CONTRASTE DE HIPÓTESIS

Índice:

1. Contraste de hipótesis.

Antes de comenzar con un ejemplo que podamos ilustrar en que consiste un contraste de hipótesis, veamos algunas definiciones que vamos a utilizar a lo largo de este tema

- **Contraste de hipótesis**.- Es un procedimiento estadístico mediante el cual se investiga la verdad o falsedad de una hipótesis acerca de una población o poblaciones.
- **Hipótesis nula** H_0 .- Es la hipótesis que se fórmula y que se quiere contrastar; es, por tanto, la hipótesis que se acepta o se rechaza como consecuencia del contraste.
- **Hipótesis alternativa** H_a .- Es cualquier otra hipótesis que difiere o es contraria a la hipótesis nula, de forma que si se rechaza H_0 se acepta H_a y si se acepta H_a se rechaza H_0 .
- **Estadístico del contraste**.- Es función de valores muestrales. Es una variable aleatoria que sigue una distribución en el muestreo. Toma un valor para cada muestra.
- **Región de aceptación**.- Es la formada por el conjunto de puntos tales que los valores del estadístico del contraste nos lleva a aceptar la hipótesis nula.
- **Región de crítica**.- Es la formada por el conjunto de puntos tales que los valores del estadístico del contraste nos lleva a rechazar la hipótesis nula.
- **Contraste bilateral**.- Cuando la región crítica está formada por dos conjuntos de puntos disjuntos.
- **Contraste unilateral**.- Cuando la región crítica está formada por un solo conjunto de puntos disjuntos.

 # Ejemplo.- Si en una investigación anterior se ha estimado que la temperatura media del cuerpo humano para los adultos sanos se distribuye según una $N(\mu,\sigma)=N(37^{o}C\,;0{,}9\,^{o}C)$.

Si denominamos,

Hipótesis nula: $H_0\ :\ \mu=37^{o}C$.

Hipótesis alternativa: $H_a\ :\ \mu\neq 37^{o}C$.

Para contrastar la hipótesis nula, elegimos una muestra aleatoria de 10 personas adultas sanas, obteniendo las siguientes temperaturas en grados centígrados:

37,7; 36,7; 37,5; 37; 37,9; 37,6; 37,1; 37; 36,8; 37,1

Si calculemos la media muestral y obtenemos $\bar{x}=37{,}24\,^{o}C$ *, que es un valor particular del estadístico* $\bar{X}$ *, que sigue una distribución* $N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)=N\left(37,\frac{0{,}9}{\sqrt{10}}\right)=N(37,0{,}285)$.

Por tanto, la diferencia $|\mu-\bar{x}|=|37^{o}\,C-37{,}24\,^{o}C|=0{,}24\,^{o}C$ *, puede ser debida al azar, en cuyo caso diremos que* ***no es significativa****, o puede ser debida o otras causas, en cuyo caso diremos que* ***es significativa****. ¿Como podemos saber cuando es significativa o no?.*

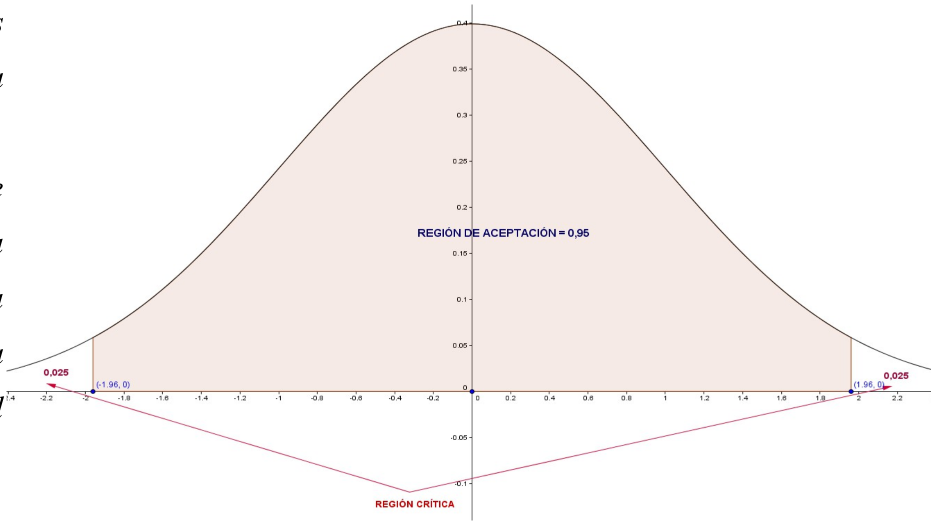
Fijamos previamente el nivel de confianza que queremos, por ejemplo $1-\alpha=0{,}95$ *, y tipificando la variable* $\bar{X}$ *, para el valor particular* $\bar{x}$ *, obtenemos* $\frac{37{,}24-37}{0{,}285}=0{,}93$.

Y existirá un valor $z_{\frac{\alpha}{2}}$ tal que $P\left(-z_{\frac{\alpha}{2}}\le\frac{37{,}24-37}{0{,}285}\le z_{\frac{\alpha}{2}}\right)=0{,}95$ *, que equivale a* $P\left(0{,}93\le z_{\frac{\alpha}{2}}\right)=0{,}975$ *, que mirando en la tabla normal* $N(0{,}1)$ *se obtiene* $z_{\frac{\alpha}{2}}=1{,}96$.

Luego, aceptaremos la hipótesis nula si el estadístico del contraste $\frac{37{,}24-37}{0{,}285}=0{,}93$ *cae dentro del intervalo* $\left(-z_{\frac{\alpha}{2}},z_{\frac{\alpha}{2}}\right)$ *, es decir, del intervalo* $(-1{,}96\,,1{,}96)$.

Como 0,93 cae dentro de dicho intervalo, aceptaremos la hipótesis nula.

Es decir, la muestra es realmente compatible con la población en el 95% de los casos. O también decimos, que a partir de los datos muestrales se acepta la hipótesis de que la temperatura media del cuerpo humano para adultos sanos es 37º C, con un nivel de confianza del 95%



Los valores críticos α mas usuales para contrastar hipótesis son

Nivel de significación α	***Valores críticos: contrastes unilaterales*** z_{α}	***Valores críticos: contrastes bilaterales*** $z_{\frac{\alpha}{2}}$
0,100	$\pm1{,}28$	$\pm1{,}645$
0,050	$\pm1{,}64$	$\pm1{,}96$
0,010	$\pm2{,}33$	$\pm2{,}58$
0,005	$\pm2{,}58$	$\pm2{,}81$
0	$\pm2{,}88$	$\pm3{,}08$

2. Errores de tipo I y tipo II

Todo contraste de hipótesis lleva a aceptar o rechazar la hipótesis nula planteada. Ahora bien, pueden ocurrir los siguientes casos:

1º Aceptar la hipótesis nula siendo verdadera. Ésta es una decisión correcta.

2º Rechazar la hipótesis nula siendo falsa. Esta es otra decisión correcta.

3º Rechazar la hipótesis nula siendo verdadera. Cometemos un **error de tipo I**.

4º Aceptar la hipótesis nula siendo falsa. Cometemos un **error de tipo II**.

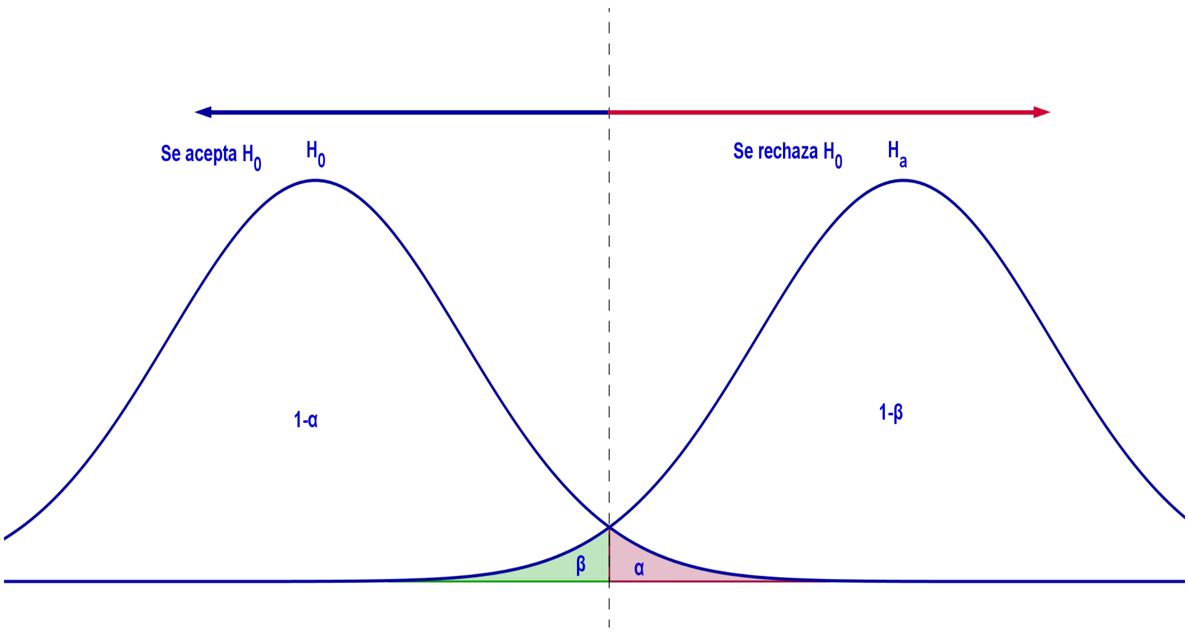
La probabilidad P, de cometer el error de tipo I se llama **nivel de significación** y lo representamos por α. Este determinado valor es conocido de antemano, ya que se debe de fijar antes de efectuar el contraste.

La probabilidad P, de cometer el error de tipo II la denominamos por β. Y la probabilidad $1-\beta$ se llama nivel **potencia del contraste,** que es la probabilidad de rechazar la hipótesis nula cuando es falsa.

Podemos resumir este tipo de errores en la siguiente tabla

	H_0 : verdadera	H_0 : falsa
Aceptar H_0	Decisión correcta $P=1-\alpha$	Cometer error de tipo II $P=\beta$
Rechazar H_0	Cometer error de tipo I $P=\alpha$	Decisión correcta $P=1-\beta$ Potencia del contraste

Y que podemos representar mediante las áreas correspondientes



3. Contraste para el parámetro p de una distribución binomial

Supongamos que partimos de una distribución binomial de parámetros $B(n,p)$ y que queremos contrastar el valor $p=p_0$, fijado un nivel de significación α, si elegimos una muestra para la cual un valor determinado de p es $\hat{p}$. Entonces, se tiene

H_0	H_a	Tipo de contraste	Estadístico	Región de aceptación
$p=p_0$	$p\neq p_0$	**bilateral**	$Z=\dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0.(1-p_0)}{n}}}$ que sigue una distribución $N(0,1)$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$
$p\leq p_0$	$p>p_0$	**unilateral**		$(-\infty, z_\alpha)$
$p\geq p_0$	$p<p_0$	**unilateral**		$(-z_\alpha, +\infty)$

 # Ejemplo.- El ayuntamiento de una ciudad afirma que el 65% de los accidentes juveniles de los fines de semana son debidos al alcohol. Un investigador decide contrastar dicha hipótesis, para lo cual toma una muestra formada por 35 accidentes y observa que 24 de ellos han sido debidos al alcohol. ¿Qué podemos decir sobre la afirmación del ayuntamiento?

Efectuamos los siguientes pasos

1º PASO.- formulamos las hipótesis

$H_0: p=0{,}65=p_0$

$H_a: p\neq 0{,}65$

2º PASO.- elegimos un nivel de significación α *, por ejemplo* $\alpha=0{,}01$.

3º PASO.- elegimos el estadístico de contraste $Z=\dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0.(1-p_0)}{n}}}$

4º PASO.- elegimos la región de aceptación. Que como es un contraste bilateral, para $\alpha=0{,}01$ *, se tiene el intervalo* $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)=(-2{,}58;2{,}58)$

5º PASO.- Calculamos

$\hat{p}=\dfrac{24}{35}=0{,}686$; $n=35$

$$z=\frac{0{,}686-0{,}65}{\sqrt{\dfrac{0{,}65.(1-0{,}65)}{35}}}=0{,}444$$

6º PASO.- Como $0{,}444\in(-2{,}58;2{,}58)$ *, se acepta la hipótesis nula y decimos que el nivel* 1 *se acepta la proporción de accidentes debidos al alcohol es del* 65

 # Ejemplo.- Un entrenador asegura que sus jugadores en los entrenamientos encestan más del 92% de los tiros libres. Con el fin de contrastar esta afirmación de ha elegido aleatoriamente una muestra de 60 lanzamientos, de los que 42 han entrado en la canasta. Estos resultados, ¿ponen en duda al entrenador o no?

Efectuamos los siguientes pasos

1º PASO.- formulamos las hipótesis

$H_0 : p \geq 0{,}92 = p_0$

$H_a : p < 0{,}92$

2º PASO.- elegimos un nivel de significación α *, por ejemplo* $\alpha = 0{,}1$.

3º PASO.- elegimos el estadístico de contraste $Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 . (1 - p_0)}{n}}}$

4º PASO.- elegimos la región de aceptación. Que como es un contraste unilateral, para $\alpha = 0{,}1$ *, se tiene el intervalo* $(-z_\alpha , +\infty) = (-1{,}288 ; +\infty)$

5º PASO.- Calculamos

$\hat{p} = \dfrac{42}{60} = 0{,}7$; $n = 60$

$$z = \frac{0{,}7 - 0{,}92}{\sqrt{\dfrac{0{,}92 . (1 - 0{,}92)}{60}}} = -6{,}28$$

6º PASO.- Como $-6{,}38 \notin (-1{,}28 ; +\infty)$ *, se rechaza la hipótesis nula; por tanto, al nivel de 10% se pone en duda la afirmación del entrenador.*

4. Contraste para la media de una población normal

Si σ es conocida

Supongamos que partimos de una distribución $N(\mu , \sigma)$, donde σ es conocida, y queremos contrastar el valor $\mu = \mu_0$. Para ello, fijado un determinado nivel de significación α , elegimos una muestra de tamaño n para la cual la media muestral $\bar{x}$ se tiene

H_0	H_a	**Tipo de contraste**	**Estadístico**	**Región de aceptación**
$\mu = \mu_0$	$\mu \neq \mu_0$	**bilateral**	$Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ **que sigue una distribución** $N(0,1)$	$(-z_{\frac{\alpha}{2}} , z_{\frac{\alpha}{2}})$
$\mu \leq \mu_0$	$\mu > \mu_0$	**unilateral**		$(-\infty , z_\alpha)$
$\mu \geq \mu_0$	$\mu < \mu_0$	**unilateral**		$(-z_\alpha , +\infty)$

<u>Si σ es desconocida y el tamaño de la muestra n es grande ($n \geq 30$)</u>

Sustituiremos la varianza σ^2 por la cuasivarianza muestral $\hat{s}^2$. Por tanto el estadístico del contraste $z = \dfrac{\bar{x} - \mu_0}{\dfrac{\hat{s}}{\sqrt{n}}}$ *sigue* $N(0,1)$.

 # Ejemplo.- Se cree que el tiempo medio de ocio que dedican al día los estudiantes de Bachillerato sigue una distribución normal de media 350 minutos y desviación típica poblacional de 60 minutos. Para contrastar esta hipótesis, se toma una muestra aleatoria formada por 100 alumnos, y se observa que el tiempo medio es 320 minutos. ¿Qué se puede decir de esta afirmación al nivel 10%?

1º PASO.- formulamos las hipótesis

$H_0 : \mu = 350 = \mu_0$

$H_a : \mu \neq 350$

2º PASO.- elegimos un nivel de significación α *, por ejemplo* $\alpha = 0{,}1$.

3º PASO.- elegimos el estadístico de contraste $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

4º PASO.- elegimos la región de aceptación. Que como es un contraste bilateral, para $\alpha = 0{,}1$ *, se tiene el intervalo* $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right) = (-1{,}645 ; 1{,}645)$

5º PASO.- Calculamos

$\bar{x} = 320$; $n = 100$; $\sigma = 60$

$$z = \frac{320 - 350}{\dfrac{60}{\sqrt{100}}} = -5$$

6º PASO.- Como $-5 \notin (-1{,}645 ; 1{,}645)$ *, se rechaza la hipótesis nula; por tanto, al nivel del 10% se pone en duda que el tiempo medio diario dedicado al ocio por los alumnos de Bachillerato sea 350 minutos.*

 # Ejemplo.- Se quiere contrastar el contenido de azúcar de distintos cargamentos de remolacha. Se sabe que el contenido medio de azúcar para remolacha de regadío es de 18% y con media superior para el de secano, siendo la desviación típica 6% para ambos casos. Se toma una muestra de 20 cargamentos. ¿Qué valor de la media permitirá tomar la decisión si es de secano o

de regadío al nivel del 5%?

1º PASO.- formulamos las hipótesis

$H_0 : \mu \leq 18 = \mu_0$

$H_a : \mu > 18$

2º PASO.- elegimos un nivel de significación α *, por ejemplo* $\alpha = 0{,}05$.

3º PASO.- elegimos el estadístico de contraste $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

4º PASO.- elegimos la región de aceptación. Que como es un contraste unilateral, para $\alpha = 0{,}05$ *, se tiene el intervalo* $(-\infty , z_\alpha) = (-\infty ; 1{,}645)$

5º PASO.- Calculamos

$\bar{x} = ?$; $n = 20$; $\sigma = 6$

$$z = \frac{\bar{x} - 18}{\dfrac{6}{\sqrt{20}}}$$

6º PASO.- Como $z = \dfrac{\bar{x} - 18}{\dfrac{6}{\sqrt{20}}} < 1{,}645 \Rightarrow \bar{x} < 20{,}2\,\%$

Luego el 20,2 % es el punto crítico. Por tanto,

Si $\bar{x} < 20{,}2\,\% \Rightarrow$ *remolacha de regadío*

Si $\bar{x} \geq 20{,}2\,\% \Rightarrow$ *remolacha de secano*

5. Analogías entre el contraste de hipótesis e intervalos de confianza

Para ver la analogía entre el contraste de hipótesis y los intervalos de confianza, veamos el siguiente ejemplo.

 # Ejemplo.- Supongamos que el gremio de restaurantes de Madrid afirma que el precio medio del menú del día es 11 € y queremos contrastar esta hipótesis. Para lo cual efectuamos un contraste de hipótesis, mediante los siguiente pasos

1º PASO.- formulamos las hipótesis

$H_0 : \mu = 11\,€ = \mu_0$

$H_a : \mu \neq 11\,€$

2º PASO.- elegimos un nivel de significación α *, por ejemplo* $\alpha = 0{,}05$ *, de donde* $1 - \alpha = 0{,}95$.

3º PASO.- elegimos el estadístico de contraste $Z=\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ *que sigue una distribución* $N(0,1)$

4º PASO.- elegimos la región de aceptación. Que como es un contraste bilateral, para $\alpha=0{,}05$ *, se tiene el intervalo* $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)=(-1{,}96 ; 1{,}96)$

5º PASO.- Elegimos una muestra, por ejemplo de 40 restaurantes y hallamos el precio medio de la muestra y la desviación típica muestral; sean por ejemplo:

$\bar{x}=11{,}50\,€$; $n=40$; $s=1{,}2\,€$

Y efectuamos los cálculo para el caso particular de $\bar{X}$; $\bar{x}=11{,}50\,€$

$$z=\frac{11{,}50-11}{\frac{1{,}2}{\sqrt{40}}}=2{,}635$$

6º PASO.- Como $2{,}635 \notin (-1{,}96 ; 1{,}96)$ *, se rechaza la hipótesis nula; y decimos que "al nivel de significación del 5% existe evidencia suficiente de que el precio del menú del día es distinto de 11 €.*

Sin hallamos un intervalo de confianza para la media poblacional al nivel del 5 % para el enunciado anterior, será

$$IC=\left(\bar{x}\pm z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)=\left(11{,}5\pm 1{,}96\,.\frac{1{,}2}{\sqrt{40}}\right)=(11{,}13 ; 11{,}87)$$

Y dado que $\mu_0=11\,€\notin(11{,}13 ; 11{,}87)$ al nivel de significación del 5%, vemos que existe una relación entre el contraste de hipótesis y le intervalo de confianza.

El error de tipo I del contraste de hipótesis se corresponde con la probabilidad de que el intervalo de confianza no cubra el valor parámetro.

El error de tipo I del contraste de hipótesis se corresponde con la probabilidad de que el intervalo de confianza no cubra valores erróneos.