CONTRASTE DE HIPÓTESIS PARAMÉTRICOS

1.Introducción .

El problema del contraste de hipótesis consiste básicamente en comprobar cotejar, decidir , en definitiva , sobre la veracidad de una hipótesis prefijada previamente como supuestamente cierta . En términos estadísticos , la o las hipótesis que formulamos lo serán lógicamente sobre la población. Bien afectando a algún parámetro de ésta , lo que da origen a los contrastes paramétricos o bien a otras características de la mismas que no lo sean estrictamente, lo que origina contrates "no" paramétricos . Si bien este capítulo esta dedicado a los contrastes paramétricos , esta introducción puede considerarse común a ambos tipos de contrastes .

La solución estadística del problema de contrastación se basará en los datos muestrales y la base estadística (probabilística) de la que arrancará el contraste va a ser la distribución de algún estadístico muestral

Supongamos que deseamos hacer un contraste acerca de un parámetro θ , de la población .Para llevarlo acabo consideraremos la distribución de algún estadístico muestral que de alguna manera se corresponda y se relacione con el parámetro θ ; designemos en general a este estadístico como T. Si con los datos muestrales obtenemos un valor concreto para T tal que pertenezca a una determinada región del campo de variación de T optaremos por no rechazar la hipótesis y en caso contrario por rechazarla. Obviamente la clave del problema será delimitar la región del campo de variación de T que consideraremos como zona de aceptación de la hipótesis .Esto se resolverá por un criterio probabílistico partiendo de la distribución muestral de T.

Pasemos a definir los principales conceptos implicados en nuestro problema :

Región crítica . Será aquella región del campo de variación del estadístico tal que si contiene al valor evaluado del mismo con los datos muestrales nos llevará a *rechazar* la hipótesis . La designaremos por R_1

Región de aceptación. Es la región complementaria de la anterior .Si el valor evaluado del estadístico pertenece a ella *No rechazamos* la hipótesis.(Las hipótesis nunca se aceptan de forma definitiva, sólo se aceptan provisionalmente, es decir ,no se rechazan, a la espera de una nueva información que eventualmente pueda llevarnos a rechazarla en el futuro). La designaremos por R_0 . Evidentemente los conjuntos de puntos que forman ambas regiones son disjuntos.

Una ***hipótesis estadística*** (paramétrica) es una conjetura sobre el valor concreto que tiene en realidad. El establecer una hipótesis sobre un parámetro θ , supone dividir los posibles valores del parámetro en dos grupos disjuntos tales que unos son hipotéticamente ciertos(θ_0) y los otros(θ_1) no lo son . A la hipótesis que se desea contrastar se la denomina ***"hipótesis nula"*** **,** siendo , por tanto, el valor o valores θ_0 que hipotéticamente consideramos reales , dicha hipótesis viene expresada como H_0. Alternativamente y consecuentemente se establece la denominada ***"hipótesis alternativa "*** (H_1)compuesta ésta por el valor o valores θ_1 que en consecuencia de la elección y de la complementariedad de los de la hipótesis nula , son los que , en principio, no consideramos cómo hipotéticamente reales.

El hecho de que las hipótesis , tanto la nula cómo la alternativa puedan recoger en sus planteamientos uno o varios valores , da lugar a ***hipótesis*** de carácter ***simple*** , si el número de valores plausibles e hipotéticos es de uno en ambas , o bien a ***hipótesis compuestas*** si dicho valor no es único en alguna de ellas.

Teniendo en cuenta lo dicho anteriormente , el problema de rechazar o aceptar una hipótesis puede plantearse como un problema de decisión , en el que evidentemente existe la posibilidad de fracasar o acertar en la elección o decisión a la hora de concluir que la hipótesis, bien nula bien alternativa, son rechazables o no , dado , claro está ,que no conocemos la verdad.

El problema de decisión :rechazo/no rechazo, vendría expresado en las siguientes opciones en forma de tabla :

HIPÓTESIS/ACCIÓN-DECISIÓN	NO RECHAZAMOS	RECHAZAMOS
ES CIERTA	CORRECTO	ERROR TIPO I
ES FALSA	ERROR TIPO II	CORRECTO

Así :
Si la hipótesis nula (H_0) es cierta y nuestra decisión es no rechazarla, la decisión ha sido correcta Si la hipótesis nula (H_0) es cierta y nuestra decisión es rechazarla, la decisión provoca un error . Dicho error se denomina error tipo I.

Si la hipótesis nula (H_0) es falsa y nuestra decisión es no rechazarla, la decisión provoca un error . Dicho error se denomina error tipo II.

Si la hipótesis nula (H_0) es falsa y nuestra decisión es rechazarla , la decisión ha sido correcta.

Las situaciones anteriores , tanto las que plantean un error cómo las que plantean acierto , son incontrolables e incomprobables (en ese momento) por lo que la única manera de abordar el problema radica en conocer o establecer las ***probabilidades*** con las que se pueden cometer los errores .Así tendríamos que :

α ,es la probabilidad de cometer un error tipo I , es decir , la probabilidad de rechazar una hipótesis que en verdad es cierta . Puede recordarse del tema anterior que así definíamos y denotábamos al nivel de significación. Es también α el tamaño de la región crítica del contraste. Si es la probabilidad de cometer un error evidentemente deberá ser un valor pequeño , próximo a 0.

β ,es la probabilidad de cometer un error tipo II , es decir , la probabilidad de no rechazar una hipótesis que en la realidad es falsa , no tiene nombre concreto que la denote , si bien su complementaria **1-β** (probabilidad de rechazar una hipótesis que es falsa) se denomina **potencia del contraste** para el caso de contrastes simple y función de potencia para el caso en el que el contraste fuera del tipo compuesto. Analíticamente lo expresado anteriormente quedaría de la siguiente manera :

Siendo θ el parámetro desconocido de la población sobre el que queremos realizar un contraste Siendo T un estadístico relacionado con θ y al que la muestra específica realizada a concretado un valor T .

Siendo R_1 la región crítica del contraste ; región de valores que dan lugar a rechazar la hipótesis.
Siendo θ_0 el valor asignado hipotéticamente a θ , y que constituye , por tanto la H_0

Tendremos que

$$\alpha = P\left[T \in R_1 / \theta = \theta_0\right]$$

=P[error tipo I]=

=P[T esté en la zona de rechazo siendo cierta la hipótesis]=

=P[T no pertenezca a la zona de aceptación siendo cierta la hipótesis]=

$$= P\left[T \notin R_0 / \theta = \theta_0\right] = \alpha$$

Tendremos también que

$$\beta = P\left[T \in R_0 / \theta \neq \theta_0\right]$$

=P[error tipo II]=
=P[T esté en la zona de No rechazo siendo falsa la hipótesis]=
=P[T no esté en la zona de rechazo siendo falsa la hipótesis]=

$$= P\left[T \notin R_1 / \theta \neq \theta_0\right] = \beta = P\left[T \notin R_1 / H_0 \, falsa\right]$$

Además

$$1 - \beta = P\left[T \in R_1 / \theta \neq \theta_0\right] =$$

=P[T esté en la zona de rechazo siendo falsa la hipótesis]=
=P[T no esté en la zona de aceptación siendo falsa la hipótesis]=
=P[T esté en la zona de rechazo siendo cierta la hipótesis alternativa H_1]=

$$= P\left[T \in R_1 / H_1 \, cierta\right]$$

Por último

$$1 - \alpha = P\left[T \in R_0 / \theta = \theta_0\right] = P\left[T \in R_0 / H_0 \, cierta\right] =$$

=P[T esté en la zona de no rechazo siendo cierta la hipótesis]

Es evidente que dado que no conocemos la realidad de la población no podemos conocer en que situación nos encontramos. De ahí que intentemos controlar la aparición de errores en el sentido de hacer que la probabilidad de cometerlos sea mínima ; minimizando el error tipo uno(nivel de significación) y el error tipo dos , maximizando ,en este caso, la potencia del contraste dada su complementariedad. Ambos planteamientos idóneos (mínimo α , mínimo β (máximo 1-β) no son posibles dado que nivel de significación y potencia del contraste no son independientes. Por lo que debemos plantearnos prefijar como mínimo(α ó β) el error que suponga menor coste a la hora de las consecuencias de su aparición .

Comentado en términos genéricos la realización del contrates quedaría en la siguiente forma :

Conocemos que la distribución de los estadísticos muestrales depende de la distribución y parámetros poblacionales . En consecuencia la distribución de T dependerá del valor del parámetro a contrastar θ . De forma que podemos construir la distribución del estadístico T condicionada al valor hipotético de θ , es decir , $\theta = \theta_0$ siendo esta la hipótesis nula del contraste.

Conocida la distribución y una vez prefijado el nivel de significación del contraste , tendremos que delimitar la región

R_0 [a , b] en la que se verifique que :

$$P\left[T \notin R_0 / \theta = \theta_0\right] = \alpha$$

Determinar esta región (zona) [a, b] es un problema análogo , pero recíproco al de la construcción de un intervalo de confianza 1-α . dado que :

$$P\left[T \notin R_1 / \theta = \theta_0\right] \rightarrow P\left[T \in R_o / \theta = \theta_0\right] = 1 - \alpha$$ es decir

$$P\left[(a < T < b) / \theta = \theta_0\right] = 1 - \alpha$$

Si en la construcción del intervalo de confianza nos interesaba que la amplitud del intervalo fuera lo menor posible para disponer de estimaciones más precisas, aquí nos interesa que los intervalos externos (región crítica o de rechazo) cuya expresión es : $]-\infty ,a]$ $[b,\infty[$ sean lo más grandes posibles para apurar más (hilar más fino) a la hora de aceptar la hipótesis.(Se trata de poner difícil la aceptación de la hipótesis :hacer un contraste duro).Y si la región crítica debe ser lo mayor posible la de aceptación [a,b] deberá ser lo menor posible. Y como ya conocemos, si la distribución es simétrica y unimodal el intervalo de menor amplitud (y , por tanto ,de mayor densidad media de probabilidad) de todos los intervalos que cumplen que $P[T \in [a, b]] = 1-\alpha$ será el intervalo centrado en la media.

Una vez determinadas las regiones crítica y de aceptación (una vez determinados a y b) y de manera sencilla el contraste se realiza de la siguiente manera
Si los datos muestrales dan lugar a un estadístico T tal que si :

T (con hipótesis) $\in$ [a,b] $\rightarrow$NO **rechazamos hipótesis : $\theta = \theta_0$**

T (con hipótesis) $\notin$ [a,b] $\rightarrow$ rechazamos hipótesis : $\theta = \theta_0$

Tratemos a partir de ahora de establecer algunos contrastes de interés

2.-Contrastar que la media de una población normal toma un determinado valor.

Deseamos contrastar la hipótesis de que el parámetro poblacional $\theta = \mu$ toma un determinado valor μ_0 . Conocemos que la población se distribuye normalmente y conocemos también su varianza , o bien si nos es desconocida ,el tamaño muestral es lo suficientemente grande cómo para poder utilizar la muestral cómo poblacional. Hemos determinado un nivel de significación para la realización del contraste y vamos a plantearlo en el supuesto de realizar una muestra aleatoria de tamaño n.

así : conocemos que

$$\bar{x} \Rightarrow N\left[\mu, \sigma/\sqrt{n}\right]$$

de lo que deducimos que

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \Rightarrow N[0;1]$$

de forma que bajo la hipótesis nula :

$H_0: \mu = \mu_0$ tendremos el intervalo.

$$P\left[-\lambda_{\alpha/2} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}\right] = 1 - \alpha$$

realizada la muestra y conocidas n y $\bar{x}$ conoceremos el valor T del estadístico

$$T = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

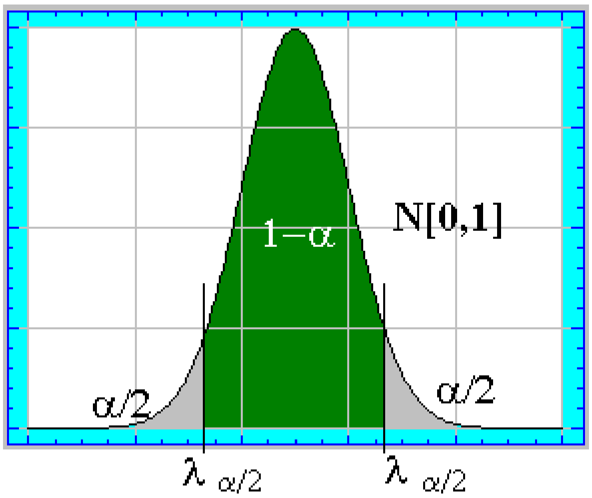
de manera que establecemos

si $T = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \in \left[-\lambda_{\alpha/2}; \lambda_{\alpha/2}\right] = R_0 \mapsto$

No rechazamos $H_0: \mu = \mu_0$

si $T = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \notin \left[-\lambda_{\alpha/2}; \lambda_{\alpha/2}\right] = R_0 \mapsto$

rechazamos $H_0: \mu = \mu_0$



donde como ya es conocido

$$\lambda_{\alpha/2}$$

es el valor de la N[0 ;1] para el que se verifica que

$$F\left(\lambda_{\alpha/2}\right) = P\left[z \leq \lambda_{\alpha/2}\right] = 1 - \alpha$$

ejemplo 1.

De 100 observaciones de una población normal se obtiene que $\bar{x} = 5$ y que S =2.Contrastar con un nivel de significación del 5% la hipótesis de que la media de la población sea 7.

Tendremos que : $H_0 : \mu = 7$
$H_1 : \mu \neq 7$ conocemos que $\alpha = 0.05$

$\bar{x} = 5$ S= 2 n=100 (mayor que 30) luego podemos tomar σ = S= 2 el estadístico T

será T= $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{5-7}{2/10} = -10$

la región o zona de no rechazo para α =0.05 será

$$\left[-\lambda_{\alpha/2} ; \lambda_{\alpha/2}\right] = R_0 \quad =[-1'96 ;1'96]$$

dado que T=-10 no pertenece a la región de aceptación estamos en condiciones de rechazar la hipótesis nula luego aceptar la alternativa : $\mu \neq 7$

3.-Contrastar que la media de una población normal toma un valor μ_0 desconociendo la varianza de la población , para un nivel de significación α , y realizada una muestra pequeña

Deseamos contrastar la hipótesis de que el parámetro poblacional $\theta = \mu$ toma un determinado valor μ_0 .Desconocemos la varianza de la población y, dado que el tamaño muestral es pequeño , no podemos utilizar la muestral en su lugar.

Hemos determinado un nivel de significación para la realización del contraste y vamos a plantearlo en el supuesto de realizar una muestra aleatoria de tamaño n.

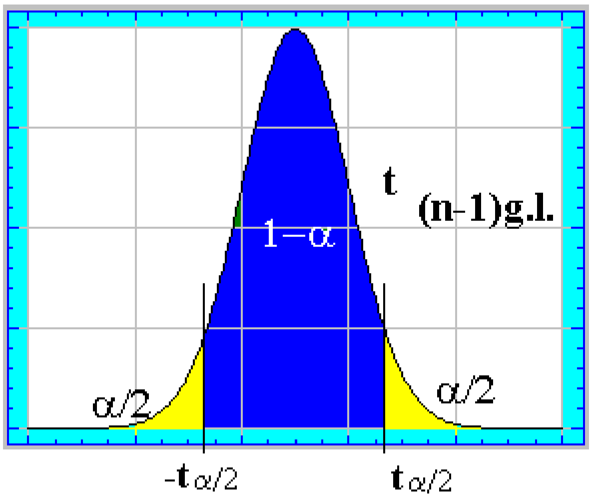
Conocemos que

$$\frac{\bar{x}-\mu}{s}\sqrt{n-1} \Rightarrow t_{n-1}$$

bajo la hipótesis planteada $H_0: \mu = \mu_0$

tendremos que

$$P\left[-t_{\alpha/2} \leq \frac{\bar{x}-\mu}{s}\sqrt{n-1} \leq t_{\alpha/2}\right] = 1-\alpha$$



Donde $t_{\alpha/2}$ es el valor de la tabla de la t de Student de (n-1) grados de libertad tal que cumple lo que se explicita en la imagen contigua.
Partiendo del intervalo expuesto y concretada la muestra obtendremos el estadístico T

tal que T= $\frac{\bar{x}-\mu_0}{s}\sqrt{n-1}$

de manera que si el estadístico T :

Si $\frac{\bar{x}-\mu_0}{s}\sqrt{n-1} \in \left[-t_{\alpha/2}; t_{\alpha/2}\right] = R_0 \mapsto$

No rechazamos la hipótesis $H_0: \mu = \mu_0$

Si T= $\frac{\bar{x}-\mu_0}{s}\sqrt{n-1} \notin \left[-t_{\alpha/2}; t_{\alpha/2}\right] = R_0 \mapsto$

Rechazamos la hipótesis $H_0: \mu = \mu_0$

ejemplo 2.

Se escoge a 17 individuos al azar y se les mide resultando que su estatura media es de 1,71 metros con desviación típica de 0,02 .Contrastar la hipótesis de que la estatura media nacional sea de 1.75 metros si utilizamos un nivel del significación del 5%. Se supone normalidad

Conocemos que n=17 α =0,05

$\bar{x} = 1{,}71$

S=0,02

deseamos contratar
H_0: μ =1,75

H_1: $\mu \neq 1{,}75$

para una t de Student de n-1=16 g.l tendremos que
$t_{\alpha/2} = t_{0.025} = 2{,}12$

por lo que el estadístico será $T = \frac{\bar{x} - \mu_0}{s}\sqrt{n-1} = \frac{1{,}71 - 1{,}75}{0{,}02} \bullet 4 = -8 = T$

dado que -8=T$\notin$ [-2,12;2,12] luego rechazamos la hipótesis nula

μ =1,75

4.-Contrastar que la diferencia de dos medias toma un determinado valor μ_0 siendo las poblaciones normales y las varianza conocidas (o tamaños muestrales grandes) para un determinado nivel de significación α

conocemos que

$$x \Rightarrow N\left[\mu_x ; \sigma_x\right]$$

con varianza conocida

conocemos también que

$$y \Rightarrow N\left[\mu_y ; \sigma_y\right]$$

con varianza conocida

siendo los tamaños muestrales n_x y n_y se formaliza la hipótesis a contrastar como :

$$H_0 : \mu_x - \mu_y = \mu_0$$

$$H_1 : \mu_x - \mu_y \neq \mu_0$$

con nivel de significación α

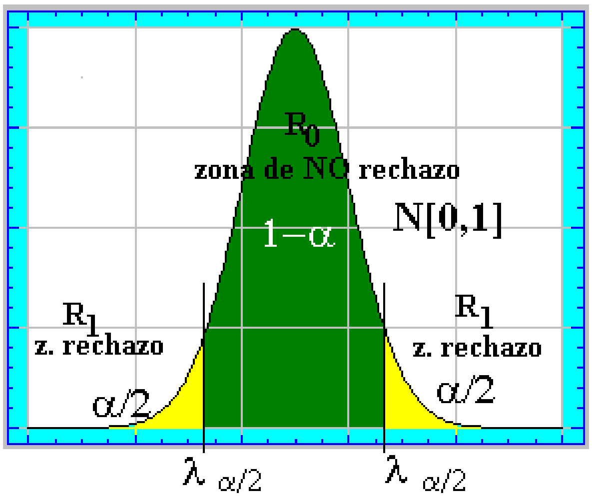
conocemos que la

$$L\left[\bar{x} - \bar{y}\right] \Rightarrow N\left[\mu_x - \mu_y ; \sqrt{\sigma_x^2 / n_x + \sigma_y^2 / n_y}\right]$$

luego bajo la hipótesis

tendremos que

$$P\left[- \lambda_{\alpha/2} \leq \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \leq \lambda_{\alpha/2}\right] = 1 - \alpha$$



de manera que para un valor concreto del nivel de significación tendremos que las zonas de decisión para el estadístico serán :

$$T= \frac{\bar{x}-\bar{y}-\mu_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \in R_0=\left[-\lambda_{\alpha/2};\lambda_{\alpha/2}\right]$$

NO rechazamos la hipótesis nula

$$T= \frac{\bar{x}-\bar{y}-\mu_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \notin R_0=\left[-\lambda_{\alpha/2};\lambda_{\alpha/2}\right]$$

rechazamos hipótesis nula , aceptando H_1

ejemplo 3.

Para comprobar si dos máquinas producen tornillos de la misma longitud se recoge una muestra aleatoria de 40 piezas producidas por la máquina A y 50 por la máquina B . Resultando que la media de las 40 piezas de A es de 10 cm. mientras que la de las 50 de B es de 9.5 cm. siendo las desviaciones típicas de las dos muestras 1 y 2 cm . respectivamente. Contrastar con un nivel de significación del 2% la hipótesis de que ambas máquinas fabrican piezas con la misma longitud media.

La hipótesis planteada será :

$$H_0: \mu_x = \mu_y \rightarrow \mu_x - \mu_y = 0$$
$$H_1: \mu_x - \mu_y \neq 0$$

Nivel de significación 0.02 por lo que la zona de no rechazo según tablas de la N[0 ;1]

quedaría R_0=[-2.33 ;2.33] siendo el estadístico T

T= $\dfrac{\bar{x}-\bar{y}-\mu_0}{\sqrt{\dfrac{\sigma_x^2}{n_x}+\dfrac{\sigma_y^2}{n_y}}} = \dfrac{10-9.5-0}{\sqrt{\dfrac{1}{40}+\dfrac{4}{50}}} = 1.5430 \in R_0[-2.33, 2.33]$

por lo que No rechazamos la igualdad de medias en la longitud de las piezas producidas por ambas máquinas

5. Contrastar que la proporción con la que se da una característica toma un determinado valor para un nivel de significación determinado y con muestreo aleatorio simple.

Según lo planteado en el enunciado la hipótesis a contrastar quedaría explicitada :

H_0 : $p=p_0$ cómo hipótesis nula frente a la alternativa :
H_1 : $p\neq p_0$
conocemos que :

$$L[\hat{p}] \Rightarrow N\left[p ; \sqrt{\frac{pq}{n}}\right]$$

por lo que para la hipótesis quedaría :

$$L[\hat{p}] \Rightarrow N\left[p_0 ; \sqrt{\frac{p_0 q_0}{n}}\right]$$

por lo que para un determinado nivel de confianza tendríamos el intervalo :

$$P\left[-\lambda_{\alpha/2} \leq \frac{\hat{p}-p_0}{\sqrt{p_0 q_0 / n}} \leq \lambda_{\alpha/2}\right] = 1-\alpha$$

por lo que al aplicar los valores muestrales la resolución del contraste quedaría :

T= $\dfrac{\hat{p}-p_0}{\sqrt{p_0 q_0 / n}} \in R_0 = \left[-\lambda_{\alpha/2} ; \lambda_{\alpha/2}\right]$

no rechazaríamos la hipótesis nula luego $P=P_0$ por el contrario si

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \notin R_0 = \left[-\lambda_{\alpha/2} ; \lambda_{\alpha/2} \right]$$

rechazaremos la hipótesis H_0

** El contraste que hemos construído lo es para muestreo aleatorio simple , es evidente que el muestreo podría haber sido irrestricto , no sólo en este caso si no también en los anteriores y posteriores que estudiamos y estudiaremos . Como en el caso de construcción de intervalos la diferencia entre m.a.s. e irrestricto radicará en la aplicación del coeficiente de exhaustividad o factor corrector de poblaciones finitas en consonancia a la varianza de la ley de probabilidad utilizada para la creación del estadístico del contraste ; así , y como ejemplo en el caso que nos ocupa pero para muestreo irrestricto, el estadístico quedaría cómo

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n} \cdot \sqrt{N - n / N - 1}}$$

manteniendo el resto consecuencias y decisiones.

ejemplo 4.

Una empresa de publicidad desea comprobar si un determinado programa de televisión es visto por el 30% de la audiencia potencial .Para ello se escoge al azar una muestra de 200 familias resultando que de ellas 50 lo ven asiduamente . Contrastar la hipótesis con un nivel de significación del 5%.

La hipótesis planteada quedaría como :

H_0 : p=0.3
H_1 : $p \neq 0.3$ con nivel de significación de 0.05 por lo que la zona o región de no rechazo quedaría según la tabla de la normal [0 ;1] como :R_0=[-1.96 ;1.96] y dado que :

$$\hat{p} = 50/200 = 0.25$$

el estadístico

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.25 - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{200}}} = -1.5430 \in \left[-1.96 ; 1.96 \right]$$

por lo que no podemos rechazar la hipótesis de que el porcentaje real sea del 30%

En este caso, evidentemente ,el hecho de muestreo irrestricto o no es improcedente ; dado que podemos considerar que la población es infinita (audiencia potencial de televisión) por lo que el hecho del reemplazamiento o no sería irrelevante, y no procede , por tanto ,tenerlo en cuenta.

6. Contrastar la hipótesis que la diferencia entre dos proporciones de dos características ,o de la misma en situaciones diferentes ,tome un determinado valor para un nivel de significación también determinado

tenemos dos poblaciones x e y , que se han muestreado con valores n_x y n_y
la hipótesis a contrastar será :

$$H_0: p_x - p_y = p_0$$

$$H_1: p_x - p_y \neq p_0$$

dado que se trata de proporciones conocemos que

$$L\left[\hat{p}_x - \hat{p}_y\right] \Rightarrow N\left[p_x - p_y ; \sqrt{\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}}\right]$$

que bajo la hipótesis planteada formaría dado que se trata de una normal el siguiente intervalo

:

$$P\left[-\lambda_{\alpha/2} \leq \frac{\hat{p}_x - \hat{p}_y - p_0}{\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}} \leq \lambda_{\alpha/2}\right] = 1 - \alpha$$

por lo que construiremos el estadístico

$$T = \frac{\hat{p}_x - \hat{p}_y - p_0}{\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}}$$

tal que si pertenece a la región (zona) de no rechazo $R_0 = \left[-\lambda_{\alpha/2} ; \lambda_{\alpha/2}\right]$ No rechazaríamos la hipótesis nula , en caso contrario y el estadístico T no perteneciera a dicha región rechazaríamos la hipótesis nula aceptando la alternativa H_1.

ejemplo 5.

Se desea saber si el coeficiente de penetración en el mercado de dos productos competidores es el mismo.Para ello se muestrean 100 familias de las cuales 20 compran el producto A y de otras 100 familias seleccionadas al azar también, e independientemente 30 compran el producto B.Contrastar la hipótesis con un nivel de significación del 2'5 %

las hipótesis planteadas serían :

$$H_0 : p_a - p_b = 0$$
$$H_1 : p_a - p_b \neq 0$$

dado que el nivel de significación es del 0.025 la zona de no rechazo $R_0 = \left[-\lambda_{\alpha/2} ; \lambda_{\alpha/2} \right]$ sería [-2.24 ;2.24]

conociendo además que

$$\hat{p}_a = {}^{20}\!/_{100} = 0.2 \quad \hat{p}_b = {}^{30}\!/_{100} = 0.3$$

el estadístico T quedaría como :

$$\frac{\hat{p}_x - \hat{p}_y - p_0}{\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}} = \frac{0.2 - 0.3 - 0}{\sqrt{\frac{0.2 \bullet 0.8}{100} + \frac{0.3 \bullet 0.7}{100}}} = -1.6439$$

dado que dicho valor pertenece al intervalo o zona antes

citado : [-2'24,2'24] no podemos rechazar la hipótesis de que ambos productos tienen la misma penetración en mercado

7.-Contrastes de una Cola

Hasta ahora hemos visto visto contrastes en los que la hipótesis nula era una hipótesis simple (la hipótesis de que el parámetro tomase un determinado valor), $H_0: \theta = \theta_0$ que se enfrentaba a la hipótesis alternativa (compuesta) $H_1: \theta \neq \theta_0$ (lo que equivalía a no establecer una alternativa concreta a la hipótesis sujeta a contraste).

Sin embargo,en muchos casos prácticos concretos nos interesará contrastar $H_0: \theta = \theta_0$

frente a la hipótesis alternativa de $\theta > \theta_0$ o bien $\theta < \theta_0$

> Cuando nos interese saber si podemos considerar que $\theta = \theta_0$ y nos vaya a reportar las mismas,o aún mejores , consecuencias el que $\theta > \theta_0$, nos interesará hacer el contraste $\begin{matrix} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{matrix}$
>
> Cuando nos interese saber si podemos considerar que $\theta = \theta_0$ y nos vaya a reportar las mismas ,o aún mejores consecuencias el que $\theta < \theta_0$,nos interesará hacer el contraste $\begin{matrix} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{matrix}$

En estos dos nuevos casos la base teórica para la decisión de aceptación o no de H_0 va a ser la misma que en el caso ya estudiado de

$$\begin{matrix} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{matrix}$$

Pero, en la medida en que el rechazo de la hipótesis nula supone la aceptación de una hipótesis alternativa bien diferente ,el diseño de los criterios de aceptación van a diferir :va a diferir la construcción de la **región crítica** y de la **región de aceptación** (no rechazo).

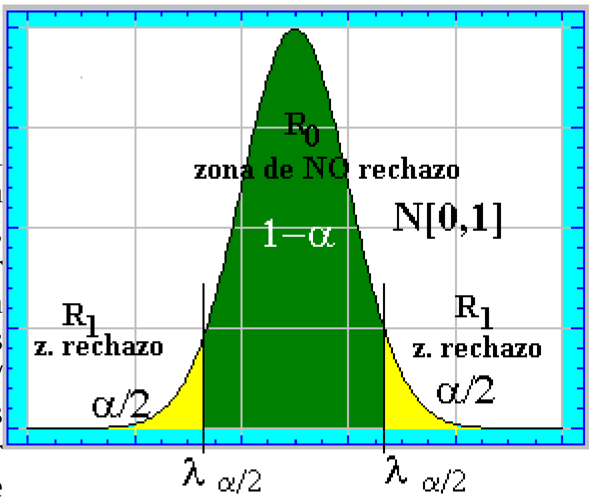
De esta manera. Dado un nivel de significación , prefijado, trabajaremos con la distribución muestral de un estadístico adecuado ,T, cuya distribución dependa del parámetro sujeto a contraste para determinar la región crítica y de aceptación ,de forma que, como en el caso ya estudiado:

La región de aceptación verifique que

$P[T \in R_0 / \theta = \theta_0] = 1 - \alpha$ y la región crítica que $P[T \in R_1 / \theta = \theta_0] = \alpha$,siendo, evidentemente, ambas complementarias.

Si los datos muestrales concretos son tales que: $T \in R_0$ no rechazamos H_0

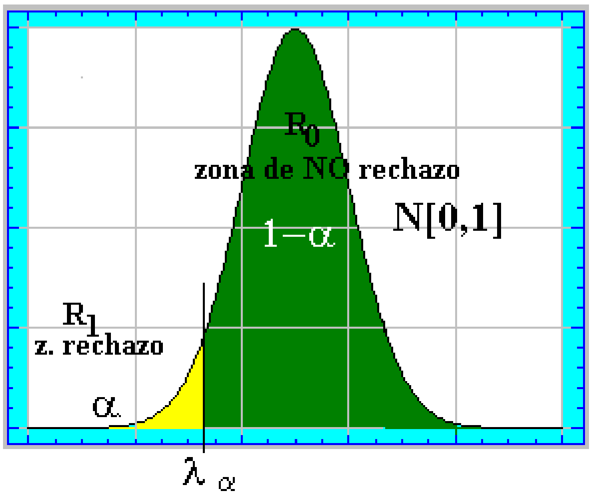
mientras que si $T \in R_1$ rechazamos H_0 luego aceptamos H_1



En el caso estudiado anteriormente, en el medida en que la distribución de T solía ser simétrica y en la que, como es lógico, estábamos interesados en hacer constrastes "duros"(severos con la hipótesis nula) elegíamos de todos los posibles pares de regiones "aceptación / crítica" precisamente aquél que nos ofrecía una región crítica de mayor amplitud y una región de aceptación de menor amplitud. Esto nos llevaba a elegir el intervalo centrado de probabilidad 1-α como región de aceptación, y como región crítica (zona de rechazo) las dos colas simétricas de probabilidad α /2.

En la medida en que rechazar $H_0: \theta = \theta_0$ suponía aceptar $H_1: \theta \neq \theta_0$ y teniendo en cuenta que la mayor parte de las distribuciones de los estadísticos están centradas en el auténtico valor del parámetro, esta determinación de las zonas de aceptación y rechazo era consecuentemente consistente con las características de nuestro contraste: un resultado muestral alejado de la zona central tanto por la izquierda como por la derecha nos da cuenta de que θ debe ser significativamente distinto de θ_0 (por defecto o por exceso). El rechazo de H_0 supone aceptar que $\theta \neq \theta_0$, e igual nos da que lo más verosímil sea que $\theta > \theta_0$ (los datos muestrales caen en la cola de la derecha) que sea $\theta < \theta_0$ (los datos muestrales caen en la cola de la izquierda).

Sin embargo al considerar el contraste del tipo : $\begin{matrix} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{matrix}$
el hecho que los datos muestrales (valor de T) caigan en la región critica R_1 nos lleva a considerar más verosímil H_1 , y, evidentemente, sólo es más verosímil H_1 frente a H_0 si los datos muestrales difieren significativamente por defecto de la zona central.



Esta argumentación nos conduce a que diseñemos una región crítica de "*una sola cola*" (cola de la izquierda). Región crítica que, recordemos debe seguir verificando que $P[T \in R_1 / \theta = \theta_0] = \alpha$. Es decir :

Ante el ejemplo de contraste :

$$H : \mu = \mu_0$$
$$H : \mu < \mu_0$$

En una población normal y varianzaconocida para muestreo aleatorio simple con un determinado nivel de significación ; tendremos que si el estadístico

$$T = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < -\lambda_\alpha$$

rechazaremos H_0 luego aceptaremos H_1
en caso contrario si

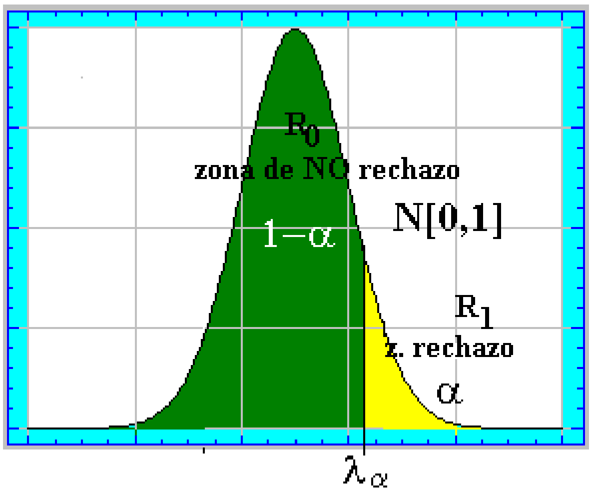
$$T = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < -\lambda_\alpha$$

no rechazaremos H_0

Evidentemente si esto ocurre con el contrate especificado de ésta manera será por analogía lo contrario en el caso de plantearse el contraste de la siguiente forma :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

Así que los datos muestrales den lugar a un valor que se encuentre en la región crítica R_1 , nos lleva considerar más verosímil H_1 y ,lógicamente , sólo es más verosímil H_1 frente a H_0 si los datos muestrales difieren significativamente por exceso de la zona central.



De forma análoga al caso anterior, esta argumentación nos conduce a que diseñemos una región crítica "de una sola cola" (cola de la derecha). Región crítica que deberá seguir cumpliendo que

$$P\left[T \in R_1 / \theta = \theta_0\right] = \alpha$$

y, por tanto, la región de aceptación será su complementaria y cumplirá que

$$P\left[T \in R_0 / \theta = \theta_0\right] = 1 - \alpha$$

Así, por ejemplo, ante un contraste:

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0$$

en una población normal con varianza desconocida y muestra pequeña

si $T = \frac{\bar{x} - \mu_0}{s}\sqrt{n-1} > t_\alpha$

rechazaremos la hipótesis nula aceptando la alternativa

en cambio si

$$T = \frac{\bar{x} - \mu_0}{s}\sqrt{n-1} < t_\alpha$$

no rechazaremos la hipótesis nula rechazando ,obviamente , la alternativa de que la media es mayor que el valor hipotético planteado. Evidentemente en este caso las zonas de aceptación y rechazo parten de la t de Student (n-1 grados de libertad) pues , como ya vimos , es lo debe utilizarse si la población es normal , conocemos la varianza y la muestra es pequeña.

ejemplo 6

Un fabricante de refrescos sin burbujas desea sacar al mercado una variedad de su producto que tenga burbujas. Su director comercial opina que al menos el 50 % de los consumidores verá con buenos ojos la innovación. Se realiza un sondeo de mercado y resulta que de 100 consumidores encuestados 40 son favorables a la innovación.

a) Contrastar la hipótesis del director comercial frente a la alternativa de que el % de aceptación es inferior, con un nivel de significación del 2,5 %.

b) Si el aceptable la hipótesis de que el % de aceptación del nuevo producto es inferior o igual al 30 % el fabricante decidirá no fabricarlo. Si es aceptable el criterio del director comercial entonces sí fabricarán el refresco con burbujas. Y si ninguna de las 2 hipótesis es aceptable procederán a hacer otro sondeo. Para tomar esta decisión trabajarán con un nivel de significación del 2,5 %. ¿ Por qué opción se decantarán ?.

se plantea la hipótesis $H_0: p=0.5$

$H_1: p> 0.5$ conocemos que $\alpha = 0.025$ por lo que

$1-\alpha = 0.975$ siendo $-\lambda\ \alpha = -1.96$

dado que

$$\hat{p} = 0.4$$

n=100 tendremos que

$$T=\frac{\hat{p}-p_0}{\sqrt{p_0 q_0 / n}} = \frac{-0.1}{0.05} = -2$$

dado que : $-2 < -1.96$ rechazamos que p=0.5 aceptando que es menor que dicho valor

b) la hipótesis planteada será :

$H_0: p=0.3$ frente a la alternativa

$H_1: p> 0.3$ conocemos que $\alpha = 0.025$ por lo que

$1-\alpha = 0.975$ siendo $\lambda\ \alpha = 1.96$ dado que $\hat{p} = 0.4$

n=100 tendremos que

$$T=\frac{\hat{p}-p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.1}{0.05} = 2$$

dado que : 2 > 1.96 rechazamos que la hipótesis sea 0.3 y aceptamos que es mayor que dicho valor.

Uniendo los dos resultados llegamos a la conclusión de que es recomendable otro sondeo que nos explicíte la proporción , si bien podemos aventurar que parece encontrarse entre el 0.5 y el 0.3

8. Contraste de igualdad de varianzas en dos poblaciones normales.

Contrastaremos la hipótesis nula de que las varianzas de dos variables X e Y son iguales frente a la alternativa de que la varianza de X es mayor a la de Y , Tomando como variable X aquella cuya varianza muestral sea mayor.

El contraste quedaría definido :

$$H_0: \sigma_x^2 = \sigma_y^2$$
$$H_1: \sigma_x^2 > \sigma_y^x$$

para un determinado nivel de significación y con tamaños muestrales n_x para la variable X y n_y para la variable Y de manera que conocemos que :

$$\frac{n_x S_x^2 \sigma_y^2 (n_y - 1)}{n_y S_y^2 \sigma_x^2 (n_x - 1)} \Rightarrow F(n_x - 1)(n_y - 1)$$

bajo la hipótesis nula

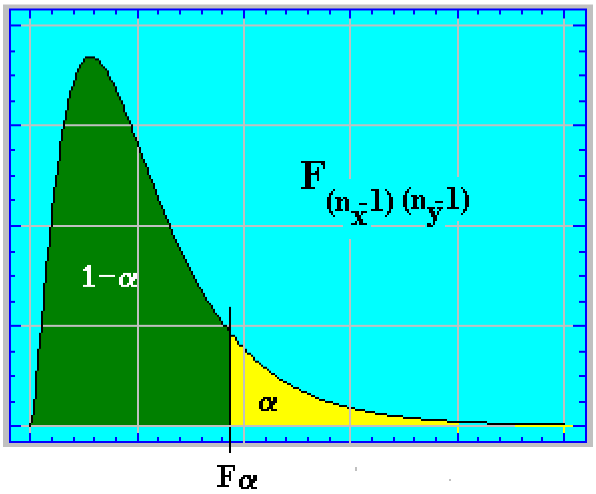
$$\left(\sigma_x^2 = \sigma_y^2\right) \rightarrow \frac{\sigma_y^2}{\sigma_x^2} = 1$$

luego tendríamos que

$$\frac{n_x S_x^2 (n_y - 1)}{n_y S_y^2 (n_x - 1)} \Rightarrow F(n_x - 1)(n_y - 1)$$

Para un nivel de significación α tendremos :

$$P\left(\frac{n_x S_x^2 (n_y - 1)}{n_y S_y^2 (n_x - 1)} < F_\alpha \right) = 1 - \alpha$$



donde Fα es el valor crítico de las tablas para
n_x-1 grados de libertad en el numerador y
n_y-1 grados de libertad en el denominador
y un nivel de significación α

luego , en definitiva el contraste quedaría :

si

$$\frac{n_x S_x^2 (n_y - 1)}{n_y S_y^2 (n_x - 1)} < F_\alpha$$

No rechazamos la hipótesis de igualdad de varianzas

si

$$\frac{n_x S_x^2 (n_y - 1)}{n_y S_y^2 (n_x - 1)} > F_\alpha$$

Rechazaríamos la hipótesis de igualdad de varianzas aceptando la alternativa , por tanto , varianza de X superior a varianza de Y

ejemplo 7.

Contrastar la hipótesis de que dos poblaciones tienen la misma dispersión con un nivel de significación del 1 % y sabiendo que la desviación típica de una de una muestra realizada sobre la primera población era 12 con un tamaño muestral de 25 y que en una muestra sobre la segunda de tamaño 30 la desviación típica resultó ser 7 .

Considérese que ambas poblaciones son normales.

El contraste sería :

$$H_0 : \sigma_x^2 = \sigma_y^2$$
$$H_1 : \sigma_x^2 > \sigma_y^x$$

conocemos que el nivel de significación es 0.01

y :

$S_x^2 = 144; S_y^2 = 49; n_x = 25; n_y = 30$

tomando cómo X la de mayor varianza muestral el estadístico T será :

T= $\dfrac{n_x S_x^2 (n_y - 1)}{n_y S_y^2 (n_x - 1)} = \dfrac{25 \bullet 144 \bullet 29}{30 \bullet 49 \bullet 24} = 2.95918$

el valor de $F_{24,29,(\alpha=0.01)} = 2.49$ dado que T > Fα 2.95918>2.49 rechazamos igualdad de varianzas aceptando que la primera tiene una varianza mayor .

9.-Contraste de incorrelación

Dadas dos poblaciones normales sobre las que se realiza un muestreo conjunto de tamaño n puede demostrarse con relativa sencillez que, bajo la hipótesis

H_0 ; ρ =0 es decir, bajo la hipótesis de que ambas poblaciones están incorrelacionadas y por tanto que los coeficientes de regresión de las dos regresiones entre las poblaciones [Y/X: Y= α + β X ; X/Y: X= α '+β 'Y] , β y β ' son

nulos ; la distribución del estadístico:

$$F = \frac{r^2}{1-r^2}(n-2) \Rightarrow F_{1,(n-2)}$$

(Sigue una distribución F de Snedecor con 1 g.l. en el numerador y n - 2 g.l. en el denominador.)

Donde **r** es el coeficiente de correlación muestral y **n** el tamaño muestral
De otra forma el estadístico puede plantearse :

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \Rightarrow t_{n-2}$$

(sigue una distribución t de Student con n-2 g.l.partiendo de estos argumentos planteamos el contraste bilateral de manera que:

$H_0: \rho = 0$
$H_1: \rho \neq 0$

para un determinado nivel de significación α : diseñaremos el contraste de la siguiente manera :
En el caso de utilización de la F de Snedecor sería:

si

$$F = \frac{r^2}{1-r^2}(n-2) < F_{1,(n-2),\alpha}$$

No rechazaríamos la hipótesis nula , por tanto la incorrelación.

Si

$$F = \frac{r^2}{1-r^2}(n-2) > F_{1,(n-2),\alpha}$$

rechazaríamos H_0 por lo que aceptaríamos que están correladas.

en el caso de utilizar la t de Student sería:

si

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \in \left[-t_{\alpha/2}; t_{\alpha/2}\right]$$

no rechazaríamos incorrelación

si

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \notin \left[-t_{\alpha/2}; t_{\alpha/2}\right]$$

rechazaríamos incorrelación , aceptando que están correladas

También podemos plantearnos la realización de un contraste de una cola, en el que es conveniente la utilización del estadístico relacionado con la T de student dado que plantea, por la simetría de la distribución, la posibilidad de trabajar con positivos y negativos , cuestión más acorde con el problema de la correlación que así lo exige. Tendremos ,por tanto , dos posibles opciones:

Cuando el coeficiente de correlación muestral sea positivo

el contraste supondrá establecer la hipótesis de incorrelación ante la alternativa de que el coeficiente de correlación poblacional es positivo. Así:

$$H_0: \rho = 0$$
$$H_1: \rho > 0$$

estableciendo que si $T < t_\alpha$ aceptaremos la H_0

y si T > t_α rechazaremos la H_0 ; aceptando correlación negativa

Cuando el coeficiente de correlación muestral sea negativo

el contraste supondrá establecer la hipótesis de incorrelación ante la alternativa de que el coeficiente de correlación poblacional es negativo. Así:

$$H_0: \rho = 0$$
$$H_1: \rho < 0$$

estableciendo que si T > t_α aceptaremos la H_0

y si T < t_α rechazaremos la H_0 ; aceptando correlación positiva

ejemplo 8

Contrastar la existencia de correlación positiva entre la renta y el consumo de un país, con un nivel de significación del 5 % , si de los datos de diez años se desprende que el coeficiente de correlación de dichos años fue de r=0.87634

Nos planteamos el contraste de la hipótesis de incorrelación frente a la alternativa de que exista correlación positiva:

Si T < t_α aceptaremos que existe incorrelación (o que podría ser, incluso, negativa)
Si T > t_α rechazamos la incorrelación ,el dato muestral es significativo, aceptamos la correlación positiva.

Según la información que poseemos , tendremos que :

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} = \frac{0.87634}{\sqrt{1-0.8734^2}}\sqrt{10-2} = 5.681$$

dado que el valor de t_α con 8 grados de libertad y para alfa 0.05 es 1.86 por tanto el estadístico es mayor. Aceptaremos que existe correlación positiva dado que rechazamos la hipótesis nula de incorrelación.