BLOC C

BC : PS1 du 15/11
PS2 du 22/11

Inferència estadística

- Per fer inferència cal:
  - Dades de una mostra aleatòria
  - Hem d'assumir simetria

Paràmetres poblacionals (θ) (100% població)

- μ: media (valor esperat)
- σ: desviació (saber si pot variar molt)
- π: proporció de persones fan algo (és un exemple)

que es inferen

θ pot ser qualsevol dels 3, englobo tots.

Estadístics ($\hat{\theta}$) (mesures del 100% de la població)

- $\bar{X}$: media mostral
- S: Variació estandar
- P: proporció de persones amb una característica

que estimen

$\hat{\theta}$ pot ser qualsevol dels 3

$$S = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x - \bar{x})^2$$

- La pregunta: com arribem als estadístics de valoració?

<u>Estimador</u> (estadístic que posa info sobre algún paràmetre)

$[\hat{\mu} = \bar{X}$ estimador de $\mu]$

Que estimem

Que estimem info de

Propietats estimadors

→ <u>Biaix</u> (sesgo): em interessa que vigui proper a 0, per tal assegurar resultats fiables

→ <u>Eficiència</u>: em interessa que vigui <u>alta</u>

Estimació de μ amb σ coneguda

<u>Exemple: embotelladora funciona bé?</u>

- volem omplir ampolles amb 1L ≡ 1000 cc
- Observem $\bar{X} = 995$ cc ($= \hat{\mu}$) → (amb només això no podem garantir res, és imprescindible:
  - mitja mostral: μ
  - var dispersió: σ)

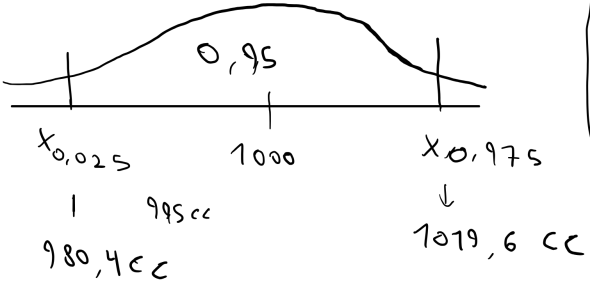
- Objectiu: calcular intervals de confiança, amb un cert nivell de confiança de saber que no te errar gran:

$$IC(\mu, 1-\alpha) = [L, R] \text{ tal que } P(L \leq \mu \leq R) = \underline{\underline{1-\alpha}}$$

μ → sobre quin paràmetre

$1-\alpha$ → nivell confiança

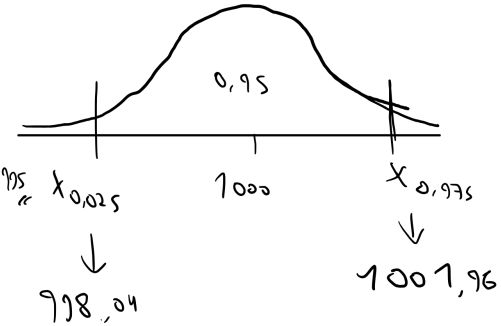
Suposem $X \sim N(1000, 10)$



[Cap problema que es amplí amb 995 cc!]

Ara amb $n = 100$, i volem mitjana de les 100:

$$\bar{X}_{100} \sim N\left(1000, \frac{10}{\sqrt{100}}\right)$$



[Problema amb la mitjana de 995 cc!! en vez de 95%!!]

• Cómo obtenemos μ a partir de $\bar{x} = 995$ cc?

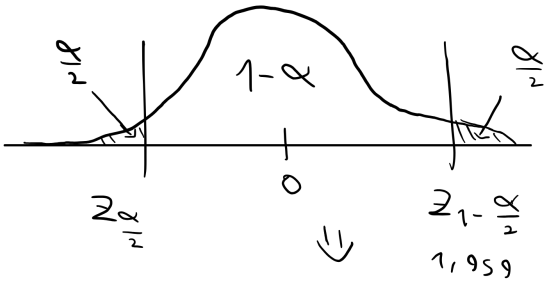
$$\left.\begin{array}{l} n = 100 \\ \bar{x}_n = 995 \\ \sigma = 10 \end{array}\right\} IC(\mu, 0.95) = ¿?$$

<u>TCL</u>

¿?

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$



$$\Downarrow$$

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Downarrow$$

$$P\left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

4,223

$$IC(\mu, 0.95) = \bar{X}_n \mp z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 995 \mp 1{,}96 \cdot \frac{10}{\sqrt{100}} = [993{,}04 \,,\, 996{,}96]$$

$$\boxed{\bullet\ IC(\mu, 1-\alpha) = \bar{X}_n \mp z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$$

– Te diu que les μ que desconeixem, amb una infinitat, les μ estan amb un 95% probabilitat entre aquell interval, però no sabem el valor de μ.

$$\left.\begin{array}{l} \boxed{\frac{\sigma}{\sqrt{n}} = \text{se (standard error)}} \\ \\ \boxed{z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = \text{marge d'error}} \end{array}\right\} \text{Són part de la fórmula}$$

tepé

- → σ gran, tenen més marge d'error (- precisió)
- → n gran, tenen més petit marge d'error (+ precisió)
- → $1-\alpha$ gran, tenen més marge d'error (- precisió)

Exercici

$\bar{X}_n = 995$, $n = 50$, $1-\alpha = 0{,}99$, $\alpha = 0{,}01$, $\sigma = 10$

$N(0,1)$

$-z_{1-\frac{\alpha}{2}}$ $\quad 0 \quad$ $z_{1-\frac{\alpha}{2}}$

$\Downarrow$ $\qquad$ $0{,}995$

$-2{,}5758$ $\qquad$ $\Downarrow$

$2{,}5758$

$IC(\mu, 0{,}99) = [991.36, 998.64]$

$1-\alpha = 0{,}99$ $\qquad \sigma = 10$

$n = ¿?$

$\uparrow_y$

margen error de 1

ex: $[994, 996]$

$n = 666$

$$\int (X+1) = X + 2{,}58 \cdot \frac{10}{\sqrt{n}}$$

$$\Downarrow$$

$$1 = \frac{25{,}8}{\sqrt{n}}$$

$$\sqrt{n} = 25{,}8$$

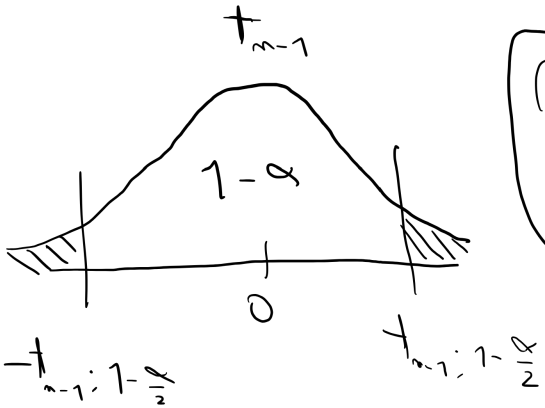
$$\boxed{n = 666}$$

Estimación de μ sin saber σ (t-student)

$$\hat{\sigma} = s = \sqrt{\frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2}$$

$$\boxed{\frac{\bar{X}_n - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}}$$

t := Es una distribución muy similar a una normal, de ser posibilidades son de grados de libertad



(cuando σ desconocido)

$$IC(\mu, 1-\alpha) = \bar{X} \pm t_{n-1;\,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

Ej: $\bar{x} = 995$, $\underline{S = 10}$, $n = 50$, $1-\alpha = 0{,}99$

- Daría un intervalo <u>más</u> <u>grande</u> porque tomamos una estimación de σ

$$IC(\mu, 0.99) = 995{,}5 \mp \underbrace{t_{49,\,0.995}}_{2{,}68} \cdot \frac{10}{\sqrt{50}} = [991{,}2,\ 998{,}8]$$

[t grans de g.l. ⇒ més igual a $N(0,1)$]

Estimació de π

$IC(\pi, 1-\alpha) = \uparrow \pm$ marge d'error

$X_1 \sim B(1, \pi)$

$\mu = 1 \cdot \pi = \pi$

$\sigma^2 = 1 \cdot \pi \cdot (1-\pi)$

$\hat{\pi} = \bar{X} = \uparrow \Rightarrow$ per el TCL $\uparrow = \tilde{X}_n \sim N\left(\pi, \sigma^2 = \frac{\pi(1-\pi)}{n}\right)$

$\Downarrow$

$$\frac{\uparrow - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

$\hat{\pi} = \uparrow$

si n gran

$$\boxed{t_{n-1} \sim \frac{\uparrow - \pi}{\sqrt{\frac{\uparrow(1-\uparrow)}{n}}}} \approx \boxed{N(0,1)}$$

0,06

0,0778

$$\boxed{IC(\pi, 1-\alpha) = p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}}$$

p = proporció observada

Exemple:

$IC(\pi, 0.95)$, $n = 1023$, marge d'error?

$\hookrightarrow \alpha = 0{,}05$ $\quad z_{1-\frac{0{,}05}{2}} = 1{,}96$

$1 - \frac{0{,}05}{2} = 0{,}975$

$$1.e = 1{,}96 \cdot \frac{\sqrt{0{,}5 \cdot 0{,}5}}{\sqrt{1000}} = \underline{0{,}031} \checkmark$$

Exemple 2:

$$IC(\mu, 1-\alpha) = \bar{X}_n \mp t_{n-1;\, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$1 - \alpha = 0{,}9 \qquad \alpha = 0{,}1 \qquad 1 - \frac{0{,}1}{2} = 0{,}95$

X: hores de deporte per setmana en estudiantes de la FIB

$\bar{X}_n = 5{,}639$

σ

$n = 133$

$$IC(\mu, 0.9) = 5{,}639 \mp 1{,}656 \cdot \frac{4{,}387}{\sqrt{133}}$$

$$= [5 \;,\; 6{,}27]$$

$t_{132;\, 0{,}95} = 1{,}656$

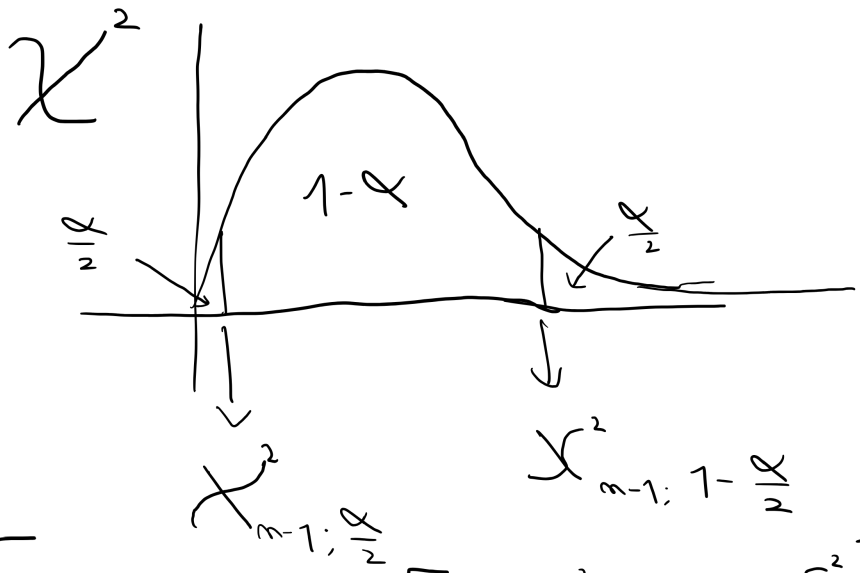
Estimació de σ^2

$n = 30$ $\qquad S^2 \not\sim N$

$\hat{\sigma} = S = 10$ c.c $\qquad S^2 \sim ¿?$

- Es pot demostrar que S: $X_1, X_n \sim N(\mu, \sigma)$ i independents $\Rightarrow$

$$\left[(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} \right]$$



$$\left[IC(\sigma^2, 1-\alpha) = \left[\frac{(n-1)\cdot S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} , \frac{(n-1)\cdot S^2}{\chi^2_{n-1;\frac{\alpha}{2}}} \right] \right]$$

$n = 30, \quad \hat{\sigma} = S = 10, \quad IC(\sigma, 0.95) \Rightarrow \alpha = 0{,}05$

$\frac{\alpha}{2} = 0{,}025 \qquad \chi^2_{29, \frac{\alpha}{2}} = 16{,}05$

$1 - \frac{\alpha}{2} = 0{,}975 \qquad \chi_{29, 1-\frac{\alpha}{2}} = 45{,}72$

$$IC(\sigma^2, 0.95) = \left[\frac{29 \cdot 10^2}{45{,}72}, \frac{29 \cdot 10^2}{16{,}05}\right] = [7.96, 13{,}44]$$

IC por 2 parámetros

$\mu_1 - \mu_2$ independientes

- Ejemplo: programas, con T: tiempo de ejecución con distintos bancos de datos

Prog 1:

$\bar{Y}_1 = 24$ ms, $S_1 = 8$ ms, $n_1 = 50$

son independientes

Prog 2:

$\bar{Y}_2 = 21$ ms, $S_2 = 6$ ms, $n_2 = 100$

$\hat{\mu}_1 - \hat{\mu}_2 = \bar{Y}_1 - \bar{Y}_2 = 24 - 21 = 3$ ms

- <u>Varianzgleichheit</u>: $\sigma_1 = \sigma_2$

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}\right)$$

$$\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

$\Downarrow$ ro Ersetzen σ

$$\boxed{\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\hat{S} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}}$$

$t_{n_1 + n_2 - 2}$

$$\boxed{IC(\mu_1-\mu_2, 1-\alpha) = \bar{Y}_1 - \bar{Y}_2 \mp t_{n_1+n_2-2,\, 1-\frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Calcular $\hat{\sigma}$ (variança conjunta)

$$\hat{\sigma} = \sqrt{\frac{1}{n_1+n_2-2} \cdot \left((n_1-1)\cdot \hat{\sigma}_1^2 + (n_2-1)\cdot \hat{\sigma}_2^2\right)}$$

en l'exemple:

$$\hat{\sigma} = \sqrt{\frac{1}{50+100-2} \cdot \left((50-1)\cdot 8^2 + (100-1)\cdot 6^2\right)} = 6{,}728$$

$$IC(3m2,\ 0.95) = 3 \mp 1.976 \cdot 6{,}728 \cdot \sqrt{\frac{1}{50}+\frac{1}{100}} = \boxed{0.7\ ,\ 5.3}$$

$\alpha = 0{,}05$

$1-\frac{\alpha}{2} = 0{,}975$

$t_{100+50-2,\ 0.975} = 1{,}976$

Em és t, el may 2 és millor

$(-3, 7)$ no es pot saber

$\mu_1 - \mu_2$ emparellades

Y_1 var $E = \mu_1$ i σ_1 $\quad i, \dots, n$

Y_2 var $E = \mu_2$ i σ_2

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n} - 2 \cdot \mathrm{cov}(X, Y)}\right)$$

Definim $D = Y_1 - Y_2 \rightsquigarrow$ 1 muestra : $d_1 = Y_{11} - Y_{21}, \dots, d_n = Y_{1n} - Y_{2n}$

$$\boxed{IC(\mu_D, 1-\alpha) = \bar{d} \mp t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}}$$

Ejemplo:

$n = 6$

$\bar{d} = 2{,}95$

$\alpha = 1$

$t_{5, 0.975} = 2.571$

$$2{,}95 \pm 2.571 \cdot \frac{6}{\sqrt{6}} = [1.9 \ , \ 4]$$

$$\frac{\sigma_1^2}{\sigma_2^2}$$

Exemple:

$S_1 = 8$, $n_1 = 50$

$S_2 = 6$, $n_2 = 100$

Homocedasticitat??

$\sigma_1^2 = \sigma_2^2 = \sigma^2$??

$$IC\left(\frac{\sigma_1^2}{\sigma_2^2}, 1-\alpha\right)$$

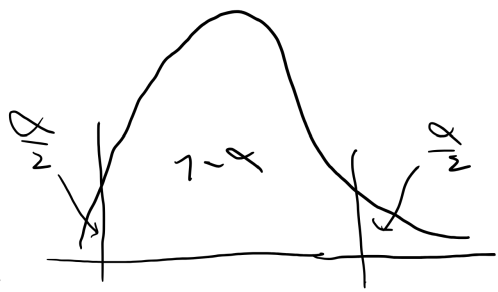
$\downarrow$

Hauria de contindre el 1 l'IC

Estimació puntual:

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{S_1^2}{S_2^2} = \frac{64}{36} = 1.78$$

$$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F_{n_1-1,\, n_2-1}$$



$$P\left(f_{n_1-1,n_2-1,\frac{\alpha}{2}} \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq f_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Downarrow$$

$$IC\left(\frac{\sigma_1^2}{\sigma_2^2}, 1-\alpha\right) = \left[\frac{\frac{S_1^2}{S_2^2}}{f_{n_1-1,n_2-1,1-\frac{\alpha}{2}}}, \frac{\frac{S_1^2}{S_2^2}}{f_{n_1-1,n_2-1,\frac{\alpha}{2}}}\right]$$

$$f_{49,49,0.025} = 0.6 \qquad f_{49,49,0.975} = 1,6$$

$$\left[\sqrt{\frac{1,78}{1,6}}, \sqrt{\frac{1,78}{0,6}}\right] = \left[1.05, 1.72\right]$$

IC de $\pi_1 - \pi_2$

$$IC(\pi_1 - \pi_2, 1-\alpha) = (p_1 - p_2) \pm z_{1-\frac{\alpha}{2}} \cdot se$$

$$se = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$