Statistique Discriptive:

• Mesures de tendance centrale.

Moyenne d'échantillon $\bar{X} = \frac{\sum x_i}{N}$

" de populat° $= \bar{\mu} = \frac{\sum x_i}{N}$.

Mediane : $i = \frac{N}{2}$

décimal → on prend $> \frac{N}{2}$.

entiers → $\frac{i + (i+1)}{2}$.

Mode. La valeur la plus fréquente.

Percentile, ex: 85% d'observat°.

$\frac{85}{100} \times (N) \Rightarrow$ Nombre d'observat°.

Nombre avec les virgules

décimal → la valeurs entiers suivantes.

ex: 10,2 ⇒ 11ème observ.

Entiers → $\frac{obser. \; i + (i+1)}{2}$.

ex: 6 ⇒ on prend la moyen de 6ème et 7ème observ/2.

Quartile. 25°, 50° et 75°

les 25% d'une observation.

$i = \frac{25}{100} \times N \Rightarrow$ même chose pour le résultat.

• Mesures de dispersion:

Etendue. Plus grand valeur – Plus petite valeur.

Etendue inter quartile. (EIQ)

3ème Quartile – 1er Quartile.

La variance:

De populat°:

$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$

De échantillon:

$V(x) = \sum x^2 - (\bar{x})^2$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

On interprete le $(\sqrt{V(x)})$ aus "ecart type" non la variance $(V(x))$ ou (s^2) car la $(V(x))$ est au carré, d'ou les résultat sans sens ex: eleve² → sans sens

L'écart-type: $s = \sqrt{s^2}$ ou $\sigma = \sqrt{\sigma^2}$.

Coefficient de variance : $\frac{\sqrt{s^2}}{moyenne} \times 100$

ex: 20% ⇒ s correspond à 20% de la moyenne

Mesure la dispersion relative de rapport de écart-type à la moyenne × 100.

• Détection des valeurs singulières

• Forme d'une distribut°: Degree d'asymétrie sert à detecter les valeurs singulières.

Les donnes à gauche, sont d'un asymétrie négatif

Les donnes à droite sont d'un asymétrie positif

$Y_1 = \frac{\mu_3}{\sigma^3}$ avec $\mu_3 = \frac{1}{n} \sum n_i (x_i - \bar{x})^3$

• Variable centrée réduite Z

$z_i = \frac{x_i - \bar{x}}{s}$ → mesure la distance entre s et $(x_i - \bar{x})$. appliquer pour tout les lois

• Théoreme de chebyshev: permet de déduire le pourcentage d'obs. dans un interval $\left(1 - \frac{1}{z^2}\right)$ avec $[\bar{x} - z s; \bar{x} + z s]$ avec $z > 1$

ex. $\bar{x} = 70; s = 5; [60; 80]$.

60 —10— 70 —10— 80

$z \times 5 = 10 \Rightarrow z = 2$.

$70 - z \times 5 = 2$ $70 + z \times 5 = 2$.

d'ou : $z = 2$,

$1 - \frac{1}{z^2} = 1 - \frac{1}{4} = 0{,}75$ observ.

75% d'etudiant obtenu une note entre 60 et 80

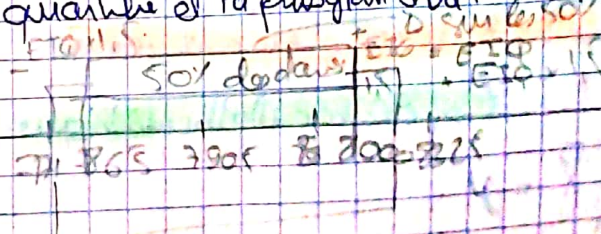
• Règle empirique: déduire le % pour 2 ou 3 s (appliquer seulement pour la loi normale)

• Analyse exploratoire des donnees: Résumer en 5 % en utilisant la plus petite, 1er quartile, médiane, 3ème quartile et la plus grand valeur.

50% des donnes

EIQ 1,5

EIQ 1,5



- Échantillon :

→ $P = \frac{n}{P}$ ⇒ proportion de population

→ Type d'échantillon : – Aléatoire :

- Aléatoire stratifié : devise la pop. en sous groupe homogène.
  ex : 1ère année, 2ème année, 3ème année, 5ème année } des sous grp homogène

- Par grappe : chaque grappe est une mini population contient toutes les caractéristique de la pop ≠ homogène.
  ex : 100 avion.
  Chaque avion contient des jeunes, des vieux, femme, hommes, éco-class, business class, moyen class.

- Systématique !
  ex : pop : 5000, $n = 50$
  1. – Calcul $K = \frac{N}{n} = \frac{5000}{50} = 100$
  2. – Choisir un valeur aléatoire $0 \leq n < 50$ — 20
  3. – Cinquante de nb choisit 100 jusqu'à le 50ème nbre. ⇒ $\underbrace{20, 120, \ldots, n}_{50}$

– Non aléatoire : Raisonnable

- De commodité : des échantillon disponible
- Subjectif : un échantillon le plus représentatif de la pop.
  ex : une thème politique → échantillon des gens politiques.
- Participat° volontaire : Par volonté.

⇒ Estimation ponctuelle :

– Distribution :

D'après TD.

- Distribution d'échantillon $\bar{x}$:

1 – L'espérance Mathématique ; Moyen.

$\bar{X} = \mu$.

2 – L'écart-type :

facteur type

- Pop finie : $\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$

Erreur type

- Pop infinie : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

⚠ : Pop infinie lorsque :
– $\frac{n}{N} \leq 0{,}05$; $n \leq 5/N$
– Aucun infos sur la pop.

- Distribution de la proportat° d'échant.

1 – $\bar{p} = \frac{n}{n}$.

2 – $E(\bar{p}) = P$.

3 – L'écart-type.

- Pop finie : $\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \times \sqrt{\frac{P(1-P)}{n}}$

- Pop infinie : $\sigma_{\bar{p}} = \sqrt{\frac{P(1-P)}{n}}$

La probabilité :

- La probabilité que $\bar{x}$ s'écarte de y de la μ

↳ $\mu - y \leq \bar{X} \leq \mu + y$

$\frac{\mu - \mu - y}{\sigma} \leq \frac{\bar{X} - \mu}{\sigma} < \frac{\mu + y - \mu}{\sigma}$

- La probabilité que $\bar{x}$ s'écarte de y et z

$P(y < \bar{X} \leq z)$.

$\frac{y - \mu}{\sigma} < \frac{\bar{X} - \mu}{\sigma} < \frac{z - \mu}{\sigma}$.

⇒ Pour agrandir la probabilité soit grandir la condition ex : Seq, soit diminuer $\sigma_{\bar{X}}$, soit agrandir la taille, pour avoir des résultat pertinente.

⇒ Estimation par intervalle :

• De la moyenne de pop :

• $\bar{X} \pm Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$ ⇒ σ connu de pop

• $\bar{X} \pm Z_{\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}$ ⇒ σ inconnu. → écart type d'échantillon → en cherchant pour (n-1)

La marge d'erreur :

$$E = Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$1-\alpha$: coefficient de confiance.

• De la proportion de pop :

$$\bar{p} \pm Z_{\sigma/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

⇒ Taille d'échantillon

Par moyenne

$$n = \frac{(Z_{\alpha/2})^2 \times \sigma^2}{E^2}$$

Par proportion :

$$n = \frac{(Z_{\alpha/2})^2 \times p^\circ(1-p^\circ)}{E^2}$$

$p^\circ = P(1-P) = 0{,}5$
Peut être car 0,5 est le max.

Coef de détermination :

$$\frac{SC\ reg}{SC\ T} = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(\hat{y}-\bar{y})^2 + \sum(y_i-\hat{y})^2}$$

Coef de corrélat° :

r_{xy} = signe de $b_1 \sqrt{r^2}$

$$r^2 = (r_{xy})^2 = \left(\frac{cov}{\sigma_x \times \sigma_y}\right)^2$$

$$Cov = \frac{\sum(x-\bar{x})(y-\bar{y})}{n}$$

• Regression linéaire :

$Y = \beta_0 + \beta_1 x + \varepsilon$ ⇒ erreur

$y = ax + b$

→ $\hat{y} = b_0 + b_1 x$

⇒ Par la méthode moindre carré

$$b_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x_i-\bar{x})^2}$$

$b_0 = \bar{y} - b_1 \bar{x}$.

$SC\ res = \sum(y_i - \hat{y}_i)^2$.

⇒ Coef de détermination : fournit une mesure de l'adéquation de l'équation estimée de la regression.

$SCT = \sum(y_i - \bar{y})^2$ $SC_{res} = \sum(y-\hat{y})^2$

$SC\ reg = \sum(\hat{y}-\bar{y})^2$.

. Coef de corrélation :

r_{xy} = (signe de b_1) $\sqrt{r^2}$

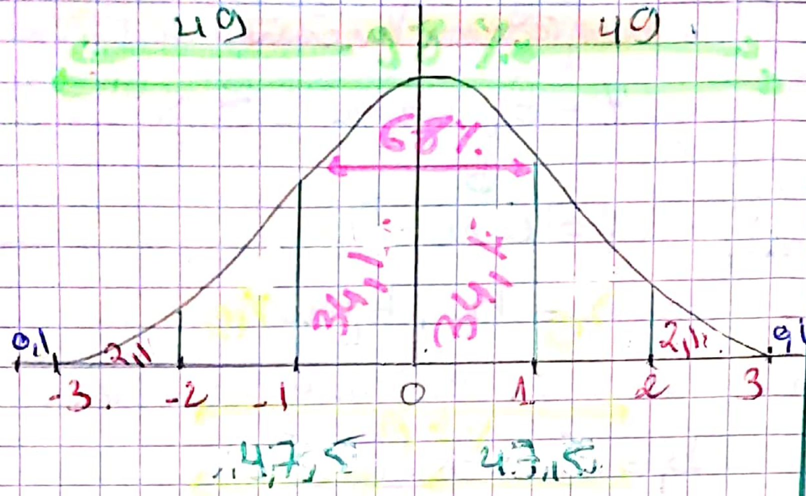
. $SCT = SC\ reg + SC\ res$.

↳ Cas idéal $SCT = 0$, sans erreur ou des écarts.

. Hypothèse :

$$\sigma_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

$$a = \frac{\sum(x-\bar{x})\sum(y-\bar{y})}{\sum(x-\bar{x})^2}$$



Loi binomiale.

$np \geq 5$

$n(1-p) \geq 5$

- Hypothèse du modèle.

Terme d'erreur ε

1_ $E(\varepsilon) = 0 \Rightarrow E(Y) = \beta_0 + \beta_1 x$

2_ La variance de ε notée σ^2 est la même pour toutes les valeurs de x

3_ Les valeurs de ε sont indépendantes entre elles

4_ Le terme d'erreur ε est une v.a. normalement distribuée (et donc y aussi).

. Test de signification:

$E(Y) = \beta_0 + \beta_1 x$.

$\beta_1 = 0$ $E(Y) = \beta_0$ (avec $\beta_1 \neq 0$).

H_0 : A rejeter l'opposé

H_1 : Hypothèse à être valable.

Test inférieur	test supérieur	test bilatéral
$z \leq -z_\alpha$	$z \geq z_\alpha$	$z \leq -z_{\frac{\alpha}{2}}$ ou $z \geq z_{\frac{\alpha}{2}}$

valeur critique

avec σ connu, si non on fait la même chose mais avec la loi de student.

- $Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$. Statistique de test

1_ Test de student:

Concernant β_1

$H_0 = \beta_1 = 0$; H_a : $\beta_1 \neq 0$.

avec β_1 et β_0 sont échantillonnage, avec leur propre distribution d'échantillon.

_ Espérance : $E(b_1) = \beta_1$

- $\sigma_{\beta_1} = \sigma_{b_1} = \frac{\sigma}{\sqrt{\sum(x_i - \bar{X})^2}}$

Si σ non connu :

$S_{b_1} = \frac{S}{\sqrt{\sum(x_i - \bar{X})^2}}$

Estimation σ^2 :

$s^2 = MCres = \frac{SCres}{n-2}$.

une estimat° sans biais

. Erreur type d'estimat°. $b_0 = b$

$s = \sqrt{MCres} = \sqrt{\frac{SCres}{n-2}}$

Statistique de test = $t = \frac{b_1}{S_{b_1}}$ a

$b_1 \rightarrow y = b_1 x + b_0$

Rejet de H_0

App. par VL critique $t \leq -t_{\frac{\alpha}{2}}$ ou $t \geq t_{\frac{\alpha}{2}}$

Intervalle de confiance pour β_1

$\beta_1 \in [b_1 \pm t_{\frac{\alpha}{2}} \times S_{b_1}]$

2_ Test de Fisher:

F : Statistique du Test = $\frac{MCreg}{MCres}$.

avec :

- $MCreg = \frac{SCreg}{1} \rightarrow$ Nb de ddl.
- $MCres = \frac{SCres}{N-2}$

Rejet de H_0.

$F \geq F_\alpha \rightarrow$ Table Fisher.

Loi binomiale.

$E(x) = n \times p \rightarrow$ Moyenne.

$\sigma^2 = np(1-p)$

Scanné avec CamScanner