omega})$, entonces una mejor explicación para los datos observados se puede hallar dentro de Ω_0 y no deberíamos rechazar la hipótesis nula $H_0 : \Theta \in \Omega_0$; pero si $L(\widehat{\Omega}_0) < L(\widehat{\Omega})$, entonces la mejor explicación para los datos observados se puede hallar dentro de Ω_a y deberíamos considerar rechazar H_0 a favor de H_a. Esto es la base de lo siguiente.

Prueba de razón de probabilidades
Defina λ mediante

$$\lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)}$$

La prueba de razón de verosimilitudes de $H_0 : \Theta \in \Omega_0$ contra $H_a : \Theta \in \Omega_a$ emplea λ como un estadístico de prueba, y la región de rechazo está determinada por $\lambda \leq k$. El valor real de k se selecciona de modo que α alcance el valor deseado.

Desafortunadamente, el método de razón de probabilidades no siempre produce un estadístico de prueba con distribución de probabilidad conocida. Sin embargo, si el tamaño muestral es grande, podemos obtener una aproximación a la distribución de λ si algunas ’ condiciones de regularidad ’ razonables son satisfechas por la(s) distribución(es) poblacional(es) básica(s). Las condiciones de regularidad comprenden básicamente la existencia de derivadas, con respecto a los parámetros, de la función de probabilidad. Otra condición clave es que la región sobre la cual la función de probabilidad es positiva no puede depender de valores paramétricos desconocidos.

Teorema 10.2

Sean $Y_1, Y_2, \ldots, Y_n$ que tienen una función de probabilidad conjunta $L(\Theta)$. Denotemos con r_0 el número de parámetros libres que están especificados por $H_0 : \Theta \in \Omega_0$ y denotemos con r el número de parámetros libres especificados por el enunciado $\Theta \in \Omega$. Entonces, para n grande, $-2\ln(\lambda)$ tiene aproximadamente una distribución χ^2 con $r_0 - r$ grados de libertad.

Observe que $-2\ln(\lambda)$ es una función decreciente de λ. Como la prueba de razón de probabilidades especifica que usemos RR : $\{\lambda < k\}$, este rechazo se puede reescribir como RR : $\{-2\ln(\lambda) > -2\ln(k) = k^*\}$. Para tamaños muestrales grandes, se deseamos una prueba de nivel α, el Teorema 10.2 implica que $k^* \approx \chi^2_\alpha$. Esto es, una prueba de razón de probabilidades para una muestra grande tiene la región de rechazo dada por

$$-2\ln(\lambda) > \chi^2_\alpha \quad \text{donde } \chi^2_\alpha \text{ está basada en } r_0 - r \text{ grados de libertad}$$