RÉPONSES EXERCICES SÉRIE MIXTE : CHAPITRES 2, 3 & 4

Exercice 1 → chapitre 3 estimation d'une moyenne

On cherche $\boldsymbol{\bar{y}}$ $(moyenne\ des\ poids\ observée\ \text{à}\ partir\ des\ 36\ données\ de\ l'\text{é}\boldsymbol{chantillon})\ \boldsymbol{tel\ que}$

$$\boldsymbol{P(\mu > 65|données) \geq\ 0.95}$$

$$dans\ la\ \boldsymbol{population}, la\ moyenne\ des\ poids\ soit\ supérieure\ \text{à}\ 65kg$$

$avec\ une\ probabilité\ supérieure\ ou\ égale$ à 0.95 $(seuil\ critique\ de\ significativité)$.

n = 36 (n > 20→ approximation normale), s=0.4

- A posteriori pour μ : $(\mu \Big| données)\ \approx N\ (\bar{y}\ ; \frac{s^2}{n} = \frac{0.4^2}{36} = 0.004444)$
- On sait que : $P(\mu > 65|données)\ doit\ \text{ê}tre \geq\ 0.95$

$\Leftrightarrow P(Z > \frac{65-\bar{y}}{\sqrt{0.004444}})$ = 0.95

Table de la normale centrée réduite : $\frac{65-\bar{y}}{\sqrt{0.004444}}$ = - 1.64 (attention signe !!!)

$\bar{y} = 65 + 1.64\ \times \sqrt{0.004444}$ = 65.1093

Exercice 2 → chapitre 3 ANOVA 1

- Hyp : Y = nombre d'heures de sport pratiquées par semaine par des étudiants universitaires $Y \sim N(\mu_k;\ \sigma^2 constante)$
- K = 3 (étudiants classés en 3 groupes en fonction du nombre de fruits qu'ils mangent par jour) ; n = 28
- Évaluer la plausibilité de l'hypothèse $\mu_1\ =\ \mu_2\ =\ \mu_3$:

$$F_{obs} = \frac{\sum_{k=1}^{K} \frac{n_k(\bar{y}_k - \bar{y})^2}{K-1}}{s_p^2} \quad avec\ s_p^2 = \frac{(n_1 - 1)s_1^2 + \cdots + (n_k - 1)s_k^2}{n - K}$$

- $s_p^2 = 3.3751$; $\bar{y} = 2.68$
- $F_{obs} = 2.158 <\ F_{2,25}(0.95) = 3.39$ → hypothèse $\mu_1\ =\ \mu_2\ =\ \mu_3$ plausible

Avec les données à disposition, on ne peut affirmer que le nombre d'heures de sport pratiquées par semaine et le nombre de fruits mangés par jour sont liés dans la population étudiée.

Exercice 3 → chapitre 2 estimation d'une proportion et comparaison de 2 proportions (échantillons indépendants)

a. Estimation d'une proportion avec a priori non informatif
   → A posteriori pour $\pi = (\pi|T = 24) \sim Beta\ (25; 46) \approx N\ (0.3478; 0.0032877)$
   - $P(\pi \geq 0.5|T = 24)\ \ = P\ (Z \geq 2.65)$ = 1 – $P\ (Z \leq 2.65) = 1 - 0.996 = 0.004$

Il y a seulement 0.4% de chance qu'au moins un étudiant sur 2 en Sciences Humaines croit en une forme de vie après la mort. Autrement dit, la probabilité qu'il y en ait moins d'un sur deux qui y croit est suffisamment grande pour conserver plutôt cette hypothèse.

- IC (95%) pour π : [0.2354 ; 0.4602] $\nexists\, 0.5\; et\; V < 0.5$ → même conclusion avec pour information supplémentaire que 95% des V plausibles pour la vraie proportion (population) d'étudiants qui croient en une forme de vie après la mort sont situées entre 23.5 et 46%. (La marge d'erreur est importante).

b. Comparaison de 2 proportions (toujours avec *a priori* non informatifs). Si :
   - $\pi_1 =$ $proportion\ d'\acute{e}tudiants\ qui\ croient\ en\ une\ forme\ de\ vie\ apr\grave{e}s\ la\ mort\ en\ SA$
   - $\pi_2 =$ $proportion\ d'\acute{e}tudiants\ qui\ croient\ en\ une\ forme\ de\ vie\ apr\grave{e}s\ la\ mort\ en\ SHS$

   A posteriori pour $\pi_1 - \pi_2$: $(\pi_1 - \pi_2 | donn\acute{e}es) \approx N\,(0.425; 0.0133212)$
   IC (95%) pour $\pi_1 - \pi_2$: [0.199 ; 0.651] $\nexists\, 0\; et\; V > 0$ → la proportion d'étudiants qui croient en une forme de vie après la mort est plus importante chez les SA que chez les SHS mais l'estimation de cette différence est relativement imprécise puisque les valeurs plausibles s'étalent entre 20 et 65% (avec encore 5% de risque de se tromper).

c. Estimation d'une proportion avec a priori informatif :
   - A priori : $\pi \sim Beta\,(30, 32)$
   - A posteriori : $(\pi | T = 24) \sim Beta\,(54, 77) \approx N\,(0.411; 0.001876)$
   - $P(\pi \geq 0.5 | T = 24) = 0.0202$ → on reste en dessous du seuil de 5% de chance de se tromper en soutenant l'hypothèse selon laquelle moins d'un étudiant sur 2 croit en une forme de vie après la mort.
   - IC (95%) pour π : [0.326 ; 0.496] $\nexists\, 0.5\; et\; V < 0.5$ → même conclusion mais l'on observe qu'avec un a priori informatif la précision de l'estimation augmente.

Exercice 4 → chapitre 3 estimation d'une moyenne et comparaison de 2 moyennes (données pairées)

n = 24 > 20 → approximation normale

a. $(\mu_{SMS\ re\c{c}us} \big| donn\acute{e}es) \approx N\,(\bar{y} = 8.583\,; \frac{s^2}{n} = \frac{40.955}{24} = 1.7065)$

   $(\mu_{SMS\ envoy\acute{e}s} \Big| donn\acute{e}es) \approx N\,(\bar{y} = 8.542\,; \frac{s^2}{n} = \frac{37.8108}{24} = 1.5758)$

b. $P(\mu_{SMS\ envoy\acute{e}s} > 5 | donn\acute{e}es)$ = 0.9976 → oui, il est très probable (P > 0.95) que l'abonnement soit utile en moyenne aux étudiants.

c. IC (95%) pour $\mu_{SMS\ envoy\acute{e}s}$: [6 ; 11] $\nexists\, 5\; et\; V > 5$ → La « vraie » moyenne de SMS envoyés dans la population de ces étudiants est située entre 6 et 11 SMS par jour (en gardant 5% de chances de se tromper). Même conclusion qu'en b.

d. Définition d'une nouvelle variable D = SMS reçus – SMS envoyés

Étudiant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
D	0	5	-5	0	0	0	0	0	0	0	0	0	5	0	0	0	2	0	0	0	0	-4	-2	0

- A posteriori pour μ_D: $(\mu_D \left| données\right) \approx N\left(\bar{d} = 0.04167 ; \frac{s^2{}_D}{n_D} = \frac{4.3025}{24} = 0.1793\right)$
- $P(\mu_D > 0|données) = 0.5398$; IC (95%) pour μ_D : [-0.788 ; 0.872]→ on ne peut pas se prononcer ; il n'apparaît pas qu'il y ait une différence significative entre la moyenne du nombre de SMS reçus et envoyés par ces étudiants.

Exercice 5 → <u>chapitre 4 comparaison de 2 proportions (données pairées)</u> Si :

- $\pi_1 = proportion\ d'Américains\ prêts$ à $payer\ plus\ d'impôts$
- $\pi_2 = proportion\ d'Américains\ prêts$ à $diminuer\ leur\ niveau\ de\ vie$

A posteriori pour $\pi_2 - \pi_1$: $(\pi_2 - \pi_1|données) \approx N\,(-0.02185; 0.0001822)$

Le test de McNemar peut nous dire s'il existe une différence significative entre ces deux proportions mais, si tel est le cas, il ne permet pas de déterminer si $\pi_2 > \pi_1$ ou inversement.

McNemar : 2.615 < $X_1^2(0.95) = 3.84$ → 0 fait partie des valeurs plausibles pour la différence de proportions → on ne peut pas affirmer qu'il y ait une différence significative. Ici en l'occurrence, ce type de test est suffisant pour répondre à la question.

Le test d'hypothèse ou l'intervalle de crédibilité apportent par contre directement les 2 informations qui nous intéressent pour répondre à la question : le caractère significatif ou non de la différence et, le cas échéant, le sens de la différence ($\pi_2 > \pi_1$ ou inversement).

P $(\pi_2 - \pi_1 < 0|données) = 0.0526$

IC (95%) pour $\pi_2 - \pi_1$: [-0.0483 ; 0.0046] → 0 fait partie des valeurs plausibles donc il n'y a pas de différence significative entre la proportion d'Américains qui sont prêts à diminuer leur niveau de vie et ceux qui sont prêts à payer plus d'impôts. Cela n'a donc pas de sens de chercher à déterminer le sens de la différence puisqu'il est plausible qu'il n'y en ait pas au niveau de la population.

Exercice 7 → <u>chapitre 3 ANOVA1</u>

- Hyp : Y = somme d'argent de poche reçue par semaine par les étudiants de l'ULg
  $Y \sim N(\mu_k;\ \sigma^2 constante)$
- K = 3 (étudiants classés en trois groupes en fonction de la fréquence des sorties/semaine)
- n = 28, $n_1 = 6;\ \bar{y}_1 = 11.3;\ s_1 = 7.9$; $n_2 = 15;\ \bar{y}_2 = 23.9;\ s_2 = 15.5$; $n_3 = 7;\ \bar{y}_3 = 86.4;\ s_3 = 101.8$

  Complétez la table :

- ddl (Nombre de sorties) = K-1 = 2
- ddl (erreur) = n-K = 25
- MC (ord. Orig) = SC (ord orig)/ddl (ord orig) = 39367.64
- SC (nombre de sorties) = MC (nombre de sortie) x ddl (nombre de sorties) = 23615.44
- MC (erreur) = SC (erreur)/ddl(erreur) = 2635.12
- F (ord orig) = MC (ord. Orig)/MC (erreur) = 14.94

- F (nombre de sortie) = F_{obs} = MC (nombre de sorties) / MC (erreur) = 4.481

$F_{obs} =$ 4.481 à $> F_{2,25}(0.95) = 3.39$ → rejet de l'hypothèse $\mu_1 = \mu_2 = \mu_3$: l'argent de poche et le nombre de sorties sont bien significativement liés (mais il se peut que tous les contrastes de moyennes 2 à 2 ne soient pas significatifs).

Exercice 7 → chapitre 3 comparaison de 2 moyennes (échantillons indépendants)

$n_{NF} = 14;\ n_F = 8\ \ < 20$ → distribution de student

a. $(\mu_{NF}\,\Big|\,données) \sim t_{13}\ (\bar{y}_{NF} = 22.657\,; \frac{s^2}{n_{NF}} = \frac{7.142}{14} = 0.51)$
   $(\mu_F\,\Big|\,données) \sim t_7\ (\bar{y}_F = 21.325\,; \frac{s^2}{n_F} = \frac{7.0393}{8} = 0.88)$
b. Si $\theta = \mu_{NF} - \mu_F$: $(\theta\,\big|\,données) \sim t_{20}\ (1.332\,; 1.39)$
- P $(\theta > 0\,\big|\,données) = [0.75; 0.90]$: probabilité trop faible que pour se prononcer.
- IC (95%) pour θ : [-1.13 ; 3.79] → il n'y a pas de différence significative entre le BMI moyen des fumeurs et des non-fumeurs. Les chercheurs ne peuvent pas valider leur hypothèse.

Exercice 8 → chapitre 2 estimation d'une proportion et comparaison de 2 proportions (échantillons indépendants)

a. Les a priori sont non informatifs → A posteriori pour $\pi_A - \pi_B$:
   $(\pi_A - \pi_B | T_A = 43;\ T_B = 6)\ \approx N\ (0.4767; 0.01068)$
   - ➔ P $(\pi_A - \pi_B > 0) = P\ (Z \leq 4.61)\ \approx 1$ → On peut affirmer avec une chance infime de se tromper que les étudiants américains sont proportionnellement plus nombreux à soutenir une action caritative que les étudiants belges.
   - ➔ IC (95%) pour $\pi_A - \pi_B$: $[0.274\ ;\ 0.679] \not\ni 0\ et\ V > 0$ → $\pi_A > \pi_B$ (même conclusion mais estimation de la différence relativement imprécise vu la « largeur » de l'IC)
b. A priori informatif pour π_A :
   - A priori : $\pi \sim Beta\ (44, 18)$
   - A posteriori : $(\pi_{2010} | T = 32) \sim Beta\ (76; 57) \approx N\ (0.5725; 0.0018683)$
   - IC (95%) pour π_{2010} : [0.488 ; 0.657] → la « vraie » proportion d'étudiants américains soutenant une action caritative dans la population américaine est située entre 48.8% et 65.7% (en gardant 5% de chance de se tromper).

Exercice 9 → chapitre 4 comparaison de 2 proportions (données pairées). Si

- $\pi_1 = proportion\ d'Américains\ croyant\ au\ paradis$
- $\pi_2 = proportion\ d'Américains\ croyant\ à\ l'enfer$

A posteriori pour $\pi_2 - \pi_1$: $(\pi_2 - \pi_1\,\Big|\,données)\ \approx N\ (-0.1122; \frac{0.10317}{1123} = 9.187\ \times 10^{-5})$

- McNemar : 122.123 > $X_1^2(0.95) = 3.84$ → on peut affirmer qu'il existe une différence significative entre ces proportions mais on ne peut pas déterminer sur base de ce seul test si $\pi_2 > \pi_1$ ou inversement.

- $P(\pi_2 - \pi_1 < 0|données) = P(Z \leq 11.71) \approx 1$ → on peut dire qu'il existe une différence significative et qu'elle va dans le sens $\pi_2 < \pi_1$: la proportion d'Américains qui croient en l'enfer est inférieure à celle des Américains qui croient au paradis.
- IC (95%) pour $\pi_2 - \pi_1$: [-0.131 ; -0.093] $\nexists\, 0\ et\ V < 0$ même conclusion (avec information supplémentaire : il y a environ entre 9 et 13% d'Américains en moins qui croient en l'enfer par rapport à ceux qui croient au paradis).