4 85 — Estimation, échantillonnage et test

Dans une série : $SCE_x = SC - n \times m^2$ $\qquad K = \frac{N-n}{N-1}$

$s_x^2 = \frac{SCE_x}{n}$

Echantillon : $\sigma_m = \frac{\sigma_x}{\sqrt{n}} \sqrt{K}$ $\qquad P\left[\mu - t_{1-\frac{\alpha}{2}} \sigma_m < m < \mu + t_{1-\frac{\alpha}{2}} \sigma_m\right] = 1-\alpha$

$\sigma_f = \sqrt{\frac{pq}{n}} \sqrt{K}$

Estimation variance :

$$\frac{SCE}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma_x^2 < \frac{SCE}{\chi^2_{\frac{\alpha}{2}}(\nu)}$$

$ME_{rel} = ME_{abs} \times \frac{100}{m}$

$ME_{abs} = t \times \sigma_m = t \times \frac{\sigma_x}{\sqrt{n}}$

↳ on en déduit $\sigma_x$ !

$\hat{\sigma}_x^2 = \frac{SCE_x}{n-1}$

→ Choix entre 2 paramètres

$t_{1-\alpha \text{ ou } \alpha} = \frac{\Pi - \mu_1}{\sigma_1(m)} \qquad t_{\beta \text{ ou } 1-\beta} = \frac{\Pi - \mu_2}{\sigma_2(m)}$

→ Comparaison moyenne à valeur donnée

* $\sigma$ connu $t_{calc} = \frac{m - E_m}{\sigma_m}$ loi Normale si $\mathcal{L}(x) \to \mathcal{N}$ ou si $n > 30$

si on rejette $H_0$ → $\alpha$ risque de se tromper

pvalue $= P(T > |t_{calc}|)$ ou $2P(T > |t_{calc}|)$ ... bilat

* $\sigma$ inconnu loi Student (~~seul~~ si $\mathcal{L}(x) \to \mathcal{N}$ et $n < 30$)

→ Comparaison proportion à valeur donnée

$t_{calc} = \frac{f - E_f}{\sigma_f} \qquad \sigma_f = \sqrt{\frac{pq}{n}} \sqrt{K}$

→ Comparaison d'une variance à une valeur donnée

$\chi^2_{calc} = \frac{SCE_x}{\sigma^2}$ → on accepte si $\chi^2_{\frac{\alpha}{2}} < \chi^2 < \chi^2_{1-\frac{\alpha}{2}}$

→ Comparaison de 2 variances

$F_{calc} = \frac{\sigma^2 \text{ le + grd}}{\sigma^2 \text{ le + petit}}$ → on accepte si calc < théo

→ Comparaison de 2 séries

$t_{calc} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sigma_d(m)} \qquad \sigma_d(m) = \sqrt{\sigma_1^2(m) + \sigma_2^2(m)}$

↳ si on rejette $H_0$ car $|t_{calc}| > t_{théo}$ ⇒ risque < $\alpha$ %

# COMPARAISONS

## moyenne / valeur

$\hookrightarrow$ si $\sigma_x$ connue : $t_{calc} = \dfrac{m - E_m}{\sigma_m}$ avec $\sigma_m = \dfrac{\sigma_x}{\sqrt{n}}\sqrt{K} \rightarrow \mathcal{N}$

$\hookrightarrow$ si $\sigma_x$ inconnue : " avec $\sigma_x = \dfrac{SCE_x}{n-1} \longrightarrow St$

## proportion / valeur

$$t_{calc} = \frac{f - E_f}{\sigma_f} \qquad \sigma_f = \frac{\sqrt{pq}}{\sqrt{n}}\sqrt{K} \longrightarrow \mathcal{N}$$

## variance / valeur

$$\chi^2_{calc} = \frac{SCE_x}{\sigma_0^2} \longrightarrow \chi^2$$

## variance / variance

$$F_{calc} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > 1$$

## série / série

$\hookrightarrow$ éch. indpd : pop $\mathcal{N}$ + $\sigma$ connue $\rightarrow \mathcal{N}$
pop $\mathcal{N}$ + $n < 30 \rightarrow St$
pop ? + $n > 30 \rightarrow \mathcal{N}$

$$\left.\right\} t_{calc} = \frac{\Delta m - \Delta\mu}{\sigma_d(m)}$$

avec $\sigma_d(m) = \sqrt{\sigma_2^2 + \sigma_1^2}$

$\hookrightarrow$ échan. appariés $t_{calc} = \dfrac{\bar{d} - \delta}{\sigma(\bar{d})}$

$\bar{d}$ — différence moyenne des moyennes !
$\delta \rightarrow$ différence théorique

$$\sigma(\bar{d}) = \frac{\hat{\sigma}_d}{\sqrt{n}}\sqrt{K}$$

## prop / prop

$$t_{calc} = \frac{f_1 - f_2}{\sigma_d(f)} \neq \qquad \sigma_d(f) = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{\frac{pq}{n_1}K_1 + \frac{pq}{n_2}K_2} \qquad p_0 = \frac{k_1 + k_2}{n_1 + n_2}$$

## Etapes test d'hypothèse

1. Variable ? Loi ? Infos ?
2. Hypothèses ?
3. Considérer $\alpha$
4. critère stat ?
5. Définir $\Pi$
6. Conclure

## Bartlett, pour comparer + de 2 variances

$$\chi^2_{calc} = \left(\frac{2{,}3026}{C}\right) \times \left[\nu \times \log\hat{\sigma}^2 - \sum_1^k \left(\nu_j \times \log\hat{\sigma}_j^2\right)\right]$$

Cette formule ↑ n'est pas à savoir.
Ses composantes, si :

$$C = 1 + \frac{1}{3(k-1)} \times \left[\sum_1^k \frac{1}{\nu_j} - \frac{1}{\nu}\right]$$

NB : k est le nb d'échant.

Pour les autres composantes faire un tableau malin :

| Éch | $n$ | $\nu$ | $SCE$ | $\hat{\sigma}^2$ | $\log\hat{\sigma}_j^2$ | $\nu_j\hat{\sigma}_j^2$ |
|---|---|---|---|---|---|---|
| $U_1$ | $n_1$ | | | | | |
| $\vdots$ | | | | | | |
| $U_j$ | $n_j$ | $n_j - 1$ | | | | |
| $\vdots$ | | | | | | |

et $\nu = \sum \nu_j$

$$\hat{\sigma}^2 = \frac{1}{\nu} \times \sum \nu_j\hat{\sigma}_j^2$$

TADA !

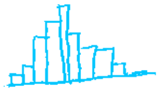
# les Stats... dit autrement

## Le(s) facteur(s) étudié a-t-il (ont-ils) un effet ?

① → <u>Modèle</u> : – combien de facteurs ? modalités ? répétitions ?
– combien de traitements ? unités d'expérience ?
⟹ QUELLE EST LA REPONSE ETUDIEE ?

② → <u>Etude des résidus</u> :

– suivent-ils une loi normale ?

- <u>HISTOGRAMME</u>. si il est symétrique et en cloche, genre  alors l'hypothèse de normalité n'est pas réfutée

- <u>COEF. DE PEARSON</u>
  - ↳ asymétrie : si $\beta_1 \simeq 0$, c'est plutôt symétrique
    si $\beta_1 \neq 0$, c'est pas très symétrique
    la p-value, c'est le risque d'avoir tort en disant "c'est pas symétrique"
  - ↳ aplatissement : si $\beta_2 \simeq 3$, c'est plutôt mésocurtique
    si $\beta_2 \neq 3$, c'est pas très mésocurtique
    la p-value, c'est le risque d'avoir tort en disant "c'est pas en cloche"

– sont-ils bien aléatoires (sans aucun [illegible])

- CARTOGRAPHIE (seulement pour les études de terrain)

– les facteurs ne font-ils bien varier <u>que</u> la moyenne ?
(soit) les variances sont elles à peu près les mêmes

- HYPOTHESE D'HOMOSEDASTICITE

→ avec Bartlett
le $\chi^2$ (> 2 modalités)
le F (= 2 modalités)

③ → test ANOVA

| Source de variance | SCE | ddl | CM | $F_{calc}$ | $F_{théo}$ (ddl, ddle) |
|---|---|---|---|---|---|

si $F_c < F_t$ ou pval $> 0{,}05$ → pas d'effect facteur

si $F_c > F_t$ ou pval $< 0{,}05$ → effet facteur

④ → PPDS Quels groupes sont vraiment différents ?

$$ppds = t_{1-\frac{\alpha}{2}} \sqrt{\frac{CM_e}{n_i} \times 2}\ *$$

donné par l'ANOVA

NB : ranger les moyennes par ordre ↗

table de Student avec d.d.l.e

Si la différence entre les 2 moyennes est < ppds, alors c'est m̂ groupe.

* si les effectifs varient

$$ppds = t_{1-\frac{\alpha}{2}} \sqrt{\frac{CM_e}{\text{[illegible]}} \quad \frac{CM_e}{\text{[illegible]}}}$$

## Echantillonnage / estimation

$\sigma_x^2 = \frac{SCE}{n-1}$

$$\boxed{\begin{array}{l} \text{Si } \frac{n}{N} \leqslant 10\% \text{ ou si remise alors } K = 1 \\ \text{Sinon, } \quad K = \frac{N-n}{N-1} \end{array}}$$

$$P\left[E(m) - t_{\frac{1-\alpha}{2}}\,\sigma(m) < m_n < E(m) + t_{\frac{1+\alpha}{2}}\,\sigma(m)\right] = 1 - \alpha$$

$\mu$ ou $p$ — selon la loi — $\frac{\sigma_x \text{ ou } \sqrt{pq}}{\sqrt{n}}\sqrt{K}$

($m = p$ pour la freq.)

→ estimation d'une proportion (si c'est pas normal)

$\omega = n - k$

**FISHER**

$$p_1 = \frac{k}{k + (\omega + 1) \times F_{1-\frac{\alpha}{2}}(\nu_1, \nu_2)} \qquad \nu_1 = 2(\omega + 1), \quad \nu_2 = 2k$$

$$p_2 = \frac{k}{\omega + (k+1) \times F_{1-\frac{\alpha}{2}}(\nu_1, \nu_2)} \qquad \nu_1 = 2(k+1), \quad \nu_2 = 2\omega$$

**$\chi^2$**

$$p_1 = \frac{1}{2n}\,\chi^2_{\frac{\alpha}{2}}(2k)$$

$$p_2 = \frac{1}{2n}\,\chi^2_{1-\frac{\alpha}{2}}(2k+2)$$

→ estimation d'une variance

$$p_1 = \frac{\sum_1^n (x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}$$

$$p_2 = \frac{\sum_1^n (x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

campudia

C'est pas

Exercice 3

1 → tableau à compléter

2 → $\alpha = 0{,}05$

ANOVA

| Facteur | ~~981,326~~ SCE | ddl | CM | $F_{calc}$ | $F_{théo}(ddl_{\hat{y}} ; ddl_e)$ |
|---|---|---|---|---|---|
| $\hat{y}/U_1$ | 981,326 | 1 | 981,326 | 25,7581 | 4,35 |
| résid | 761,955 | 20 | 38,0977 | | |
| tot | 1743,281 | 21 | | | |

$F_{calc} \gg F_{théo}$ donc on rejète $H_0$ et on admet $H_1$: que ce modèle est satisfaisant (modèle additif linéaire de 1er degré) !!

$t = \sqrt{F}$

$t = \frac{b_2}{\sigma(b_2)}$

p-value < 0,5 → on garde $H_0$

p-value > 0,5 → on rejete $H_0$