Universidad Autónoma de Madrid

Gestión Aeronáutica: Estadística Teórica
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández

CONTRASTES
ERROR TIPO I, II Y POTENCIA DEL CONTRASTE

RELACIÓN: INTERVALOS CONFIANZA - CONTRASTE HIPÓTESIS

a)

Intervalo de confianza para la media μ de una distribución normal $N(\mu, \sigma)$ de varianza conocida:

$$N(\mu, \sigma) \qquad \boxed{\bar{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

$$I_{1-\alpha}(\mu) = \left\{ \overbrace{\bar{x}}^{\text{media muestral}} \pm \overbrace{z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}^{\text{error muestral}} \right\}$$

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \mapsto \quad -z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu \le z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \mapsto \quad \left| \bar{x} - \mu \right| \le z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Hipótesis sobre la media de una población con σ^2 conocida: REGIÓN DE RECHAZO

$H_0: \ \mu = \mu_0 \quad H_1: \ \mu \ne \mu_0 \qquad R = \left\{ \left| \bar{x} - \mu_0 \right| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$ bilateral (compuesta)

$H_0: \ \mu = \mu_0 \quad H_1: \ \mu_1 > \mu_0 \qquad R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\}$ unilateral (simple)

$H_0: \ \mu = \mu_0 \quad H_1: \ \mu_1 < \mu_0 \qquad R = \left\{ \bar{x} - \mu_0 \overbrace{<}^{z_\alpha = -z_{1-\alpha}} z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$ unilateral (simple)

b)

Intervalo de confianza para la media μ de una distribución normal $N(\mu, \sigma)$ de varianza desconocida con muestras pequeñas $n \le 30$

$$N(\mu, \sigma) \qquad \boxed{\frac{\bar{x} - \mu}{s_x / \sqrt{n}} \approx t_{n-1} \quad n \le 30}$$

$$I_{1-\alpha}(\mu) = \left\{ \overbrace{\bar{x}}^{\text{media muestral}} \pm \overbrace{t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}}}^{\text{error muestral}} \right\} \underbrace{\overbrace{\equiv}^{n\,\sigma_x^2 = (n-1)\, s_x^2}}_{\frac{s_x^2}{n} = \frac{\sigma_x^2}{n-1}} \left\{ \bar{x} \pm t_{\frac{\alpha}{2}, (n-1)} \frac{\sigma_x}{\sqrt{n-1}} \right\}$$

$$\bar{x} - t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}} \quad \mapsto \quad t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}} \le \bar{x} - \mu \le t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}} \quad \mapsto \quad \left| \bar{x} - \mu \right| \le t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}}$$

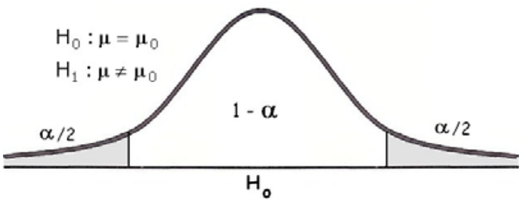
$$I_{1-\alpha}(\mu) = \left\{ \bar{x} \pm t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}} \right\} \quad \mapsto \quad \left| \bar{x} - \mu \right| \le t_{\frac{\alpha}{2}, (n-1)} \frac{s_x}{\sqrt{n}}$$

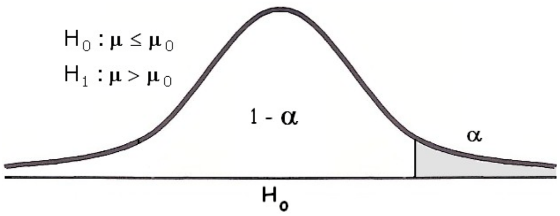
HIPÓTESIS SOBRE LA MEDIA DE UNA POBLACIÓN CON σ^2 DESCONOCIDA: REGIÓN DE RECHAZO

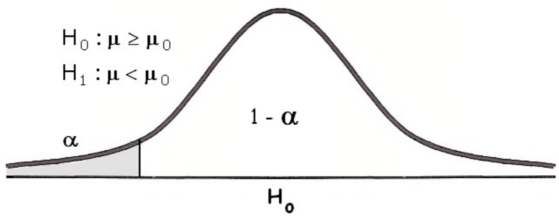
$$H_0:\ \mu=\mu_0 \quad H_1:\ \mu\neq\mu_0 \qquad R=\left\{\ \left|\bar{x}-\mu_0\right| > t_{\alpha/2\,;\,(n-1)}\,\frac{s_x}{\sqrt{n}}\right\}$$ bilateral (compuesta)

$$H_0:\ \mu=\mu_0 \quad H_1:\ \mu_1>\mu_0 \qquad R=\left\{\ \bar{x}-\mu_0 > t_{\alpha\,;(n-1)}\,\frac{s_x}{\sqrt{n}}\right\}$$ unilateral (simple)

$$H_0:\ \mu=\mu_0 \quad H_1:\ \mu_1<\mu_0 \qquad R=\left\{\ \bar{x}-\mu_0 \overbrace{<}^{t_{\alpha;\,n}=-t_{1-\alpha;\,n}} t_{1-\alpha\,;(n-1)}\,\frac{s_x}{\sqrt{n}}\right\}$$ unilateral (simple)

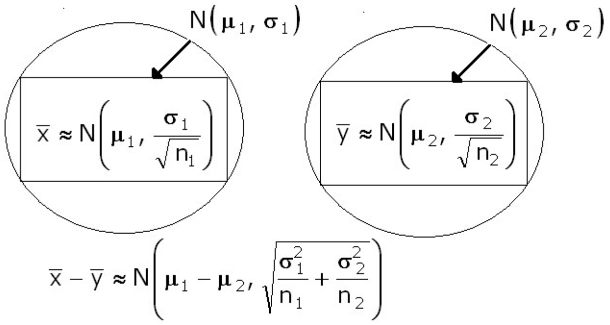






c)

Intervalo de confianza para la diferencia de medias $(\mu_1-\mu_2)$ de dos distribuciones normales $N(\mu_1,\,\sigma_1)$, $N(\mu_2,\,\sigma_2)$ con varianzas poblacionales conocidas:



$$I_{1-\alpha}\,(\mu_1-\mu_2)=\left\{\ \overbrace{(\bar{x}-\bar{y})}^{\text{diferencia muestral}}\ \pm\ \overbrace{z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}^{\text{error muestral}}\right\}$$

$$(\bar{x}-\bar{y})-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le(\mu_1-\mu_2)\le(\bar{x}-\bar{y})+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$\left|(\bar{x}-\bar{y})-(\mu_1-\mu_2)\right|\le z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

CONTRASTE DE IGUALDAD DE MEDIAS DE DOS POBLACIONES NORMALES CON VARIANZAS CONOCIDAS: REGIÓN DE RECHAZO

$H_0 : \mu_1 = \mu_2 \qquad R = \left\{ \left|(\bar{x} - \bar{y}) - 0\right| > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ bilateral

$H_0 : \mu_1 - \mu_2 = k \qquad R = \left\{ \left|\bar{x} - \bar{y} - k\right| > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ bilateral

$H_0 : \mu_1 - \mu_2 \leq k \qquad R = \left\{ \bar{x} - \bar{y} - k > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ unilateral

$H_0 : \mu_1 - \mu_2 \geq k \qquad R = \left\{ \bar{x} - \bar{y} - k < z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ unilateral

$H_0 : \mu_1 \leq \mu_2 \qquad R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ unilateral

$H_0 : \mu_1 \geq \mu_2 \qquad R = \left\{ \bar{x} - \bar{y} \overbrace{>}^{z_\alpha = -z_{1-\alpha}} z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ unilateral

CONTRASTE DE HIPÓTESIS

CONTRASTE DE LA MEDIA DE UNA POBLACIÓN NORMAL $N(\mu, \sigma)$ CON VARIANZA σ^2 CONOCIDA:

a) CONTRASTE BILATERAL o DE DOS COLAS

Hipótesis nula: $H_0 : \mu = \mu_0$ Hipótesis alternativa: $H_1 : \mu \neq \mu_0$

La hipótesis alternativa es $\mu \neq \mu_0$ es la decisión que se ha de tomar de tomar, deberán ser válidos los valores de μ mayores o menores que un valor dado μ_0, por lo cual el contraste debe ser bilateral o de dos colas.

$$\text{Regla decisión} \begin{cases} \text{Si } |\bar{x}| \leq k \text{ se acepta la hipótesis nula } H_0 \text{ (Región Aceptación)} \\ \text{Si } |\bar{x}| > k \text{ se rechaza la hipótesis nula } H_0 \text{ (Región Rechazo)} \end{cases}$$

De otra parte, en la distribución del muestreo $\bar{x} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$, que bajo la hipótesis nula $\bar{x} \sim N\left(\mu_0, \dfrac{\sigma}{\sqrt{n}}\right)$, con lo que la variable $\dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ es N(0, 1).

El **valor crítico k** se calcula mediante el error de significación α :

$$\alpha = P\left[\text{Rechazar } H_0 / H_0 \text{ es cierta}\right] = P\left[|\bar{x}| > k / N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)\right] = P\left[\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > K\right] =$$

$$= P\left[\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -K\right) \cup \left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > K\right)\right] = P\left[\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -K\right] + P\left[\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > K\right] \overset{\overbrace{\text{simetría } N(0,1)}}{=} \frac{\alpha}{2} + \frac{\alpha}{2}$$

La región crítica será $\left|\dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2} \Rightarrow R = \left\{|\bar{x} - \mu_0| > z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}\right\}$

En otras palabras,

Se acepta H_0 si $\dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$ Se rechaza H_0 si $\dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}$

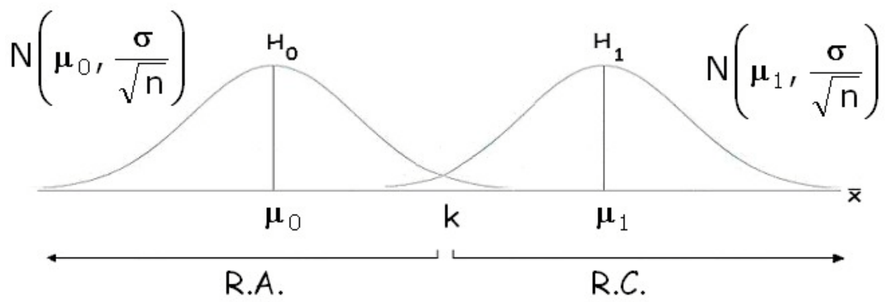
b) CONTRASTE UNILATERAL o DE UNA COLA

Hipótesis nula: $H_0 : \mu = \mu_0$ Hipótesis alternativa: $H_1 :\ \mu_1 > \mu_0$

La hipótesis alternativa es $\mu \neq \mu_0$ es la decisión que se ha de tomar de tomar solo son válidos los valores de μ_1 son mayores que un valor dado μ_0, por lo cual el contraste debe ser unilateral o de una cola.

Regla de decisión $\begin{cases} \text{Si } \overline{x} \leq k & \text{se acepta la hipótesis nula } H_0 \quad \text{(Región Aceptación)} \\ \text{Si } \overline{x} > k & \text{se rechaza la hipótesis nula } H_0 \quad \text{(Región Rechazo)} \end{cases}$

De otra parte, en la distribución del muestreo $\overline{x} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$



Bajo la hipótesis nula: $\overline{x} \sim N\left(\mu_0, \dfrac{\sigma}{\sqrt{n}}\right)$ Bajo la hipótesis alternativa: $\overline{x} \sim N\left(\mu_1, \dfrac{\sigma}{\sqrt{n}}\right)$

Para hallar el valor crítico K se recurre al Error Tipo I:

$$\alpha = P\left[\text{ET I}\right] = P\left[\text{Rechazar } H_0 \,/\, H_0 \text{ es cierta}\right] = P\left[\overline{x} > k \,/\, N\left(\mu_0, \dfrac{\sigma}{\sqrt{n}}\right)\right] = P\left[\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > K\right]$$

La región crítica será $\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \;\Rightarrow\; R = \left\{\overline{x} - \mu_0 > z_\alpha\, \dfrac{\sigma}{\sqrt{n}}\right\}$

En otras palabras,

Se acepta H_0 si $\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \leq z_\alpha$ Se rechaza H_0 si $\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$

CONTRASTE DE IGUALDAD DE MEDIAS DE DOS POBLACIONES NORMALES CON VARIANZAS CONOCIDAS

a) CONTRASTE BILATERAL o DE DOS COLAS

Hipótesis nula: $H_0: \mu_1 - \mu_2 = 0$ Hipótesis alternativa: $H_1: \mu_1 - \mu_2 \neq 0$

La regla de decisión será: $\begin{cases} \text{Si } \left|\bar{x} - \bar{y}\right| \le k \text{ no se rechaza } H_0 \mapsto \text{(R.A)} \\ \text{Si } \left|\bar{x} - \bar{y}\right| > k \text{ se rechaza } H_0 \mapsto \text{(R.C)} \end{cases}$

La región crítica de dos colas $\left|\bar{x} - \bar{y}\right| > k$ es función de la diferencia de las medias muestrales. En esta línea, las distribuciones en el muestreo de las medias son:

$\bar{x} \sim N\left[\mu_1, \dfrac{\sigma_1}{\sqrt{n_1}}\right]$, $\bar{y} \sim N\left[\mu_2, \dfrac{\sigma_2}{\sqrt{n_2}}\right]$, con lo cual, la diferencia de medias muestrales,

bajo la hipótesis nula $H_0: \mu_1 - \mu_2 = 0$ se tiene que $\bar{x} - \bar{y} \sim N\left[0, \sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{n_2}}\right]$

El valor crítico k se determina mediante el error tipo I:

$$\alpha = P(\text{ET I}) = P(\text{Rechazar } H_0 \,|\, H_0 \text{ cierta}) = P\left[\,\left|\bar{x} - \bar{y}\right| > k / H_0: \mu_1 - \mu_2 = 0\right] =$$

$$= P\left[\left|\frac{(\bar{x} - \bar{y}) - 0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}\right| > K\right] =$$

$$= P\left[\left(\frac{\bar{x} - \bar{y}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} < -K\right) \cup \left(\frac{\bar{x} - \bar{y}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} > K\right)\right] =$$

$$= P\left[\left(\frac{\bar{x} - \bar{y}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} < -K\right)\right] + P\left[\left(\frac{\bar{x} - \bar{y}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} > K\right)\right] = \frac{\alpha}{2} + \frac{\alpha}{2} \quad \text{(simetría)}$$

La región crítica es

$$\left|\frac{\bar{x} - \bar{y}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}\right| > z_{\alpha/2} \quad \mapsto \quad R = \left\{ \left|(\bar{x} - \bar{y})\right| > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$$

En otras palabras,

Se acepta la hipótesis nula H_0 si: $\underbrace{\dfrac{\left|\bar{x} - \bar{y}\right|}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}}_{\text{estadístico observado}} \le \underbrace{z_{\alpha/2}}_{\substack{\text{estadístico} \\ \text{teórico}}}$

Se rechaza la hipótesis nula H_0 si: $\underbrace{\dfrac{\left|\bar{x} - \bar{y}\right|}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}}_{\text{estadístico observado}} > \underbrace{z_{\alpha/2}}_{\substack{\text{estadístico} \\ \text{teórico}}}$

b) CONTRASTE UNILATERAL o DE UNA COLA

Hipótesis nula: $H_0: \mu_1 - \mu_2 = K_0$ Hipótesis alternativa: $H_1: \mu_1 - \mu_2 > K_0$

La regla de decisión será: $\begin{cases} \text{Si } (\overline{x} - \overline{y}) \le k \text{ no se rechaza } H_0 \mapsto \text{(R.A)} \\ \text{Si } (\overline{x} - \overline{y}) > k \text{ se rechaza } H_0 \mapsto \text{(R.C)} \end{cases}$

La región crítica de una cola $(\overline{x} - \overline{y}) > k$ es función de la diferencia de las medias muestrales. En esta línea, las distribuciones en el muestreo de las medias son:

$\overline{x} \sim N\left[\mu_1, \dfrac{\sigma_1}{\sqrt{n_1}}\right]$, $\overline{y} \sim N\left[\mu_2, \dfrac{\sigma_2}{\sqrt{n_2}}\right]$, con lo cual, la diferencia de medias muestrales

$$\overline{x} - \overline{y} \sim N\left[(\mu_1 - \mu_2), \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Bajo la hipótesis nula $H_0: \mu_1 - \mu_2 = K_0$ se tiene $\overline{x} - \overline{y} \sim N\left[K_0, \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\right]$

El valor crítico K se determina mediante el nivel de significación α:

$$\alpha = P(\text{ET I}) = P(\text{Rechazar } H_0 \mid H_0 \text{ cierta}) = P\left[\, (\overline{x} - \overline{y}) > k / H_0 : \mu_1 - \mu_2 = K_0 \,\right] =$$

$$= P\left[\frac{(\overline{x} - \overline{y}) - K_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} > \frac{k - K_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}\right] = P\left[z > \frac{k - K_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}\right]$$

con lo cual, el valor crítico se despeja $\dfrac{k - K_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} = z_\alpha$

Comprobando después si se verifica o no la evidencia empírica $(\overline{x} - \overline{y}) > k$

De otra parte, la región crítica $\dfrac{(\overline{x} - \overline{y}) - K_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} > z_\alpha$

por tanto, la región de rechazo: $R = \left[(\overline{x} - \overline{y}) - K_0 > z_\alpha \sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}\right]$

ERROR TIPO I, TIPO II Y POTENCIA

$\alpha \equiv$ Probabilidad de rechazar la hipótesis nula H_0 siendo cierta (Error Tipo I)

$\beta \equiv$ Probabilidad de aceptar la hipótesis nula H_0 siendo falsa (Error Tipo II)

$1-\beta \equiv$ probabilidad de rechazar la hipótesis nula H_0 siendo falsa (Potencia Contraste)

Los errores están relacionados, al disminuir el uno aumenta el otro:

$\alpha = P(\text{Error Tipo I}) = 0 \ \mapsto$ Rechazar siempre $H_1 \Leftrightarrow \ \beta = P(\text{Error Tipo II}) = 1$

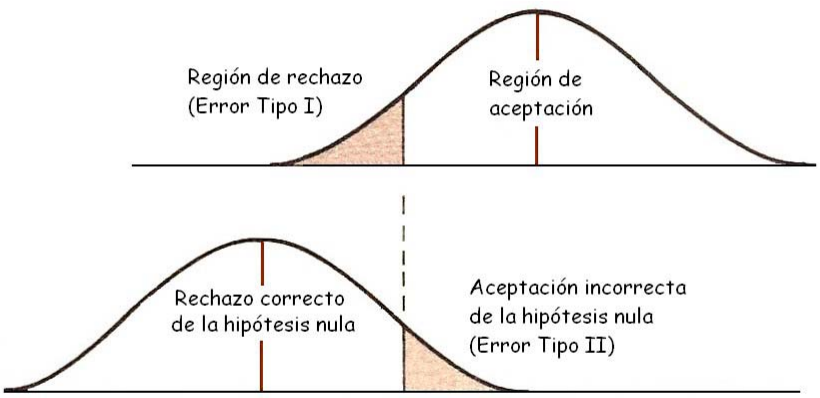
$\beta = P(\text{Error Tipo II}) = 0 \ \mapsto$ Rechazar siempre $H_0 \Leftrightarrow \ \alpha = P(\text{Error Tipo I}) = 1$

Un contraste debería buscar simultáneamente el nivel de significación α más bajo posible y la potencia $1-\beta$ más alta posible.

Fijado el nivel de significación, se determina la región de rechazo cuya potencia es mayor entre todos los contrastes cuyo tamaño sea el fijado a priori.

La única posibilidad para conseguir que un contraste mejore su potencia $1-\beta$, sin aumentar el nivel de significación α, es incrementar el tamaño de la muestra.

Al aumentar el tamaño de la muestra, varía la ley de distribución del estadístico de contraste, y generalmente disminuye la varianza. Generalmente, las propiedades del contraste mejoran.



Antes de la universalización del ordenador se utilizaban como más representativos los valores del 1%, 5%, y 10%. La metodología más razonable es tomar un nivel de significación α de acuerdo con la experiencia y después obtener el llamado p-valor.

$$\text{p-valor} \equiv \begin{cases} \text{nivel de significación } \alpha \text{ más pequeño posible que se puede escoger, para} \\ \text{el que todavía se rechazaría la hipótesis nula } H_0 \text{ con la muestra actual.} \end{cases}$$

$$\begin{cases} \text{Si } \ \alpha < \text{p-valor} \ \mapsto \ \text{Se acepta } H_0 \\ \text{Si } \ \alpha \geq \text{p-valor} \ \mapsto \ \text{Se rechaza } H_0 \end{cases}$$

El p-valor es el menor α que permite aceptar la hipótesis alternativa H_1.

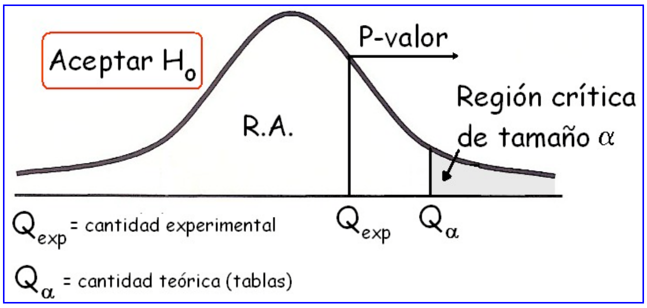
El p-valor tiene la ventaja de permitir que se decida que hipótesis se acepta, esto no es posible cuando se indica sólo el resultado del contraste (si se acepta o se rechaza H_0 con un α fijo).

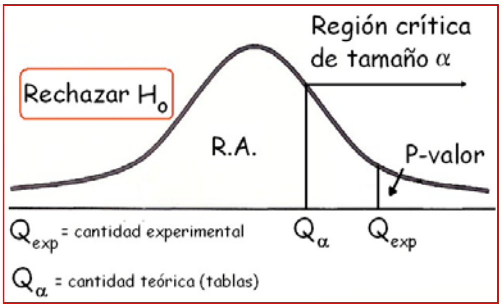
CRITERIOS GENERALES para CONTRASTES:

Calcular una cantidad experimental Q_{exp} a partir de los datos
Calcular una cantidad teórica Q_α a partir de las tablas

Si $Q_{exp} < Q_\alpha \Rightarrow$ Se acepta H_0	Si $Q_{exp} \geq Q_\alpha \Rightarrow$ Se rechaza H_0

EL NIVEL MÍNIMO DE SIGNIFICACIÓN (p-valor) es el error de la primera región crítica de rechazo. Es decir, el área que deja a la derecha la cantidad experimental Q_{exp}





CÁLCULO DEL ERROR TIPO I, TIPO II Y POTENCIA.

1.- Sea una variable aleatoria X procedente de una población con densidad de probabilidad $N(\mu, 5)$. Efectuadas dos hipótesis sobre el valor de μ

$$H_0 : \mu = \mu_0 = 12 \qquad H_1 : \mu = \mu_1 = 15$$

mediante un muestreo aleatorio simple de tamaño 25, se contrasta la hipótesis H_0 respecto de la hipótesis H_1, estableciéndose que si la media muestral es menor que 14 se aceptaría la hipótesis nula. Determinar:

a) La probabilidad de cometer el error tipo I

b) La probabilidad de cometer el error tipo II

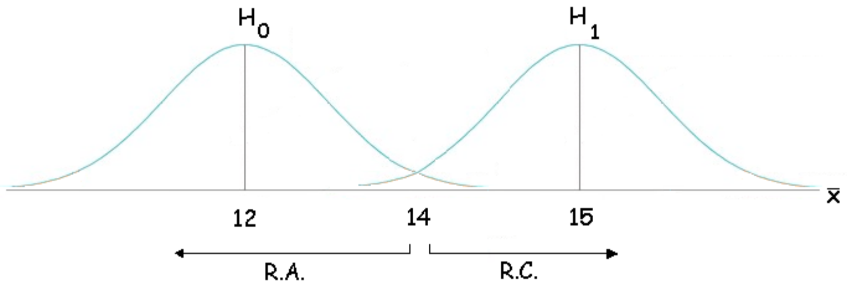
c) La potencia del contraste

Solución:

a) Sea la variable aleatoria $X \sim N(\mu, 5)$

Las hipótesis sobre la media poblacional (contraste unilateral):

$$H_0 : \mu_0 = 12$$
$$H_1 : \mu_1 = 15$$



Regla de decisión $\begin{cases} \bar{x} < 14 & \text{se acepta } H_0 \equiv \text{región aceptación R.A.} \\ \bar{x} \geq 14 & \text{se rechaza } H_0 \equiv \text{región crítica R.C.} \end{cases}$

La distribución de la media muestral $\bar{x}$, de tamaño 25, con la varianza poblacional $\sigma^2 = 25$ conocida:

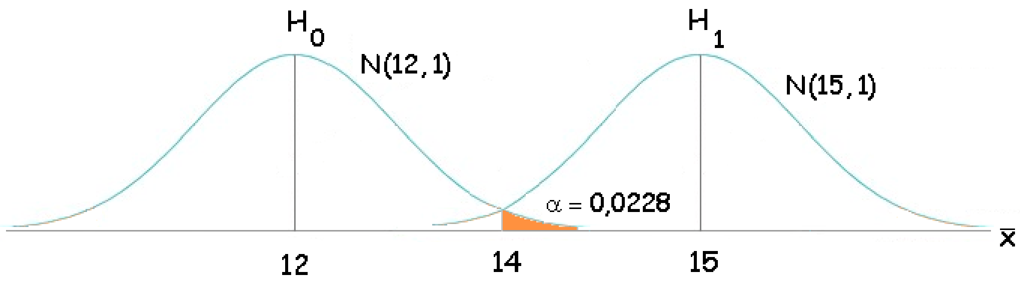
$$H_0 : \quad \bar{x} \sim N\left[\mu_0, \frac{\sigma}{\sqrt{n}}\right] = N\left[12, \frac{5}{\sqrt{25}}\right] = N(12, 1)$$

$$H_1 : \quad \bar{x} \sim N\left[\mu_1, \frac{\sigma}{\sqrt{n}}\right] = N\left[15, \frac{5}{\sqrt{25}}\right] = N(15, 1)$$

Error Tipo I: $\alpha = P(\text{ET I}) = P(\text{Rechazar } H_0 / H_0 \text{ cierta})$

$$\alpha = P(\text{ET I}) = P\left[\bar{x} \geq 14 \mid H_0 : \mu_0 = 12\right] = P\left[z \geq \frac{14-12}{1}\right] = P(z \geq 2) = 0{,}0228$$

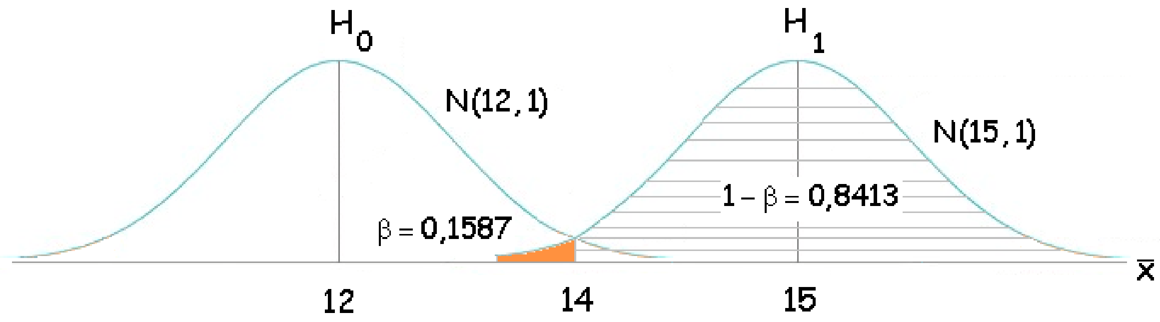
H_0 se rechaza cuando es cierta el 2,28% de los casos



b) Error Tipo II: $\beta = P(\text{ET II}) = P(\text{Aceptar } H_0 \,/\, H_0 \text{ falsa})$

$$\beta = P(\text{ET II}) = P\left[\bar{x} < 14 \mid H_1 : \mu_1 = 15\right] = P\left[z < \frac{14-15}{1}\right] = P(z < -1) = P(z > 1) = 0{,}1587$$

H_0 se acepta cuando es falsa el 15,87% de los casos



c) Potencia del Contraste: $\text{Potencia} = P\left[\text{Rechazar } H_0 \mid H_0 \text{ falsa}\right] = 1-\beta$

$1-\beta = P\left[\text{Rechazar } H_0 \mid H_0 \text{ falsa}\right] = 1 - 0{,}1587 = 0{,}8413$ (bondad del tipo de ajuste)

o bien,

$$\text{Potencia} = P\left[\bar{x} \geq 14 \mid N(15,1)\right] = P\left[\frac{\bar{x}-15}{1} \geq \frac{14-15}{1}\right] = P\left[z \geq -1\right] = 1 - P\left[z \leq -1\right] =$$

$$= 1 - P\left[z \leq -1\right] = 1 - P\left[z \geq 1\right] = 1 - 0{,}1587 = 0{,}8413$$

H_0 se rechaza cuando es falsa el 84,13% de los casos

Resaltar que es más grave cometer un Error Tipo I (α) que un Error Tipo II (β).

2.- Las latas de mejillones de una determinada marca indican que el peso escurrido de dicho producto es de 250 gr. No obstante, un consumidor está convencido de que el peso escurrido medio de dicho producto es menor que el que indican las latas. Si el peso escurrido sigue una ley normal con desviación típica 9 gr.

a) Determinar, si existe, la mejor región crítica para contrastar, con un nivel de significación del 5% y muestras aleatorias simples de tamaño 100.

a) Tomar una decisión acerca del rechazo o no de la hipótesis nula a partir de una muestra aleatoria simple de tamaño 100 en la cual se ha observado un peso escurrido promedio de 245 gr.

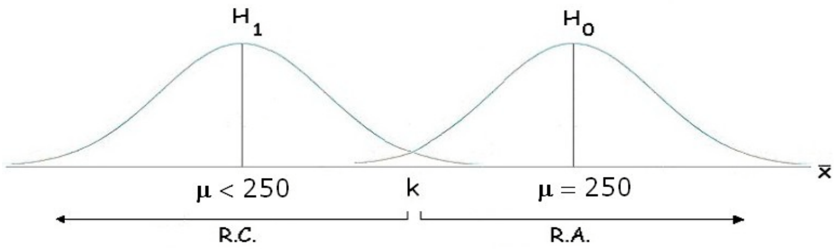
b) Determinar la función de potencia del contraste.

Solución:

a) Sea la variable aleatoria X = "peso escurrido de las latas de mejillones"

Contraste unilateral:

$H_0: \mu = 250 \quad H_1: \mu < 250$



La regla de decisión del muestreo: $\begin{cases} \bar{x} > k & \text{se acepta } H_0 \text{ (R.A.)} \\ \bar{x} \le k & \text{se rechaza } H_0 \text{ (R.C.)} \end{cases}$

Bajo la hipótesis nula, $\quad X \sim N\left[250,\ \dfrac{9}{\sqrt{100}}\right]$

El valor crítico k, bajo la hipótesis nula, se determina con el nivel de significación α:

$$\alpha = P[\text{ET I}] = P[\text{Rechazar } H_0 / H_0 \text{ cierta}] = P[\bar{x} \le k / H_0 \text{ cierta}] =$$

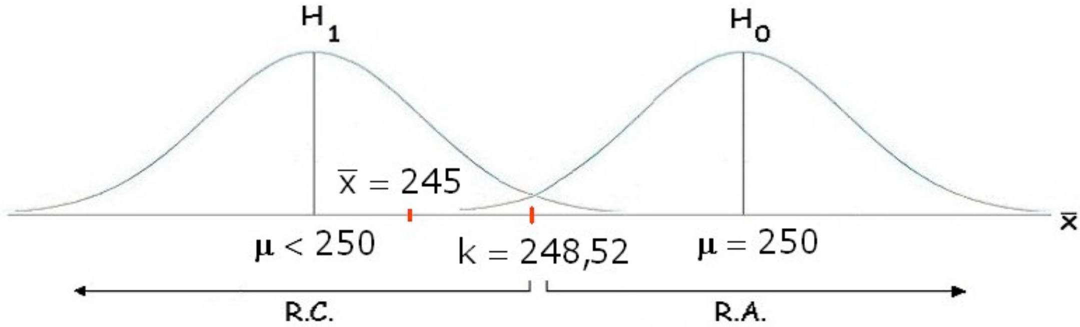
$$= P[\ \bar{x} \le k / N(250; 0,9)\ ] = P\left[\frac{\bar{x} - 250}{0,9} \le \frac{k-250}{0,9}\right] = P\left[z \le \frac{k-250}{0,9}\right] = P\left[z \ge \frac{250-k}{0,9}\right] = 0,05$$

Observando las tablas de la N(0,1) se tiene: $\dfrac{250-k}{0,9} = 1,645 \ \mapsto\ k = 248,52$

La región crítica más potente, para muestras de tamaño 100, es $\bar{x} \le 248,52$

b) Dado que $\bar{x} = 245 < 248,52$, el peso escurrido promedio se encuentra en la región de rechazo de la hipótesis nula.

$\begin{cases} \bar{x} > 248,52 & \text{se acepta } H_0 \text{ (R.A.)} \\ \bar{x} \le 248,52 & \text{se rechaza } H_0 \text{ (R.C.)} \end{cases}$



c) La función potencia del contraste se establece como:

$$P(\mu) = P(\bar{x} \le 248{,}52) = P\left[\ \bar{x} \le k \,/\, N(\mu;\ \frac{9}{\sqrt{100}})\ \right] = P\left[\frac{\bar{x}-\mu}{0{,}9} \le \frac{248{,}52-\mu}{0{,}9}\right] =$$

$$= P\left[\ z \le \frac{248{,}52-\mu}{0{,}9}\right]_{\mu \le 250} \qquad \text{(bondad del tipo de ajuste)}$$

3.- Sea una variable aleatoria X procedente de una población con densidad de probabilidad $N(\mu, 4)$. Se quiere contrastar la hipótesis nula $H_0 : \mu = \mu_0 = 10$ frente a la hipótesis alternativa $H_1 : \mu = \mu_1 = 12$, con un nivel de significación $\alpha = 0{,}05$, con un muestreo simple de tamaño 25. Determinar:

a) La probabilidad de cometer el error tipo II

b) La potencia del contraste

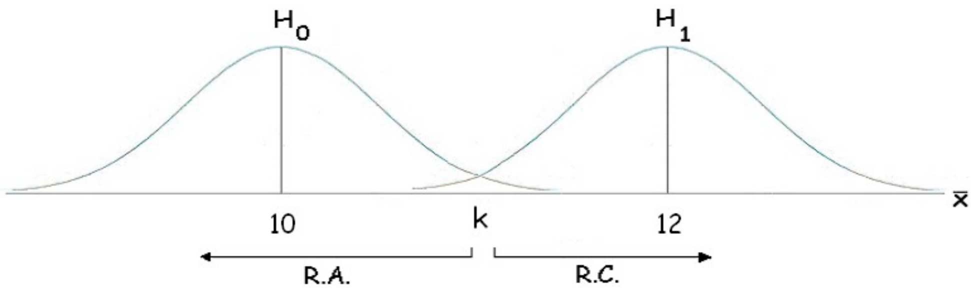
Solución:

Sea la variable aleatoria $X \sim N(\mu, 4)$

Las hipótesis sobre la media poblacional (contraste unilateral):

$H_0 : \mu_0 = 10$

$H_1 : \mu_1 = 12$



$$\text{Regla de decisión} \begin{cases} \bar{x} \le k & \text{se acepta } H_0 \quad \mapsto \quad \text{región aceptación (R.A.)} \\ \bar{x} > k & \text{se rechaza } H_0 \quad \mapsto \quad \text{región crítica (R.C.)} \end{cases}$$

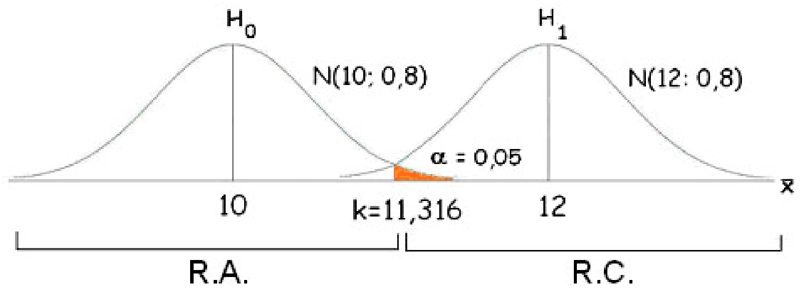
La distribución de la media muestral $\bar{x}$, de tamaño 25, con la varianza poblacional $\sigma^2 = 16$ conocida:

$$H_0 : \quad \bar{x} \sim N\left[\mu_0, \frac{\sigma}{\sqrt{n}}\right] = N\left[10, \frac{4}{\sqrt{25}}\right] = N(10; 0{,}8)$$

$$H_1: \quad \bar{x} \sim N\left[\mu_1, \frac{\sigma}{\sqrt{n}}\right] = N\left[12, \frac{4}{\sqrt{25}}\right] = N(12; 0,8)$$

a) Para hallar el valor crítico 'k' recurrimos al Error Tipo I:

$$\alpha = P[\text{ET I}] = P[\text{Rechazar } H_0 / H_0 \text{ cierta}] = 0,05$$



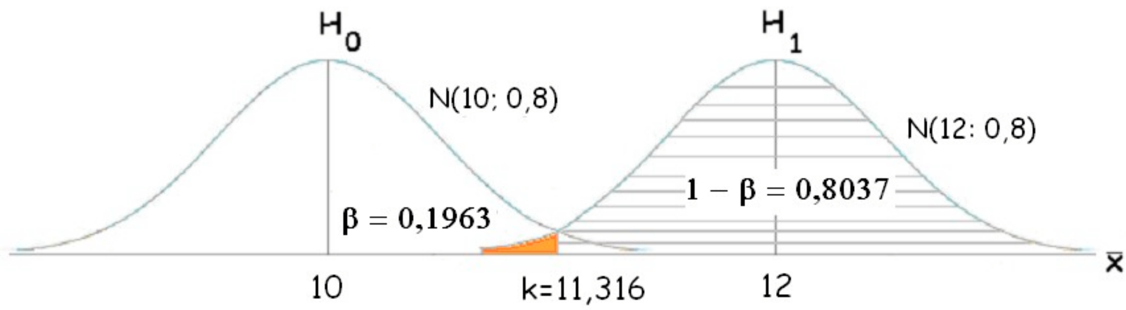
$$\alpha = P[\text{ET I}] = P[\bar{x} > k \mid H_0 : \mu_0 = 10] = P\left[z > \frac{k-10}{0,8}\right] = 0,05$$

$$\frac{k-10}{0,8} = 1,645 \quad \Rightarrow \quad K = 11,316$$

Error Tipo II: $\beta = P[\text{ET II}] = P[\text{Aceptar } H_0 / H_0 \text{ falsa}]$

$$\beta = P[\text{ET II}] = P[\bar{x} \leq 11,316 \mid H_1 : \mu_1 = 12] = P\left[z \leq \frac{11,316 - 12}{0,8}\right] = P(z \leq -0,855) =$$

$$= P(z \geq 0,855) = 0,1963$$



b) Potencia del Contraste: $\text{Potencia} = P[\text{Rechazar } H_0 \mid H_0 \text{ falsa}] = 1 - \beta$

$$1 - \beta = P[\text{Rechazar } H_0 \mid H_0 \text{ falsa}] = 1 - 0,1963 = 0,8037$$

4.- Un agricultor sabe que el peso en kg. de las patatas sigue una distribución $N(\mu, 1)$. Una muestra de patatas dio un peso medio de 330 gramos. Con la muestra se realizó un contraste, con un nivel de significación del 5% y una potencia de 0,6406, en el que la hipótesis nula era $\mu = 0,4$ Kg y la alternativa $\mu = 0,3$ Kg. Se pide:

a) ¿Cuál es el tamaño de la muestra utilizada por el agricultor?.

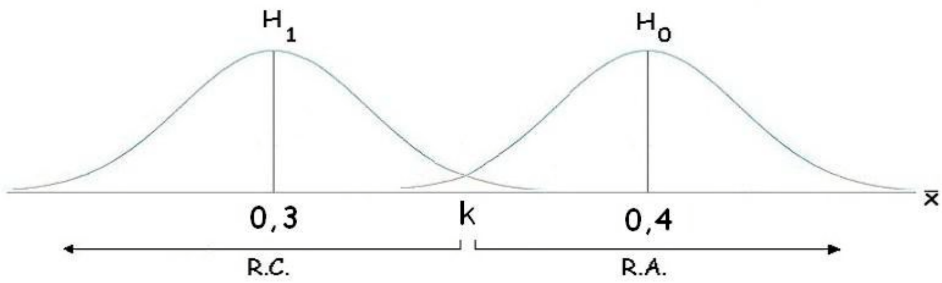
b) Qué hipótesis fue aceptada

Solución:

a) Sea la variable aleatoria X = 'peso en kg. de las patatas". $X \sim N(\mu, 1)$

Hipótesis sobre μ: $\begin{cases} H_0 : \mu_0 = 0,4 \\ H_1 : \mu_1 = 0,3 \end{cases}$

Regla de decisión $\begin{cases} \text{Si } \bar{x} > k \mapsto \text{R.A : se acepta } H_0 \\ \text{Si } \bar{x} \le k \mapsto \text{R.C : se rechaza } H_0 \end{cases}$



La distribución de la media muestral, bajo la hipótesis nula, sigue una ley $N\left[0,4, \frac{1}{\sqrt{n}}\right]$

A partir del nivel de significación, se obtiene el valor crítico k:

$$\alpha = P\left[\text{Rechazar } H_0 \middle| H_0 \text{ cierta}\right] = P\left[\bar{x} \le k \middle| H_0 \text{ cierta}\right] = P\left[\bar{x} \le k \middle| \mu = 0,4\right] =$$

$$= P\left[\frac{\bar{x} - 0,4}{1/\sqrt{n}} \le \frac{k - 0,4}{1/\sqrt{n}}\right] = P\left[\frac{\bar{x} - 0,4}{1/\sqrt{n}} \ge \frac{0,4 - k}{1/\sqrt{n}}\right] = 0,05 \xrightarrow{N(0;1)} \frac{0,4 - k}{1/\sqrt{n}} = 1,64$$

Por otro lado, como la potencia del contraste es 0,6406:

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \middle| H_0 \text{ falsa}\right] \equiv P\left[\text{Rechazar } H_0 \middle| H_1 \text{ cierta}\right]$$

La media muestral, bajo la hipótesis alternativa, sigue una ley $N\left(0,3, \frac{1}{\sqrt{n}}\right)$, con lo cual,

$$\text{Potencia} = P\left[\frac{\bar{x} - 0,3}{1/\sqrt{n}} \le \frac{k - 0,3}{1/\sqrt{n}}\right] = 0,6406 \mapsto P\left[z > \frac{k - 0,3}{1/\sqrt{n}}\right] = 0,3594 \xrightarrow{N(0,1)} \frac{k - 0,3}{1/\sqrt{n}} = 0,36$$

Resolviendo el sistema: $\begin{cases} 0,4-k=\dfrac{1,64}{\sqrt{n}} \\ \\ k-0,3=\dfrac{0,36}{\sqrt{n}} \end{cases}$ $\quad \sqrt{n}=2/0,1 \mapsto n=400$

El tamaño de la muestra de patatas utilizada por el agricultor es de 400 patatas.

Se determina el valor crítico k: $\dfrac{0,4-k}{1/\sqrt{400}}=1,64 \longrightarrow k=0,318$

b) Regla de decisión $\begin{cases} \text{Si } \bar{x}>0,318 \mapsto \text{R.A : se acepta } H_0 \\ \text{Si } \bar{x}\le 0,318 \mapsto \text{R.C : se rechaza } H_0 \end{cases}$

Siendo $\bar{x}=0,330>0,318$ se acepta la hipótesis nula de que el peso medio de las patatas es de 400 gramos.

CONTRASTE UNILATERAL DE LA PROPORCIÓN. HIPÓTESIS SIMPLES.

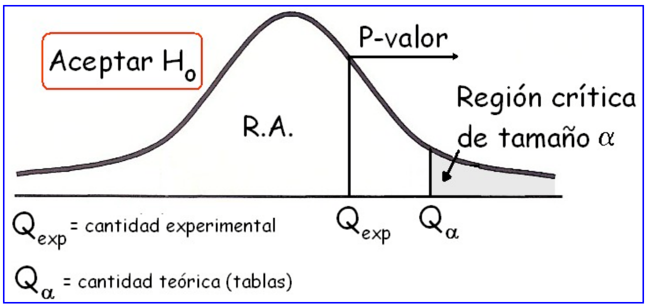
5.- Define p-valor y pon un ejemplo práctico en un contraste bilateral con una muestra pequeña.

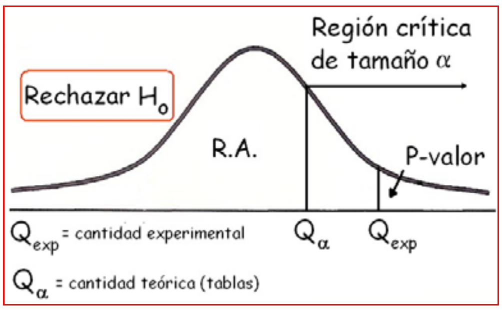
Solución:

El p–valor (α_p) es la probabilidad del error que se comete al rechazar la hipótesis nula, cuantificando el riesgo. Es decir, El p–valor (α_p) es el menor nivel de significación para el que se rechaza la hipótesis nula.

$$p-\text{valor}=\alpha_p<\alpha \quad \mapsto \quad \text{Se rechaza } H_0$$
$$p-\text{valor}=\alpha_p>\alpha \quad \mapsto \quad \text{Se acepta } H_0$$

El p-valor es el error de la primera región crítica de rechazo. Es decir, el área que deja a la derecha la cantidad experimental Q_{exp} a partir de los datos, siendo Q_α la cantidad teórica a partir de las tablas.



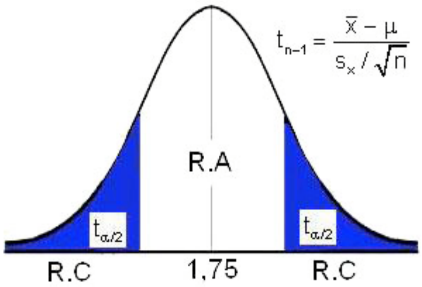


PRÁCTICO: Calcular el p-valor del contraste: $H_0:\mu=1,75$, $H_1:\mu\neq 1,75$, suponiendo la muestra $x_1=1,5$, $x_1=2$ y $x_3=2,5$

Con los datos de la muestra: $\bar{x} = 2 \quad s_x = 0,5$

El estadístico muestral $t_{n-1} = t_2 = \dfrac{|\bar{x} - \mu_0|}{s_x / \sqrt{n}} = \dfrac{|2 - 1,75|}{0,5 / \sqrt{3}} = 0,866$

Al ser el contraste bilateral, el p–valor (α_p) viene dado por la expresión $P(t_2 \geq 0,866) = \dfrac{\alpha_p}{2}$



En la tabla de la t de Student: $P(t_2 \geq 0,617) = 0,3$ y $P(t_2 \geq 1,061) = 0,2$

Interpolando se tiene:

0,3	$\alpha_p / 2$	0,2
0,617	0,866	1,061

$$0,3 - 0,2 \longrightarrow 0,617 - 1,061$$
$$\alpha_p / 2 - 0,2 \longrightarrow 0,866 - 1,061$$

$$\left(\frac{\alpha_p}{2} - 0,2\right) \text{x} (0,617 - 1,061) = (0,3 - 0,2) \text{ x } (0,866 - 1,061) \mapsto \frac{\alpha_p}{2} = 0,2439 \mapsto \alpha_p = 0,4878$$

Para un nivel de significación $\alpha = 0,05$, siendo $\alpha_p = 0,4878 > \alpha = 0,05$, se acepta la hipótesis nula $H_0 : \mu = 1,75$.

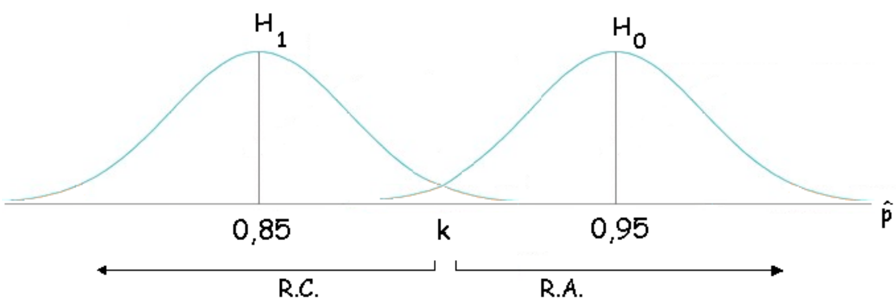
6.- Un laboratorio farmacéutico quiere lanzar un nuevo medicamento para la hipertensión, llamado Hipotensil. El director de dicho laboratorio cree que la eficacia del medicamento sería de un 95%, medida ésta como la proporción de pacientes a los que se les suministra y que experimentan una mejoría. Sin embargo, el inspector de sanidad del Ministerio no es tan optimista y opina que la eficacia es sólo del 85%. Para analizar la eficacia del medicamento antes de su comercialización, se selecciona una muestra aleatoria de 500 pacientes, a los que se les administra Hipotensil, de los cuales mejoran 467. ¿Tiene razón el director del laboratorio?.
Suponga un nivel de significación del 5%.

Solución:

Sea la variable aleatoria X = 'eficacia Hipotensil' $\sim B(1, p)$

Las hipótesis sobre la proporción (contraste unilateral):

$$H_0: \ p_0 = 0,95 \qquad H_1: \ p_1 = 0,85$$



Regla de decisión para el valor crítico k $\begin{cases} \hat{p} > k & \text{se acepta } H_0 \ \ (R.A.) \\ \hat{p} \le k & \text{se rechaza } H_0 \ \ (R.C.) \end{cases}$

La distribución en el muestreo del estadístico $\hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n} \overset{T.C.L.}{\sim} N\left[p, \sqrt{\dfrac{p.q}{n}}\right]$

En consecuencia:

Bajo la hipótesis $H_0 : \hat{p}_0 \sim N\left[0,95, \sqrt{\dfrac{0,95.0,05}{500}}\right] = N(0,95,\ 0,00974)$

Bajo la hipótesis $H_1 : \hat{p}_1 \sim N\left[0,85, \sqrt{\dfrac{0,85.0,15}{500}}\right] = N(0,85,\ 0,01597)$

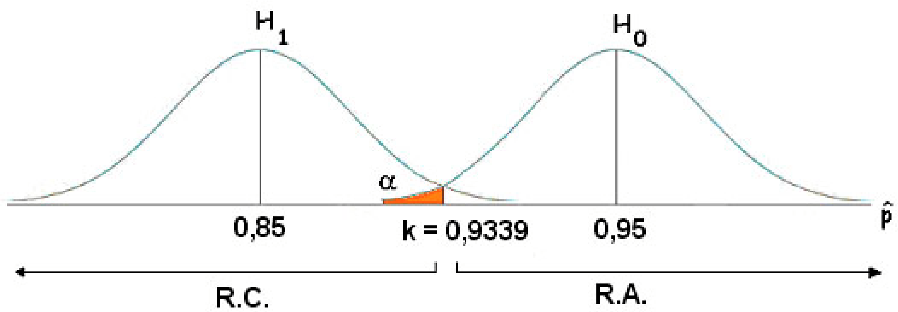
Se determina el valor crítico k a partir del nivel de significación α:

$$\alpha = P[ET\ I] = P\left[\text{Rechazar } H_0 \middle| H_0 \text{ cierta}\right] = P\left[\hat{p} \le k\right] = P\left[z \le \frac{k - 0,95}{0,00974}\right] = 0,05 \quad \mapsto$$

$$\mapsto \quad P\left[z \ge \frac{0,95 - k}{0,00974}\right] = 0,05$$

Observando en las tablas de la normal $N(0,1)$, resulta:

$$\frac{0,95 - k}{0,00974} = 1,645 \ \Rightarrow \ k = 0,95 - (0,00974 \ .\ 1,645) = 0,9339$$



Regla de decisión $\begin{cases} \hat{p} > 0,9339 & \text{se acepta } H_0 \ \ (R.A.) \\ \hat{p} \le 0,9339 & \text{se rechaza } H_0 \ \ (R.C.) \end{cases}$

El valor del estadístico muestral $\hat{p}$ (evidencia empírica), $\hat{p} = \dfrac{467}{500} = 0,934$, se tiene:

$\hat{p} = 0,934 > 0,9339$, concluyendo que existe evidencia empírica suficiente para aceptar la hipótesis nula H_0. Es decir, el Hipotensil es eficaz en un 95% de los casos.

7.- Se trata de determinar si en una ciudad el 20% o el 30% de las familias dispone de lavavajillas; para dilucidarlo se toma al azar una muestra de 400 familias de la mencionada ciudad y se adopta el criterio de si en la muestra hay menos de 100 familias con lavavajillas, se rechaza que el 20% de las familias poseen el mencionado electrodoméstico. Se pide:

a) Nivel de significación del test.

b) Potencia del test.

Solución:

a) Sea el parámetro p = "proporción de familias con lavavajillas"

Al realizar un contraste sobre una proporción, partimos de una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de tamaño $n = 400$, donde $X \sim B(1, p)$.

Para calcular la probabilidad interesa conocer la distribución del parámetro muestral

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{donde} \quad x_i = \begin{cases} 1 & \text{si la familia tiene lavavajillas} \\ 0 & \text{si la familia no tiene lavavajillas} \end{cases}$$

Al ser el tamaño suficientemente grande $n = 400$ y estar definido $\hat{p}$ como suma de variables independientes según una distribución de Bernouilli $B(1, p)$, se puede aproximar la distribución muestral de $\hat{p}$ como :

$$\overline{x} = \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} \overbrace{\sim}^{\text{T.C.L.}} N\left[p, \sqrt{\frac{p.q}{n}}\right]$$

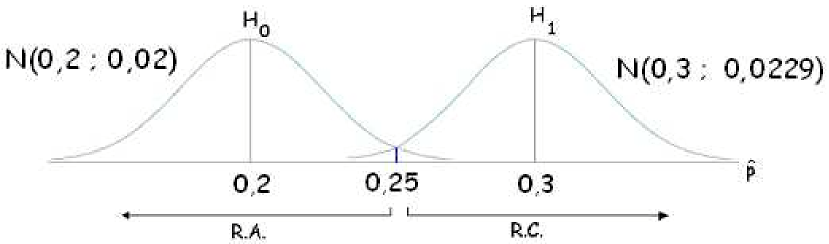
En el contraste: La hipótesis nula $H_0: p = 0,2$ y la hipótesis alternativa $H_1: p = 0,3$

Por el lema de Neyman-Pearson, la regla de decisión del muestreo ($\hat{p} = \frac{100}{400} = 0,25$):

$$\begin{cases} \hat{p} < 0,25 & \text{se acepta } H_0 \text{ (R.A.)} \\ \hat{p} \geq 0,25 & \text{se rechaza } H_0 \text{ (R.C.)} \end{cases}$$

Con la hipótesis nula $H_0: p = 0,2: \ \hat{p} \sim N\left[0,2\ , \ \sqrt{\frac{0,2.0,8}{400}}\right] = N(0,2\ ,\ 0,02)$

Con la hipótesis alternativa $H_1: p = 0,3: \ \hat{p} \sim N\left[0,3\ , \ \sqrt{\frac{0,3.0,7}{400}}\right] = N(0,3\ ,\ 0,0229)$



Bajo la hipótesis nula, con el valor crítico k = 0,25, se determina el nivel de significación α:

$$\alpha = P[ET\ I] = P[\text{Rechazar } H_0 / H_0 \text{ cierta}] = P[\hat{p} > 0,25 / H_0 \text{ cierta}] =$$

$$= P[\hat{p} \geq 0,25 / N(0,2;\, 0,02)] = P\left[\frac{\hat{p} - 0,2}{0,02} \geq \frac{0,25 - 0,2}{0,02}\right] = P[\, z \geq 2,5] = 0,00621$$

b) Potencia del Contraste: $\text{Potencia} = P[\text{Rechazar } H_0 \mid H_0 \text{ falsa}] = 1 - \beta$

Error Tipo II: $\beta = P[ET\ II] = P[\text{Aceptar } H_0 / H_0 \text{ falsa}]$

$$\beta = P[ET\ \ II] = P[\,\hat{p} < 0,25\ / N(0,3,\, 0,0229)] = P\left[\frac{\hat{p} - 0,3}{0,0229} < \frac{0,25 - 0,3}{0,0229}\right] = P[\, z < -2,1822] =$$

$$= P[\, z > 2,1822] = 0,0144$$

En consecuencia, bondad del tipo de ajuste:

$$\text{Pot} = 1 - \beta = P[\text{Rechazar } H_0 \mid H_0 \text{ falsa}] = 1 - 0,0144 = 0,9856$$

CONTRASTE UNILATERAL DE LA VARIANZA CON MEDIA POBLACIONAL CONOCIDA.

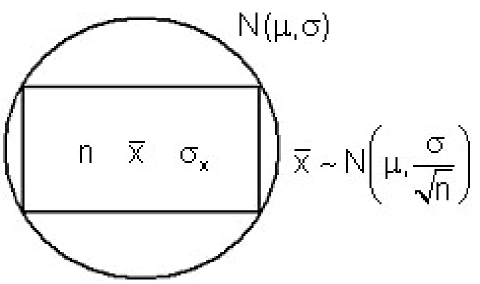
8.- Las especificaciones de un tipo de báscula aseguran que los errores de los pesajes siguen una distribución $N(0,\, \sigma)$. Se quiere contrastar la afirmación sobre la dispersión que es igual a la unidad, frente a una hipótesis alternativa de que es el doble. Para ello se realizan 5 pesajes en las que el error cometido resultó ser:

1	0,9	- 0,2	1,4	- 0,7

Para un nivel de significación del 5% se pide enunciar una regla de decisión (obtener la región crítica) e indicar que hipótesis resulta aceptada.

Solución:

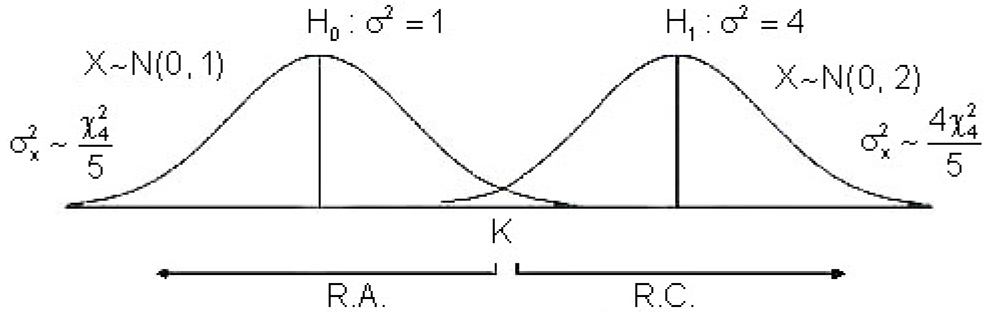
Sea la variable aleatoria X = 'Errores en el peso' $X \sim N(0,\, \sigma)$



Lema de Fisher:

$$\chi^2_{n-1} \sim \frac{\text{varianza muestral}}{\text{varianza teórica}} = \frac{\sigma_x^2}{(\sigma / \sqrt{n})^2} \quad \mapsto \quad \sigma_x^2 = \frac{\sigma^2\, \chi^2_{n-1}}{n}$$

Hipótesis sobre σ^2: $H_0 : \sigma^2 = 1 \qquad H_1 : \sigma^2 = 4$



$$\chi_{n-1}^2 = \frac{\text{varianza muestral}}{\text{varianza teórica}} = \frac{\sigma_x^2}{\left(\sigma / \sqrt{n}\right)^2} \quad \mapsto \quad \sigma_x^2 = \frac{\sigma^2\,\chi_{n-1}^2}{n}$$

Regla de decisión: $\begin{cases} \text{Si } \sigma_x^2 < k \ \mapsto \ \text{R.A : Aceptar } H_0 \\ \text{Si } \sigma_x^2 \geq k \ \mapsto \ \text{R.C : Rechazar } H_0 \end{cases}$

Por el Lema de Fisher, bajo la hipótesis nula, $\sigma_x^2 \sim \dfrac{\sigma^2.\chi_{n-1}^2}{n} = \dfrac{1.\ \chi_{5-1}^2}{5} = \dfrac{\chi_4^2}{5}$

El valor crítico k se determina a partir de α (Error Tipo I):

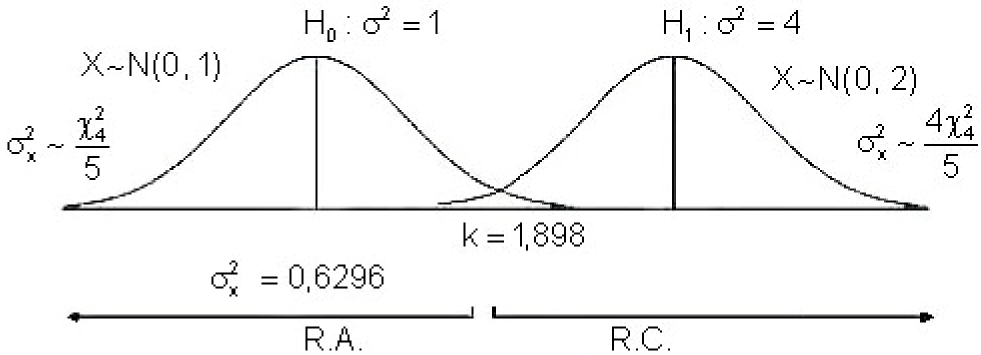
$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 \geq k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 \geq k \,\middle|\, \sigma^2 = 1\right] =$$

$$= P\left[\frac{\chi_4^2}{5} \geq k\right] = P\left[\chi_4^2 \geq 5k\right] = 0{,}05 \xrightarrow{\chi^2 \text{ Pearson}} 5k = 9{,}488 \ \mapsto \ k = 1{,}898$$

Se compara el valor crítico $(k = 1{,}898)$ con la evidencia muestral

x_i	1	0,9	- 0,2	1,4	- 0,7	$\bar{x} = \sum x_i / 5 = 0{,}48$
x_i^2	1	0,81	0,04	1,96	0,49	$\sum x_i^2 / 5 = 0{,}86$

$$\sigma_x^2 = \left[\sum x_i^2 / 5\right] - \left[\sum x_i / 5\right]^2 = 0{,}6296$$



Como $\sigma_x^2 = O{,}6296 < 1{,}898$ se sitúa en la región de aceptación (R.A), no se puede rechazar la hipótesis de que la dispersión sea 1, con un nivel de confianza del 95%.

CONTRASTE UNILATERAL DE LA VARIANZA CON LA MEDIA POBLACIONAL CONOCIDA. POTENCIA DEL CONTRASTE.

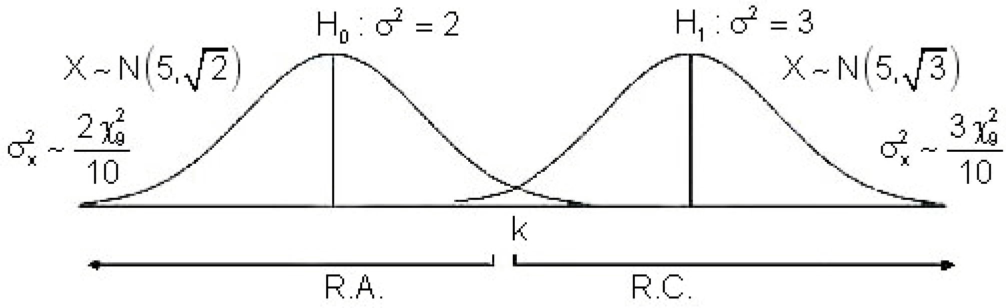
9.- En una población $N(5,\sigma)$ se quiere contrastar la hipótesis nula $H_0 : \sigma^2 = 2$ frente a la hipótesis alternativa $H_1 : \sigma^2 = 3$, con un nivel de significación $\alpha = 0{,}025$, con una muestra aleatoria simple de tamaño 10:

5,1	6,2	4	2,8	2,9	5,6	3,7	3,4	2,5	5,2

Hallar la potencia del contraste.

Solución:

Hipótesis sobre σ^2 : $H_0 : \sigma^2 = 2$ $\quad H_1 : \sigma^2 = 3$



$$\chi_{n-1}^2 = \frac{\text{varianza muestral}}{\text{varianza teórica}} = \frac{\sigma_x^2}{(\sigma/\sqrt{n})^2} \mapsto \sigma_x^2 = \frac{\sigma^2\,\chi_{n-1}^2}{n}$$

Regla de decisión: $\begin{cases} \text{Si } \sigma_x^2 > k \mapsto \text{R.C : Rechazar } H_0 \\ \text{Si } \sigma_x^2 \le k \mapsto \text{R.A : Aceptar } H_0 \end{cases}$

En el muestreo, bajo la hipótesis nula $\sigma^2 = 2$, con tamaño muestral $n = 10$, se tiene:

$$\sigma_x^2 \sim \frac{\sigma^2 \cdot \chi_{n-1}^2}{n} = \frac{\overbrace{2 \cdot \chi_9^2}^{H_0}}{10} \qquad \sigma_x^2 \sim \frac{3\chi_9^2}{10}$$

La determinación del valor crítico k a partir de α (Error Tipo I):

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, \sigma^2 = 2\right] =$$

$$= P\left[\frac{2.\chi_9^2}{10} > k\right] = P\left[\chi_9^2 > 5k\right] = 0{,}025 \xrightarrow{\text{Tablas } \chi^2 \text{ Pearson}} 5k = 19{,}023 \mapsto k = 3{,}8046$$

Se compara el valor crítico $(k = 3{,}8046)$ con la evidencia muestral $\sigma_x^2 = 1{,}5204$

Como $\sigma_x^2 = 1{,}5204 < 3{,}8046$ nos situamos en la región de aceptación (R.A), aceptando la hipótesis de que la varianza es 2, con un nivel de confianza del 97,5%.

De otra parte, la potencia del contraste:

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ falsa}\right] \equiv P\left[\text{Rechazar } H_0 \,\middle|\, H_1 \text{ cierta}\right]$$

En el muestreo, bajo la hipótesis alternativa $\sigma_0^2 = 3$, con tamaño muestral $n = 10$, se tiene:

$$\sigma_x^2 \sim \frac{\sigma_0^2 \cdot \chi_{n-1}^2}{n} = \frac{\overbrace{3 \cdot \chi_9^2}^{H_1}}{10}$$

$$\text{Potencia} = P\left[\frac{3 \cdot \chi_9^2}{10} > 3{,}8046 \,/\, H_1 \text{ cierta}\right] = P\left[\chi_9^2 \geq 12{,}682\right] = 0{,}252$$

Abscisas	Áreas
4,168 – 14,684	0,90 – 0,10
12,682 – 14,684	x – 0,10

$$x = 0{,}10 + \frac{2{,}002 \,.\, 0{,}80}{10{,}516} = 0{,}252$$

CONTRASTE UNILATERAL DE LA VARIANZA CON MEDIA POBLACIONAL DESCONOCIDA. POTENCIA DEL CONTRASTE.

10.- En una población con distribución $N(\mu, \sigma)$ se desea contrastar la hipótesis nula $H_0 : \sigma^2 = 25$ frente a la hipótesis alternativa $H_1 : \sigma^2 = 36$, se toma una muestra aleatoria de tamaño 16, con varianza igual a 27.

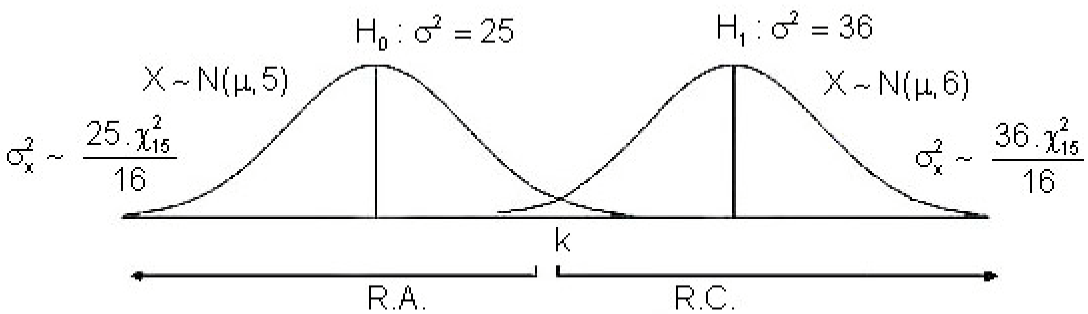
a) Tiene sentido utilizar el nivel de significación del 1%

b) En caso negativo, utiliza el nivel de confianza más grande posible

c) Hallar la potencia del contraste

Solución:

Hipótesis sobre σ^2 : $H_0 : \sigma^2 = 25$ $\quad H_1 : \sigma^2 = 36$ $\quad \sigma_1^2 > \sigma_0^2$



$$\chi_{n-1}^2 = \frac{\text{varianza muestral}}{\text{varianza teórica}} = \frac{\sigma_x^2}{\left(\sigma/\sqrt{n}\right)^2} \;\mapsto\; \sigma_x^2 = \frac{\sigma^2\, \chi_{n-1}^2}{n}$$

Regla de decisión: $\begin{cases} \text{Si } \sigma_x^2 > k \;\mapsto\; \text{R.C : Rechazar } H_0 \\ \text{Si } \sigma_x^2 \leq k \;\mapsto\; \text{R.A : Aceptar } H_0 \end{cases}$

Por el Lema de Fisher-Cochran: $\sigma_x^2 \sim \dfrac{\sigma^2 \,.\, \chi_{n-1}^2}{n}$. Bajo la hipótesis nula $\sigma^2 = 25$, con tamaño muestral $n = 16$: $\sigma_x^2 \sim \dfrac{25 \cdot \chi_{15}^2}{16}$

La determinación del valor crítico k a partir de $\alpha = 0,01$

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, \sigma^2 = 2\right] =$$

$$= P\left[\frac{25.\chi_{15}^2}{16} > k\right] = P\left[\chi_{15}^2 > \frac{16}{25}\,k\right] = 0,01 \xrightarrow{\chi^2 \text{ Pearson}} \frac{16}{25}\,k = 30,578 \;\mapsto\; k = 47,778$$

Adviértase que el valor crítico k se encuentra fuera del intervalo $[25, 36]$ con lo que no tiene sentido utilizar un nivel de confianza del 99%

a) Para que el valor crítico se encuentre dentro del rango deseado se puede tomar un nivel de confianza del 90%, con lo cual $\alpha = 0,10$, resultando:

$$\alpha = P\left[\frac{25.\chi_{15}^2}{16} > k\right] = P\left[\chi_{15}^2 > \frac{16}{25}\,k\right] = 0,10 \xrightarrow{\chi^2 \text{ Pearson}} \frac{16}{25}\,k = 22,307 \;\mapsto\; k = 34,855$$

Se compara el valor crítico $(k = 34,855)$ con la evidencia muestral $\sigma_x^2 = 27$

Como $\sigma_x^2 = 27 < 34,855$ nos situamos en la región de aceptación (R.A), aceptando la hipótesis de que la varianza es 25, con un nivel de confianza del 90%.

b) La potencia del contraste con $\alpha = 0,10$

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ falsa}\right] \equiv P\left[\text{Rechazar } H_0 \,\middle|\, H_1 \text{ cierta}\right]$$

En el muestreo, bajo la hipótesis alternativa $\sigma^2 = 36$, con tamaño muestral $n = 16$, se tiene: $\sigma_x^2 \sim \dfrac{\sigma^2.\chi_{n-1}^2}{n} = \dfrac{\overbrace{36}^{H_1}.\chi_{15}^2}{16}$

$$\text{Potencia} = P\left[\frac{36.\chi_{15}^2}{16} > 34,855 \,/\, H_1 \text{ cierta}\right] = P\left[\chi_{15}^2 \geq 15,491\right] = 0,496$$

Abscisas	Áreas
8,547 − 22,307	0,90 − 0,10
15,491 − 22,307	x − 0,10

$$x = 0,10 + \frac{6,816 . 0,80}{13,76} = 0,496$$

CONTRASTE UNILATERAL DE LA MEDIA CON VARIANZA DESCONOCIDA. CALCULAR Y REPRESENTAR LA FUNCIÓN DE POTENCIA.

11.- El propietario de un automóvil sospecha que su vehículo tiene un consumo medio de combustible en carretera superior a los 5,6 litros /100 Km., que es lo que el fabricante indica en su publicidad. Para apoyar empíricamente su sospecha observa el consumo medio en 11 viajes seleccionados aleatoriamente entre todos los que realiza en el año, obteniendo los siguientes resultados:

6,1	6,5	5,1	6	5,9	5,2	5,8	5,3	6,2	5,9	6,3

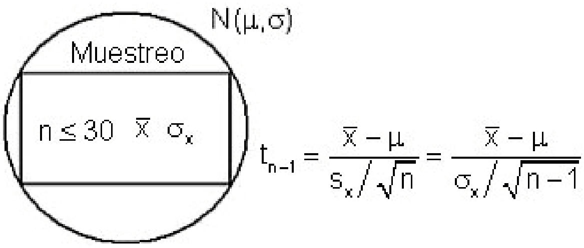
Se pide:

a) ¿Están fundadas las sospechas del propietario a un nivel de significación del 1%?

b) ¿En cuántas ocasiones debería observarse el consumo medio para que con un nivel de confianza del 99% se detectase un consumo medio de 5,9 litros/100 km.?

Solución:

a) Se supone que el consumo medio del automóvil sigue una distribución normal $N(\mu, \sigma)$, ambos parámetros desconocidos.



En el muestreo de una población normal con varianza desconocida, desviación típica muestral (σ_x), cuasidesviación típica (s_x), la variable: $t_{n-1} = \dfrac{\bar{x}-\mu}{s_x/\sqrt{n}} = \dfrac{\bar{x}-\mu}{\sigma_x/\sqrt{n-1}}$

El fabricante afirma que $H_0: \mu = 5,6$ y el propietario del vehículo cree que $H_1: \mu > 5,6$.

Se trata, pues, de un contraste unilateral, donde H_1 es compuesta.

Regla decisión: $\begin{cases} \text{Si } \bar{x} \le k \mapsto \text{R.A : Aceptar } H_0 \\ \text{Si } \bar{x} > k \mapsto \text{R.C : Rechazar } H_0 \end{cases}$

Bajo la hipótesis nula, con los datos muestrales ($\bar{x} = 5,8454$, $s_x = 0,4612$), el muestreo sigue una distribución $t_{10}\left[5,6; \dfrac{0,4612}{\sqrt{11}}\right] \equiv t_{10}(5,6\ ;\ 0,1390)$

El valor crítico k, bajo la hipótesis nula, se calcula a partir del nivel de significación α:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\bar{x} > k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\bar{x} > k \,\middle|\, \mu_0 = 5,6\right] =$$

$$= P\left[\frac{\bar{x}-5,6}{0,1390} > \frac{k-5,6}{0,1390}\right] = P\left[t_{10} > \frac{k-5,6}{0,1390}\right] = 0,01 \Rightarrow \frac{k-5,6}{0,1390} = 2,764 \quad \mapsto \quad \boxed{k = 5,9842}$$

Siendo $\bar{x} = 5,8454 < 5,9842$ no se puede rechazar la hipótesis nula H_0, con lo que se acepta las afirmaciones del fabricante sobre el consumo medio del automóvil.

b) En esta ocasión se plantea la cuestión $P(\overline{x} > k \,/\, \mu = 5,9)$

Como el tamaño muestral es desconocido, el estadístico $\dfrac{\overline{x} - 5,9}{s_x \big/ \sqrt{n}} \neq t_{10}$ no sigue una t-Student con 10 grados de libertad.

Por tanto, conociendo que $P\left[\dfrac{\overline{x} - 5,6}{s_x \big/ \sqrt{n}} > 2,764\right] = 0,01$, se recurre a la siguiente estrategia:

$$P\left[\overline{x} > k \,/\, \mu = 5,9\right] = P\left[\frac{\overline{x} - 5,6}{s_x \big/ \sqrt{n}} > 2,764 \,\Big|\, \mu = 5,9\right] P\left[\frac{\overline{x} - 5,9 + 0,3}{s_x \big/ \sqrt{n}} > 2,764\right] =$$

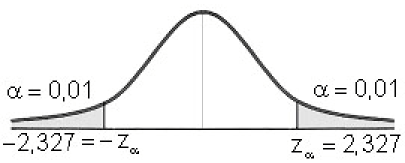
$$= P\left[\frac{\overline{x} - 5,9}{s_x \big/ \sqrt{n}} + \frac{0,3}{s_x \big/ \sqrt{n}} > 2,764\right] = P\left[\frac{\overline{x} - 5,9}{s_x \big/ \sqrt{n}} > 2,764 - \frac{0,3}{s_x \big/ \sqrt{n}}\right] = 0,99$$

o bien, $P\left[\dfrac{\overline{x} - 5,9}{s_x \big/ \sqrt{n}} \leq 2,764 - \dfrac{0,3}{s_x \big/ \sqrt{n}}\right] = 0,01$

En un punto donde no se pueden utilizar las tablas de la t-Sudent porque no se conoce el tamaño de la muestra.

Para ello, se supone que el tamaño es suficientemente grande para que pueda ser aceptable la aproximación de la t mediante la distribución normal.

Con la aproximación normal, y con la simetría de la N(0, 1):



$$P\left[z \leq 2,764 - \frac{0,3}{s_x \big/ \sqrt{n}}\right] = 0,01 \;\Rightarrow\; 2,764 - \frac{0,3}{s_x \big/ \sqrt{n}} = -2,327$$

$$\frac{0,3}{s_x \big/ \sqrt{n}} = 5,091 \;\mapsto\; n = \left[\frac{(5,091)\,.\,s_x}{0,3}\right]^2$$

Otra vez en un callejón sin salida, aún es necesario conocer la cuasidesviación típica de una muestra sin saber su tamaño.

Se puede dar una salida, suponiendo que la cuasidesviación típica de esta nueva muestra es igual a la obtenida en la muestra anterior $s_x = 0,4612$.

En este caso, $n = \left[\dfrac{(5,091).(0,4612)}{0,3}\right]^2 = 61,255$

Para mayor seguridad en el logro del objetivo, se redondea con el entero inmediato superior, esto es, el tamaño de la muestra es 62.

12.- El directorio de uno de los grandes operadores de Internet está considerando la posibilidad de ofrecer tarifa plana a sus clientes. Según sus conocimientos sobre el tema, sabe que está trabajando con una variable aleatoria que se distribuye como una normal. Mantiene la hipótesis de que los hogares que tienen Internet se conectan con una media de 5 horas mensuales. No obstante, existen otros estudios que sostienen que el tiempo de conexión es más alto. Para evaluar, a un 10% de significación, dicha hipótesis, el directorio decide encuestar a una muestra aleatoria de 30 hogares, obteniendo una media de 5,34 horas de conexión, con una dispersión de 7,24 horas.

a) Formular el contraste a realizar.

b) Determinar la mejor región crítica del contraste.

c) ¿Se puede rechazar la hipótesis nula?

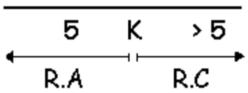
d) Calcular y representar la función de potencia. ¿Qué representa la función de potencia?. Para facilitar los cálculos, suponer que el tamaño muestral es 300 hogares.

Solución:

a) Sea la variable aleatoria X = 'Conexión a Internet por hogares' $X \sim N(5, \sigma)$

Las hipótesis sobre la media poblacional μ con σ^2 desconocida:

Hipótesis sobre μ : $\begin{cases} H_0 : \mu_0 = 5 \\ H_1 : \mu_1 > 5 \end{cases}$



b) Se trata de un contraste unilateral, siendo H_1 compuesta.

Regla de decisión: $\begin{cases} \text{Si } \ \overline{x} > k \ \mapsto \ \text{R.C : Rechazar } H_0 \\ \text{Si } \ \overline{x} \leq k \ \mapsto \ \text{R.A : Aceptar } H_0 \end{cases}$

En el muestreo de una población normal con varianza desconocida, y desviación típica muestral σ_x, la variable: $\dfrac{\overline{x} - \mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} = \dfrac{\overline{x} - \mu}{\dfrac{s_x}{\sqrt{n}}} = t_{n-1} \quad \mapsto \quad \overline{x} = \mu + t_{n-1} \dfrac{\sigma_x}{\sqrt{n-1}}$

c) Determinación de k a partir del nivel de significación α:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\overline{x} > k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\overline{x} > k \,\middle|\, \mu_0 = 5\right] =$$

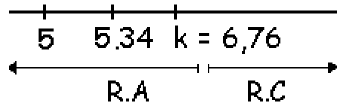
$$P\left[\mu_0 + t_{n-1} \frac{\sigma_x}{\sqrt{n-1}} > k \,\middle|\, \mu_0 = 5\right] = P\left[5 + t_{30-1} \frac{7,24}{\sqrt{30-1}} > k\right] = P\left[t_{29} > \frac{(k-5).\,\sqrt{29}}{7,24}\right] = 0,10$$

En las tablas de la t de Student: $t_{0,10\,;\,29} = 1,311$, con lo cual,

$$\frac{(k-5).\,\sqrt{29}}{7,24} = 1,311 \;\Rightarrow\; k = 5 + \frac{1,311.\,7,24}{\sqrt{29}} = 6,76$$

Comparando k con el valor del estadístico muestral $\overline{x}_0 = 5,34$

El valor $\overline{x}_0 = 5{,}34 < 6{,}76$, no pudiendo rechazar H_0, por lo que tiene razón el Directivo, el tiempo medio de conexión es de 5 horas.



Alternativamente, con el estadístico de contraste (valor experimental), la región de rechazo R viene dada por la expresión:

$$R = \left\{ \overline{x} - \mu_0 > t_{\alpha;\,(n-1)} \frac{s_x}{\sqrt{n}} \right\} \equiv \left\{ \overline{x} - \mu_0 > t_{\alpha;\,(n-1)} \frac{\sigma_x}{\sqrt{n-1}} \right\}$$

$$R = \left\{ 5{,}34 - 5 > \overbrace{1{,}311}^{t_{0,10\,;29}} \frac{7{,}24}{\sqrt{29}} \right\} \equiv \left[0{,}34 > 1{,}763 \right]$$

Como no se verifica la región de rechazo, no existe evidencia significativa para rechazar la hipótesis H_0 con un nivel de confianza del 90%.

Una forma análoga de enfocar el problema consistiría en aceptar la hipótesis H_0 cuando el estadístico experimental fuera menor o igual que el estadístico teórico, es decir

Se acepta H_0 cuando se verifica: $\overbrace{t_{n-1} = \dfrac{\overline{x} - \mu}{\dfrac{\sigma_x}{\sqrt{n-1}}}}^{\text{estadístico experimental}} \leq \overbrace{t_{\alpha;\,n-1}}^{\text{estadístico teórico}}$

En este caso, $t_{29} = \dfrac{5{,}34 - 5}{7{,}24 \big/ \sqrt{29}} = 0{,}2537 < 1{,}311 = t_{0{,}10\,;\,29}$

Se acepta la hipótesis H_0

d) Potencia $= (1-\beta) = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ falsa}\right]$

La hipótesis H_1 es compuesta (existen infinitos valores tal que $\mu > 5$), se construye una función, es decir:

$H_1 : \mu_1 > 5$	β	$1-\beta$
Distintos valores		

$$\beta = P\left[ET\,II\right] = P\left[\text{Rechazar } H_1 \,\middle|\, H_1 \text{ cierta}\right] = P\left[\text{Aceptar } H_0 \,\middle|\, H_0 \text{ falsa}\right] = P\left[\overline{x} \overbrace{<}^{\substack{\text{Región} \\ \text{Decisión}}} k \,\middle|\, H_1 \right]$$

En el muestreo de una población normal con varianza desconocida, y desviación típica muestral σ_x, la variable: $(\overline{x} - \mu) \Big/ \dfrac{\sigma_x}{\sqrt{n-1}} = t_{n-1}$, siendo la muestra $n = 300$, la distribución se aproxima a una $N(0,1)$

$$\beta = P(\overline{x} < k \mid H_1) = P\left[\mu_1 + t_{n-1} \frac{\sigma_x}{\sqrt{n-1}} < k \,\middle|\, H_1 \right] = P\left[t_{n-1} < \frac{(k - \mu_1).\,\sqrt{n-1}}{\sigma_x} \right] \overbrace{\underbrace{=}_{n>30}}^{k=6{,}76}$$

$$= P\left[z < \frac{(6,76 - \mu_1) . \sqrt{n-1}}{\sigma_x} \right] \text{, en consecuencia:}$$

$H_1 : \mu_1 = 5,5$

$$\beta = P\left[z < \frac{(6,76 - \mu_1) . \sqrt{n-1}}{\sigma_x} \right] = P\left[z < \frac{(6,76 - 5,5) . \sqrt{299}}{7,24} \right] = P\left[z < 3\right] = 0,9986$$

$H_1 : \mu_1 = 6$

$$\beta = P\left[z < \frac{(6,76 - 6) . \sqrt{299}}{7,24} \right] = P\left[z < 1,82\right] = 0,9656$$

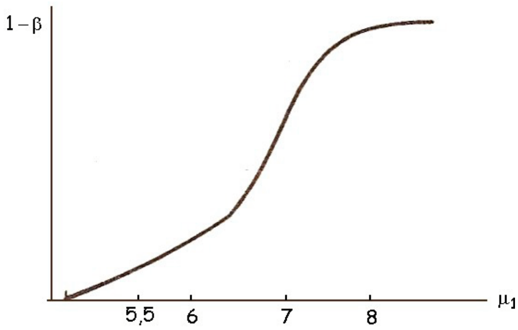
$H_1 : \mu_1 = 7$

$$\beta = P\left[z < \frac{(6,76 - 7) . \sqrt{299}}{7,24} \right] = P\left[z < -0,57\right] = 0,2843$$

$H_1 : \mu_1 = 8$

$$\beta = P\left[z < \frac{(6,76 - 8) . \sqrt{299}}{7,24} \right] = P\left[z < -2,96\right] = 0,00154$$

$H_1 : \mu_1 > 5$	$\mathbb{Z}$	$1-\beta$
5,5	0,9986	0,0014
6	0,9656	0,0034
7	0,2843	0,7157
8	0,0015	0,9985



A medida que se alejan las hipótesis H_0 y H_1 aumenta la potencia del contraste. Es decir, cuanto más alejadas se encuentren las hipótesis para contrastar, mayor será la probabilidad de que se rechace H_0 cuando sea falsa, algo deseable (constante α y k).

CONTRASTE UNILATERAL DE LA MEDIA CON VARIANZA CONOCIDA.

13.- El número de averías de un determinado tipo de avión se considera una variable aleatoria con distribución de Poisson de media 2 averías al mes. El equipo de mantenimiento intenta reducir esta media incorporando algunas mejoras. Para comprobar si con estas medidas se reduce el número medio de averías, se decide observar el número medio de averías en los 25 meses siguientes a la introducción de las mejoras. Si el número medio de averías en esos 25 meses fue de 1,5.
¿Qué decisión debe adoptar el servicio técnico a un nivel de significación del 1%?. ¿Y si el servicio técnico relaja su nivel de exigencia al 85% de confianza?. ¿Cambiaría su decisión?.

Solución:

Sea la variable aleatoria X = 'Número de averías al mes' $X \sim P(\lambda = 2) \overset{n\to\infty}{\overbrace{\to}} N(2, \sqrt{2})$

En la muestra de tamaño $n = 25$ meses $\hat{\lambda} = \bar{x} = 1,5$ averías.

La distribución en el muestreo, bajo la hipótesis nula H_0, $\bar{x} \sim N\left[\mu, \frac{\sigma}{\sqrt{n}}\right] \sim N\left[2, \frac{2}{\sqrt{25}}\right]$

Es un contraste unilateral, donde se plantean las hipótesis: $\begin{cases} H_0 : \mu = 2 \\ H_1 : \mu < 2 \text{ (compuesta)} \end{cases}$

La regla de decisión $\begin{cases} \text{Si } \bar{x} > k \text{ se acepta } H_0 \text{ (R A)} \\ \text{Si } \bar{x} \le k \text{ se rechaza } H_0 \text{ (R C)} \end{cases}$

Se determina el valor de k considerando el nivel de significación α:

$$\alpha = P\left[\text{Rechazar } H_0 \middle| H_0 \text{ cierta}\right] = P\left[\bar{x} \le k \ / H_0 : \mu = 2\right] = P\left[z \le \frac{k-2}{\sqrt{2/25}}\right] = 0,01 \mapsto$$

$$\mapsto P\left[z \ge \frac{2-k}{\sqrt{2/25}}\right] = 0,01 \Rightarrow \frac{2-k}{\sqrt{2/25}} = 2,32 = z_{0,01} \mapsto k = 1,34$$

Se advierte que $(1,5 > 1,34)$, cae en la región de aceptación, con lo cual se acepta la hipótesis nula, esto es, con un nivel de significación del 1% se afirma que las averías mensuales se mantienen siendo 2, y en consecuencia, las mejorías no son operativas.

Si $\alpha = 0,15$ se replantean los cálculos: $\dfrac{2-k}{\sqrt{2/25}} = 1,04 = z_{0,15} \mapsto k = 1,71$

Como $1,71 < 1,34$, cae en la región de rechazo, con lo que no se acepta la hipótesis nula, y se concluye que las mejoras son operativas.

CONTRASTE UNILATERAL DE LA MEDIA CON VARIANZA POBLACIONAL DESCONOCIDA.

14.- La concentración media de dióxido de carbono en el aire en una determinada zona no es habitualmente mayor que 355 p.p.m.v (partes por millón en volumen). Se sospecha que esta concentración es mayor en la capa de aire más próxima a la superficie.
Para contrastar esta hipótesis se analiza aire en 20 puntos elegidos aleatoriamente a una misma altura del suelo. Los datos recogidos tienen una media muestral de 580 p.p.m.v y una cuasidesviación típica muestral de 180.
Suponiendo que las mediciones siguen una distribución normal, ¿podemos afirmar a un nivel de 0,01, que los datos proporcionan suficiente evidencia estadística a favor de la hipótesis de que la concentración es mayor cerca del suelo?.
Indicar razonadamente si el p-valor es mayor o menor que 0,01

Solución:

Se tiene una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de tamaño $n = 20$, donde la variable aleatoria X = "concentración de dióxido de carbono en puntos cercanos al suelo" sigue una distribución normal $X \sim N(\mu, \sigma)$ con varianza poblacional desconocida.

Se desea comprobar si hay suficiente evidencia estadística a favor de que $\mu > 335$. Para ello, con un nivel de significación $\alpha = 0{,}01$, se plantea el contraste:

Hipótesis nula $H_0 : \mu \leq 335$ frente a la hipótesis alternativa $H_1 : \mu > 335$

La cuestión queda ante un contraste unilateral (una cola) para la media poblacional con varianza poblacional desconocida.

Se rechaza la hipótesis nula en la región: $R = \left\{ \bar{x} - \mu_0 > t_{\alpha;\,(n-1)} \dfrac{s_x}{\sqrt{n}} \right\}$

$$\bar{x} - \mu_0 = 580 - 355 = 225 \qquad\qquad t_{\alpha;\,(n-1)} = t_{0,01;\,19} \frac{s_x}{\sqrt{n}} = 2{,}539 \frac{\sqrt{180}}{\sqrt{20}} = 102{,}19$$

Como $R = \{ 225 > 102{,}19 \}$ se verifica la condición de rechazo, por tanto, se rechaza la hipótesis nula H_0.

En consecuencia, existe suficiencia estadística (con un nivel de significación 0,01) para concluir que la concentración media de dióxido de carbono es superior a 355 cerca del suelo.

Por otra parte, el p-valor se interpreta como el apoyo que los datos proporcionan a la hipótesis nula H_0. En otras palabras, cuando el $p\text{-valor} < \alpha \;\mapsto\;$ se rechaza H_0

Habiendo rechazado H_0 con $\alpha = 0{,}01 \;\mapsto\; p-\text{valor} < 0{,}01$

CONTRASTE BILATERAL DE LA PROPORCIÓN.

15.- Un dentista afirma que el 40% de los niños de 10 años presentan indicios de caries dental. Tomada una muestra de 100 niños, se observó que 36 presentaban indicios de caries.
Contrastar la hipótesis del dentista para un nivel de confianza del 90%.

Solución:

Sea el parámetro p = 'proporción de niños que presentan indicios de caries dental'.

Como siempre que se desea hacer un contraste sobre una proporción, se parte de una muestra aleatoria $(X_1, X_2, \cdots, X_n)$, de tamaño $n = 100$, donde $X \sim B(1, p)$

La distribución en el muestreo del estadístico $\bar{x} = \hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n} \overset{T.C.L.}{\sim} N\left[p, \sqrt{\dfrac{p.q}{n}}\right]$

Con un nivel de significación $\alpha = 0{,}10$ se plantea el contraste:

Hipótesis nula $H_0: p = 0{,}40$ frente a la Hipótesis alternativa $H_1: p \neq 0{,}40$

La hipótesis nula se rechaza en la región: $R = \left\{ \left|\bar{x} - p_0\right| > z_{\alpha/2}\sqrt{\dfrac{p_0(1-p_0)}{n}} \right\}$

donde, $\left|\bar{x} - p_0\right| = \left|\hat{p} - 0{,}40\right| = \left|\dfrac{36}{100} - 0{,}40\right| = 0{,}04$

$$z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} = z_{0,05}\sqrt{\frac{0{,}40(1-0{,}40)}{100}} = (1{,}64)\,.\,(0{,}0489898) = 0{,}08$$

Siendo $R = \{0{,}04 > 0{,}08\}$ no se verifica la condición de rechazo y se acepta la hipótesis nula H_0. En consecuencia, con un nivel de significación $\alpha = 0{,}10$, se puede afirmar que el 40% de los niños presenta indicios de caries dental.

16.- En los días previos a unas elecciones municipales, el candidato de un partido político está convencido de obtener el 60% de los votos electorales. No obstante, su partido encarga una encuesta entre 100 votantes potenciales, resultando que el 52% de ellos dijeron tener intención de votar a dicho candidato.
Con un nivel de significación del 5%, se pide contrastar:

a) $H_0: p = 0,60$ frente a $H_1: p = 0,50$

b) $H_0: p = 0,60$ frente a $H_1: p \neq 0,60$

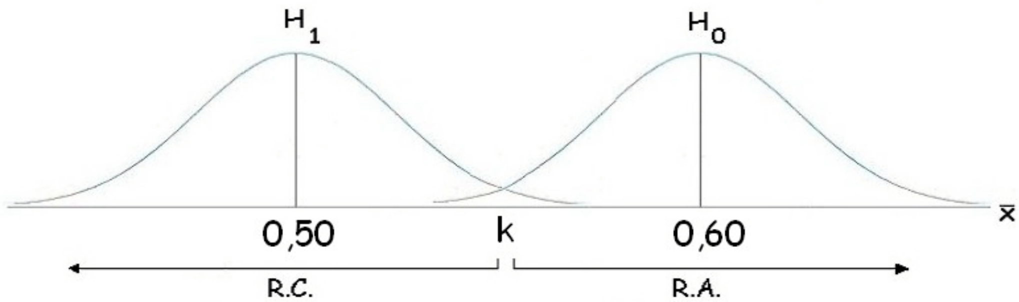
c) Potencia del contraste efectuado en el apartado (a).

Solución:

a) Sea la variable X ='% de votos al candidato'

Se trata de un contraste de hipótesis nula simple frente a una hipótesis alternativa simple: $H_0: p = 0,60$ frente a $H_1: p = 0,50$

Regla de decisión: $\begin{cases} \hat{p} > k & \text{Se acepta } H_0 \ \text{(R.A.)} \\ \hat{p} \leq k & \text{Se rechaza } H_0 \ \text{(R.C.)} \end{cases}$



La distribución en el muestreo de $\hat{p} = \dfrac{1}{100}\sum_{i=1}^{100} x_i$ siendo $x_i = \begin{cases} 1 & \text{vota Si} \\ 0 & \text{vota No} \end{cases}$

donde se conoce $E(\hat{p}) = p$ y $V(\hat{p}) = \dfrac{p(1-p)}{n}$.

Al ser el tamaño suficientemente grande $n = 100$ y estar definido $\hat{p}$ como suma de variables aleatorias independientes, según una distribución de Bernouilli $B(1, p)$ se puede aproximar la distribución exacta de $\hat{p} \sim N\left[\ p,\ \sqrt{p.(1-p)/n}\ \right]$

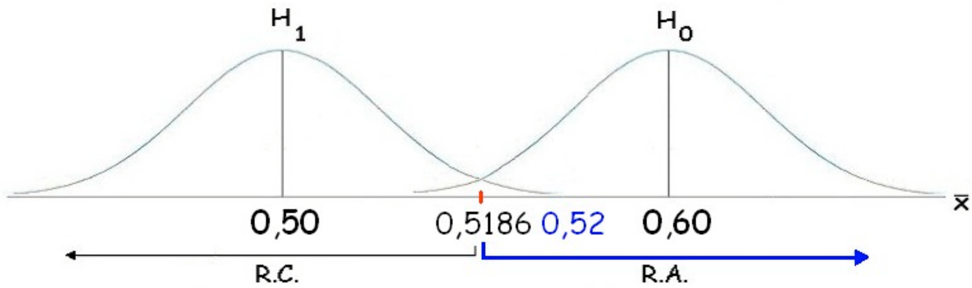
Bajo la hipótesis nula: $\hat{p} \sim N\left[0,60\,,\sqrt{\dfrac{0,60.(1-0,60)}{100}}\right] \equiv N(0,60\,,\,0,049)$

El valor crítico K se determina, bajo la hipótesis nula, por el nivel de significación:

$$\alpha = P\left[ET\ I\right] = P\left[\text{Rechazar } H_0 \,/\, H_0 \text{ cierta}\right] = 0,05$$

$$\alpha = P\left[ET\ \ I\right] = P\left[\hat{p} \leq k \,\middle|\, H_0: p = 0,6\right] = P\left[\frac{\hat{p}-0,60}{0,049} \leq \frac{k-0,60}{0,049}\right] = P\left[z \leq \frac{k-0,60}{0,049}\right] = 0,05$$

$$\frac{0,60-k}{0,049}=1,645 \;\mapsto\; k=0,5186$$



Como la proporción muestral $\hat{p}=\bar{x}=52/100=0,52$, $\hat{p}=0,52>0,5186$, se encuentra dentro de la región de aceptación de la hipótesis nula.

b) En este caso, se trata de un contraste bilateral con hipótesis nula simple frente a una hipótesis alternativa compuesta: $H_0: p=0,60$ frente a $H_1: p\neq 0,60$

Regla de decisión: $\begin{cases} |\hat{p}|>k & \text{Se acepta } H_0 \quad \text{(R.A.)} \\ |\hat{p}|\leq k & \text{Se rechaza } H_0 \quad \text{(R.C.)} \end{cases}$

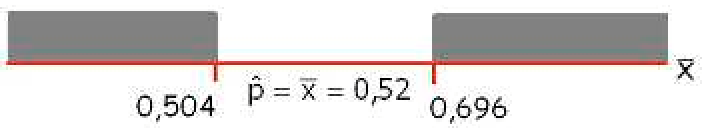
Bajo la hipótesis nula: $\hat{p}\sim N(0,60\,,\,0,049)$

Para determinar el crítico K, bajo la hipótesis nula, con el nivel de significación:

$$\alpha = P\left[ET\ I\right]=P\left[\text{Rechazar } H_0\ /\ H_0 \text{ cierta}\right]=0,05$$

$$\alpha = P\left[ET\ \ I\right]=P\left[\,|\hat{p}|\leq K\ \middle|\ H_0: p=0,6\right]=P\left[\left|\frac{\hat{p}-0,60}{0,049}\right|\leq\frac{K-0,60}{0,049}\right]=0,05$$

La región crítica: $\left|\dfrac{\hat{p}-0,60}{0,049}\right|>1,96=z_{0,025} \;\mapsto\; \begin{array}{l}\hat{p}-0,60>(0,049)(1,96)=0,696\\ \hat{p}-0,60<-(0,049)(1,96)=0,504\end{array}$



Como $\hat{p}=\bar{x}=52/100=0,52$, se encuentra en la región de aceptación, concluyendo que se acepta la hipótesis nula.

c) Potencia del Contraste: $\text{Potencia}=P\left[\text{Rechazar } H_0\ \middle|\ H_0 \text{ falsa}\right]=1-\beta$

Error Tipo II: $\beta=P\left[ET\ II\right]=P\left[\text{Aceptar } H_0\ /\ H_0 \text{ falsa}\right]$

Con la hipótesis alternativa $H_1: p=0,50$: $\hat{p}\sim N\left[0,5\ ,\ \sqrt{\dfrac{0,5\,.\,0,5}{100}}\right]=N(0,5\ ,\ 0,05)$

$$\beta = P\left[ET\ \ II\right] = P\left[\ \hat{p} > 0,5186 / \ N(0,5,\,0,05)\right] = P\left[\frac{\hat{p}-0,5}{0,05} > \frac{0,5186-0,5}{0,05}\right] =$$

$$= P\,(z > 0,372) = 0,35496$$

Abscisas	Areas
0,37 - 0,38	0,3557 - 0,3520
0,372 - 0,38	x - 0,3520

$$x = 0,3520 + \frac{(0,008)\,(0,0037)}{0,01} = 0,35496$$

$$\text{Potencia} = 1-\beta = 1-0,35496 = 0,6450$$

También se podía haber realizado por la definición de potencia de un contraste:

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \ \middle|\, H_0 \ \text{falsa}\right] = P\left[\ \hat{p} \leq 0,5186 / \ N(0,5\ ,\,0,05)\right] =$$

$$= P\left[\frac{\hat{p}-0,5}{0,05} \leq \frac{0,5186-0,5}{0,05}\right] = P\,(z \leq 0,372) = 1-P\,(z \geq 0,372) = 0,6450$$

17.- El dueño de los cines CINEFILÓN considera que, dado el aforo de la sala, una afluencia diaria a la misma del 85% sería óptima, en el sentido de que los clientes se sientan cómodos y para que a la vez no haya pérdidas económicas. Durante un período de tiempo, se analiza la afluencia a los cines, observándose que, en media, se ocupan 171 de las 200 butacas.

¿Con qué confianza podrá afirmar el dueño de CINEFILÓN que la asistencia a sus cines es óptima?

¿Qué pasaría si el dueño quisiera estar más seguro de su decisión, y ampliar el nivel de confianza al 99%?

Nota: *La regla de la decisión adoptada es que si hay una desviación inferior al 5% de la cantidad óptima, se aceptaría la hipótesis de que la afluencia es, efectivamente, óptima.*

Solución:

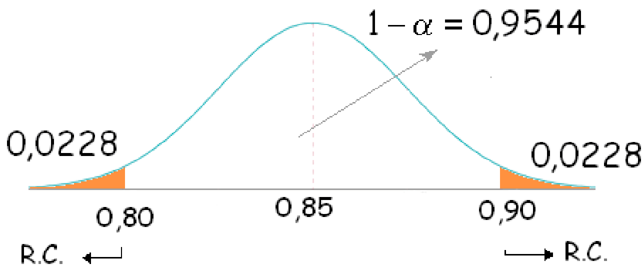
La variable aleatoria X = '% asistencia diaria a los cines CINEFILÓN', $X \sim B(1,\ p)$.

- En el contraste bilateral se establecen las hipótesis: $H_0 : p_0 = 0,85 \quad H_1 : p_1 \neq 0,85$

- En el muestreo, la proporción que acude a los cines (200 butacas) en un día ($\hat{p} = \frac{X}{n}$), bajo la hipótesis nula, por el TCL (Teorema Central Límite), sigue una distribución

$$N\left[p\,,\sqrt{\frac{p.q}{n}}\right] \equiv N\left[0,85\,,\sqrt{\frac{0,85.0,15}{200}}\right] = N(0,85\ ,\ 0,025)$$

- Se determina $(1-\alpha)$ a partir de los valores críticos, considerando que una desviación inferior al 5% de la cantidad óptima significa que $0,80 < \hat{p} < 0,90$



$$\alpha = P\left[\text{Rechazar } H_0 / H_0 \text{ cierta}\right] = P\left[\left(\frac{0{,}80-0{,}85}{0{,}025} > z\right) \cup \left(z > \frac{0{,}90-0{,}85}{0{,}025}\right)\right] =$$

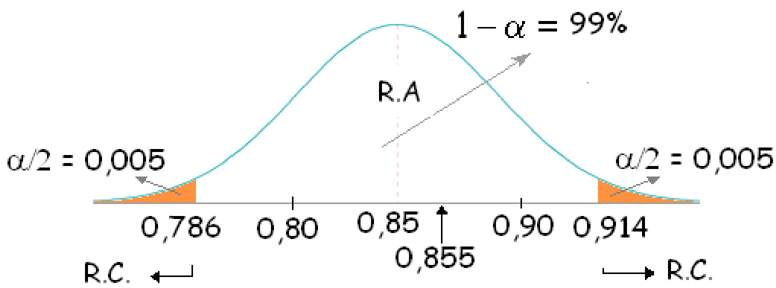
$$= P\left[-2 > z\right] + P\left[z > 2\right] = 2.0{,}028 = 0{,}0456$$

El nivel de confianza: $1-\alpha = 1-0{,}0456 = 0{,}9544$

Si el nivel de confianza $\left[1-\alpha\right] = 99\% \;\Rightarrow\; \alpha = 1\%$ (mayor exigencia).

El valor crítico k se determina a partir del nivel de significación $\alpha = 0{,}01$

$$\alpha = P\left[\text{Rechazar } H_0 / H_0 \text{ cierta}\right] = P\left[\left(\frac{k_1-0{,}85}{0{,}025} > z\right) \cup \left(z < \frac{k_2-0{,}85}{0{,}025}\right)\right] = 0{,}01$$



$$\left|\frac{k-0{,}85}{0{,}025}\right| = 2{,}57 = z_{0{,}005} \;\Rightarrow\; \begin{array}{l} k_1 = 0{,}914 \\ k_2 = 0{,}786 \end{array}$$

Comparando los valores críticos k con el valor del estadístico muestral (evidencia empírica) $\hat{p} = \frac{171}{200} = 0{,}855$, se observa que cae dentro de la región de aceptación (R.A.), concluyendo que se acepta la hipótesis nula siendo óptima la afluencia a los cines CINEFILÓN.

CONTRASTE BILATERAL DE LA VARIANZA CON MEDIA POBLACIONAL CONOCIDA.

18.- En una población $N(10\,,\,\sigma)$ se desea contrastar la hipótesis nula $H_0:\sigma^2=3$ frente a la hipótesis alternativa $H_1:\sigma^2\neq 3$ con un nivel de significación del 5%. En una muestra aleatoria simple de tamaño 4 se obtuvieron los resultados:

10	8	12	14

Solución:

Hipótesis compuesta sobre σ^2 : $\begin{cases} H_0:\sigma^2=3 \\ H_1:\sigma^2\neq 3\end{cases}$

Como la hipótesis alternativa es $\sigma^2\neq 3$ en la decisión deberán ser válidos los valores de σ^2 tanto mayores o menores que 3, por lo que el contraste debe ser bilateral o de dos colas.

Regla de decisión: $\begin{cases} \text{Si } \left|\displaystyle\sum_{i=1}^{4}(x_i-\mu)^2\right| \geq k \;\mapsto\; \text{R.C : Rechazar } H_0 \\ \text{Si } \left|\displaystyle\sum_{i=1}^{4}(x_i-\mu)^2\right| < k \;\mapsto\; \text{R.A : Aceptar } H_0 \end{cases}$

En la distribución en el muestreo del estadístico $\sum_{i=1}^{4}(x_i-\mu)^2$, las variables aleatorias independientes x_i se distribuyen $N(10\,,\,\sigma)$.

Bajo la hipótesis nula $N(10\,,\,\sqrt{3})$: $\displaystyle\sum_{i=1}^{4}\left(\frac{x_i-\mu}{\sigma_0}\right)^2=\sum_{i=1}^{4}\left(\frac{x_i-10}{\sqrt{3}}\right)^2=\chi_4^2$

(Adviértase que es la suma de 4 variables aleatorias $N(0\,,\,1)$, independientes entre sí)

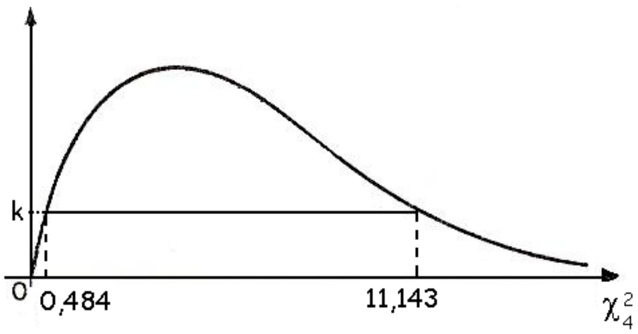
Fijado el nivel de significación α (Error Tipo I), se halla el valor de la constante k:

$$\alpha = P\big[\text{Rechazar } H_0\,/\,H_0 \text{ cierta}\big] = P\left[\,\left|\sum_{i=1}^{4}\left(\frac{x_i-10}{\sqrt{3}}\right)^2\right| \geq k\,/\,H_0 \text{ cierta}\right] = P\big[\,\left|\chi_4^2\right|\geq k\,\big] =$$

$$= P\big[(\chi_4^2\leq k_1)\cup(\chi_4^2\geq k_2)\big] = P\big[\chi_4^2\leq k_1\big] + P\big[\chi_4^2\geq k_2\big] = \alpha_1+\alpha_2 = 0{,}05$$

$$P\big[\chi_4^2\leq k_1\big]=1-P\big[\chi_4^2\geq k_1\big]=1-\alpha_1=0{,}975 \;\Rightarrow\; k_1=\chi^2_{0,975\,;\,4}=0{,}484$$

$$P\big[\chi_4^2\geq k_2\big]=\alpha_2=0{,}025 \Rightarrow\; k_2=\chi^2_{0,025\,;\,4}=11{,}143$$



Región de rechazo: $(\chi^2_4 < 0{,}484) \cup (\chi^2_4 > 11{,}143)$

La región de aceptación es el intervalo $\left[0{,}484\ ;\ 11{,}143\right]$

Bajo la hipótesis nula, el estadístico muestral $\sigma_0^2 = \dfrac{\sum_{i=1}^{4}(x_i - 10)^2}{3} = 8$, se encuentra en la región de aceptación, $0{,}484 \leq 8 \leq 11{,}143$, concluyendo que la varianza de la distribución es 3.

Análogamente, se podría haber resuelto considerando la región de rechazo:

$$R = \left[\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma_0^2} \notin \left[\chi^2_{1-\alpha/2\,;\,n}\ ;\ \chi^2_{\alpha/2\,;\,n}\right] \right]$$

El valor del estadístico muestral $\sum_{i=1}^{4}(x_i - \mu)^2 = 24$, bajo la hipótesis nula resulta:

$$\chi^2_4 = \frac{\sum_{i=1}^{4}(x_i - 10)^2}{3} = \frac{24}{3} = 8\,.$$

Los valores tabulares son $\begin{array}{l} \chi^2_{1-\alpha/2\,;\,n} = \chi^2_{0{,}975\,;\,4} = 0{,}484 \\ \chi^2_{\alpha/2\,;\,n} = \chi^2_{0{,}025\,;\,4} = 11{,}143 \end{array}$

La región de aceptación de la hipótesis nula es el intervalo $\left[0{,}484\ ;\ 11{,}143\right]$.

Al pertenecer el valor muestral del estadístico $\dfrac{\sum_{i=1}^{4}(x_i - \mu)^2}{\sigma_0^2} = 8$ a la región de aceptación, se concluye que la varianza de la distribución es 3.

CONTRASTE BILATERAL DE LA MEDIA CON VARIANZA POBLACIONAL CONOCIDA.

19.- En una población $N(\mu, 5)$ se quiere contrastar la hipótesis nula $H_0 : \mu = 18$ frente a la hipótesis alternativa $H_1 : \mu \neq 18$, con un nivel de significación $\alpha = 0{,}01$, con una muestra de tamaño 10:

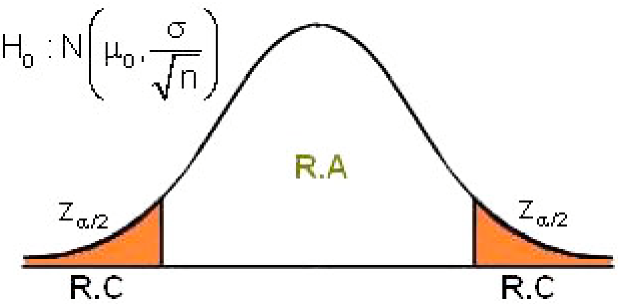
16	12	15	16	20	25	14	18	17	22

Solución:

En el contraste se establecen las hipótesis: $\begin{cases} H_0 : \mu = 18 \\ H_1 : \mu \neq 18 \end{cases}$

Como la hipótesis alternativa es $\mu \neq 18$ en la decisión deberán ser válidos valores de μ tanto mayores o menores que 18, por lo cual el contraste debe ser bilateral o de dos colas.

Regla decisión $\begin{cases} \text{Si } \ |\bar{x}| > k \ \text{ se rechaza } H_0 \text{ (R.C.)} \\ \text{Si } \ |\bar{x}| \leq k \ \text{ se acepta } H_0 \text{ (R.A.)} \end{cases}$



La muestra de tamaño 10, con media $\bar{x} = \dfrac{\sum_{i=1}^{10} x_i}{10} = 17{,}5$, siendo la varianza poblacional conocida $\sigma^2 = 25$, bajo la hipótesis nula sigue una distribución $N\left[18, \dfrac{5}{\sqrt{10}}\right]$, con lo que la variable $\dfrac{\bar{x}-18}{\frac{5}{\sqrt{10}}} \sim N(0,1)$

El valor de k se calcula mediante el nivel de significación $\alpha = 0{,}01$:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[|\bar{x}| > k\right] = P\left[\left|\frac{\bar{x}-18}{\frac{5}{\sqrt{10}}}\right| > K\right] =$$

$$= P\left[\left(\frac{\bar{x}-18}{\frac{5}{\sqrt{10}}} < -K\right) \cup \left(\frac{\bar{x}-18}{\frac{5}{\sqrt{10}}} > K\right)\right] = P\left[\left(\frac{\bar{x}-18}{\frac{5}{\sqrt{10}}} < -K\right)\right] + P\left[\left(\frac{\bar{x}-18}{\frac{5}{\sqrt{10}}} > K\right)\right] = 0{,}01$$

La región crítica será:

$$\left|\frac{\overline{x}-18}{\frac{5}{\sqrt{10}}}\right| > 2,575 = z_{\alpha/2} = z_{0,005} \Rightarrow \begin{cases} \overline{x} > 18 + \dfrac{5}{\sqrt{10}}\,(2,575) = 22,07 \\ \overline{x} < 18 - \dfrac{5}{\sqrt{10}}\,(2,575) = 13,93 \end{cases}$$

En consecuencia, la región de aceptación: $13,93 < \overline{x} < 22,07$

Como la media muestral $\overline{x} = \dfrac{\sum_{i=1}^{10} x_i}{10} = 17,5$ está contenida en la región de aceptación, no se rechaza la hipótesis nula $\mu = 18$, con un nivel de significación de 0,01.

Análogamente, se podría haber resuelto con la región de rechazo:

$$R_{rechazo} = \left\{ \left|\overline{x} - \mu_0\right| > z_{\alpha/2} \,.\, \frac{\sigma}{\sqrt{n}} \right\}$$

$$R_{rechazo} = \left\{ \left|17,5 - 18\right| > z_{0,05} \,.\, \frac{5}{\sqrt{10}} \right\} = \left\{ \left|17,5 - 18\right| > (2,575).\, \frac{5}{\sqrt{10}} \right\} = \left\{ 0,5 > 4,071 \right\}$$

No se verifica la región de rechazo, en consecuencia se admite la hipótesis nula $\mu = 18$, con un nivel de significación de 0,01.

CONTRASTE BILATERAL DE LA MEDIA CON VARIANZA POBLACIONAL DESCONOCIDA.

20.- En una población $N(\mu, \sigma)$ se quiere contrastar la hipótesis nula $H_0 : \mu = 2$ frente a la hipótesis alternativa $H_1 : \mu \neq 2$, con un nivel de significación $\alpha = 0{,}01$, con una muestra aleatoria simple de tamaño 15:

2,1	2,25	3,01	2,92	2,98	3,08	3,8	3,95
2,75	2,74	3,16	2,56	3,15	2,65	3,12	

Solución:

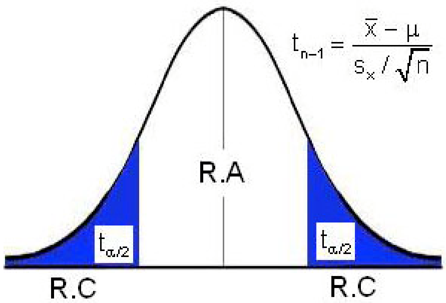
En la muestra se obtiene:

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = 2{,}948 \qquad \sigma_x^2 = \frac{\sum_{i=1}^{15} (x_i - \bar{x})^2}{15} = 0{,}2264 \qquad s_x^2 = \frac{\sum_{i=1}^{15} (x_i - \bar{x})^2}{14} = 0{,}2426$$

$$\sigma_x = 0{,}476 \qquad s_x = 0{,}493 \qquad n = 15$$

Como la hipótesis alternativa es $\mu \neq 2$ en la decisión deberán ser válidos valores de μ bien sean mayores o menores que 2, por lo cual el contraste debe ser bilateral o de dos colas.

Regla decisión $\begin{cases} \text{Si } \ |\bar{x}| > k \ \text{ se rechaza } H_0 \text{ (R.C.)} \\ \text{Si } \ |\bar{x}| \leq k \ \text{ se acepta } H_0 \text{ (R.A.)} \end{cases}$



En el muestreo de una población normal con varianza desconocida, y desviación típica muestral σ_x, la variable $\dfrac{\bar{x} - \mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} = \dfrac{\bar{x} - \mu}{\dfrac{s_x}{\sqrt{n}}} = t_{n-1}$

Con el nivel de significación $\alpha = 0{,}01$ se determina la región crítica:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[|\bar{x}| > k \ / \ H_0\right] = P\left[\left|\frac{\bar{x} - \mu}{\sigma_x / \sqrt{n-1}}\right| > K \ / \ \mu = 2\right] =$$

$$= P\left[\left|\underbrace{\frac{\bar{x} - 2}{0{,}476 / \sqrt{14}}}_{t_{14}}\right| > K\right] = P\left[\left(\frac{\bar{x} - 2}{0{,}476 / \sqrt{14}} < -K\right) \cup \left(\frac{\bar{x} - 2}{0{,}476 / \sqrt{14}} > K\right)\right] =$$

$$= P\left[\underbrace{\frac{\overline{x}-2}{0,476\big/\sqrt{14}}}_{t_{14}} < -K\right] + P\left[\underbrace{\frac{\overline{x}-2}{0,476\big/\sqrt{14}}}_{t_{14}} > K\right] = \frac{\alpha}{2} + \frac{\alpha}{2} = 0,01$$

La región crítica: $\left|\dfrac{\overline{x}-2}{0,476\big/\sqrt{14}}\right| > 2,977 = t_{\alpha/2\,;\,(n-1)} = t_{0,005\,;\,14}$

$$\left|\frac{\overline{x}-2}{0,476\big/\sqrt{14}}\right| > 2,977 \;\Rightarrow\; \begin{cases} \overline{x} \;>\; 2 + 2,977\,.(0,476\big/\sqrt{14}) = 2,378 \\ \overline{x} \;<\; 2 - 2,977\,.(0,476\big/\sqrt{14}) = 1,622 \end{cases}$$

La región de aceptación será: $1,622 < \overline{x} < 2,378$

Se observa que el valor del estadístico muestral $\overline{x} = 2,948$ no se encuentra en la región de aceptación, por tanto, se rechaza la hipótesis nula y se acepta que la media de la distribución poblacional es distinta de 2.

Análogamente, se podría haber resuelto considerando la región de rechazo:

$$R = \left\{\left|\,\overline{x}-\mu_0\,\right| \;>\; t_{\alpha/2\,;\,(n-1)}\,\frac{s_x}{\sqrt{n}}\right\} = \left\{\;\left|\,2,948-2\,\right| > \overbrace{2,977}^{t_{0,005\,;\,14}}\,\frac{0,493}{\sqrt{15}}\right\} = \left\{0,948 > 0,379\right\}$$

Al verificarse la región de rechazo, no se acepta la hipótesis nula, concluyendo que se acepta la hipótesis alternativa de que la media de la distribución poblacional es distinta de 2.

Una forma análoga de enfocar la cuestión consiste en aceptar la hipótesis nula $H_0 : \mu = 2$ cuando el estadístico experimental fuera menor o igual que el estadístico teórico, esto es:

$$\overbrace{t_{n-1} = \frac{\left|\,\overline{x}-\mu_0\,\right|}{s_x\big/\sqrt{n}}}^{\text{estadístico experimental}} \quad\leq\quad \overbrace{t_{\alpha/2\,;\,(n-1)}}^{\text{estadístico teórico}}$$

El estadístico de contraste: $t_{n-1} = \dfrac{\left|\,\overline{x}-\mu_0\,\right|}{s_x\big/\sqrt{n}} = \dfrac{\left|\,2,948 - 2\,\right|}{0,493\,/\sqrt{15}} = 7,447$

El valor tabular del estadístico teórico: $t_{\alpha/2\,;\,(n-1)} = t_{0,005\,;\,14} = 2,977$

Como $t_{n-1} = 7,447 > 2,977 = t_{0,005\,;\,14}$ se rechaza la hipótesis nula, concluyendo que la media de la distribución poblacional es distinta de 2.

21.- Una empresa de neumáticos afirma que una nueva gama en promedio duran más de 28.000 km. Las pruebas con 64 neumáticos dan como resultado una duración media de 27.800 km, con una desviación típica de 1.000 km.

a) Si se usa un nivel de significación del 5%, comprobar si hay evidencia suficiente para rechazar la afirmación de la empresa.

b) ¿Cuál es el p-valor?

Solución:

a) Sobre la población de los neumáticos se define la variable aleatoria X = "duración en kilómetros", donde los valores poblacionales $N(28000, \sigma)$

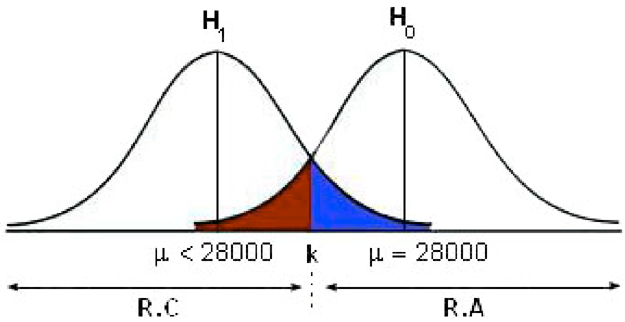
Las hipótesis sobre la media poblacional μ con σ^2 desconocida:

Hipótesis sobre $\mu: \begin{cases} H_0 : \mu_0 = 28000 \\ H_1 : \mu_1 < 28000 \end{cases}$

Se trata de un contraste unilateral por la izquierda.

Regla de decisión:

$\begin{cases} \text{Si } \bar{x} > k & \text{Aceptar } H_0 \text{ (R.A)} \\ \text{Si } \bar{x} \le k & \text{Rechazar } H_0 \text{ (R.C)} \end{cases}$



Bajo la hipótesis nula, la muestra sigue una distribución $N\left(28000, \frac{1000}{\sqrt{64}}\right) \equiv N(28000, 125)$

El valor crítico k, bajo la hipótesis nula, se determina con el nivel de significación $\alpha = 0,05$:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\bar{x} \le k \,\middle|\, H_0 \text{ cierta}\right] = P\left[\bar{x} \le k \,\middle|\, \mu_0 = 28000\right] =$$

$$= P\left(\frac{\bar{x} - 28000}{125} \le \frac{k - 28000}{125}\right) = P\left[z \le \frac{k - 28000}{125}\right] = 0,05$$

$$P\left[z \le \frac{k - 28000}{125}\right] = P\left[z \ge -\frac{k - 28000}{125}\right] = 0,05$$

$$-\frac{k - 28000}{125} = 1,645 \quad \mapsto \quad k = 28000 - 125 \text{ x } 1,645 = 27794,375$$

Como $\bar{x} = 27800 > 27794,375$ se acepta la hipótesis nula, por tanto, no hay evidencias suficientes para rechazar la afirmación de la empresa.

b) El p–valor (α_p) es el menor nivel de significación para el que se rechaza la hipótesis nula.

El valor del estadístico que se obtiene con la muestra: $\dfrac{\overline{x}-28000}{125}=\dfrac{27800-28000}{125}=-1,6$

Como el contraste es unilateral por la izquierda, el p–valor viene dado: $P(z<-1,6)=\alpha_p$

$\alpha_p = P(z\leq -1,6)=P(z\geq 1,6)=0,0548$

Para un nivel de significación $\alpha=0,05$, siendo $\alpha_p = 0,0548>\alpha=0,05$, se acepta la hipótesis nula.

22.- Una línea de producción sigue una distribución normal, funciona con un peso de llenado de 16 gramos por envase. El exceso o defecto de peso en el llenado presenta graves problemas, debiendo parar la línea de producción. Un inspector de calidad toma una muestra de 25 artículos, y de acuerdo con los resultados obtenidos, ¿qué decisión debe tomar?.

a) Para un nivel de significación de 0,05, en la muestra se obtiene $\overline{x}=16,32$ gramos y una desviación $s_x = 0,8$ gramos.

b) ¿Cuál es el p-valor?.

Solución:

a) Sea la variable aleatoria X = “peso de llenado”, $X\sim N(16,\sigma)$

Las hipótesis sobre la media poblacional μ con σ^2 desconocida:

Hipótesis sobre μ : $\begin{cases} H_0 : \mu_0 = 16 \\ H_1 : \mu_1 \neq 16 \end{cases}$

Adviértase que como la hipótesis alternativa $H_1 : \mu \neq 16$, en la decisión deberán ser válidos valores de μ tanto mayores o menores que 16, por lo que el contraste debe ser bilateral o de dos colas.

En esta línea, se rechaza la hipótesis nula si se verifica la región de rechazo:

$$R=\left\{ \left|\overline{x}-\mu_0\right| > t_{\alpha/2\,,\,(n-1)}\frac{s_x}{\sqrt{n}} \right\} \quad \text{siendo,} \quad \left|\overline{x}-\mu_0\right| = \left|16,32-16\right| = 0,32$$

$$t_{\alpha/2\,;\,(n-1)}\frac{s_x}{\sqrt{n}} = t_{0,025\,;\,24}\frac{s_x}{\sqrt{n}} = 2,064\frac{0,8}{\sqrt{25}} = 0,33$$

Como no se verifica la región de rechazo $R=\left\{0,32 \not> 0,33\right\}$, se acepta la hipótesis nula H_0, afirmando que el peso de llenado son 16 gr, con un nivel de significación de 0,05.

b) El p–valor (α_p) es el menor nivel de significación para el que se rechaza la hipótesis nula.

El estadístico muestral: $t_{n-1} = t_{24} = \dfrac{|\bar{x} - \mu_0|}{s_x / \sqrt{n}} = \dfrac{|16,32 - 16|}{0,8 / \sqrt{25}} = 2$

Al ser el contraste bilateral, el p–valor (α_p) viene dado por la expresión $P(t_{24} \geq 2) = \dfrac{\alpha_p}{2}$

En la tabla de la t de Student: $P(t_{24} \geq 1,711) = 0,05$ y $P(t_{24} \geq 2,064) = 0,025$

Interpolando se tiene:

0,05	$\alpha_p / 2$	0,025
1,711	2	2,064

$$0,05 - 0,025 \longrightarrow 1,711 - 2,064$$
$$\alpha_p / 2 - 0,025 \longrightarrow 2 - 2,064$$

$$\left(\frac{\alpha_p}{2} - 0,025\right) \times (1,711 - 2,064) = (0,05 - 0,025) \times (2 - 2,064) \quad \mapsto \quad \frac{\alpha_p}{2} = 0,0295 \quad \mapsto \quad \alpha_p = 0,059$$

Para un nivel de significación $\alpha = 0,05$, siendo $\alpha_p = 0,059 > \alpha = 0,05$, se acepta la hipótesis nula.

23.- Se recibe un envío de latas de conserva de las que se afirma que el peso medio son 1.000 gr. Examinada una muestra de 5 latas se obtiene un peso medio de 995 gr. con una cuasivarianza $s_x^2 = 19,6$. Al nivel de confianza del 95%, ¿se puede aceptar que el peso medio son 1.000 gr.?

Solución:

Hay una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de tamaño $n = 5$, donde la variable aleatoria X = 'peso de una lata de conserva' se supone que sigue una distribución normal $X \sim N(\mu, \sigma)$ con varianza poblacional desconocida.

Se desea saber si el resultado de que $\mu = 1000$ es aceptable. Para ello, con un nivel de significación $\alpha = 0,05$, se plantea el contraste:

Hipótesis nula $H_0 : \mu = 1000$ frente a la Hipótesis alternativa $H_1 : \mu \neq 1000$

Adviértase que como la hipótesis alternativa $H_1 : \mu \neq 1000$, en la decisión deberán ser válidos valores de μ tanto mayores o menores que 1000, por lo que el contraste debe ser bilateral o de dos colas.

En esta línea, se rechaza la hipótesis nula si se verifica la región de rechazo:

$$R = \left\{ \, |\bar{x} - \mu_0| \; > \; t_{\alpha/2\,,\,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$$

siendo, $|\bar{x} - \mu_0| = |995 - 1000| = 5$

$$t_{\alpha/2\,;\,(n-1)} \frac{s_x}{\sqrt{n}} = t_{0,025\,;\,4} \frac{s_x}{\sqrt{n}} = 2,776 \frac{\sqrt{19,6}}{\sqrt{5}} = 5,50$$

Como no se verifica la región de rechazo $R \neq \{ 5 > 5{,}5 \}$, se acepta la hipótesis nula H_0, afirmando que el peso medio son 1000 gramos, con un nivel de significación de 0,05.

CONTRASTE BILATERAL DE LA DIFERENCIA DE MEDIAS CON VARIANZAS POBLACIONALES CONOCIDAS.

24.- El análisis laboral que la U.E ha realizado para toda Europa, señala que en España, el salario mensual de los varones, en algunos sectores económicos, supera en más de 100 euros el salario de las mujeres que desempeñan las mismas tareas.
El Ministerio de Trabajo español decide considerar el salario mensual como una variable aleatoria normalmente distribuida con desviación típica de 39,6 euros para los trabajadores masculinos y de 36 euros para las trabajadoras de dichos sectores, siendo el salario de cada población independiente del de la otra. Para tratar de verificar lo publicado, se elige una muestra aleatoria simple de 500 trabajadores y de 700 trabajadoras, obteniéndose unos salarios medios mensuales de 1.500 y 1.370 euros respectivamente.
¿Está fundamentada las conclusiones de la U.E al 1% de significación?

Solución:

Sean las variables aleatorias, respectivamente, X = 'Salario mensual de los varones' e Y = 'Salario mensual de las mujeres', donde $X \sim N(\mu_x, 39{,}6)$ e $Y \sim N(\mu_y, 36)$

En las muestras se obtuvieron los resultados: $\begin{cases} \bar{x} = 1.500\,€ & n_x = 500 \\ \bar{y} = 1.370\,€ & n_y = 700 \end{cases}$

En el contraste se establecen las hipótesis:

$H_0 : \mu_x - \mu_y = 100 \qquad H_1 : \mu_x - \mu_y > 100$ (compuesta)

La regla de decisión será: $\begin{cases} \text{Si } (\bar{x} - \bar{y}) > k & \text{se rechaza } H_0 \text{ (RC)} \\ \text{Si } (\bar{x} - \bar{y}) \leq k & \text{se acepta } H_0 \text{ (RA)} \end{cases}$

La diferencia de medias muestrales $(\bar{x} - \bar{y})$, siendo las varianzas muestrales conocidas, bajo la hipótesis nula, sigue una distribución:

$$N\left[(\mu_x - \mu_y);\ \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right] \equiv N\left[100;\ \sqrt{\frac{39{,}6^2}{500} + \frac{36^2}{700}} \right] = N(100;\ 2{,}23)$$

Se determina el valor de k mediante el nivel de significación α:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right] = P\left[\, \bar{x} - \bar{y} > k \ / H_0 : \mu_x - \mu_y = 100 \right] = P\left[z > \frac{k - 100}{2{,}23} \right] = 0{,}01$$

$$\Rightarrow \ \frac{k - 100}{2{,}23} = 2{,}32 = z_{0{,}01} \quad \mapsto \quad k = 105{,}2\ €$$

La evidencia empírica $(\bar{x} - \bar{y}) = 1.500 - 1.370 = 130\ €$.

Se advierte que $(\overline{x}-\overline{y})=130>105{,}2$, esto es, la diferencia de medias muestrales cae en la región de rechazo, con lo cual se rechaza la hipótesis nula, con un nivel de significación del 1%.
Afirmando que con el mismo trabajo, las diferencias salarias entre hombres y mujeres en algunos sectores económicos españoles, son superiores a 100 euros mensuales.

25.- Con un nivel de significación del 4,72%, se desea contrastar la hipótesis nula de igualdad de medias de dos poblaciones $N(\mu_1,\,4)$ y $N(\mu_2,\,4{,}5)$. Para ello, se han tomado dos muestras aleatorias simples e independientes, respectivamente, obteniéndose los siguientes valores:

x_i	20,4	10,2	7,3	12,8	13,4	9,4
y_j	19,8	9,7	14,6	15.7	8,4	

Solución:

En el contraste bilateral se establecen las hipótesis: $\begin{cases} H_0: \mu_1-\mu_2=0 \\ H_1: \mu_1-\mu_2\neq 0 \end{cases}$

Regla de decisión $\begin{cases} \text{Si } \left|\overline{x}-\overline{y}\right|>k \ \text{ se rechaza } H_0 \\ \text{Si } \left|\overline{x}-\overline{y}\right|\leq k \ \text{ se acepta } H_0 \end{cases}$

Para analizar el contraste, se realizan los cálculos muestrales:

$$\overline{x}=\frac{\sum_{i=1}^{6} x_i}{6}=12,25 \qquad \overline{y}=\frac{\sum_{j=1}^{5} y_j}{5}=13,64 \qquad n_1=6 \qquad n_2=5$$

La región crítica de dos colas $\left|\overline{x}-\overline{y}\right|>k$ es función de la diferencia de las medias muestrales. En esta línea, las distribuciones en el muestreo de las medias son:

$$\overline{x}\sim N\left[\mu_1,\,\frac{\sigma_1}{\sqrt{n_1}}\right],\ \overline{y}\sim N\left[\mu_2,\,\frac{\sigma_2}{\sqrt{n_2}}\right]$$

Bajo la hipótesis nula $H_0:\mu_1-\mu_2=0$, la diferencia de medias muestrales se distribuye:

$$(\overline{x}-\overline{y})\sim N\left[0,\,\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] \equiv N\left[0,\,\sqrt{\frac{4^2}{6}+\frac{4{,}5^2}{5}}\right] = N\left[0\,,\,2{,}59\right]$$

Se determina el valor de k mediante el nivel de significación α:

$$\alpha=P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ cierta}\right]=P\left[\,\left|\overline{x}-\overline{y}\right|>k\ /\ H_0:\mu_1-\mu_2=0\,\right]=$$

$$=P\left[\left|\frac{(\overline{x}-\overline{y})-0}{2{,}59}\right|>K\right]=P\left[\left(\frac{\overline{x}-\overline{y}}{2{,}59}<-K\right)\cup\left(\frac{\overline{x}-\overline{y}}{2{,}59}>K\right)\right]=$$

$$= P\left[\frac{\overline{x}-\overline{y}}{2,59} < -K\right] + P\left[\frac{\overline{x}-\overline{y}}{2,59} > K\right] = \frac{\alpha}{2} + \frac{\alpha}{2} = 0,05 \ \text{ por simetría } N(0,1)$$

La región crítica es $\left|\dfrac{\overline{x}-\overline{y}}{2,59}\right| > 1,995 = z_{0,0236} \quad \mapsto \quad \begin{cases} \overline{x}-\overline{y} > 5,17 \\ \overline{x}-\overline{y} < -5,17 \end{cases}$

En consecuencia, la región de aceptación: $-5,17 \leq \overline{x}-\overline{y} \leq 5,17$

La evidencia empírica $\left|\overline{x}-\overline{y}\right| = \left|12,25 - 13,64\right| = 1,39$, valor que se encuentra en la región de aceptación por lo que se acepta la hipótesis nula de igualdad de medias, con un nivel de significación del 4,72%.

Análogamente, se podría haber resuelto considerando la región de rechazo de la hipótesis nula H_0:

$$R = \left\{ \left|\overline{x}-\overline{y}\right| > z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\} \mapsto R = \left\{ \left|12,25 - 13,64\right| > (1,995)\sqrt{\frac{4^2}{6} + \frac{4,5^2}{5}} \right\}$$

La región de rechazo de la hipótesis nula no se cumple, $R \neq \left\{1,39 > 5,17\right\}$, concluyendo que existe igualdad entre las medias poblacionales.

Cálculo de $z_{0,0236}$:

Abscisas	Áreas
1,98 − 1,99	0,0239 − 0,0233
x − 1,99	0,0236 − 0,0233

$$x = 1,99 + \frac{0,01 \,.\, 0,0003}{0,0006} = 1,995$$