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

ERROR TIPO I - ERROR TIPO II - POTENCIA CONTRASTE
CONTRASTES DE HIPÓTESIS
EJERCICIOS TEÓRICOS CON SPSS

RELACIÓN: INTERVALOS CONFIANZA - CONTRASTE HIPÓTESIS

a)

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

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

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

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

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 derecha

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

b)

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

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

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

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

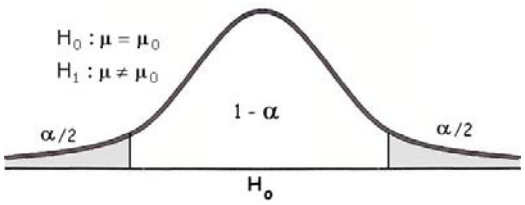
$$I_{1-\alpha}(\mu) = \left\{ \bar{x} \pm t_{\frac{\alpha}{2},(n-1)} \frac{s_x}{\sqrt{n}} \right\} \mapsto \left|\bar{x} - \mu\right| \le t_{\frac{\alpha}{2},(n-1)} \frac{s_x}{\sqrt{n}}$$

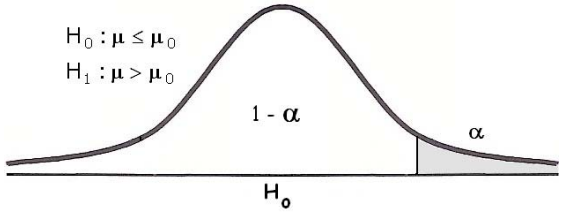
HIPÓTESIS SOBRE LA MEDIA DE UNA POBLACIÓN CON σ^2 DESCONOCIDA: REGIÓN DE RECHAZO

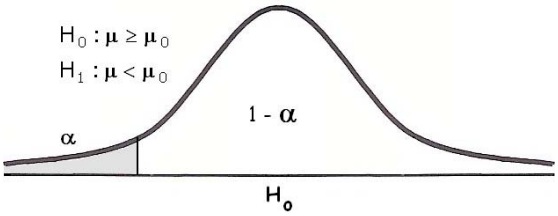
$$H_0:\ \mu=\mu_0 \quad H_1:\ \mu\neq\mu_0 \qquad R=\left\{\ \left|\bar{x}-\mu_0\right| > t_{\alpha/2;(n-1)}\frac{s_x}{\sqrt{n}}\right\} \text{ bilateral (compuesta)}$$

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

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

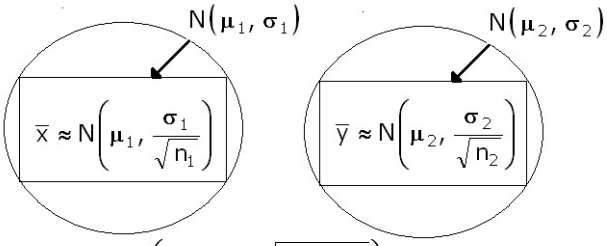






c)

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:



$$\bar{x}\approx N\left(\mu_1,\frac{\sigma_1}{\sqrt{n_1}}\right) \qquad \bar{y}\approx N\left(\mu_2,\frac{\sigma_2}{\sqrt{n_2}}\right)$$

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

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

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

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

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

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

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 $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, que bajo la hipótesis nula $\bar{x} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$, con lo que la variable $\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ es N(0, 1).

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

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

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

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

En otras palabras,

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

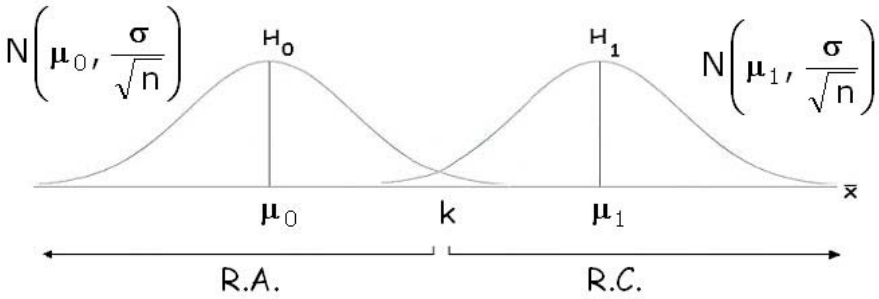
b) CONTRASTE UNILATERAL o DE UNA COLA

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

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



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

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

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

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

En otras palabras,

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

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: $\begin{cases} \text{Si } \left|\overline{x} - \overline{y}\right| \leq k \text{ no se rechaza } H_0 \mapsto \text{(R.A)} \\ \text{Si } \left|\overline{x} - \overline{y}\right| > k \text{ se rechaza } H_0 \mapsto \text{(R.C)} \end{cases}$

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:

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

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

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

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

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

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

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

La región crítica es

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

En otras palabras,

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

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

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: $\begin{cases} \text{Si } (\bar{x} - \bar{y}) \leq k \text{ no se rechaza } H_0 \mapsto \text{(R.A)} \\ \text{Si } (\bar{x} - \bar{y}) > k \text{ se rechaza } H_0 \mapsto \text{(R.C)} \end{cases}$

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

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

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

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

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

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

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

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

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

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

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

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:

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

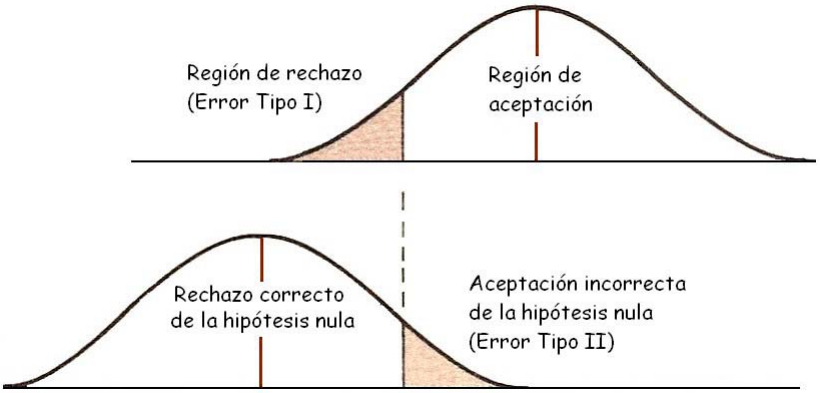
$\beta = P(\text{Error Tipo II}) = 0 \;\mapsto\; \text{Rechazar siempre } H_0 \Leftrightarrow \alpha = P(\text{Error Tipo I}) = 1$

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.

$$\alpha_p = \text{p}-\text{valor} = P\left[\text{Rechazar el estadístico muestral observado / } H_0 \text{ es cierta}\right]$$

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).

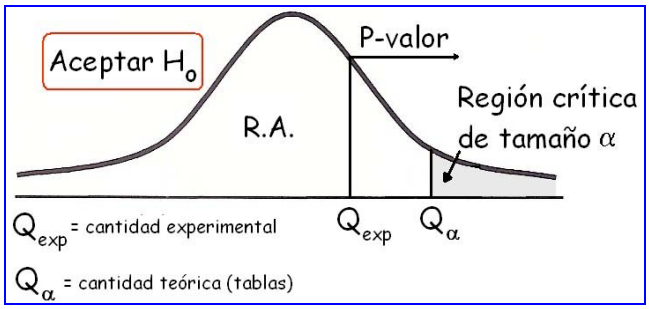
CRITERIOS GENERALES para CONTRASTES:

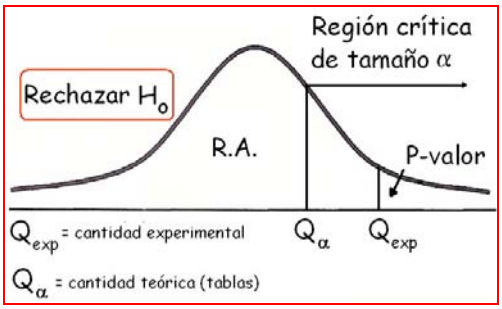
Calcular una cantidad experimental Q_{exp} a partir de los datos

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}





Definir el Error de Tipo I y el Error de Tipo II. Analice la verdad o no de las siguientes afirmaciones y explique por qué:

a) P(Error tipo I) + P(Error Tipo II) = 1

b) Cuando disminuye la probabilidad de cometer un Error de Tipo I, también disminuye la probabilidad de cometer un Error de Tipo II.

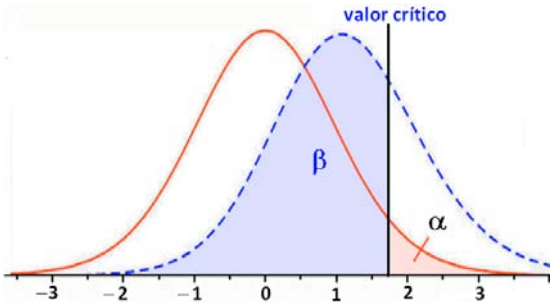
Solución:

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

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

No es verdad que $P(\text{Error tipo I}) + P(\text{Error Tipo II}) = 1$ porque no son complementarios

b) No es verdad. Se encuentran inversamente relacionados.



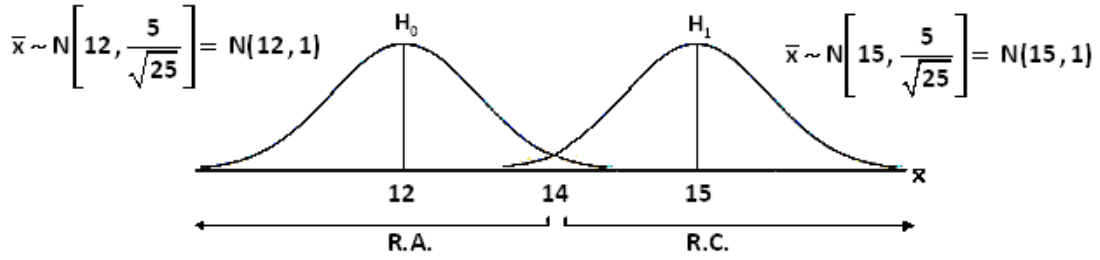
1. Sea una variable aleatoria X procedente de una población con densidad de probabilidad $N(\mu, 5)$. Efectuadas dos hipótesis sobre el valor de μ : $H_0 : \mu = 12 \quad H_1 : \mu = 15$
Mediante un muestreo aleatorio simple de tamaño 25, se contrasta la hipótesis H_0 respecto de la hipótesis H_1, estableciéndose 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) Se trata de un contraste unilateral de hipótesis sobre la media poblacional

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

$H_0: \mu = 12 \quad H_1: \mu = 15$



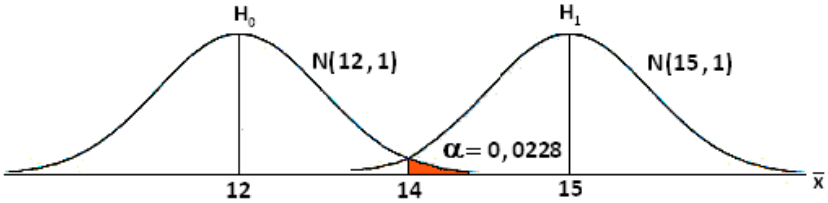
La distribución de la media muestral $\bar{x}$ de tamaño 25 bajo la hipótesis nula y la hipótesis alternativa viene dada:

$$H_0: \ \bar{x} \sim N\left[\mu_0, \frac{\sigma}{\sqrt{n}}\right] = N\left[12, \frac{5}{\sqrt{25}}\right] = N(12,1) \qquad H_1: \ \bar{x} \sim N\left[\mu_1, \frac{\sigma}{\sqrt{n}}\right] = N\left[15, \frac{5}{\sqrt{25}}\right] = N(15,1)$$

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

$$\alpha = P(\text{ET I}) = P\left[\bar{x} > 14 \,\middle|\, H_0 : \mu = 12\right] = P\left[z \ge \frac{14-12}{1}\right] = P(z \ge 2) = 0{,}0228$$

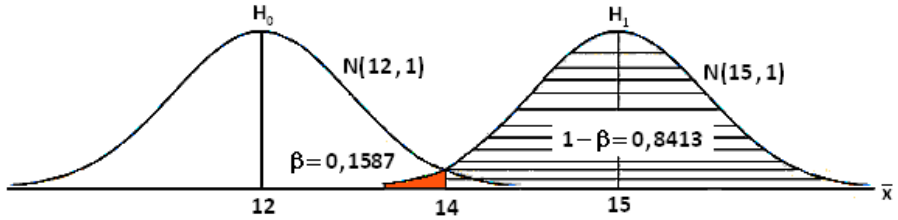
H_0 se rechaza cuando es cierta el 2,28% de los casos



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

$$\beta = P(\text{ET II}) = P\left[\bar{x} \le 14 \,\middle|\, H_1 : \mu = 15\right] = P\left[z \le \frac{14-15}{1}\right] = P(z \le -1) = P(z \ge 1) = 0{,}1587$$

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



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

$1-\beta = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Falsa}\right] = 1-0,1587 = 0,8413$ (Bondad del tipo de ajuste)

O bien,

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

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

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

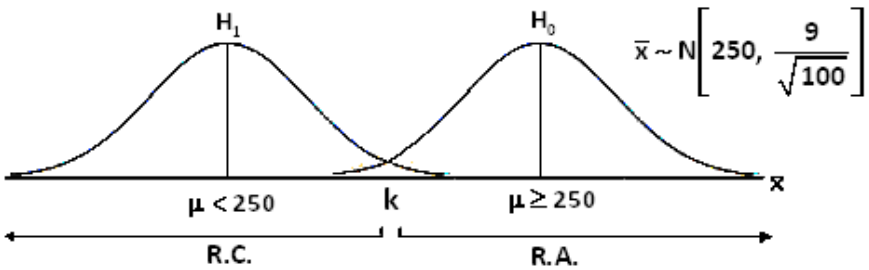
Resaltar que es más grave cometer un Error Tipo I (α) que un Error Tipo II (β)

2. Las latas de mejillones de una determinada marca indican que el peso escurrido de dicho producto es de 250 gramos. 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 gramos. Se pide:
a) Determinar, si existe, la mejor región crítica para contrastar, con un nivel de significación del 5% y muestras aleatorias simples de tamaño 100.
b) 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 gramos.
c) Determinar la función de potencia del contraste.

Solución:

a) Sea la variable aleatoria X = " Peso escurrido de las latas de mejillones"
Contraste unilateral: $H_0: \mu \geq 250$ $H_1: \mu < 250$

Regla de decisión del muestreo: $\begin{cases} \bar{x} \geq k & \text{se acepta } H_0 \text{ (R.A.)} \\ \bar{x} < k & \text{se rechaza } H_0 \text{ (R.C.)} \end{cases}$



Bajo la hipótesis nula $\bar{x} \sim N\left[250, \frac{9}{\sqrt{100}}\right] \equiv N(250, 0,9)$

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

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

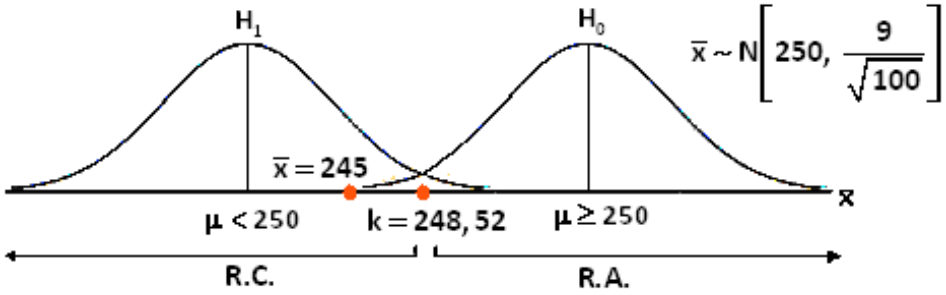
$$= P[\bar{x} \le k / N(250; 0,9)] = P\left[\frac{\bar{x}-250}{0,9} \le \frac{k-250}{0,9}\right] = P\left[z \le \frac{k-250}{0,9}\right] = P\left[z \ge \frac{250-k}{0,9}\right] = 0,05$$

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

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

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

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



c) La función potencia del contraste se establece como:

$$P(\mu) = P(\bar{x} \le 248,52) = P\left[\bar{x} \le k / N\left(\mu; \frac{9}{\sqrt{100}}\right)\right] = P\left[\frac{\bar{x}-\mu}{0,9} \le \frac{248,52-\mu}{0,9}\right] == P\left[z \le \frac{248,52-\mu}{0,9}\right]_{\mu \le 250}$$

(bondad del tipo de ajuste)

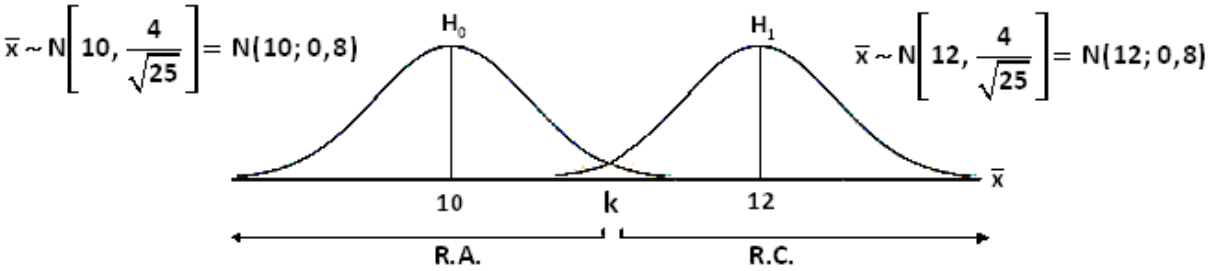
3. Sea una variable aleatoria X procedente de una población con densidad de probabilidad $N(\mu, 4)$. Se quiere contrastar la hipótesis nula H_0: $\mu = 10$ frente a la hipótesis alternativa H_1: $\mu = 12$, con un nivel de significación $\alpha = 0,05$, con un muestreo simple de tamaño 25. Determinar:

a) La probabilidad de cometer el error tipo II
b) La potencia del contraste

Solución:

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

Las hipótesis sobre la media poblacional (contraste unilateral):



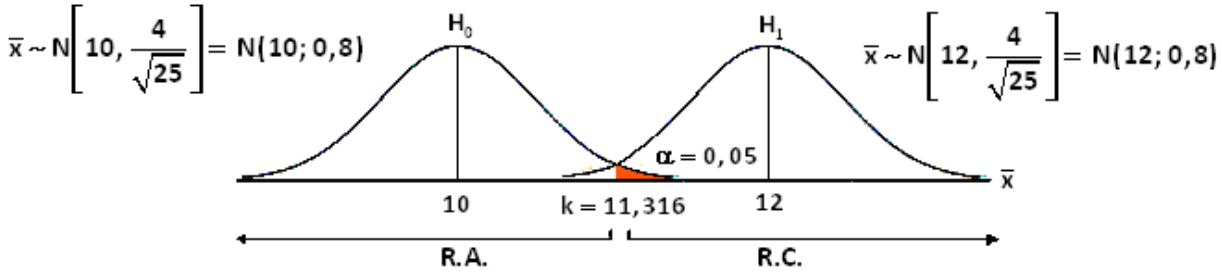
La distribución de la media muestral $\bar{x}$ de tamaño 25 bajo la hipótesis nula y alternativa, respectivamente:

$$\bar{x} \sim N\left[\mu_0, \frac{\sigma}{\sqrt{n}}\right] = N\left[10, \frac{4}{\sqrt{25}}\right] = N(10; 0,8) \qquad \bar{x} \sim N\left[\mu_1, \frac{\sigma}{\sqrt{n}}\right] = N\left[12, \frac{4}{\sqrt{25}}\right] = N(12; 0,8)$$

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

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

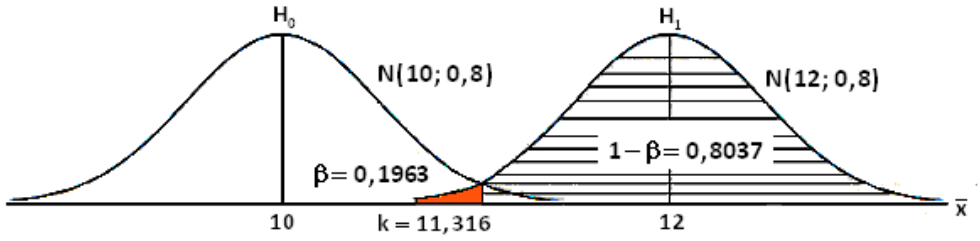
$$\alpha = P[ET\ I] = P[\bar{x} > k \mid H_0 : \mu_0 = 10] = P\left[z > \frac{k-10}{0,8}\right] = 0,05 \rightarrow \frac{k-10}{0,8} = 1,645 \Rightarrow k = 11,316$$



$$\beta = P[ET\ II] = P[\text{Aceptar } H_0 / H_0 \text{ Falsa}]$$

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

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



b) Potencia del Contraste: $\text{Potencia} = P[\text{Rechazar } H_0 \mid H_0 \text{ Falsa}] = 1 - \beta = 1 - 0,1963 = 0,8037$

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:

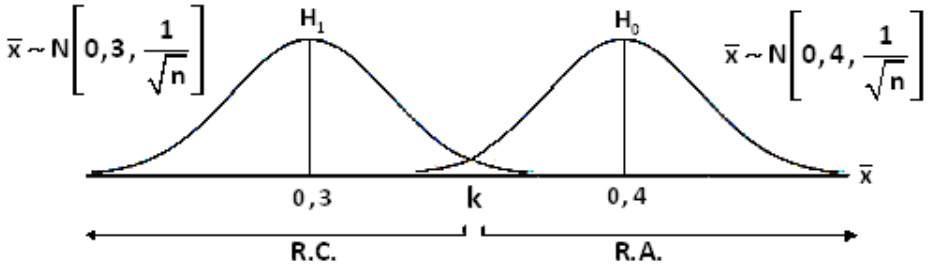
a) ¿Cuál es el tamaño de la muestra utilizada por el agricultor?
b) Qué hipótesis fue aceptada

Solución:

a) La variable aleatoria X = ' Peso en kg. de las patatas, $X \sim N(\mu, 1)$

Hipótesis sobre μ: $H_0: \mu_0 = 0,4 \quad H_1: \mu_1 = 0,3$

Regla de decisión $\begin{cases} \bar{x} \ge k & \rightarrow \text{ Se acepta } H_0 \text{ (R.A.)} \\ \bar{x} < k & \rightarrow \text{ Se rechaza } H_0 \text{ (R.C.)} \end{cases}$



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

El valor crítico 'k' se obtiene a partir del nivel de significación:

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

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

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,

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

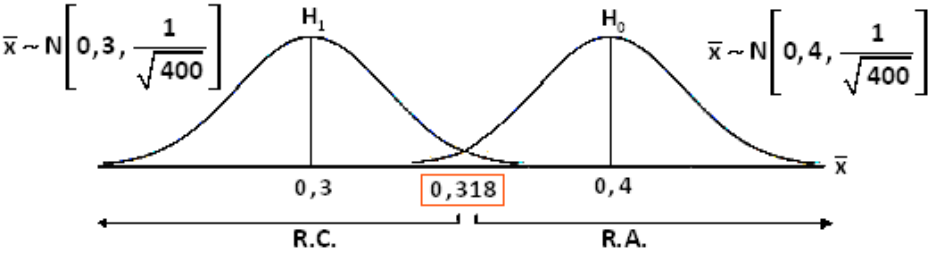
Resolviendo el sistema:

$$\begin{cases} 0,4-k = 1,64/\sqrt{n} \\ k-0,3 = 0,36/\sqrt{n} \end{cases} \rightarrow 0,1 = 2/\sqrt{n} \rightarrow \sqrt{n} = 2/0,1 \rightarrow n = 400$$

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

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

b) Regla de decisión $\begin{cases} \bar{x} \geq 0,318 \rightarrow \text{Se acepta } H_0 \text{ (R.A.)} \\ \bar{x} < 0,318 \rightarrow \text{Se rechaza } H_0 \text{ (R.C.)} \end{cases}$



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

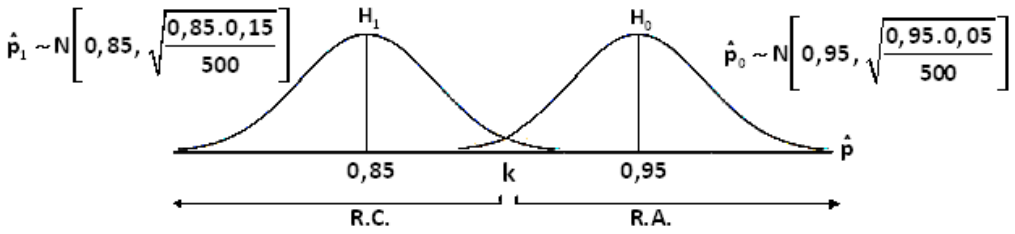
5. Un laboratorio farmacéutico quiere lanzar un nuevo medicamento para la hipertensión, llamado Hipotensil. El director de dicho laboratorio cree que la eficacia del medicamento sería de un 95%, medida ésta como la proporción de pacientes a los que se les suministra y que experimentan una mejoría. Sin 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 izquierda): H_0: $p_0 = 0{,}95$ H_1: $p_1 = 0{,}85$

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



La distribución en el muestreo del estadístico $\hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n} \overbrace{\sim}^{\text{Teorema Central Límite}} N\left[p, \sqrt{\dfrac{p.q}{n}}\right]$

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

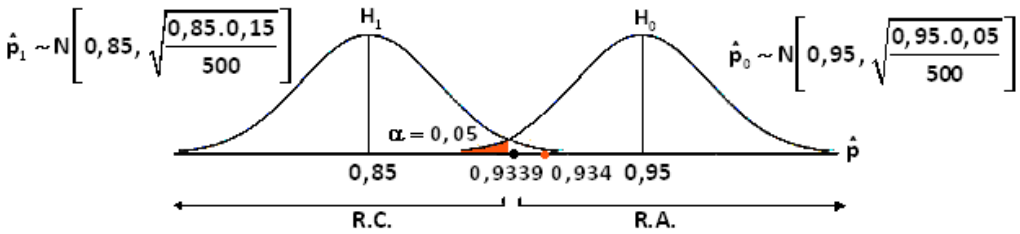
Bajo la hipótesis H_1: $\hat{p}_1 \sim N\left[0{,}85, \sqrt{\dfrac{0{,}85.0{,}15}{500}}\right] = N(0{,}85, 0{,}01597)$

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

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

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

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



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

El valor del estadístico muestral $\hat{p}$ (evidencia empírica) es $\hat{p} = \frac{467}{500} = 0,934$, siendo $\hat{p} = 0,934 > 0,9339$, 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.

6. Un portal e-business sabe que el 60% de todos sus visitantes a la web están interesados en adquirir sus productos pero no reacios al comercio electrónico y no realizan finalmente la compra vía internet. Sin embargo, en la dirección del portal se piensa que en el último año, el porcentaje de gente que está dispuesta a comprar por internet ha aumentado y esto se debe reflejar en sus resultados empresariales.
a) Contrastar con un nivel de significación del 2%, si en el último año se ha reducido el porcentaje de gente que no está dispuesta a comprar por internet, si para ello se tomó una muestra de 500 visitantes para conocer su opinión y se observó que el 55% no estaba dispuesto a realizar compras vía on-line.
b) Realizar el contraste con el p_valor

Solución:

Sea el parámetro p = "Proporción del número de visitantes al portal". Al realizar el contraste sobre la proporción, se parte de una muestra aleatoria $(x_1, x_2, \cdots, x_{500})$, donde $X \sim B(1, p)$

La distribución del parámetro muestral $\hat{p} = \sum_{i=1}^{500} x_i / n$ al ser de tamaño suficientemente grande $n = 500$ y estar $\hat{p}$ definido como suma de variables independientes entre sí según una distribución de Bernouilli $B(1, p)$, por el teorema central del límite (TCL) se puede aproximar

$$\hat{p} = \frac{\sum_{i=1}^{500} x_i}{n} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Se establecen las hipótesis: $H_0: p = 0,6 \quad H_1: p < 0,6$

Se trata de un contraste unilateral por la izquierda

Regla de decisión: $\begin{cases} \hat{p} \geq k & \text{se acepta } H_0 \\ \hat{p} < k & \text{se rechaza } H_0 \end{cases}$

Bajo la hipótesis nula $\hat{p} \sim N\left(0,6, \sqrt{\frac{0,6 \times 0,4}{500}}\right) = N(0,6,\ 0,022)$

A partir del nivel de significación $\alpha = 0,02$ se determina el valor crítico k:

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}\right] = P\left[\hat{p} < k \mid N(0,6,\ 0,022)\right] = P\left[\frac{\hat{p} - 0,6}{0,022} < \frac{k - 0,6}{0,022}\right] =$$

$$= P\left[z < \frac{k - 0,6}{0,022}\right] = P\left[z \geq \frac{0,6 - k}{0,022}\right] = 0,02 \quad \rightarrow \quad \frac{0,6 - k}{0,022} = 2,055 \quad \Rightarrow \quad k = 0,555$$

El valor del estadístico muestral (evidencia empírica) es $\hat{p} = 0,55 < \in 0,555$, rechazando la hipótesis nula. En conclusión, existe evidencia empírica que la proporción de visitantes al portal

que están dispuestos a comprar on-line ha disminuido, es decir, el porcentaje de visitantes que son reacios a comprar por internet ha aumentado.

b) El p_valor es el nivel de significación más pequeño posible que puede escogerse para el que todavía se aceptaría la hipótesis alternativa con las observaciones actuales. Paquetes Estadísticos proporcionan el *significance level* (nivel de significación) cuando se refieren en realidad al p_valor (p_value).

$$\alpha_p = p_valor = P\left[\text{Rechazar el estadístico muestral observado / } H_0 \text{ es Cierta}\right]$$

$$\alpha_p = P\left[\hat{p} < 0{,}55 \,\middle|\, N(0{,}6,\ 0{,}022)\right] = P\left[\frac{\hat{p}-0{,}6}{0{,}022} < \frac{0{,}55-0{,}6}{0{,}022}\right] = P(z < -2{,}27) = P(z \geq 2{,}27) = 0{,}0116$$

Siendo $\alpha_p = 0{,}0116 < \alpha = 0{,}02$ se rechaza la hipótesis nula a un nivel de significación del 2%

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 sí 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 se parte 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 tiene:

La distribución en el muestreo del estadístico $\hat{p} = \dfrac{\sum_{i=1}^{n} X_i}{n} \overbrace{\sim}^{\text{Teorema Central Límite}} N\left[p, \sqrt{\dfrac{p.q}{n}}\right]$

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

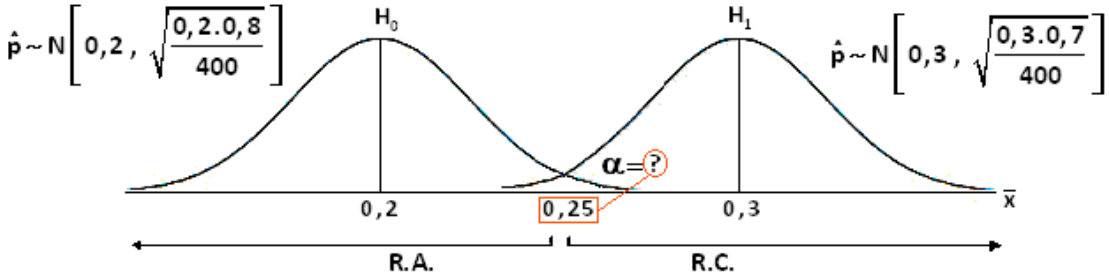
Bajo la hipótesis H_1: $\hat{p} \sim N\left[0{,}3\ ,\ \sqrt{\dfrac{0{,}3.0{,}7}{400}}\right] = N(0{,}3\ ,\ 0{,}0229)$

Se establecen las hipótesis: $H_0: p = 0{,}2 \quad H_1: p < 0{,}3$

Por el lema de Neyman-Pearson, la regla de decisión del muestreo $k = \dfrac{100}{400} = 0{,}25$ es:

$$\begin{cases} \hat{p} \le 0,25 & \text{Se acepta } H_0 \text{ (R.A.)} \\ \hat{p} > 0,25 & \text{Se rechaza } H_0 \text{ (R.C.)} \end{cases}$$

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



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

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

b) $\text{Potencia} = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Falsa}\right] = 1-\beta = 1 - P[\text{Aceptar } H_0 \,/\, H_0 \text{ Falsa}]$

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

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

En consecuencia, bondad del tipo de ajuste:

$$\text{Pot} = 1-\beta = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Falsa}\right] = 1 - 0,0144 = 0,9856$$

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: $1 \quad 0,9 \quad -0,2 \quad 1,4 \quad -0,7$

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:

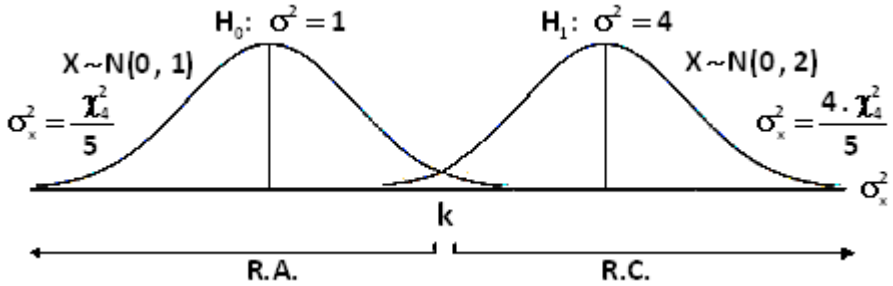
Sea la variable aleatoria X = ' Errores en el peso' $X \sim N(0,\sigma)$

Distribución en el muestreo: $\chi^2_{n-1} = \dfrac{\text{Varianza muestral}}{\text{Varianza teórica}} = \dfrac{\sigma_x^2}{(\sigma/\sqrt{n})^2} \rightarrow \sigma_x^2 = \dfrac{\sigma^2 \cdot \chi^2_{n-1}}{n}$

Se establecen las hipótesis: H_0: $\sigma^2 = 1$ $\quad H_1$: $\sigma^2 = 4$

Bajo la hipótesis: H_0: $\sigma_x^2 = \dfrac{\chi^2_4}{5}$ $\quad H_1$: H_0: $\sigma_x^2 = \dfrac{4 \cdot \chi^2_4}{5}$

Regla de decisión: $\begin{cases} \sigma_x^2 \le k & \text{se acepta } H_0 \\ \sigma_x^2 > k & \text{se rechaza } H_0 \end{cases}$



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

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

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

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

x_i	1	0,9	−0,2	1,4	−0,7	$\bar{x} = \sum_{i=1}^{5} \frac{x_i}{5} = 0{,}48$
x_i^2	1	0,81	0,04	1,96	0,49	$\sum_{i=1}^{5} \frac{x_i^2}{5} = 0{,}86$

$$\sigma_x^2 = \sum_{i=1}^{5} \frac{x_i^2}{5} - \left(\sum_{i=1}^{5} \frac{x_i}{5}\right)^2 = 0{,}6296$$

Como $\sigma_x^2 = 0{,}6296 < 1{,}898$ se sitúa en la región de aceptación (R.A) , con un nivel de confianza del 95% no se puede rechazar la hipótesis de que la dispersión sea 1.

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.

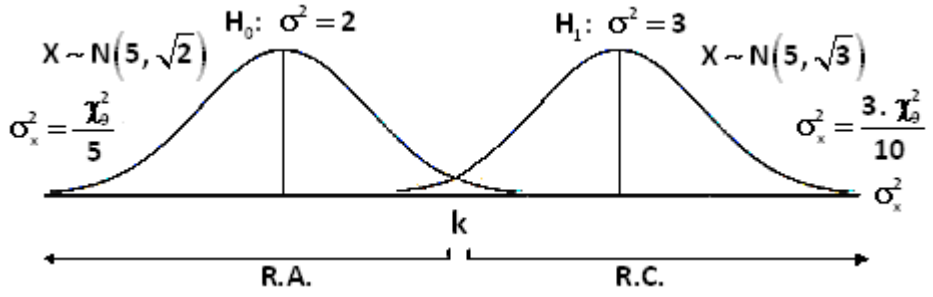
Solución:

Distribución en el muestreo: $\chi_{n-1}^2 = \dfrac{\text{Varianza muestral}}{\text{Varianza teórica}} = \dfrac{\sigma_x^2}{(\sigma / \sqrt{n})^2} \rightarrow \sigma_x^2 = \dfrac{\sigma^2 \cdot \chi_{n-1}^2}{n}$

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

Bajo la hipótesis: $H_0: \sigma_x^2 = \dfrac{2 \cdot \chi_9^2}{10} = \dfrac{\chi_9^2}{5}$ $\quad H_1: \sigma_x^2 = \dfrac{3 \cdot \chi_9^2}{10}$

Regla de decisión: $\begin{cases} \sigma_x^2 \leq k & \text{se acepta } H_0 \\ \sigma_x^2 > k & \text{se rechaza } H_0 \end{cases}$



Determinación del valor crítico k a partir del nivel de significación α (Error Tipo I):

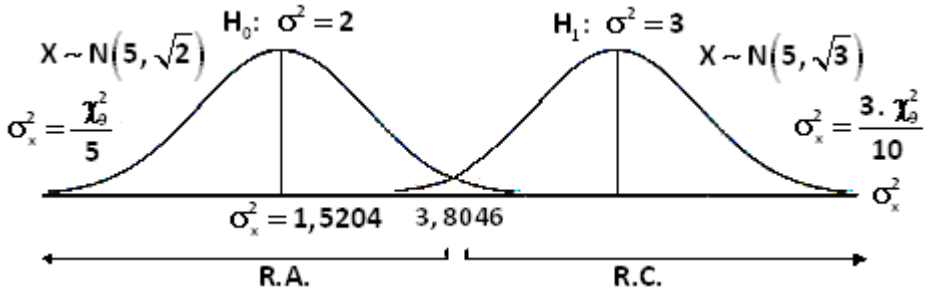
$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, \sigma^2 = 2\right] =$$

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

Se compara el valor crítico $k = 3{,}8046$ con la evidencia muestral:

x_i	5,1	6,2	4	2,8	2,9	5,6	3,7	3,4	2,5	5,2	$\bar{x} = \sum_{i=1}^{10} \frac{x_i}{10} = 4{,}14$
x_i^2	26,01	38,44	16	7,84	8,41	31,36	13,69	11,56	6,25	27,04	$\sum_{i=1}^{10} \frac{x_i^2}{10} = 18{,}66$

$$\sigma_x^2 = \sum_{i=1}^{10} \frac{x_i^2}{10} - \left(\sum_{i=1}^{10} \frac{x_i}{10}\right)^2 = 18{,}66 - 4{,}14^2 = 1{,}5204$$



Siendo $\sigma_x^2 = 1{,}5204 < 3{,}8046$ se sitúa en la región de aceptación (R.A) , con un nivel de confianza del 97,5% se acepta que la la varianza es 2.

De otra parte, la potencia del contraste:

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Falsa}\right] \equiv P\left[\text{Aceptar } H_1 \,\middle|\, H_1 \text{ Cierta}\right] = P\left[\frac{3 . \chi_9^2}{10} > 3{,}8046\right] =$$

$$= P\left[\chi_9^2 \geq 12{,}682\right] = 0{,}252$$

Abscisas		Áreas						
4,168 – 14,684	$\longleftrightarrow$	0,90 – 0,10	$-10{,}516$	$\longleftrightarrow$	0,80	$\rightarrow$	$x = 0{,}10 + \frac{2{,}002 . 0{,}80}{10{,}516} = 0{,}252$
12,682 – 14,684	$\longleftrightarrow$	x – 0,10	$-2{,}002$	$\longleftrightarrow$	x – 0,10		

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:

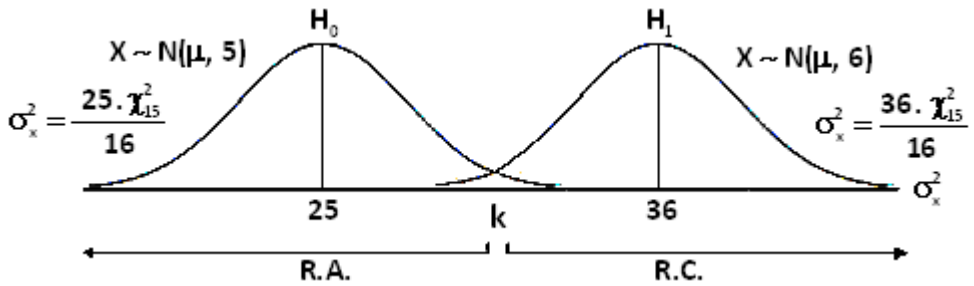
a) Por el Lema de Fisher-Cochran, la distribución en el muestreo:

$$\chi^2_{n-1} = \frac{\text{Varianza muestral}}{\text{Varianza teórica}} = \frac{\sigma_x^2}{(\sigma/\sqrt{n})^2} \rightarrow \sigma_x^2 = \frac{\sigma^2 \cdot \chi^2_{n-1}}{n}$$

Hipótesis: $H_0: \sigma^2 = 25 \qquad H_1: \sigma^2 = 36$

Bajo la hipótesis: $H_0: \sigma_x^2 = \dfrac{25.\chi^2_{15}}{16} \qquad H_1: \sigma_x^2 = \dfrac{36.\chi^2_{15}}{16}$

Regla de decisión: $\begin{cases} \sigma_x^2 \le k & \text{se acepta } H_0 \\ \sigma_x^2 > k & \text{se rechaza } H_0 \end{cases}$



La determinación del valor crítico k a partir del nivel de significación $\alpha = 0{,}01$:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\sigma_x^2 > k \,\middle|\, \sigma^2 = 25\right] =$$

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

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%

b) 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:

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

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

Como $\sigma_x^2 = 27 < 34{,}855$ se sitúa en la región de aceptación (R.A), con un nivel de confianza del 90% se acepta que la la varianza es 25.

c) Potencia del contraste con $\alpha = 0,10$

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Falsa}\right] \equiv P\left[\text{Rechazar } H_0 \mid H_1 \text{ Cierta}\right] =$$

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

Abscisas		Áreas						
8,547 – 22,307	$\longleftrightarrow$	0,90 – 0,10	$-13,76$	$\longleftrightarrow$	0,80	$\rightarrow$	$x = 0,10 + \frac{6,816 \,.\, 0,80}{13,76} = 0,496$
15,491 – 22,307	$\longleftrightarrow$	x – 0,10	$-6,816$	$\longleftrightarrow$	x – 0,10		

11. El propietario de un automóvil sospecha que su vehículo según una ley normal 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) Calcula el p_valor

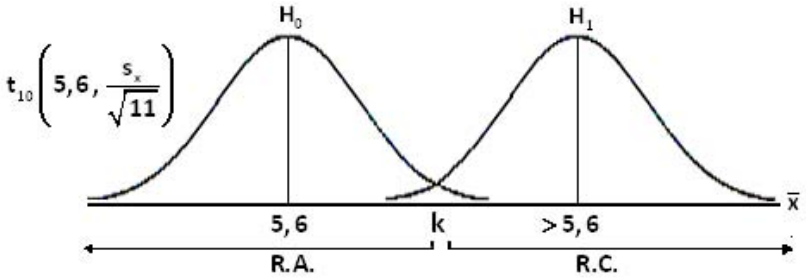
c) ¿En cuántas ocasiones debería observarse el consumo medio para que con un nivel de confianza del 99% se detectase un consumo medio de 5,9 litros/100 km?

Solución:

a) En el muestreo de una población normal $N(\mu, \sigma)$ con varianza poblacional desconocida, con un tamaño muestral pequeño $n \leq 30$, la media muestral $\bar{x} \sim t_{n-1}\left(\mu, \frac{s_x}{\sqrt{n}}\right) \equiv t_{n-1}\left(\mu, \frac{\sigma_x}{\sqrt{n-1}}\right)$

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

Se trata, pues, de un contraste unilateral (cola a la derecha)



Regla de decisión:

$$\begin{cases} \bar{x} \leq k & \text{se acepta } H_0 \\ \bar{x} > k & \text{se rechaza } H_0 \end{cases}$$

La media muestral bajo la hipótesis nula sigue una distribución $t-\text{Student}$ con $(n-1)$ grados de libertad: $t_{10}\left(5,6, \frac{0,4612}{\sqrt{11}}\right) \equiv t_{10}(5,6, 0,139)$

Con los datos muestrales:

x_i	6,1	6,5	5,1	6	5,9	5,2	5,8	5,3	6,2	5,9	6,3
x_i^2	37,21	42,25	26,01	36	34,81	27,04	33,64	28,09	38,44	34,81	39,69

$$\bar{x} = \sum_{i=1}^{11} \frac{x_i}{11} = \frac{64,3}{11} = 5,8454 \qquad \sigma_x^2 = \sum_{i=1}^{11} \frac{x_i^2}{11} - \left(\sum_{i=1}^{11} \frac{x_i}{11} \right)^2 = \frac{377,99}{11} - \left(\frac{64,3}{11} \right)^2 = 0,19339$$

$$n \,.\, \sigma_x^2 = (n-1) \,.\, s_x^2 \quad \rightarrow \quad s_x^2 = \frac{n \,.\, \sigma_x^2}{(n-1)} = \frac{11 \,.\, 0,19339}{10} = 0,21273 \quad \rightarrow \quad s_x = \sqrt{0,21273} = 0,4612$$

El valor crítico k, bajo la hipótesis nula, se calcula a partir del nivel de significación $\alpha = 0,01$:

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

$$= P\left[\frac{\bar{x} - 5,6}{0,139} > \frac{k - 5,6}{0,139} \right] = P\left[t_{10} > \frac{k - 5,6}{0,139} \right] = 0,01 \quad \rightarrow \quad \frac{k - 5,6}{0,139} = 2,764 \quad \rightarrow \quad k = 5,9842$$

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) El p_valor (α_p) es el nivel de significación más pequeño posible que puede escogerse para el que todavía se aceptaría la hipótesis alternativa con las observaciones actuales.

Si $\alpha_p > \alpha$ se acepta la hipótesis nula H_0

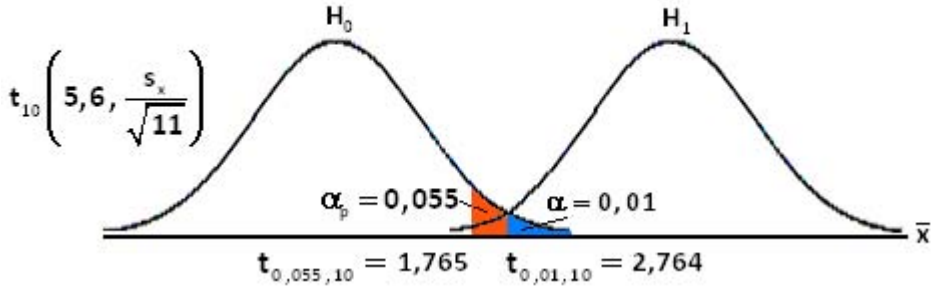
$$\alpha_p = p_valor = P\left[\text{Rechazar el estadístico muestral observado / } H_0 \text{ Cierta}\right] =$$

$$= P\left[\bar{x} > 5,8454 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\bar{x} > 5,8454 \,\middle|\, t_{10}(5,6\ ,\ 0,139)\right] =$$

$$= P\left[\frac{\bar{x} - 5,6}{0,139} > \frac{5,8454 - 5,6}{0,139} \right] = P\left[t_{10} > 1,765\right] = 0,055$$

En la tabla de la $t-$Student $\quad \dfrac{1,372 - 1,812}{0,10 - 0,05} = \dfrac{1,765 - 1,812}{x - 0,05} \quad \rightarrow \quad x = 0,055$

Si $\alpha_p = 0,055 > \alpha = 0,01$ se acepta la hipótesis nula H_0



Para un nivel de significación $\alpha = 0,01$, el p-valor $\alpha_p = 0,055 > \alpha = 0,01$, aceptando la hipótesis nula. Es decir, con un nivel de confianza del 99% se acepta que el consumo medio de combustible en carretera superior es de 5,6 litros /100 km.

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

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

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

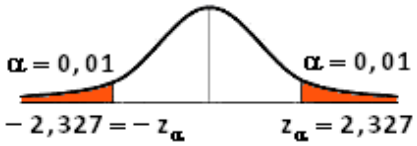
$$P\left[\bar{x} > k / \mu = 5{,}9\right] = P\left[\frac{\bar{x} - 5{,}6}{s_x / \sqrt{n}} > 2{,}764 \mid \mu = 5{,}9\right] = P\left[\frac{\bar{x} - 5{,}9 + 0{,}3}{s_x / \sqrt{n}} > 2{,}764\right] =$$

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

O bien, $P\left[\frac{\bar{x} - 5{,}9}{s_x / \sqrt{n}} \leq 2{,}764 - \frac{0{,}3}{s_x / \sqrt{n}}\right] = 0{,}01$

En un punto donde no se pueden utilizar las tablas de la $t-$Student 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):



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

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

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

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:

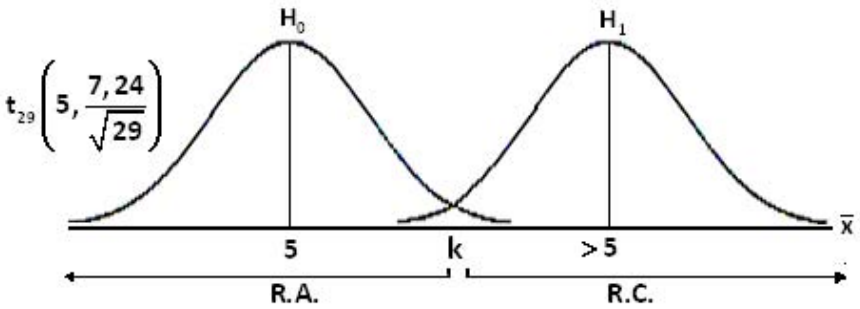
a) Sea la variable aleatoria X = ' Conexión a Internet por hogares' $X \sim N(5, \sigma)$

Hipótesis de contraste: $H_0: \mu_0 \geq 5$ $H_1: \mu_1 > 5$

b) Se trata de un contraste unilateral (cola derecha). Regla de decisión: $\begin{cases} \bar{x} \leq k & \text{se acepta } H_0 \\ \bar{x} > k & \text{se rechaza } H_0 \end{cases}$

En el muestreo de una población normal $N(\mu, \sigma)$ con varianza poblacional desconocida, con un tamaño muestral pequeño $n \leq 30$, la media muestral bajo la hipótesis nula sigue una distribución

$$\bar{x} \sim t_{n-1}\left(\mu, \frac{\sigma_x}{\sqrt{n-1}}\right) \equiv t_{29}\left(5, \frac{7,24}{\sqrt{29}}\right)$$



c) Determinación del valor crítico k a partir del nivel de significación $\alpha = 0,10$:

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}\right] = P\left[\bar{x} > k \mid H_0 \text{ Cierta}\right] = P\left[\bar{x} > k \mid t_{29}\left(5, \frac{7,24}{\sqrt{29}}\right)\right] =$$

$$= P\left[\frac{\bar{x}-5}{7,24/\sqrt{29}} > \frac{k-5}{7,24/\sqrt{29}}\right] = P\left[t_{29} > \frac{k-5}{7,24/\sqrt{29}}\right] = 0,10 \rightarrow \frac{\sqrt{29}\,.\,(k-5)}{7,24} = 1,311$$

$$k = 5 + \frac{1,311\,.\,7,24}{\sqrt{29}} = 6,76$$

Regla de decisión: $\begin{cases} \bar{x} \leq 6,76 & \text{se acepta } H_0 \\ \bar{x} > 6.76 & \text{se rechaza } H_0 \end{cases}$

La media muestral $\bar{x}_0 = 5{,}34 < 6{,}76$ se encuentra en la región de aceptación, por lo que se acepta la hipótesis nula H_0, concluyendo que tiene razón el Directivo, el tiempo medio de conexión es de 5 horas.

✓ Alternativamente, el problema se puede plantear desde la región de rechazo:

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

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

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: $\overbrace{t_{\alpha_p/2} = \frac{\bar{x} - \mu_o}{\frac{\sigma_x}{\sqrt{n-1}}}}^{\text{estadístico experimental}} \leq \overbrace{t_{\alpha, n-1}}^{\text{estadístico teórico}}$

En este caso, $t_{\alpha_p/2} = \frac{5{,}34 - 5}{7{,}24/\sqrt{29}} = 0{,}2537 < 1{,}311 = t_{0{,}10\,;\,29} \rightarrow$ Se acepta la hipótesis H_0

d) Potencia $= 1 - \beta = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Falsa}\right] = P\left[\text{Rechazar } H_0 \mid H_1 \text{ Cierta}\right]$

La hipótesis alternativa 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		

$\beta = P\left[\text{ETII}\right] = P\left[\text{Aceptar } H_0 \mid H_0 \text{ Falsa}\right] = P\left[\text{Rechazar } H_1 \mid H_1 \text{ Cierta}\right] = P\left[\bar{x} \leq k \mid H_1 \text{ Cierta}\right]$

En el muestreo de una población normal $N(\mu, \sigma)$ con varianza poblacional desconocida, cuando el tamaño muestral es grande $n = 300 > 30$, la distribución de la media muestral se aproxima a una distribución normal.

En esta línea, bajo la hipótesis alternativa $\bar{x} \sim N\left(\mu_1, \frac{\sigma_x}{\sqrt{n-1}}\right) \equiv N\left(\mu_1, \frac{7{,}24}{\sqrt{300-1}}\right)$

$$\beta = P\left[\bar{x} \leq k \mid H_1 \text{ Cierta}\right] = P\left[\bar{x} \leq 6{,}76 \mid N\left(\mu_1, \frac{7{,}24}{\sqrt{300-1}}\right)\right] = P\left[\frac{\bar{x} - \mu_1}{7{,}24/\sqrt{299}} \leq \frac{6{,}76 - \mu_1}{7{,}24/\sqrt{299}}\right] =$$

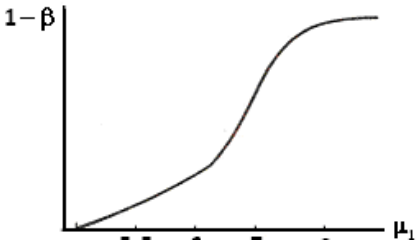
$$= P\left[z \leq \frac{6{,}76 - \mu_1}{7{,}24/\sqrt{299}}\right]$$

En consecuencia:

▭ $H_1: \mu_1 = 5{,}5 \rightarrow \beta = P\left[z \leq \frac{6{,}76 - 5{,}5}{7{,}24/\sqrt{299}}\right] = P(z \leq 3) = 0{,}9986$

- $H_1: \mu_1 = 6 \rightarrow \beta = P\left[z \le \dfrac{6,76-6}{7,24/\sqrt{299}}\right] = P(z \le 1,82) = 0,9656$
- $H_1: \mu_1 = 7 \rightarrow \beta = P\left[z \le \dfrac{6,76-7}{7,24/\sqrt{299}}\right] = P(z \le -0,57) = 0,2843$
- $H_1: \mu_1 = 8 \rightarrow \beta = P\left[z \le \dfrac{6,76-8}{7,24/\sqrt{299}}\right] = P(z \le -2,96) = 0,00154$

H_1: $\mu_1 > 5$	z	$1-\beta$
5,5	0,9986	0,0014
6	0,9656	0,0034
7	0,2843	0,7157
8	0,00154	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).

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" $X \sim P(\lambda = 2) \xrightarrow{\lambda > 10} N(2, \sqrt{2})$

Es un contraste unilateral, donde se plantean las hipótesis: $H_0: \lambda \ge 2 \quad H_1: \lambda < 2$

En el muestreo, bajo la hipótesis nula H_0 $\bar{x} \sim N\left[\lambda, \sqrt{\dfrac{\lambda}{n}}\right] = N\left[2, \sqrt{\dfrac{2}{25}}\right] \equiv N(2, 0,2828)$

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

Regla de decisión $\begin{cases} \text{Si } \bar{x} \ge k \text{ se acepta } H \\ \text{Si } \bar{x} < k \text{ se rechaza } H_0 \end{cases}$

Se determina el valor de k considerando el nivel de significación $\alpha = 0,01$:

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}\right] = P\left[\bar{x} < k \ / \ \lambda = 2\right] = P\left[\frac{\bar{x}-2}{0,2828} < \frac{k-2}{0,2828}\right] = P\left[z < \frac{k-2}{0,2828}\right] = 0,01$$

$$P\left[z \ge \frac{2-k}{0,2828}\right] = 0,01 \rightarrow \frac{2-k}{0,2828} = 2,32 \rightarrow k = 1,3439$$

Se advierte que $1,5 > 1,3439$, cae en la región de aceptación, con lo cual se acepta la hipótesis nula. Es decir, con 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: $\dfrac{2-k}{0,2828} = 1,04 = z_{0,15} \quad \rightarrow \quad k = 1,71$

Como $1,5 < 1,71$ la media muestral observada cae en la región de rechazo, con lo que no se acepta la hipótesis nula, y se concluye que las mejoras son operativas.

14. 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 = ' Porcentaje de votos al candidato'

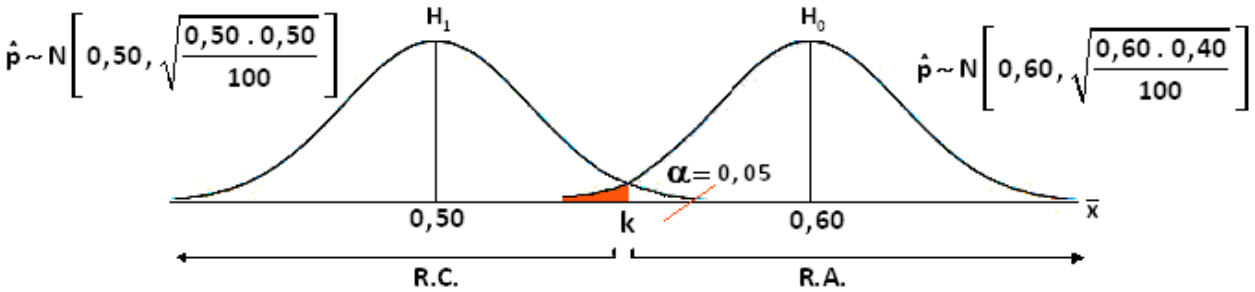
Al ser el tamaño suficientemente grande $n = 100$ y estar definido p como suma de variables aleatorias independientes, según una distribución de Bernouilli $B(1, p)$, se puede aproximar a la normal $N\left[np, \sqrt{npq}\right]$.

La distribución en el muestreo $\hat{p} = \dfrac{1}{100}\sum_{i=1}^{100} x_i$ siendo $x_i = \begin{cases} 1 & \text{vota Si} \\ 0 & \text{vota No} \end{cases}$ $\quad \hat{p} \sim N\left[p, \sqrt{\dfrac{pq}{n}}\right]$

Contraste unilateral a la izquierda (cola a la izquierda): $H_0: p = 0,60 \quad H_1: p = 0,50$

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

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



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

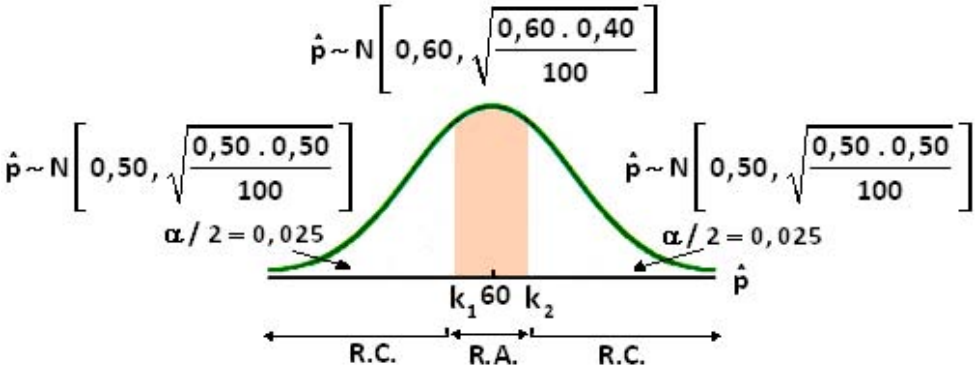
$$\alpha = P[ET\,I] = P[\text{Rechazar } H_0 \ / \ H_0 \text{ Cierta}] = P\left[\hat{p} < k \mid N(0,60, 0,049)\right] = P\left[\frac{\hat{p} - 0,60}{0,049} < \frac{k - 0,60}{0,049}\right] =$$

$$=P\left[\frac{\hat{p}-0,60}{0,049}\geq\frac{0,60-k}{0,049}\right]=P\left[z\geq\frac{0,60-k}{0,049}\right]=0,05\ \rightarrow\ \frac{0,60-k}{0,049}=1,645\ \rightarrow\ k=0,5193$$

La proporción muestral observada $\hat{p}=\overline{x}=52/100=0,52>0,5193$ se encuentra dentro de la región de aceptación, es decir, se acepta la hipótesis nula con un nivel de confianza del 95%.

b) En este caso se trata de un contraste bilateral (dos colas): $H_0: p=0,60 \quad H_1: p\neq 0,60$

Regla de decisión: $\begin{cases}|\hat{p}|\leq k & \text{Se acepta } H_0 \text{ (R.A.)}\\ |\hat{p}|>k & \text{Se rechaza } H_0 \text{ (R.C.)}\end{cases} \rightarrow \begin{cases}k_1\leq\hat{p}\leq k_2 & \text{Se acepta } H_0\\ \hat{p}<k_1\cup\hat{p}>k_2 & \text{Se rechaza } H_0\end{cases}$



Los valores críticos k se determinan por el nivel de significación $\alpha/2=0,05/2=0,025$:

$$\alpha=P[\text{ET I}]=P[\text{Rechazar } H_0\ /\ H_0 \text{ Cierta}]=P[|\hat{p}|>k\ |\ H_0: p=0,6]=P\left[\left|\frac{\hat{p}-0,60}{0,049}\right|>\frac{k-0,60}{0,049}\right]=$$

$$=P\left[|z|>\frac{k-0,60}{0,049}\right]=P\left[z<\frac{k_1-0,60}{0,049}\right]+P\left[z>\frac{k_2-0,60}{0,049}\right]=0,025+0,025=0,05$$

$$P\left[z<\frac{k_1-0,60}{0,049}\right]=P\left[z>\frac{-k_1+0,60}{0,049}\right]=0,025\ \rightarrow\ \frac{-k_1+0,60}{0,049}=1,96\ \rightarrow\ k_1=0,504$$

$$P\left[z>\frac{k_2-0,60}{0,049}\right]=0,025\ \rightarrow\ \frac{k_2-0,60}{0,049}=1,96\ \rightarrow\ k_2=0,696$$

La proporción muestral observada $\hat{p}=\overline{x}=52/100=0,52$ se encuentra dentro de la región de aceptación $0,504<\hat{p}=0,52<0,696$, concluyendo que se acepta la hipótesis nula.

✓ Otra forma de enfocar el problema consistiría en aceptar la hipótesis H_0 cuando el estadístico de contraste fuera menor o igual que el estadístico teórico, es decir:

Se acepta H_0 cuando se verifica: $\overbrace{z=\frac{|\hat{p}-p_0|}{\sqrt{\frac{\hat{p}.\hat{q}}{n}}}}^{\text{estadístico contraste}}\leq\overbrace{z_{\alpha/2}}^{\text{estadístico teórico}}$

Siendo $z=\frac{|0,52-0,60|}{\sqrt{\frac{0,52.0,48}{100}}}=1,6028<1,96=z_{0,025}$ se acepta la hipótesis nula.

✓ Se puede elaborar un intervalo de confianza para el parámetro poblacional p de una distribución binomial y observar si el valor $p = 0,6$ de la hipótesis nula se encuentra dentro, con lo que se aceptaría la hipótesis.

$$I.C\,(p) = \left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,.\,\hat{q}}{n}}\right] = \left[0,52 \pm 1,96 \cdot \sqrt{\frac{0,52\,.\,0,48}{100}}\right] = \left[0,4221\,,\ 0,6179\right]$$

$0,6 \in \left[0,4221\,,\ 0,6179\right] \ \rightarrow$ Se acepta H_0

c) Potencia $= 1 - \beta = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Falsa}\right] = P\left[\text{Rechazar } H_0 \mid H_1 \text{ Cierta}\right]$

Regla de decisión: $\begin{cases} \hat{p} \geq 0,5193 & \text{Se acepta } H_0 \\ \hat{p} < 0,5193 & \text{Se rechaza } H_0 \end{cases}$

Con la hipótesis alternativa $H_1: p = 0,50 \qquad \hat{p} \sim N\left[0,5\,,\ \sqrt{\frac{0,50.0,50}{100}}\right] = N(0,5\,,\,0,05)$

$$\text{Potencia} = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}\right] = P\left[\hat{p} < 0,5193 \ / \ N(0,5\,,\,0,05)\right] = P\left[\frac{\hat{p} - 0,5}{0,05} < \frac{0,5193 - 0,5}{0,05}\right] =$$

$$= P(z < 0,386) = 1 - P(z \geq 0,386) = 1 - 0,34978 = 0,6502$$

Abscisas		Áreas					
0,38 – 0,39	$\longleftrightarrow$	0,3520 – 0,3483	$-0,01$	$\longleftrightarrow$	0,0037	$\rightarrow$	$x = 0,3483 + \frac{0,004\,.\,0,0037}{0,01} = 0,34978$
0,386 – 0,39	$\longleftrightarrow$	x – 0,3483	$-0,004$	$\longleftrightarrow$	x – 0,3483		

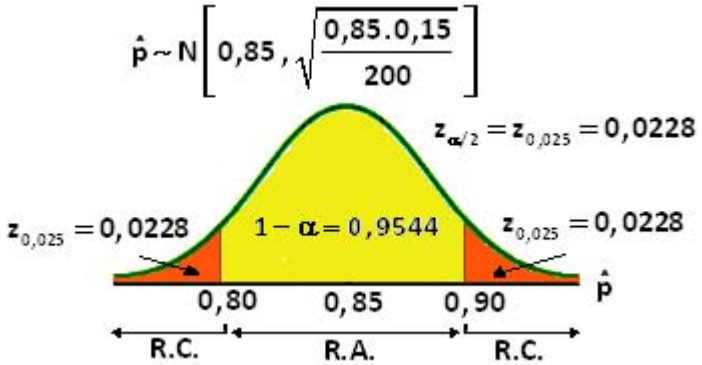
15. El dueño de los cines CINEFILÓN considera que, dado el aforo de la sala, una afluencia diaria a la misma del 85% sería óptima, en el sentido de que los clientes se sientan cómodos y para que a la vez no haya pérdidas económicas. Durante un período de tiempo, se analiza la afluencia a los cines, observándose que, en media, se ocupan 171 de las 200 butacas.
a) ¿Con qué confianza podrá afirmar el dueño de CINEFILÓN que la asistencia a sus cines es óptima?
b) ¿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:

a) La variable aleatoria X = ' Porcentaje asistencia diaria a los cines CINEFILÓN' $X \sim B(1\,,\,p)$.
En el contraste bilateral (dos colas) se establecen las hipótesis: $H_0: p_0 = 0,85 \qquad H_1: p_1 \neq 0,85$

- 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

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



Considerando que una desviación inferior al 5% de la cantidad óptima significa que $0{,}80 < \hat{p} < 0{,}90$

El nivel de confianza $(1-\alpha)$ se determina a partir de los valores críticos:

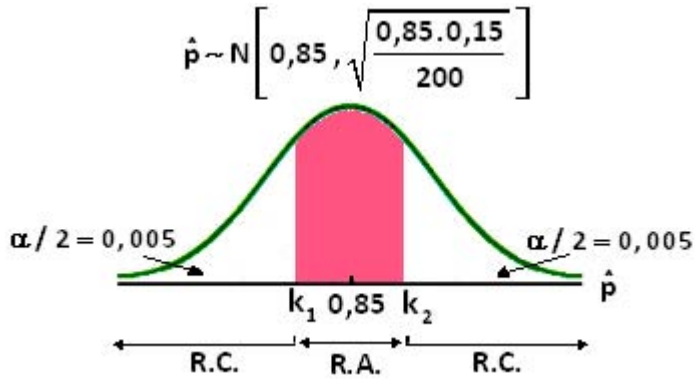
$$\alpha = P[\text{ET I}] = P[\text{Rechazar } H_0 \ / \ H_0 \text{ Cierta}] = P\left[\left(\frac{0{,}80-0{,}85}{0{,}025} > z\right) \cup \left(z > \frac{0{,}90-0{,}85}{0{,}025}\right)\right] =$$

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

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

b) Sí el nivel de confianza $(1-\alpha) = 99\%$ $\rightarrow$ $\alpha = 1\%$ mayor exigencia

Regla de decisión: $\begin{cases} |\hat{p}| \le k & \text{Se acepta } H_0 \ \text{(R.A.)} \\ |\hat{p}| > k & \text{Se rechaza } H_0 \ \text{(R.C.)} \end{cases} \rightarrow \begin{cases} k_1 \le \hat{p} \le k_2 & \text{Se acepta } H_0 \\ \hat{p} < k_1 \text{ o } \hat{p} > k_2 & \text{Se rechaza } H_0 \end{cases}$



Los valores críticos k se determina a partir del nivel de significación $\alpha = 0{,}01$

$$\alpha = P[\text{ET I}] = P[\text{Rechazar } H_0 \ / \ H_0 \text{ Cierta}] = P[|\hat{p}| < k \mid N(0{,}85\ ,\ 0{,}025)] = P\left[\left|\frac{\hat{p}-0{,}85}{0{,}025}\right| < \frac{k-0{,}85}{0{,}025}\right] =$$

$$= P\left[|z| < \frac{k-0{,}85}{0{,}025}\right] = P\left[|z| \ge \frac{0{,}85-k}{0{,}025}\right] = P\left[\left(z \ge \frac{0{,}85-k_1}{0{,}025}\right) \cup \left(z \ge -\frac{0{,}85-k_2}{0{,}025}\right)\right] = 0{,}005 + 0{,}005 = 0{,}01$$

$$\frac{0{,}85-k_1}{0{,}025} = 2{,}57 = z_{0{,}005} \quad \rightarrow \quad k_1 = 0{,}78575$$

$$-\frac{0{,}85-k_2}{0{,}025} = 2{,}57 = z_{0{,}005} \quad \rightarrow \quad k_2 = 0{,}91425$$

O bien $\left|\frac{0{,}85-k}{0{,}025}\right| = 2{,}57 = z_{0{,}005} \quad \rightarrow \quad \begin{array}{lll} \frac{0{,}85-k_1}{0{,}025} = 2{,}57 & \rightarrow & k_1 = 0{,}78575 \\ \frac{0{,}85-k_2}{0{,}025} = -2{,}57 & \rightarrow & k_2 = 0{,}91425 \end{array}$

Comparando el estadístico muestral (evidencia empírica) $\hat{p} = \frac{171}{200} = 0{,}855$ se observa que cae dentro de la región de aceptación $0{,}78575 < \hat{p} = 0{,}855 < 0{,}91425$ concluyendo que se acepta la hipótesis nula siendo óptima la afluencia a los cines CINEFILÓN.

16. 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 σ^2: $H_0: \sigma^2 = 3$ $H_1: \sigma^2 \neq 3$

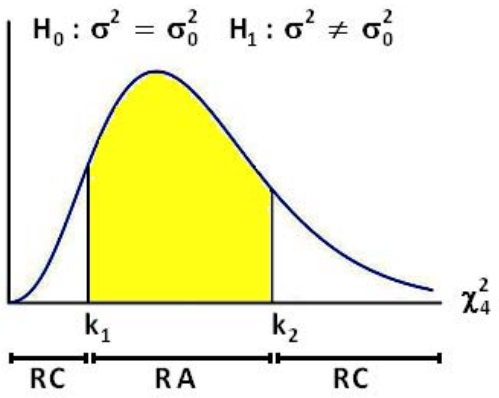
Como la hipótesis alternativa es $\sigma^2 \neq 3$ en la decisión deberán ser válidos los valores de σ^2 tanto mayores o menores que 3, por lo que el contraste debe ser bilateral o de dos colas.
Bajo la hipótesis nula, las variables aleatorias independientes $X_i \sim N\left(10, \sqrt{3}\right)$ con lo que

$$\frac{X_i - 10}{\sqrt{3}} \sim N(0,\ 1) \ \rightarrow \ \sum_{i=1}^{4} \left(\frac{x_i - 10}{\sqrt{3}} \right)^2 = \chi_4^2$$

Adviértase que la suma de 4 variables aleatorias $N(0, 1)$ independientes entre sí es una distribución Chi-cuadrado con 4 grados de libertad.

Regla decisión:

$$\left| \chi_4^2 = \sum_{i=1}^{4} \left(\frac{x_i - 10}{\sqrt{3}} \right)^2 \right| \leq k \rightarrow \text{Se acepta } H_0 \qquad \left| \chi_4^2 = \sum_{i=1}^{4} \left(\frac{x_i - 10}{\sqrt{3}} \right)^2 \right| > k \rightarrow \text{Se rechaza } H_0$$



Los valores críticos k se determina a partir del nivel de significación $\alpha = 0{,}05$

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

$$= P\left[(\chi_4^2 < k_1) \cup (\chi_4^2 > k_2) \right] = P\left[\chi_4^2 < k_1 \right] + P\left[\chi_4^2 > k_2 \right] = \alpha_1 + \alpha_2 = 0{,}025 + 0{,}025 = 0{,}05$$

$$P\left[\chi_4^2 < k_1 \right] = 1 - P\left[\chi_4^2 > k_1 \right] = 1 - 0{,}025 = 0{,}975 \ \rightarrow \ k_1 = \chi^2_{0{,}975\,,\,4} = 0{,}484$$

$$P\left[\chi_4^2 > k_2 \right] = 0{,}025 \ \rightarrow \ k_2 = \chi^2_{0{,}025\,,\,4} = 11{,}143$$

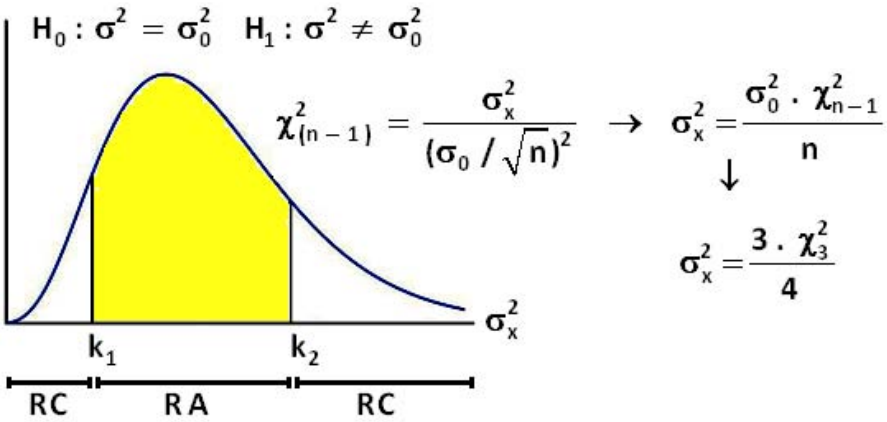
Región de aceptación: $\left[0{,}484\ ,\ 11{,}143\right]$

La varianza muestral observada $\sigma_x^2 = \dfrac{\sum_{i=1}^{4}(x_i - \overline{x})^2}{4} = \dfrac{\sum_{i=1}^{4}(x_i - 11)^2}{4} = \dfrac{20}{4} = 5$ se encuentra dentro de la región de aceptación $0{,}484 < \sigma_x^2 = 5 < 11{,}143$, aceptando la hipótesis nula con varianza poblacional 3.

✓ Se puede plantear que bajo la hipótesis nula la distribución en el muestreo:

$$\chi^2_{(n-1)} = \frac{\text{Varianza muestral}}{\text{Varianza teórica}} = \frac{\sigma_x^2}{(\sigma/\sqrt{n})^2} \rightarrow \sigma_x^2 = \frac{\sigma^2 \cdot \chi^2_{n-1}}{n}$$

Regla de decisión: $\begin{cases} \left|\sigma_x^2\right| \le k & \text{se acepta } H_0 \\ \left|\sigma_x^2\right| > k & \text{se rechaza } H_0 \end{cases} \rightarrow \begin{cases} k_1 \le \sigma_x^2 \le k_2 & \text{se acepta } H_0 \\ \left(\sigma_x^2 < k_1\right) \cup \left(\sigma_x^2 > k_2\right) & \text{se rechaza } H_0 \end{cases}$



$$\alpha = P\left[\text{Rechazar } H_0 \,/\, H_0 \text{ Cierta}\right] = P\left[\left(\sigma_x^2 < k_1\right) \cup \left(\sigma_x^2 > k_2\right) \,\middle|\, \sigma_x^2 = \frac{3}{4}\chi_3^2\right] =$$

$$= P\left[\frac{3}{4}\chi_3^2 < k_1\right] + P\left[\frac{3}{4}\chi_3^2 > k_2\right] = \alpha_1 + \alpha_2 = 0{,}025 + 0{,}025 = 0{,}05$$

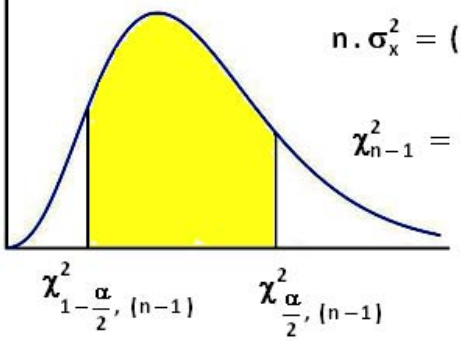
$$P\left[\frac{3}{4}\chi_3^2 < k_1\right] = P\left[\chi_3^2 > -\frac{4}{3}k_1\right] = 0{,}025 \rightarrow -\frac{4}{3}k_1 = 9{,}384 \rightarrow k_1 = -7{,}038$$

$$P\left[\frac{3}{4}\chi_3^2 > k_2\right] = P\left[\chi_3^2 > \frac{4}{3}k_2\right] = 0{,}025 \rightarrow \frac{4}{3}k_2 = 9{,}384 \rightarrow k_2 = 7{,}038$$

La varianza muestral observada $\sigma_x^2 = 5$ se encuentra dentro de la región de aceptación, aceptando la hipótesis nula.

✓ Se podría haber elaborado un intervalo de confianza para la varianza σ^2 de una población normal, verificando si se acepta la hipótesis nula al encontrarse la varianza $\sigma^2 = 3$ cubierta en el intervalo.

$H_0 : \sigma^2 = \sigma_0^2 \quad H_1 : \sigma^2 \neq \sigma_0^2$



$$n.\sigma_x^2 = (n-1).s_x^2$$

$$\chi^2_{n-1} = \frac{n.\sigma_x^2}{\sigma_0^2} = \frac{(n-1).s_x^2}{\sigma_0^2}$$

Se acepta H_0 sí $\chi^2_{n-1} = \dfrac{(n-1).s_x^2}{\sigma_0^2} \in \left[\chi^2_{1-\frac{\alpha}{2},\,(n-1)},\ \chi^2_{\frac{\alpha}{2},\,(n-1)}\right]$

$$\Updownarrow$$

$$I(\sigma_0^2) = \left[\frac{(n-1).s_x^2}{\chi^2_{\frac{\alpha}{2},\,(n-1)}},\ \frac{(n-1).s_x^2}{\chi^2_{1-\frac{\alpha}{2},\,(n-1)}}\right]$$

$$I.C\,(\sigma_0^2) = \left[\frac{(n-1).s_x^2}{\chi^2_{\alpha/2,\,(n-1)}},\ \frac{(n-1).s_x^2}{\chi^2_{1-\alpha/2,\,(n-1)}}\right]$$

$$(n-1).s_x^2 = n.\sigma_x^2 = 4.5 = 20$$

$$I.C\,(\sigma_0^2) = \left[\frac{20}{9{,}348},\ \frac{20}{0{,}216}\right] = [2{,}139,\ 92{,}592]$$

Como $3 \in [2{,}139,\ 92{,}592] \rightarrow$ Se acepta H_0

17. En una población $N(\mu, 5)$ se formula 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 con los datos recogidos: 16 12 15 16 20 25 14 18 17 22

a) Contrastar la hipótesis nula
b) Realizar el contraste con el p_valor
c) Realizar el contraste con métodos afines al p_ valor

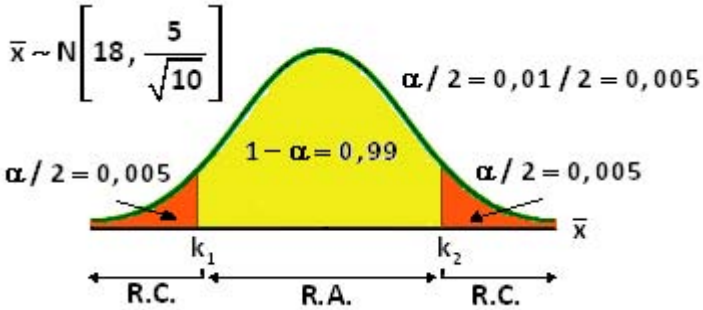
Solución:

a) Se establecen las hipótesis: $H_0: \mu = 18 \quad H_1: \mu \neq 18$

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.

La muestra bajo la hipótesis nula sigue una distribución $N\left[18, \dfrac{5}{\sqrt{10}}\right] = N(18, 1{,}58)$

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



Los valores críticos k se calculan mediante el nivel de significación $\alpha = 0,01$

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}\right] = P\left[|\bar{x}| > k \mid N(18, 1,58)\right] = P\left[(\bar{x} < k_1) \cup (\bar{x} > k_2\right] =$$
$$= P(\bar{x} < k_1) + P(\bar{x} > k_2) = 0,005 + 0,005 = 0,01$$

$$P(\bar{x} < k_1) = P\left[\frac{\bar{x}-18}{1,58} < \frac{k_1-18}{1,58}\right] = P\left[z < \frac{k_1-18}{1,58}\right] = P\left[z > -\frac{k_1-18}{1,58}\right] = 0,005$$

$$-\frac{k_1-18}{1,58} = 2,575 \quad \rightarrow \quad k_1 = 13,9315$$

$$P(\bar{x} > k_2) = P\left[\frac{\bar{x}-18}{1,58} > \frac{k_2-18}{1,58}\right] = P\left[z > \frac{k_2-18}{1,58}\right] = 0,005$$

$$\frac{k_2-18}{1,58} = 2,575 \quad \mapsto \quad k_2 = 22,0685$$

En consecuencia, la región de aceptación: $13,9315 < \bar{x} < 22,0685$

La media muestral observada (evidencia empírica) $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 17,5$ se encuentra dentro de la región de aceptación, afirmando con un nivel de confianza del 99%, que se verifica la hipótesis nula.

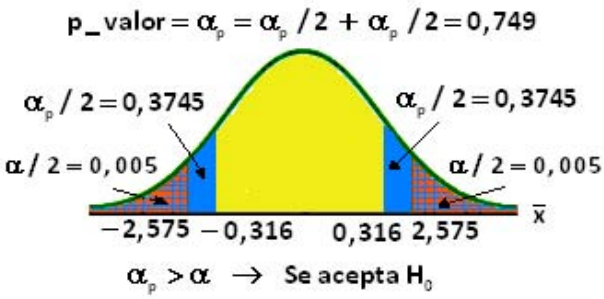
b) El p_valor (α_p) es el nivel de significación más pequeño posible que puede escogerse para el que todavía se aceptaría la hipótesis alternativa con las observaciones actuales.

$$\alpha_p = p_valor = P\left[\text{Rechazar el estadístico muestral observado } / \ H_0 \text{ es Cierta}\right] \quad \boxed{\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}}}$$

$$\alpha_p = P\left[|\bar{x}| > 17,5 \mid N(18, 1,58)\right] = P\left[\left|\frac{\bar{x}-18}{1,58}\right| > \frac{17,5-18}{1,58}\right] = P\left[|z| > -0,316\right] =$$
$$= P\left[|z| > 0,316\right] = P\left[z < -0,316\right] + P\left[z > 0,316\right] = 2P\left[z > 0,316\right] = 2(0,3745) = 0,749$$

Tabla N(0,1) $\quad \frac{0,31-0,32}{0,3783-0,3745} = \frac{0,316-0,32}{x-0,3745} \quad \rightarrow \quad x = 0,3745$

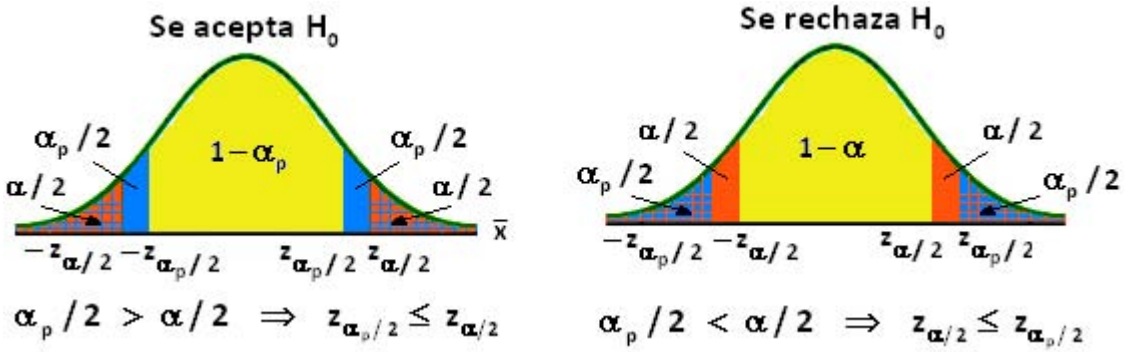
Como $\alpha_p = 0,749 > 0,01 = \alpha$ se acepta la hipótesis nula.



c) Métodos consecuencia del p_valor:

▭ El p_valor $= \alpha_p$ es la abscisa de la distribución normal $N(0,1)$ que origina el estadístico de contraste (estadístico empírico u observado), de modo que en un contraste bilateral (o de dos colas) para la media poblacional se acepta la hipótesis nula cuando $z_{\alpha_p/2} \leq z_{\alpha/2}$

$$z_{\alpha_p/2} = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \quad \rightarrow \quad \alpha_p/2 > \alpha/2 \quad \Rightarrow \quad z_{\alpha_p/2} \leq z_{\alpha/2}$$



$$z_{\alpha_p/2} = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} = \frac{|17{,}5 - 18|}{1{,}58} = 0{,}316 < 2{,}575 = z_{0{,}005} \quad \rightarrow \quad \text{Se acepta } H_0$$

Señalar que $\alpha_p/2 = P\left[z_{\alpha_p/2} > 0{,}316\right] = 0{,}3745 \quad \rightarrow \quad \alpha_p = 0{,}749$

▭ Región de Rechazo: $R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right\}$

$$R = \left\{ |17{,}5 - 18| > 2{,}575 \cdot \frac{5}{\sqrt{10}} \right\} = \left\{ 0{,}5 \not> 4{,}07 \right\} \quad \rightarrow$$

No se verifica la región de rechazo, aceptando la hipótesis nula $\mu = 18$

▭ Observando si la media de la hipótesis nula $\mu = 18$ se encuentra dentro del intervalo de confianza para la media poblacional μ: $\text{I.C}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$

$$\text{I.C}(\mu) = \left[17{,}5 \pm 2{,}575 \frac{5}{\sqrt{10}} \right] = \left[13{,}43 \,,\, 21{,}57 \right] \quad \rightarrow \quad \mu = 18 \in \left[13{,}43 \,,\, 21{,}57 \right]$$

Se acepta la hipótesis nula.

18. Una empresa de neumáticos afirma que una nueva gama sigue una ley normal y 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 estándar 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?

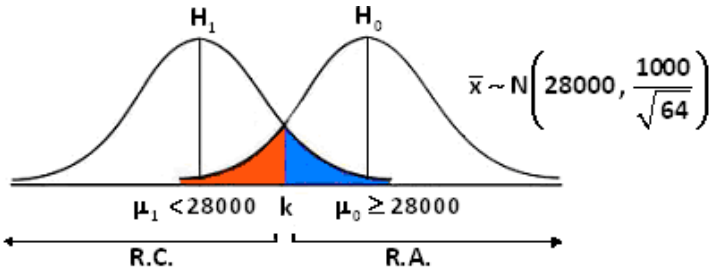
Solución:

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

Hipótesis sobre la media poblacional μ con varianza σ^2 desconocida:
$H_0: \mu_0 \geq 28000 \qquad H_1: \mu_1 < 28000$

Se trata de un contraste unilateral por la izquierda (cola por la izquierda)

Regla decisión $\begin{cases} \bar{x} \geq k & \text{Se acepta } H_0 \\ \bar{x} < k & \text{Se rechaza } H_0 \end{cases}$



En el muestreo bajo la hipótesis nula: $\bar{x} \sim N\left(28000, \frac{1000}{\sqrt{64}}\right) \equiv N(28000, 125)$

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

$$\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ Cierta}] = P[\bar{x} < k \mid N(28000,125)] = P\left(\frac{\bar{x}-28000}{125} < \frac{k-28000}{125}\right) =$$

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

$$\frac{28000-k}{125} = 1,645 \quad \rightarrow \quad k = 28000 - 125 \times 1,645 = 27794,375$$

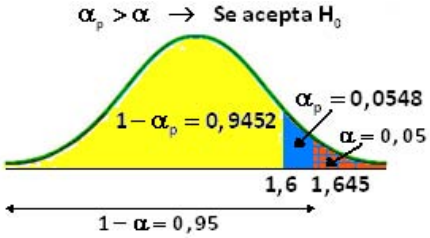
Siendo $\bar{x} = 27800 > 27794,375$ se acepta la hipótesis nula, por tanto, es buena la afirmación de la empresa con un nivel de confianza del 95%.

b) El p_valor (α_p) es el nivel de significación más pequeño posible que puede escogerse para el que todavía se aceptaría la hipótesis alternativa con las observaciones actuales.
$\alpha_p = p_valor = P[\text{Rechazar el estadístico muestral observado / } H_0 \text{ es Cierta}]$

$$\alpha_p = P[\bar{x} < 27800 \mid H_0 \text{ Cierta}] = P[\bar{x} < 27800 \mid N(28000,125)] = P\left[\frac{\bar{x}-28000}{125} < \frac{27800-28000}{125}\right] =$$

$$= P[z < -1,6] = P[z > 1,6] = 0,0548$$

Siendo $\alpha_p = 0{,}0548 > \alpha = 0{,}05$ se acepta la hipótesis nula. Es decir, con una fiabilidad del 95% se acepta que la duración media de los neumáticos es de 28.000 km.



19. 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", $X \sim N(16, \sigma)$

Hipótesis sobre la media poblacional μ con varianza σ^2 desconocida: H_0: $\mu_0 = 16$ $\quad H_1$: $\mu_1 \neq 16$

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 el muestreo de una población normal $N(\mu, \sigma)$ con varianza poblacional desconocida, con un tamaño muestral pequeño $n = 16 \leq 30$, la media muestral bajo la hipótesis nula sigue una distribución $\overline{x} \sim t_{n-1}\left(\mu, \frac{s_x}{\sqrt{n}}\right) \equiv t_{24}\left(16, \frac{0{,}8}{\sqrt{25}}\right) \equiv t_{24}(16, 0{,}16)$

Para realizar el contraste se compara el estadístico observado (abscisa del p_valor) con el estadistico teórico, aceptándose la hipótesis nula cuando $t_{\alpha_p/2} = \frac{|\overline{x} - \mu_0|}{s_x / \sqrt{n}} \leq t_{\alpha/2, n-1}$

$$t_{\alpha_p/2} = \frac{|\overline{x} - \mu_0|}{s_x / \sqrt{n}} = \frac{|16{,}32 - 16|}{0{,}16} = 2 \leq 2{,}064 = t_{0{,}025, 24} \quad \rightarrow \quad \text{Se acepta } H_0$$

✓ Señalar que una forma análoga de proceder sería aceptar la hipótesis nula si no verifica la región de rechazo: $R = \left\{ |\overline{x} - \mu_0| > t_{\alpha/2, n-1} \frac{s_x}{\sqrt{n}} \right\}$

Se acepta H_0 porque no se verifica $R = \{ |16{,}32 - 16| > 2{,}064 \,.\, 0{,}16 \} \equiv \{0{,}32 \not> 0{,}33\}$ afirmando que el peso de llenado son 16 gramos, con un nivel de significación de 0,05.

b) El p_valor (α_p) es el nivel de significación más pequeño posible que puede escogerse para el que todavía se aceptaría la hipótesis alternativa con las observaciones actuales.

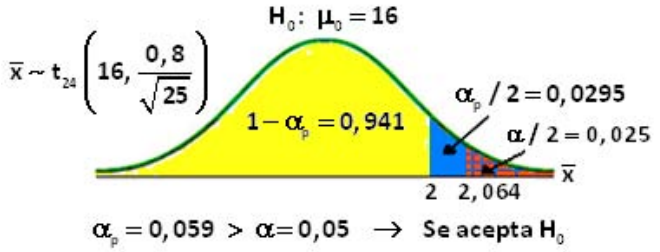
$$t_{\alpha_p/2} = \frac{|\overline{x} - \mu_0|}{s_x / \sqrt{n}} = \frac{|16{,}32 - 16|}{0{,}16} = 2$$

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

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:

$$\begin{array}{ccc|ccc|ccc} 0{,}05 & \alpha_p/2 & 0{,}025 & 0{,}05-0{,}025 & \longleftrightarrow & 1{,}711-2{,}064 & 0{,}025 & \longleftrightarrow & -0{,}353 \\ 1{,}711 & 2 & 2{,}064 & \alpha_p/2-0{,}025 & \longleftrightarrow & 2-2{,}064 & \alpha_p/2-0{,}025 & \longleftrightarrow & -0{,}064 \end{array}$$

$$(\alpha_p/2-0{,}025) \times 0{,}353 = 0{,}025 \times 0{,}064 \;\rightarrow\; \alpha_p/2 = 0{,}025 + \frac{0{,}025 \times 0{,}064}{0{,}353} = 0{,}0295 \;\rightarrow\; \alpha_p = 0{,}059$$



Siendo $\alpha_p = 0{,}059 > \alpha = 0{,}05$ se acepta la hipótesis nula.

20. 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.
¿Están fundamentadas 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 $Y \sim N(\mu_y, 36)$

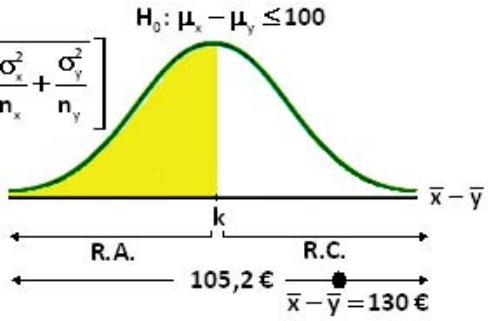
En las muestras se obtuvieron los resultados: $\begin{cases} \bar{x} = 1.500\,€ & n_x = 500 \\ \bar{y} = 1.370\,€ & n_y = 700 \end{cases}$

En el contraste se establecen las hipótesis: $H_0: \mu_x - \mu_y \leq 100$ $\quad H_1: \mu_x - \mu_y > 100$

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

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

$$(\bar{x}-\bar{y}) \sim N\left[(\mu_x-\mu_y),\ \sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right]$$



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

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

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

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

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

Se advierte que $(\bar{x}-\bar{y}) = 130 > 105{,}2$, 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.

21. 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	

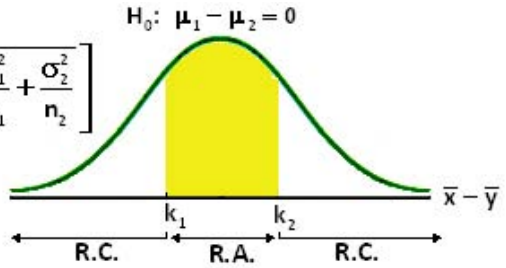
Solución:

En el contraste bilateral se establecen las hipótesis: H_0: $\mu_1-\mu_2=0$ $\quad H_1$: $\mu_1-\mu_2 \neq 0$

Regla de decisión $\begin{cases} |\bar{x}-\bar{y}| \le k & \text{se acepta } H_0 \\ |\bar{x}-\bar{y}| > k & \text{se rechaza } H_0 \end{cases}$

Para analizar el contraste se calculan las medias muestrales $\bar{x}$ e $\bar{y}$

$$(\bar{x}-\bar{y}) \sim N\left[0,\ \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$



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

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

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

Los valores críticos k se determinan bajo el nivel de significación $\alpha = 0,0472$, considerando que se trata de un contraste bilateral (o de dos colas):

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\,|\bar{x}-\bar{y}| > k \;/\; N(0\,,\,2,59)\,\right] = P\left[\left|\frac{(\bar{x}-\bar{y})-0}{2,59}\right| > \frac{k-0}{2,59}\right] =$$

$$= P\left[\,|z| > \frac{k-0}{2,59}\right] = P\left[\left(z \le \frac{k_1}{2,59}\right) \cup \left(z \ge \frac{k_2}{2,59}\right)\right] = P\left[\left(z \ge \frac{-k_1}{2,59}\right) \cup \left(z \ge \frac{k_2}{2,59}\right)\right] =$$

$$= 0,0236 + 0,0236 = 0,0472$$

$$\frac{-k_1}{2,59} = 1,995 = z_{0,\,0236} \quad \rightarrow \quad k_1 = -\,2,59 \times 1,995 = -5,17$$

$$\frac{k_2}{2,59} = 1,995 = z_{0,\,0236} \quad \rightarrow \quad k_2 = 2,59 \times 1,995 = 5,17$$

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

La evidencia empírica $|\bar{x}-\bar{y}| = |12,25-13,64| = 1,39$ 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%.

✓ Se puede resolver mediante el p_valor:

$$\alpha_p = p_valor = P\left[\text{Rechazar el estadístico muestral observado } / \; H_0 \text{ es Cierta}\right] =$$

$$= P\left[\,|\bar{x}-\bar{y}| > 1,39 \,\middle|\, N(0\,,\,2,59)\right] = P\left[\left|\frac{(\bar{x}-\bar{y})-0}{2,59}\right| > \frac{1,39-0}{2,59}\right] = P\left[\,|z| > \frac{1,39-0}{2,59}\right] =$$

$$= P\left[\,|z| > 0,5367\,\right] = P\left[\,z < -\,0,5367\,\right] + P\left[\,z > 0,5367\,\right] = 2\,.\,P\left[z > 0,5367\right] = 2\;.\,0,095 = 0,19$$

O bien aplicando consecuencias directas:

▭ Aceptando la hipótesis nula H_0 cuando el estadístico observado o de contraste es menor que el estadístico teórico:

$$z_{\alpha_p/2} = \frac{|\bar{x}-\bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \le z_{\alpha/2} = z_{0,\,0236} \quad \rightarrow \quad z_{\alpha_p/2} = \frac{1,39}{2,59} = 0,5367 \le 1,995 = z_{0,\,0236}$$

$$\alpha_p / 2 = P\left[z_{\alpha_p/2} \ge 0,5367\right] = 0,095 \quad \longrightarrow \quad \alpha_p = 0,19 > 0,0472 = \alpha \quad \text{Se acepta } H_0$$

▭ Aceptando la hipótesis nula H_0 cuando no se verifica la región de rechazo:

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

La región de rechazo de la hipótesis nula no se cumple $R = \left\{ 1,39 \not> 5,17 \right\}$, concluyendo que existe igualdad entre las medias poblacionales.

Verificando que el intervalo de confianza cubre el valor 0, en este caso existe igualdad de medias poblacionales.

$$I(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \longrightarrow I(\mu_1 - \mu_2) = \left[(12,25 - 13,64) \pm 1,995 \cdot \sqrt{\frac{4^2}{6} + \frac{4,5^2}{5}} \right]$$

$$I(\mu_1 - \mu_2) = \left[-1,39 \pm 5,17 \right] = \left[-6,56 \, , \, 3,78 \right] \longrightarrow 0 \in \left[-6,56 \, , \, 3,78 \right] \text{ Se acepta } H_0$$

Cálculo de $z_{0,0236}$:

Abscisas	Áreas			
1,98 – 1,99	0,0239 – 0,0233	0,01	0,006	$\rightarrow x = 1,99 + \dfrac{0,01 \, . \, 0,0003}{0,0006} = 1,995$
x – 1,99	0,0236 – 0,0233	x – 1,99	0,0003	

22. La directora del departamento de personal de una corporación está buscando empleados para un puesto en el extranjero Durante el proceso de selección, la administración le pregunta cómo va la incorporación de empleados, y ella contesta que sigue una ley normal y que la puntuación promedio en la prueba de aptitudes será de 90 puntos.
Cuando la administración revisa 19 de los resultados de la prueba, encuentra que la puntuación media es de 86,25 puntos con una desviación estándar de 11.
Con un nivel de confianza del 90%, ¿lleva razón la directora?. Calcula su p_valor.

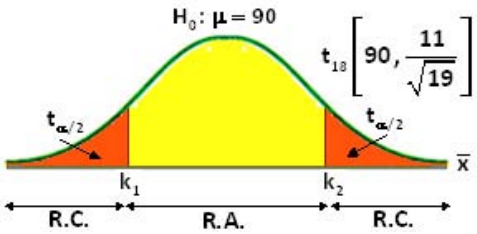
Solución:

Se establecen las hipótesis: $H_0: \mu = 90 \qquad H_1: \mu \neq 90$

En el muestreo de la población normal con varianza desconocida, con muestras pequeñas $n = 19 < 30$, la media muestral $\bar{x} \sim t_{18}\left(\mu, \dfrac{s_x}{\sqrt{19}} \right) \equiv \left(\mu, \dfrac{\sigma_x}{\sqrt{18}} \right)$

Como la hipótesis alternativa es $\mu \neq 90$ en la decisión deberán ser válidos valores de μ tanto mayores o menores que 90, por lo cual el contraste debe ser bilateral o de dos colas.

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



En el muestreo, bajo la hipótesis nula $\bar{x} \sim t_{18}\left[90, \dfrac{11}{\sqrt{19}} \right] \equiv t_{18}(90 \, , \, 2,524)$

Los valores críticos k se determinan con el nivel de significación $\alpha = 0,10$

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ Cierta} \right] = P\left[|\bar{x}| > k \mid t_{18}(90 \, , \, 2,524) \right] = P\left[(\bar{x} < k_1) \cup (\bar{x} > k_2 \right] =$$
$$= P(\bar{x} < k_1) + P(\bar{x} > k_2) = 0,05 + 0,05 = 0,10$$

$$P(\bar{x} < k_1) = P\left[\frac{\bar{x}-90}{2,524} < \frac{k_1-90}{2,524}\right] = P\left[t_{18} < \frac{k_1-90}{2,524}\right] = P\left[t_{18} \geq -\frac{k_1-90}{2,524}\right] = 0,05$$

$$-\frac{k_1-90}{2,524} = 1,734 = t_{0,05,18} \quad \rightarrow \quad k_1 = 85,62$$

$$P(\bar{x} > k_2) = P\left[\frac{\bar{x}-90}{2,524} > \frac{k_2-90}{2,524}\right] = P\left[t_{18} > \frac{k_2-90}{2,524}\right] = 0,05$$

$$\frac{k_2-90}{2,524} = 1,734 = t_{0,05,18} \quad \rightarrow \quad k_2 = 94,37$$

La media muestral observada $\bar{x} = 86,25$ se encuentra en la región de aceptación, con un nivel de significación de 0,10, se acepta la manifestación de la directora de la corporación.

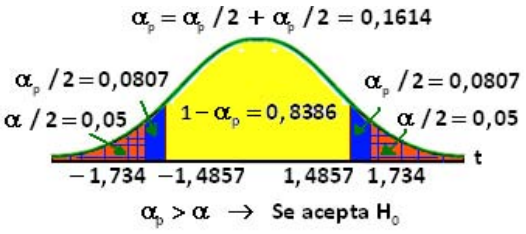
$$\alpha_p = p_valor = P[\text{Rechazar el estadístico muestral observado } / \ H_0 \text{ es Cierta}] =$$

$$= P\left[\,|\bar{x}| > 86,25 \,|\, t_{18}(90\ ,\ 2,524)\right] = P\left[\left|\frac{\bar{x}-90}{2,524}\right| > \frac{86,25-90}{2,524}\right] = P\left[\,|t_{18}| > -1,4857\right] =$$

$$= P\left[\,\|t_{18}\| > 2,6774\right] = P\left[\,t_{18} < -1,4857\right] + P\left[\,t_{18} > 1,4857\right] = 2P\left[\,t_{18} > 1,4857\right] = 2(0,0807) = 0,1614$$

Tabla t-Student: $\dfrac{1,330-1,734}{0,1-0,05} = \dfrac{1,4857-1,734}{x-0,05} \quad \rightarrow \quad x = 0,05 + \dfrac{0,05\,.\,0,2483}{0,404} = 0,0807$

Como $\alpha_p = 0,1614 > 0,10 = \alpha$ se acepta la hipótesis nula y en consecuencia es cierta la manifestación de la directora de la corporación.



O bien aplicando consecuencias directas:

▭ Aceptando la hipótesis nula H_0 cuando el estadístico observado o de contraste es menor que el estadístico teórico:

$$t_{\alpha_p/2} = \frac{|\bar{x}-\mu_0|}{s_x/\sqrt{n}} \leq t_{\alpha/2,n-1} = t_{0,05,18} \quad \rightarrow \quad t_{\alpha_p/2} = \frac{|86,25-90|}{2,524} = 1,4857 < 1,734 = t_{0,05,18} \quad \rightarrow$$

$\rightarrow$ Se acepta la hipótesis nula H_0

Adviértase que $\alpha_p/2 = P\left[t_{\alpha_p/2} > 1,4857\right] = 0,0807$

En la tabla de la t de Student: $P\left[t_{18} \geq 1,330\right] = 0,1$ y $P\left[t_{18} \geq 1,734\right] = 0,05$, interpolando se tiene:

0,10	$\alpha_p/2$	0,05	$0,10-0,05$	$\longleftrightarrow$	$1,330-1,734$	0,05	$\longleftrightarrow$	$-0,404$
1,330	1,4857	1,734	$\alpha_p/2-0,05$	$\longleftrightarrow$	$1,4857-1,734$	$\alpha_p/2-0,05$	$\longleftrightarrow$	$-0,2483$

$$(\alpha_p/2-0,05) \times 0,404 = 0,05 \times 0,2483 \quad \rightarrow \quad \alpha_p/2 = 0,05 + \frac{0,05 \times 0,2483}{0,404} = 0,0807 \rightarrow \quad \alpha_p = 0,1614$$

Alternativamente, el problema se puede plantear desde la región de rechazo:

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

Región de rechazo: $R=\left\{\ |86{,}25-90| > \overbrace{1{,}734}^{t_{0{,}05,18}}\ \frac{11}{\sqrt{19}}\right\}=\left[3{,}75 \not> 4{,}3758\right]$

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%.

Si el intervalo de confianza para la media poblacional μ cubre el valor $\mu=90$ se acepta la hipótesis nula: $I.C=\left[\bar{x}\ \pm\ t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}}\right]=\left[86{,}25\ \pm\ 1{,}734\frac{11}{\sqrt{19}}\right]=\left[81{,}873\ ,\ 90{,}627\right]$

$90\in\left[81{,}873\ ,\ 90{,}627\right]\ \rightarrow\ 81{,}873<90<90{,}627\ \rightarrow$ Se acepta H_0

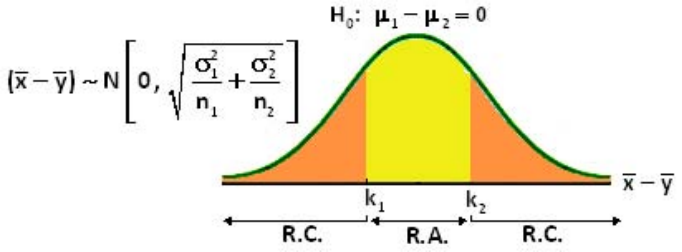
23. Una empresa ubicada en Madrid tiene dos conductores para trasladar a los empleados a Segovia. Los conductores deben anotar la duración de cada trayecto. En una muestra aleatoria simple de 50 partes de incidencias por conductor, el conductor A registra un tiempo medio de trayecto de 62,30 minutos con una desviación típica de 10,325 minutos, mientras que el conductor B tiene un tiempo medio de trayecto de 60,02 minutos con una desviación típica de 8,625 minutos. El tiempo medio empleado por el conductor A en el trayecto sigue una ley normal $N(\mu_1,9)$ y el empleado por el conductor B se distribuye según una ley $N(\mu_2,8)$.
Con un nivel de significación del 5%, se pide contrastar:
a) H_0: $\mu_1=\mu_2$ frente a H_1: $\mu_1\neq\mu_2$
b) H_0: $\mu_1-\mu_2\geq 2$ frente a H_1: $\mu_1-\mu_2<2$

Solución:

a) Para realizar el contraste bilateral planteado H_0: $\mu_1=\mu_2$ se recurre al test razón de verosimilitud, en este caso el lema de Neyman-Pearson no proporciona una región crítica óptima.

La regla de decisión que proporciona el test de razón de verosimilitud es:

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



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

$$\bar{x}\sim N\left(\mu_1,\frac{\sigma_1}{\sqrt{n_1}}\right)\quad \bar{y}\sim N\left(\mu_2,\frac{\sigma_2}{\sqrt{n_2}}\right)$$

Con lo cual, la diferencia de medias muestrales, bajo la hipótesis nula, se distribuye:

$$(\bar{x}-\bar{y}) \sim N\left[0, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] \equiv N\left[0, \sqrt{\frac{9^2}{50}+\frac{8^2}{50}}\right] = N[0, 1{,}7]$$

Los valores críticos k se determinan a partir del nivel de significación $\alpha = 0{,}05$

$$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ Cierta}) = P\left[\,|\bar{x}-\bar{y}| > k \;/\; N[0, 1{,}7]\right] = P\left[\left|\frac{\bar{x}-\bar{y}-0}{1{,}7}\right| > \frac{k-0}{1{,}7}\right] =$$

$$= P\left[|z| > \frac{k}{1{,}7}\right] = P\left[\left(z < \frac{k_1}{1{,}7}\right) \cup \left(z > \frac{k_2}{1{,}7}\right)\right] = P\left[z > -\frac{k_1}{1{,}7}\right] + P\left[z > \frac{k_2}{1{,}7}\right] = 0{,}025 + 0{,}025 = 0{,}05$$

Por simetría: $\frac{-k_1}{1{,}7} = 1{,}96 = z_{0,025} \;\rightarrow\; k_1 = -3{,}33 \qquad \frac{k_2}{1{,}7} = 1{,}96 = z_{0,025} \;\rightarrow\; k_2 = 3{,}33$

La región de aceptación: $-3{,}33 \le \bar{x}-\bar{y} \le 3{,}33$

La evidencia empírica $|\bar{x}-\bar{y}| = |62{,}30 - 60{,}02| = 2{,}28$ valor que se encuentra en la región de aceptación, por lo que se acepta la hipótesis nula de igualdad de medias.
Los dos conductores emplean en promedio el mismo tiempo en el trayecto Madrid – Segovia, con una fiabilidad del 95%.

▭ Se podía haber decidido con el p_valor o métodos aplicando sus consecuencias directas:

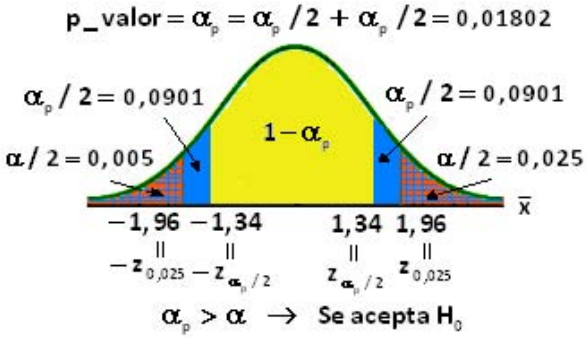
$$\alpha_p = p_valor = P[\text{Rechazar el estadístico muestral observado } / \; H_0 \text{ es Cierta}] =$$

$$= P\left[\,|\bar{x}-\bar{y}| > 2{,}28 \mid N[0, 1{,}7]\right] = P\left[\left|\frac{\bar{x}-\bar{y}-0}{1{,}7}\right| > \frac{2{,}28-0}{1{,}7}\right] = P[|z| > 1{,}34] =$$

$$= P[z < -1{,}34] + P[z > 1{,}34] = 2P[z > 1{,}34] = 2(0{,}0901) = 0{,}1802$$

Siendo $\alpha_p = 0{,}1802 > 0{,}05 = \alpha \;\rightarrow\;$ Se acepta H_0

El p_valor es el nivel de significación más pequeño posible que puede elegirse para el que se aceptaría la hipótesis alternativa.



▭ Aceptando la hipótesis nula H_0 cuando el estadístico observado o de contraste es menor que el estadístico teórico:

$$z_{\alpha_p/2} = \frac{|\bar{x}-\bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \le z_{\alpha/2} \;\rightarrow\; z_{\alpha_p/2} = \frac{|62{,}30-60{,}02|}{1{,}7} = 1{,}34 < 1{,}96 = z_{0,025} \;\rightarrow$$

$\rightarrow$ Se acepta la hipótesis nula H_0 $\Rightarrow$ Los dos conductores emplean en promedio el mismo tiempo en el trayecto Madrid – Segovia.

Adviértase que $\alpha_p / 2 = P\left[z_{\alpha_p/2} > 1{,}34\right] = 0{,}0901 \;\rightarrow\; \alpha_p = 0{,}1802$

Alternativamente, el problema se puede plantear desde la región de rechazo:

$$R=\left\{\left|\bar{x}-\bar{y}\right|>z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right\} \rightarrow R=\left\{\left|62,30-60,02\right|>1,96\sqrt{\frac{9^2}{50}+\frac{8^2}{50}}\right\}=\left\{2,28\not>3.332\right\}$$

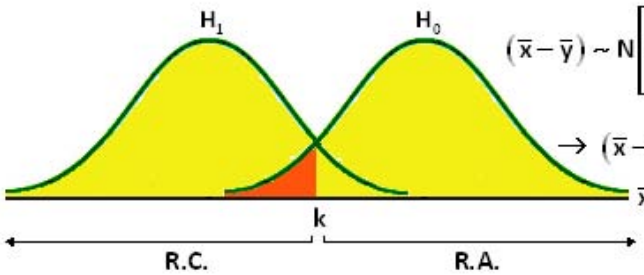
Al no verificarse la región de rechazo se acepta la hipótesis nula.

Si el intervalo de confianza para la diferencia de medias poblacionales $\mu_1-\mu_2$ cubre el valor 0 se acepta la hipótesis nula: $I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\ \pm\ z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$

$$I_{0,95}(\mu_1-\mu_2)=\left[2,28\ \pm\ 1,96\times 1,7\right]=\left[-1,05\ ,\ 5,61\right]$$

Como el intervalo cubre el cero $-1,05<0<5,61$ no existe diferencia significativa entre el tiempo empleado por los dos conductores en el trayecto Madrid – Segovia.

b) Hipótesis nula planteada H_0: $\mu_1-\mu_2-2\geq 0$. Se trata de un contraste unilateral con cola a la izquierda H_1: $\mu_1-\mu_2-2<0$



$$(\bar{x}-\bar{y})\sim N\left[0,\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] \rightarrow$$

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

$\bar{x}-\bar{y}-k$

Bajo la hipótesis nula $(\bar{x}-\bar{y})\sim N\left[0,1,7\right] \rightarrow (\bar{x}-\bar{y}-2)\sim N\left[-2,1,7\right]$

$$\text{Regla decisión}\begin{cases}\bar{x}-\bar{y}-2\ \geq\ k & \text{se acepta } H_0 \text{ (R.A.)}\\ \bar{x}-\bar{y}-2\ <\ k & \text{se rechaza } H_0 \text{ (R.C.)}\end{cases} \rightarrow \begin{cases}\bar{x}-\bar{y}\ \geq\ 2+k & \text{se acepta } H_0 \text{ (R.A.)}\\ \bar{x}-\bar{y}\ <\ 2+k & \text{se rechaza } H_0 \text{ (R.C.)}\end{cases}$$

$$\alpha=P(\text{Rechazar } H_0 \mid H_0 \text{ Cierta})=P\left[\bar{x}-\bar{y}-2<k\ /\ N\left[0,1,7\right]\right]=P\left[\bar{x}-\bar{y}<k+2\ /\ N\left[0,1,7\right]\right]=$$

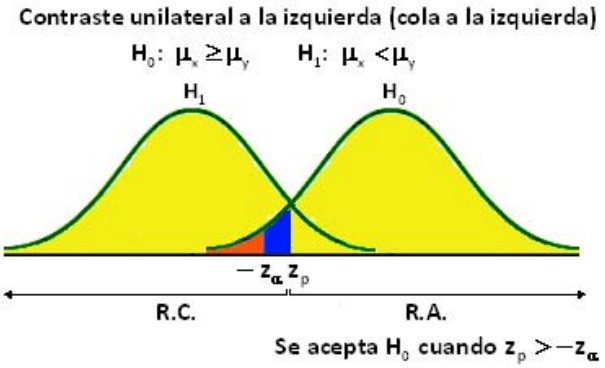
$$=P\left[\frac{\bar{x}-\bar{y}-0}{1,7}<\frac{k+2-0}{1,7}\right]=P\left[z<\frac{k+2}{1,7}\right]=P\left[z>\frac{-k-2}{1,7}\right]=0,05 \rightarrow \frac{-k-2}{1,7}=1,645$$

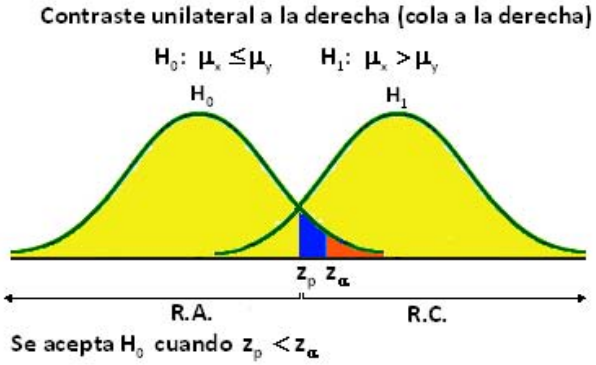
$$\rightarrow k+2=-2,7965$$

$$\text{Regla decisión}\begin{cases}\bar{x}-\bar{y}\geq -2,7965 & \text{se acepta } H_0\\ \bar{x}-\bar{y}<-2,7965 & \text{se rechaza } H_0\end{cases} \quad \text{siendo } \bar{x}-\bar{y}=2,28 \text{ se acepta } H_0$$

Se acepta la hipótesis nula sí $z_p=z_{observado}>-z_\alpha$

$$z_p=\frac{62,30-60,02-2}{\sqrt{\frac{9^2}{50}+\frac{8^2}{50}}}=0,164>-1,645=z_{0,05} \rightarrow \text{Se acepta } H_0$$





24. Una compañía de refrescos presenta un nuevo producto en el mercado afirmando que posee menos calorías que su homólogo más antiguo y conserva el resto de propiedades. Para verificar la afirmación de la compañía se eligieron al azar 14 botes del refresco nuevo y se calculó su media, 20 calorías por bote, y su desviación estándar muestral, tres calorías. De modo independiente, se tomó otra muestra aleatoria de 16 botes del refresco antiguo, obteniéndose una media de 28 calorías por bote con una desviación estándar muestral de 5 calorías. Suponiendo que la cantidad de calorías por bote sigue una distribución normal en ambos refrescos, pero con desviaciones típicas diferentes, ¿existe alguna razón para no creer en la afirmación de la compañía con un nivel de significación del 2,5%?

Solución:

Denotando por X e Y las variables aleatorias que representan la cantidad de calorías por bote e el nuevo producto y en el antiguo, respectivamente: $X \sim N(\mu_x, \sigma_x)$ e $Y \sim N(\mu_y, \sigma_y)$ siendo X e Y independientes y con $\sigma_x \neq \sigma_y$

Se pretenden contrastar las hipótesis:

H_0: $\mu_x - \mu_y \geq 0$

H_1: $\mu_x - \mu_y < 0$ (el nuevo producto posee menos calorías que el antiguo)

Se trata de un contraste unilateral a la izquierda.

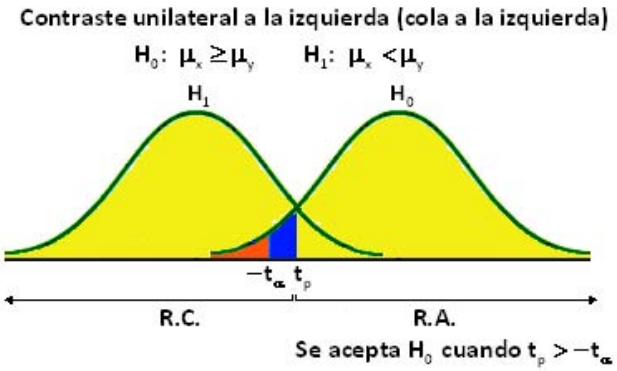
Se acepta H_0 sí $t_p = t_{observado} > -t_\alpha$

Estadístico de contraste:

$$t_p = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{20 - 28}{\sqrt{\frac{9}{14} + \frac{25}{16}}} = -5{,}387$$

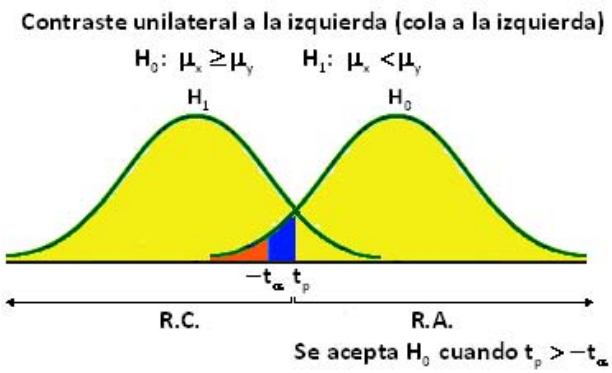
$$f = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x + 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y + 1}} - 2 = \frac{\left(\frac{9}{14} + \frac{25}{16}\right)^2}{\frac{\left(\frac{9}{14}\right)^2}{15} + \frac{\left(\frac{25}{16}\right)^2}{17}} - 2 \simeq 26$$

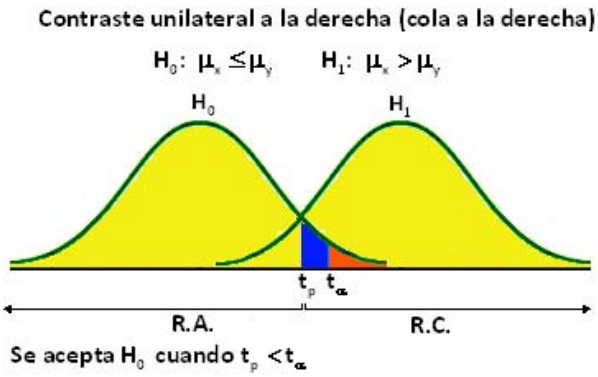
Estadístico teórico: $-t_{0{,}025\,,\,26} = -2{,}056$



El p_valor es el nivel de significación más pequeño para el que se acepta la hipótesis alternativa.

Como $t_{observado} = -5,387 < -2,056 = -t_{0,025\,,\,26}$ se rechaza la hipótesis nula concluyendo que no existen razones para no creer en la afirmación de la compañía.

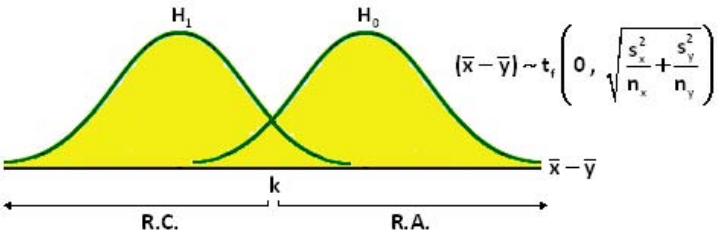




Otro procedimiento

Regla de decisión:

$$\begin{cases} \bar{x} - \bar{y} \geq k & \text{Se acepta } H_0 \\ \bar{x} - \bar{y} < k & \text{Se rechaza } H_0 \end{cases}$$



Bajo la hipótesis nula $(\bar{x} - \bar{y}) \sim t_f\left(0\,,\,\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right) = t_f\left(0\,,\,\sqrt{\frac{9}{14} + \frac{25}{16}}\right) \equiv t_{26}(0,\ 1,485)$

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[(\bar{x} - \bar{y}) < k \,\middle|\, t_{26}(0,\ 1,485)\right] = P\left[t_{26} < \frac{k}{1,485}\right] =$$

$$= P\left[t_{26} > \frac{-k}{1,485}\right] = 0,025 \quad \rightarrow \quad \frac{-k}{1,485} = 2,056 \quad \rightarrow \quad k = -3,053$$

Siendo $\bar{x} - \bar{y} = 20 - 28 = -8 < -3,053$ se rechaza la hipótesis nula, con lo que no se puede corroborar lo que dice la compañía con un nivel de significación del 2,5%.

25. El estacionamiento de corta estancia en el Aeropuerto Adolfo Suárez está cerca de la terminal, de modo que al recoger a un pasajero solo hay que caminar una corta distancia al área de recuperación de equipajes. Para decidir si el estacionamiento tiene suficientes lugares, el gerente del aeropuerto necesita saber si el tiempo medio de permanencia es superior a 40 minutos. Una muestra aleatoria de una población normal mostró que 12 clientes estuvieron en el estacionamiento los siguientes lapsos de tiempo, en minutos:

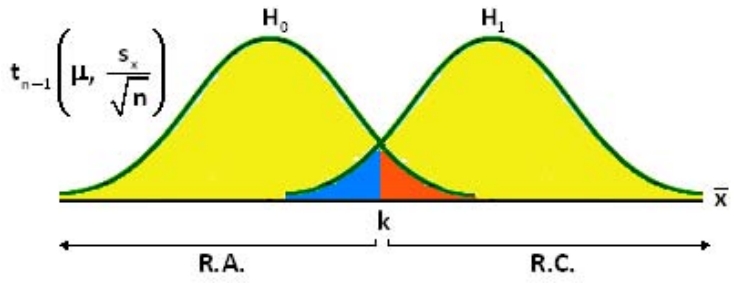
55	47	27	64	53	56
49	53	39	48	48	37

Con un nivel de significación 0,05, ¿se puede afirmar que el tiempo medio en el estacionamiento es superior a 40 minutos?

Solución:

Se establecen las hipótesis nula y alternativa: $H_0 : \mu \leq 40 \quad H_1 : \mu > 40$

Se trata de un contraste unilateral a la derecha (cola a la derecha)



Regla de decisión:

$$\begin{cases} \text{Si } \bar{x} \leq k & \text{Se acepta } H_0 \text{ (R.A)} \\ \text{Si } \bar{x} > k & \text{Se rechaza } H_0 \text{ (R.C)} \end{cases}$$

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

En la muestra: $\bar{x} = \dfrac{1}{12}\sum_{i=1}^{12} x_i = \dfrac{576}{12} = 48 \qquad s_x^2 = \dfrac{1}{11}\sum_{i=1}^{12}(x_i - \bar{x})^2 = \dfrac{1064}{11} = 96,727$

$$s_x = \sqrt{96,73} = 9,835 \quad \rightarrow \quad t_{11} = \frac{\bar{x}-\mu}{\frac{9,835}{\sqrt{12}}} = \frac{\bar{x}-\mu}{2,839}$$

Bajo la hipótesis nula, la muestra sigue una distribución $t_{11} = \dfrac{\bar{x}-40}{2,839}$

A partir del nivel de significación α se determina el valor crítico k:

$$\alpha = P\left[\text{Rechazar } H_0 \,\middle|\, H_0 \text{ Cierta}\right] = P\left[\bar{x} > k \,\middle|\, t_{n-1}\left(40, \frac{s_x}{\sqrt{n}}\right)\right] =$$

$$= P\left[\frac{\bar{x}-40}{9,835/\sqrt{12}} > \frac{k-40}{9,835/\sqrt{12}}\right] = P\left[t_{11} > \frac{k-40}{2,839}\right] = 0,05 \quad \rightarrow \quad \frac{k-40}{2,839} = 1,796 \quad \Rightarrow k = 45,098$$

$$\begin{cases} \text{Si } \bar{x} \leq 45,098 & \text{Se acepta } H_0 \text{ (R.A)} \\ \text{Si } \bar{x} > 45,098 & \text{Se rechaza } H_0 \text{ (R.C)} \end{cases}$$

Siendo $\bar{x} = 48 > 45,098$ se rechaza la hipótesis nula, concluyendo que el tiempo medio que los clientes pasan en el estacionamiento es superior a 40 minutos. El aeropuerto necesita ampliar su estacionamiento.

26. Muebles Quintana manifiesta que la satisfacción de sus empleados sigue una ley normal y es idéntica en las distintas secciones. El jefe de recursos humanos para decidir si la afirmación es correcta toma una muestra aleatoria de 100 trabajadores de barnizado encontrando 80 satisfechos y otra de 200 trabajadores de cortado encontrando 140 empleados satisfechos. Con un nivel de significación del 5% se puede admitir la afirmación de Muebles Quintana.

Solución:

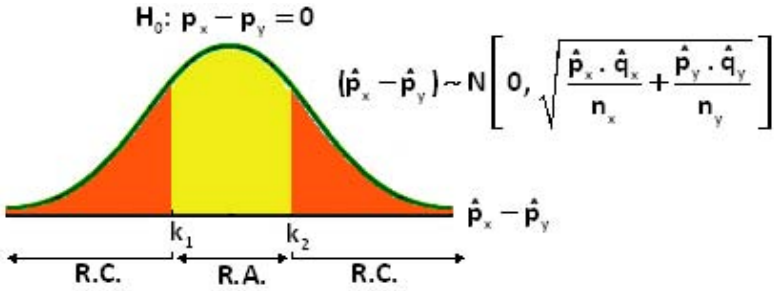
Se contrasta la hipótesis nula de que no existen diferencias entre las proporciones de trabajadores satisfechos en ambas secciones: $H_0: p_x = p_y$ frente a la hipótesis alternativa de que si existen diferencias $H_1: p_x \neq p_y$. El contraste es bilateral o de dos colas.

Las variables aleatorias $X \equiv$ " Trabajador de la sección de barnizado" e $Y \equiv$ " Trabajador de la sección de cortado', respectivamente, siguen una distribución de Bernouilli que toman el valor uno si el trabajador está satisfecho, y el valor cero en caso contrario.

Proporciones muestrales obtenidas son: $\hat{p}_x = \dfrac{80}{100} = 0,8 \quad \hat{p}_y = \dfrac{140}{200} = 0,7$

Bajo la hipótesis nula la diferencia de las proporciones muestrales $(\hat{p}_x - \hat{p}_y)$, teniendo en cuenta el tamaño de las muestras (TCL) siguen una distribución

$$(\hat{p}_x - \hat{p}_y) \sim N\left[0, \sqrt{\frac{\hat{p}_x \cdot \hat{q}_x}{n_x} + \frac{\hat{p}_y \cdot \hat{q}_y}{n_y}}\right] \equiv N\left[0, \sqrt{\frac{0,8 \cdot 0,2}{100} + \frac{0,7 \cdot 0,3}{200}}\right] \equiv N\left[0, 0,05\right]$$



Regla de decisión:

$$\begin{cases} |\hat{p}_x - \hat{p}_y| \leq k & \text{Se acepta } H_0 \text{ (RA)} \\ |\hat{p}_x - \hat{p}_y| > k & \text{Se rechaza } H_0 \text{ (RC)} \end{cases} \rightarrow \begin{cases} k_1 \leq \hat{p}_x - \hat{p}_y \leq k_2 & \text{Se acepta } H_0 \text{ (RA)} \\ \hat{p}_x - \hat{p}_y < k_1 \text{ ó } \hat{p}_x - \hat{p}_y > k_2 & \text{Se rechaza } H_0 \text{ (RC)} \end{cases}$$

Los valores críticos k se determinan mediante el nivel de significación $\alpha = 0,05$

$$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ Cierta}) = P\left[\,|\hat{p}_x - \hat{p}_y| > k \;/\; N\left[0, 0,05\right]\right] = P\left[\left|\frac{\hat{p}_x - \hat{p}_y - 0}{0,05}\right| > \frac{k - 0}{0,05}\right] =$$

$$= P\left[|z| > \frac{k}{0,05}\right] = P\left[\left(z < \frac{k_1}{0,05}\right) \cup \left(z > \frac{k_2}{0,05}\right)\right] = P\left[z > -\frac{k_1}{0,05}\right] + P\left[z > \frac{k_2}{0,05}\right] = 0,025 + 0,025 = 0,05$$

$$P\left[z > -\frac{k_1}{0,05}\right] = 0,025 \;\rightarrow\; -\frac{k_1}{0,05} = 1,96 \;\rightarrow\; k_1 = -0,098$$

$$P\left[z > \frac{k_2}{0,05}\right] = 0,025 \;\rightarrow\; \frac{k_2}{0,05} = 1,96 \;\rightarrow\; k_2 = 0,098$$

Región de aceptación: $-0,098 < \hat{p}_x - \hat{p}_y < 0,098$

La evidencia empírica (estadístico observado) $\left|\hat{p}_x - \hat{p}_y\right| = \left|0,8 - 0,7\right| = 0,1$ no se encuentra en la región de aceptación, por lo que no se admite la hipótesis nula. La satisfacción entre los trabajadores de las distintas secciones no es la misma.

▭ Aceptando la hipótesis nula H_0 cuando el estadístico observado o de contraste es menor que el estadístico teórico:

$$z_{\alpha_p/2} = \frac{\left|\hat{p}_x - \hat{p}_y\right|}{\sqrt{\frac{\hat{p}_x \cdot \hat{q}_x}{n_x} + \frac{\hat{p}_y \cdot \hat{q}_y}{n_y}}} \le z_{\alpha/2} \;\rightarrow\; z_{\alpha_p/2} = \frac{\left|0,8 - 0,7\right|}{0,05} = 2 \not< 1,96 = z_{0,025} \;\rightarrow$$

$\rightarrow$ Se rechaza la hipótesis nula H_0 $\Rightarrow$ La satisfacción de los trabajadores no es la misma en las secciones de Muebles Quintana

▭ Rechazando la hipótesis nula H_0 al verificarse la región de rechazo:

$$R.C = \left\{\left|\hat{p}_x - \hat{p}_y\right| > z_{\alpha/2}\sqrt{\frac{\hat{p}_x \cdot \hat{q}_x}{n_x} + \frac{\hat{p}_y \cdot \hat{q}_y}{n_y}}\right\} \;\rightarrow\; \left\{(0,8 - 0,7) > 1,96\sqrt{\frac{0,8 \cdot 0,2}{100} + \frac{0,7 \cdot 0,3}{200}}\right\} =$$

$= (0,1 > 0,098)$ $\rightarrow$ Se acepta la región de rechazo de H_0

▭ Elaborando un intervalo de confianza para la diferencia de parámetros $(\hat{p}_x - \hat{p}_y)$ de dos distribuciones binomiales

$$I(p_x - p_y) = \left[\hat{p}_x - \hat{p}_y \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x \cdot \hat{q}_x}{n_x} + \frac{\hat{p}_y \cdot \hat{q}_y}{n_y}}\right] = \left[0,1 \pm 1,96\sqrt{\frac{0,8 \cdot 0,2}{100} + \frac{0,7 \cdot 0,3}{200}}\right] =$$

$= [0,002\,,\, 0,198] \;\rightarrow\; 0 \notin [0,002\,,\, 0,198] \;\rightarrow$ Se rechaza H_0

27. Se quiere comparar las puntuaciones en estadística teórica de dos Universidades que siguen, respectivamente, una distribución normal, con varianzas desconocidas y se supone que estadísticamente iguales. Se toman al azar dos muestras iguales de estudiantes de ambas Universidades. Los resultados obtenidos en SPSS son:

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
Calificación	Se han asumido varianzas iguales	,152	,699	,443	28	,661	2,200	4,962	−7,964	12,364

Con un nivel de significación de 0,05, se pide:

a) ¿Existe diferencia significativa sobre las puntuaciones de estadística teórica en las dos Universidades?

b) ¿Cuál es la media ponderada de las cuasivarianzas muestrales? . Obtener el p-valor.

c) Con los datos de la tabla calcula el valor de la t de Student.

d) Cuál sería el p-valor si se hubieran contrastado las puntuaciones

$H_0 : \mu_{Madrid} \leq \mu_{Salamanca} \qquad H_1 : \mu_{Madrid} > \mu_{Salamanca}$

Solución:

a) En los resultados obtenidos en SPSS la Prueba de Levene permite decidir si las varianzas poblacionales son iguales. Como la probabilidad asociada al estadístico de Levene (0,699) es mayor que 0,05 se acepta la hipótesis nula de que las varianzas poblacionales son iguales.

$$I_{1-\alpha}(\mu_1-\mu_2) = \left[(\bar{x}-\bar{y}) \pm t_{\alpha/2,\,(n_1+n_2-2)} \cdot \underbrace{s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}_{\text{error típico diferencia}} \right]$$

$$LI = 12{,}364-(-7{,}964) = 20{,}328 \qquad \in = s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 4{,}962 \qquad LI = 2 \,.\, t_{0{,}025\,,\,28} \,.\, \in$$

Las dos Universidades tienen la misma puntuación en estadística teórica a un nivel de significación del 0,05 ya que el intervalo de confianza cubre el cero.

Un intervalo de confianza sirve para tomar una decisión sobre la hipótesis nula que permite contrastar el estadístico T (t de Student).

Por otra parte,

p-valor = Sig.(bilateral) $= 0{,}661 > 0{,}05 \quad \rightarrow$ Se acepta la hipótesis nula que establece

H_0: $\mu_{Madrid} = \mu_{Salamanca}$

b) $s_p^2 \equiv$ Media ponderada cuasivarianzas muestrales: $s_p^2 = \dfrac{(n_1-1)\,.\,s_1^2+(n_2-1)\,.\,s_2^2}{n_1+n_2-2}$

Error típico de la diferencia $= s_p \cdot \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}} \quad \rightarrow \quad 4{,}962 = s_p \cdot \sqrt{\dfrac{1}{15}+\dfrac{1}{15}}$

$$s_p = \frac{4{,}962}{\sqrt{\frac{1}{15}+\frac{1}{15}}} = \frac{4{,}962}{0{,}3651} = 13{,}588 \quad \rightarrow \quad s_p^2 = 184{,}633$$

$t \equiv$ valor experimental del estadístico de contraste:

$$t_{28} = \frac{\text{Diferencia de medias}}{\text{Error típico diferencia}} = \frac{|\bar{x}-\bar{y}|}{13{,}588 \cdot \sqrt{\frac{1}{15}+\frac{1}{15}}} = \frac{2{,}200}{4{,}962} = 0{,}443$$

$\alpha_p = 2.P(t_{28} \geq 0{,}443) = 2.0{,}3305 = 0{,}661$

Cálculo $P(t_{28} \geq 0{,}443)$:

Abscisas	Áreas			
$0{,}256 - 0{,}530$	$0{,}40 - 0{,}30$	$-0{,}274$	$0{,}10$	$\rightarrow \; x = 0{,}40 - \dfrac{0{,}187 \,.\, 0{,}10}{0{,}274} = 0{,}3305$
$0{,}256 - 0{,}443$	$0{,}40 - x$	$-0{,}187$	$0{,}40 - x$	

El p-valor = Sig.(bilateral) $= 0{,}661$, Significación bilateral, muestra el grado de compatibilidad entre el valor poblacional propuesto y la información muestral disponible.

Si el nivel crítico es pequeño (generalmente menor que 0,05), se concluye que la información recogida en la muestra es incompatible con la hipótesis nula de que las medias poblacionales son iguales.

La Prueba T se basa en el supuesto de normalidad, el cumplimiento de este supuesto sólo es exigible con muestras pequeñas. El supuesto de normalidad carece de relevancia en muestras grandes.

c) $LI = 20{,}328 = 2 \,.\, t_{0{,}025\,,\,28} \,.\, 4{,}962 \;\rightarrow\; t_{0{,}025\,,\,28} = \dfrac{20{,}328}{2 \,.\, 4{,}962} = 2{,}048$

d) SPSS siempre calcula el p-valor para un contraste bilateral o de dos colas, si se desea realizar un contraste unilateral o de una cola hay que dividir entre 2 el contraste bilateral
(p-valor = 0,661 / 2 = 0,3305)
Contraste unilateral (cola a la derecha): $H_1 : \mu_{Madrid} > \mu_{Salamanca}$)
p-valor = Sig.(unilateral derecha) $= 0{,}3305 > 0{,}05 \;\rightarrow\;$ Se acepta H_0

28. Las puntuaciones de estadística teórica se distribuyen según una ley normal, con μ y σ desconocidas. Se desea contrastar H_0: $\mu = 74$ frente a H_1: $\mu \neq 74$. Para ello, se introducen en SPSS las puntuaciones de quince alumnos escogidos aleatoriamente. Los resultados han sido:

Prueba para una muestra

	Valor de prueba = 74					
					95% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Autónoma	1,923	14	,075	7,200	-,83	15,23

Con un nivel de significación de 0,05, se pide:

a) ¿Qué se puede decir del contraste?
b) Media y cuasidesviación típica muestral
c) Define el p-valor y como calcular la significación bilateral
d) Calcula el p-valor al contrastar H_0: $\mu \leq 74$ frente a H_1: $\mu > 74$
e) Calcula el p-valor al contrastar H_0: $\mu \geq 74$ frente a H_1: $\mu < 74$

Solución:

a) El intervalo de confianza $[-0{,}83\,,\; 15{,}23]$ cubre el cero, se puede afirmar que la muestra procede de una población $N(74, \sigma)$, con un nivel de significación de 0,05

Por otro lado,

p-valor = Sig.(bilateral) $= 0{,}075 > 0{,}05$ y se acepta la hipótesis nula H_0: $\mu = 74$

b) $I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{\alpha/2,(n-1)} \frac{s_x}{\sqrt{n}} \right]$

Diferencia de medias: $\bar{x} - 74 = 7,2 \quad \rightarrow \quad \bar{x} = 81,2$

$t \equiv$ valor experimental del estadístico de contraste:

$$t = \frac{|\bar{x} - 74|}{s/\sqrt{15}} \quad \rightarrow \quad \frac{s_x}{\sqrt{15}} = \frac{|\bar{x} - 74|}{t} = \frac{7,2}{1,923} = 3,743$$

Error típico de la media: $\frac{s_x}{\sqrt{15}} = 3,743 \quad \rightarrow \quad s_x = 14,496$

c) La Significación bilateral (p_valor) muestra el grado de compatibilidad entre el valor poblacional propuesto y la información muestral disponible.

Si el nivel crítico es pequeño (generalmente menor que 0,05) se concluye que la información recogida en la muestra es incompatible con la hipótesis nula de que la muestra procede de la población.

$$\alpha_p = \text{p-valor} = P\left[\text{Rechazar } \bar{x} \mid H_0 \text{ es Cierta}\right]$$

$$\alpha_p = \text{p-valor} = P\left[\,|t_{14}| > 1,923\right] = 2.P\left[t_{14} > 1,923\right] = 0,075$$

d) SPSS siempre calcula el p_valor para un contraste bilateral o de dos colas, si se desea realizar un contraste unilateral o de una cola hay que dividir entre 2 el contraste bilateral
(p-valor $= 0,075 / 2 = 0,0375$)

$$\alpha_p = \text{p-valor} = P\left[t_{14} > 1,923\right] = 0,0375$$

Se trata de un contraste unilateral (cola a la derecha: H_1: $\mu > 74$)

p-valor = Sig.(Unilateral derecha) $= 0,0375 < 0,05 \quad \rightarrow \quad$ Se rechaza H_0

e) Se trata de un contraste unilateral (cola a la izquierda: H_1: $\mu < 74$)

p-valor = Sig.(Unilateral izquierda) $= 1 - 0,0375 = 0,9625 > 0,05 \quad \rightarrow \quad$ Se acepta H_0

29. Se quiere comparar las puntuaciones en un test de inteligencia de dos estratos sociales que siguen, respectivamente, una distribución normal pasando la prueba de Kolmogórov-Smirnov, con varianzas desconocidas. Se han tomado al azar dos muestras iguales en ambos estratos. Los resultados obtenidos en SPSS son:

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Puntuaciones	Se han asumido varianzas iguales	,048	,829	,330	28	,744	1,600	4,843	-8,321	11,521

Con un nivel de significación de 0,05, se pide:

a) Calcula con los datos el valor de la t de Student, sin recurrir a las tablas

b) ¿Cuál es la media ponderada de las cuasivarianzas muestrales?

Se aceptaría el contraste $H_0: \mu_{Estrato\,1} \geq \mu_{Estrato\,2}$ $H_1: \mu_{Estrato\,1} < \mu_{Estrato\,2}$ calculando el p-valor

Solución:

a) Diferencia medias $= \bar{x}_{Estrato\,1} - \bar{y}_{Estrato\,2} = 80,6 - 79 = 1,6$

$$\text{Estadístico } t = \frac{\left|\bar{x}_{Estrato\,1} - \bar{y}_{Estrato\,2}\right|}{\in} = \frac{1,6}{4,843} = 0,330$$

$$I(\mu_{Estrato\,1} - \mu_{Estrato\,2}) = \left[(\bar{x}_{Estrato\,1} - \bar{y}_{Estrato\,2}) \pm t_{0,025\,,\,28} \cdot \in\right] = \left[-8,321\,,\ 11,521\right]$$

$$\text{Longitud Intervalo} = 11,521 - (-8,321) = 19,842 = 2 \,.\, t_{0,025\,,\,28} \,.\, \in = 2 \,.\, t_{0,025\,,\,28} \,.\, 4,843$$

$$t_{0,025\,,\,28} = \frac{19,842}{2 \,.\, 4,843} = 2,048$$

b) $\in =$ Error típico diferencias $= 4,843 = s_p \cdot \sqrt{\frac{1}{15} + \frac{1}{15}} \rightarrow s_p = \frac{4,843}{\sqrt{\frac{1}{15} + \frac{1}{15}}} = 13,2632$

c) SPSS siempre calcula el p-valor para un contraste bilateral o de dos colas, si se desea realizar un contraste unilateral o de una cola hay que dividir entre 2 el contraste bilateral, teniendo en cuenta que es un contraste unilateral a la izquierda $H_0: \mu_{Estrato\,1} \geq \mu_{Estrato\,2}$ el p-valor $= 1 - 0,744 / 2 = 0,628$

p-valor = Sig.(unilateral izquierda) $= 0,628 > 0,05 \rightarrow$ Se acepta H_0

30. Dos agencias de viajes siguen distintas estrategias para la venta de paquetes de viajes, siguiendo la característica en estudio una ley normal, se han introducido los datos en SPSS y aparece en el Visor la tabla siguiente:

Prueba de muestras independientes

		Prueba T para la igualdad de medias						
		t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
							Inferior	Superior
Estrategia Ventas	Se han asumido varianzas iguales	,734	28	,469	,63867	,87045	-1,14436	

Mediante cálculo razonado, se pide:

a) Signatura bilateral (p-valor)
b) Valor de la t de Student
c) ¿Cuál es el diagnóstico atendiendo al intervalo de confianza?

Solución:

a) $$t_{p/2} = \frac{|\bar{x} - \bar{y}|}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0,63867}{0,87045} = 0,7337$$

$$\alpha_p = P\left[\,|t_{28}| > 0,734\right] = 2 \,.\, P\left[t_{28} > 0,734\right] = 2 \,.\, 0,235 = 0,47$$

$$\frac{0,8546 - 0,6834}{0,200 - 0,250} = \frac{0,734 - 0,6834}{x - 0,250} \rightarrow \frac{0,1712}{-0,05} = \frac{0,0506}{x - 0,250}$$

$$x = 0,250 - \frac{0,05 \,.\, 0,0506}{0,1712} = 0,235$$

b) $$L.I = -1,14436 = 0,63867 - t_{\alpha/2,28} \,.\, 0,87045 \rightarrow t_{\alpha/2,28} = 2,0484 \rightarrow \alpha = 0,05$$

c) $$L.S = 0,63867 + 2,0484 \,.\, 0,87045 = 2,42170$$

El intervalo de confianza $I = \left[-1,14436\,,\ 2,42170\right]$ cubre el 0, con lo que con una significación $\alpha = 0,05$, se puede afirmar que no existe diferencia entre la venta de paquetes de viajes de las dos agencias.

Región de rechazo
Hipótesis sobre los parámetros poblacionales

HIPÓTESIS SOBRE LA MEDIA DE UNA POBLACIÓN: REGIÓN DE RECHAZO

$X \approx N(\mu,\sigma)$ σ^2 conocida $z_\alpha = -z_{1-\alpha}$			
$H_0: \mu = \mu_0$	$H_1: \mu_1 > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$	unilateral simple derecha
$H_0: \mu = \mu_0$ $H_1: \mu_1 < \mu_0$		$R = \left\{ (\bar{x}-\mu_0) < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$	unilateral simple izquierda
$H_0: \mu = \mu_0$	$H_1: \mu > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$	unilateral compuesta derecha
$H_0: \mu = \mu_0$	$H_1: \mu < \mu_0$	$R = \left\{ (\bar{x}-\mu_0) < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$	unilateral compuesta izquierda
$H_0: \mu = \mu_0$	$H_1: \mu \neq \mu_0$	$R = \left\{ \lvert \bar{x}-\mu_0 \rvert > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$	bilateral compuesta
$H_0: \mu \leq \mu_0$	$H_1: \mu > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$	unilateral compuesta derecha
$H_0: \mu \geq \mu_0$	$H_1: \mu < \mu_0$	$R = \left\{ (\bar{x}-\mu_0) < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$	unilateral compuesta izquierda

$X \approx N(\mu,\sigma)$ σ^2 desconocida $n.\sigma_x^2 = (n-1).s_x^2 \rightarrow \frac{s_x}{\sqrt{n}} = \frac{\sigma_x}{\sqrt{n-1}}$ $t_{\alpha,n} = -t_{1-\alpha,n}$			
$H_0: \mu = \mu_0$	$H_1: \mu_1 > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > t_{\alpha,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	unilateral simple derecha
$H_0: \mu = \mu_0$	$H_1: \mu_1 < \mu_0$	$R = \left\{ (\bar{x}-\mu_0) < t_{1-\alpha,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	unilateral simple izquierda
$H_0: \mu = \mu_0$	$H_1: \mu > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > t_{\alpha,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	unilateral compuesta derecha
$H_0: \mu = \mu_0$	$H_1: \mu \neq \mu_0$	$R = \left\{ \lvert \bar{x}-\mu_0 \rvert > t_{\alpha/2,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	bilateral compuesta
$H_0: \mu \leq \mu_0$	$H_1: \mu > \mu_0$	$R = \left\{ (\bar{x}-\mu_0) > t_{\alpha,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	unilateral compuesta derecha
$H_0: \mu \leq \mu_0$	$H_1: \mu < \mu_0$	$R = \left\{ (\bar{x}-\mu_0) < t_{1-\alpha,(n-1)} \frac{s_x}{\sqrt{n}} \right\}$	unilateral compuesta izquierda

HIPÓTESIS SOBRE LA VARIANZA DE UNA POBLACIÓN: REGIÓN DE RECHAZO

Media poblacional conocida	Región de rechazo hipótesis nula
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma_1^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha;n} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma_1^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha;n} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha;n} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha;n} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 \neq \sigma_0^2$	$R = \left\{ \begin{array}{l} \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha/2;n} \\ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha/2;n} \end{array} \right\}$
$H_0: \sigma^2 \leq \sigma_0^2 \quad H_1: \sigma^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha;n} \right\}$
$H_0: \sigma^2 \leq \sigma_0^2 \quad H_1: \sigma^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha;n} \right\}$

HIPÓTESIS SOBRE LA VARIANZA DE UNA POBLACIÓN: REGIÓN DE RECHAZO

Media poblacional desconocida	Región de rechazo hipótesis nula
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma_1^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha,(n-1)} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma_1^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha,(n-1)} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha,(n-1)} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha,(n-1)} \right\}$
$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 \neq \sigma_0^2$	$R = \left\{ \begin{array}{l} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha/2,(n-1)} \\ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha/2,(n-1)} \end{array} \right\}$
$H_0: \sigma^2 \leq \sigma_0^2 \quad H_1: \sigma^2 > \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \geq \chi^2_{\alpha/2,(n-1)} \right\}$
$H_0: \sigma^2 \leq \sigma_0^2 \quad H_1: \sigma^2 < \sigma_0^2$	$R = \left\{ \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma_0^2 \right] \leq \chi^2_{1-\alpha,(n-1)} \right\}$

HIPÓTESIS SOBRE LA IGUALDAD DE MEDIAS: REGIÓN DE RECHAZO

Igualdad de medias: $X \approx N(\mu_1, \sigma_1)$, $Y \approx N(\mu_2, \sigma_2)$	
$H_0: \mu_1 = \mu_2$ $((\sigma_1, \sigma_2$ conocidas)	$R = \left\{ \lvert \bar{x} - \bar{y} \rvert > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: \mu_1 - \mu_2 = k$ $(\sigma_1, \sigma_2$ conocidas)	$R = \left\{ \lvert \bar{x} - \bar{y} - k \rvert > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: \mu_1 = \mu_2$ σ_1, σ_2 desconocidas pero iguales	$R = \left\{ \lvert \bar{x} - \bar{y} \rvert > t_{\alpha/2, (n_1 + n_2 - 2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$
$H_0: \mu_1 = \mu_2$ σ_1, σ_2 desconocidas y distintas	$R = \left\{ \lvert \bar{x} - \bar{y} \rvert > t_{\alpha/2, f}\, s_p \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$
$H_0: \mu_1 \leq \mu_2$ $((\sigma_1, \sigma_2$ conocidas)	$R = \left\{ (\bar{x} - \bar{y}) > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: (\mu_1 - \mu_2) \leq k$ $(\sigma_1, \sigma_2$ conocidas)	$R = \left\{ (\bar{x} - \bar{y} - k) > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: \mu_1 \leq \mu_2$ $(\sigma_1 = \sigma_2)$	$R = \left\{ (\bar{x} - \bar{y}) > t_{\alpha, (n_1 + n_2 - 2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$
$H_0: \mu_1 \leq \mu_2$ $(\sigma_1 \neq \sigma_2)$	$R = \left\{ (\bar{x} - \bar{y}) > t_{\alpha, f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$
$H_0: \mu_1 \geq \mu_2$ $(\sigma_1, \sigma_2$ conocidas)	$R = \left\{ (\bar{x} - \bar{y}) > z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: (\mu_1 - \mu_2) \geq k$ $(\sigma_1, \sigma_2$ conocidas)	$R = \left\{ (\bar{x} - \bar{y} - k) < -z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$
$H_0: \mu_1 \geq \mu_2$ $(\sigma_1 = \sigma_2)$	$R = \left\{ (\bar{x} - \bar{y}) > t_{1-\alpha, (n_1 + n_2 - 2)}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

HIPÓTESIS SOBRE LA IGUALDAD DE MEDIAS: REGIÓN DE RECHAZO

Igualdad de medias: $X \approx N(\mu_1, \sigma_1)$, $Y \approx N(\mu_2, \sigma_2)$	
$H_0: \mu_1 \geq \mu_2 \quad (\sigma_1 = \sigma_2)$	$R = \left\{ (\bar{x} - \bar{y}) > t_{1-\alpha, f}\ s_p \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$

HIPÓTESIS SOBRE LA IGUALDAD DE VARIANZAS: REGIÓN DE RECHAZO

$H_0: \sigma_1 = \sigma_2$	$R = \left\{ \frac{s_1^2}{s_2^2} \notin \left[F_{1-\alpha/2, (n_1-1), (n_2-1)}, F_{\alpha/2, (n_1-1), (n_2-1)} \right] \right\}$
$H_0: \sigma_1 \leq \sigma_2$	$R = \left\{ \frac{s_1^2}{s_2^2} > F_{\alpha, (n_1-1), (n_2-1)} \right\}$
$H_0: \sigma_1 \geq \sigma_2$	$R = \left\{ \frac{s_1^2}{s_2^2} < F_{1-\alpha, (n_1-1), (n_2-1)} \right\}$

HIPÓTESIS SOBRE EL PARÁMETRO P DE UNA DISTRIBUCIÓN BINOMIAL

$X \approx B(n, p)$ (muestras grandes)	Región de rechazo hipótesis nula
$H_0: p = p_0$	$R = \left\{ \lvert \hat{p} - p_0 \rvert > z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right\}$
$H_0: p \leq p_0$	$R = \left\{ (\hat{p} - p_0) > z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right\}$
$H_0: p \geq p_0$	$R = \left\{ (\hat{p} - p_0) < z_{1-\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right\}$

HIPÓTESIS SOBRE EL PARÁMETRO LAMBDA DE UNA DISTRIBUCIÓN DE POISSON

$X \approx \text{Poisson}(\lambda)$ (muestras grandes)	Región de rechazo hipótesis nula
$H_0: \lambda = \lambda_0$	$R = \left\{ \lvert \hat{\lambda} - \lambda_0 \rvert > z_{\alpha/2} \sqrt{\frac{\hat{\lambda}}{n}} \right\}$
$H_0: \lambda \leq \lambda_0$	$R = \left\{ (\hat{\lambda} - \lambda_0) > z_{\alpha} \sqrt{\frac{\hat{\lambda}}{n}} \right\}$
$H_0: \lambda \geq \lambda_0$	$R = \left\{ (\hat{\lambda} - \lambda_0) < z_{1-\alpha} \sqrt{\frac{\hat{\lambda}}{n}} \right\}$

HIPÓTESIS SOBRE LA IGUALDAD DE PARÁMETROS DE DOS DISTRIBUCIONES BINOMIALES

$X \approx B(n_1, p_1)$ $Y \approx B(n_2, p_2)$	Región de rechazo hipótesis nula
$H_0: p_1 = p_2$	$R = \left\{ \lvert \hat{p}_1 - \hat{p}_2 \rvert > z_{\alpha/2} \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}} \right\}$
$H_0: p_1 \leq p_2$	$R = \left\{ (\hat{p}_1 - \hat{p}_2) > z_{\alpha} \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}} \right\}$
$H_0: p_1 \geq p_2$	$R = \left\{ (\hat{p}_1 - \hat{p}_2) < z_{1-\alpha} \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}} \right\}$

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} \qquad f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \text{ entero más próximo}$$

EJERCICIOS
RESUELTOS
CON SPSS

CONTRASTE PARA LA MEDIA POBLACIONAL DE UNA DISTRIBUCIÓN NORMAL

1. Se realiza un experimento para analizar el tiempo (en minutos) que requiere el lagarto para alcanzar 45^0 partiendo de la temperatura normal de su cuerpo mientras que está en la sombra, el proceso sigue una ley normal. Se han obtenido las siguientes observaciones:

10,1 12,5 12,2 10,2 12,8 12,1 11,2 11,4 10,7 14,9 13,9 13,3

¿Puede concluirse que el tiempo medio requerido es inferior a 13 minutos, con una confianza del 95%?

Solución:

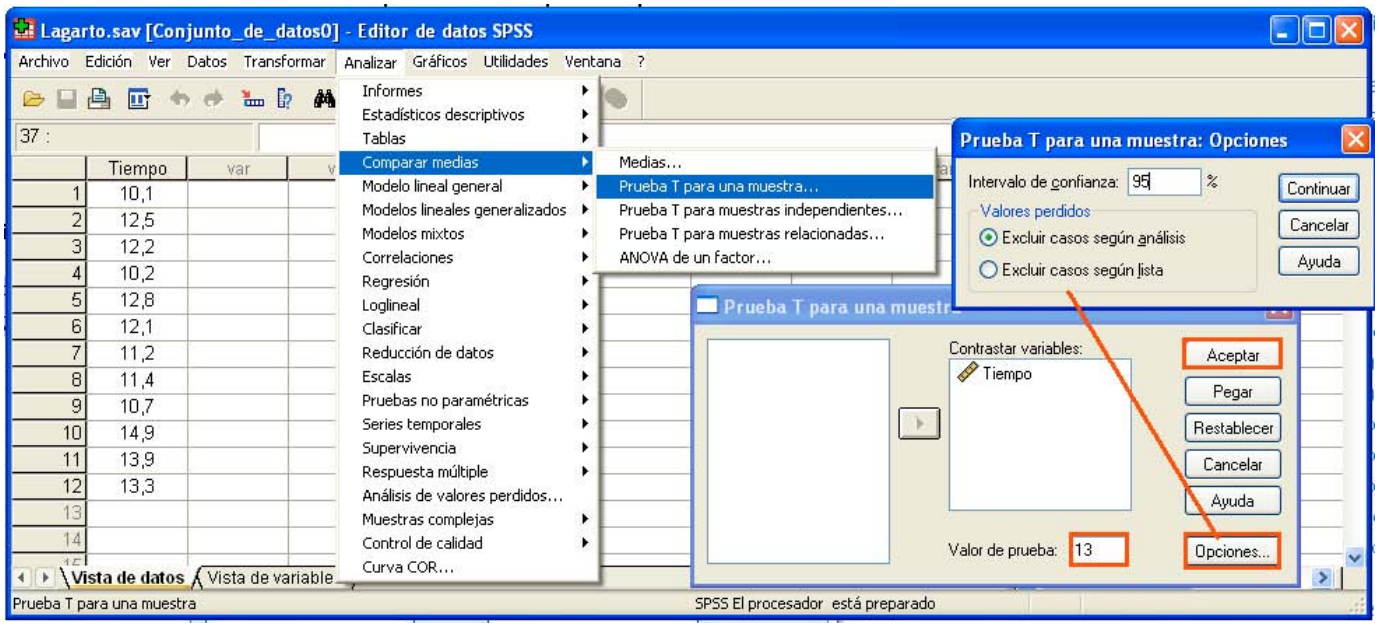
Se trata de un contraste unilateral a la derecha, se establecen las hipótesis: $H_0 : \mu \leq 13$ $H_1 : \mu > 13$

Lagarto.sav

SPSS: En "Vista de variable" se define la variable. Posteriormente, en "Vista de datos" se introducen los datos.

En el menú principal: Analizar/Comparar medias/Prueba T para una muestra...

En la pantalla emergente se selecciona Tiempo para la Variable para contrastar y en el Valor de la prueba se pone 13. En Opciones se determina el nivel de confianza.



Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
Tiempo	12	12,108	1,4786	,4268

Prueba para una muestra

	Valor de prueba = 13					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Tiempo	-2,089	11	,061	-,8917	-1,831	,048

- En pantalla sale el contraste bilateral: $H_0 : \mu = 13$ $H_1 : \mu \neq 13$

Estadístico experimental: $t_{11} = \dfrac{\text{Diferencia de medias}}{\text{Error típico de la media}} = \dfrac{|\bar{x} - \mu_0|}{s_x / \sqrt{n}} = \dfrac{|12{,}108 - 13|}{0{,}4268} = 2{,}089$

α_p = p-valor (Significación bilateral) $= 2.\ P\left[t_{11} \geq 2{,}089\right] = 2.\ 0{,}0305 = 0{,}061$

siendo $\alpha_p = 0{,}061 > 0{,}05 = \alpha$ se acepta la hipótesis nula H_0 concluyendo que el tiempo medio requerido es de 13 minutos.

Por otra parte, el Intervalo de confianza para la diferencia $-1{,}831 \leq \mu - \overline{x} \leq 0{,}048$ cubre el 0, por lo que se acepta que la media poblacional sea 13.

- En el contraste unilateral a la derecha: $H_0 : \mu \leq 13 \quad H_1 : \mu > 13$

El valor del estadístico de contraste experimental bilateral (p-valor = 0,061) deja a la derecha una área

$$\alpha_{p/2} = \frac{0{,}061}{2} = 0{,}0305 < 0{,}05 = \alpha$$

En consecuencia, se rechaza la hipótesis nula y se concluye que el tiempo medio requerido no es inferior a 13 minutos.

CONTRASTE PARA LA DIFERENCIA DE MEDIAS POBLACIONALES DE DOS DISTRIBUCIONES NORMALES.

2. Se quieren comparar dos poblaciones de ranas pipiens aisladas geográficamente. Para ello se toman dos muestras de ambas poblaciones de tamaño 12 y 10 y se les mide la longitud del cuerpo expresado en milímetros. Se supone que la longitud se distribuye según una ley normal.

Población 1	20,1	22,5	22,2	30,2	22,8	22,1	21,2	21,4	20,7	24,9	23,9	23,3
Población 2	25,3	31,2	22,4	23,1	26,4	28,2	21,3	31,1	26,2	21,4		

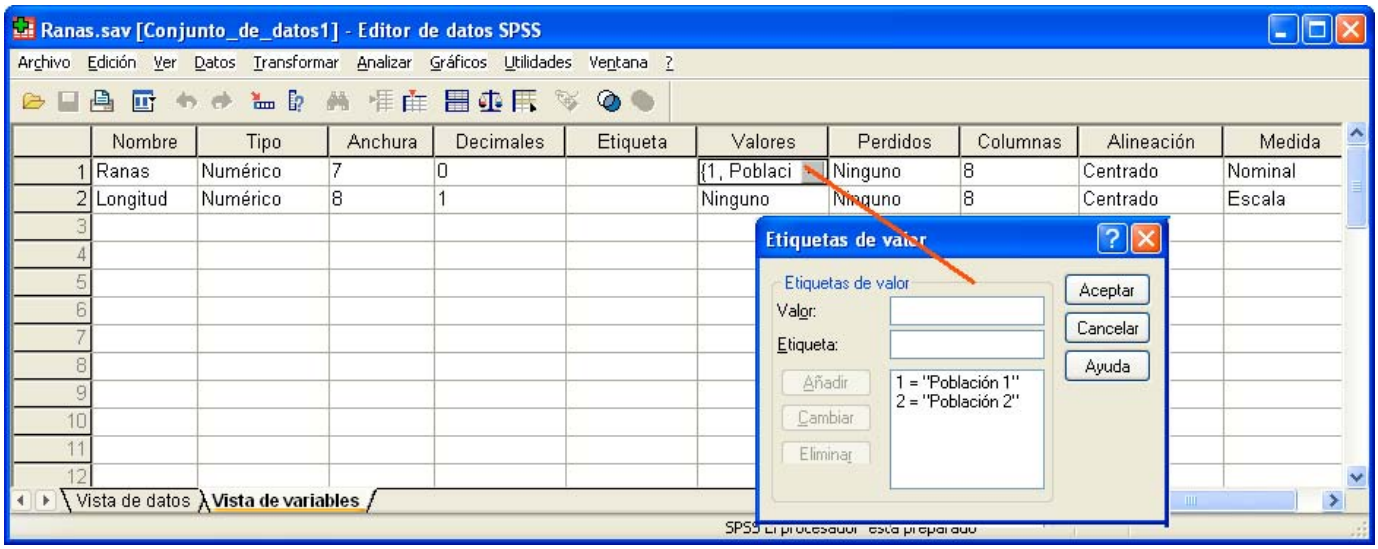
Contrastar la hipótesis de igualdad de medias a un nivel de significación del 1%.

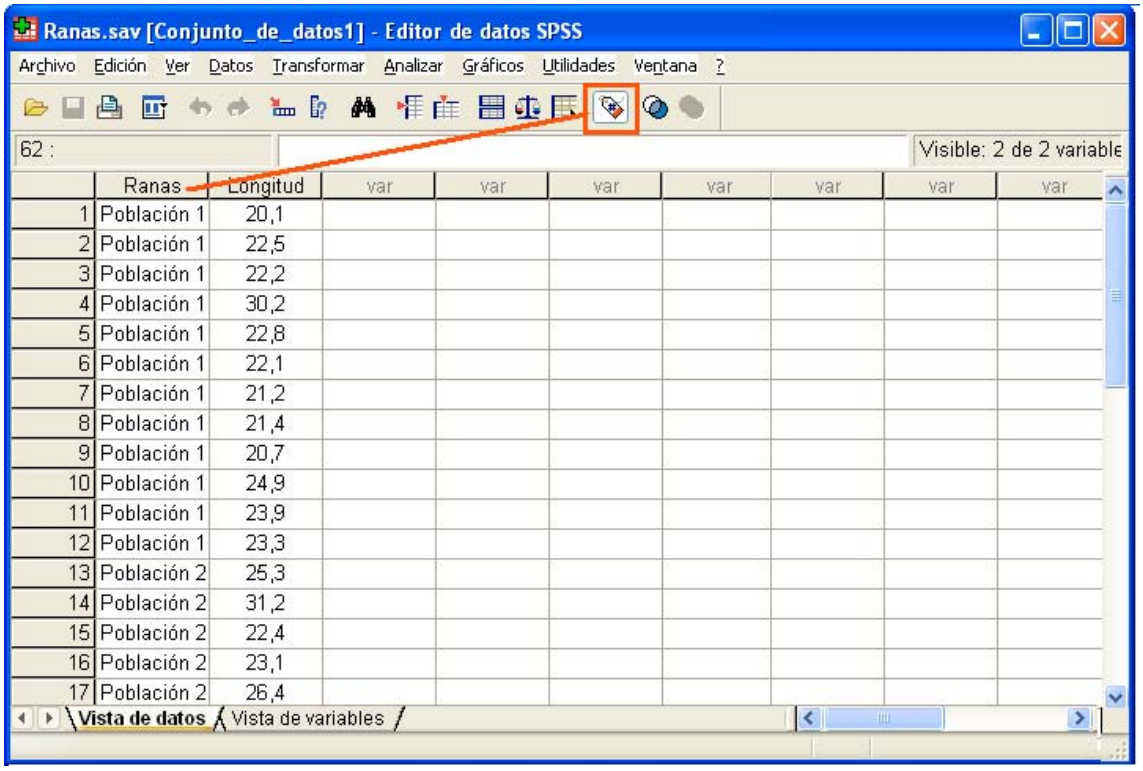
Solución:

Se trata de un contraste bilateral, se establecen las hipótesis: $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$

Ranas.sav

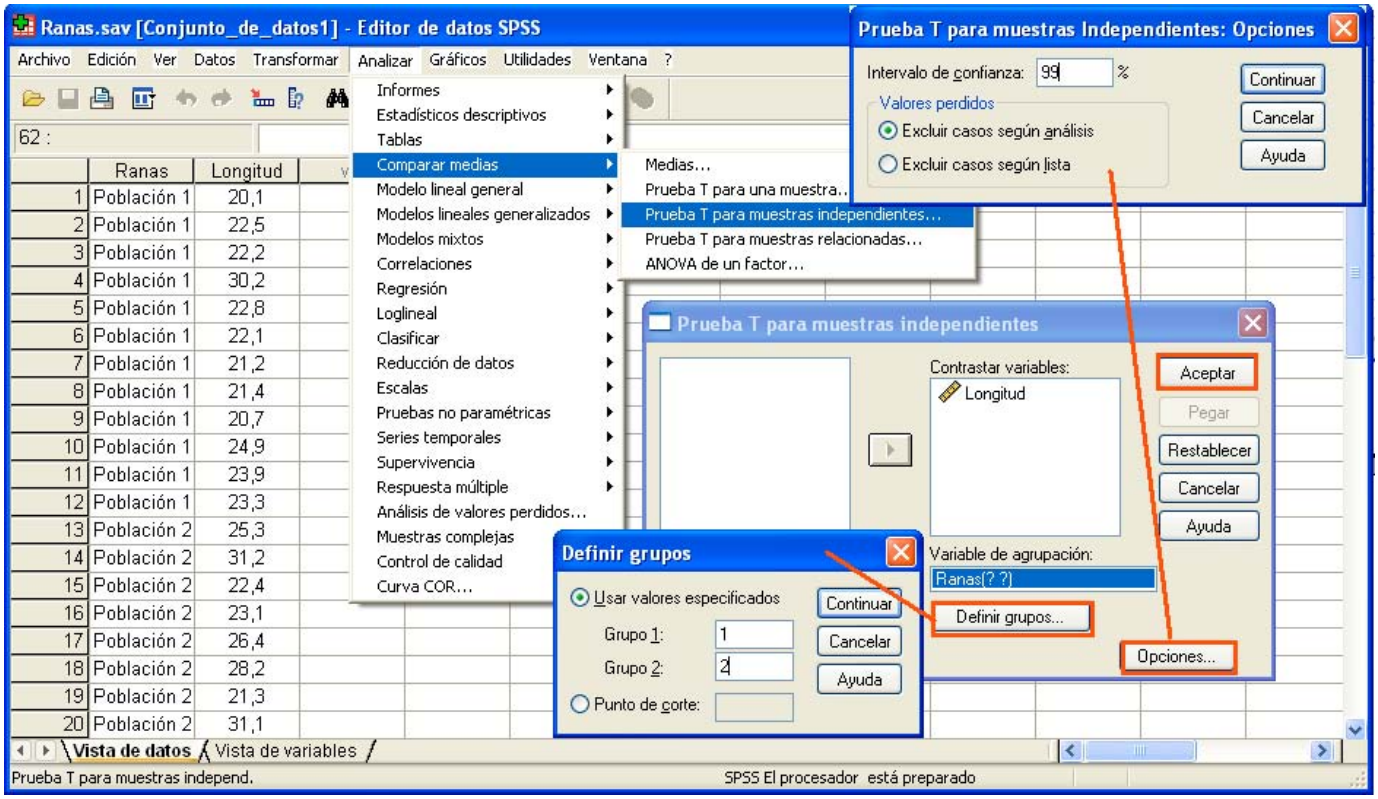
SPSS: En "Vista de variable" se definen las variables y sus Valores.
Posteriormente, en "Vista de datos" se introducen los datos.





En el menú principal: Analizar/Comparar medias/Prueba T para muestras independientes

En la pantalla emergente se selecciona Longitud para Contrastar variables y Ranas en Variable de agrupación, definiendo los grupos. En Opciones se determina el nivel de confianza.



Estadísticos de grupo

	Ranas	N	Media	Desviación típ.	Error típ. de la media
Longitud	Población 1	12	22,942	2,6552	,7665
	Población 2	10	25,660	3,6825	1,1645

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	99% Intervalo de confianza para la diferencia	
									Inferior	Superior
Longitud	Se han asumido varianzas iguales	2,110	,162	-2,010	20	,058	-2,7183	1,3526	-6,5670	1,1304
	No se han asumido varianzas iguales			-1,950	16,026	,069	-2,7183	1,3941	-6,7893	1,3527

El valor experimental del estadístico de Levene $F = 2{,}110$ presenta un p-valor (Sig) $= 0{,}162 > 0{,}05$, por lo que se acepta la hipótesis nula de igualdad de varianzas.

La Prueba T para la igualdad de medias tiene un p-valor (Sig. bilateral) $= 0{,}058 > 0{,}05 = \alpha$ aceptando la igualdad de medias en las dos poblaciones de ranas.

Por otra parte, el Intervalo de confianza para la diferencia de medias $-6{,}5870 \leq \mu_1 - \mu_2 \leq 1{,}1304$ cubre el 0, concluyendo que las medias de las dos poblaciones de ranas son iguales.

$$t_{20} = \frac{\text{Diferencia de medias}}{\text{Error típico de la difencia}} = \frac{|22{,}942 - 25{,}660|}{1{,}3528} = 2{,}010$$

p-valor (Significación bilateral) $= \alpha_p = 2.P[t_{20} \geq 2{,}010] = 2.0{,}029 = 0{,}058$

Error típico diferencia $= 1{,}3528 = s_p \cdot \sqrt{\frac{1}{12} + \frac{1}{10}} \rightarrow 1{,}3528 = s_p \cdot 0{,}4282 \rightarrow s_p = 3{,}1592$

Media ponderada de cuasivarianzas muestrales: $s_p^2 = 3{,}1592^2 = 9{,}980$

$LI = 7{,}6974 = 2 \cdot t_{0{,}005\,,\,20} \cdot 1{,}3526 \rightarrow t_{0{,}005\,,\,20} = \frac{7{,}6974}{2 \cdot 1{,}3526} = 2{,}8454$

PRUEBAS DE NORMALIDAD.
CONTRASTE PARA LA DIFERENCIA DE MEDIAS POBLACIONALES DE DOS DISTRIBUCIONES NORMALES.

3. En un máster universitario de Gestión Aeronáutica las calificaciones obtenidas en una muestra han sido:

Autónoma $\begin{cases} 99 & 98 & 97 & 99 & 68 & 60 & 61 & 56 \\ 79 & 82 & 84 & 81 & 79 & 85 & 90 \end{cases}$

Complutense $\begin{cases} 69 & 72 & 73 & 75 & 62 & 60 & 85 & 96 \\ 95 & 80 & 96 & 98 & 78 & 79 & 67 \end{cases}$

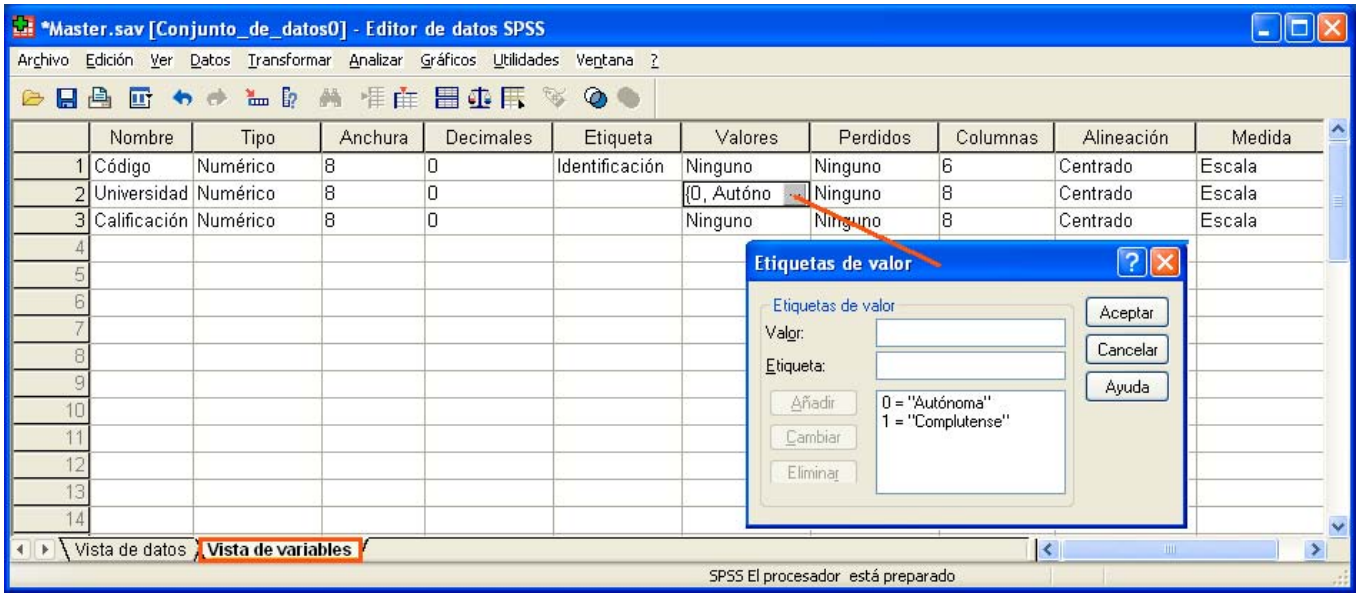
a) Analizar la normalidad de las calificaciones obtenidas.

b) Contrastar la igualdad de medias poblaciones de las dos Universidades, con una fiabilidad del 90%

Solución:

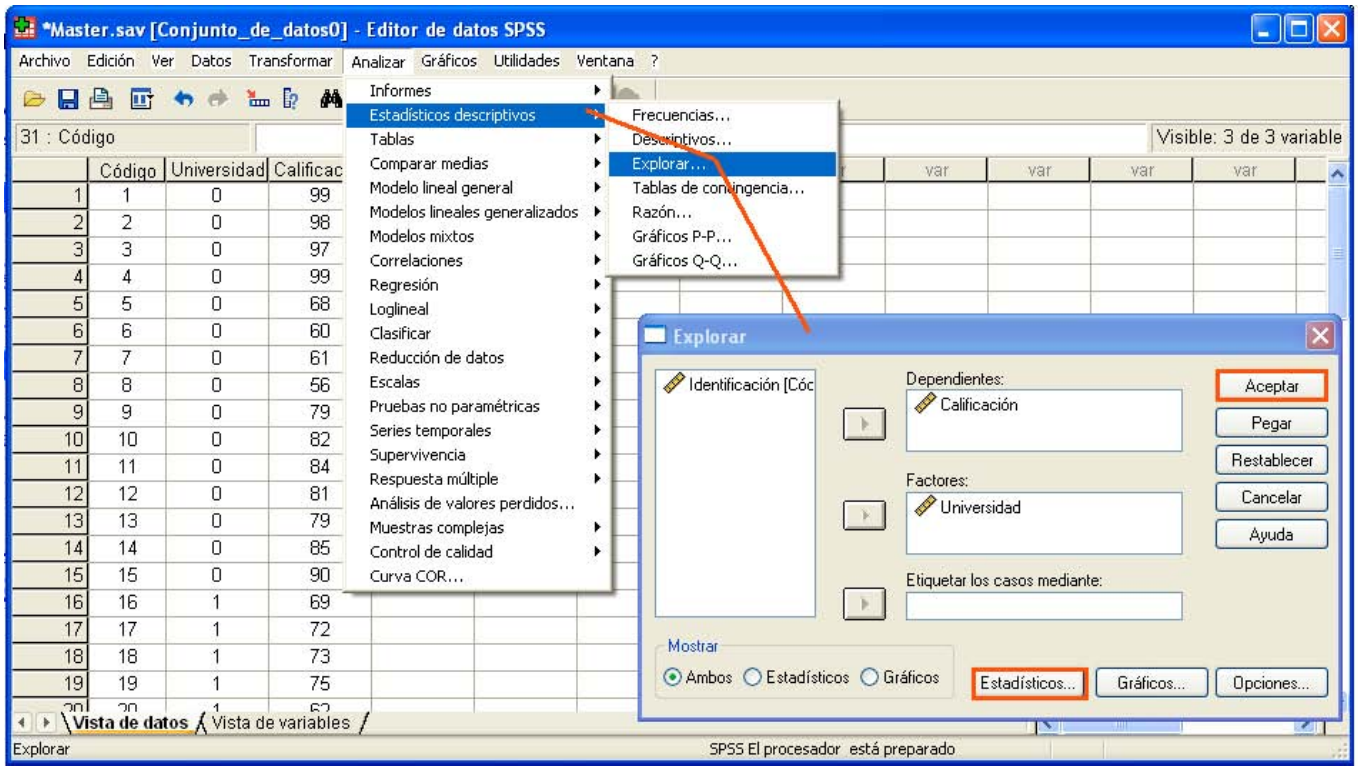
a) Se establece la hipótesis nula H_0 : "Los datos proceden de distribuciones normales"

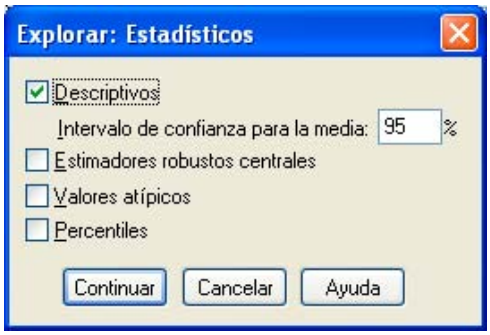
SPSS: En "Vista de variable" se definen las variables y se introducen sus valores. Posteriormente, en "Vista de datos" se introducen los datos.



En el menú principal: Analizar/Estadísticos descriptivos/Explorar…

En la ventana emergente se traspasa la variable Calificación como variable Dependiente y Universidad a Factores.





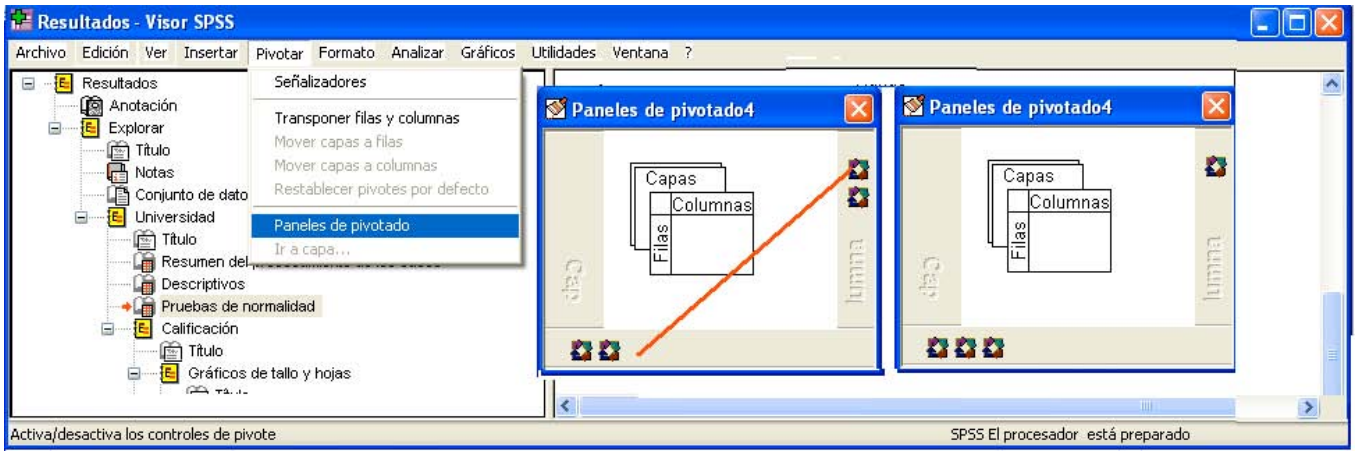
En la opción Estadísticos… se define el nivel de confianza.

Pruebas de normalidad

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Universidad	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Calificación	Autónoma	,173	15	,200*	,914	15	,156
	Complutense	,164	15	,200*	,928	15	,257

*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors



Pruebas de normalidad

	Universidad		Estadístico	gl	Sig.
Calificación	Autónoma	Kolmogorov-Smirnov[a]	,173	15	,200*
		Shapiro-Wilk	,914	15	,156
	Complutense	Kolmogorov-Smirnov[a]	,164	15	,200*
		Shapiro-Wilk	,928	15	,257

*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors

Las pruebas de normalidad para una muestra es un procedimiento de "bondad de ajuste", que permite medir el grado de concordancia existente entre la distribución de un conjunto de datos y una distribución teórica normal.

Para comprobar que la muestra ha sido extraída de una población con distribución normal se recurre a un estudio gráfico y/o analítico.

Cuando la muestra $n > 50$ se aplica el contraste de Kolmogórov-Smirnov, si la muestra es pequeña $n \leq 50$ se aplica el contraste de Shapiro-Wilk.

Aplicando el estadístico de Shapiro-Wilk:

Autónoma: p-valor $= 0{,}156 > 0{,}05 = \alpha$ $\rightarrow$ Se acepta H_0, aceptando que los datos proceden de una distribución normal.

Complutense: p-valor $= 0{,}257 > 0{,}05 = \alpha$ $\rightarrow$ Se acepta H_0, aceptando que los datos proceden de una distribución normal.

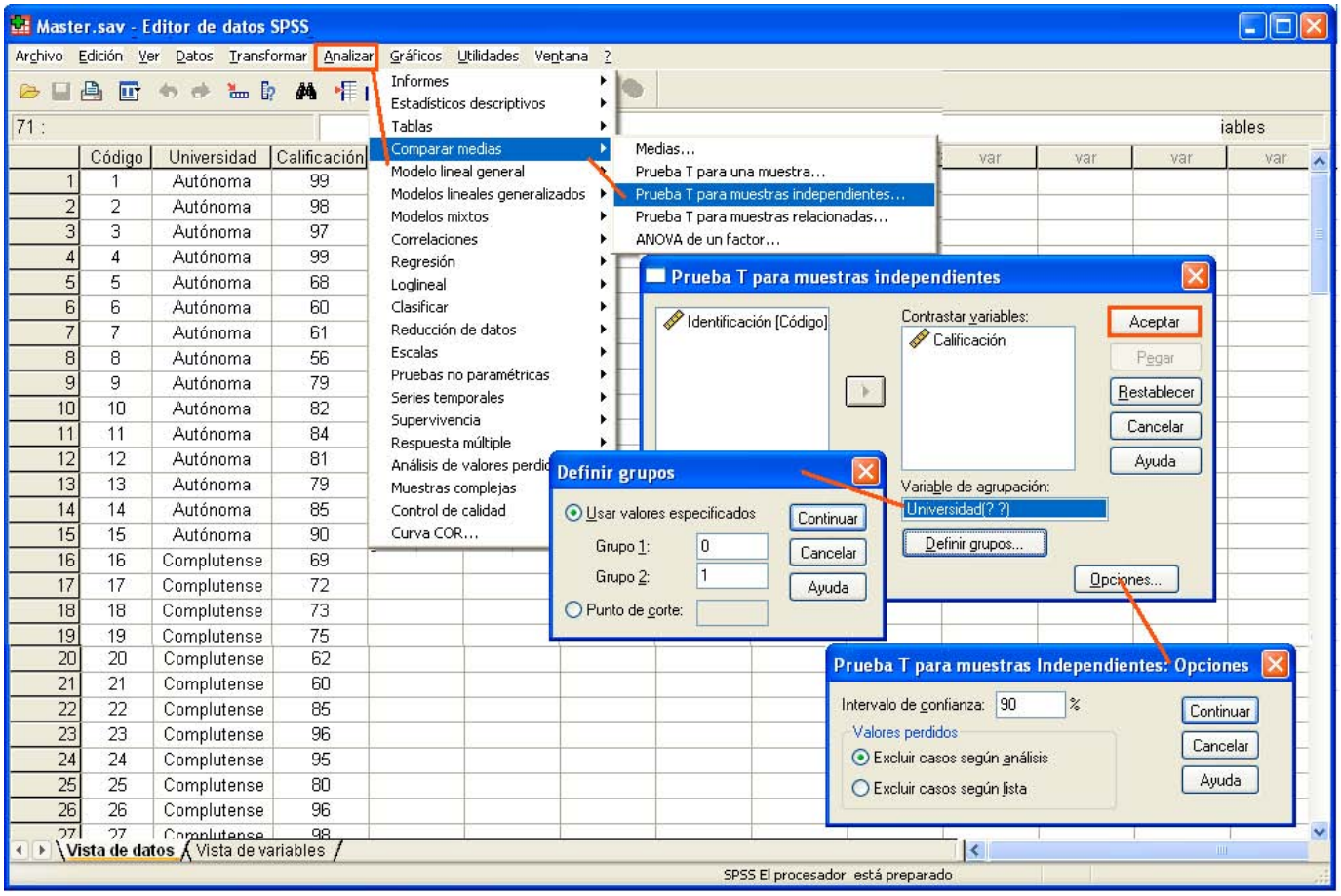
Descriptivos

	Universidad			Estadístico	Error típ.
Calificación	Autónoma	Media		81,20	3,743
		Intervalo de confianza para la media al 95%	Límite inferior	73,17	
			Límite superior	89,23	
		Media recortada al 5%		81,61	
		Mediana		82,00	
		Varianza		210,171	
		Desv. típ.		14,497	
		Mínimo		56	
		Máximo		99	
		Rango		43	
		Amplitud intercuartil		29	
		Asimetría		-,416	,580
		Curtosis		-,905	1,121
	Complutense	Media		79,00	3,257
		Intervalo de confianza para la media al 95%	Límite inferior	72,01	
			Límite superior	85,99	
		Media recortada al 5%		79,00	
		Mediana		78,00	
		Varianza		159,143	
		Desv. típ.		12,615	
		Mínimo		60	
		Máximo		98	
		Rango		38	
		Amplitud intercuartil		26	
		Asimetría		,241	,580
		Curtosis		-1,111	1,121

b) Contrastar la igualdad de medias poblaciones de las dos Universidades, con una fiabilidad del 90%

En el menú principal, Analizar/Comparar medias/Prueba T para muestras independientes...

En la ventana emergente se traspasa la variable Calificación a la columna Contrastar variables.
En variable de agrupación se pasa la variable Universidad definiendo los Grupos.
En Opciones se define el nivel de confianza (90%).



Estadísticos de grupo

	Universidad	N	Media	Desviación típ.	Error típ. de la media
Calificación	Autónoma	15	81,20	14,497	3,743
	Complutense	15	79,00	12,615	3,257

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	90% Intervalo de confianza para la diferencia	
									Inferior	Superior
Calificación	Se han asumido varianzas iguales	,152	,699	,443	28	,661	2,200	4,962	-6,241	10,641
	No se han asumido varianzas iguales			,443	27,475	,661	2,200	4,962	-6,246	10,646

La Prueba de Levene para la igualdad de varianzas indica si se puede o no suponer varianzas iguales.

Sí la probabilidad asociada al estadístico Levene es p-valor $= 0{,}699 > 0{,}05 = \alpha$ se asumen varianzas iguales.

Después de asumir las varianzas iguales e observa el estadístico t con su nivel de significación bilateral, este valor informa sobre el grado de compatibilidad entre la hipótesis nula de igualdad de medias:

$H_0 : \mu_{Autónoma} - \mu_{Complutense} = 0$

En este caso, p-valor $= 0{,}661 > 0{,}05 = \alpha \quad \rightarrow$ Se acepta la Hipótesis nula de igualdad de medias poblacionales.

El intervalo de confianza, con una fiabilidad del 90%, para la diferencia de medias poblacionales: $I(\mu_{Autónoma} - \mu_{Complutense}) = \left[-6{,}241\ ,\ 10{,}641\right]$ cubre el valor 0 , indicando que no existe diferencia significativa entre las medias poblacionales de las dos Universidades.

Por otra parte, $t_{28} = \dfrac{\left|\overline{x} - \overline{y}\right|}{s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{2{,}2}{4{,}962} = 0{,}443$

p-valor (Significación bilateral) $= 2.\,P\left[t_{28} \geq 0{,}443\right] = 2.0{,}3305 = 0{,}661$

CONTRASTE DE LA PROPORCIÓN DE UNA DISTRIBUCIÓN BINOMIAL

4. Se ignora la proporción de familias numerosas y con el fin de determinar dicha proporción se toma una muestra de 800 familias siendo la proporción observada de 0,18. Se puede afirmar que la proporción de familias numerosas es 0.20.

Solución:

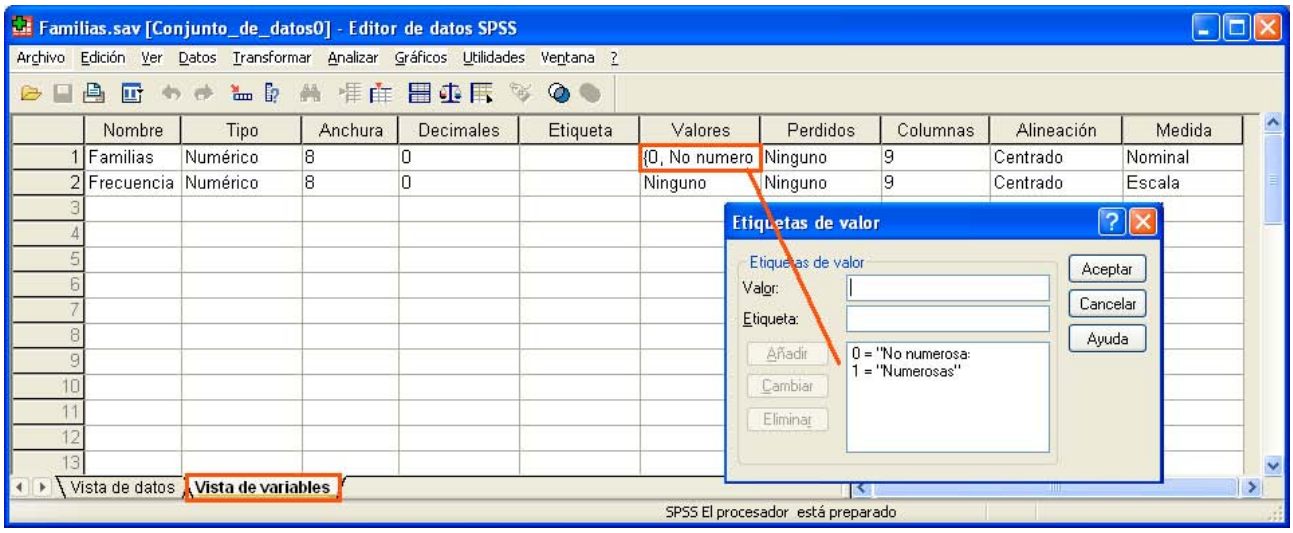
Se realiza un contraste bilateral para la proporción de una distribución Binomial. Es decir, se establecen las hipótesis: $H_0: p = 0{,}20 \quad H_a: p \neq 0{,}20$

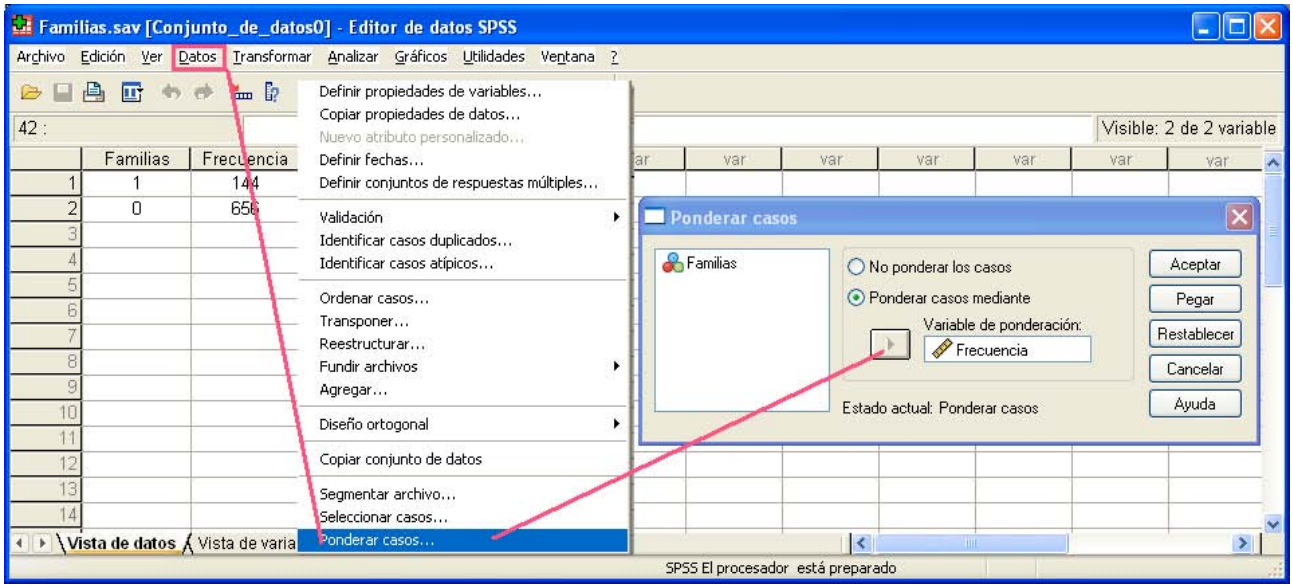
En una muestra de 800 familias, la proporción observada de familias numerosas es 0,18.

$800 \times 0{,}18 = 144$ **familias numerosas y** $800 - 144 = 656$ **familias no numerosas**

Familias.sav

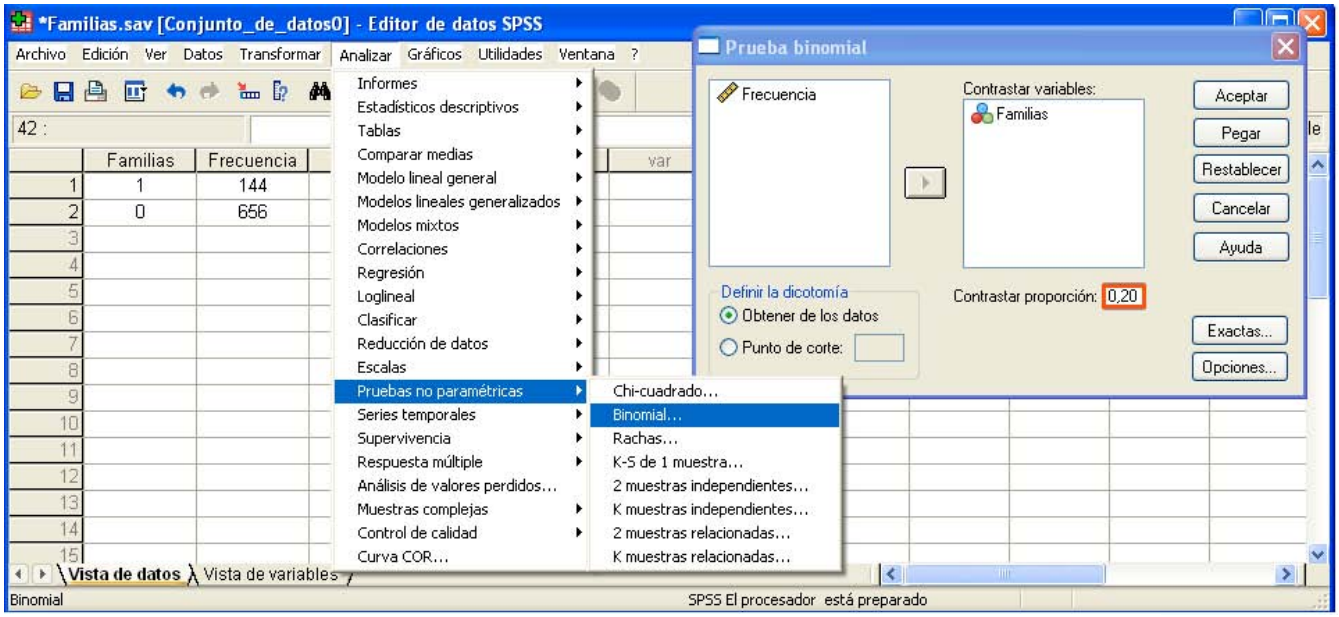
SPSS: Se define la variable y se ponderan los datos con la Frecuencia.





En el menú principal: **Analizar/Pruebas no paramétricas/Binomial.**

En la ventana que aparece se introduce Familias en Contrastar variables y 0,20 en Contrastar Proporción.



Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
Familias	Grupo 1	Numerosas	144	,2	,2	,084[a,b]
	Grupo 2	No numerosas	656	,8		
	Total		800	1,0		

a. La hipótesis alternativa establece que la proporción de casos del primer grupo sea < ,2.

b. Basado en la aproximación Z.

El p-valor de la prueba (Sig. asintótica unilateral) es 0,084 mayor que 0,05. En consecuencia, no se rechaza la Hipótesis nula, pudiendo afirmar que la proporción de familias numerosas es 0.20.

Se acepta H_0 si $z_{p_{\alpha/2}} = \dfrac{|\hat{p} - p_0|}{\sqrt{\dfrac{\hat{p}\,.\,(1-\hat{p})}{n}}} \leq z_{\alpha/2} \rightarrow z_{p_{\alpha/2}} = \dfrac{|0{,}18 - 0{,}20|}{\sqrt{\dfrac{0{,}18\,.\,0{,}82}{800}}} = 1{,}4724 \leq 1{,}96 = z_{0{,}025}$

p-valor $= \alpha_p = 1 - 2.\,P[z \geq 1{,}4724] = 0{,}084$ siendo $0{,}084 > 0{,}05$ se acepta H_0

CONTRASTE DE LA PROPORCIÓN DE UNA DISTRIBUCIÓN BINOMIAL

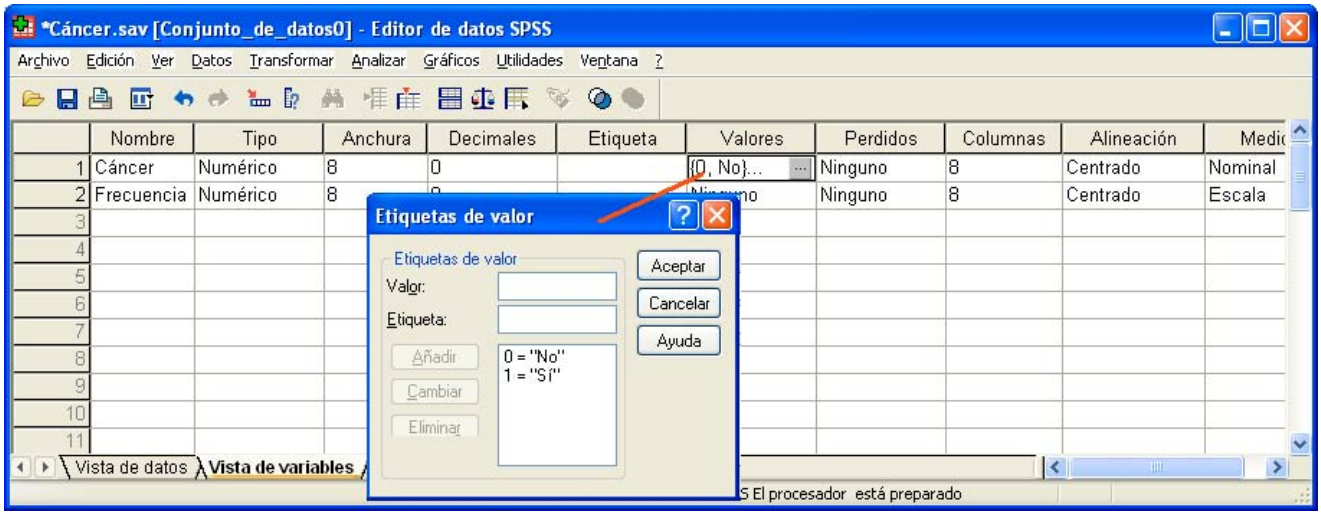
5. Entre los pacientes con cáncer de pulmón, el 90% o más muere generalmente en el espacio de tres años. Como resultado de nuevas formas de tratamiento, se cree que esta tasa se ha reducido.
En un reciente estudio sobre 150 pacientes diagnosticados de cáncer de pulmón, 128 murieron en el espacio de tres años.
¿Se puede afirmar que realmente ha disminuido la tasa de mortalidad?, con una significación de 0,05

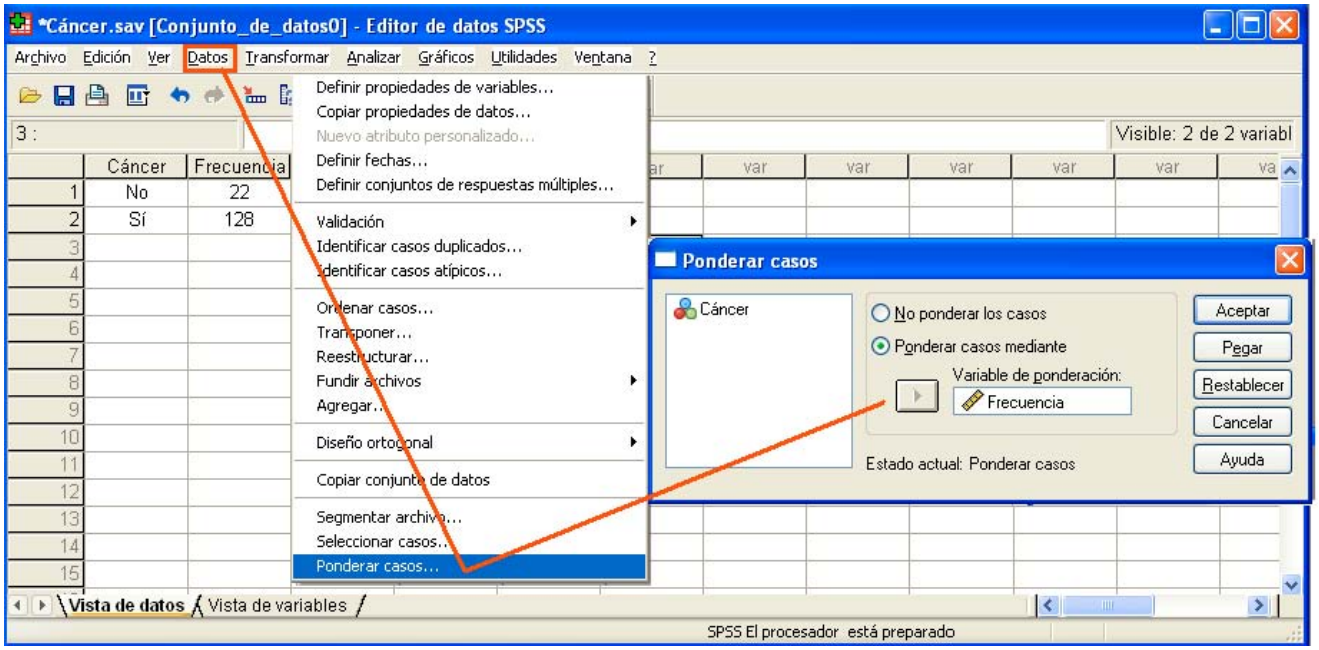
Solución:

Se realiza un contraste unilateral a la izquierda para la proporción de una distribución Binomial. Es decir, se establecen las hipótesis: $H_0: p \geq 0{,}90$ $H_a: p < 0{,}90$

Cáncer.sav

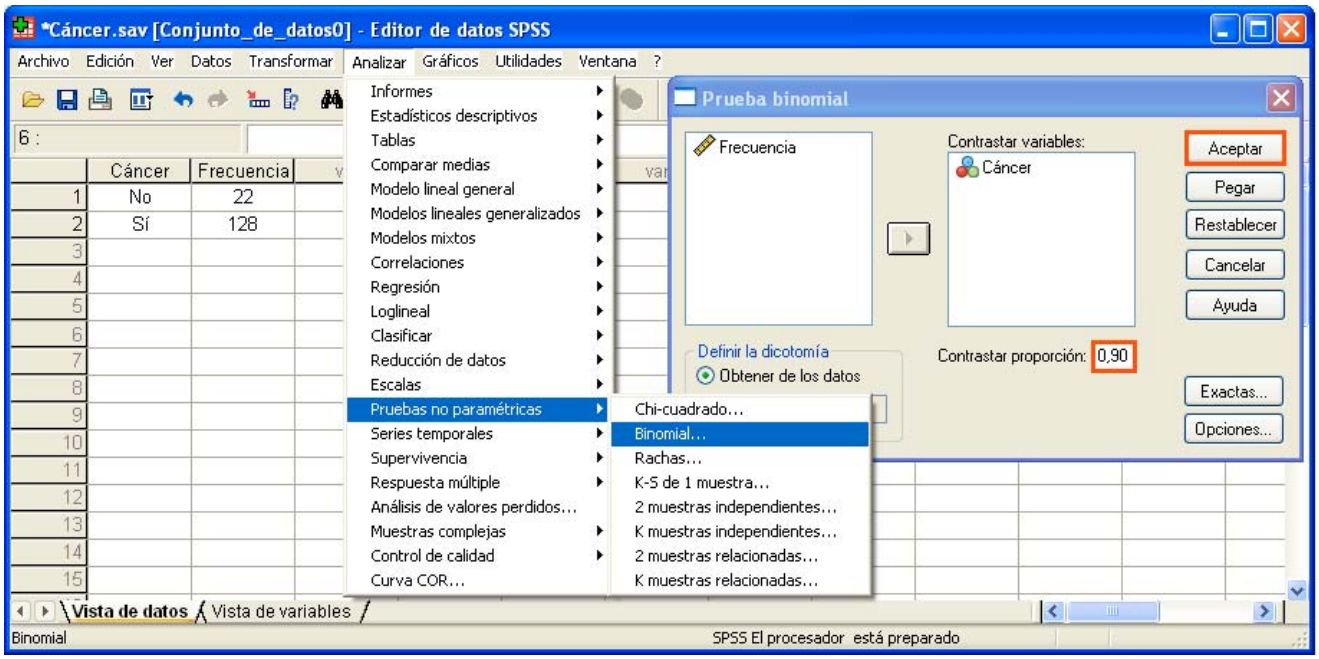
SPSS: Se define la variable y se ponderan los datos con la Frecuencia.





En el menú principal: Analizar/Pruebas no paramétricas/Binomial.

En la ventana que aparece se introduce Cáncer en Contrastar variables y 0,90 en Contrastar Proporción.



Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
Cáncer	Grupo 1	No	22	,1	,9	,000[a,b]
	Grupo 2	Sí	128	,9		
	Total		150	1,0		

a. La hipótesis alternativa establece que la proporción de casos del primer grupo sea < ,9.

b. Basado en la aproximación Z.

El p-valor del contraste (Sig. asintót. unilateral) es 0,000 < 0,05 , indicando que debe rechazarse la hipótesis nula.

En consecuencia, se puede afirmar que ha disminuido la tasa de mortalidad.

CONTRASTE DE LA PROPORCIÓN DE UNA DISTRIBUCIÓN BINOMIAL

6. De cien vástagos de un cierto cruce de conejos, quince son negros y los otros no lo son. De acuerdo con el modelo genético, éstos deberán estar en proporción 3/13 (negros/no negros). Se pide contratar la hipótesis de consistencia de los datos Estudiar el modelo para un nivel de significación de 0,05.

Solución:

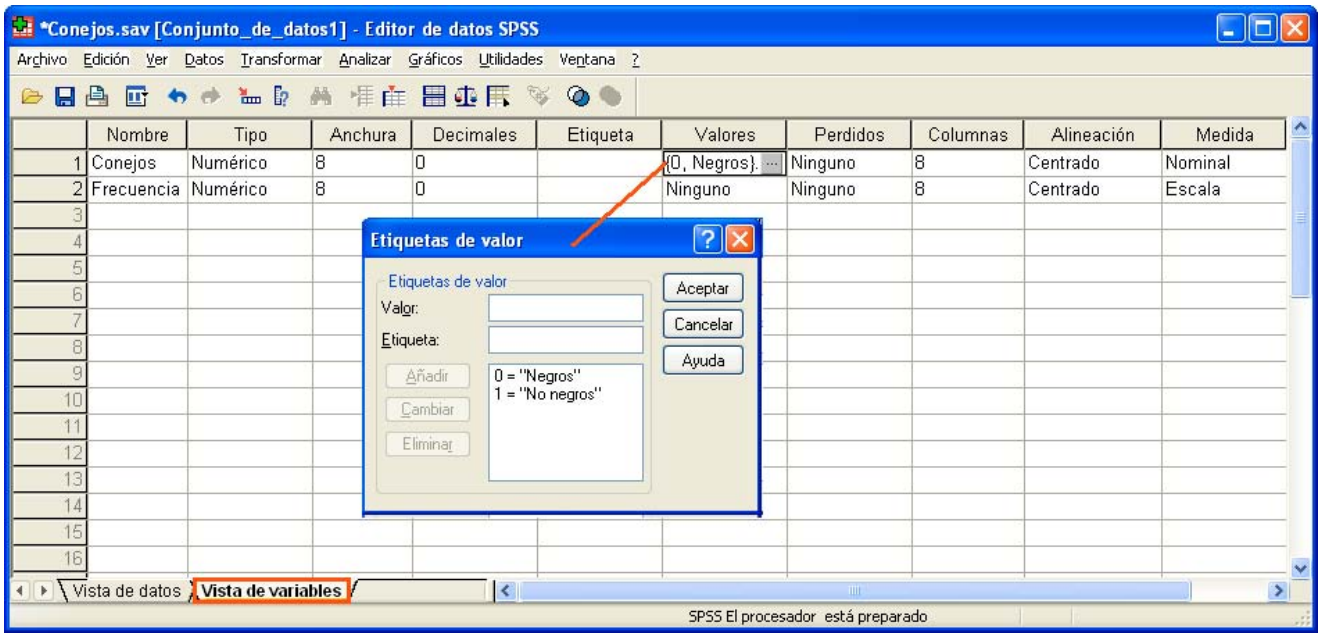
Es un modelo dicotómico, y, por tanto, frente a una distribución binomial $B(n, p)$, donde $n = 100$ y se trata de muestras grandes.

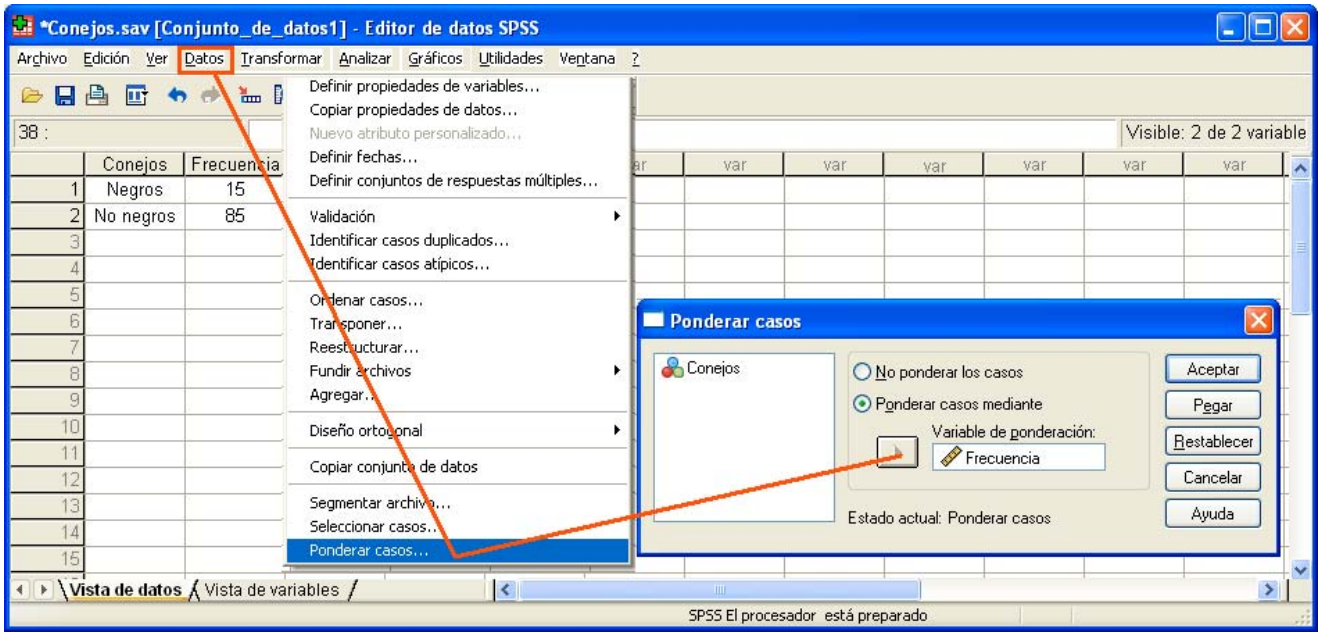
Para contrastar la hipótesis de consistencia de los datos se plantea el contraste bilateral:

$$H_0 : p = \frac{3}{13} \qquad H_a : p \neq \frac{3}{13}$$

Conejos.sav

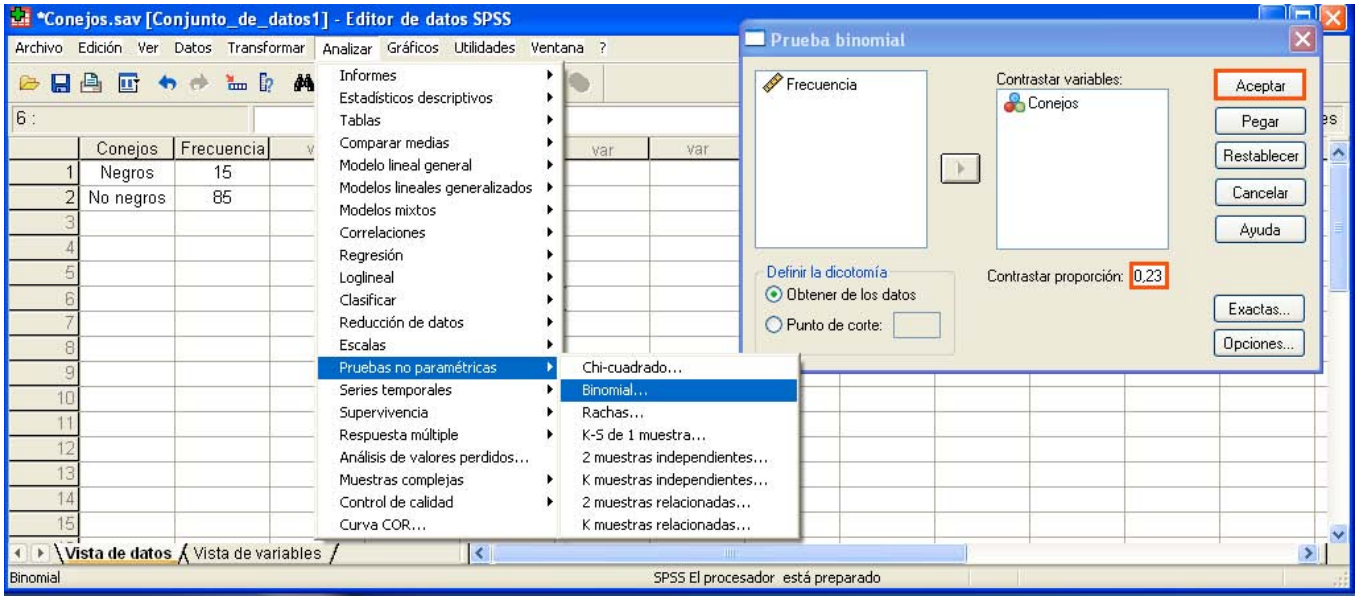
SPSS: Se define la variable y se ponderan los datos con la Frecuencia.





En el menú principal: Analizar/Pruebas no paramétricas/Binomial.

En la ventana que aparece se introduce Conejos en Contrastar variables y 0,23 en Contrastar Proporción.



Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
Conejos	Grupo 1	Negros	15	,150	,230	,033[a,b]
	Grupo 2	No negros	85	,850		
	Total		100	1,000		

a. La hipótesis alternativa establece que la proporción de casos del primer grupo sea < ,230

b. Basado en la aproximación Z.

El p-valor del contraste (Sig. asintót. unilateral) es 0,033 < 0,05 indicando que debe rechazarse la hipótesis nula.

En consecuencia, los datos son consistentes con el modelo genético a un nivel de significación 0,05.

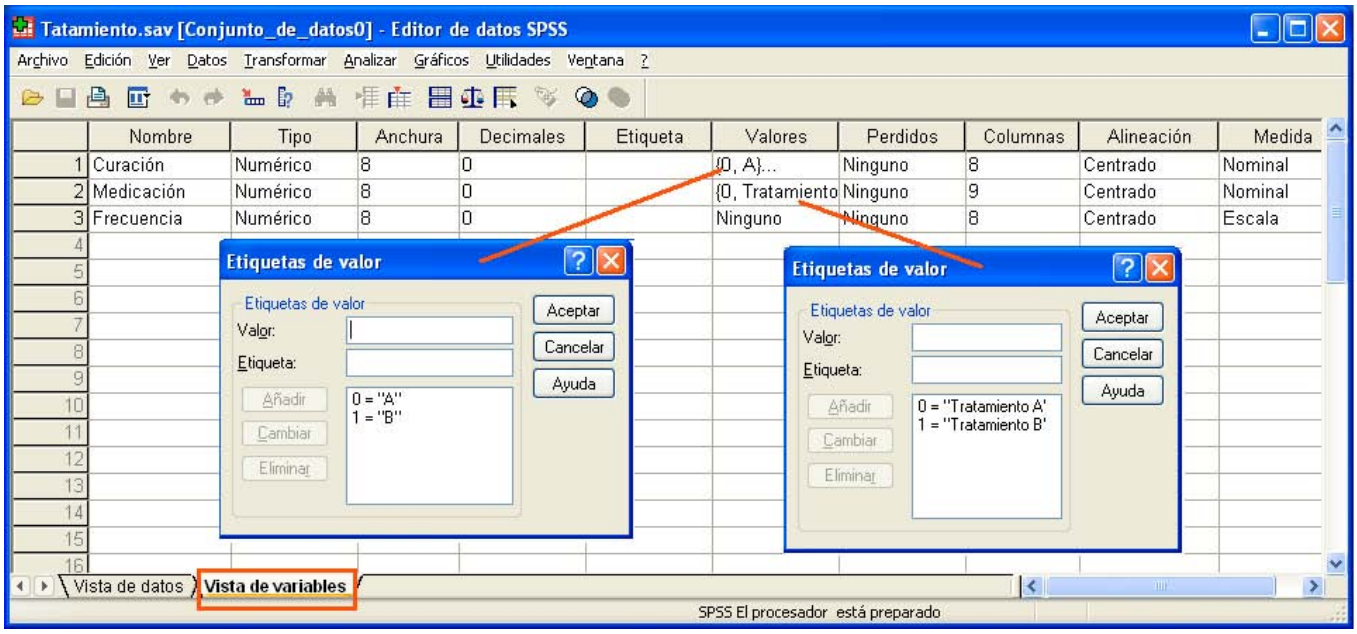
CONTRASTE PARA LA DIFERENCIA DE PROPORCIONES POBLACIONALES

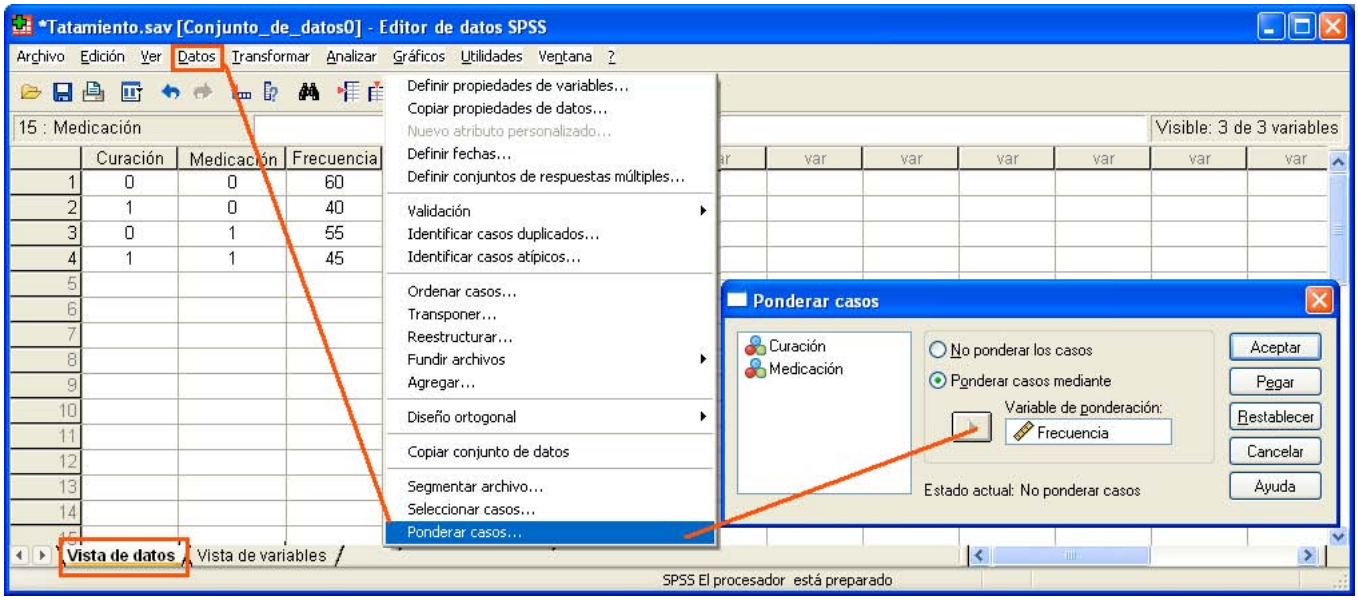
7. Un responsable médico afirma que los dos tipos de medicamentos (A y B) obtienen igual número de curaciones. Para confirmarlo, se eligen dos grupos de enfermos de 100 individuos cada uno.
En el grupo que se suministra el medicamento A se curan 60 enfermos y en el grupo que se suministra el medicamento B se curan 55 enfermos.
¿Tienen los dos tipos de tratamientos la misma efectividad en la curación de la enfermedad, con una confianza del 95%?

Solución:

Se realiza un contraste bilateral para la igualdad de los parámetros de dos distribuciones binomiales. Es decir, se establecen las hipótesis: $H_0: p_A - p_B = 0 \quad H_a: p_A - p_B \neq 0$

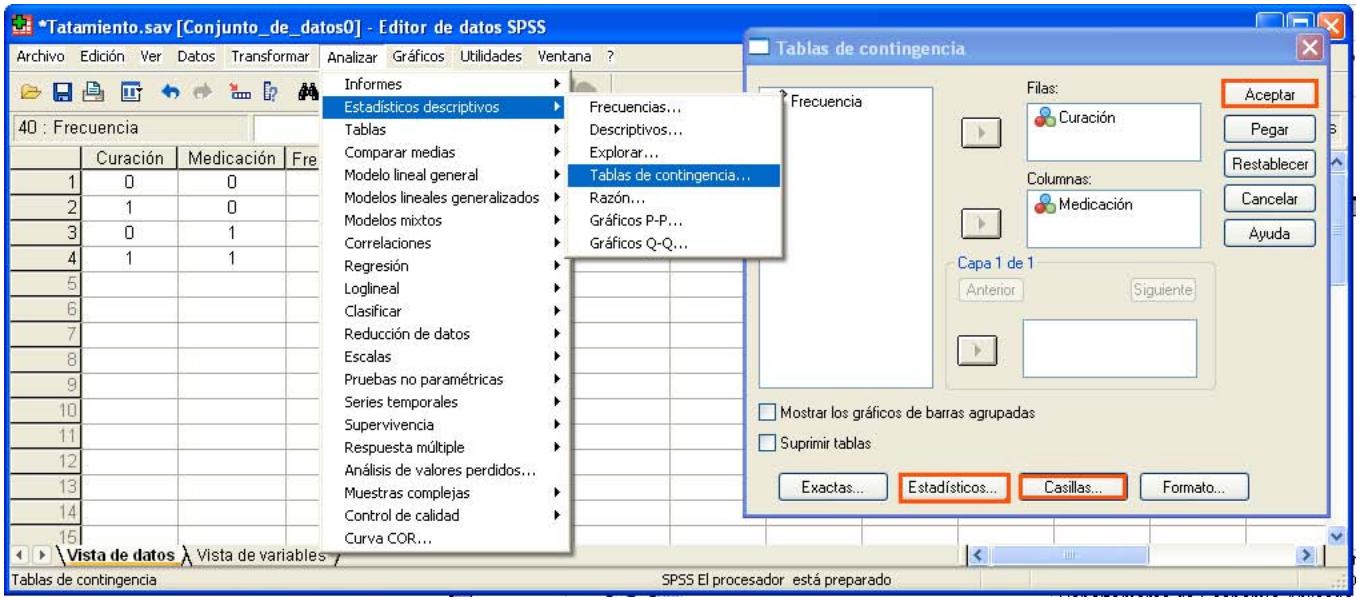
SPSS: En "Vista de variable" se definen las variables y se introducen sus valores. Posteriormente, en "Vista de datos" se introducen los datos y se ponderan con la "Frecuencia".



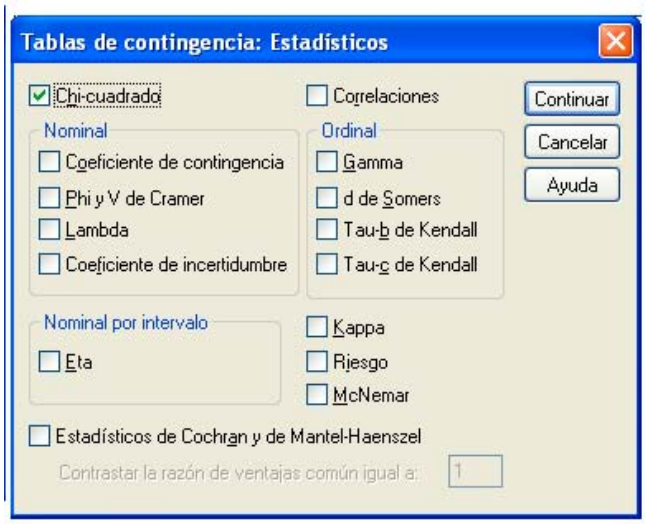


En el menú principal: Analizar/Estadísticos descriptivos/Tablas de contingencia...

En la ventana emergente se seleccionan las variables dicotómicas que se van a contrastar. La variable "Curación" en filas y la variable "Medicación" en columnas.



En Estadísticos se selecciona "Chi-cuadrado" y en Casillas las opciones "Observadas" y Porcentajes en "Columna".



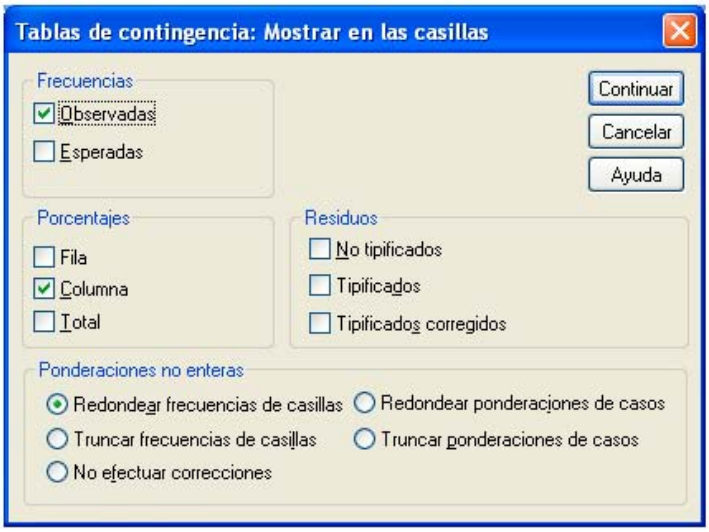


Tabla de contingencia Curación * Medicación

			Medicación		
			Tratamiento A	Tratamiento B	Total
Curación	A	Recuento	60	55	115
		% de Medicación	60,0%	55,0%	57,5%
	B	Recuento	40	45	85
		% de Medicación	40,0%	45,0%	42,5%
Total		Recuento	100	100	200
		% de Medicación	100,0%	100,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,512[b]	1	,4745		
Corrección por continuidad[a]	,327	1	,5672		
Razón de verosimilitudes	,512	1	,4744		
Estadístico exacto de Fisher				,56734	,28367
Asociación lineal por lineal	,509	1	,4756		
N de casos válidos	200				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 42,50.

Se elige el estadístico Chi-cuadrado", que tiene menor nivel de significación, con un p-valor = 0,4745 , siendo p-valor = 0,4745 > 0,05 = α se acepta la hipótesis nula $H_0: p_A - p_B = 0 \rightarrow H_0: p_A = p_B$ que establece que la efectividad de los dos tratamientos es igual, con un nivel de confianza del 95%.

De otra parte, se acepta H_0 sí $z_{p_{\alpha/2}} = \dfrac{|\hat{p}_A - \hat{p}_B|}{\sqrt{\dfrac{\hat{p}_A \cdot \hat{q}_A}{n_A} + \dfrac{\hat{p}_B \cdot \hat{q}_B}{n_B}}} \leq z_{\alpha/2}$

Se acepta H_0: $z_{p_{\alpha/2}} = \dfrac{|0,6 - 0,55|}{\sqrt{\dfrac{0,6 \,.\, 0,4}{100} + \dfrac{0,55 \,.\, 0,45}{100}}} = 0,7163 \leq 1,96 = z_{0,025}$

$\alpha_p = 2.\, P[z \geq 0,7163] = 2 \,.\, 0,23725 = 0,4745$ **siendo** $\alpha_p = 0,4745 > 0,05 = \alpha$ **se acepta** H_0

⊗ **El Estadístico exacto de Fisher presenta un p-valor (Sig. asintótica bilateral) = 0,56734**

- **Si el contraste fuera unilateral a la derecha** $H_0: p_A - p_B \leq 0 \quad H_a: p_A - p_B > 0$

  p-valor (Sig. exacta unilateral de la derecha) $= \dfrac{0,56734}{2} = 0,28367$

- **Si el contraste fuera unilateral a la izquierda** $H_0: p_A - p_B \geq 0 \quad H_a: p_A - p_B < 0$

  p-valor (Sig. exacta unilateral de la izquierda) $= 1 - \dfrac{0,56734}{2} = 0,71633$

CONTRASTE PARA LA DIFERENCIA DE PROPORCIONES POBLACIONALES

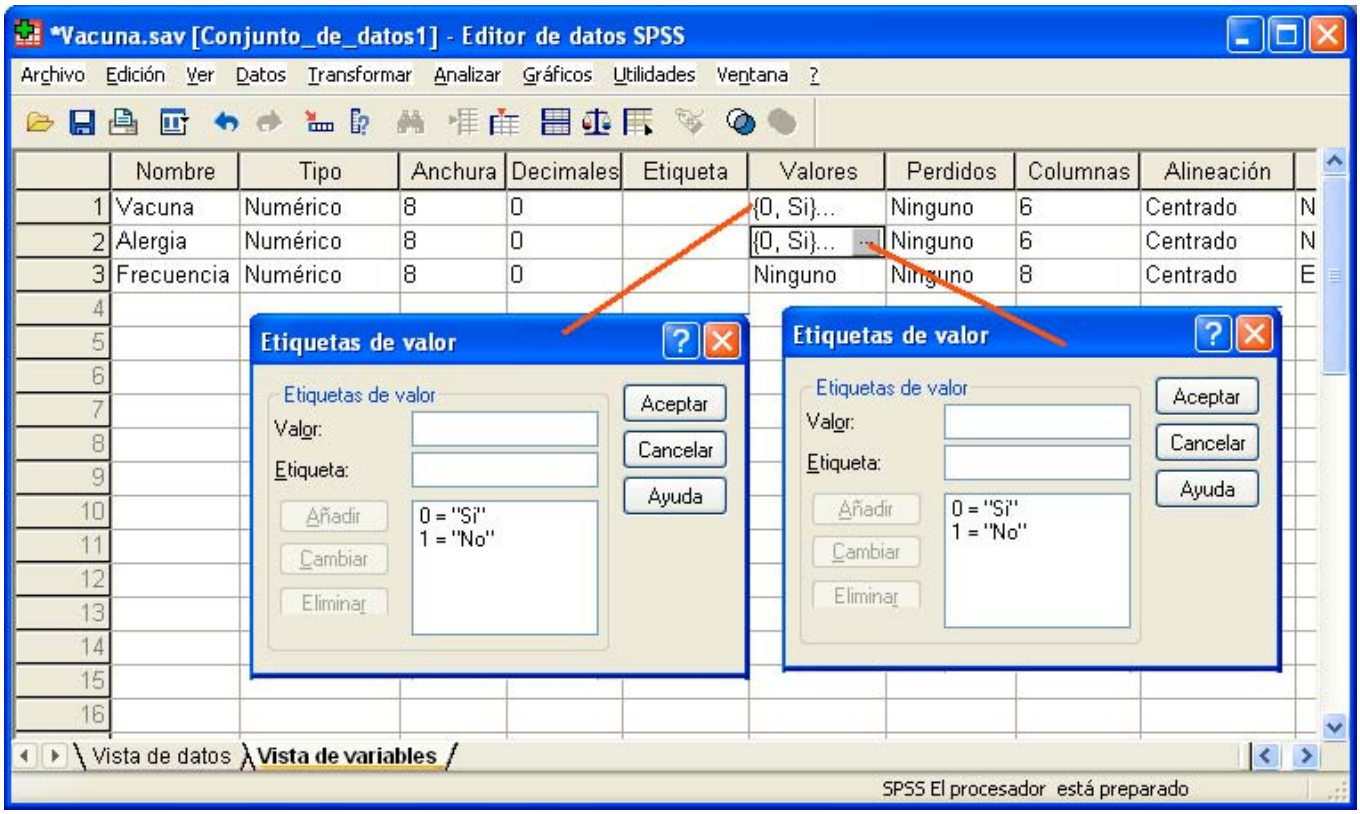
8. Se quiere comprobar la efectividad de una vacuna contra una alergia. Para ello se suministró la vacuna a cien pacientes y se les comparó con un grupo testigo de cien pacientes afectados también por la misma alergia en épocas anteriores. Entre los vacunados, ocho sufrieron alergia y entre los no vacunados veinticinco volvieron a sufrir alergia.
¿Se puede afirmar que la vacuna es eficaz en disminuir la alergia?. Utilizar un nivel de significación de 0,05.

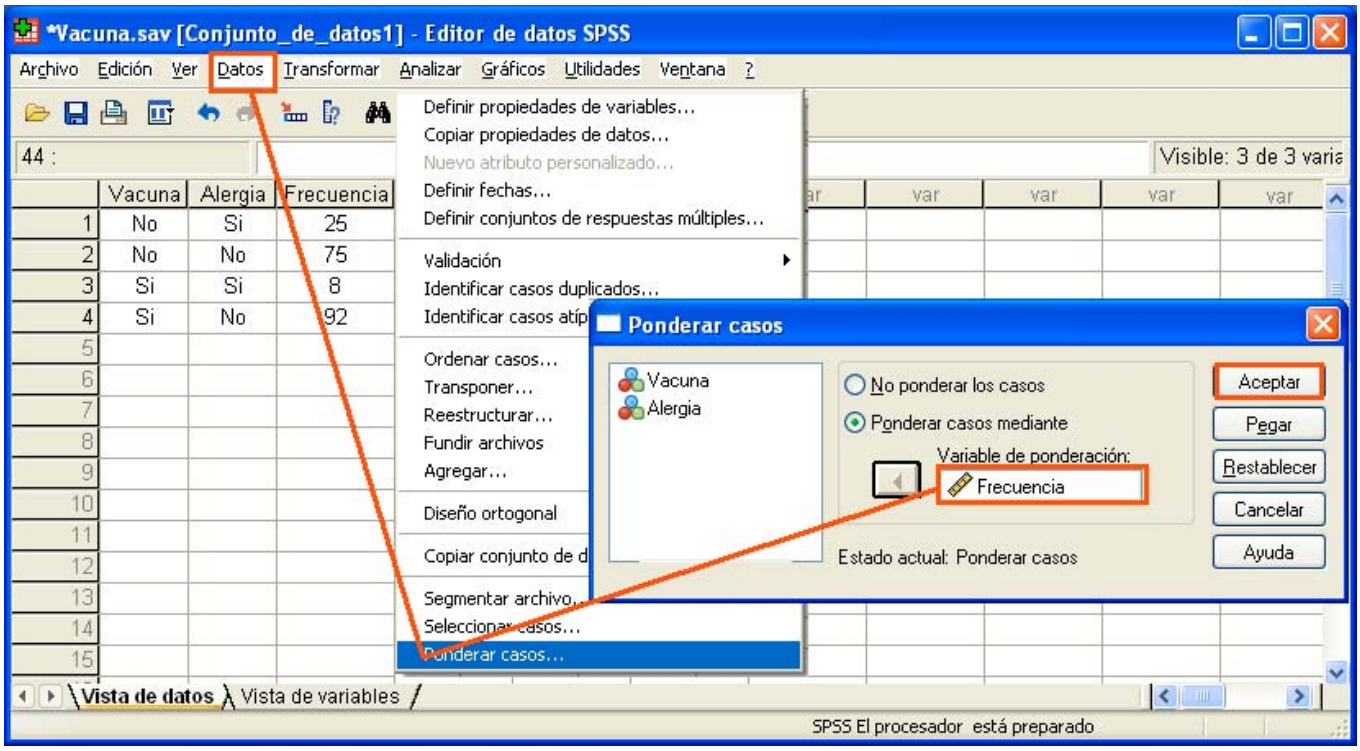
Solución:

Se establece el modelo: Los pacientes vacunados afectados por la alergia es una variable binomial $B(n_1, p_1)$ y el número de pacientes no vacunados es otra binomial $B(n_2, p_2)$, en donde $n_1 = n_2 = 100$
Para ver si la vacuna es eficaz hay que contrastar $p_2 \leq p_1$.

Se realiza un contraste unilateral a la derecha: $H_0: p_2 - p_1 \leq 0$ $\quad H_a: p_2 - p_1 > 0$

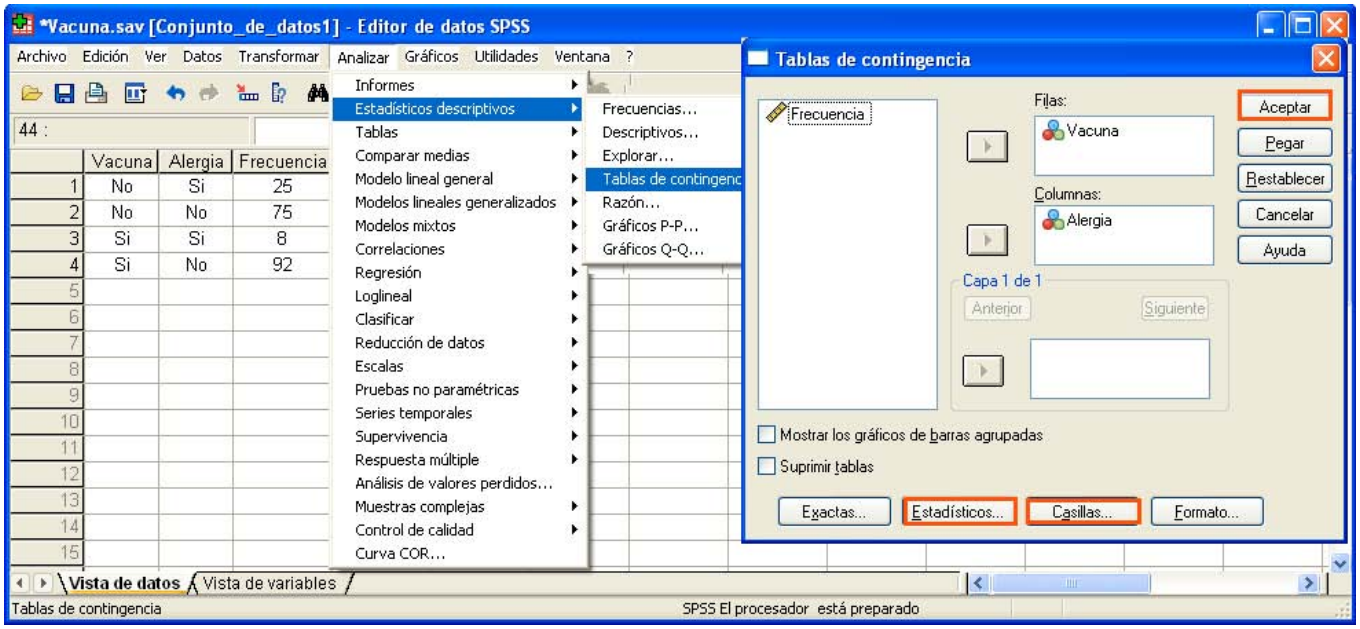
SPSS: En "Vista de variable" se definen las variables y se introducen sus valores. Posteriormente, en "Vista de datos" se introducen los datos y se ponderan con la "Frecuencia".



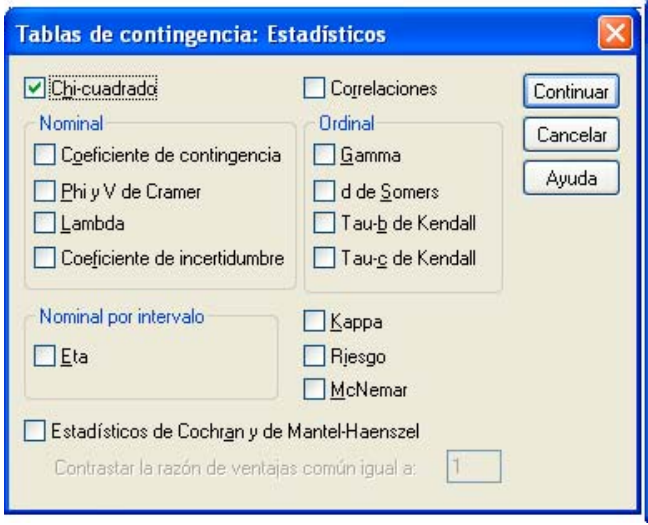


En el menú principal: Analizar/Estadísticos descriptivos/Tablas de contingencia...

En la ventana emergente se seleccionan las variables dicotómicas que se van a contrastar. La variable "Curación" en filas y la variable "Medicación" en columnas.



En Estadísticos se selecciona "Chi-cuadrado" y en Casillas las opciones "Observadas" y Porcentajes en "Columna".



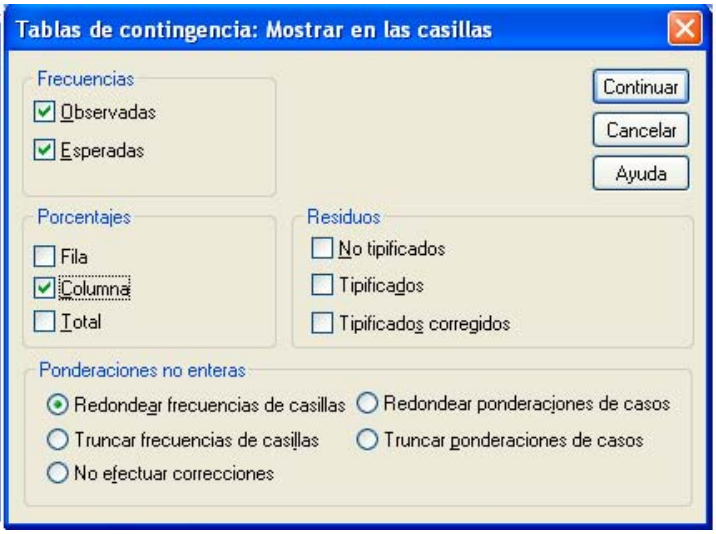


Tabla de contingencia Vacuna * Alergia

			Alergia		Total
			Si	No	
Vacuna	Si	Recuento	8	92	100
		Frecuencia esperada	16,5	83,5	100,0
		% de Alergia	24,2%	55,1%	50,0%
	No	Recuento	25	75	100
		Frecuencia esperada	16,5	83,5	100,0
		% de Alergia	75,8%	44,9%	50,0%
Total		Recuento	33	167	200
		Frecuencia esperada	33,0	167,0	200,0
		% de Alergia	100,0%	100,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	10,488[b]	1	,001		
Corrección por continuidad[a]	9,291	1	,002		
Razón de verosimilitudes	10,927	1	,001		
Estadístico exacto de Fisher				,002	,001
Asociación lineal por lineal	10,436	1	,001		
N de casos válidos	200				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 16,50.

Se elige el estadístico Chi-cuadrado", que tiene menor nivel de significación, con un p-valor $= 0{,}001$, siendo p-valor $= 0{,}001 < 0{,}05 = \alpha$ se rechaza la hipótesis nula del contraste bilateral

$H_0 : p_1 - p_1 = 0 \rightarrow H_0 : p_2 = p_1$ que establece que con la vacuna la proporción de pacientes afectados es la misma.

- **Con el contraste fuera unilateral a la derecha: $H_0 : p_2 - p_1 \leq 0 \quad H_a : p_2 - p_1 > 0$**

  p-valor (Sig. exacta unilateral de la derecha) $= \dfrac{0{,}001}{2} = 0{,}0005$

p-valor $= 0{,}0005 < 0{,}05 = \alpha$ se rechaza la hipótesis nula, aceptando por tanto la hipótesis $H_a : p_2 > p_1$, afirmando que la vacuna es eficaz, con una confianza del 95%.

De otra parte, se acepta H_0 sí $z_p = \dfrac{\hat{p}_2 - \hat{p}_1}{\sqrt{\dfrac{\hat{p}_2 \cdot \hat{q}_2}{n_2} + \dfrac{\hat{p}_1 \cdot \hat{q}_1}{n_1}}} \leq z_\alpha$

$$z_p = \frac{0{,}25 - 0{,}08}{\sqrt{\dfrac{0{,}25 \cdot 0{,}75}{100} + \dfrac{0{,}08 \cdot 0{,}92}{100}}} = 3{,}4 > 1{,}65 = z_{0{,}025}$$

Se rechaza la hipótesis nula $H_0 : p_2 \leq p_1$ y se puede concluir que la vacuna es eficaz

- Llevando la información a una tabla de contingencia de 2×2

Se calculan los valores de la χ^2 correspondientes a las dos observaciones, siendo la frecuencia esperada

$$e_{ij} = p_{ij} \times n = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$$

$e_{11} = \dfrac{100 \cdot 33}{200} = 16{,}5$ $\quad e_{12} = \dfrac{100 \cdot 167}{200} = 83{,}5$ $\quad e_{21} = \dfrac{100 \cdot 33}{200} = 16{,}5$ $\quad e_{22} = \dfrac{100 \cdot 167}{200} = 83{,}5$

Vacunados	Alergia		Total
	Sí	No	
Sí	$n_{11} = 8$ $e_{11} = 16{,}5$	$n_{12} = 92$ $e_{12} = 83{,}5$	$n_{1\bullet} = 100$ 100
No	$n_{21} = 25$ $e_{21} = 16{,}5$	$n_{22} = 75$ $e_{22} = 83{,}5$	$n_{2\bullet} = 100$ 100
Total	$n_{\bullet 1} = 33$	$n_{\bullet 2} = 167$	$n = 200$

$$\chi_1^2 = \frac{n \times (n_{11} \times n_{22} - n_{12} \times n_{21})^2}{n_{1\bullet} \times n_{2\bullet} \times n_{\bullet 1} \times n_{\bullet 2}} = \frac{200 \times (8 \times 75 - 92 \times 25)^2}{100 \times 100 \times 33 \times 167} = 10{,}488$$

O bien, $\chi^2_{(2-1).(2-1)} = \chi_1^2 = \displaystyle\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{n_{ij}^2}{e_{ij}} - n = \frac{8^2}{16{,}5} + \frac{92^2}{83{,}5} + \frac{25^2}{16{,}5} + \frac{75^2}{83{,}5} - 200 = 10{,}488$

Como $\chi_1^2 = 10{,}488 > 3{,}841 = \chi^2_{0{,}05,\,1}$ se rechaza la hipótesis nula $H_0 : p_2 = p_1$, concluyendo que la vacuna es eficaz.

Por otra parte, p-valor $= \alpha_p = P\left(\chi_1^2 > 10{,}488\right) = 0{,}001287$

Como $\alpha_p = 0{,}001287 < 0{,}05 = \alpha \rightarrow$ Se rechaza $H_0 : p_2 = p_1$

CONTRASTE PARA LA DIFERENCIA DE MEDIAS DE DOS DISTRIBUCIONES NORMALES APAREADAS.

9. Se quiere probar la efectividad de un antitérmico en reducir la fiebre. Con tal fin se tomó la temperatura a diez niños, antes o inmediatamente después de la administración del antitérmico, y los resultados fueron:

Temperatura antes	39,3	39,7	39,9	40,0	38,9	39,7	39,3	39,6	39,9	40,0
Temperatura después	38,1	38,0	38,3	38,3	37,9	38,7	38,3	37,0	36,9	39,0

Suponiendo que la característica en estudio sea normal. Contrastar si el antitérmico reduce la fiebre en los niños, con un nivel de confianza del 90%.

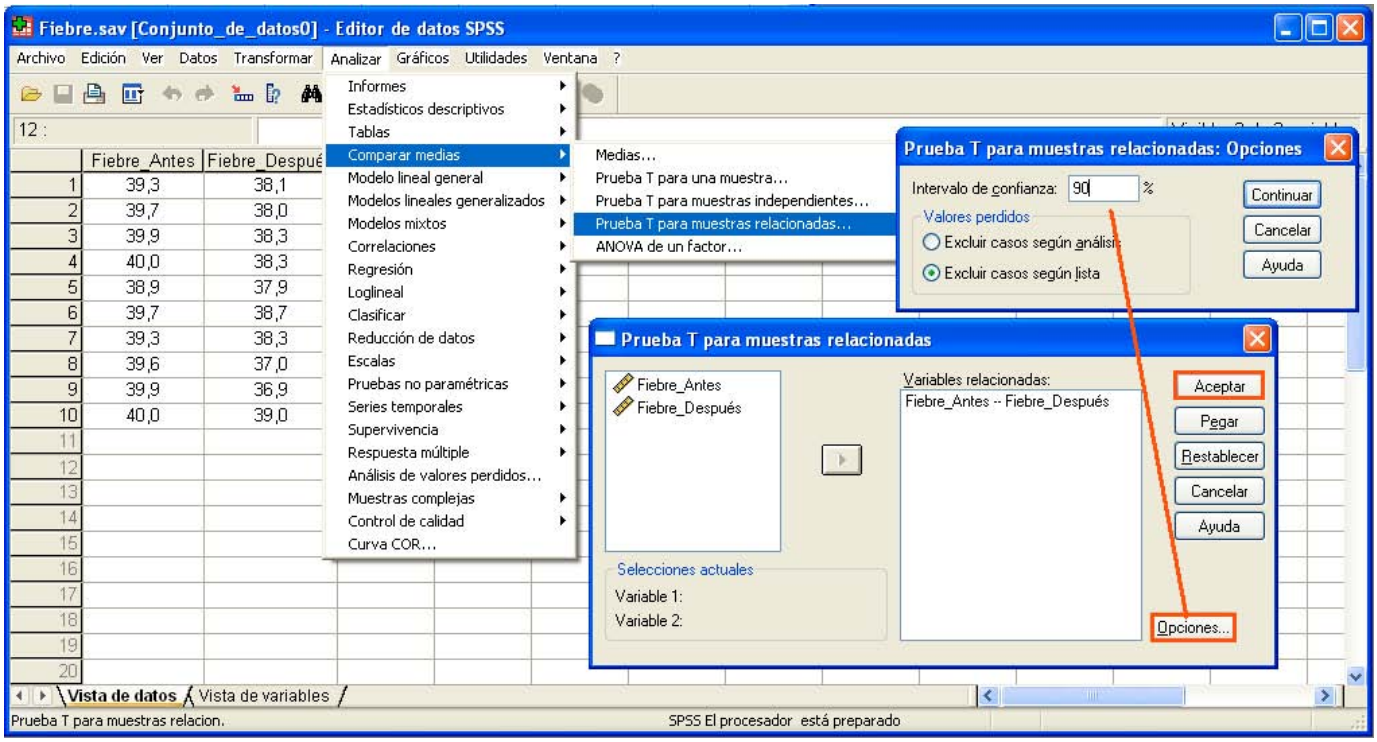
Solución:

Es un experimento de datos apareados. Se establece las hipótesis de las diferencias:

Contraste unilateral a la derecha $H_0 : d = \mu_a - \mu_d \leq 0$ $H_1 : d = \mu_a - \mu_d > 0$

Fiebre.sav

Menú principal: Analizar/Comparar medias/Prueba T para muestras relacionadas...



Correlaciones de muestras relacionadas

		N	Correlación	Sig.
Par 1	Fiebre_Antes y Fiebre_Después	10	,132	,717

El coeficiente de correlación $r = 0{,}132$ tiene un p-valor $= 0{,}717 > 0{,}05$, aceptando la hipótesis nula $H_0 : r = 0$ por lo que no existe una correlación significativa entre las variables.

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	90% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Fiebre_Antes - Fiebre_Después	1,5800	,7131	,2255	1,1667	1,9933	7,007	9	,000

El valor experimental del estadístico de contraste (t = 7,007) tiene un α_p = p-valor (Sig. bilateral) de 0,000. Como se trata de un contraste unilateral a la derecha $\alpha_{p/2} = \frac{0,000}{2} = 0,000 < 0,01$ con lo que se rechaza la hipótesis nula, el antitérmico no reduce la fiebre.

CONTRASTE PARA LA DIFERENCIA DE MEDIAS DE DOS DISTRIBUCIONES NORMALES APAREADAS.

10. Una empresa farmacéutica está interesada en la investigación preliminar de un nuevo medicamento que parece tener propiedades reductoras del colesterol en sangre. A tal fin se toma una muestra al azar de seis personas comparables, y se determina el contenido en colesterol antes y después del tratamiento. Los resultados han sido los siguientes:

Antes	217	252	229	200	209	213
Después	209	241	230	208	206	211

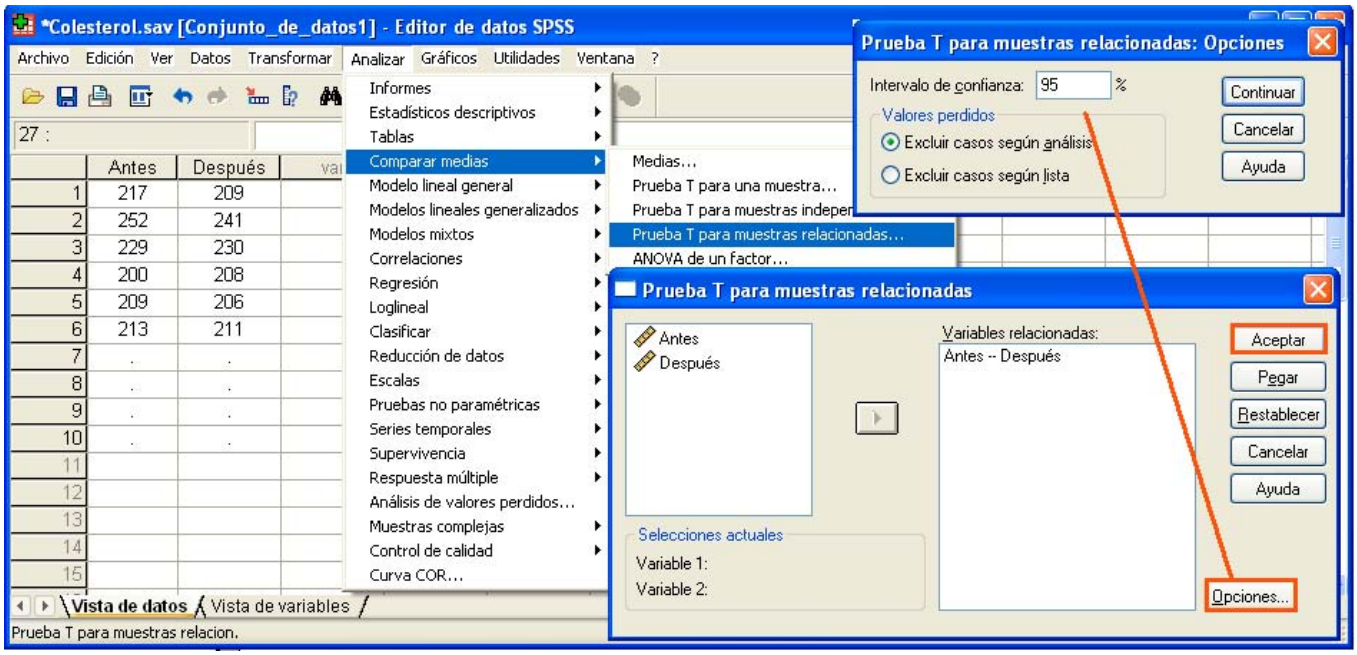
Suponiendo que la característica en estudio sea normal. Con un nivel de confianza del 95%, ¿se puede afirmar estadísticamente la bondad del medicamento?.

Solución:

Es un experimento de datos apareados. Se establece las hipótesis de las diferencias:

Contraste unilateral a la derecha $H_0 : d = \mu_a - \mu_d = 0$ $H_1 : d = \mu_a - \mu_d \neq 0$

Menú principal: Analizar/Comparar medias/Prueba T para muestras relacionadas...



Correlaciones de muestras relacionadas

		N	Correlación	Sig.
Par 1	Antes y Después	6	,944	,005

El coeficiente de correlación $r = 0{,}944$ tiene un p-valor $= 0{,}005 < 0{,}05$, aceptando la hipótesis nula $H_0 : r = 0$ por lo que no existe una correlación significativa entre las variables.

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Antes - Después	2,500	6,716	2,742	-4,548	9,548	,912	5	,404

El valor experimental del estadístico de contraste (t = 0,912) tiene un α_p = p-valor (Sig. bilateral) de 0,404. Siendo $\alpha_p = 0,404 > 0,05$ se acepta la hipótesis nula, concluyendo que estadísticamente no se puede afirmar la bondad del medicamento en la reducción del colesterol.

Tratamiento de Datos
Anova de un Factor con Medidas repetidas.
Análisis de la Varianza con varios factores.
Análisis Multivariable

El Análisis de la varianza con varios factores presenta una media teórica definida por tres términos
Análisis de la Varianza con varios factores

ANÁLISIS VARIANZA

En cualquier experimento se pueden distinguir dos fases. Una primera, que consiste en la observación y análisis de los hechos que acontecen, y otra segunda, que interpreta y obtiene conclusiones.
La Estadística Descriptiva se encarga de dar una descripción numérica, ordenar y simplificar la información recogida en la primera fase.
La Estadística Teórica, con la información obtenida en el muestreo de la primera fase, pretende generalizar o predecir los resultados para toda la población.

AF
ACP
Análisis Multivariante
El Análisis de Componentes Principales es una técnica de reducción de la dimensionalidad.
Actualmente es el método más utilizado en la investigación social y comercial.
El Análisis de Componentes Principales (ACP) tiene como objetivo simplificar la estructura de los datos transformando las variables en análisis en el menor número posible de factores comunes o componentes principales que sean combinaciones lineales de las variables.

El índice de Gini verifica:
Principio del Anonimato
Principio de la Renta Relativa
Principio de Dalton
Principio de la Población
La curva de Lorenz es la representación de la concentración de recursos entre individuos.
Desigualdad
Curva Lorenz
Igualdad
Rentas
Población
El índice de Gini expresa el grado de concentración de los recursos.
$I_G = 1 - \frac{\sum_{i=1}^{n-1} q_i}{\sum_{i=1}^{n} p_i}$
El cambio de escala no afecta al índice
El cambio de origen afecta al índice

Estadística Aplicada

Análisis Multivariante
INVESTIGACIÓN en Cómic
Análisis Factorial
Componentes Principales (ACP)
Análisis Cluster
SPSS