→ 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 $I = \left(\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ • σ desconocida $I = \left(\bar{x} \pm t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right)$

• para σ^2: $I = \left(\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right)$ error es el semiperiodo (lo del ±)

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

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

4) Dos normales (muestras independientes) no independientes se toma nueva V.A $D = X - Y$ y se calc sobre esa $(X_1, \dots, X_m)$ $X \sim N(\mu_1, \sigma_1)$ $\bar{x}, s_1^2$ $(Y_1, \dots, Y_n)$ $Y \sim N(\mu_2, \sigma_2)$ $\bar{y}, s_2^2$

$S_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$; Intervalo de confianza $1-\alpha$ para $\mu_1 - \mu_2$

• σ_1^2 y σ_2^2 conocidas $I = \left(\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right)$

• σ_1^2 y σ_2^2 desconocidas, $\sigma_1^2 = \sigma_2^2$ → $I = \left(\bar{x} - \bar{y} \pm t_{m+n-2;\alpha/2}\, S_p\sqrt{\frac{1}{m} + \frac{1}{n}}\right)$

• σ_1^2 y σ_2^2 desc ; $\sigma_1^2 \neq \sigma_2^2$ $I = \left(\bar{x} - \bar{y} \pm t_{f;\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}\right)$ f entero + próximo $\frac{(s_1^2/m + s_2^2/n)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$

• Intervalo para σ_1^2/σ_2^2 $I = \left(\frac{s_1^2/s_2^2}{F_{m-1;n-1;\alpha/2}}, \frac{s_1^2/s_2^2}{F_{m-1;n-1;1-\alpha/2}}\right)$

5) Comparación de proporciones
$X \sim$ Bernoulli(p_1) $Y \sim$ Bernoulli(p_2)
$I = \left(\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}\right)$ intervalo de confianza $1-\alpha$ para $p_1 - p_2$

→ Contrastes de hipótesis
α = nivel 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) $X \sim$ 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) $X \sim$ Poisson(λ)
$H_0: \lambda = \lambda_0$ $R = \{|\bar{x} - \lambda_0| > z_{\alpha/2}\sqrt{\lambda_0/n}\}$
$H_0: \lambda \leq \lambda_0$ $R = \{\bar{x} - \lambda_0 > z_\alpha\sqrt{\lambda_0/n}\}$ $H_0: \lambda \geq \lambda_0$ $R = \{\bar{x} - \lambda_0 < z_{1-\alpha}\sqrt{\lambda_0/n}\}$

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 $\sigma_1 = \sigma_2$ $R = \{s_1^2/s_2^2 \notin (F_{m-1;n-1;1-\alpha/2}, F_{m-1;n-1;\alpha/2})\}$
H_0 $\sigma_1 \leq \sigma_2$ $R = \{s_1^2/s_2^2 > F_{m-1;n-1;\alpha}\}$
H_0 $\sigma_1 \geq \sigma_2$ $R = \{s_1^2/s_2^2 < F_{m-1;n-1;1-\alpha}\}$

5) Comparación
H_0 $p_1 = p_2$ $R = \{|\bar{x} - \bar{y}| > z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m} + \frac{1}{n})}\}$ $\bar{p} = \frac{\sum x_i + \sum y_i}{m+n} = \frac{m\bar{x} + n\bar{y}}{m+n}$
H_0 $p_1 \leq p_2$ $R = \{\bar{x} - \bar{y} > z_\alpha\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m} + \frac{1}{n})}\}$
H_0 $p_1 \geq p_2$ $R = \{\bar{x} - \bar{y} < z_{1-\alpha}\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m} + \frac{1}{n})}\}$

Una variable

- Media muestral $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ \underset{V.A.}{=} \ \frac{1}{n}\sum_{i=1}^{n} P_X(x_i)\, x_i$ • mediana = 2º cuartil $= y_{\frac{n+1}{4}\cdot 2}$
- σ^2 • Varianza $= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = E(x^2) - E(x)^2$ cuasi es lo mismo pero en vez de $\frac{1}{n}$: $\frac{1}{n-1}$ – S^2
- Desviación típica $\sigma = \sqrt{\sigma^2}$ • Valores atípicos $x \notin [Q_1 - 1'5(IQR),\ Q_3 + 1'5(IQR)]$ $IQR = Q_3 - Q_1$
- Histograma barras proporcionales a frecuencia • Cajas y bigotes $Q_1 - 1'5IQR$ Q_1 Q_2 Q_3 $Q_3+1'5IQR$
- Tallos y hojas: datos en una tabla todas las cifras menos la última a la izquierda de la tabla, el resto a la derecha $5'26$ $5'29$ $5'35$ → $5\,2\,|\,6\,9$; $5\,3\,|\,5$

Dos variables

- $\left[\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right] - \bar{x}\bar{y}$
- Covarianza $cov_{x,y}: \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$ si >0 → asociación +, si <0 → asociación −, si $\simeq 0$ no parece existir relación lineal
- Modelo de regresión, la recta de regresión es la que minimiza $ECM = \frac{1}{n}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = V_y(1-r^2)$

$$y - \bar{y} = \frac{cov_{x,y}}{V_x}(x-\bar{x}) \to y = a + bx = \bar{y} - \frac{cov_{x,y}}{V_x}\bar{x} + \frac{cov_{x,y}}{V_x}x \qquad b = r\frac{\sigma_y}{\sigma_x}$$

- Coeficiente de variación $CV = \frac{\sqrt{V_x}}{|\bar{x}|}$ • coeficiente correlación $r_{x,y} = \frac{cov_{x,y}}{\sqrt{V_x V_y}}$

$r>0$ → relación directa
$r<0$ → relación inversa
$r \simeq 0$ → relación débil
$r \simeq |1|$ → relación fuerte

Probabilidad

- Función de masa → $P_X(x_i) = P(X = x_i)$ • Función distribución $F_X(x_i) = P(X \le x_i)$ • función densidad $\int_{\mathbb{R}} f(x)dx = 1$
- → Sean a y b constantes
$E(a) = a$; $E(aX) = aE(X)$; $E(aX+b) = aE(X)+b$; $E(g(x) \pm h(x)) = E(g(x)) \pm E(h(x))$
$Var(g(x)) = E(g(x)^2) - E(g(x))^2$; $Var(a) = 0$; $Var(aX) = a^2 Var(X)$
- $P(A \setminus B) = P(A) - P(A\cap B)$; $P(A^c) = 1 - P(A)$; $P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j) - \sum_{1\le j<i\le n} P(A_i \cap A_j) + \sum_{1\le j<i<k\le n} P(A_i\cap A_j\cap A_k) \pm \ldots (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$
- $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B|A)\cdot P(A)}{P(B)}$ si son independientes $P(A|B) = P(A)$ (sucesos incompatibles 2 a 2) $\sum_{i=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ $\sum_{i=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$
- $P(A|B) = 1 - P(A^c|B)$ • Probabilidad total $P(B) = P(A_1)\cdot P(B|A_1) + P(A_2)\cdot P(B|A_2) + \ldots P(B|A_n)$

Vectores

- Distr. marginales distribuciones de variables por separado $P_X(x) = \sum P(X=x, Y=y)$ $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$
si son independientes → $f_T(x,y) = F_X(x)\,F_Y(y)$ 2 V.A. son ind. si $P(X=x_i | Y=y_i) = P(X=x_i)$ y viceversa

Distribuciones

- → Binomial, n pruebas independientes con probabilidad de éxito p, el número de éxitos
$P(X=K) = \binom{n}{k} p^k (1-p)^{n-k}$ $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ $E(X) = np$ $V(X) = np(1-p)$ $B(n,p)$ $X+Y = B(n_1+n_2, p)$ $X\sim B(n_1,p)$ $Y\sim B(n_2,p)$
- → Geométrica fallos hasta el 1º éxito $P(X=k) = (1-p)^k p$ $E(X) = \frac{1-p}{p}$ $V(X) = \frac{1-p}{p^2}$ $G(p)$
- → Poisson $P(X=K) = e^{-\lambda}\frac{\lambda^k}{k!}$ $E(X) = V(X) = \lambda$ $X\sim P(\lambda_1)$ $Y\sim P(\lambda_2)$ → $X+Y \sim P(\lambda_1+\lambda_2)$
- → Uniforme $F(x) = \begin{cases} \frac{1}{b-a} & x\in[a,b] \\ 0 \end{cases}$ $E(X) = \frac{a+b}{2}$ $V(X) = \frac{(b-a)^2}{2}$
- → Exponencial $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x>0 \\ 0 \end{cases}$ $\mu = \frac{1}{\lambda}$ $\sigma^2 = \frac{1}{\lambda^2}$ $\int \lambda e^{-\lambda x}dx = \frac{\lambda e^{-\lambda x}}{-\lambda}$ $exp(\lambda)$
- → Normal $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x-\mu)^2}{2\sigma^2}$ si $X\sim N(\mu,\sigma) \to \frac{x-\mu}{\sigma} \sim N(0,1)$
$P(Z < \frac{x-\mu}{\sigma}) =$
$= P(Z > \frac{\mu - x}{\sigma})$ si $X \to N(\mu,\sigma^2)$ y $a\in\mathbb{R} \to aX \to N(a\mu, a\sigma)$ $X_1+X_2 \to N(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2})$

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 $\lambda \ge 5$
- $B(n,p) \simeq P(\lambda)$ $\lambda = np$

acuérdate de pasar de desviación a varianza

- → TCL: la suma de n variables aleatorias de media μ y varianza σ^2 con misma distribución $S_n = X_1+X_2+\ldots+X_n \sim N(n\mu, \sqrt{n\sigma^2})$
$\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n} \sim N\left(\mu, \sqrt{\frac{\sigma^2}{n}}\right)$

— Método de momentos: igualamos momentos con análogos $E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$ $E(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \ldots$ → $\sum_{i=1}^{n} P_X(i)\cdot i$
Máxima verosimilitud $L(\theta, x_1, x_2, \ldots, x_n) = P_\theta(x_1)\ldots P_\theta(x_n)$
tomamos logaritmos derivamos igualamos a 0 y calculamos

+ $X\sim Bern(p)$ $L(p) = \prod_{j=1}^{n} p^{x_j}(1-p)^{1-x_j} = p^{\sum_{j=1}^{n} x_j}(1-p)^{\sum_{j=1}^{n} 1-x_j}$ $\log(L(p))' = \frac{\sum_{j=1}^{n} x_j}{p} + \frac{-n+\sum_{j=1}^{n} x_j}{(1-p)}$ $\log(L(p))' = 0$, $\hat{p} = \frac{1}{n}\sum_{j=1}^{n} x_j = \bar{x}$
+ $X\sim P(\lambda)$ $L(\lambda) = \prod_{j=1}^{n} e^{-\lambda}\cdot\frac{\lambda^{x_i}}{x_i!} = \frac{e^{\sum_{j=1}^{n} -\lambda}\cdot\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$; $\log(L(\lambda)) = -n\lambda + \log(\lambda)\cdot\sum x_i - \log\prod_{i=1}^{n} x_i!$ $\log(L(\lambda))' = 0$ $-n + \frac{\sum x_i}{\lambda} = 0$ $\hat{\lambda} = \bar{x}$
+ $X\sim Exp(\lambda)$ $L(\lambda) = \prod_{j=1}^{n} \lambda e^{-\lambda x_i}$ $L(\lambda) = \lambda^{\sum 1} e^{\sum -\lambda x_i}$ $\log(L(\lambda)) = n\lambda + \sum -\lambda x_i$ $\log(L(\lambda))' = \frac{n}{\lambda} - \sum x_i$ $\hat{\lambda} = \bar{x}^{-1}$
+ $X\sim N(\mu,\sigma)$ $L(\mu,\sigma) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{(\sigma\sqrt{2\pi})^{\sum_{i=1}^{n} 1}} \cdot e^{-\frac{\sum(x-\mu)^2}{2\sigma^2}} =$
$= \frac{1}{(\sigma\sqrt{2\pi})^n}\cdot e^{-\frac{1}{2\sigma^2}\sum(x-\mu)^2}$ $\log(L(\mu,\sigma)) = -\frac{1}{2\sigma^2}\sum(x-\mu)^2 \log(e)^{1} + \log(\sigma\sqrt{2\pi})^{-n} =$
$= -\frac{1}{2\sigma^2}\sum(x-\mu)^2 - n[\log\sigma + \log\sqrt{2\pi}]$
$0 = \frac{\partial}{\partial\mu}(\log(L(\mu,\sigma))) = \frac{(\sum_{j=1}^{n} x_j) - n\mu}{\sigma^2} = 0$ $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
$0 = \frac{\partial}{\partial\sigma}(\log(L(\mu,\sigma))) = \frac{1}{\sigma^3}(-n\sigma^2 + \sum_{j=1}^{n}(x_j-\mu)^2)$ $\hat{\sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \hat{\mu})$

+ $X\sim Geom(p)$
$\hat{p} = \frac{n}{\sum x_i + n}$
recuerda en x_i no contar el acierto