

Distribución χ^2 Pearson

Estimación y Contraste

- Varianza poblacional σ^2

X_1, X_2, \dots, X_n v.a $N(0,1)$

$$\chi_{n-1}^2 = \frac{(n-1)s_x^2}{\sigma_x^2}$$

Distribución Normal

Estimación y Contrastes

- Media, Diferencia Medias
- Proporción p
- Diferencia Proporciones
- Parámetro λ de Poisson

Hay muchas variables asociadas a fenómenos naturales que siguen una distribución normal; además, es una buena aproximación de otras distribuciones (Binomial, Poisson, etc)

Normal tipificada
 $z \sim N(0, 1)$

$$z = \frac{X - \mu}{\sigma}$$

$X \sim N(\mu, \sigma)$

Distribución F Fisher-Snedecor

Estimación y Contraste

- Igualdad varianzas población

χ_1^2 y χ_2^2 v.a χ^2 de Pearson

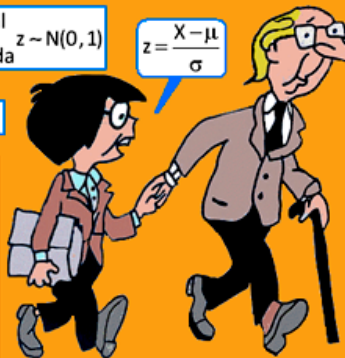
$$F_{n_1, n_2} = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$$

Distribución t de Student

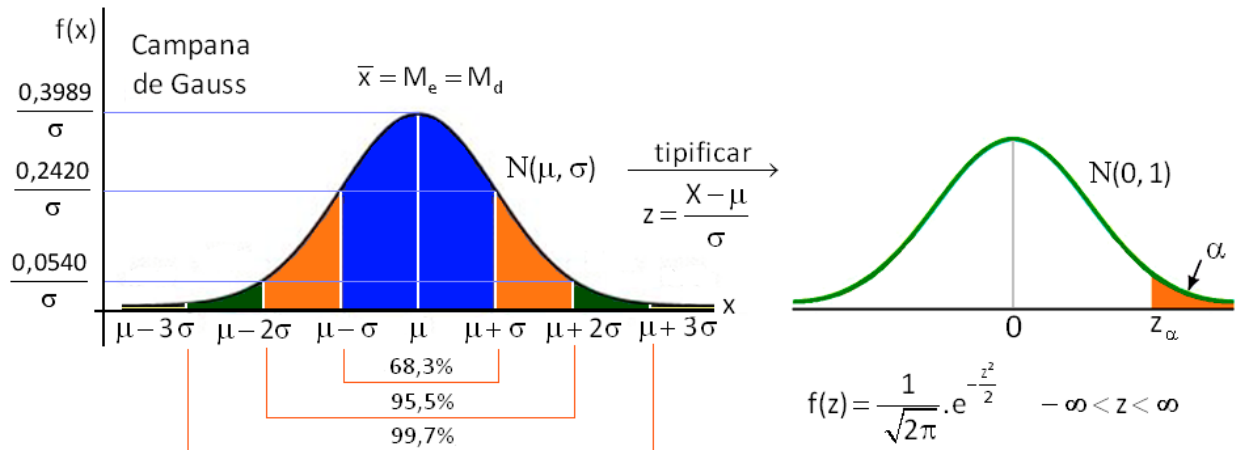
Estimación y Contrastes

- Media ($n \leq 30$)
- Diferencias medias ($n \leq 30$)
- Datos apareados

$$t_{n-1} = \frac{\bar{x} - \mu}{s_x / \sqrt{n}} = \frac{\bar{x} - \mu}{\sigma_x / \sqrt{n-1}}$$



DISTRIBUCIÓN NORMAL



El área bajo la curva surge de la función de densidad:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \mapsto \quad \text{Área} = \int_{-\infty}^{\infty} f(x) dx = 1$$

Función de distribución:

$$F(x) = P(X \leq x_0) = \int_{-\infty}^x f(x) \cdot dx = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot dx \quad -\infty < x < \infty$$

- Si una variable aleatoria X es $N(\mu_1, \sigma_1)$ y otra variable Y es $N(\mu_2, \sigma_2)$ independientes entre sí, entonces la nueva variable $W = X \pm Y$ sigue también una distribución normal $N(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

La propiedad se puede generalizar a n variables aleatorias independientes.

Sean n variables aleatorias independientes (X_1, X_2, \dots, X_n) que siguen

una distribución normal $N(\mu, \sigma)$, entonces $\sum_{i=1}^n X_i$ sigue también una

distribución normal $N(n \cdot \mu, \sqrt{n \cdot \sigma^2})$

- Sea X una variable aleatoria binomial de parámetros n y p , cuando n es grande y ni p ni q son próximos a cero, se puede considerar que X sigue aproximadamente una distribución normal $N(n \cdot p, \sqrt{n \cdot p \cdot q})$

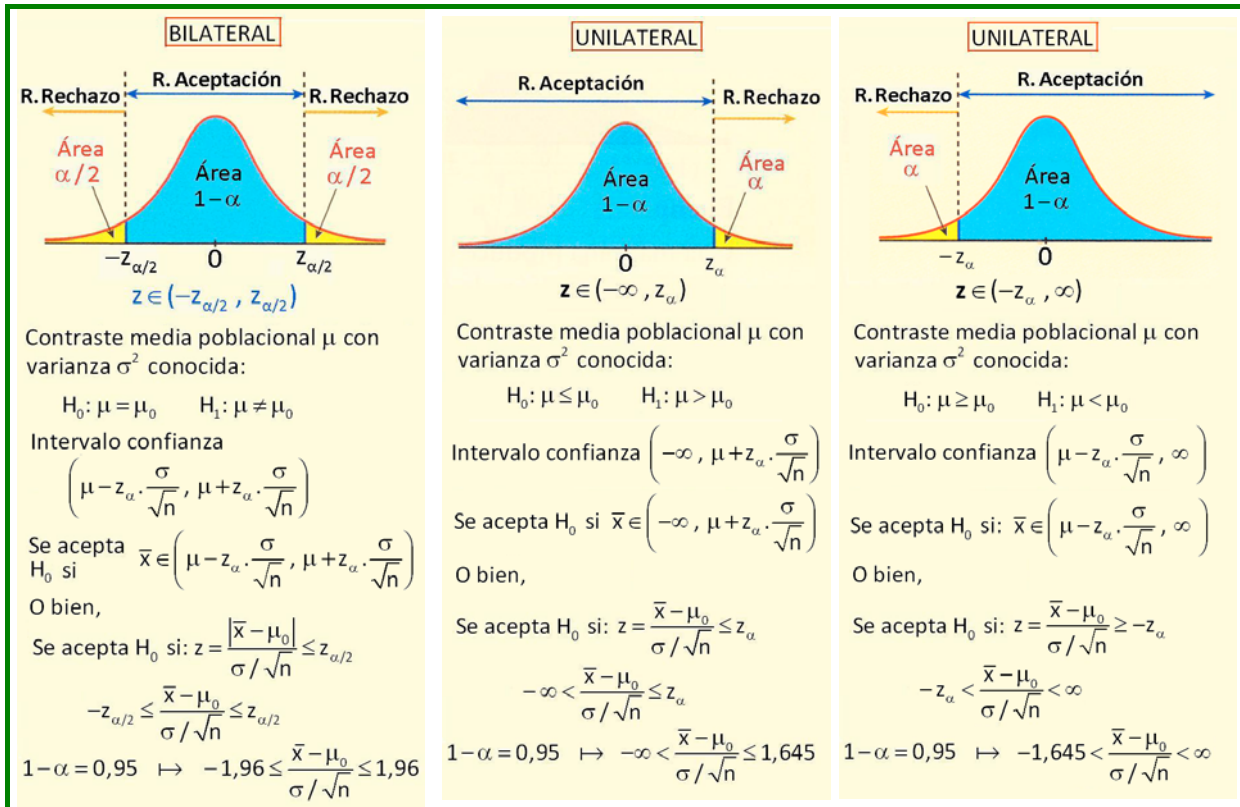
Por el *teorema de Moivre*: $z = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot q}} \quad z \sim N(0, 1)$

Para utilizar correctamente esta transformación de una variable aleatoria discreta X (con distribución binomial) en una variable continua z (con distribución normal) es necesario hacer una corrección de continuidad.

$$P(X = a) = P(a - 0,5 \leq X \leq a + 0,5)$$

$$P(a < X < b) = P(a + 0,5 \leq X \leq b - 0,5) \quad P(a \leq X \leq b) = P(a - 0,5 \leq X \leq b + 0,5)$$

En general, es aceptable la transformación cuando $\begin{cases} p \leq 0,5 & n \cdot p > 5 \\ q \leq 0,5 & n \cdot q > 5 \end{cases}$



TEÓRICO El contenido de una proteína se distribuye en individuos sanos según una $N(20, 0,5)$. En individuos con individuos con la enfermedad E se cree que la distribución normal sigue siendo razonable pero se sospecha que su valor medio podría alterarse. Supondremos que no cambia la variabilidad, por lo que el contenido de la proteína en presencia de E, variable aleatoria X, sigue una normal $N(\mu, 0,5)$. Por lo costoso de las determinaciones de dicha proteína se toma una muestra de 5 individuos con la enfermedad y se obtienen los siguientes valores: 22,2 , 21 , 18,8 , 21,5 , 20,5.

a) Con un nivel de confianza del 95%, ¿se puede admitir que la media de la proteína para los individuos con la enfermedad E se altera?.

- $H_0: \mu = 20$ $H_1: \mu \neq 20$ *contraste bilateral*

$\alpha = 0,05 \mapsto \alpha / 2 = 0,025 \mapsto z_{0,025} = 1,96$ *estadístico teórico*

Muestra: $n = 5$ $\bar{x} = 20,8$ $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \equiv N\left(\mu, \frac{0,5}{\sqrt{5}}\right)$

Intervalo de confianza: $\left(\mu - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) \mapsto$
 $\mapsto \left(20 - 1,96 \cdot \frac{0,5}{\sqrt{5}}, 20 + 1,96 \cdot \frac{0,5}{\sqrt{5}} \right) = (19,56, 20,44)$

$\bar{x} = 20,8 \notin (19,56, 20,44) \mapsto$ No se acepta la hipótesis nula H_0 , con un nivel de confianza del 95%

- También se podría haber realizado de la siguiente forma:

$$\alpha = 0,05 \mapsto z_{\alpha/2} = z_{0,025} = 1,96 \mapsto \begin{cases} \text{Región aceptación: } (-1,96, 1,96) \\ \text{Región crítica: } (-\infty, -1,96) \cup (1,96, \infty) \end{cases}$$

Bajo la hipótesis nula $H_0: \mu = 20$, $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \equiv N\left(20, \frac{0,5}{\sqrt{5}}\right) \equiv N(20, 0,22)$

En consecuencia, $P\left[-1,96 \leq z = \frac{\bar{x} - 20}{0,22} \leq 1,96\right] = 1 - \alpha = 0,95$

Hay una probabilidad de 0,95 de que el valor $\frac{\bar{x} - 20}{0,22}$ se encuentre entre $-1,96$ y $1,96$.

estadístico contraste

$$z = \frac{\bar{x} - 20}{0,22} = \frac{20,8 - 20}{0,22} = 3,64 \notin (-1,96, 1,96) \mapsto \text{Se rechaza } H_0$$

En definitiva, se toma la decisión de rechazar la hipótesis nula, rechazando que la media es 20 y aceptando la hipótesis alternativa que es distinta de 20.

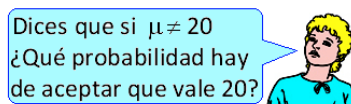
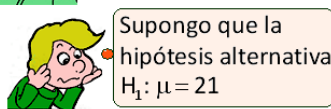
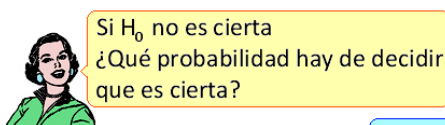
Las probabilidades conocidas, a priori, en esta forma de decidir son:

$$P[\text{Rechazar } H_0: \mu=20 \text{ siendo cierta } H_0] = 0,05 = \alpha$$

$$P[\text{Aceptar } H_0: \mu=20 \text{ siendo cierta } H_0] = 0,95 = 1 - \alpha$$

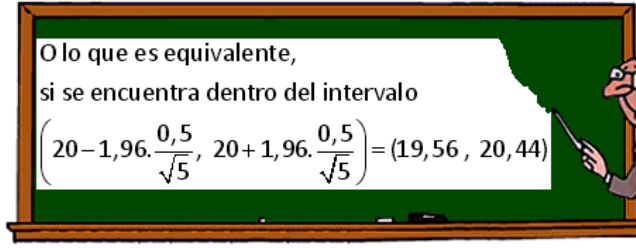
Pudiendo rellenar parte de la tabla siguiente:

		Decisión	
		H_0	H_1
Realidad	Cierta H_0	$1 - \alpha = 0,95$	$\alpha = 0,05$ Error tipo I
	Cierta H_1		





Hay que pensar si el estadístico del contraste $z = \frac{\bar{x} - \mu}{0,22}$ se encuentra en la región de aceptación $(-1,96, 1,96)$ bajo la hipótesis $H_1: \mu = 21$



Bajo la hipótesis $H_1: \mu = 21$, $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \equiv N\left(21, \frac{0,5}{\sqrt{5}}\right) \equiv N(21, 0,22)$

$$P[19,56 < \bar{x} < 20,44] = P\left[\frac{19,56 - 21}{0,22} < \frac{\bar{x} - 21}{0,22} < \frac{20,44 - 21}{0,22}\right] = P[-6,54 < z < -2,54] =$$

$$= P[2,54 < z < 6,54] = P[z > 2,54] - P[z > 6,54] = 0,00554$$

Adviértase que $P[\text{Aceptar } H_0 | H_0 \text{ es falsa}] \equiv P[\text{Aceptar } H_0 | H_1 \text{ es cierta}] = 0,00554 = \beta$

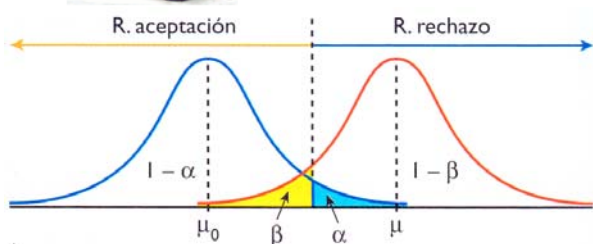
Se define β como la probabilidad de cometer un Error Tipo II:
 $\beta = P[\text{ET II}] = P[\text{Aceptar } H_0 | H_0 \text{ es falsa}] \equiv P[\text{Aceptar } H_0 | H_1 \text{ es cierta}]$

Se define la **Potencia del contraste** como $\text{Pot} = 1 - \beta$
 $\text{Pot} = 1 - \beta = P[\text{Rechazar } H_0 | H_0 \text{ es falsa}] \equiv P[\text{Aceptar } H_1 | H_1 \text{ es cierta}]$

$\beta = 0,00554 \rightarrow \text{Pot} = 1 - \beta = 1 - 0,00554 = 0,9945$

Pudiendo rellenar la tabla:

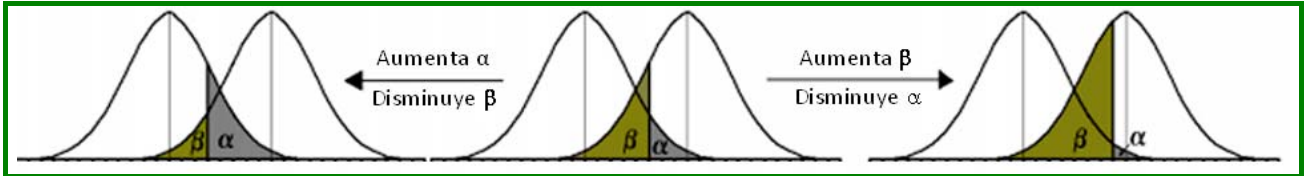
		Decisión	
		H_0	H_1
Realidad	Cierta H_0	$1 - \alpha = 0,95$	$\alpha = 0,05$ Error tipo I
	Cierta H_1	$\beta = 0,00554$ Error tipo II	$\text{Pot} = 1 - \beta = 0,9945$ Potencia contraste



Decisión	H_0 verdadera	H_0 falsa
Se acepta H_0	Decisión correcta Probabilidad $1 - \alpha$	Error tipo II Probabilidad β
Se rechaza H_0	Error tipo I Probabilidad α	Decisión correcta Probabilidad $1 - \beta$ Potencia contraste

Tanto α como β son probabilidades condicionadas. Las probabilidades de ambos errores no pueden calcularse en sentido absoluto. Para calcular α es necesario asumir que H_0 es cierta y para calcular β se asume que H_1 es cierta.

En cualquier prueba de hipótesis lo más conveniente será que ambos tipos de errores sean lo más pequeño posible, pero esto no es fácil de lograr, porque al intentar disminuir uno el otro aumenta proporcionalmente.



Al aumentar el tamaño n de la muestra es posible disminuir la probabilidad de cometer un Error Tipo II, manteniendo constante la probabilidad de cometer un Error Tipo I, en la figura adjunta se refleja como al aumentar el tamaño de la muestra se reduce la varianza de las distribuciones e igualmente el valor de β , mientras que el valor de α se mantiene en 0,05.



POTENCIA $Pot = 1 - P(ET II) = 1 - \beta$

- La potencia es mayor cuánto más alejada se encuentre la media verdadera μ_1 del valor supuesto μ_0 .
- Cuanto menor sea α menor es la potencia, es decir, si se reduce la probabilidad de cometer un error Tipo I, se incrementa la probabilidad de un error Tipo II.
- Cuanto mayor es la varianza de la población menor es la potencia, cuando se tiene más variabilidad resulta más difícil detectar pequeñas desviaciones del valor real respecto del valor supuesto μ_0 .
- Cuanto mayor sea el tamaño muestral mayor es la potencia del contraste, esto es, cuanto más información se tiene sobre la población más sencillo resulta detectar pequeñas desviaciones del valor real respecto de la hipótesis nula.

Se sabe que la desviación típica de las notas de cierto examen es 2,4. Para una muestra de 36 estudiantes se obtuvo una nota media de 5,6. ¿Sirven estos datos para confirmar la hipótesis de que la nota media del examen fue 6, con un nivel de confianza del 95%?

Solución:

- $H_0: \mu = 6$ $H_1: \mu \neq 6$ *contraste bilateral*

$$\alpha = 0,05 \quad \mapsto \quad z_{0,025} = 1,96 \quad \mapsto \quad \text{Estadístico teórico: } z_{0,025} = 1,96$$

$$\text{Intervalo de confianza: } \left(\mu - z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right) \quad \mapsto$$

$$\mapsto \left(6 - 1,96 \cdot \frac{2,4}{\sqrt{36}}, 6 + 1,96 \cdot \frac{2,4}{\sqrt{36}} \right) = (5,22, 6,78)$$

$\bar{x} = 5,6 \in (5,22, 6,78) \quad \mapsto \quad$ Se acepta la hipótesis nula H_0 , con un nivel de confianza del 95%

- Estadístico de contraste: $z = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}}$

$$\text{Se acepta } H_0 \text{ si: } z = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \leq z_{\alpha/2} \quad \mapsto \quad -z_{\alpha/2} \leq \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq z_{\alpha/2}$$

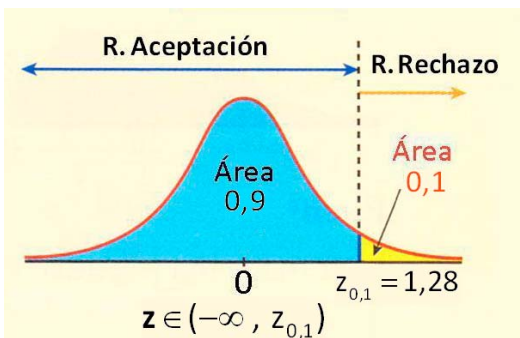
Es decir: $\alpha = 0,05 \quad \mapsto \quad \alpha/2 = 0,025 \quad \mapsto \quad z_{0,025} = 1,96 \quad \mapsto \quad$ Región aceptación: $(-1,96, 1,96)$

$$-1,96 \leq \frac{5,6 - 6}{2,4 / \sqrt{36}} \leq 1,96 \quad \mapsto \quad -1,96 \leq -1 \leq 1,96$$

Se acepta que la nota media del examen fue de 6, con un nivel de confianza del 95%

Un informe indica que el precio medio del billete de avión entre Canarias y Madrid es, como máximo, de 120 euros con una desviación típica de 40 euros. Se toma una muestra de 100 viajeros y se obtiene que la media de los precios de sus billetes es de 128 euros. ¿Se puede aceptar, con un nivel de significación de 0,1, la afirmación de partida?

Solución:



- $H_0: \mu \leq 120$ $H_1: \mu > 120$ *contraste unilateral*

$$\alpha = 0,1 \quad \mapsto \quad z_{0,1} = 1,28$$

$$\text{Intervalo de confianza: } \left(-\infty, \mu + z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right) \quad \mapsto$$

$$\mapsto \left(-\infty, 120 + 1,28 \cdot \frac{40}{\sqrt{100}} \right) = (-\infty, 125,12)$$

$128 \notin (-\infty, 125, 12) \mapsto$ No se acepta la hipótesis nula H_0 con un nivel de confianza del 90%

• Estadístico de contraste: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ Estadístico teórico: z_α

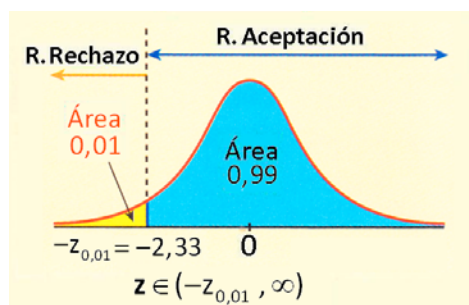
Se acepta H_0 si: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq z_\alpha \mapsto z = \frac{128 - 120}{40 / \sqrt{100}} = 2 \not\leq 1,28 = z_\alpha$

Es decir: $\alpha = 0,1 \mapsto z_{0,1} = 1,28 \mapsto$ Región aceptación: $(-\infty, 1,28)$

Como $2 \notin (-\infty, 1,28) \mapsto$ No se acepta la hipótesis nula H_0 con un nivel de confianza del 90%

La duración de las bombillas de 100 W que fabrica una empresa sigue una distribución normal con una desviación típica de 120 horas de duración. Su vida media está garantizada durante un mínimo de 800 horas. Se escoge al azar una muestra de 50 bombillas de un lote y, después de comprobarlas, se obtiene una vida media de 750 horas. Con un nivel de significación de 0,01, ¿habría que rechazar el lote por no cumplir la garantía?

Solución:



• $H_0: \mu \geq 800$ $H_1: \mu < 800$ *contraste unilateral*

$\alpha = 0,01 \mapsto z_{0,01} = 2,33$

Intervalo de confianza: $\left(\mu - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty \right) \mapsto$

$\mapsto \left(800 - 2,33 \cdot \frac{120}{\sqrt{50}}, \infty \right) = (760,46, \infty)$

$\bar{x} = 750 \notin (760,46, \infty) \mapsto$ No se acepta le informe con una fiabilidad del 99%

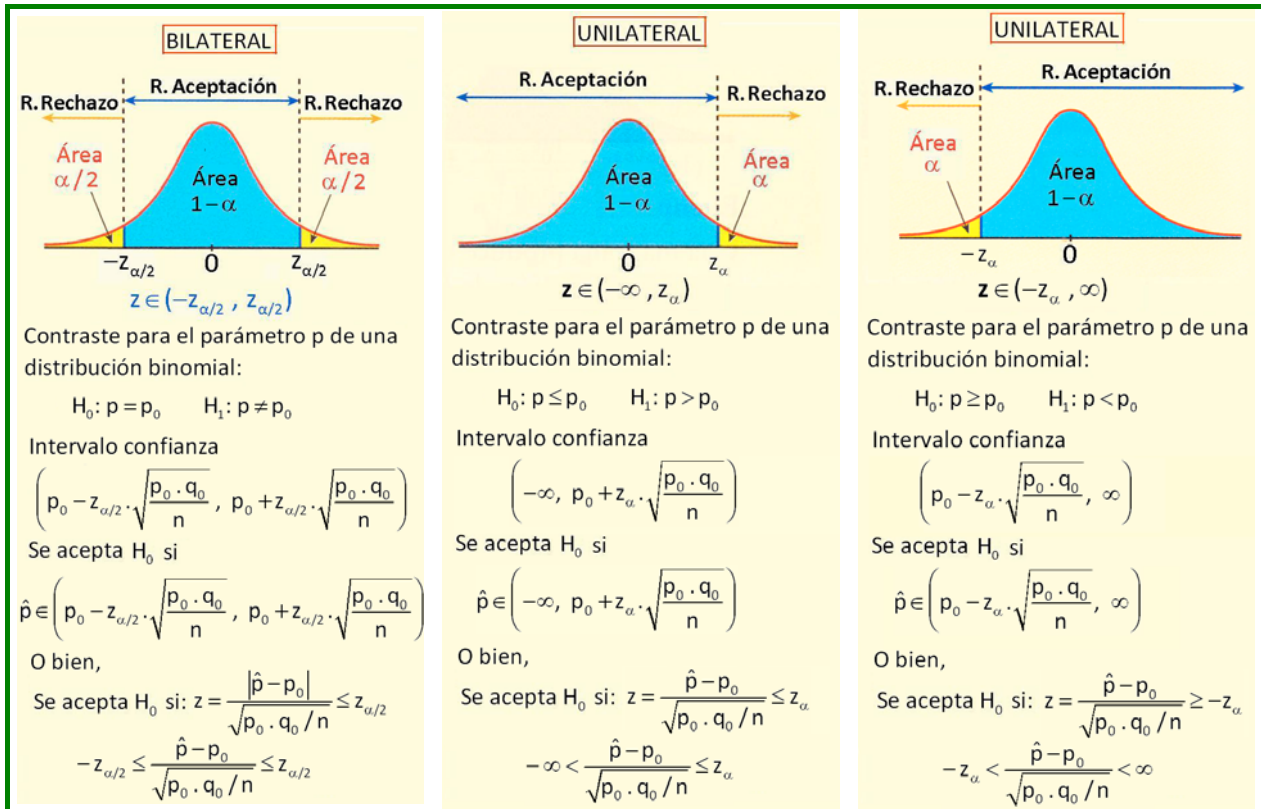
• Estadístico de contraste: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{750 - 800}{120 / \sqrt{50}} = -2,946$

Se acepta H_0 si: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq -z_\alpha \mapsto z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq z_\alpha \mapsto -2,946 \not\geq 2,33 = z_{0,01}$

Es decir,

$\alpha = 0,01 \mapsto z_{0,01} = 2,33 \mapsto -z_{0,01} = -2,33 \mapsto$ Región aceptación: $(-2,33, \infty)$

Como $-2,946 \notin (-2,33, \infty) \mapsto$ No se acepta la hipótesis nula H_0 con un nivel de confianza del 99%



Un laboratorio desea estudiar el porcentaje de personas que tienen somnolencia al tomar un medicamento. Se ha realizado un estudio con una muestra de 120 individuos, y se ha obtenido que el 15% ha tenido dichos efectos secundarios. El laboratorio desea afirmar que solo un 10% de pacientes tiene dichos efectos. ¿Puede realizar esta afirmación con un nivel de confianza del 99%?

Solución:

- $H_0: p = 0,1 \quad H_1: p \neq 0,1$ *contraste bilateral*

$$1 - \alpha = 0,99 \quad \mapsto \quad \alpha = 0,01 \quad \mapsto \quad z_{0,01/2} = z_{0,005} = 2,58 \quad \text{estadístico teórico}$$

$$\begin{aligned} \text{Intervalo de confianza: } & \left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \right) \mapsto \\ & \mapsto \left(0,1 - 2,58 \cdot \sqrt{\frac{0,1 \cdot 0,9}{120}}, 0,1 + 2,58 \cdot \sqrt{\frac{0,1 \cdot 0,9}{120}} \right) = (0,03, 2,65) \end{aligned}$$

$\hat{p} = 0,15 \in (0,03, 2,65) \mapsto$ Se acepta la hipótesis nula H_0 (10% de los pacientes tendrá efectos secundarios) con un nivel de confianza del 99%

- Estadístico de contraste: $z = \frac{|\hat{p} - p_0|}{\sqrt{p_0 \cdot q_0 / n}}$

$$\text{Se acepta } H_0 \text{ si: } z = \frac{|\hat{p} - p_0|}{\sqrt{p_0 \cdot q_0 / n}} \leq z_{\alpha/2} \quad \mapsto \quad -z_{\alpha/2} \leq \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot q_0 / n}} \leq z_{\alpha/2}$$

$$\alpha = 0,01 \quad \mapsto \quad z_{0,01/2} = z_{0,005} = 2,58 \quad \mapsto \quad \text{región aceptación: } (-2,58, 2,58)$$

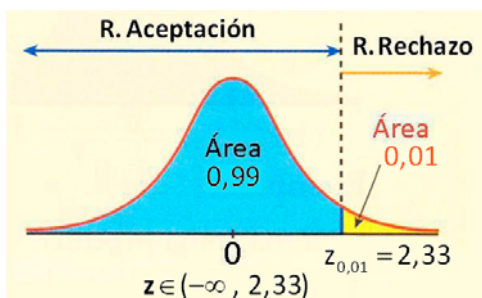
$$-2,58 \leq \frac{0,15 - 0,1}{\sqrt{\frac{0,1 \cdot 0,9}{120}}} \leq 2,58 \quad \mapsto \quad -2,58 \leq 1,83 \leq 2,58$$

Se acepta que la hipótesis nula H_0 con un nivel de confianza del 99%

Una marca de nueces afirma que, como máximo, el 6% de las nueces están vacías. Se eligieron 300 nueces al azar y se detectaron 21 vacías.

- Con un nivel de significación del 1%, ¿se puede aceptar la afirmación de la marca?
- Si se mantiene el porcentaje muestral de nueces que están vacías y el nivel de confianza es del 95%, ¿qué tamaño muestral se necesitaría para estimar la proporción de nueces con un error menor del 1%?

Solución:



- $H_0: p \leq 0,06 \quad H_1: p > 0,06 \quad \text{contraste unilateral}$

$$\alpha = 0,01 \quad \mapsto \quad z_{0,01} = 2,33 \quad \text{estadístico teórico}$$

$$\text{Intervalo de confianza: } \left(-\infty, p_0 + z_{\alpha} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \right) \quad \mapsto$$

$$\mapsto \left(-\infty, 0,06 + 2,33 \cdot \sqrt{\frac{0,06 \cdot 0,94}{300}} \right) = (-\infty, 0,092)$$

$$\hat{p} = \frac{21}{300} = 0,07 \in (-\infty, 0,092) \quad \mapsto \quad \text{Se acepta la hipótesis nula } H_0$$

con un nivel de confianza del 99%

- Estadístico de contraste: $z = \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot q_0 / n}}$

$$\text{Se acepta } H_0 \text{ si: } z = \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot q_0 / n}} \leq z_{\alpha} \quad \mapsto \quad -\infty < \frac{\hat{p} - p_0}{\sqrt{p_0 \cdot q_0 / n}} < z_{\alpha}$$

$$\alpha = 0,01 \mapsto z_{0,01} = 2,33 \mapsto \text{región aceptación: } (-\infty, 2,33)$$

$$\frac{0,07 - 0,06}{\sqrt{\frac{0,06 \cdot 0,94}{300}}} = 0,729 \leq 2,33 \mapsto \text{Se acepta que la hipótesis nula } H_0 \text{ con un nivel de confianza del 99\%}$$

$$\text{b) } I_{1-\alpha}(p) = \left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right] \quad 1 - \alpha = 0,95 \mapsto z_{\alpha/2} = z_{0,025} = 1,96$$

$$\xi = 0,01 = 1,96 \cdot \sqrt{\frac{0,07 \cdot 0,93}{n}} \mapsto \frac{0,01}{1,96} = \sqrt{\frac{0,07 \cdot 0,93}{n}} \mapsto 0,0051 = \sqrt{\frac{0,0651}{n}}$$

$$0,0051 = \sqrt{\frac{0,0651}{n}} \mapsto n = \frac{0,0651}{0,0051^2} \approx 2.500 \mapsto n \geq 2.501$$

DISTRIBUCIÓN t de Student

Sean $n+1$ variables aleatorias independientes X_1, X_2, \dots, X_n, X que siguen una distribución normal $N(0, \sigma)$, entonces la variable:

$$t_n = \frac{X}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (X_i)^2}}$$

se denomina *t de Student* con n grados de libertad.

Dividiendo, en el segundo término, numerador y denominador por la desviación típica σ , resulta:

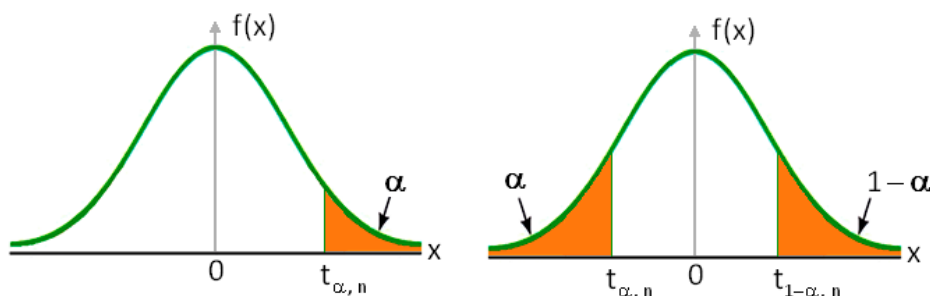
$$t_n = \frac{\frac{X}{\sigma}}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2}} = \frac{z}{\sqrt{\frac{1}{n} \cdot \chi_n^2}}$$

La función de densidad es:

$$f_{t_n}(x) = \frac{1}{\sqrt{n} \cdot \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}} \quad \begin{array}{l} n=1, 2, \dots \\ -\infty < x < \infty \end{array}$$

En análisis se define la función $\beta(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)}$

$$\Gamma(p) = \int_0^{\infty} x^{p-1} \cdot e^{-x} \cdot dx \quad \Gamma(p) \cdot \Gamma(p-1) = \frac{\pi}{\text{sen } p\pi} \quad \Gamma(1/2) = \sqrt{\pi}$$



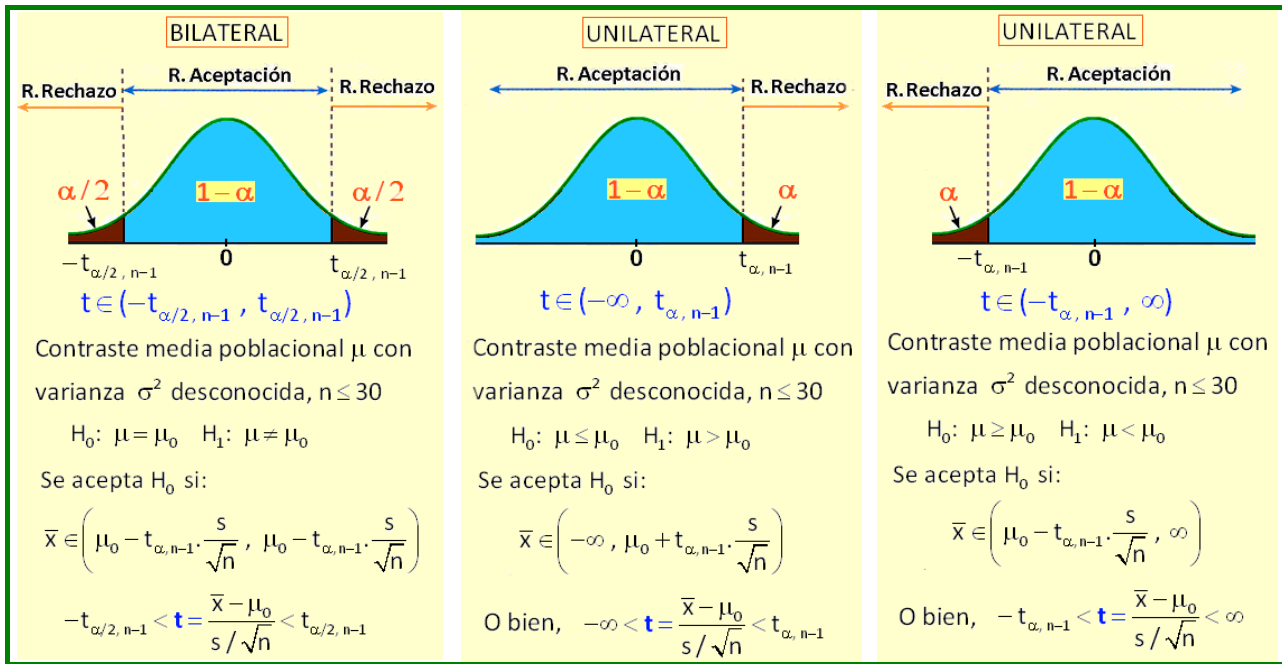
$$P(t_n \geq t_{\alpha, n}) = \alpha$$

$$P(t_n \leq t_{\alpha, n}) = 1 - P(t_n \geq t_{\alpha, n}) = 1 - \alpha$$

$$P(t_n \geq t_{\alpha, n}) = 1 - P(t_n \leq t_{\alpha, n}) = 1 - P(t_n \geq t_{1-\alpha, n}) = 1 - (1 - \alpha) = \alpha$$

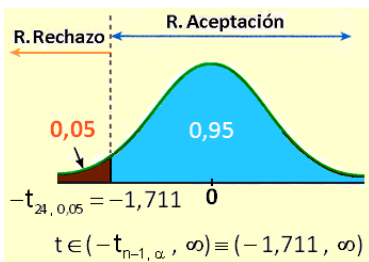
• En la estimación y en el contraste de hipótesis sobre la media poblacional μ se utiliza mucho la t de Student cuando las muestras son pequeñas $n \leq 30$ y la varianza poblacional σ^2 desconocida:

$$t_{n-1} = \frac{\bar{x} - \mu}{\sigma_x / \sqrt{n-1}} = \frac{\bar{x} - \mu}{s_x / \sqrt{n}} \quad \text{Muestra: } n \cdot \sigma_x^2 = (n-1) \cdot s_x^2$$



Se quiere estudiar el tiempo medio de reacción a un determinado estímulo visual. Sabiendo que el tiempo medio de reacción para una muestra de 25 observaciones resultó ser 2,9 con una cuasi-desviación típica de 0,8 sg, contrastar en hipótesis de normalidad, si el tiempo medio es menor de 3 sg, para un nivel de confianza del 95%

Solución:



• $H_0: \mu = 3$ $H_1: \mu < 3$ *contraste unilateral*

$\alpha = 0,05$
 $n = 25 \Rightarrow t_{\alpha, n-1} = t_{0,05, 24} = 1,711$ estadístico teórico

región aceptación: $(-1,711, \infty)$

• Intervalo de confianza: $\left(\mu_0 - t_{\alpha, n-1} \cdot \frac{s}{\sqrt{n}}, \infty \right) \mapsto$

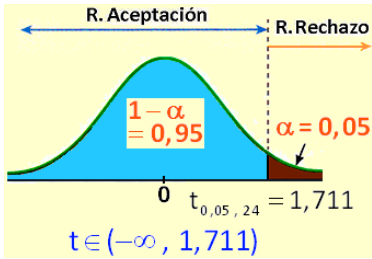
$$\mapsto \left(3 - 1,711 \cdot \frac{0,8}{\sqrt{25}} \right) = (2,73, \infty)$$

$\bar{x} = 2,9 \in (2,73, \infty) \mapsto$ Se acepta la hipótesis nula H_0 con una fiabilidad del 95%

• Estadístico de contraste: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2,9 - 3}{0,8/\sqrt{25}} = -0,625$

$t = -0,625 \in (-1,711, \infty) \mapsto$ Se acepta la hipótesis nula H_0 con un nivel de confianza del 95%. El tiempo medio de reacción es de 3 sg.

OTRO MÉTODO



• $H_0: \mu \leq 3$ $H_1: \mu > 3$ *contraste unilateral*

$\alpha = 0,05$
 $n = 25 \quad \mapsto \quad t_{\alpha, n-1} = t_{0,05, 24} = 1,711$ estadístico teórico

región aceptación: $(-\infty, 1,711)$

• Intervalo de confianza: $\left(-\infty, \mu_0 + t_{\alpha, n-1} \cdot \frac{s}{\sqrt{n}}\right) \mapsto$
 $\mapsto \left(-\infty, 3 + 1,711 \cdot \frac{0,8}{\sqrt{25}}\right) = (-\infty, 3,27)$

$\bar{x} = 2,9 \in (-\infty, 3,27) \mapsto$ Se acepta la hipótesis nula H_0 con una fiabilidad del 95%.

El tiempo medio de reacción es de 3 sg.

• Estadístico de contraste: $t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{2,9 - 3}{0,8 / \sqrt{25}} = -0,625$

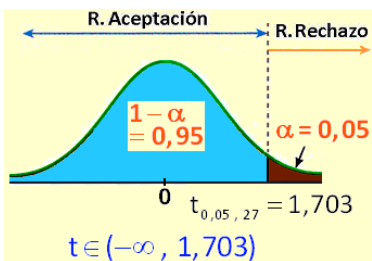
$t = -0,625 \in (-\infty, 1,711) \mapsto$ Se acepta la hipótesis nula H_0 con un nivel de confianza del 95%. El tiempo medio de reacción es de 3 sg.

Las puntuaciones en un test que mide la creatividad siguen, en la población general de adolescentes, una distribución normal de media 11,5. En centro escolar que ha implado un programa de estimulación de la creatividad de 28 alumnos ha proporcionado las siguientes puntuaciones:

9 11 17 8 11 24 17 17 5 6 14 23 15 7
 8 23 11 9 10 15 19 17 6 7 17 21 11 9

A un nivel de confianza del 95%, ¿puede afirmarse que el programa es efectivo?

Solución:



• $H_0: \mu = 11,5$ $H_1: \mu > 11,5$ *contraste unilateral*

$\alpha = 0,05$
 $n = 28 \quad \mapsto \quad t_{\alpha, n-1} = t_{0,05, 27} = 1,703$ estadístico teórico

región aceptación: $(-\infty, 1,703)$

Muestra: $\bar{x} = 13,10$ $\sigma_x^2 = 30,95$ $s_x^2 = 32,01$ $s_x = 5,66$

- Intervalo de confianza: $\left(-\infty, \mu_0 + t_{\alpha, n-1} \cdot \frac{s}{\sqrt{n}}\right) \mapsto$
 $\mapsto \left(-\infty, 11,5 + 1,703 \cdot \frac{5,67}{\sqrt{28}}\right) = (-\infty, 13,32)$

$13,10 \in (-\infty, 13,32) \mapsto$ Se acepta la hipótesis nula H_0 con una fiabilidad del 95%

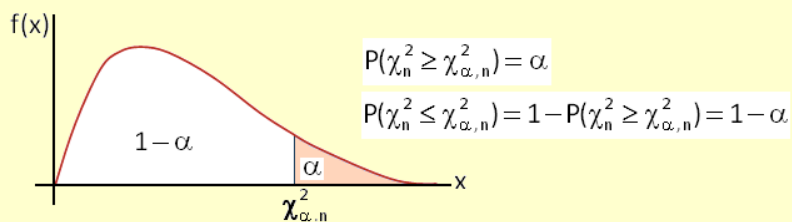
- Estadístico de contraste: $t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{13,1 - 11,5}{5,66 / \sqrt{28}} = 1,49$

$t = 1,49 \in (-\infty, 1,703) \mapsto$ Se acepta la hipótesis nula H_0 con un nivel de confianza del 95%, es decir, no hay evidencia de que el programa de estímulo de creatividad sea efectivo.

DISTRIBUCIÓN χ^2 de PEARSON

Sean n variables independientes entre sí (X_1, X_2, \dots, X_n) que siguen una distribución $N(0, 1)$

La variable $\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$ recibe el nombre de χ^2 (chi-cuadrado, ji-cuadrado) de Pearson con n grados de libertad.



■ Media, varianza y desviación típica: $\mu_{\chi^2} = n \quad \sigma_{\chi^2}^2 = 2n \quad \sigma_{\chi^2} = \sqrt{2n}$

■ Si $\chi_{n_1}^2$ y $\chi_{n_2}^2$ son dos χ^2 (chi-cuadrado, ji-cuadrado) de Pearson, respectivamente con n_1 y n_2 grados de libertad, independientes entre sí, la suma $\chi_{n_1+n_2}^2 = \chi_{n_1}^2 + \chi_{n_2}^2$ es también una χ^2 con $n_1 + n_2$ grados de libertad. Se puede generalizar para n variables aleatorias independientes.

■ Al aumentar el número de grados de libertad, la χ^2 se aproxima asintóticamente a la distribución normal:

$$\sqrt{2}\chi^2 \xrightarrow{n \rightarrow \infty} N(\sqrt{2n-1}, 1)$$

■ En una población $N(\mu, \sigma)$, el muestreo con media \bar{x} y

cuasivarianza s^2 , la variable $\chi_{n-1}^2 = \frac{(n-1) \cdot s^2}{\sigma^2}$

Esta propiedad es muy utilizada en la estimación y en el contraste de hipótesis sobre la varianza poblacional σ^2 .

Contraste para la varianza de una población normal

BILATERAL

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

Se acepta H_0 si $\chi_{n-1}^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2} \in (\chi_{1-\alpha/2, n-1}^2, \chi_{\alpha/2, n-1}^2)$

UNILATERAL

$$\bullet H_0: \sigma^2 \leq \sigma_0^2 \quad H_1: \sigma^2 > \sigma_0^2$$

Se acepta H_0 si $\chi_{n-1}^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2} < \chi_{\alpha, n-1}^2$

UNILATERAL

$$\bullet H_0: \sigma^2 \geq \sigma_0^2 \quad H_1: \sigma^2 < \sigma_0^2$$

Se acepta H_0 si $\chi_{n-1}^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2} > \chi_{\alpha, n-1}^2$

Una agencia de alquiler de automóviles, bajo la hipótesis de normalidad, necesita conocer con un 90 % de fiabilidad si la flota presenta una desviación típica de 7 km/día.

A tal fin, en varios días de la semana toma los recorridos de 25 vehículos de la flota y obtiene que la media muestral es de 165 km/día, y la cuasivarianza estándar muestral de 6 km/día.

Solución:

MÉTODO I

El intervalo de confianza para la varianza poblacional σ^2 viene dado por:

$$I(\sigma^2) = \left[\frac{(n-1) \cdot s_x^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1) \cdot s_x^2}{\chi_{1-\alpha/2, n-1}^2} \right] \quad s_x^2 \equiv \text{cuasivarianza muestral} \\ (n-1) \cdot s_x^2 = n \cdot \sigma_x^2$$

$$n = 25 \quad \alpha/2 = 0,05 \quad \chi_{\alpha/2, n-1}^2 = \chi_{0,05, 24}^2 = 36,415 \quad \chi_{1-\alpha/2, n-1}^2 = \chi_{0,95, 24}^2 = 13,848$$

$$s_x^2 = 36$$

$$I(\sigma^2) = \left[\frac{(n-1) \cdot s_x^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1) \cdot s_x^2}{\chi_{1-\alpha/2, n-1}^2} \right] = \left[\frac{24 \cdot 36}{36,415}, \frac{24 \cdot 36}{13,848} \right] = [23,73, 62,39]$$

El intervalo de confianza para la desviación típica poblacional σ :

$$I(\sigma) = \left[\sqrt{\frac{(n-1) \cdot s_x^2}{\chi_{\alpha/2, n-1}^2}}, \sqrt{\frac{(n-1) \cdot s_x^2}{\chi_{1-\alpha/2, n-1}^2}} \right] = \left[\sqrt{23,73}, \sqrt{62,39} \right] = [4,87, 7,9]$$

Como la desviación típica poblacional $\sigma = 7 \in [4,87, 7,9]$ se puede afirmar con un 90% de fiabilidad que la desviación típica poblacional es de 7 km/día.

MÉTODO II

El contraste de hipótesis bilateral para la varianza σ^2 de la población:

- $H_0: \sigma^2 = 49$ $H_1: \sigma^2 \neq 49$

Se acepta H_0 si $\chi_{n-1}^2 = \frac{(n-1) \cdot s_x^2}{\sigma_0^2} \in (\chi_{1-\alpha/2, n-1}^2, \chi_{\alpha/2, n-1}^2)$

$$\chi_{n-1}^2 = \frac{(n-1) \cdot s_x^2}{\sigma_0^2} = \frac{24 \cdot 36}{49} = 17,63 \in (13,848, 36,415)$$

Con un nivel de confianza del 90% se acepta que la varianza poblacional es 49, en consecuencia se acepta que la desviación típica poblacional es de 7 km/día.

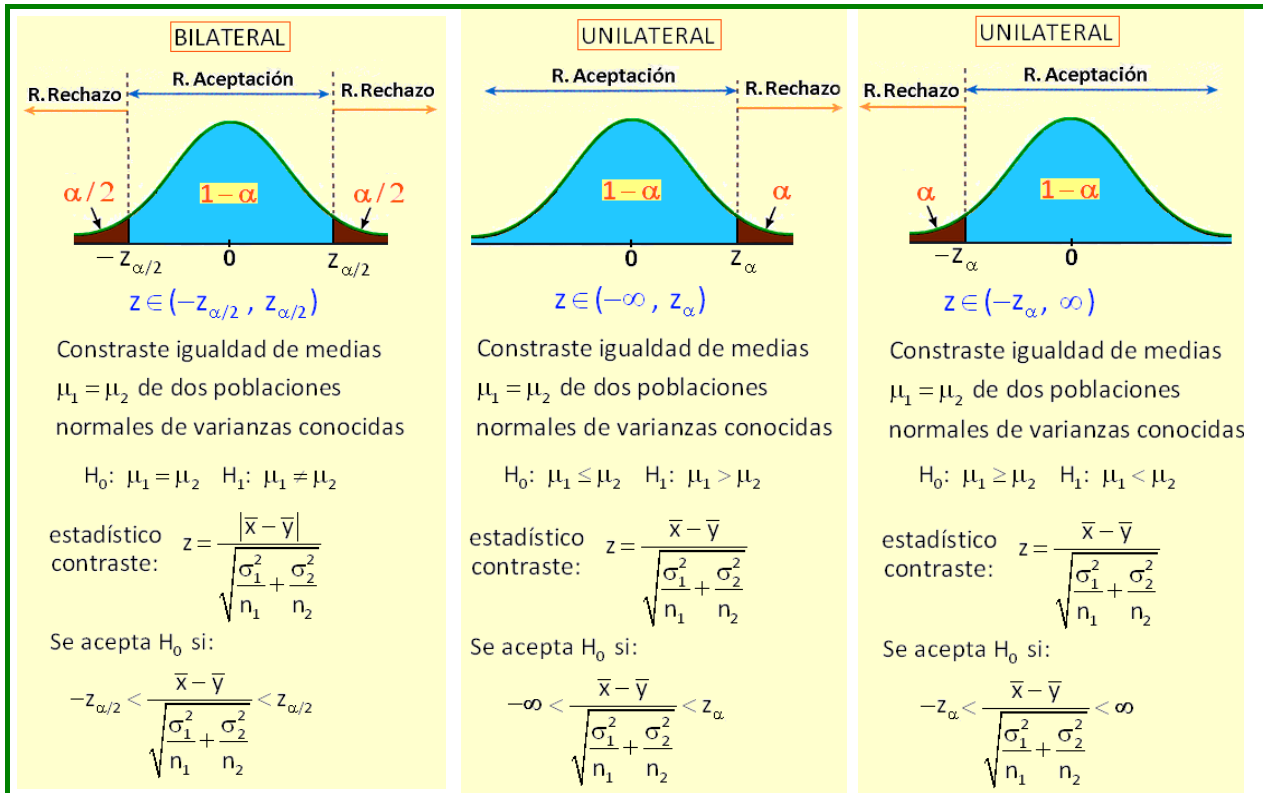


Si la muestra de 25 vehículos procede de una población $N(\mu, 7)$. ¿Cuál es la probabilidad de que la cuasivarianza muestral sea superior a 6 km/día?

La cuasivarianza muestral $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ es tal que el estadístico $\frac{(n-1) \cdot s_x^2}{\sigma^2} \sim \chi_{n-1}^2$

$$P[s_x^2 \geq 6] = P\left[\frac{(n-1) \cdot s_x^2}{\sigma^2} \geq \frac{(n-1) \cdot 36}{49}\right] = P\left[\chi_{24}^2 \geq \frac{24 \cdot 36}{49}\right] = P[\chi_{24}^2 \geq 17,63] = 0,81$$

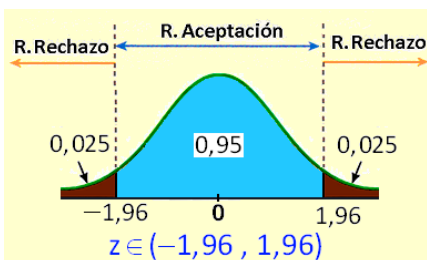
$$\begin{cases} 15,659 - 33,196 & 0,90 - 0,10 \\ 17,63 - 33,196 & x - 0,10 \end{cases} \begin{vmatrix} -17,537 & 0,8 \\ -15,566 & x - 0,10 \end{vmatrix} \mapsto x = 0,1 + \frac{0,8 \cdot 15,566}{17,537} = 0,81$$



A 100 alumnos de una clase se les separa en dos grupos, aquellos que practican habitualmente un deporte y los que no practican ninguno, formando cada grupo 60 y 40 alumnos, respectivamente. Se miden las alturas, y se obtiene para el primer grupo una media de 1,80 m y desviación típica 0,08 m, y para el segundo grupo, una media de 1,76 m y desviación típica 0,10 m. Suponiendo que la variable aleatoria altura sigue una distribución normal en los dos grupos, ¿es posible afirmar con un nivel de confianza del 95% que hay diferencia de altura entre los alumnos que practican algún deporte y los que no?

Solución:

Se trata de un contraste bilateral para la igualdad de medias poblacionales con muestras grandes y varianzas poblacionales desconocidas.



• $H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$

$1 - \alpha = 0,95 \Rightarrow \alpha / 2 = 0,025 \Rightarrow z_{0,025} = 1,96$

región aceptación: $(-1,96, 1,96)$

Muestra $\begin{cases} \bar{x}_1 = 1,80 \text{ m} & n_1 = 60 & \sigma_1 = 0,08 \text{ m} \\ \bar{x}_2 = 1,76 \text{ m} & n_2 = 40 & \sigma_2 = 0,10 \text{ m} \end{cases}$

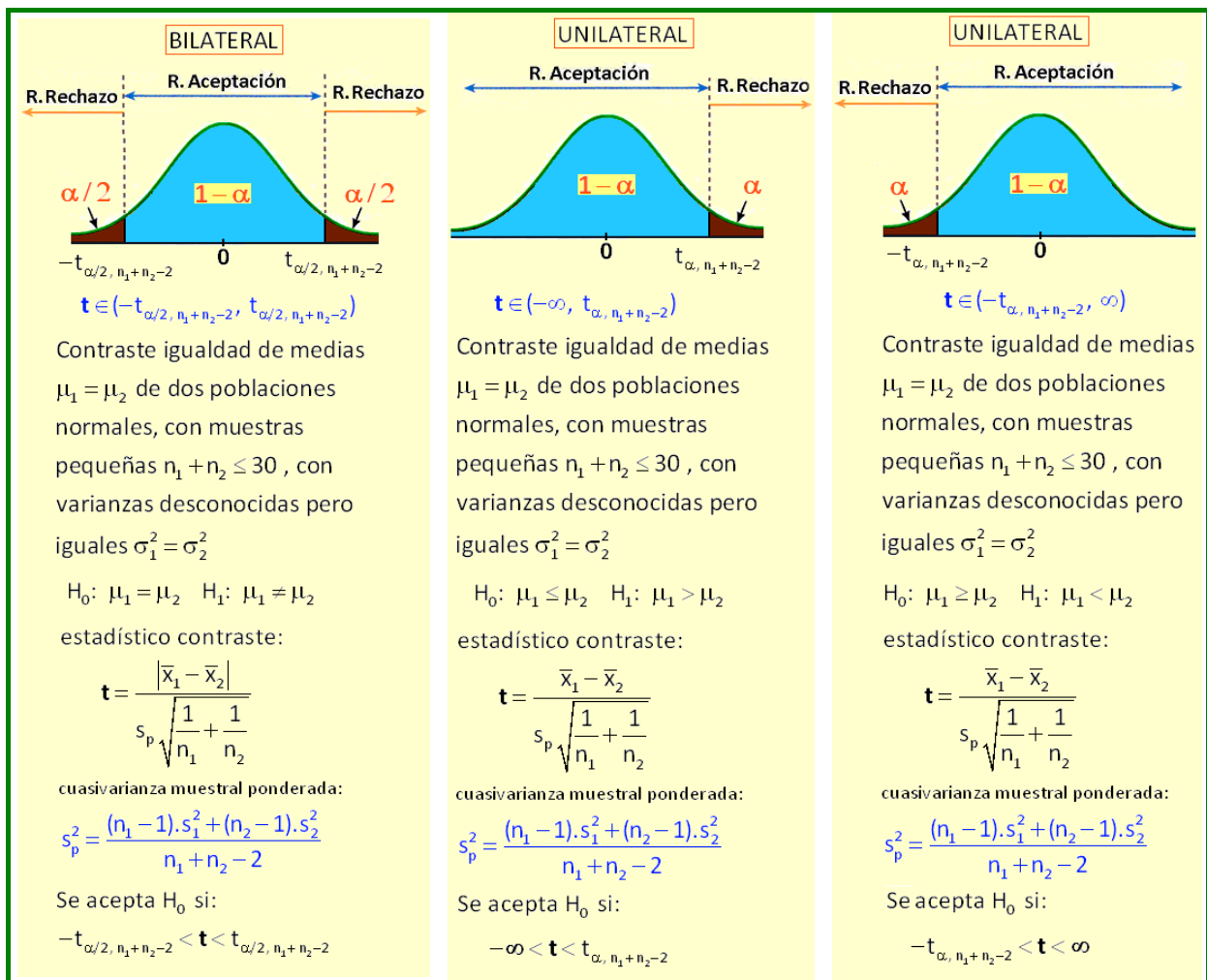
Cuasivarianzas y cuasidesviaciones típicas muestrales:

$$\begin{cases} n \cdot \sigma_1^2 = (n-1) \cdot s_1^2 \mapsto 60 \cdot 0,08^2 = 59 \cdot s_1^2 \mapsto s_1^2 = 0,0065 \mapsto s_1 = 0,081 \\ n \cdot \sigma_2^2 = (n-1) \cdot s_2^2 \mapsto 40 \cdot 0,10^2 = 39 \cdot s_2^2 \mapsto s_2^2 = 0,0103 \mapsto s_2 = 0,101 \end{cases}$$

Estadístico de contraste:
$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|1,80 - 1,76|}{\sqrt{\frac{0,0065}{60} + \frac{0,0103}{40}}} = 2,091$$

Se acepta la hipótesis nula H_0 si el estadístico de contraste se encuentra en la región de aceptación: $2,091 \notin (-1,96, 1,96)$

Concluyendo que, con un nivel de confianza del 95%, la media poblacional de los alumnos que practican algún deporte y de los que no lo practican es diferente.



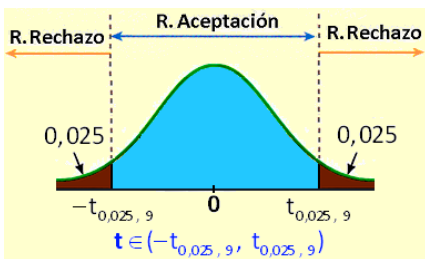
Las notas obtenidas en el Análisis de Datos de cinco individuos elegidos al azar del grupo G1 y de 6 individuos, elegidos también al azar, del grupo G2 son:

G1:	10	6	4	5	4	
G2:	4	8	6	6	2	3

Suponiendo que la variable aleatoria en estudio sigue una distribución normal en los dos grupos, ¿puede concluirse a un nivel de confianza del 95% que las puntuaciones medias de ambos grupos son iguales?

Solución:

Se trata de un contraste bilateral para la igualdad de medias poblacionales con muestras pequeñas $n_1 + n_2 = 11$ siendo $n_1 \approx n_2$, varianzas poblacionales desconocidas que se suponen iguales.



• $H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$

$1 - \alpha = 0,95 \quad \mapsto \quad t_{\alpha/2, n_1+n_2-2} = t_{0,025, 9} = 2,262$

región aceptación: $(-2,262, 2,262)$

Muestra $\begin{cases} \text{G1: } \bar{x}_1 = 5,8 & n_1 = 5 & s_1^2 = 6,2 \\ \text{G2: } \bar{x}_2 = 4,83 & n_2 = 6 & s_2^2 = 4,96 \end{cases}$

Media ponderada de las cuasivarianzas muestrales: $s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$

$s_p^2 = \frac{4 \cdot 6,2 + 5 \cdot 4,96}{9} = 5,51 \quad \mapsto \quad s_p = \sqrt{5,51} = 2,35$

Estadístico de contraste: $t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|5,8 - 4,83|}{2,35 \cdot \sqrt{\frac{1}{5} + \frac{1}{6}}} = 0,681$

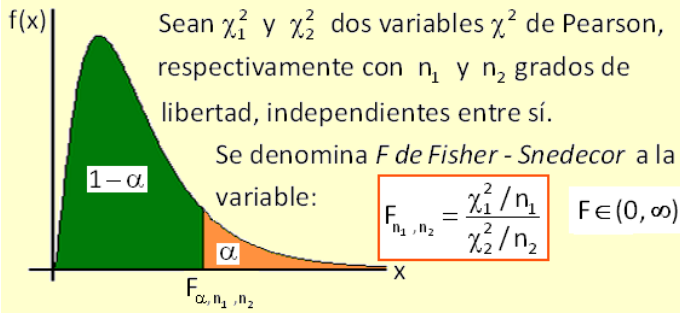
Siendo, $0,681 \in (-2,262, 2,262) \equiv -2,262 < 0,681 < 2,262$

Se acepta la hipótesis nula H_0 , por consiguiente se puede afirmar con una fiabilidad del 95% que no existe diferencia entre la media poblacional de ambos grupos.

▪ F de Fisher - Snedecor

En ocasiones es necesario contrastar la igualdad de varianzas de dos poblaciones normales.

DISTRIBUCIÓN F de Fisher-Snedecor



Es muy utilizada en el análisis de la varianza

$$P(F_{n_1, n_2} \geq F_{\alpha, n_1, n_2}) = \alpha$$

$$F_{\alpha, n_1, n_2} = \frac{1}{F_{1-\alpha, n_2, n_1}}$$

$$P(F_{n_1, n_2} < F_{\alpha, n_1, n_2}) = 1 - P(F_{n_1, n_2} \geq F_{\alpha, n_1, n_2}) = 1 - \alpha$$

Contraste de igualdad de varianzas de dos poblaciones normales.

BILATERAL

- $H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 \neq \sigma_2^2$

Se acepta H_0 si $\frac{S_1^2}{S_2^2} \in [F_{1-\alpha/2, n_1-1, n_2-1}; F_{\alpha/2, n_1-1, n_2-1}]$

UNILATERAL

- $H_0: \sigma_1^2 \leq \sigma_2^2 \quad H_1: \sigma_1^2 > \sigma_2^2$

Se acepta H_0 si $\frac{S_1^2}{S_2^2} \leq F_{\alpha, n_1-1, n_2-1}$

UNILATERAL

- $H_0: \sigma_1^2 \geq \sigma_2^2 \quad H_1: \sigma_1^2 < \sigma_2^2$

Se acepta H_0 si $\frac{S_1^2}{S_2^2} \geq F_{\alpha, n_1-1, n_2-1}$

Una central lechera recibe diariamente leche de dos granjas, A y B.

Para analizar la calidad de la leche se eligen dos muestras al azar de la cantidad suministrada por cada una de las granjas, se analiza el contenido de la grasa y se obtienen los datos:

	Número muestras	Media muestral	Cuasivarianza
Granja A	$n_1 = 12$	$\bar{x}_1 = 0,305$	$s_1^2 = 0,034$
Granja B	$n_2 = 16$	$\bar{x}_2 = 0,318$	$s_2^2 = 0,027$

Bajo la hipótesis de normalidad de los datos, con una fiabilidad del 90%, ¿Son las dos granjas semejantes desde el punto de vista del contenido en grasa de la leche que emiten a la central lechera?

Solución:

Se trata de un contraste de igualdad de medias, con muestras pequeñas y se necesita comprobar si las varianzas poblacionales σ_1^2 y σ_2^2 son o no iguales.

- Contraste de igualdad de varianzas de dos poblaciones normales:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$\text{Estadístico de contraste: } F = \frac{s_1^2}{s_2^2} = \frac{0,034}{0,027} = 1,26$$

$$\text{Se acepta } H_0 \text{ si } F = \frac{s_1^2}{s_2^2} \in [F_{1-\alpha/2, n_1-1, n_2-1}; F_{\alpha/2, n_1-1, n_2-1}]$$

$$\alpha = 0,1 \quad \alpha/2 = 0,05 \quad \begin{cases} F_{1-\alpha/2, n_1-1, n_2-1} = F_{0,95, 11, 15} = \frac{1}{F_{0,05, 15, 11}} = \frac{1}{2,72} = 0,37 \\ F_{\alpha/2, n_1-1, n_2-1} = F_{0,05, 11, 15} = 2,51 \end{cases}$$

Siendo $F = \frac{s_1^2}{s_2^2} = 1,26 \in [0,37; 2,51]$ se acepta la hipótesis de que $\sigma_1^2 = \sigma_2^2$,

es decir, las varianzas son iguales.

- Es un contraste de igualdad de medias poblacionales con varianzas iguales y muestras pequeñas $n_1 + n_2 = 12 + 16 = 28 < 30$, se establecen las hipótesis:

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

$$\text{Media ponderada de las cuasivarianzas muestrales: } s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$$

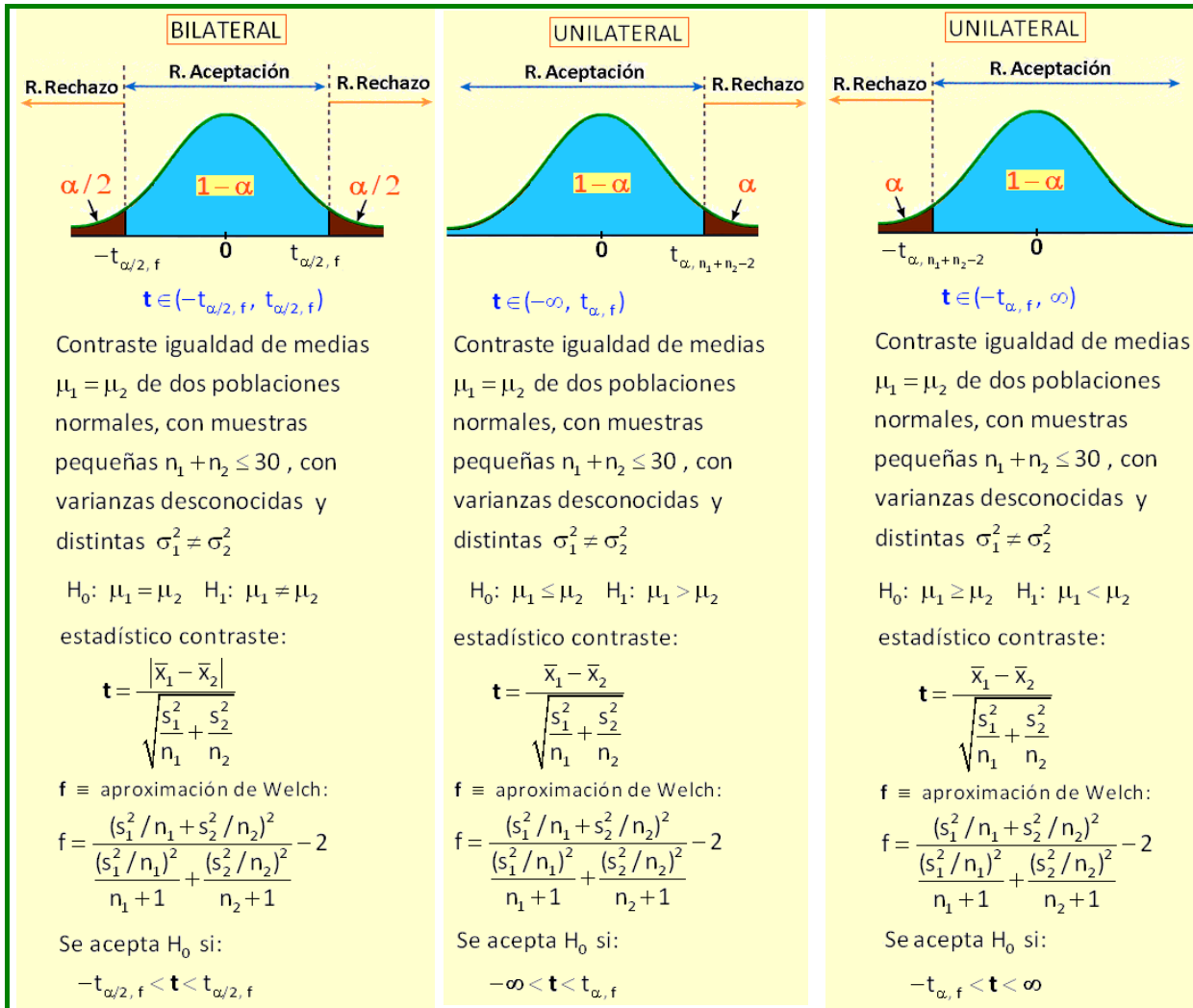
$$s_p^2 = \frac{11 \cdot 0,034 + 15 \cdot 0,027}{26} = 0,03 \quad \mapsto \quad s_p = \sqrt{0,03} = 0,1732$$

$$\text{Estadístico de contraste: } t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|0,305 - 0,318|}{0,1732 \cdot \sqrt{\frac{1}{12} + \frac{1}{16}}} = \frac{|-0,013|}{0,0661} = 0,1967$$

$$\text{Estadístico teórico: } t_{\alpha/2, n_1 + n_2 - 2} = t_{0,025, 26} = 2,056$$

$$\text{Como } 0,1967 \in (-2,056, 2,056) \equiv -2,056 < 0,1967 < 2,056$$

Se acepta la hipótesis nula H_0 y se concluye que las dos granajas son semejantes desde el punto de vista del contenido en grasa de la leche, con un nivel de confianza del 90%.



Un Instituto de Alimentación animal quiere comparar estadísticamente dos tipos de dietas. Selecciona al azar una muestra de veinte animales de una población de animales comparables. A doce de ellos se les suministra la dieta primera y a los ocho restantes la dieta segunda. Los resultados del aumento de peso en una semana son:

	Número muestras	Media muestral	Cuasivarianza
Dieta primera	$n_1 = 12$	$\bar{x}_1 = 4,3$ kg	$s_1^2 = 0,81$
Dieta segunda	$n_2 = 8$	$\bar{x}_2 = 3,6$ kg	$s_2^2 = 3,61$

Con un nivel de confianza del 90%, ¿se puede afirmar que la primera dieta es mejor que la segunda?

Solución:

Se trata de un contraste unilateral de igualdad de medias, con muestras pequeñas y se necesita comprobar si las varianzas poblacionales σ_1^2 y σ_2^2 son o no iguales.

- Contraste de igualdad de varianzas de dos poblaciones normales:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\text{Estadístico de contraste: } F = \frac{s_1^2}{s_2^2} = \frac{0,81}{3,61} = 0,22$$

$$\text{Se acepta } H_0 \text{ si } F = \frac{s_1^2}{s_2^2} \in [F_{1-\alpha/2, n_1-1, n_2-1}; F_{\alpha/2, n_1-1, n_2-1}]$$

$$\alpha = 0,1 \quad \alpha/2 = 0,05 \quad \begin{cases} F_{1-\alpha/2, n_1-1, n_2-1} = F_{0,95, 11, 7} = \frac{1}{F_{0,05, 7, 11}} = \frac{1}{3,01} = 0,33 \\ F_{\alpha/2, n_1-1, n_2-1} = F_{0,05, 11, 7} = 3,60 \end{cases}$$

Siendo $F = \frac{s_1^2}{s_2^2} = 0,22 \notin [0,33; 3,60]$ no se acepta la hipótesis de que $\sigma_1^2 = \sigma_2^2$,

es decir, las varianzas poblacionales son distintas.

- Es un contraste de unilateral de medias poblacionales con varianzas poblacionales distintas $\sigma_1^2 \neq \sigma_2^2$, se establecen las hipótesis:

$$H_0: \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2$$

$$\text{Estadístico de contraste: } t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{4,3 - 3,6}{\sqrt{\frac{0,81}{12} + \frac{3,61}{8}}} = 0,97$$

Estadístico teórico: $t_{\alpha, f}$ donde f es la aproximación de Welch

$$f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2+1}} - 2 = \frac{\left(\frac{0,81}{12} + \frac{3,61}{8}\right)^2}{\frac{\left(\frac{0,81}{12}\right)^2}{13} + \frac{\left(\frac{3,61}{8}\right)^2}{9}} - 2 = 9,71 \approx 10$$

$$t_{\alpha, f} = t_{0,1, 10} = 1,372$$

Se acepta H_0 si $t = 0,97 < t_{0,1, 10} = 1,372$

Se acepta la hipótesis nula H_0 y se concluye que no existe evidencia estadística de que la primera dieta sea mejor que la segunda.

Contraste igualdad de medias para datos apareados

BILATERAL

- $H_0: d=0 \Leftrightarrow \mu_1 = \mu_2$ $H_1: d \neq 0 \Leftrightarrow \mu_1 \neq \mu_2$

Se acepta H_0 si $t = \frac{|\bar{d}|}{s_d / \sqrt{n}} \leq t_{\alpha/2, n-1}$

UNILATERAL

- $H_0: d \leq 0 \Leftrightarrow \mu_1 \leq \mu_2$ $H_1: d > 0 \Leftrightarrow \mu_1 > \mu_2$

Se acepta H_0 si $t = \frac{\bar{d}}{s_d / \sqrt{n}} \leq t_{\alpha, n-1}$

UNILATERAL

- $H_0: d \geq 0 \Leftrightarrow \mu_1 \geq \mu_2$ $H_1: d < 0 \Leftrightarrow \mu_1 < \mu_2$

Se acepta H_0 si $t = \frac{\bar{d}}{s_d / \sqrt{n}} \geq t_{\alpha, n-1}$

siendo

$$d_i = x_i - y_i \quad \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$$

Una empresa farmacéutica está interesada en la investigación preliminar de un nuevo medicamento con posibles propiedades reductoras del colesterol. Para ello, toma una muestra al azar de seis personas comparables, determinando el contenido en colesterol antes y después del tratamiento. Los resultados obtenidos han sido:

Antes: 217 252 229 200 209 213

Después: 209 241 230 208 206 211

Bajo el supuesto de normalidad de la variable aleatoria en estudio, ¿se puede afirmar la bondad del medicamento, con un nivel de confianza del 95%?

Solución:

Se trata de un experimento de datos apareados, siendo los valores de las diferencias:

$x_i \equiv$ Antes 217 252 229 200 209 213

$y_i \equiv$ Después 209 241 230 208 206 211

$d_i = x_i - y_i$ 8 11 -1 -8 3 2

$$\bar{d} = \frac{1}{6} \sum_{i=1}^6 d_i = 2,5 \quad s_d^2 = \frac{1}{5} \sum_{i=1}^6 (d_i - \bar{d})^2 = 45,1 \quad \mapsto \quad s_d = \sqrt{45,1} = 6,71$$

- Es un contraste bilateral de igualdad de medias en el caso de datos apareados. Se establecen las hipótesis:

$$H_0 : d=0 \Leftrightarrow \mu_1 = \mu_2 \quad H_1 : d \neq 0 \Leftrightarrow \mu_1 \neq \mu_2$$

$$\text{Estadístico de contraste: } t = \frac{|\bar{d}|}{s_d / \sqrt{n}} = \frac{|2,5|}{6,71 / \sqrt{6}} = 0,91$$

$$\text{Estadístico teórico, } n=6, \alpha/2=0,025 \quad \mapsto \quad t_{\alpha/2, n-1} = t_{0,025, 5} = 2,015$$

Siendo $t = 0,91 < 2,015 = t_{0,025, 5}$ se acepta la hipótesis nula y se concluye que no se puede afirmar estadísticamente la bondad del medicamento en la reducción del colesterol.



Test de hipótesis más frecuentes

α = nivel de significación del contraste

H_0 = hipótesis nula

R = región crítica o

n = tamaño de la muestra

de rechazo de H_0

$X \sim N(\mu, \sigma)$

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

$$H_0 : \mu = \mu_0 \ (\sigma \text{ desconocida}) \quad R = \left\{ |\bar{x} - \mu_0| > t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \mu \leq \mu_0 \ (\sigma \text{ conocida}) \quad R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

$$H_0 : \mu \leq \mu_0 \ (\sigma \text{ desconocida}) \quad R = \left\{ \bar{x} - \mu_0 > t_{\alpha, n-1} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \mu \geq \mu_0 \ (\sigma \text{ conocida}) \quad R = \left\{ \bar{x} - \mu_0 < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

$$H_0 : \mu \geq \mu_0 \ (\sigma \text{ desconocida}) \quad R = \left\{ \bar{x} - \mu_0 < t_{1-\alpha, n-1} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \sigma = \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 \notin \left[\chi_{1-\alpha/2, n-1}^2, \chi_{\alpha/2, n-1}^2 \right] \right\}$$

$$H_0 : \sigma \leq \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi_{\alpha, n-1}^2 \right\}$$

$$H_0 : \sigma \geq \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 < \chi_{1-\alpha, n-1}^2 \right\}$$

$X \sim B(1, p)$ muestras grandes

$$H_0 : p = p_0 \quad R = \left\{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

$$H_0 : p \leq p_0 \quad R = \left\{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

$$H_0 : p \geq p_0 \quad R = \left\{ \bar{x} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

$X \sim P(\lambda)$ muestras grandes

$$H_0 : \lambda = \lambda_0 \quad R = \left\{ |\bar{x} - \lambda_0| > z_{\alpha/2} \sqrt{\lambda_0/n} \right\}$$

$$H_0 : \lambda \leq \lambda_0 \quad R = \left\{ \bar{x} - \lambda_0 > z_{\alpha} \sqrt{\lambda_0/n} \right\}$$

$$H_0 : \lambda \geq \lambda_0 \quad R = \left\{ \bar{x} - \lambda_0 < z_{1-\alpha} \sqrt{\lambda_0/n} \right\}$$



Test de hipótesis más frecuentes

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Dos poblaciones normales independientes

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

$$H_0 : \mu_1 = \mu_2 \ (\sigma_1 = \sigma_2) \quad R = \left\{ |\bar{x} - \bar{y}| > t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

$$H_0 : \mu_1 = \mu_2 \ (\sigma_1 \neq \sigma_2) \quad R = \left\{ |\bar{x} - \bar{y}| > t_{\alpha/2, f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$$

$$H_0 : \mu_1 \leq \mu_2 \ (\sigma_1, \sigma_2 \text{ conocidas}) \quad R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$$

$$H_0 : \mu_1 \leq \mu_2 \ (\sigma_1 = \sigma_2) \quad R = \left\{ \bar{x} - \bar{y} > t_{\alpha, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

$$H_0 : \mu_1 \leq \mu_2 \ (\sigma_1 \neq \sigma_2) \quad R = \left\{ \bar{x} - \bar{y} > t_{\alpha, f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$$

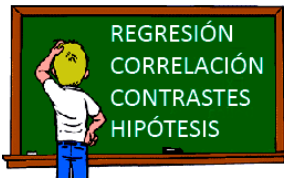
$$H_0 : \mu_1 \geq \mu_2 \ (\sigma_1, \sigma_2 \text{ conocidas}) \quad R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$$

$$H_0 : \mu_1 \geq \mu_2 \ (\sigma_1 = \sigma_2) \quad R = \left\{ \bar{x} - \bar{y} < t_{1-\alpha, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

$$H_0 : \sigma_1 = \sigma_2 \quad R = \left\{ s_1^2/s_2^2 \notin [F_{1-\alpha/2, n_1-1, n_2-1}, F_{\alpha/2, n_1-1, n_2-1}] \right\}$$

$$H_0 : \sigma_1 \leq \sigma_2 \quad R = \left\{ s_1^2/s_2^2 > F_{\alpha, n_1-1, n_2-1} \right\}$$

$$H_0 : \sigma_1 \geq \sigma_2 \quad R = \left\{ s_1^2/s_2^2 < F_{1-\alpha/2, n_1-1, n_2-1} \right\}$$



En un test de 27 niños se ha calculado el coeficiente de correlación de Pearson entre sus años y las respuestas inadecuadas que tiene, obteniéndose una correlación inversa de 0,74. Con un nivel de confianza del 95%, ¿están las dos variables incorreladas?

Solución:

Bajo los supuestos:

- Las variables se distribuyen normalmente en la población
- Se trata de una muestra aleatoria simple

$\rho \equiv$ correlación poblacional
 $r \equiv$ correlación muestral

$H_0: \rho = \rho_0 \quad H_1: \rho \neq \rho_0$
 Se acepta H_0 si: $t = \frac{\bar{x} - \rho_0}{\sqrt{\frac{1-r^2}{n-2}}} \leq t_{\alpha/2, n-2}$
 $\rho = 0$ variables incorreladas
 $\rho \neq 0$ variables correladas

- $H_0: \rho = 0 \quad H_1: \rho \neq 0$ variables incorreladas

Estadístico de contraste: $t = \frac{r - \rho_0}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0,74 - 0}{\sqrt{\frac{1-0,74^2}{27-2}}} = 5,5$

Estadístico teórico: $\alpha = 0,05$, $n = 27 \mapsto t_{0,025, 25} = 2,06$

$t = 5,5 > 2,06 = t_{0,025, 25} \mapsto$ Se rechaza la hipótesis nula con una fiabilidad del 95%
 Los años y las respuestas inadecuadas están correlados

En la tabla adjunta se muestran las puntuaciones que obtuvieron en un examen de aptitud previo a su contratación (X) y la venta mensual de los agentes comerciales (Y) en de diez mil euros.

X: 82 51 70 59 57 52 93 99 90 43 68 71 82 98 95

Y: 10 8 13 10 12 6 12 12 10 8 10 13 14 14 13

a) ¿Se deduce de estos datos que la puntuación en el examen previo influye en la venta mensual?, ¿con que grado de predicción?

b) Con una fiabilidad del 95%, ¿es ocasional la correlación de las variables X e Y?

Solución:

a) La recta de regresión de Y sobre X: $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x})$

$$\bullet \quad a_{10} = \bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = \frac{1110}{15} = 74 \quad a_{20} = \frac{\sum_{i=1}^{15} x_i^2}{15} = \frac{87076}{15} = 5805,07$$

$$\sigma_x^2 = a_{20} - a_{10}^2 = 5805,07 - 74^2 = 329,07 \quad \mapsto \quad \sigma_x = \sqrt{329,07} = 18,14$$

$$\bullet \quad a_{01} = \bar{y} = \frac{\sum_{i=1}^{15} y_i}{15} = \frac{165}{15} = 11 \quad a_{02} = \frac{\sum_{i=1}^{15} y_i^2}{15} = \frac{1895}{15} = 126,33$$

$$\sigma_y^2 = a_{02} - a_{01}^2 = 126,33 - 11^2 = 5,33 \quad \mapsto \quad \sigma_y = \sqrt{5,33} = 2,3$$

$$\bullet \quad a_{11} = \bar{xy} = \frac{\sum_{i=1}^{15} x_i \cdot y_i}{15} = \frac{12630}{15} = 842 \quad \sigma_{xy} = a_{11} - a_{10} \cdot a_{01} = 842 - 74 \cdot 11 = 28$$

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) \quad \mapsto \quad y - 11 = \frac{28}{329,07} (x - 74) \quad \mapsto \quad \boxed{y = 4,71 + 0,085 x}$$

La pendiente $b_{yx} = 0,085$ indica un aumento de 850 euros mensuales de venta.

El grado de predicción, indicando un grado de asociación directa entre las puntuaciones del examen de aptitud y las ventas mensuales,

viene determinado por el coeficiente de correlación lineal $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad \mapsto \quad r = \frac{28}{18,14 \cdot 2,3} = 0,67$$

b) Para aceptar o rechazar el hecho de que el grado de predicción, o asociación lineal de las variables, no es meramente una coincidencia, es necesario realizar un contraste de hipótesis.


Bajo el supuesto de que las variables en estudio son una muestra aleatoria simple y que se distribuyen normalmente en la población, se establece el contraste:

- $H_0 : \rho = 0$ $H_1 : \rho \neq 0$ *variables incorreladas*

$$\text{Estadístico de contraste: } t = \frac{r - \rho_0}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0,67 - 0}{\sqrt{\frac{1-0,67^2}{15-2}}} = 3,25$$

$$\text{Estadístico teórico: } \alpha = 0,05 \quad , \quad n = 15 \quad \mapsto \quad t_{0,025, 13} = 2,16$$

Se rechaza la hipótesis nula con una fiabilidad del 95%
 $t = 3,25 > 2,16 = t_{0,025, 13} \mapsto$ indicando que la correlación $r = 0,67$ es significativa y las variables sí tienen un estrecho grado de predicción



El error estándar de la regresión mide la dispersión de la ecuación de regresión $y = a + b x$

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{15} y_i^2 - a \sum_{i=1}^{15} y_i - b \sum_{i=1}^{15} x_i \cdot y_i}{n-2}}$$

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{15} y_i^2 - a \sum_{i=1}^{15} y_i - b \sum_{i=1}^{15} x_i \cdot y_i}{n-2}} = \sqrt{\frac{1895 - 4,71 \cdot 165 - 0,085 \cdot 12630}{15-2}} = 1,84$$

El error estándar de la regresión es menor de 18400 euros mensuales.

Para medir la dispersión de la ecuación de regresión también se utiliza la desviación estándar de la variable dependiente s_y , aunque es más preciso el error estándar de la regresión s_{yx}

$$\text{desviación estándar de Y: } s_y = \sqrt{\frac{\sum_{i=1}^{15} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{15} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{15} y_i \right)^2}{n-1}}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{15} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{15} y_i \right)^2}{n-1}} = \sqrt{\frac{1895 - \frac{1}{15} \cdot 165^2}{15-1}} = 2,39$$

La desviación estándar de las ventas es aproximadamente de 23900 euros.

Es menor el error estándar de la regresión $s_{yx} = 18400$ euros ($s_{yx} < s_y$)



Recta de regresión de Y sobre X: $y = 4,71 + 0,085 x$

Entonces el agente que obtuvo 100 puntos en el examen vendería: $y = 4,71 + 0,085 \cdot 100 = 13,21$
es decir, cerca de 132.100 euros mensuales

Es una muestra y las variables en la población siguen una normal

Se puede ampliar el concepto para generalizar o hacer inferencias

Fijando un porcentaje de confianza...

¿Cuánto vendería al mes un vendedor promedio?

¿Cuánto vendería al mes un vendedor cualquiera?



Se construye un intervalo de confianza para el *vendedor promedio* μ_{YX} :

$$\mu_{YX} \in (\hat{y} \pm t_{\alpha/2, n-2} \cdot s_{\hat{y}})$$

Se construye un intervalo de confianza para *cualquier vendedor* Y:

$$Y \in (\hat{y} \pm t_{\alpha/2, n-2} \cdot s_{y \cdot})$$



Para un valor dado x^* se obtiene el valor $\hat{y} = a + b \cdot x^*$

donde,

$$s_{\hat{y}} = s_{YX} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}}$$

$$s_{y \cdot} = s_{YX} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}}$$

- Un *vendedor promedio* que saque 77 puntos en el examen venderá mensualmente (μ_{YX}), con un nivel de confianza del 95%:

$$x^* = 77 \mapsto \hat{y} = 4,71 + 0,085 \cdot x^* = 4,71 + 0,085 \cdot 77 = 11,255$$

$$s_{\hat{y}} = s_{YX} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}} = 1,84 \cdot \sqrt{\frac{1}{15} + \frac{(77 - 74)^2}{87076 - \frac{1}{15} 1110^2}} = 0,482$$

Estadístico teórico: $\alpha = 0,05$, $n = 15 \mapsto t_{0,025, 13} = 2,16$

$$\mu_{YX} \in (\hat{y} \pm t_{\alpha/2, n-2} \cdot s_{\hat{y}}) \mapsto \mu_{YX} \in (11,255 \pm 2,16 \cdot 0,482) \mapsto 10,21 < \mu_{YX} < 12,30$$

Un vendedor promedio que saque 77 puntos en el examen venderá mensualmente entre 102100 y 123000 euros, con un nivel de confianza del 95%.

- *Cualquier vendedor* que saque 77 puntos en el examen venderá mensualmente (Y), con un nivel de confianza del 95%:

$$s_{y^*} = s_{yx} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}} = 1,84 \cdot \sqrt{1 + \frac{1}{15} + \frac{(77 - 74)^2}{87076 - \frac{1}{15} 1110^2}} = 1,902$$

$$Y \in (\hat{y} \pm t_{\alpha/2, n-2} \cdot s_{y^*}) \mapsto Y \in (11,255 \pm 2,16 \cdot 1,902) \mapsto 7,15 < Y < 15,36$$

Cualquier vendedor que saque 77 puntos en el examen venderá mensualmente entre 71500 y 153600 euros, con un nivel de confianza del 95%.

