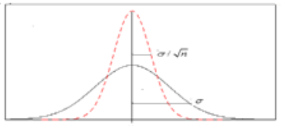
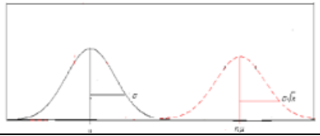
Propietats de les mostres i Intervals de Confiança

TCL: $X_1,\ldots,X_n$ i.i.d. ($n\to\infty$), amb $E(X_i)=\mu$ i $V(X_i)=\sigma^2$, llavors $\dfrac{\sum_{i=1}^n X_i}{n}=\overline{X}_n \approx N(\mu,\sigma^2/n)$ **(i també** $\sum_{i=1}^n X_i \approx N(n\mu,\sigma^2 n)$ **)**





Estadístic mitjana mostral ($\overline{x}$): $\dfrac{(\overline{x}-\mu)}{\sqrt{\sigma^2/n}} \approx N(0,1)$ $\dfrac{(\overline{x}-\mu)}{\sqrt{s^2/n}} \approx t_{n-1}$ on $\overline{x}=\sum_{i=1}^n x_i \Big/ n$

Estadístic variància mostral (s^2): $s^2\dfrac{n-1}{\sigma^2} \approx \chi^2_{n-1}$ on $s^2=\dfrac{\sum_{i=1}^n (x_i-\overline{x})^2}{n-1}=\dfrac{\sum_{i=1}^n x_i^2-n(\overline{x})^2}{n-1}=\dfrac{\sum_{i=1}^n x_i^2-\dfrac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}$

Paràmetre	Estadístic	Premisses	Distribució	Interval de Confiança 1-α (Risc α)
μ	$\hat{z}=\dfrac{(\overline{x}-\mu)}{\sqrt{\sigma^2/n}}$	[X ~ N o $n \geq \approx 30$] i σ coneguda	$\hat{z}$ ~ N(0,1)	$\mu \in (\overline{x} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sigma^2}{n}})$
μ	$\hat{t}=\dfrac{(\overline{x}-\mu)}{\sqrt{s^2/n}}$	X ~ N	$\hat{t}$ ~ t_{n-1}	$\mu \in (\overline{x} \pm t_{n-1,1-\alpha/2}\sqrt{\dfrac{s^2}{n}})$
μ	$\hat{z}=\dfrac{(\overline{x}-\mu)}{\sqrt{s^2/n}}$	$n \geq \approx 100$	$\hat{z}$ ~ N(0,1)	$\mu \in (\overline{x} \pm z_{1-\alpha/2}\sqrt{\dfrac{s^2}{n}})$
σ (normal)	$\hat{X}^2=\dfrac{s^2(n-1)}{\sigma^2}$	X ~ N	$\hat{X}^2$ ~ χ^2_{n-1}	$\sigma^2 \in \left(\dfrac{S^2(n-1)}{\chi^2_{n-1,1-\alpha/2}},\dfrac{S^2(n-1)}{\chi^2_{n-1,\alpha/2}}\right)$
π (Binomial)	$\hat{z}=\dfrac{(p-\pi)}{\sqrt{\pi(1-\pi)/n}}$	$(1-\pi)n \geq \approx 5$ $\pi n \geq \approx 5$	$\hat{z}$ ~ N(0,1)	$\pi \in (P \pm z_{1-\alpha/2}\sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}})$ $\hat{\pi}=P$ o $\hat{\pi}=0.5$
λ (Poisson)	$\hat{z}=\dfrac{(L-\lambda)}{\sqrt{\lambda}}$	$\lambda \geq \approx 5$	$\hat{z}$ ~ N(0,1)	$\lambda \in (L \pm z_{1-\alpha/2}\sqrt{L})$

Proves d'Hipòtesis

Paràmetre	Hipòtesi nul·la	Estadístic	Premisses	Distribució sota H0	Criteri Decisió (Risc α)
	$H_0 : \mu = \mu_0$	$\hat{z} = \dfrac{(\bar{y}-\mu_0)}{\sqrt{\sigma^2/n}}$	Y~ N o n≥≈*30* i σ coneguda	$\hat{z} \sim N(0,1)$	Rebutjar H_0 si $\lvert \hat{z} \rvert > z_{1-\alpha/2}$ ($\lvert\hat{z}\rvert>1'96$ amb $\alpha=5\%$)
μ	$H_0 : \mu = \mu_0$	$\hat{t} = \dfrac{(\bar{y}-\mu_0)}{\sqrt{S^2/n}}$	Y ~ N	$\hat{t} \sim t_{n-1}$	Rebutjar H_0 si $\lvert \hat{t} \rvert > t_{n-1,1-\alpha/2}$ ($\lvert \hat{t} \rvert > t_{n-1,0'975}$ amb $\alpha=5\%$)
μ	$H_0 : \mu = \mu_0$	$\hat{z} = \dfrac{(\bar{y}-\mu_0)}{\sqrt{S^2/n}}$	n ≥ ≈*100*	$\hat{z} \sim N(0,1)$	Rebutjar H_0 si $\lvert \hat{z} \rvert > z_{1-\alpha/2}$
π (Binomial)	$H_0 : \pi = \pi_0$	$\hat{z} = \dfrac{(p-\pi_0)}{\sqrt{\pi_0(1-\pi_0)/n}}$	$(1-\pi_0)n \geq \approx 5$ $\pi_0 n \geq \approx 5$	$\hat{z} \sim N(0,1)$	Rebutjar H_0 si $\lvert \hat{z} \rvert > z_{1-\alpha/2}$ ($\lvert\hat{z}\rvert>1'96$ amb $\alpha=5\%$)
Anexe: λ (Poisson)	$H_0 : \lambda = \lambda_0$	$\hat{z} = \dfrac{(f-\lambda_0)}{\sqrt{\lambda_0}}$	$\lambda_0 \geq \approx 5$	$\hat{z} \sim N(0,1)$	Rebutjar H_0 si $\lvert \hat{z} \rvert > z_{1-\alpha/2}$ ($\lvert\hat{z}\rvert>1'96$ amb $\alpha=5\%$)
σ (normal)	$H_0 : \sigma = \sigma_0$	$\hat{X}^2 = \dfrac{S^2(n-1)}{\sigma^2}$	Y ~ N	$\hat{X}^2 \sim \chi^2_{n-1}$	Rebutjar H_0 si $\hat{X}^2 < \chi^2_{n-1,\alpha/2}$ o $\hat{X}^2 > \chi^2_{n-1,1-\alpha/2}$

En les proves unilaterals s'acumula el valor de *P* a un sol costat	**H_0: $\mu \leq \mu_0$ → Rebutjar H_0 si $\hat{z} > z_{1-\alpha}$** **H_0: $\mu \geq \mu_0$ → Rebutjar H_0 si $\hat{z} < -z_{1-\alpha}$**

Proves de μ i σ en 2 mostres

Paràmetre	Hipòtesi nul·la	Estadístic	Premisses	Distrib. sota H_0	Decisió (Risc α)
μ	$H_0 : \mu_1 = \mu_2$	$\hat{z} = \dfrac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	[Y_1 Y_2 ~ N *o* n_1 n_2 $\geq\approx$*30*] m.a.s. ind. i σ_1 σ_2 conegudes	$\hat{z}$ ~ N(0,1)	Rebutjar si $\lvert \hat{z} \rvert >$ **$Z_{1-\alpha/2}$** ($\lvert\hat{z}\rvert$>1'96 amb α=5%)
μ	$H_0 : \mu_1 = \mu_2$	$\hat{t} = \dfrac{(\bar{y}_1 - \bar{y}_2)}{S\sqrt{1/n_1 + 1/n_2}}$ $S^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$	Y_1 , Y_2 ~ N $\sigma_1 = \sigma_2$ m.a.s indep.	$\hat{t}$ ~ $t_{n1+n2-2}$	Rebutjar si $\lvert \hat{t} \rvert > t_{n1+n2-2,\ 1-\alpha/2}$ $\lvert\hat{t}\rvert > t_{n1+n2-2,0'975}$ amb α=5%
μ	$H_0 : \mu_1 = \mu_2$	$\hat{z} = \dfrac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	n_1 , n_2 $\geq\approx$*100* m.a.s indep	$\hat{z}$ ~ N(0,1)	Rebutjar si $\lvert \hat{z} \rvert >$ **$Z_{1-\alpha/2}$**
μ	$H_0 : \mu_1 = \mu_2$	$\hat{t} = \dfrac{\bar{D} - \mu_0}{S_D/\sqrt{n}}$	D ~ N m.a. aparellada	$\hat{t}$ ~ t_{n-1}	Rebutjar si $\lvert \hat{t} \rvert > t_{n-1,\ 1-\alpha/2}$
σ (normal)	$H_0 : \sigma^2_1 = \sigma^2_2$	$\hat{F} = S_A^2 / S_B^2$ Sent $S_A^2 > S_B^2$	Y_1 , Y_2 ~ N m.a.s. indep	$\hat{F}$ ~ $F_{nA-1,nB-1}$	Rebutjar si $\hat{F} > F_{nA-1,nB-1,\ 1-\alpha/2}$

Les corresponents proves unilaterals es fan acumulant el risc α a un costat

Proves de π en 2 mostres

Proves de Comparació de 2 Paràmetres més usuals				
Hipòtesis	Estadístic	Premisses	Distrib.(H0)	Decisió (α=0.05)
H_0: $\pi_1 = \pi_2 = \pi$ H_1: $\pi_1 \neq \pi_2$	$z = \frac{(P_1 - P_2)}{\sqrt{P(1-P)/n_1 + P(1-P)/n_2}}$ $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$	$e_{ij} \geq 5 \ \forall$ ij m.a.s indep.	$\hat{z} \sim N(0,1)$	Rebutjar si $\lvert \hat{z} \rvert > 1.96$
	$\hat{x}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$		$\hat{\chi}^2 \sim \chi^2_{(I-1)(J-1)}$	Rebutjar si $\hat{\chi}^2 > \chi^2_{(I-1)(J-1),0.95}$
Homogeneïtat $H_0 : \pi(A_j \mid B_i) = \pi(A_j \mid B_{i'}) \quad \forall$ i, i', j $H_1 : \exists$ j t.q. $\pi(A_j \mid B_i) \neq \pi(A_j \mid B_{i'})$ Independència $H_0 : \pi(A_j \cap B_i) = \pi(A_j)\ \pi(B_i) \quad \forall$ i, j $H_1 : \exists$ ij t.q. $\pi(A_j \cap B_i) \neq \pi(A_j)\ \pi(B_i)$	$\hat{x}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$	$e_{ij} \geq 5 \ \forall$ ij m.a.s indep.	$\hat{\chi}^2 \sim \chi^2_{(I-1)(J-1)}$	Rebutjar si $\hat{\chi}^2 > \chi^2_{(I-1)(J-1),0.95}$
H_0: $\pi_1 = \pi_2 = \pi$ H_1: $\pi_1 \neq \pi_2$	$\hat{x}^2 = \frac{(a-b)^2}{(a+b)}$	a, b $\geq$ 5 m.a. aparellades	$\hat{\chi}^2 \sim \chi^2_1$	Rebutjar si $\hat{\chi}^2 > \chi^2_{1,0.95}$
Les corresponents proves unilaterals es fan acumulant el risc α a un costat.				

Model lineal (quantitativa vs quantitativa)

$$b_1 = \frac{S_{XY}}{S_X^2} = r\,\frac{S_Y}{S_X}$$

$$b_0 = \overline{Y} - b_1\,\overline{X}$$

$$S^2 = \frac{\sum e_i^2}{n-2} = \frac{(n-1)\,S_Y^2\,(1-r^2)}{n-2} = \frac{(n-1)\left(S_Y^2 - b_1 S_{XY}\right)}{n-2}$$

$$\overline{y} = \frac{\sum y_i}{n} \qquad s_y^2 = \frac{\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}}{n-1} \qquad s_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{n-1}$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y}) \big/ (n-1)}{S_X S_Y} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}}$$

Estimació i inferència dels paràmetres [ML]

Paràmetre	β_0	β_1	σ^2
Estimador	$b_0 = \bar{Y} - b_1\bar{X}$	$b_1 = S_{XY} / S^2_X$	$S^2 = \Sigma e_i^2/(n-2)$
Esperança	$E(b_0) = \beta_0$	$E(b_1) = \beta_1$	$E(S^2) = \sigma^2$
Error tipus	$S_{b_0} = \sqrt{S^2\left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)S_x^2}\right)}$	$S_{b_1} = \sqrt{\frac{S^2}{(n-1)S_x^2}}$	
Distribució	$(b_0-\beta_0) / S_{b0} \sim t_{n-2}$	$(b_1-\beta_1) / S_{b1} \sim t_{n-2}$	$(n-2)S^2/\sigma^2 \sim \chi^2_{n-2}$
Interval de Confiança	$IC(\beta_0,95\%) =$ $= b_0 \pm t_{n-2,0.975} \cdot S_{b0}$	$IC(\beta_1,95\%) =$ $= b_1 \pm t_{n-2,0.975} \cdot S_{b1}$	$IC(\sigma^2,95\%) =$ $(n-2)S^2/\chi^2_{n-2,0.975} \leq \sigma^2$ $\leq (n-2)S^2/\chi^2_{n-2,0.025}$
H_0 usual	$\beta_0 = 0$	$\beta_1 = 0$	
Rebutjar H_0 si	$\lvert b_0 / S_{b0} \rvert > t_{n-2,0.975}$	$\lvert b_1 / S_{b1} \rvert > t_{n-2,0.975}$	

	Estimació puntual	Estimació per interval 95%	
Predicció		Per al valor esperat	Per a valors individuals
	$\hat{y}_h = b_0 + b_1 X_h$	$\hat{y}_h \pm t_{n-2,0.975} S \sqrt{\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$	$\hat{y}_h \pm t_{n-2,0.975} S \sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$