Chapitre 5 : Estimation

5.1 « Prédiction » de la valeur d'une loi normale

A) Loi normale centrée réduite

Si $Z \sim \mathcal{N}(0\,;\,1)$, alors

- $$\mathbb{P}[-1{,}96 \leqslant Z \leqslant 1{,}96] = \underbrace{F(1{,}96)}_{0{,}975} - \underbrace{F(-1{,}96)}_{1-\underbrace{F(1{,}96)}_{0{,}975}}$$
  $$\simeq 0{,}975 - (1 - 0{,}975) \simeq 0{,}95$$
- Donc avec la confiance 95%,
  Z doit être entre -1,96 et 1,96
- $\mathbb{P}[-2{,}576 \leqslant Z \leqslant 2{,}576] \simeq 0{,}99$
- Donc avec la confiance 99%,
  Z doit être entre -2,576 et 2,576
- Pour d'autres niveaux de confiance : table en bas de la page 10 du formulaire.

B) Loi normale générale

Si $X \sim \mathcal{N}(\mu\,;\,\sigma)$, alors

- $$\mathbb{P}\left[\mu - 1{,}96\sigma \leqslant X \leqslant \mu + 1{,}96\sigma\right]$$
  $$= \mathbb{P}\Big[\underbrace{\frac{\mu - 1{,}96\sigma - \mu}{\sigma}}_{-1{,}96} \leqslant \underbrace{\frac{X-\mu}{\sigma}}_{Z} \leqslant \underbrace{\frac{\mu + 1{,}96\sigma - \mu}{\sigma}}_{1{,}96}\Big]$$
  $$= \mathbb{P}\left[-1{,}96 \leqslant Z \leqslant 1{,}96\right] \simeq 0{,}95$$
- Donc avec la confiance 95%,
  X doit être entre $\mu - 1{,}96\sigma$ et $\mu + 1{,}96\sigma$
- Autres niveaux de confiance : remplacer 1,96 par la valeur issue de la table en bas de la page 10 du formulaire

5.2 Estimation d'une proportion

Question : Quelle est la proportion p des étudiant·e·s Dijonnais qui sont stressé·e·s ?

- Échantillon de 75 étudiant·e·s, parmi lesquels X sont stressé·e·s

A) Ce qu'on pourrait dire si on connaissait la proportion p

Par exemple si on sait que $p = 0{,}4$, alors
$X \approx \mathcal{B}(75; 0{,}4) \approx \mathcal{N}(\underbrace{30}_{75\times 0{,}4}\,;\,\underbrace{4{,}24}_{\sqrt{75\times 0{,}4\times 0{,}6}})$.

Avec la confiance 95%,
X est donc dans l'intervalle $[\underbrace{22}_{30-1{,}96\times 4{,}24}\,;\,\underbrace{38}_{30+1{,}96\times 4{,}24}]$

Conclusion : si dans notre échantillon le nombre d'individus stressés n'est pas dans cet intervalle, on conclura avec la confiance 95% que ces probabilités sont incorrectes, et que la proportion p n'était pas égale à 0,4.

B) Si on ne connaît pas p (on essaie de l'estimer)

$X \approx \mathcal{B}(75,p) \approx \mathcal{N}(75\,p\,;\,\sqrt{75 \times p \times (1-p)})$

$P = \frac{X}{75}$: proportion qui sont stressé·e·s parmi l'échantillon

$m(P) = \frac{m(X)}{75} = p$ et $s(P) = \frac{s(X)}{75} = \sqrt{\frac{p\times(1-p)}{75}}$, et on peut montrer que

$$P \approx \mathcal{N}(p\,;\,\underbrace{\sqrt{\frac{p\times(1-p)}{75}}}_{s}).$$

Avec la confiance 95% P est dans l'intervalle

$$[p - 1{,}96\,s\,;\,p + 1{,}96\,s].$$

C) Estimation

Supposons qu'on réalise l'expérience : parmi 75 étudiant·e·s choisi·e·s au hasard, 18 sont stressé·e·s

- $p_e = 6326$: proportion expérimentale
- p_e devrait être dans $[p - 1{,}96\, s\,;\ p + 1{,}96\, s]$ (avec confiance 95%)
- donc p devrait être dans $[p_e - 1{,}96\, s\,;\ p_e + 1{,}96\, s]$ (avec confiance 95%)
- on approxime $s \simeq 2{,}55$.
- donc $p_e - 1{,}96s \simeq 1{,}3$ et $p_e + 1{,}96s \simeq$
  On lit sur la table inverse que pour la confiance 0,95 1,96 d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0{,}03204$
  On estime donc que μ est dans l'intervalle $[2{,}55 - 0{,}032\,04\,;\ 2{,}55 + 0{,}032\,04] \simeq [2{,}52\,;\ 2{,}58]$, avec la confiance $c = 0{,}95$..
- Avec la confiance 95%, on affirme que la proportion p est dans l'intervalle $[0{,}144\,;\ 0{,}336]$

Remarque : Cet intervalle s'appelle « intervalle de confiance »

D) **Procédure générale** (page 5 du formulaire)

- Échantillon de taille n, au sein d'une population qui contient beaucoup d'individus
- p_e : proportion au sein de l'échantillon
- c niveau de confiance que l'on souhaite

$\rightsquigarrow$ risque d'erreur $\alpha = 1 - c$ (exemple : la confiance $c = 95\%$ signifie un risque d'erreur $\alpha = 5\%$).

Hypothèses à vérifier

$$n \geqslant 30,\ n\,p_e \geqslant 5 \text{ et } n(1-p_e) \geqslant 5.$$

Dans ce cas, le formulaire indique comment calculer un intervalle de confiance :

- La table du formulaire indique un nombre z_α selon le niveau de confiance
- On calcule $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}}$, puis on calcule les nombres $p_e - a_\alpha$ et $p_e + a_\alpha$.
- Avec la confiance $c = 1 - \alpha$, on peut affirmer que le proportion sur l'ensemble de la population se trouve dans l'intervalle $[p_e - a_\alpha, p_e + a_\alpha]$.

E) **Taille de l'échantillon** (page 5 du formulaire)

- Pour avoir une précision h (c'est à dire $a_\alpha < h$), prendre un échantillon de taille $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2}$.
- Si on ne connaît pas p_e, il suffit de prendre $n > z_\alpha^2 \frac{1}{4h^2}$.

5.3 Estimation d'une moyenne

A) Procédure du formulaire (page 5 du formulaire)

- Échantillon de taille n, au sein d'une population qui contient beaucoup d'individus
- m_e et s_e : moyenne et écart type au sein de l'échantillon
- c niveau de confiance que l'on souhaite
- On veut estimer la moyenne μ au sein de la grande population

Premier cas : si $n \geqslant 30$

- La table du formulaire indique un nombre z_α selon le niveau de confiance
- On calcule $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}}$, puis on calcule les nombres $m_e - a_\alpha$ et $m_e + a_\alpha$.
- Avec la confiance $c = 1 - \alpha$, on peut affirmer que μ se trouve dans l'intervalle $[m_e - a_\alpha, m_e + a_\alpha]$.

Deuxième cas : si $n < 30$

- On doit supposer que la variable suit une loi normale
- On lit un seuil t_α dans la table de la loi de student à $n-1$ degrés de liberté : page 11, ligne $n-1$, colonne $p = \alpha/2$.
- On calcule $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}}$, puis on calcule les nombres $m_e - a_\alpha$ et $m_e + a_\alpha$.
- Avec la confiance $c = 1 - \alpha$, on peut affirmer que μ se trouve dans l'intervalle $[m_e - a_\alpha, m_e + a_\alpha]$.

B) Exemples

Données de l'exercice 8

- On a déjà $n = 16$, que $m_e = 28{,}94$ et $s_e = 7{,}09$.
- On utilise la confiance $c = 0{,}95$ et on applique la procédure du cas où $n \geqslant 30$:
  On cherche t_α dans la table page 11 du formuaire, en ligne 15 (car 16 -1 = 15) et colonne 0,025 (car $c = 0{,}95$),
  On lit $t_\alpha \simeq 2{,}1314$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3{,}901\,8$.
  On estime donc que μ est dans l'intervalle $[28{,}94 - 3{,}901\,8\,;\, 28{,}94 + 3{,}901\,8] \simeq [25{,}04\,;\, 32{,}84]$ avec la confiance $c = 0{,}95$

Données de l'exercice 25

- Niveaux d'anxiété avant thérapie : $n = 7{,}09$, que
  $m_e =$
  On lit dans la table du χ^2 à 15 degrés de liberté la v
  6,262 correspondant à $p = 0{,}025$ et la valeur $x_2 =$ 27,49 correspondant à $q = 0{,}025$
  Cela donne $s_e\sqrt{\frac{n}{x_2}} \simeq 7{,}09\sqrt{\frac{16}{27{,}49}} \simeq 5{,}41$
  $ets_e\sqrt{\frac{n}{x_1}} \simeq 7{,}09\sqrt{\frac{16}{6{,}262}} \simeq 11{,}3$
  On estime donc que $l'ecarttype$ est dans l'intervalle [
  0,95. et $s_e = 16$.
- On prend la confiance $c = 0{,}95$, et en supposant qu'on ait une loi normale, on applique la procédure pour $n < 30$:
- 7,09

C) **Taille de l'échantillon** (page 5 du formulaire)

- Pour avoir une précision h (c'est à dire $a_\alpha < h$), prendre un échantillon de taille $n > z_\alpha^2 \frac{(s_e)^2}{h^2}$.

5.4 Estimation d'un écart type

A) Procédure du formulaire (page 6 du formulaire)

- On lit dans la table du χ^2 à $n-1$ degrés de libertés deux nombres x_1 et x_2 :
  - Ligne $n-1$ de la table en page 12 du formulaire.
  - x_1 est à la colonne $q = \frac{1-c}{2}$ et x_2 à la colonne $p = \frac{1-c}{2}$.
- Avec la confiance c, on peut affirmer que l'écart type sur l'ensemble de la population est dans l'intervalle $\left[s_e\sqrt{\frac{n}{x_2}}; s_e\sqrt{\frac{n}{x_1}}\right]$.

B) Exemple (exercice 25)

- $n = ??$, et $s_e = ??$.
- ??

5.5 Que conclure d'un intervalle de confiance ?

- Si la question est par exemple « L'anxiété moyenne est-elle plus grande que 30 ? », ou « La proportion d'étudiants stressés est-elle plus petite que 40% ? »
  - On parvient à conclure uniquement si la valeur de l'énoncer est en dehors de l'intervalle de confiance.
- Si la question est par exemple « L'anxiété moyenne est-elle plus petite après la thérapie qu'avant ? »
  - Si les deux intervalles se chevauchent, on ne peut pas conclure. S'ils sont disjoints, on peut conclure.
  - Dans le cas de variables appariées, on peut calculer leur différence et se demander si la moyenne de la différence est positive ou négative.