UNIDAD 4.- **INFERENCIA ESTADÍSTICA II**

1. ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Consideraremos una variable aleatoria X con una media μ desconocida y una desviación típica σ conocida (parámetros poblacionales).

Lo que deseamos es estimar el valor de μ a partir de la información de una muestra de tamaño n, con parámetros $\bar{x}$ y $\bar{\sigma}$ (parámetros de la muestra).

¿Qué es un intervalo de confianza?

Para estimar la media poblacional μ, vamos a dar un intervalo alrededor de la media muestral $\bar{x}$, que llamaremos *intervalo de confianza* $[\bar{x}-E, \bar{x}+E]$.

Observaciones:

- Cada posible muestra generará un intervalo de confianza.
- Los diferentes intervalos de confianza puede que contengan a la media poblacional o no.
- El valor E es el *error máximo* cometido si el intervalo de confianza contiene a la media poblacional (μ).

Llamaremos nivel de confianza $N_c = (1-\alpha)$, si el $(1-\alpha)\cdot 100\%$ de los intervalos de confianza contienen a la media poblacional o al parámetro poblacional que se está estimando. Al valor α se le denomina nivel de riesgo o nivel de significación y representa el porcentaje de intervalos de confianza que no contendrán a la media.

Por ejemplo si el nivel de confianza es del 95% ($1-\alpha = 0,95$), quiere decir que el 95% de las muestras generarán intervalos de confianza que contendrán a la media. Y en este caso, el nivel de riesgo o de significación es del 0,05% ($\alpha = 0,05$) y es el porcentaje de intervalos de confianza que no contendrán a la media

Así pues:

Dada una muestra de tamaño n, con una media $\bar{x}$, llamaremos *intervalo de confianza* con un nivel de confianza $1-\alpha$ a un intervalo alrededor de $\bar{x}$, $[\bar{x}-E, \bar{x}+E]$, tal que la media poblacional, μ, estará en ese intervalo en el $(1-\alpha)\cdot 100\%$ de las posibles muestras.

$$P(\overline{X}-E \le \mu \le \overline{X}+E) = 1-\alpha$$

Nota: es un error decir que dado un intervalo de confianza, éste contendrá a la media con una cierta probabilidad, porque una vez dado el intervalo, éste contiene la media o no la contiene.

Ejemplo:

Si tenemos una población de 60 individuos, con una media desconocida.

Si tomamos muestras de tamaño 5, tendríamos $60^5 = 777.600.000$ posibles muestras distintas.

Cada una de las muestras tendrá su media, $\bar{x}_i$, si fijamos un nivel de confianza del 95% y construimos los $777.600.000$ intervalos de confianza (cada uno alrededor de su media muestral), lo que sabremos es que el 95% de esos intervalos, es decir, 738.720.000 contendrán a la media de la población.

El paso siguiente es construir los intervalos de confianza, para ello nos va a hacer falta un valor numérico que se conoce como *VALOR CRÍTICO* que se representa por $z_{\alpha/2}$. Veamos cómo se calcula:

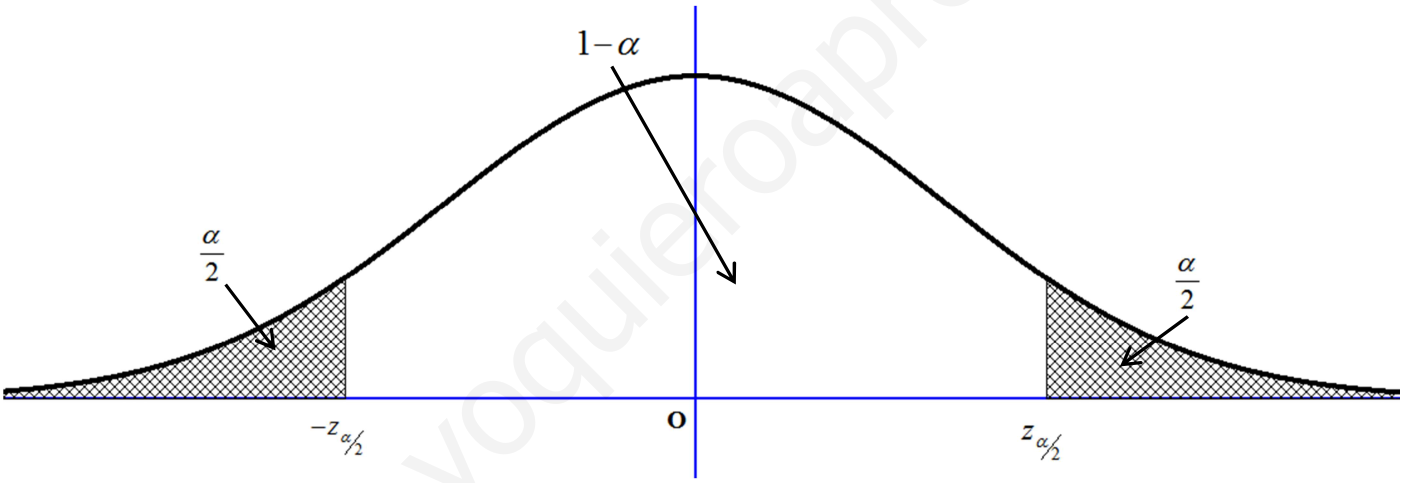
CÁLCULO DEL VALOR CRÍTICO $z_{\alpha/2}$

Dado un nivel de confianza $1-\alpha$ (o lo que es lo mismo, nivel de significación α), el número $z_{\alpha/2}$ es aquél nº real que en la distribución normal estándar, $N(0;1)$, deja a su derecha un área igual a $\frac{\alpha}{2}$. Es decir,

$$P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}$$

Como además, $P(Z \geq z_{\alpha/2}) = 1 - P(Z \leq z_{\alpha/2}) \Rightarrow 1 - P(Z \leq z_{\alpha/2}) = \frac{\alpha}{2} \Rightarrow \boxed{P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}}$ que es la usada para calcular $z_{\alpha/2}$

Gráficamente podemos observar a que corresponde cada expresión



Donde se observa fácilmente que:

$$P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2} \qquad P(Z \leq -z_{\alpha/2}) = \frac{\alpha}{2} \qquad P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$$

Esta última expresión, $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$, se usa para calcular los intervalos de confianza

Ejemplos:

- Para un nivel de confianza del 90%, $z_{\alpha/2} = 1,645$.

  ya que: $1-\alpha = 0,90$ y por tanto $\alpha = 0,10$, $\frac{\alpha}{2} = 0,05$ y $1 - \frac{\alpha}{2} = 0,95$.

  $$P(Z \leq z_{\alpha/2}) = 0,95 \Rightarrow z_{\alpha/2} = 1,645$$

- Para un nivel de confianza del 95%, $z_{\alpha/2} = 1,96$.

ya que: $1-\alpha=0{,}95$ y por tanto $\alpha=0{,}05$, $\frac{\alpha}{2}=0{,}025$ y $1-\frac{\alpha}{2}=0{,}975$.

$$P(Z\le z_{\alpha/2})=0{,}975\Rightarrow z_{\alpha/2}=1{,}96$$

- Para un nivel de confianza del 99%, $z_{\alpha/2}=2{,}575$

ya que: $1-\alpha=0{,}99$ y por tanto $\alpha=0{,}01$, $\frac{\alpha}{2}=0{,}005$ y $1-\frac{\alpha}{2}=0{,}995$

$$P(Z\le z_{\alpha/2})=0{,}995\Rightarrow z_{\alpha/2}=2{,}575$$

CÁLCULO DEL INTERVALO DE CONFIANZA

La *distribución muestral de medias*, $\overline{X}$, sigue una distribución normal $\overline{X}=N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$.

Por tanto, $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sigue una distribución $N(0,1)$.

Según se ha definido el valor $z_{\alpha/2}$ se tiene que: $P(-z_{\alpha/2}\le Z\le z_{\alpha/2})=1-\alpha$

Sustituyendo:

$$P\left(-z_{\alpha/2}\le\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le z_{\alpha/2}\right)=1-\alpha$$

Multiplicando todos los miembros de las desigualdades por $\frac{\sigma}{\sqrt{n}}$

$$P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\le\overline{X}-\mu\le z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Restando a todos los miembros de las desigualdades $\overline{X}$

$$P\left(-\overline{X}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\le-\mu\le-\overline{X}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Cambiando el signo de los miembros de las desigualdades (cambia el sentido de la desigualdad).

$$P\left(\overline{X}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ge\mu\ge\overline{X}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Ó lo que es lo mismo

$$P\left(\overline{X}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Si $\overline{x}$ es una media muestral, diremos que el intervalo:

$$\boxed{\left(\overline{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)}$$

es un intervalo de confianza para la media con un nivel de confianza del $(1-\alpha)\cdot 100\%$

Esto quiere decir, que el $(1-\alpha)\cdot 100\%$ de los intervalos construidos de esta manera contendrán a la media, es decir, el $(1-\alpha)\cdot 100\%$ de las muestras posibles generarán intervalos de confianza que contendrán a la media.

<u>Nota importante</u>: <u>No es correcto</u> decir "el intervalo de confianza dado contiene a la media con una probabilidad $(1-\alpha)$", porque una vez dado un intervalo, éste o contiene a la media o no la contiene.

Para la distribución de proporciones muestrales el proceso es el mismo (en lugar de $\overline{x}$ usaremos $\hat{p}$) y el intervalo de confianza que queda es:

$$\boxed{\left(\hat{p}-z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p}+z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)}$$

Resumiendo, para encontrar el intervalo de confianza en el que se encuentra un parámetro poblacional (media o proporción) con un nivel de confianza $(1-\alpha)\cdot 100\%$, hemos de seguir los siguientes pasos:

- Determinar el estimador muestral (media o proporción de la muestra)
- Determinar la desviación típica correspondiente al estimador
- Determinar el valor crítico $z_{\alpha/2}$ correspondiente a ese nivel de confianza $N_c=(1-\alpha)\cdot 100\%$
- El intervalo de confianza buscador es:

(estimador - $z_{\alpha/2}$ ·desviación típica, estimador + $z_{\alpha/2}$ ·desviación típica

<u>Tabla resumen</u>:

Parámetro	Intervalo de confianza
Media μ	$\left(\overline{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$
Proporción p	$\left(\hat{p}-z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}\ ,\ \hat{p}+z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}\right)$

<u>NOTA</u>: Recordemos que hemos supuesto que la media de la población (μ) era desconocida, la desviación típica (σ) era conocida y queríamos estimar la media, pero es razonable pensar que también la desviación típica sea desconocida, ¿qué hacer en esos casos?, pues para n suficientemente grande $(n\geq 30)$, se puede sustituir σ por la desviación típica de la muestra $\overline{\sigma}$ en los cálculos de los intervalos de confianza y la estimación sigue siendo satisfactoria.
Análogamente podemos decir de la proporción p en una población, pues ésta no será conocida. En su lugar usaremos la proporción de la muestra $\hat{p}$, con un tamaño de la muestra conveniente

Ejemplo: Tenemos una población con una desviación típica $\sigma = 2$. Tomamos una muestra de 100 individuos y obtenemos una media $\bar{x} = 8$. Queremos dar el intervalo de confianza para la media con un nivel de confianza del 95%.

Así: $1-\alpha = 0{,}95$; por tanto $\alpha = 0{,}05$ y $\frac{\alpha}{2} = 0{,}025$

$$P(Z \le z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0{,}975 \quad \Rightarrow \quad z_{\alpha/2} = 1{,}96$$

Por lo que el intervalo de confianza es: $\left(8 - 1{,}96 \cdot \frac{2}{\sqrt{100}}, 8 + 1{,}96 \cdot \frac{2}{\sqrt{100}}\right) = (7{,}608 ; 8{,}392)$

No sé si este intervalo de confianza contiene a la media, pero sé que el 95% de los intervalos de confianza creados de esta manera contienen a la media.

NOTA:

La tabla siguiente recoge los valores críticos $Z_{\frac{\alpha}{2}}$ más utilizados:

%	$1-\alpha$	α	$\frac{\alpha}{2}$	$Z_{\frac{\alpha}{2}}$
90 %	0,9	0,1	0,05	1,645
95 %	0,95	0,05	0,025	1,96
99 %	0,99	0,01	0,005	2,575

Ejemplo: Si conocemos que la desviación típica para la altura de las chicas de 18 años en Aragón es 10 cm. Supongamos que hemos tomado una muestra de 100 chicas sale que la altura media de la muestra: $\overline{X} = 170$. Hallar los intervalos de confianza para la altura media de las chicas de 18 años en Aragón para un nivel de confianza del 90%, 95% y 99%

Nivel de confianza del 90%: $(170 - \frac{10}{\sqrt{100}} \cdot 1{,}645 \ , \ 170 + \frac{10}{\sqrt{100}} \cdot 1{,}645)$ = (168,355 , 171,645)

Nivel de confianza del 95%: $(170 - \frac{10}{\sqrt{100}} \cdot 1{,}96 \ , \ 170 + \frac{10}{\sqrt{100}} \cdot 1{,}96)$ = (168,04, 171,96)

Nivel de confianza del 99%: $(170 - \frac{10}{\sqrt{100}} \cdot 2{,}575 \ , \ 170 + \frac{10}{\sqrt{100}} \cdot 2{,}575)$ = (167,425 , 172,575)

Ejemplo: Tomada, al azar, una muestra de 120 estudiantes de una Universidad, se encontró que 54 de ellos hablaban inglés. Halle, con un nivel de confianza del 90%, un intervalo de confianza para estimar la proporción de estudiantes que hablan el idioma inglés entre los estudiantes de esa Universidad.

Como vemos se trata de una muestra donde estudiamos la característica hablar inglés y no conocemos la proporción de la población. Por ello usaremos la proporción de la muestra $\hat{p} = \frac{54}{120} = 0{,}45$ y por tanto $\hat{q} = 1 - 0{,}45 = 0{,}55$. Tenemos que

$$1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow \frac{\alpha}{2} = 0{,}05 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}95 \Rightarrow P\left(Z \le z_{\alpha/2}\right) = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}645$$

Aplicando la fórmula $\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \ , \ \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right)$ =

$$\left(0{,}45 - 1{,}645 \cdot \sqrt{\frac{0{,}45 \cdot 0{,}55}{120}} \ , \ 0{,}45 + 1{,}645 \cdot \sqrt{\frac{0{,}45 \cdot 0{,}55}{120}}\right) = (0{,}45 - 0{,}0747, \ 0{,}45 + 0{,}0747) = (0{,}3753 \ , \ 0{,}5247)$$

2. TAMAÑO DE LAS MUESTRAS. ERROR MÁXIMO ADMISIBLE

Al estudiar los intervalos de confianza hemos visto que su amplitud depende del siguiente factor:

$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ para la media ó $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$ para la proporción.

Es decir, cuánto mayor sea el tamaño de la muestra más fiable es el intervalo elegido, pero también es mucho más caro el estudio. Las expresiones anteriores nos dan el error máximo que cometemos.

Definición: El error máximo admisible para la estimación de medias o proporciones viene dado, respectivamente, por:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \text{ para la media ó } E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \text{ para la proporción}$$

De estas expresiones deducimos que:

- A mayor nivel de confianza, mayor $z_{\alpha/2}$, y por tanto mayor es el error cometido
- A mayor tamaño muestral, menor es el error cometido

TAMAÑO DE LAS MUESTRAS

El valor $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ es el error máximo cometido al hacer una estimación de la media.

Si fijamos de antemano un error máximo E, es decir, conocemos E, debe ocurrir que

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq E$$

Despejando:

$$z_{\alpha/2} \cdot \sigma \leq E \cdot \sqrt{n} \Rightarrow \frac{z_{\alpha/2} \cdot \sigma}{E} \leq \sqrt{n} \Rightarrow \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 \leq n$$

Luego:

$$\boxed{n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2}$$

De forma análoga, para la proporción si despejamos en $E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$, nos queda que

$$\boxed{n \geq \frac{{z_{\alpha/2}}^2 \cdot \hat{p} \cdot \hat{q}}{E^2}}$$

No hace falta aprenderse estas fórmula, simplemente sustituyendo cada factor por su valor resulta más fácil despejar

Ejemplo: Tenemos una población con una desviación típica $\sigma = 2$. Queremos tomar una muestra para estimar la media con un nivel de confianza del 95%, ¿cuál debe ser el tamaño mínimo de la muestra para cometer un error menor de 0,4?

Así: $1 - \alpha = 0{,}95$; por tanto $\alpha = 0{,}05$ y $\frac{\alpha}{2} = 0{,}025 \Rightarrow P(Z \leq z_{\alpha/2}) = 0{,}975 \quad \Rightarrow \quad z_{\alpha/2} = 1{,}96$

Como se ha de cumplir que

$$z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\le E \Rightarrow 1{,}96\cdot\frac{2}{\sqrt{n}}\le 0{,}4 \Rightarrow \frac{3{,}92}{\sqrt{n}}\le 0{,}4 \Rightarrow \frac{3{,}92}{0{,}4}\le\sqrt{n}\Rightarrow\sqrt{n}\ge 9{,}8\Rightarrow n\ge 96{,}04$$

Luego la muestra debe ser al menos de 97 individuos.

Ejemplo: Deseamos conocer el nº de personas mayores de edad que sería necesario incluir en una muestra nacional con un error absoluto de 0,04 y un nivel de confianza del 99,73%. Se dispone de un valor de $p=0,45$

Para ese nivel de confianza,

$$1-\alpha=0{,}9973\Rightarrow\alpha=0{,}0027\Rightarrow\frac{\alpha}{2}=0{,}00135\Rightarrow 1-\frac{\alpha}{2}=0{,}99865\Rightarrow P\left(Z\le z_{\alpha/2}\right)=0{,}99865\Rightarrow z_{\alpha/2}=3$$

Aquí aplicamos la fórmula $n\ge\dfrac{z_{\alpha/2}{}^{2}\cdot\hat{p}\,\hat{q}}{E^2}\,n\ge\dfrac{3^2\cdot 0.45\cdot 0{,}55}{0{,}04^2}\Rightarrow n\ge 1392{,}19$

Se necesita para la muestra al menos 1393 personas

3. ESTADÍSTICA DEDUCTIVA. HIPÓTESIS ESTADÍSTICA

A partir de este punto vamos a trabajar la estadística deductiva, que se basa en tomar decisiones sobre la población a partir de los datos obtenidos de una muestra.
Trataremos el importante aspecto de la toma de decisiones, referida a decidir si un valor obtenido a partir de la muestra es probable que pertenezca a la población.
En general, la media (o proporción) en una muestra suele ser distinta a la media de la población, de la cual se extrae la muestra. Lo normal suele ser que tal diferencia entre la media muestral y poblacional sea pequeña y debida al azar, pero podría suceder que dicha diferencia no esté justificada por el azar y se deba a un cambio en la población, y debamos modificar los datos que conocemos previamente.
Ejemplos:

a) Hace algunos años, la media de estatura de los españoles adultos varones era de 170 cm y su desviación típica 9 cm. Pasado el tiempo, un muestreo realizado a 36 adultos da una medida de 172 cm. ¿Puede afirmarse que esa diferencia de 2 cm es debida al azar o realmente la estatura media ha aumentado?.

b) Supongamos que, respecto a una determinada ley, el 52% de los ciudadanos está en contra. Pasado el tiempo, una encuesta realizada a 400 personas indica que los ciudadanos en contra han descendido hasta el 49%.¿Ha cambiado realmente la opinión pública o tal resultado es debido al azar?.

Trataremos de utilizar los datos obtenidos en una muestra para tomar decisiones sobre la población.
Para ello, debemos realizar ciertos supuestos o conjeturas sobre las poblaciones. Estos supuestos, que pueden ser o no ciertos, se llaman hipótesis estadísticas.
Podemos, entonces, definir el *test de hipótesis* o *contraste de hipótesis* como el procedimiento estadístico mediante el cual se investiga la verdad o falsedad de una hipótesis acerca de una población o poblaciones.
Dichas hipótesis se formularán sobre la media poblacional μ o la proporción poblacional p.

Definiciones: Llamaremos hipótesis nula, y se representa por H_0, a la hipótesis que se formula y por tanto se quiere contrastar o rechazar, y llamaremos hipótesis alternativa, y se representa por H_1, a cualquier otra hipótesis que sea diferente de la formulada, y que sea contraria a H_0, de forma que la aceptación de la hipótesis nula H_0 implica el rechazo de la alternativa H_1 y viceversa, el rechazo de H_0 implica la aceptación de H_1.

En un problema de contraste de hipótesis, pues, siempre tiene que formularse una hipótesis nula H_0, y ha de ir acompañada de una alternativa, H_1 que es la que aspira a desplazar a la nula.
<u>*Ejemplo*</u>: Un investigador afirma que la temperatura del cuerpo humano en un adulto sano se distribuye según una normal de media μ = 37º C y desviación típica σ = 0,9º C. Formular la hipótesis nula y la hipótesis alternativa.
A la vista de los datos, el investigador afirma que la temperatura media del cuerpo humano es 37º, es decir la hipótesis o conjetura que formula es:
$H_0 : \mu = 37^{\text{o}}$ (hipótesis nula)
Como hipótesis alternativa, hemos de tomar aquella contraria a esta, que la media sea distinta de 37º C, es decir:
$H_1 : \mu \neq 37^{\text{o}}$ (hipótesis alternativa)

Si la hipótesis nula fuese del tipo $H_0 : \mu \geq \mu_0$ la hipótesis alternativa sería: $H_1 : \mu < \mu_0$.
Si la hipótesis nula fuese del tipo $H_0 : \mu \leq \mu_0$ la hipótesis alternativa sería: $H_1 : \mu > \mu_0$.

4. <u>CONTRASTE DE HIPÓTESIS</u>

Sabemos ya formular la hipótesis nula y la hipótesis alternativa. Lo que necesitamos ahora es un criterio para saber si debemos aceptar una u otra, es decir, ¿con cuál de las dos hipótesis nos quedamos?

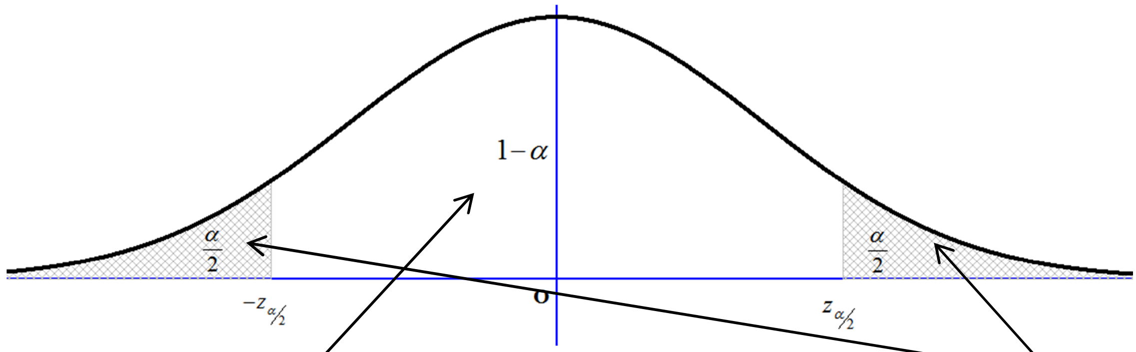
Al tener ya formulada la hipótesis nula, es necesario que las evidencias sean muy fuertes para rechazarla; es decir, puede que haya cambios debidos al azar, en cuyo caso el cambio no es significativo, y no cambiamos, pero puede que los cambios sean debidos a otras causas. En este último caso es cuando el cambio es significativo y rechazaremos.

Por lo tanto, lo primero que debemos hacer es fijar un cierto intervalo dentro del cual es normal que haya cambios, es decir, una región tal que si el parámetro se mantiene en dicho intervalo, nos seguimos quedando con H_0, pues esas pequeñas variaciones son debidas al azar. Ese intervalo o región se denomina <u>*región de aceptación*</u>, y será mayor o menor dependiendo del nivel de confianza que precisemos, $1-\alpha$.

La región que quede fuera de la región de aceptación indica que en este caso los cambios no se pueden atribuir al azar, y por tanto hemos de rechazar H_0 y aceptar H_1. Tal región se <u>llama *región crítica o de rechazo*</u>.

Llegados a este punto, hemos de distinguir entre dos tipos de contraste o test, que determinan la región de aceptación y la región de rechazo.

a) <u>**Contraste bilateral (o de dos colas)**</u>: En este caso la región de rechazo o región crítica está formada por dos conjuntos de puntos disjuntos. Dicho caso se presenta cuando la hipótesis nula es del tipo $H_0 : \mu = \mu_0$ (o bien $H_0 : p = p_0$) y la hipótesis alternativa, por tanto, es del tipo $H_1 : \mu \neq \mu_0$ (o bien $H_1 : p \neq p_0$).
La región crítica para un cierto nivel α sería, en la $N(0;1)$:



El intervalo $\left(-z_{\alpha/2}, z_{\alpha/2}\right)$ es la Región de Aceptación. La región sombreada es la Región crítica, formada por dos partes o colas.

La región de aceptación es el correspondiente intervalo de probabilidad para $\bar{x}$ o para $\hat{p}$

$\left(\mu_0 - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$ para la media ó $\left(p_0 - z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}, p_0 + z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right)$ para la proporción

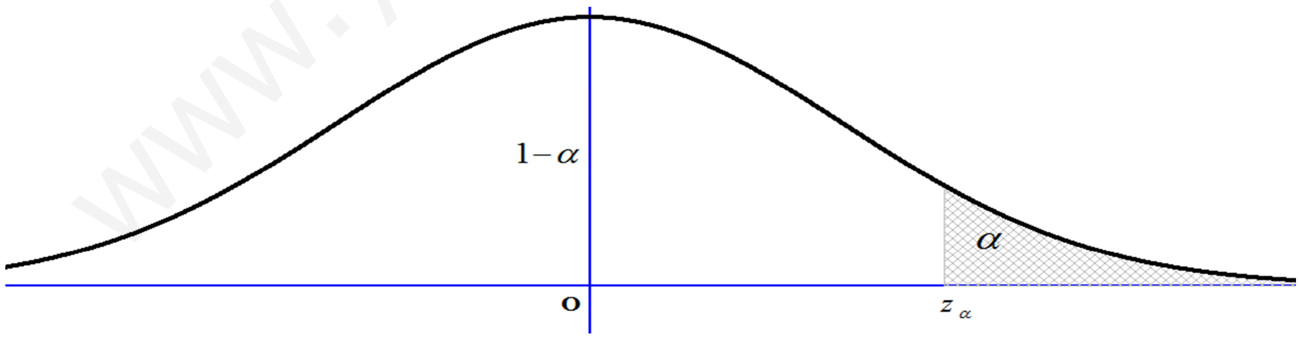
Las correspondientes regiones críticas serán:

$\left(-\infty, \mu_0 - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)\cup\left(\mu_0 + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, +\infty\right)$ para la media

$\left(-\infty, p_0 - z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right)\cup\left(p_0 + z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}, +\infty\right)$ para la proporción

b) **Contraste unilateral (o de una cola)**: En este caso la región crítica está formada por un sólo conjunto de puntos. El nivel de significación α se concentra sólo en una parte o cola.
Tenemos dos tipos de contraste unilateral:

Unilateral por la derecha o de cola derecha: Cuando la hipótesis nula es del tipo $H_0: \mu \le \mu_0$ (o bien $H_0: p \le p_0$) y la hipótesis alternativa, por tanto, es del tipo $H_1: \mu > \mu_0$ (o bien $H_1: \mu > p_0$). En este caso hay que calcular z_α (¡OJO! no es $z_{\alpha/2}$) para conocer tanto la región de aceptación como la crítica, como se puede observar en la figura:



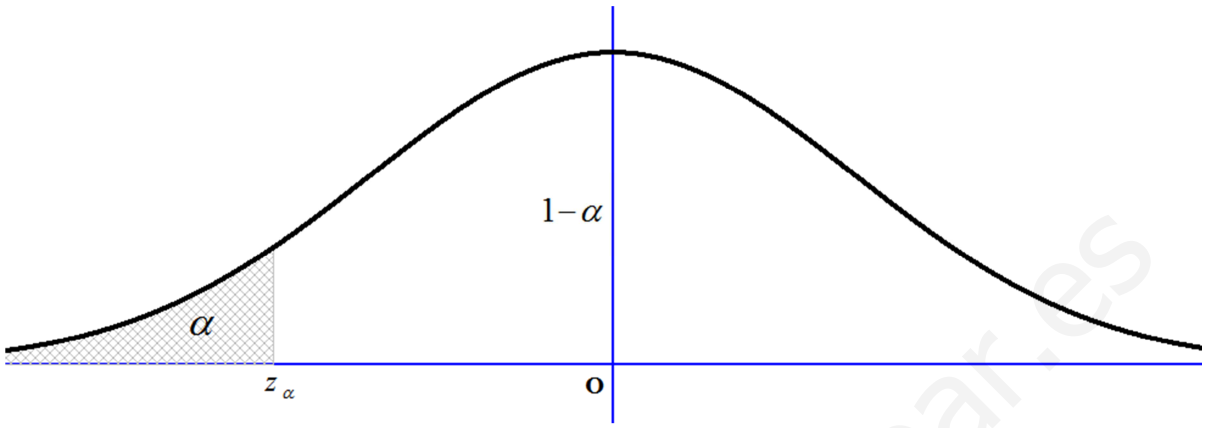
La región de aceptación es:

$\left(-\infty, \mu_0 + z_\alpha\cdot\frac{\sigma}{\sqrt{n}}\right)$ para la media o $\left(-\infty, p_0 + z_\alpha\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right)$ para la proporción.

La región crítica es:

$\left(\mu_0 + z_\alpha\cdot\frac{\sigma}{\sqrt{n}}, +\infty\right)$ para la media o $\left(p_0 + z_\alpha\cdot\sqrt{\frac{p_0\cdot q_0}{n}}, +\infty\right)$ para la proporción.

Unilateral por la izquierda o de cola izquierda: Cuando la hipótesis nula es del tipo $H_0 : \mu \geq \mu_0$ (o bien $H_0 : p \geq p_0$) y la hipótesis alternativa, por tanto, es del tipo $H_1 : \mu < \mu_0$ (o bien $H_1 : \mu < p_0$). En este caso hay que calcular z_α (¡OJO! no es $z_{\alpha/2}$) para conocer tanto la región de aceptación como la crítica, como se puede observar en la figura:



La región de aceptación es:

$\left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, +\infty\right)$ para la media o $\left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}, +\infty\right)$ para la proporción.

La región crítica es:

$\left(-\infty, \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$ para la media o $\left(-\infty, p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\right)$ para la proporción.

En todos los casos, conociendo el nivel de confianza $1-\alpha$ 1 – α, tendremos que determinar el valor $z_{\alpha/2}$ (para contrastes bilaterales) o bien z_α (para contrastes unilaterales), que separa las regiones de rechazo y aceptación.

ETAPAS PARA LOS CONTRASTES DE HIPÓTESIS

Los procedimientos seguidos en las pruebas de hipótesis correspondientes a las situaciones de decisión estadística se encuentran totalmente prefijados y se llevan a cabo en una serie de etapas que facilitan su comprensión, y que son:

1. Enunciar la hipótesis nula H_0 y la alternativa H_1. Deben ser excluyentes entre si.
   La hipótesis H_0 siempre lleva el valor igual.
   Es bilateral cuando es del tipo "=" y unilateral si es del tipo "≤" ó "≥"
2. Determinar el valor $z_{\alpha/2}$ (para contrastes bilaterales) o el valor z_α (para contrastes unilaterales) a partir del nivel de confianza o del nivel de significación.
3. Escribir las correspondientes regiones de aceptación y rechazo. Las de rechazo no se suelen usar
4. Usando el estadístico dado (la media muestral $\bar{x}$ o proporción muestral $\hat{p}$), comprobar si pertenece a la región de aceptación o no. Y dependiendo de ello, aceptar o rechazar H_0 y rechazar o aceptar H_1

5. CONTRASTE DE HIPÓTESIS PARA LA MEDIA

En estas situaciones se comienza con una suposición a priori del valor de la media de la población μ_0.
Después se utiliza una media $\bar{x}$ a partir de una muestra obtenida de forma aleatoria de la población para decidir si es probable la suposición acerca de μ. Veamos en una tabla las diferentes etapas en los diferentes tipos de contraste de hipótesis

	1ª Etapa		*2ª Etapa*	*3ª Etapa*	*4ª Etapa*
	H_0	H_1	Cálculo de $z_{\alpha/2}$ ó z_α	Determinación del intervalo de aceptación I	Decisión a partir del estadístico
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$z_{\alpha/2}$	$I = \left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$	Si $\bar{x} \in I$ se acepta H_0 Si $\bar{x} \notin I$ se rechaza H_0
Unilateral a la derecha	$\mu \leq \mu_0$	$\mu > \mu_0$	z_α	$I = \left(-\infty, \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$	
Unilateral a la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$	z_α	$I = \left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, +\infty\right)$	

Ejemplo: Una encuesta, realizada a 64 empleados de una fábrica, concluyó que el tiempo medio de duración de un empleo en la misma era de 6,5 años con una desviación típica de 4. ¿Sirve esta afirmación para aceptar, con un nivel de significación del 5%, que el tiempo medio de empleo en esa fábrica es menor o igual que 6? Justificar adecuadamente la respuesta.

El enunciado no puede ser más claro a la hora de determinar la hipótesis nula y la alternativa.
Queremos comprobar si el tiempo medio de empleo en esa fábrica es menor o igual que 6, luego la hipótesis nula será que dicho tiempo medio de empleo sea MENOR O IGUAL que 6, es decir:

$$H_0 : \mu \leq 6$$

Frente a la hipótesis alternativa:

$$H_1 : \mu > 6$$

Es claramente un contraste unilateral a la derecha o de cola derecha.
Dado que el nivel de significación es del 5 % tenemos que $\alpha = 0,05$ y con esto calculamos z_α:

$$P(Z > z_\alpha) = 0,05 \Rightarrow P(Z \leq z_\alpha) = 0,95 \Rightarrow z_\alpha = 1,645$$

Calculamos el intervalo de aceptación:

$$I = \left(-\infty, \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(-\infty, 6 + 1,645 \cdot \frac{4}{\sqrt{64}}\right) = (-\infty, 6,8225)$$

Como en la muestra resulta que $\bar{x} = 6,5 \in I = (-\infty, 6,8225)$, aceptamos la hipótesis nula $H_0 : \mu \leq 6$, la media es menor o igual a 6 años, con un 95 % de confianza

Ejemplo: (Selectividad 2012) Un índice para calibrar la madurez lectora de los alumnos de primaria se distribuye según una ley Normal con desviación típica 2. Elegida una muestra de 18 alumnos en un centro de primaria, se obtiene una media muestral de 10,8 en dicho índice. Mediante el uso de un contraste de hipótesis, ¿se puede aceptar, con un nivel de significación del 1%, la hipótesis nula de que la media del índice de madurez lectora de los alumnos de este centro no es inferior a 11?

El enunciado no puede ser más claro a la hora de determinar la hipótesis nula y la alternativa.
Queremos comprobar si la media del índice de madurez lectora en ese centro es mayor o igual (no inferior) que 11, luego la hipótesis nula será que dicho tiempo medio de empleo sea MAYOR O IGUAL que 11, es decir:

$$H_0 : \mu \geq 11$$

Frente a la hipótesis alternativa:

$$H_1 : \mu < 11$$

Es claramente un contraste unilateral a la izquierda o de cola izquierda.
Dado que el nivel de significación es del 1 % tenemos que $\alpha = 0,01$ y con esto calculamos z_α :

$$P(Z > z_\alpha) = 0,01 \Rightarrow P(Z \leq z_\alpha) = 0,99 \Rightarrow z_\alpha = 2,33$$

Calculamos el intervalo de aceptación:

$$I = \left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, +\infty \right) = \left(11 - 2,33 \cdot \frac{2}{\sqrt{18}}, +\infty \right) = (9,902, +\infty)$$

Como en la muestra resulta que $\bar{x} = 10,8 \in I = (9,902, +\infty)$, aceptamos la hipótesis nula $H_0 : \mu \geq 11$, la media es mayor o igual a 11, con un 99 % de confianza

6. CONTRASTES DE HIPÓTESIS PARA LAS PROPORCIONES

Se partirá en este caso de una suposición o porcentaje poblacional, p_0. Después utilizaremos la proporción de la muestra, $\hat{p}$ calculada a partir de una sola muestra, obtenida de forma aleatoria, para comprobar si es cierta la suposición sobre la proporción poblacional. Veamos en la tabla las diferentes etapas

	1ª Etapa		*2ª Etapa*	*3ª Etapa*	*4ª Etapa*
	H_0	H_1	Cálculo de $z_{\alpha/2}$ ó z_α	Determinación del intervalo de aceptación I	Decisión a partir del estadístico
Bilateral	$p = p_0$	$p \neq p_0$	$z_{\alpha/2}$	$I = \left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \right)$	Si $\hat{p} \in I$ se acepta H_0 Si $\hat{p} \notin I$ se rechaza H_0
Unilateral a la derecha	$p \leq p_0$	$p > p_0$	z_α	$I = \left(-\infty, p_0 + z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \right)$	
Unilateral a la izquierda	$p \geq p_0$	$p < p_0$	z_α	$I = \left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}, +\infty \right)$	

Ejemplo: (Selectividad 2011) Un estudio sociológico afirma que el 70% de las familias cena viendo la televisión. Se desea contrastar la veracidad de esta información, y para ello, se toma una muestra de 500 familias de las cuales 340 ven la televisión mientras cenan. Decida, mediante un contraste de hipótesis, si la afirmación es cierta con un nivel de significación de 0,01.

En este problema tenemos proporciones con los datos $p_0 = 0,7$ que es la proporción suposición. El nº de elementos de la muestra es $n = 500$, la proporción de la muestra $\hat{p} = \frac{340}{500} = 0,68$ y el nivel de significación es $\alpha = 0,01$.

Planteamos la hipótesis nula $H_0 : p = 0,7$ y la hipótesis alternativa $H_1 : p \neq 0,7$, que como observamos es un contraste bilateral o de doble cola.

Calculamos el valor crítico $Z_{\alpha/2}$, que verifica

$$P\left(Z > Z_{\alpha/2}\right) = \frac{\alpha}{2} \Rightarrow P\left(Z > Z_{\alpha/2}\right) = 0,005 \Rightarrow P\left(Z \leq Z_{\alpha/2}\right) = 0,995 \Rightarrow Z_{\alpha/2} = 2,575$$

El paso siguiente es calcular el intervalo de aceptación:

$$I = \left(0,7 - 2,575 \cdot \sqrt{\frac{0,7 \cdot 0,3}{500}}, 0,7 + 2,575 \cdot \sqrt{\frac{0,7 \cdot 0,3}{500}}\right) = (0,647 \ , \ 0,753)$$

Como el estadístico $\hat{p} = 0,68 \in I$, aceptamos como válida la hipótesis nula, y por tanto es cierto el estudio sociológico con un nivel de confianza del 99%

7. ERRORES EN LOS CONTRASTES DE HIPÓTESIS

Hay ocasiones en que la hipótesis nula, H_0, es cierta, pero a la vista de la muestra tengamos que rechazarla. En tal caso, estamos cometiendo un error. El error que consiste en rechazar la hipótesis nula cuando ésta es verdadera, se denomina error de tipo I.

Otro tipo de error puede ocurrir cuando, siendo H_0 falsa, las evidencias de la muestra, sin embargo, nos lleven a aceptarla. Este error, cometido al aceptar H_0 cuando ésta es falsa, se denomina error de tipo II.

Resumiendo:

		SITUACIÓN REAL	
DECISIONES		H_0 verdadera	H_1 verdadera
	Mantener H_0	Decisión correcta No se comete error Probabilidad = $1-\alpha$	Decisión incorrecta Error tipo II Probabilidad = β
	Mantener H_1	Decisión incorrecta Error tipo I Probabilidad = α	Decisión correcta No se comete error Probabilidad = $1-\beta$

- La probabilidad de cometer un error de tipo I viene dado por α, es decir, por el nivel de significación. Esta probabilidad no depende del tamaño de la muestra. $P(\text{error tipo I}) = \alpha$
- La probabilidad de cometer un error de tipi II viene dado por β, y depende del verdadero valor del parámetro y del tamaño de la muestra, siendo menor el error cuanto mayor es el tamaño de la muestra. $P(\text{error tipo II}) = \beta$

Utilizando probabilidades condicionadas:
$P(\text{error tipo I}) = P(\text{rechazar } H_0 / H_0 \text{ verdadera}) = \alpha$ y $P(\text{aceptar } H_0 / H_0 \text{ verdadera}) = 1-\alpha$
Por otra parte:
$P(\text{error tipo II}) = P(\text{aceptar } H_0 / H_0 \text{ falsa}) = \beta$ y $P(\text{rechazar } H_0 / H_0 \text{ falsa}) = 1-\beta$

A la probabilidad $1-\beta$ se le denomina *potencia del contraste*. Los contrastes que se utilizan son los que maximizan la potencia, es decir, minimizan el error de tipo II