EXERCICES SUR LES CHAPITRES 2, 3 et 4 : « MIXTE »

1. On désire comparer le poids moyen des individus d'une population à un poids de 65 kg. Sachant que l'écart-type des poids d'un échantillon d'effectif 36 de cette population est de 0,4 kg, déterminer les valeurs de la moyenne des poids observée sur cet échantillon qui conduiraient à un poids moyen significativement supérieur à 65 kg.

2. Lors d'une étude portant sur la santé générale des étudiants universitaires, il a été demandé à 28 étudiants combien d'heures de sport ils pratiquaient par semaine et combien de fruits ils mangeaient habituellement par jour. Voici leurs réponses:

Nombre de fruits/jour	Effectif	Nombre heures sport/semaine ($\bar{y}$ ± s)
0-1	15	2,40 ± 1,16
2	5	2,80 ± 1,64
3 ou plus	6	4,17 ± 3,31

   Pensez-vous, qu'en moyenne, le nombre d'heures de sport pratiquées par semaine et le nombre de fruits mangés par jour sont liés dans la population étudiée?

3. Des chercheurs ont décidé d'analyser les étudiants de la faculté des Sciences Humaines. Un échantillon aléatoire simple de 69 étudiants a été interrogé. Parmi ceux-ci, 24 étudiants ont affirmé croire dans une forme de vie après la mort.
   a. Peut-on affirmer qu'au moins un étudiant sur deux croit dans une forme de vie après la mort en Sciences Humaines? (utiliser deux méthodes différentes)
   b. Deux orientations sont proposées aux étudiants de Sciences Humaines : SA et SHS. Voici la distribution des réponses en fonction de l'orientation.

	Vie après la mort		
Section	Oui	Non	Total
SA	15	9	24
SHS	9	36	45

   Comparez la proportion d'étudiants qui croient en une forme de vie après la mort dans les deux sections et donnez un ensemble de valeurs plausibles pour la différence de ces proportions dans la population correspondante.
   c. Après avoir présenté leurs résultats, les enquêteurs ont appris que la même enquête a été réalisée aux Etats-Unis auprès d'étudiants en sociologie. Un échantillon aléatoire simple de 60 étudiants de sociologie a été interrogé aux Etats-Unis. Un total de 29 étudiants a déclaré croire dans une forme de vie après la mort. Répondez à la question a) à la lumière de ces nouvelles informations.

4. Un opérateur de GSM souhaite créer un nouvel abonnement réservé aux étudiants. Le nouvel abonnement est intéressant pour les personnes qui envoient en moyenne plus de 5 SMS par jour. Avant de lancer son nouvel abonnement, il a décidé d'étudier les habitudes téléphoniques

des étudiants et a interrogé 24 étudiants de la faculté des Sciences Humaines sélectionnés au hasard. Voici le nombre de SMS envoyés et reçus par jour pour chacun d'eux.

Etudiant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
SMS reçus	0	20	10	25	5	3	3	5	5	20	15	2	10	10	8	10	12	10	10	1	10	6	3	3
SMS envoyés	0	15	15	25	5	3	3	5	5	20	15	2	5	10	8	10	10	10	10	1	10	10	5	3

a. Déterminez la densité a posteriori du nombre moyen de SMS reçus et envoyés.
b. Pensez-vous utile que l'opérateur GSM propose son abonnement aux étudiants ?
c. Déterminez l'ensemble des valeurs plausibles pour le nombre moyen de SMS envoyés dans la population correspondante.
d. Les étudiants de Sciences Humaines reçoivent-il en moyenne plus de SMS qu'ils en envoient?

5. Durant l'étude « General Social Survey » menée en 2002 aux Etats-Unis, on a demandé à un échantillon aléatoire de 1144 Américains si, pour aider l'environnement, ils étaient prêts à payer des taxes plus élevées et à accepter une réduction de leur niveau de vie. Les Américains préfèrent-ils, à cette fin, diminuer leur niveau de vie ou payer plus d'impôts ?

	Diminuer leur niveau de vie		
Payer plus d'impôt	Oui	Non	Total
Oui	227	132	359
Non	107	678	785
Total	334	810	1144

6. Dans le cadre d'une étude sur la situation des étudiants de l'Université de Liège, des chercheurs ont voulu savoir si la fréquence des sorties et la somme d'argent de poche reçue par semaine par les étudiants étaient liés.
Les étudiants ont été séparés en 3 groupes : ceux qui ne sortent pas la semaine, ceux qui sortent une fois par semaine et ceux qui sortent plus d'une fois par semaine.
Sur les 28 étudiants interrogés, les 6 étudiants ne sortant pas reçoivent (moyenne ± écart-type) en moyenne 11,3 ± 7,9€, les 15 étudiants sortant une fois 23,9 ± 15,5€ et les 7 étudiants sortant plus d'une fois 86,4 ± 101,8€.
Complétez la table et vérifiez si l'argent de poche et le nombre de sorties sont liés.

Effet	SC	ddl	MC	F
Ord. origine	39367,64	1		
Nombre de sorties			11807,72	
Erreur	65877,98			

7. Des chercheurs se sont intéressés à l'effet du tabac sur le poids de jeunes adultes, avec l'hypothèse que le tabac a un effet « coupe-faim » et que les fumeurs devraient donc présenter un BMI (kg/m²) moyen plus petit que les non-fumeurs. Pour cela, ils ont interrogé 22 étudiants en Sciences Humaines sélectionnés au hasard. Voici leurs réponses :

	BMI (kg/m²)													
Non-fumeurs	27.0	27.3	23.0	23.3	20.6	24.4	21.0	20.3	19.8	23.2	23.6	24.2	17.8	21.7
Fumeurs	20.7	20.1	27.4	22.0	21.7	19.0	19.6	20.1						

a. Déterminez la distribution a posteriori du BMI moyen dans la population des étudiants fumeurs et non-fumeurs.
b. Quelles sont les conclusions des chercheurs ? (utiliser deux méthodes).

8. Lors d'une étude sur les étudiants en sociologie, il a été demandé à un échantillon aléatoire de 25 étudiants belges et de 60 étudiants américains s'ils soutenaient activement une action caritative. C'était le cas pour 6 étudiants sur 25 en Belgique et 43 étudiants sur 60 aux Etats-Unis.
   a. Les étudiants américains soutiennent-ils une action caritative plus fréquemment que les étudiants belges?
   b. L'étude a été réitérée en 2010 aux Etats-Unis, cette année-là, 32 étudiants sur 71 soutenaient une action caritative. Déterminez un ensemble de valeurs plausibles pour la proportion d'étudiants soutenant une action caritative dans la population correspondante.

9. Lors d'une étude aux Etats-Unis, les chercheurs ont demandé aux participants s'ils croyaient au paradis et à l'enfer. Voici leurs réponses :

	Croire en l'enfer	
Croire au paradis	Oui	Non
Oui	833	128
Non	2	160

Comparez la proportion de personnes croyant en l'enfer et au paradis de plusieurs manières.

RÉPONSES EXERCICES SÉRIE MIXTE : CHAPITRES 2, 3 & 4

Exercice 1 → chapitre 3 estimation d'une moyenne

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

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

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

$avec\ une\ probabilité\ supérieure\ ou\ égale$ 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 : $P(\mu > 65|données)\ doit\ être \geq\ 0.95$

$\Leftrightarrow P(Z > \frac{65-\bar{y}}{\sqrt{0.004444}})$ = 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 = 26
- Évaluer la plausibilité de l'hypothèse $\mu_1\ =\ \mu_2\ =\ \mu_3$:

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

- $s_p^2 = 3.6686$; $\bar{y} = 2.8854$
- $F_{obs} = 1.836 <\ F_{2,25}(0.95) = 3.39$ → 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 π : $(\pi|T = 24) \sim Beta\ (25; 46) \approx N\ (0.3478; 0.0032877)$
   - $P(\pi \geq 0.5|T = 24)\ = P\ (Z \geq 2.65)$ = 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] $\not\ni 0.5\ et\ V < 0.5$ → 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 :
   - $\pi_1 =$ $proportion\ d'étudiants\ qui\ croient\ en\ une\ forme\ de\ vie\ après\ la\ mort\ en\ SA$
   - $\pi_2 =$ $proportion\ d'étudiants\ qui\ croient\ en\ une\ forme\ de\ vie\ après\ la\ mort\ en\ SHS$

   A posteriori pour $\pi_1 - \pi_2$: $(\pi_1 - \pi_2 | données) \approx N\ (0.425; 0.0133212)$
   IC (95%) pour $\pi_1 - \pi_2$: [0.199 ; 0.651] $\not\ni 0\ et\ V > 0$ → 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] $\not\ni 0.5\ et\ V < 0.5$ → même conclusion mais l'on observe qu'avec un a priori informatif la précision de l'estimation augmente.

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

n = 24 > 20 → approximation normale

a. $(\mu_{SMS\ reçus} \Big| données) \approx N\ (\bar{y} = 8.583 ; \frac{s^2}{n} = \frac{40.955}{24} = 1.7065)$
   $(\mu_{SMS\ envoyés} \Big| données) \approx N\ (\bar{y} = 8.542 ; \frac{s^2}{n} = \frac{37.8108}{24} = 1.5758)$
b. $P(\mu_{SMS\ envoyés} > 5 | données)$ = 0.9976 → oui, il est très probable (P > 0.95) que l'abonnement soit utile en moyenne aux étudiants.
c. IC (95%) pour $\mu_{SMS\ envoyés}$: [6 ; 11] $\not\ni\ 5\ et\ V > 5$ → 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: $(\mu_D \left| données\right) \approx N\ (\bar{d} = 0.04167\ ; \frac{s^2{}_D}{n_D} = \frac{4.3025}{24} = 0.1793)$
- $P(\mu_D > 0|données) = 0.5398$; 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 → chapitre 4 comparaison de 2 proportions (données pairées) 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 6 → chapitre 3 ANOVA1

- Hyp : Y = somme d'argent de poche reçue par semaine par les étudiants de l'ULg
  $Y_k \sim N(\mu_k;\ \sigma^2 constante)$
- 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 $> F_{2,25}(0.95) = 3.39 \rightarrow$ 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 → <u>chapitre 3 comparaison de 2 moyennes (échantillons indépendants)</u>

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

a. $(\mu_{NF} \left| données) \sim t_{13}\ (\bar{y}_{NF} = 22.657\ ; \frac{s^2}{n_{NF}} = \frac{7.142}{14} = 0.51\right)$

   $(\mu_F \left| données) \sim t_7\ (\bar{y}_F = 21.325\ ; \frac{s^2}{n_F} = \frac{7.0393}{8} = 0.88\right)$

b. Si $\theta = \mu_{NF} - \mu_F$: $(\theta \middle| données) \sim t_{20}\ (1.332\ ; 1.39)$

- P $(\theta > 0 | données) = [0.75; 0.90]$: 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 → <u>chapitre 2 estimation d'une proportion et comparaison de 2 proportions (échantillons indépendants)</u>

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 $\pi_A - \pi_B$: $[0.274\ ;\ 0.679]\ \nexists\ 0\ et\ V > 0$ → $\pi_A > \pi_B$ (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 → <u>chapitre 4 comparaison de 2 proportions (données pairées)</u>. Si

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

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

- 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).