Tema 14: Inferencia estadística

La **inferencia estadística** es el proceso de sacar conclusiones de la población basados en la información de una muestra de esa población.

1. Estimación de parámetros

Cuando desconocemos los parámetros que describen una población los podemos estimar en función de la información que proporciona una muestra.
Un **estimador** es una función de los datos de la muestra que permite dar un valor aproximado de un parámetro poblacional desconocido.
En la estimación puntual, a partir de una muestra, damos un único valor al estimador del parámetro desconocido.
Dada una muestra $(x_1, x_2, \dots, x_n)$ de una población X:

$$\begin{cases} X \hookrightarrow N(\mu, \sigma), \mu \text{ desconocida}: \bar{x} = \frac{1}{n} \cdot \sum x_i \text{ es estimador de } \mu \\ X \hookrightarrow B(n, p), p \text{ desconocido}: \tilde{p} = \frac{1}{n} \cdot \sum x_i \text{ es estimador de } p \end{cases}$$

En la estimación por intervalos, se construye un intervalo que contenga el parámetro desconocido.

2. Intervalo de confianza para la media muestral

Es lógico pensar que, si la altura de una población sigue una normal, de varianza conocida y de media desconocida, para estimar dicha media, pensemos en tomar la media de las medidas que hemos conseguido con muestras..
Partimos de una población que sigue una normal $N(\mu, \sigma)$ con desviación típica conocida, y queremos estimar, mediante un intervalo, el parámetro μ. Se toma una muestra de tamaño n y se calcula su media $\bar{x}$. Sabemos que la distribución de la variable aleatoria formada por las medias $\bar{x_i}$ de todas las muestras del mismo tamaño n es una normal:

$$\bar{X} \hookrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Tipificando, la variable $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ presentará una distribución $N(0,1)$ donde, como se vio en la unidad anterior,

$$p(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

y, sustituyendo la expresión de Z, tenemos:

$$p\left(-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha \Rightarrow p\left(-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\Rightarrow p\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Por tanto el **intervalo de confianza** para el parámetro μ de una población $N(\mu, \sigma)$ a un **nivel de confianza** $1 - \alpha$ es un intervalo centrado en $\bar{x}$ y de radio $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ esto es:

$$I.C = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Es un intervalo que contiene al parámetro μ con una probabilidad $1 - \alpha$.

Ejemplo 1

Un psicólogo escolar ha estudiado que el tiempo de reacción de los alumnos de primero de Primaria se distribuye normalmente con una desviación típica de 0,04 segundos. Con una muestra de 100 alumnos, la media de tiempo de reacción fue de 45 segundos. Halla un intervalo

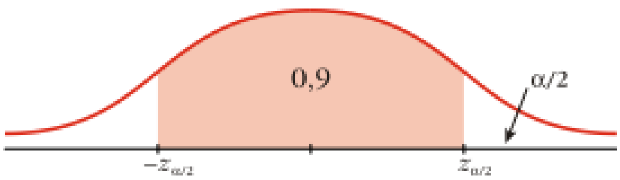
de confianza para la media de tiempos de reacción de los alumnos de primero de Primaria al nivel de confianza de:
a) 90%
b) 95%
c) Interpretar los resultados

Resolución

Media muestral $\bar{x} = 45$
Desviación típica poblacional: $\sigma = 0{,}04$

$$I.C = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

a) Calculemos $z_{\alpha/2}$ para un nivel de confianza del 90%:



Si el intervalo característico abarca un área de 0,9, fuera de él deberá haber un área de $\alpha = 0{,}1$; el área de cada una de las "colas" es $\frac{\alpha}{2}$ =0,05.
Se trata de buscar el valor de k tal que $p(Z \geq k) = 0{,}05$, esto es, $p(Z \leq k) = 0{,}95$
En las tablas encontramos:

$$p(Z \leq 1{,}64) = 0{,}9495$$
$$p(Z \leq 1{,}65) = 0{,}9505$$

El valor promedio entre 1,64 y 1,65 es 1,645. Por tanto, tomaremos $z_{\alpha/2} = 1{,}645$
El intervalo característico $[-1{,}645, 1{,}645]$ es aquel dentro del cual, en una distribución de probabilidad $N(0{,}1)$, hay un área 90% del total.
El intervalo de confianza será:

$$\left(45 - 1{,}645 \cdot \frac{0{,}04}{\sqrt{100}}, 45 + 1{,}645 \cdot \frac{0{,}04}{\sqrt{100}}\right) = (44'993, 45'007)$$

El tiempo de reacción de los alumnos de primero de Primaria está entre 44,993 y 45'007 con una confianza del 90% o lo que es lo mismo, este intervalo cubre el valor del tiempo medio de reacción con una probabilidad de 0,9.
b) Para calcular $z_{\alpha/2}$ para un nivel de confianza del 95% se procede de forma análoga obteniendo $z_{\alpha/2} = 1{,}96$
El intervalo de confianza será:

$$\left(45 - 1{,}96 \cdot \frac{0{,}04}{\sqrt{100}}, 45 + 1{,}96 \cdot \frac{0{,}04}{\sqrt{100}}\right) = (44'992, 45'008)$$

c) Cuanto mayor es el nivel de confianza, mayor es la amplitud del intervalo, con lo que aumenta el margen de error.

3. Tamaño de la muestra. Error máximo de estimación

Hasta ahora, conocido el tamaño de la muestra se calculaba el intervalo de confianza correspondiente. Se podría plantear la pregunta a la inversa: ¿cuál debe ser el tamaZo de la muestra para tener una confianza determinada?
El **error máximo** vendrá determinado por el radio del intervalo de confianza, es decir:

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

Ejemplo 2

En un determinado barrio se seleccionó al azar una muestra de 100 personas cuya media de ingresos mensuales era de 600 €. La desviación típica poblacional es de 120 €.

a) Si se toma un nivel de confianza del 95 %, ¿cuál es el intervalo de confianza para la media de los ingresos mensuales de toda la población?

b) Si se toma un nivel de confianza del 99 %, ¿cuál es el tamaño muestral necesario para estimar la media de ingresos mensuales con un error menor a 18 € ?

Resolución

a)

Sabemos que a un nivel de confianza del 95% le corresponde el valor crítico $z_{\alpha/2} = 1,96$.

El intervalo de confianza será:

$$\left(600 - 1,96 \cdot \frac{120}{\sqrt{100}}, 120 + 1,96 \cdot \frac{120}{\sqrt{100}}\right) = (576'48, 623'52)$$

b) A un nivel de confianza de 99 % le corresponde $z_{\alpha/2} = 2,575$

El error es:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Longrightarrow 18 = 2,575 \cdot \frac{120}{\sqrt{n}} \Longrightarrow \sqrt{n} = \frac{120 \cdot 2,575}{18} = 17,17 \Longrightarrow n = 294,69$$

Por tanto se necesita una muestra de, al menos, 295 personas.

Ejemplo 3

Un fabricante de pilas alcalinas sabe que el tiempo de duración, en horas, de las pilas que fabrica sigue una distribución normal de media desconocida y varianza 3600. Con una muestra de su producción, elegida al azar, y un nivel de confianza del 95% ha obtenido para la media el intervalo de confianza (372'6, 392'2).

a) Calcula el valor que obtuvo para la media de la muestra y el tamaño muestral utilizado.

b) ¿Cuál será el error de su estimación, si hubiese utilizado una muestra de tamaño 225 y un nivel de confianza del 86,9%?

Resolución

a) A un nivel de confianza del 95% le corresponde un valor $z_{\alpha/2} = 1,96$.

Varianza $\sigma^2 = 3600$; en consecuencia la desviación típica es $\sigma = \sqrt{\sigma^2} = 60$

El intervalo de confianza para la media tiene la forma

$$I.C = \left(\bar{x} - 1,96 \cdot \frac{60}{\sqrt{n}}, \bar{x} + 1,96 \cdot \frac{60}{\sqrt{n}}\right) = (372'6, 392'2)$$

Por tanto, tenemos el sistema

$$\begin{cases} \bar{x} - 1,96 \cdot \frac{60}{\sqrt{n}} = 372,6 \\ \bar{x} + 1,96 \cdot \frac{60}{\sqrt{n}} = 392,2 \end{cases}$$ y sumando miembro a miembro ambas ecuaciones, obtenemos

$$2\bar{x} = 764,8 \text{ de donde } \bar{x} = 382,4.$$

Sustituyendo este valor en la primera ecuación $382,4 - 1,96 \cdot \frac{60}{\sqrt{n}} = 372,6$ de donde $n = 144$.

El tamaño muestral utilizado es 144.

b) Para nivel de confianza de 86,9%, el intervalo abarca un área de $1 - \alpha = 0,869$; fuera de él deberá haber un área de 0,131; el área de cada una de las "colas" es 0,0655. Se trata de buscar k tal que $p(Z > k)$=0,0655, o lo que es lo mismo $p(Z \leq k) = 0,9345$ En las tablas encontramos:

$$p(Z \leq 1,51)=0,9345$$

Por tanto $z_{\alpha/2} = 1,51$ y el error $E = \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \pm 1,51 \cdot \frac{60}{\sqrt{225}} = \pm 6,04$

4. Intervalo de confianza para la proporción

En una población nos fijamos en una cierta característica de sus individuos. Queremos estimar la probabilidad p que hay de que un individuo cualquiera posea esa característica o, lo que es lo mismo, la proporción de ellos en la población. Tomamos una muestra $(x_1, x_2, \dots, x_n)$ de la población y consideramos la variable X número de individuos con la característica (éxito); se trata de una distribución binomial de parámetro p desconocido: $X \rightsquigarrow B(n, p)$ que, para n suficientemente grande, se aproxima por una normal $N(np, \sqrt{np(1-p)})$.

La proporción muestral es $\tilde{p} = \frac{individos\ en\ la\ muestra\ con\ la\ característica}{n} = \frac{X}{n}$

La distribución $\tilde{P}$ de las proporciones muestrales es una normal

$$\tilde{P} \hookrightarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Razonando de forma análoga a como lo hicimos para la media, el intervalo de confianza para la proporción de individuos que cumplen una determinada característica en una población, con un nivel de confianza $1-\alpha$, construido a partir de una muestra es

$$I.C = \left(\tilde{p} - z_{\alpha/2} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z_{\alpha/2} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right)$$

Ejemplo 3

Para hallar la proporción de inmigrantes de una ciudad, se toma una muestra de 600 personas y se observa que 36 de ellas son inmigrantes. Determina el intervalo de confianza correspondiente con un nivel de confianza del 99%.

Resolución

Sabemos que a un nivel de confianza del 99% le corresponde el valor crítico $z_{\alpha/2} = 2{,}575$. Por otro lado $\tilde{p} = \frac{36}{600} = 0{,}06$

El intervalo de confianza será:

$$\left(\tilde{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right) =$$

$$\left(0{,}06 - 2{,}575 \cdot \sqrt{\frac{0{,}06 \cdot 0{,}94}{600}}, 0{,}06 + 2{,}575 \cdot \sqrt{\frac{0{,}06 \cdot 0{,}94}{600}}\right) = (0'035, 0'085)$$

Con la muestra elegida, la probabilidad de que la proporción de inmigrantes en la población esté en el intervalo $(0'035, 0'085)$ es 0,99.

---

5. Error máximo

El **error máximo** vendrá determinado por el intervalo de confianza, es decir

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$$

Ejemplo 4

Se desea estimar la proporción, p, de individuos daltónicos de una población a través del porcentaje observado en una muestra aleatoria de individuos, de tamaño n.

a) Si el porcentaje de individuos daltónicos en la muestra es igual al 30%, calcula el valor de n para que, con un nivel de confianza de 0,95, el error cometido en la estimación sea inferior al 3,1%.

b) Si el tamaño de la muestra es de 64 individuos, y el porcentaje de individuos daltónicos en la muestra es del 35%, determina, usando un nivel de significación del 1%, el correspondiente intervalo de confianza para la proporción de daltónicos de la población.

Resolución

a) Proporción muestral: $\tilde{p} = 0{,}3$; Nivel de confianza $1 - \alpha = 0{,}95$; $E < 0{,}031$

Sabemos que a un nivel de confianza del 95% le corresponde el valor crítico $z_{\alpha/2} = 1{,}96$.

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} < 0{,}031 \Rightarrow 1{,}96 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{n}} < 0{,}031 \Rightarrow n > \left(\frac{1{,}96 \cdot \sqrt{0{,}021}}{0{,}031}\right)^2 \Rightarrow n > 83{,}948$$

La muestra ha de ser de 84 individuos.

b) $n = 64$; Nivel de significación $\alpha = 0{,}01$; Nivel de confianza $1 - \alpha = 0{,}99$

Proporción muestral $\tilde{p} = 0{,}35$

Sabemos que a un nivel de confianza del 99% le corresponde el valor crítico $z_{\alpha/2} = 2{,}575$.

$$I.C = \left(0{,}35 - 2{,}575 \cdot \sqrt{\frac{0{,}35 \cdot 0{,}65}{64}}, 0{,}35 + 2{,}575 \cdot \sqrt{\frac{0{,}35 \cdot 0{,}65}{64}}\right) = (0'196, 0'504)$$

Ejemplo 5

A partir de una muestra de tamaño 400, se estima la proporción de individuos que leen el periódico en una gran ciudad. Se obtiene una cota de error de 0,0392 con un nivel de confianza del 95%.

a) ¿Podríamos, con la misma muestra, mejorar el nivel de confianza en la estimación? ¿Qué le ocurriría a la cota de error?

b) ¿Sabrías calcular la proporción, p̃, obtenida en la muestra?

Resolución

a) Aumentando la cota de error mejoraría el nivel de confianza.

b) $n = 400$; $E = 0{,}0392$; $1 - \alpha = 0{,}95$; $z_{\alpha/2} = 1{,}96$

$$E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = 0{,}0392 \Rightarrow 1{,}96 \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{400}} = 0{,}0392 \Rightarrow$$

$$\tilde{p}(1-\tilde{p}) = \left(\frac{0{,}0392 \cdot 20}{1{,}96}\right)^2 = 0{,}16$$

Resolvemos la ecuación de segundo grado $\tilde{p}(1-\tilde{p}) = 0{,}16$:

$$\tilde{p}^2 - \tilde{p} + 0{,}16 = 0 \Rightarrow \tilde{p} = \frac{1 \pm \sqrt{1 - 0{,}64}}{2} = \begin{cases} \tilde{p} = 0{,}8 \\ \tilde{p} = 0{,}2 \end{cases}$$

---

6. Hipótesis estadística

Cuando en un estudio estadístico queremos determinar si una población cumple una determinada característica, previamente debemos plantear un test estadístico que será el procedimiento que nos permitirá evaluar, a partir de una muestra, si una determinada hipótesis formulada sobre una característica de la población se verifica o no.

Una vez concluido el test podemos considerar la hipótesis que, en principio, admitimos como válida, y que llamaremos hipótesis nula, H_0 (sobre la que queremos decidir) y una hipótesis contraria a ésta, que denominaremos hipótesis alternativa H_1, que es la que admitiremos como válida si nos vemos obligados a rechazar la hipótesis H_0.

6.1 Contraste de hipótesis

Es un procedimiento del que depende la decisión de aceptar o rechazar la hipótesis nula y esta aceptación o rechazo dependerá, a su vez, de cuál sea la discrepancia entre la hipótesis y la información muestral que tengamos. Si la discrepancia es menor que un determinado valor que consideramos aceptable, la hipótesis se dará por cierta; este valor es el **valor de significación** α.

Es evidente que una hipótesis estadística no se puede aceptar o rechazar con una certeza del 100%, sino que se define un nivel crítico α, que nos marcará los límites para aceptar o rechazar la hipótesis nula.

Así, por ejemplo, si el nivel de significación es $\alpha = 0{,}05$, rechazaremos como improbables el 5% de los casos extremos; por esta razón, en ocasiones, también se dice que estamos trabajando con un **nivel de confianza** del 95%.

Nosotros trabajaremos con hipótesis nulas relativas a la media. Los valores de la media que nos lleven a aceptar la hipótesis nula H_0, forman la **región de aceptación**, y los que nos conducen a rechazarla, constituyen la **región de rechazo**.

6.2 Fases del contraste de hipótesis

Para efectuar un contraste de hipótesis debemos seguir los pasos siguientes:

- Se debe enunciar la hipótesis nula y la alternativa. (por ejemplo, el investigador formula una hipótesis sobre un parámetro poblacional que toma un determinado valor)
- Se extrae una muestra de tamaño n y se calcula en ella el valor del parámetro estadístico que se desea encontrar.
- Se elige el nivel de significación con el que se quieren tomar las decisiones; generalmente los niveles de significación son $\alpha = 0{,}1$, $\alpha = 0{,}05$ y $\alpha = 0{,}01$.
- A continuación se construye la zona de aceptación de la hipótesis, es decir, los intervalos característicos, fuera de los cuales se encuentra el porcentaje de $\alpha \cdot 100\%$ de casos que queremos rechazar.
- Si el valor del parámetro muestral se encuentra dentro de la zona de aceptación, se acepta la hipótesis con un nivel de significación α. En caso contrario, se rechaza.

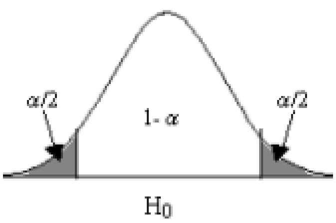
6.3 Contraste de hipótesis para la media de una población normal

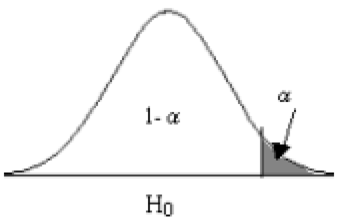
Aquí trabajaremos con hipótesis nulas relativas a la media. Los valores de la media que nos lleven a aceptar la hipótesis nula H_0, forman la **región de aceptación**, y los que nos conducen a rechazarla, constituyen la **región de rechazo**.

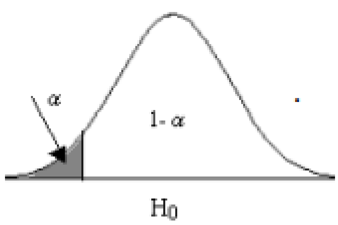
Se inicia el contraste definiendo la hipótesis nula y la alternativa.

En el momento de definir la hipótesis nula, ésta se puede plantear en términos de igualdad o de desigualdad: $H_0 \equiv \mu = \mu_0$ o bien $H_0 \equiv \mu \geq \mu_0$ o $\mu \leq \mu_0$

En el primer caso es un contraste bilateral, o de dos colas, y los otros dos, contrastes unilaterales o de una cola.



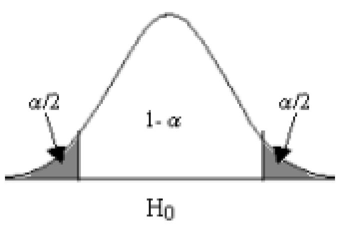




Contraste bilateral Contraste unilateral Contraste unilateral

6.3.1 Contraste bilateral

$H_0 \equiv \mu = \mu_0$



$H_1 \equiv \mu \neq \mu_0$
Dada una población X de media μ y desviación típica σ, la distribución de las medias de todas las muestras se distribuye según $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
El intervalo de aceptación para esta distribución es:

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

Si el valor de la media muestral $\bar{x}$, se encuentra en ese intervalo, se aceptará la hipótesis nula; en caso contrario, se rechazará.

Ejemplo 6

Se quiere estimar la media de la nómina mensual que reciben los directivos de las compañías multinacionales que operan en Europa.
a) Si la varianza de la nómina en la población es de 1000 €, ¿cuál es la varianza de la media muestral cuando el tamaño de la muestra es de 100?
b) Si en las condiciones del apartado anterior, la media muestral es de 4008 €, ¿se rechazaría, con un nivel de confianza del 95%, la hipótesis de que la nómina media es de 4000 €?

Resolución

a) Tamaño muestral $n = 100$
Varianza poblacional $\sigma^2 = 1000$; por tanto la desviación típica es $\sigma = \sqrt{1000}$
La desviación típica de la media muestral viene dada por $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{1000}}{\sqrt{100}} = \sqrt{10}$
La varianza de la media muestral es $\left(\frac{\sigma}{\sqrt{n}}\right)^2 = \left(\sqrt{10}\right)^2 = 10$
b) Media muestral $\bar{x} = 4008$
Se trata de un contraste bilateral para la media
$H_0 \equiv \mu = 4000$
$H_1 \equiv \mu \neq 4000$
A un nivel de confianza del 95% le corresponde el valor crítico $z_{\alpha/2} = 1{,}96$.
El intervalo de aceptación para esta distribución es:
$\left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(4000 - 1{,}96 \cdot \sqrt{10}, 4000 + 1{,}96 \cdot \sqrt{10}\right) = (3993'8, 4006'2)$
Como $\bar{x} = 4008 \notin= (3993'8, 4006'2)$ se rechaza la hipótesis de que la nómina media sea de 4000 euros con una confianza del 95%.

Ejemplo 7

El peso medio de una muestra aleatoria de 100 naranjas de una determinada variedad es de 272 g. Se sabe que la desviación típica poblacional es de 20 g. A un nivel de significación de 0,05, ¿hay suficiente evidencia para refutar la afirmación de que el peso medio de dicha variedad de naranjas es de 275 g?

Resolución

Se trata de un test de hipótesis bilateral para la media.
$H_0 \equiv \mu = 275$
$H_1 \equiv \mu \neq 275$
Nivel de significación $\alpha = 0{,}05$. Nivel de confianza $1 - \alpha = 0{,}95$. $n = 100$
Peso medio de la muestra $\bar{x} = 272\ g$. Desviación típica poblacional $\sigma = 20\ g$
La zona de aceptación de la hipótesis nula, H_0 ,es:

$$\left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = (275 - 1{,}96 \cdot 2, 275 + 1{,}96 \cdot 2) \cong (271'1, 278'9)$$

Como $\bar{x} = 272 \in (271'1, 278'9)$, no hay suficiente evidencia para refutar la afirmación de que el peso medio de dicha variedad de naranjas es de 275 g.

6.3.2 Contraste unilateral

Se plantea cuando la hipótesis nula es de la forma $H_0 \equiv \ \mu \geq \mu_0$ o $H_0 \equiv \ \mu \leq \mu_0$

El contraste unilateral ha de verificar que el área correspondiente a la región de aceptación esté toda hacia un lado de la distribución, de modo que la región rechazable quede totalmente al otro lado. Si la región de aceptación ha de ser $1-\alpha$, la región de rechazo vendrá determinada por el valor crítico z_α.

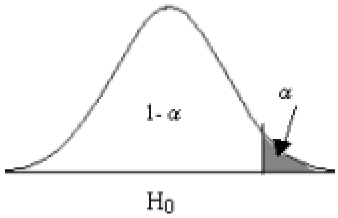
[1]

$H_0 \equiv \ \mu \leq \mu_0$

$H_1 \equiv \ \mu > \mu_0$

La zona de aceptación de la hipótesis nula, H_0 , es:

$$\left(-\infty,\ \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$$



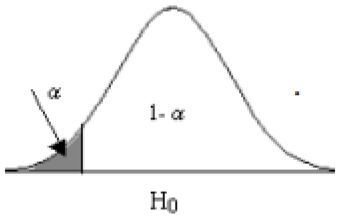
[2]

$H_0 \equiv \ \mu \geq \mu_0$

$H_1 \equiv \ \mu < \mu_0$

La zona de aceptación de la hipótesis nula, H_0 , es:

$$\left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, +\infty\right)$$



6.3.3 Errores

La siguiente tabla muestra los errores que se pueden cometer y cómo se clasifican:

	H_0 cierta	H_0 falsa
Aceptamos H_0	No hay error	Error de tipo II
Aceptamos H_1	Error de tipo I	No hay error

Ejemplo 8

En los últimos años el consumo familiar diario en electricidad (en Kw) de cierta ciudad seguía una Normal de media 6,3 y desviación típica 1,2. Sin embargo, desde hace unos meses las tarifas eléctricas han experimentado varias reducciones, y se piensa que esto ha podido repercutir en un aumento del consumo. Recientemente, para una muestra de 47 familias se ha obtenido un consumo medio diario de 6,8. Suponiendo que el consumo sigue siendo aproximadamente Normal y que la desviación típica se ha mantenido:

a) Plantea el test para contrastar que el abaratamiento de las tarifas no ha influido en el consumo, frente a que ha tenido la repercusión que se piensa, como parecen indicar los datos. Si se concluyera que la media de consumo se ha mantenido y realmente subió, ¿cómo se llama al error cometido?

b) ¿A qué conclusión se llega en el test planteado en el apartado anterior con un nivel de significación del 1%?

Resolución

a) Se trata de un test de hipótesis unilateral para la media.

$H_0 \equiv \ \mu \leq 6{,}3$

$H_1 \equiv \ \mu > 6{,}3$

Si se concluyera que la media del consumo se ha mantenido cuando realmente subió, se está aceptando que la hipótesis nula es verdadera cuando realmente es falsa. Se comete un error de tipo II.

b) Nivel de significación $\alpha = 0{,}01$. Nivel de confianza $1-\alpha = 0{,}99$.

Valor $z_\alpha = 2{,}33$. Tamaño muestral $n = 47$.

Media de la muestra $\bar{x} = 6{,}8\ g$. Desviación típica poblacional $\sigma = 1{,}2$

La zona de aceptación de la hipótesis nula, H_0 , es:

$$\left(-\infty, 6'3 + 2'33 \cdot \frac{1{,}2}{\sqrt{47}}\right) = (-\infty, 6'7)$$

Como $\bar{x} = 6{,}8 \notin (-\infty, 6'7)$, se rechaza la hipótesis, es decir, el abaratamiento de las tarifas ha repercutido en un aumento del consumo, con un nivel de significación del 1%.

6.4 Contraste de hipótesis para la comparar dos medias (diferencia de medias)

Partimos de dos poblaciones con distribuciones normales y deviaciones típicas σ_1 y σ_2 conocidas. A partir de dos muestras, de cada una de ellas, de medias $\bar{x}_1$ y $\bar{x}_2$ y tamaños n_1 y n_2, respectivamente, se construye la zona de aceptación de la misma forma que en el punto 6.3.
Los tipos de contraste, para un nivel de significación α, son:

[1] Bilateral

$H_0 \equiv \ \mu_1 - \mu_2 = 0$
$H_1 \equiv \ \mu_1 - \mu_2 = 0$
La zona de aceptación $\left(-z_{\alpha/2} \cdot \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}, z_{\alpha/2} \cdot \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}\right)$

Calculamos la diferencia de medias muestrales, $\bar{x}_1 - \ \bar{x}_2$, y comprobamos si pertenece o no a la zona de aceptación para aceptar o no la hipótesis H_0 de que las medias μ_1 y μ_2 sean iguales.

[2] Unilateral

$H_0 \equiv \ \mu_1 - \mu_2 \leq 0$
$H_1 \equiv \ \mu_1 - \mu_2 > 0$
La zona de aceptación $\left(-\infty, z_{\alpha} \cdot \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}\right)$

Calculamos la diferencia de medias muestrales, $\bar{x}_1 - \ \bar{x}_2$, y comprobamos si pertenece o no a la zona de aceptación para aceptar o no la hipótesis H_0 de que la media μ_1 es menor o igual que μ_2 .

[3] Unilateral

$H_0 \equiv \ \mu_1 - \mu_2 \geq 0$
$H_1 \equiv \ \mu_1 - \mu_2 < 0$
La zona de aceptación $\left(- z_{\alpha} \cdot \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}, +\infty\right)$

Calculamos la diferencia de medias muestrales, $\bar{x}_1 - \ \bar{x}_2$, y comprobamos si pertenece o no a la zona de aceptación para aceptar o no la hipótesis H_0 de que la media μ_1 es mayor o igual que μ_2 .

6.5 Contraste de hipótesis para la proporción (parámetro p de una binomial)

En el caso de un contraste para la proporción con nivel de significación α:

[1] Contraste bilateral

$H_0 \equiv \ p = p_0$
$H_1 \equiv \ p \neq p_0$
El intervalo de aceptación es:

$$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot (1 - p_0)}{n}}, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot (1 - p_0)}{n}}\right)$$

[2] Contraste unilateral

$H_0 \equiv \ p \leq p_0$
$H_1 \equiv \ p > p_0$
El intervalo de aceptación es:

$$\left(-\infty, p_0 + z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}}\right)$$

[3] Contraste unilateral

$H_0 \equiv \ p \geq p_0$
$H_1 \equiv \ p < p_0$
El intervalo de aceptación es:

$$\left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}}, +\infty\right)$$