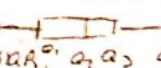


• Mediana muestral $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ • mediana - 2º cuartil = $y_{\frac{n+1}{2}}$

• Varianza $= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = E(x^2) - E(x)^2$ *cuadrado es lo mismo pero en vez de $\frac{1}{n}$, $\frac{1}{n-1} = s^2$*

• Desviación típica $\sigma = \sqrt{\sigma^2}$ • Valores atípicos $x \in [Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$ $IQR = Q_3 - Q_1$

• Histograma barra proporcional a frecuencia  • Casos y bugs 

• Tallar y habar: datos en una tabla *datos en una tabla*  *en la tabla, el resto a la derecha*  *la última a la izquierda de la izquierda de*

• Covarianza $\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ $x > 0 \rightarrow$ asociación +, $x < 0 \rightarrow$ asociación -, $x = 0$ no hay relación

• Modelo de regresión, la recta de regresión es la que minimiza $ECM = \frac{1}{n} \sum_{i=1}^n (y_i - a - bx_i)^2$ *función de coste*

$y - \bar{y} = \frac{\text{cov}(x, y)}{V_x} (x - \bar{x}) \rightarrow y = a + bx = \bar{y} - \frac{\text{cov}(x, y)}{V_x} \bar{x} + \frac{\text{cov}(x, y)}{V_x} x$ $b = r \frac{\sigma_y}{\sigma_x}$ $r > 0 \rightarrow$ relación directa, $r < 0 \rightarrow$ relación inversa, $r = 0 \rightarrow$ relación débil, $r = \pm 1 \rightarrow$ relación fuerte

• Coeficiente de variación $CV = \frac{\sqrt{V_x}}{|\bar{x}|}$ • coeficiente correlación $r_{x, y} = \frac{\text{cov}(x, y)}{\sqrt{V_x V_y}}$ *función de coste*

• Función de masa $P_X(x_i) = P(X = x_i)$ • Función distribución $F_X(x_i) = P(X \leq x_i)$

• Sean a y b constantes
 $E(a) = a$; $E(ax) = a E(x)$; $E(ax + b) = a E(x) + b$; $E(g(x) \pm h(x)) = E(g(x)) \pm E(h(x))$
 $\text{Var}(a) = 0$; $\text{Var}(ax) = a^2 \text{Var}(x)$
 $\text{Var}(g(x) \pm h(x)) = \text{Var}(g(x)) \pm 2 \text{cov}(g(x), h(x)) + \text{Var}(h(x))$

• $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; $P(A^c) = 1 - P(A)$; $P(A \cap B) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq j < k \leq n} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$

• $P(A|B) = \frac{P(A \cap B)}{P(B)}$ si son independientes $P(A|B) = P(A)$
 • $P(A|B) = 1 - P(A^c|B)$ • Probabilidad total $P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n)$

• Dist. marginales distribuciones de variables por separado $P_X(x) = \sum_y P(x, y)$ $F_X(x) = \int_{-\infty}^x f_X(x) dx$
 si son independientes $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ a la vez ind. si $P(x, y) = P(x) P(y)$ y viceversa

• Binomial, n pruebas independientes con probabilidad de éxito p, el número de éxitos
 $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ $E(X) = np$ $V(X) = np(1-p)$ $B(n, p)$ $X \sim Y$ $B(n, p) \sim B(n, p)$

• Geométrica fallar hasta el 1º éxito $P(X=k) = (1-p)^{k-1} p$ $E(X) = \frac{1}{p}$ $V(X) = \frac{1-p}{p^2}$ $GC(p)$

• Poisson $P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$ $E(X) = V(X) = \lambda$ $X \sim P(\lambda)$ $Y \sim P(\lambda_2)$ $X+Y \sim P(\lambda_1 + \lambda_2)$

• Uniforme $f(x) = \frac{1}{b-a}$ $x \in [a, b]$ $E(X) = \frac{a+b}{2}$ $V(X) = \frac{(b-a)^2}{12}$

• Exponencial $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ $\mu = \frac{1}{\lambda}$ $\sigma^2 = \frac{1}{\lambda^2}$ $\int \lambda e^{-\lambda x} dx = \frac{\lambda e^{-\lambda x}}{-\lambda} = -e^{-\lambda x}$ $e^{-\lambda x}$

• Normal $f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$ si $X \sim N(\mu, \sigma) \rightarrow \frac{X-\mu}{\sigma} \sim N(0, 1)$
 $P(Z < \frac{x-\mu}{\sigma}) =$
 $= P(Z > \frac{\mu-x}{\sigma})$ si $X \sim N(\mu, \sigma^2)$ y $a \in \mathbb{R} \rightarrow aX \sim N(a\mu, a\sigma)$ $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

• Cambio entre distribuciones
 • $B(n, p) \approx N(np, \sqrt{np(1-p)})$ si $np > 5$, $n(1-p) > 5$, $p > 0.05$, $1-p > 0.05$
 • $P(\lambda) \approx N(\lambda, \sqrt{\lambda})$ si $\lambda > 5$
 • $B(n, p) \approx P(\lambda)$ $\lambda = np$

• TCL: la suma de n variables aleatorias de media μ y varianza σ^2
 con suma distribuciones $S_n = X_1 + X_2 + \dots + X_n \sim N(n\mu, \sqrt{n\sigma^2})$
 $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N(\mu, \sqrt{\frac{\sigma^2}{n}})$

• Método de momentos: igualar los momentos con análogos
 Máxima verosimilitud $L(\theta; x_1, x_2, \dots, x_n) = P_\theta(x_1) \dots P_\theta(x_n)$
 tomar logaritmos derivar igualar a 0 y calcular

+ $X \sim \text{Bern}(p)$ $L(p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$ $\log(L(p)) = \sum x_i \log p + (n - \sum x_i) \log(1-p)$ $\frac{\partial}{\partial p} \log(L(p)) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0$ $\hat{p} = \frac{\sum x_i}{n} = \bar{x}$

+ $X \sim P(\lambda)$ $L(\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod x_i!}$ $\log(L(\lambda)) = -n\lambda + \sum x_i \log \lambda - \sum \log x_i!$ $\frac{\partial}{\partial \lambda} \log(L(\lambda)) = -n + \frac{\sum x_i}{\lambda} = 0$ $\hat{\lambda} = \frac{\sum x_i}{n} = \bar{x}$

+ $X \sim GP(\lambda)$ $L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$ $\log(L(\lambda)) = n \log \lambda - \lambda \sum x_i$ $\frac{\partial}{\partial \lambda} \log(L(\lambda)) = \frac{n}{\lambda} - \sum x_i = 0$ $\hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$

+ $X \sim N(\mu, \sigma)$ $L(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}}$
 $= \frac{1}{(\sigma \sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2}$ $\log(L(\mu, \sigma)) = -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 - n \log(\sigma \sqrt{2\pi})$
 $= -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 - n [\log \sigma + \log \sqrt{2\pi}]$
 $0 = \frac{\partial}{\partial \mu} \log(L(\mu, \sigma)) = \frac{1}{\sigma^2} \sum (x_i - \mu) = 0$ $\hat{\mu} = \frac{1}{n} \sum x_i = \bar{x}$
 $0 = \frac{\partial}{\partial \sigma} \log(L(\mu, \sigma)) = \frac{1}{\sigma^3} \sum (x_i - \mu)^2 - \frac{n}{\sigma} = 0$ $\hat{\sigma} = \frac{1}{n} \sum (x_i - \bar{x})^2$ *recuerda en xi no es el de arriba*

Intervalos de confianza X_1, \dots, X_n muestra aleatoria de X $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

1) $X \sim N(\mu, \sigma)$ Intervalo de confianza $1-\alpha$ para μ : $S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

- σ conocida $I = (\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$
- σ desconocida $I = (\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}})$
- para σ^2 : $I = (\frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}})$ error en el muestreo (total \pm)

2) $X \sim \text{Bernoulli}(p)$ Intervalo para p : $I = (\bar{x} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}})$

3) $X \sim P(\lambda)$ Intervalo para λ : $I = (\bar{x} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}}{n}})$

4) Dos normales (muestras independientes) $(X_1, \dots, X_m) \sim N(\mu_1, \sigma_1^2)$ \bar{x}, S_1^2
 $(Y_1, \dots, Y_n) \sim N(\mu_2, \sigma_2^2)$ \bar{y}, S_2^2

$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$; Intervalo de confianza $1-\alpha$ para $\mu_1 - \mu_2$:

- σ_1^2, σ_2^2 conocidas $I = (\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}})$
- σ_1^2, σ_2^2 desconocidas, $\sigma_1^2 = \sigma_2^2 \rightarrow I = (\bar{x} - \bar{y} \pm t_{m+n-2, \alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}})$
- σ_1^2, σ_2^2 desc; $\sigma_1^2 \neq \sigma_2^2$ $I = (\bar{x} - \bar{y} \pm t_{F, \alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}})$ $F_{\text{entero + potencia}}$ $(\frac{S_1^2/m + S_2^2/n}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}})^2$

5) Comparación de proporciones $X \sim \text{Bernoulli}(p_1)$ $Y \sim \text{Bernoulli}(p_2)$

$I = (\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}})$ Intervalo de confianza $1-\alpha$ para $p_1 - p_2$

Caracteres de hipótesis

α = nivel de significación n = tamaño muestra H_0 = hipótesis nula R = región crítica / rechazo H_0

- 1) $X \sim N(\mu, \sigma)$
- $H_0: \mu = \mu_0$ (σ conocida) $R = \{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \}$
 - $H_0: \mu = \mu_0$ (σ desconocida) $R = \{ |\bar{x} - \mu_0| > t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \}$
 - $H_0: \mu \leq \mu_0$ (σ conocida) $R = \{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \}$
 - $H_0: \mu \leq \mu_0$ (σ desconocida) $R = \{ \bar{x} - \mu_0 > t_{n-1, \alpha} \frac{s}{\sqrt{n}} \}$
 - $H_0: \mu \geq \mu_0$ (σ conocida) $R = \{ \bar{x} - \mu_0 < -z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \}$
 - $H_0: \mu \geq \mu_0$ (σ desconocida) $R = \{ \bar{x} - \mu_0 < -t_{n-1, 1-\alpha} \frac{s}{\sqrt{n}} \}$
 - $H_0: \sigma = \sigma_0$ $R = \{ \frac{n-1}{\sigma_0^2} S^2 \notin (\chi^2_{n-1, 1-\alpha/2}, \chi^2_{n-1, \alpha/2}) \}$
 - $H_0: \sigma \leq \sigma_0$ $R = \{ \frac{n-1}{\sigma_0^2} S^2 > \chi^2_{n-1, \alpha} \}$
 - $H_0: \sigma \geq \sigma_0$ $R = \{ \frac{n-1}{\sigma_0^2} S^2 < \chi^2_{n-1, 1-\alpha} \}$
- 2) $X \sim \text{Bernoulli}(p)$
- $H_0: p = p_0$ $R = \{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \}$
 - $H_0: p \leq p_0$ $R = \{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \}$
 - $H_0: p \geq p_0$ $R = \{ \bar{x} - p_0 < -z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \}$
- 3) $X \sim \text{Poisson}(\lambda)$
- $H_0: \lambda = \lambda_0$ $R = \{ |\bar{x} - \lambda_0| > z_{\alpha/2} \sqrt{\frac{\lambda_0}{n}} \}$
 - $H_0: \lambda \leq \lambda_0$ $R = \{ \bar{x} - \lambda_0 > z_{\alpha} \sqrt{\frac{\lambda_0}{n}} \}$
 - $H_0: \lambda \geq \lambda_0$ $R = \{ \bar{x} - \lambda_0 < -z_{1-\alpha} \sqrt{\frac{\lambda_0}{n}} \}$
- 4) 2 poblaciones normales con mismas caract que 4 de Intervalos
- $H_0: \mu_1 = \mu_2$ (σ_1^2, σ_2^2 conocidas) $R = \{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \}$
 - $H_0: \mu_1 = \mu_2$ ($\sigma_1 = \sigma_2$) $R = \{ |\bar{x} - \bar{y}| > t_{m+n-2, \alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \}$
 - $H_0: \mu_1 = \mu_2$ ($\sigma_1 \neq \sigma_2$) $R = \{ |\bar{x} - \bar{y}| > t_{F, \alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \}$
 - $H_0: \mu_1 \leq \mu_2$ (σ_1, σ_2 conocidas) $R = \{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \}$
 - $H_0: \mu_1 \leq \mu_2$ ($\sigma_1 = \sigma_2$) $R = \{ \bar{x} - \bar{y} > t_{m+n-2, \alpha} S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \}$
 - $H_0: \mu_1 \leq \mu_2$ ($\sigma_1 \neq \sigma_2$) $R = \{ \bar{x} - \bar{y} > t_{F, \alpha} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \}$
 - $H_0: \mu_1 \geq \mu_2$ (σ_1, σ_2 conocidas) $R = \{ \bar{x} - \bar{y} < -z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \}$
 - $H_0: \mu_1 \geq \mu_2$ ($\sigma_1 = \sigma_2$) $R = \{ \bar{x} - \bar{y} < -t_{m+n-2, 1-\alpha} S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \}$
 - $H_0: \mu_1 \geq \mu_2$ ($\sigma_1 \neq \sigma_2$) $R = \{ \bar{x} - \bar{y} < -t_{F, 1-\alpha} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \}$
 - $H_0: \sigma_1 = \sigma_2$ $R = \{ \frac{S_1^2}{S_2^2} \in (F_{m-1, n-1, 1-\alpha/2}, F_{m-1, n-1, \alpha/2}) \}$
 - $H_0: \sigma_1 \leq \sigma_2$ $R = \{ \frac{S_1^2}{S_2^2} > F_{m-1, n-1, \alpha} \}$
 - $H_0: \sigma_1 \geq \sigma_2$ $R = \{ \frac{S_1^2}{S_2^2} < F_{m-1, n-1, 1-\alpha} \}$
- 5) Comparación
- $H_0: p_1 = p_2$ $R = \{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}} \}$
 - $H_0: p_1 \leq p_2$ $R = \{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}} \}$
 - $H_0: p_1 \geq p_2$ $R = \{ \bar{x} - \bar{y} < -z_{1-\alpha} \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}} \}$
- $\bar{p} = \frac{\sum x_i + \sum y_i}{m+n} = \frac{m\bar{x} + n\bar{y}}{m+n}$