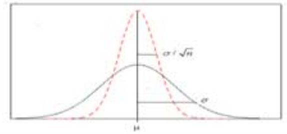
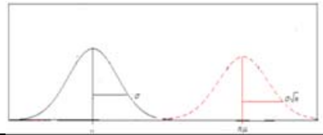
Propietats de les mostres i Intervals de Confiança

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





Estadístic mitjana mostral ($\bar{X}$): $\dfrac{(\bar{X}-\mu)}{\sqrt{\sigma^2/n}} \approx N(0,1)$ $\quad \dfrac{(\bar{X}-\mu)}{\sqrt{s^2/n}} \approx t_{n-1}$ on $\bar{X} = \sum_{i=1}^{n} X_i \big/ n$

Estadístic variància mostral (s^2): $S^2 \dfrac{n-1}{\sigma^2} \approx \chi^2_{n-1}$ on $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \dfrac{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}{n-1} = \dfrac{\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$

Paràmetre	Estadístic	Premisses	Distribució	Interval de Confiança 100(1-α)%
μ	$Z = \dfrac{(\bar{X}-\mu)}{\sqrt{\sigma^2/n}}$	[X ~ N o $n \geq \approx 30$] i σ coneguda	$Z \sim N(0,1)$	$\mu \in (\bar{X} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sigma^2}{n}})$
μ	$T = \dfrac{(\bar{X}-\mu)}{\sqrt{S^2/n}}$	X ~ N	$T \sim t_{n-1}$	$\mu \in (\bar{X} \pm t_{n-1,1-\alpha/2}\sqrt{\dfrac{S^2}{n}})$
μ	$Z = \dfrac{(\bar{X}-\mu)}{\sqrt{S^2/n}}$	$n \geq \approx 100$	$Z \sim N(0,1)$	$\mu \in (\bar{X} \pm z_{1-\alpha/2}\sqrt{\dfrac{S^2}{n}})$
σ (normal)	$X^2 = \dfrac{S^2(n-1)}{\sigma^2}$	X ~ N	$X^2 \sim \chi^2_{n-1}$	$\sigma^2 \in \left(\dfrac{S^2(n-1)}{\chi^2_{n-1,1-\alpha/2}}, \dfrac{S^2(n-1)}{\chi^2_{n-1,\alpha/2}}\right)$
π (Binomial)	$Z = \dfrac{(P-\pi)}{\sqrt{\pi(1-\pi)/n}}$	(1-π) $n \geq \approx 5$ π $n \geq \approx 5$	$Z \sim N(0,1)$	$\pi \in (P \pm z_{1-\alpha/2}\sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}})$ $\hat{\pi} = P \quad o \quad \hat{\pi} = 0.5$

Proves d'Hipòtesis

Paràmetre	Hipòtesi nul·la	Estadístic	Premisses	Distribució sota H_0	Criteri Decisió (Risc α)
μ	$H_0 : \mu = \mu_0$	$Z = \dfrac{(\bar{Y} - \mu_0)}{\sqrt{\sigma^2 / n}}$	Y~ N o $n \geq\approx 30$ i σ coneguda	$Z \sim N(0,1)$	Rebutjar H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
μ	$H_0 : \mu = \mu_0$	$T = \dfrac{(\bar{Y} - \mu_0)}{\sqrt{S^2 / n}}$	Y ~ N	$T \sim t_{n-1}$	Rebutjar H_0 si $\lvert T \rvert > t_{n-1,1-\alpha/2}$
μ	$H_0 : \mu = \mu_0$	$Z = \dfrac{(\bar{Y} - \mu_0)}{\sqrt{S^2 / n}}$	$n \geq \approx 100$	$Z \sim N(0,1)$	Rebutjar H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
π (Binomial)	$H_0 : \pi = \pi_0$	$Z = \dfrac{(P - \pi_0)}{\sqrt{\pi_0 (1 - \pi_0) / n}}$	$(1-\pi_0)\, n \geq \approx 5$ $\pi_0\, n \geq \approx 5$	$Z \sim N(0,1)$	Rebutjar H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
σ (normal)	$H_0 : \sigma = \sigma_0$	$X^2 = \dfrac{S^2 (n-1)}{\sigma^2}$	Y ~ N	$X^2 \sim \chi^2_{n-1}$	Rebutjar H_0 si $X^2 < \chi^2_{n-1,\alpha/2}$ o $X^2 > \chi^2_{n-1,1-\alpha/2}$

En les proves unilaterals s'acumula el risc a un sol costat	**H_0: $\mu \leq \mu_0$** $\rightarrow$ Rebutjar H_0 si $Z > z_{1-\alpha}$ **H_0: $\mu \geq \mu_0$** $\rightarrow$ Rebutjar H_0 si $Z < -z_{1-\alpha}$

Proves de μ i σ en 2 mostres

Paràmetres	Hipòtesi nul·la	Estadístic	Premisses	Distrib. sota H_0	Decisió (Risc α)
μ	$H_0 : \mu_1 = \mu_2$	$Z = \dfrac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	[Y_1, $Y_2 \sim N$ *o* n_1 $n_2 \geq\approx$*30*] m.a.s. indep. i σ_1 σ_2 conegudes	$Z \sim N(0,1)$	Rebutjar H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
μ	$H_0 : \mu_1 = \mu_2$	$T = \dfrac{(\bar{Y}_1 - \bar{Y}_2)}{S\sqrt{1/n_1 + 1/n_2}}$ $S^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$	Y_1, $Y_2 \sim N$ $\sigma_1 = \sigma_2$ desconegudes m.a.s indep.	$T \sim t_{n1+n2-2}$	Rebutjar H_0 si $\lvert T \rvert > t_{n1+n2-2,1-\alpha/2}$
μ	$H_0 : \mu_1 = \mu_2$	$Z = \dfrac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	n_1, $n_2 \geq\approx$*100* m.a.s indep.	$Z \sim N(0,1)$	Rebutjar H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
μ	$H_0 : \mu_D = \mu_0$ ($\mu_D = \mu_1 - \mu_2$)	$T = \dfrac{\bar{D} - \mu_0}{S_D/\sqrt{n}}$	$D \sim N$ m.a. aparellada	$T \sim t_{n-1}$	Rebutjar H_0 si $\lvert T \rvert > t_{n-1,1-\alpha/2}$
σ	$H_0 : \sigma^2_1 = \sigma^2_2$	$F = S_A^2 / S_B^2$ Sent $S_A^2 > S_B^2$	Y_1, $Y_2 \sim N$ m.a.s. indep.	$F \sim F_{nA-1,nB-1}$	Rebutjar si $F > F_{nA-1,nB-1,\ 1-\alpha/2}$

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.(H_0)	Decisió (risc α)
H_0: $\pi_1 = \pi_2 = \pi$ H_1: $\pi_1 \neq \pi_2$	$Z = \dfrac{(P_1 - P_2)}{\sqrt{P(1-P)/n_1 + P(1-P)/n_2}}$ $P = \dfrac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$	$e_{ij} \geq 5\ \forall$ i,j m.a.s indep.	$Z \sim N(0,1)$	Rebutjar * H_0 si $\lvert Z \rvert > z_{1-\alpha/2}$
	$X^2 = \sum_{\forall ij} \dfrac{(f_{ij} - e_{ij})^2}{e_{ij}}$		$X^2 \sim \chi^2_1$	Rebutjar si $X^2 > \chi^2_{1,\,1-\alpha}$
H_0 : $\pi(A_j \cap B_i) = \pi(A_j)\,\pi(B_i)$ H_1 : $\exists$ i,j t.q. $\pi(A_j \cap B_i) \neq \pi(A_j)\,\pi(B_i)$ i=1...I, j=1...J	$X^2 = \sum_{\forall ij} \dfrac{(f_{ij} - e_{ij})^2}{e_{ij}}$	$e_{ij} \geq 5\ \forall$ i,j m.a.s indep.	$X^2 \sim \chi^2_{(I-1)(J-1)}$	Rebutjar si $X^2 > \chi^2_{(I-1)(J-1),\,1-\alpha}$
* La corresponent prova unilateral es fa acumulant el risc α a un costat.				

Model lineal

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$S_Y{}^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} Y_i{}^2 - n(\overline{Y})^2}{n-1} = \frac{\sum_{i=1}^{n} Y_i{}^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}}{n-1}$$

$$S_{XY} = \frac{\sum_i x_i Y_i - \frac{\sum_i x_i \sum_i Y_i}{n}}{n-1}$$

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

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

$$b_0 = \overline{Y} - b_1 \overline{x}$$

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

Estimació i inferència dels paràmetres model lineal

Paràmetre	β_0	β_1	σ^2
Estimador	$b_0 = \overline{Y} - b_1\overline{x}$	$b_1 = S_{XY} / S^2_X$	$S^2 = \Sigma e_i^2/(n-2)$
Esperança	$E(b_0) = \beta_0$	$E(b_1) = \beta_1$	$E(S^2) = \sigma^2$
Variància	$S^2_{b_0} = S^2\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)S^2_X}\right)$	$S^2_{b_1} = \frac{S^2}{(n-1)S^2_X}$	$V(S^2) = 2\sigma^4/(n-2)$
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	$\beta_0 = \beta'_0$	$\beta_1 = \beta'_1$	
Rebutjar H_0 si	$\lvert (b_0 - \beta'_0) / S_{b0} \rvert > t_{n-2,0.975}$	$\lvert (b_1 - \beta'_1) / S_{b1} \rvert > t_{n-2,0.975}$	

Previsions	Estimació puntual	Estimació per interval 95%	
		Per al valor esperat	Per a valors individuals
	$Y_h = \hat{y}_h = b_0 + b_1 x_h$	$\hat{y}_h \pm t_{n-2,0.975} S \sqrt{\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum (x_i - \overline{x})^2}}$	$Y_h \pm t_{n-2,0.975} S \sqrt{1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum (x_i - \overline{x})^2}}$