Contrôle Terminal, 2$^{\text{ème}}$ session – 17/06/2025 – 2h00
Statistiques – L1 de psychologie

Réponses

Veuillez rendre ce sujet et votre copie.

Numéro d'anonymat :

Le formulaire et la calculatrice sont autorisés. Merci d'indiquer dans la case Numéro d'anonymat*, ci-dessus, un numéro que vous reporterez aussi sur votre copie. Vous rendrez l'énoncé et votre copie, et pouvez soit répondre sur l'énoncé, soit détailler certaines questions sur la copie si vous avez besoin de plus de place. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.*

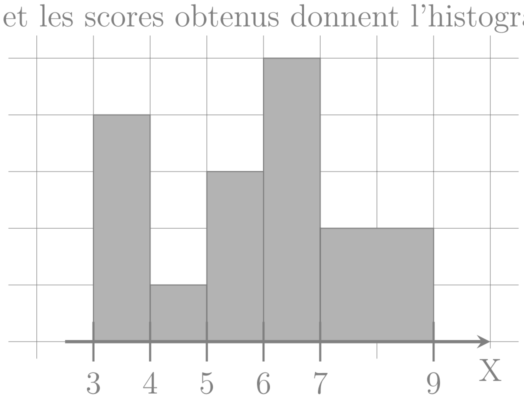
Exercice 1 : Regrets, en fonction du mode de décision

On étudie les regrets exprimés par les volontaires ayant participé à une expérience. Au cours de l'expérience, ces volontaires ont eu le choix entre deux jeux d'argent qui ne donnaient pas les mêmes chances de gain, ont choisi un jeu avec lequel ils ont perdu (ou réalisé un gain très limité) alors que l'autre jeu d'argent leur aurait rapporté plus (mais ils/elles ne le savaient pas au moment de choisir entre les deux jeux d'argent).

Certain·e·s parmi les volontaires ont choisi le jeux d'argent après avoir reçu l'avis d'une personne extérieur qui leur a conseillé un jeu plutôt que l'autre.

On s'intéresse tout particulièrement aux participant·e·s qui avaient initialement choisi le jeu rapportant le plus, puis ont changé d'avis sur conseil de la personne extérieur et ont donc perdu (ou gagné très peu). On cherche à savoir si les regrets qu'ils/elles expriment pour avoir fait un mauvais choix sont réduits parce qu'ils/elles ont simplement suivi l'avis d'une autre personne (ce qui devrait diluer le sentiment de culpabilité).

1. On met en place une procédure qui évalue les regrets exprimés par les participant·e·s. On commence par la tester sur 17 personnes, et les scores obtenus donnent l'histogramme ci-dessous :



En déduire les fréquences des différentes classes *(vous remplirez le tableau ci-dessous)*

Pour calculer ces fréquences, on a compté la surface (le nombre de cases) de chaque rectangle (4, 1, 3, 5 et 4) et la surface totale ($4 + 1 + 3 + 5 + 4 = 17$). La fréquence de chaque classe est alors la surface du rectangle correspondant, divisée par la surface totale.

Score	[3 ; 4[	[4 ; 5[	[5 ; 6[	[6 ; 7[	[7 ; 9[
fréquence	0,235	0,059	0,176	0,294	0,235

2. On considère dans cette question les participant·e·s qui n'ont pas eu d'avis extérieur, qui ont choisi par elles/eux même le jeu qui rapportait le moins et qui ont effectivement perdu (ou gagné peu). Ces participant·e·s forment le « groupe témoin », et leurs scores sont ci-dessous :

Score	[1 ; 2[	[2 ; 3[	[3 ; 4[	[4 ; 5[	[5 ; 6[	[6 ; 7[
Effectif	8	13	10	15	7	2
Fréquences	0,145	0,236	0,182	0,273	0,127	0,036
Fréq, cum,	0,145	0,382	0,564	0,836	0,964	1,0

   (a) Quelle est au sein de ce groupe témoin la proportion d'individus dont le score est plus petit que 4 ?

   Au sein de ce groupe témoin, il y a $8+13+10=31$ participant·e·s dont le score est inférieur à 4. Cela correspond à une proportion $\frac{31}{55} \simeq 0{,}563\,6$.

   (b) Quels sont la moyenne et l'écart type de ces scores ? *Dans cette question, on vous demande d'indiquer les calculs effectués.*

   moyenne : $m(X) = \frac{\sum c_i\, n_i}{n} = \frac{1{,}5\times 8+2{,}5\times 13+3{,}5\times 10+\cdots+6{,}5\times 2}{55} = \frac{198{,}5}{55} \simeq 3{,}61$
   $m(X^2) = \frac{\sum c_i^2\, n_i}{n} = \frac{1{,}5^2\times 8+2{,}5^2\times 13+3{,}5^2\times 10+\cdots+6{,}5^2\times 2}{55} = \frac{821{,}75}{55}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{821{,}75}{55} - \left(\frac{198{,}5}{55}\right)^2 \simeq 1{,}915$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}38$

   (c) Quelle est la médiane ?

   On commence par calculer les fréquences et fréquences cumulées
   (on les a insérées dans le tableau en début d'exercice).
   Classe de la médiane : [3 ; 4[
   Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 3 + \frac{4-3}{0{,}564-0{,}382}(0{,}5-0{,}382)$
   $\simeq 3{,}648$

   (d) À partir de cet échantillon, estimer le score moyen des personnes à qui on ferait passer l'expérience dans les même conditions que ce groupe témoin. *On vous demande un intervalle de confiance, correspondant à la confiance 99%.*

   Comme n= 55 >30, le formulaire indique que pour la confiance 0,99 on a $z_\alpha \simeq 2{,}576$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}576 \times \frac{1{,}38}{\sqrt{55-1}} \simeq 0{,}4838$
   On estime donc que la moyenne de ce score est dans l'intervalle $[3{,}61 - 0{,}483\,8\,;\, 3{,}61 + 0{,}483\,8] \simeq [3{,}13\,;\, 4{,}09]$, avec la confiance $c = 0{,}99$.

3. On considère dans cette question les participant·e·s qui avaient choisi le jeu le plus profitable, ont reçu un avis extérieur qui les a fait changé d'avis et ont perdu (ou gagné très peu) suite à ce changement d'avis. Ces participant·e·s forment le « groupe cible », et leurs scores sont ci-après (en haut de la page suivante) : *On pourra supposer qu'ils sont issus d'une loi normale*

6,3	5,1	7,8	5,4	7,9	5,7	6,4	7,3	3,6	5,4	5	5,1	6,5

(a) Quels sont la moyenne et l'écart type de ces scores ?

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{6{,}3+5{,}1+7{,}8+\cdots+6{,}5}{13} = \frac{77{,}5}{13} \simeq 5{,}96$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{6{,}3^2+5{,}1^2+7{,}8^2+\cdots+6{,}5^2}{13} = \frac{480{,}23}{13}$

$Var(X) = m(X^2) - m(X)^2 = \frac{480{,}23}{13} - \left(\frac{77{,}5}{13}\right)^2 \simeq 1{,}4$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}18$

(b) À partir de cet échantillon, estimer le score moyen des personnes à qui on ferait passer l'expérience dans les même conditions que ce groupe cible. *On vous demande un intervalle de confiance, correspondant à la confiance 99%.*

Comme $n = 13 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}005$ et $n-1 = 12$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 3{,}0545$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3{,}0545 \times \frac{1{,}18}{\sqrt{13-1}} \simeq 1{,}040\,5$.

On estime donc que la moyenne de ce score est dans l'intervalle $[5{,}96 - 1{,}040\,5\,;\, 5{,}96 + 1{,}040\,5] \simeq [4{,}92\,;\, 7]$ avec la confiance $c = 0{,}99$

(c) De même (toujours en vous appuyant sur cet échantillon), estimer l'écart-type du score des personnes à qui on ferait passer l'expérience dans les même conditions que ce groupe cible. *On vous demande un intervalle de confiance, correspondant à la confiance 99%.*

On lit dans la table du χ^2 à 12 degrés de liberté la valeur $x_1 = 3{,}074$ correspondant à $q = 0{,}005$ et la valeur $x_2 = 28{,}3$ correspondant à $p = 0{,}005$

On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 1{,}18\sqrt{\frac{13}{28{,}3}} \simeq 0{,}8$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 1{,}18\sqrt{\frac{13}{3{,}074}} \simeq 2{,}43$

On estime donc que σ est dans l'intervalle $[0{,}8; 2{,}43]$ avec la confiance $c = 0{,}99$.

4. Pour chacune de ces affirmations, peut-on conclure, avec la confiance 99% qu'elle soit vraie, ou fausse ? *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse à partir des intervalles des questions précédentes.*

Dans les conditions du « groupe cible », l'écart type est plus grand que 1.

☐ Vraie ☐ Fausse ☒ On ne peut rien affirmer

En effet, la valeur 1 est à l'intérieur de l'intervalle de la question 3c, cet intervalle ne dit pas si l'écart type est plus grand ou plus petit que 1.

Dans les conditions du « groupe cible », le score moyen est plus faible que dans les conditions du « groupe témoin »

☐ Vraie ☒ Fausse ☐ On ne peut rien affirmer

En effet on a vu que la moyenne dans les condition du groupe cible est dans l'intervalle $[4{,}92; 7]$, donc plus élevé que dans les conditions du groupe témoin (où elle est dans $[3{,}13; 4{,}09]$). On peut conclure car les deux intervalles ne se chevauchent pas.

Exercice 2 : Réchauffement planétaire

On s'intéresse dans cet exercice à la température moyenne sur terre : par convention on considère l'écart entre les températures moyennes d'une période fixée (par exemple une année) et les températures moyennes de la période 1850-1900. Pour les dernières années, le ministère français du développement durable indique les valeurs ci-dessous : *On a laisseé des lignes vides au cas où vous en ayiez besoin pour vos calculs*

Année (X)	13	14	15	16	17	18	19	20	21	22	23
Écart (Y)	0,98	1,03	1,18	1,29	1,2	1,12	1,25	1,28	1,12	1,16	1,43
x'_i	1	2	3	4	5	6	7	8	9	10	11
y'_i	1	2	6	10	7	3,5	8	9	3,5	5	11
$(x'_i - y'_i)^2$	0	0	9	36	4	6,25	1	1	30,25	25	0

Par exemple, la troisième colonne de ce tableau indique qu'en moyenne sur l'année 2015, la température moyenne sur terre était 1,18°C de plus que la moyenne sur la période 1850-1900.

Dans l'analyse de ces données, on désignera par X l'année, et par Y cet écart de température (que l'on appelle "anomalie de température") exprimé en degrés celsius (°C).

À titre de comparaison, les accords de Paris ont convenu de faire des efforts pour que cet écart ne dépasse pas durablement 1,5°C, et de rester bien en deça de 2°C.

1. Quel est pour ces données le coefficient de corrélation des rangs (de Spearman) entre l'année X et l'écart de température Y ? Que peut-on en conclure ? *(présenter les calculs effectués)*

On calcule tout d'abord les rangs, que l'on a ajouté dans le tableau au début de l'exercice.

Le coefficient de corrélation des rangs vaut donc $1 - \left(6 \times \frac{0+0+9+36+4+\cdots+0}{11\left(11^2-1\right)}\right) \simeq 0{,}488\,6$

Ce coefficient n'est pas très proche de 1, donc il n'y a pas un lien très fort entre l'année et la température. Mais comme il est positif (et pas trop proche de 0), il indique que plus l'année augmente, plus cet écart de température a tendance à être élevé.

2. De même, quel est le coefficient de corrélation linéaire entre l'année X et le réchauffement Y ? Que peut-on en conclure ? *Dans cette question on demande de présenter les calculs effectués.*

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{13+14+15+\cdots+23}{11} = \frac{198}{11} = 18$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{13^2+14^2+15^2+\cdots+23^2}{11} = \frac{3\,674}{11}$
$Var(X) = m(X^2) - m(X)^2 = \frac{3\,674}{11} - \left(\frac{198}{11}\right)^2 = 10$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 3{,}16$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{0{,}98+1{,}03+1{,}18+\cdots+1{,}43}{11} = \frac{13{,}04}{11} \simeq 1{,}19$
$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{0{,}98^2+1{,}03^2+1{,}18^2+\cdots+1{,}43^2}{11} = \frac{15{,}618}{11}$
$Var(Y) = m(Y^2) - m(Y)^2 = \frac{15{,}618}{11} - \left(\frac{13{,}04}{11}\right)^2 \simeq 0{,}014\,5$
Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 0{,}12$

$m(X\,Y) = \frac{\sum x_i\, y_i}{n} = \frac{13\times 0{,}98+14\times 1{,}03+\ldots+23\times 1{,}43}{11} = \frac{237{,}34}{11} \simeq 21{,}576$
$Cov(X{,}Y) = m(X\,Y) - m(X)\,m(Y) = \frac{237{,}34}{11} - \frac{198}{11}\frac{13{,}04}{11} \simeq 0{,}238\,2$
$r(X{,}Y) = \frac{Cov(X{,}Y)}{s(X)\,s(Y)} = \frac{0{,}238\,2}{3{,}16\times 0{,}12} \simeq 0{,}628\,2$

Ce coefficient n'est pas très proche de 1 donc les variables sont peu liées ou le lien n'est pas très linéaire. Comme il est positif (et pas trop proche de 0), il indique que plus l'année augmente, plus cette "anomalie de température" a tendance à être élevé.
Remarque : (ce n'était pas la réponse demandée mais) le consensus scientifique est que les températures augmentent globalement. Ce consensus découle de données sur plusieurs décennies et est beaucoup plus solide que ce que l'on pourrait conclure à partir des 11 années données dans cet énoncé.

3. Si on utilisait une droite de régression pour estimer le réchauffement futur (ce qui revient à supposer que les températures évoluent en suivant une droite et qu'elles continueront à évoluer selon la même droite que sur cette période 2013-2023), alors à quel réchauffement s'attendrait-on en 2050 ?

Il s'agit de calculer la droite $D_{Y|X}$:
On pose $a = \frac{Cov(X{,}Y)}{Var(X)} \simeq \frac{0{,}238\,2}{10} \simeq 0{,}02382$ et $b = m(Y) - a\,m(X) \simeq 1{,}185 - 0{,}023\,82 \times 18 \simeq 0{,}7562$
D'où l'équation de la droite $D_{Y|X}$: $Y = 0{,}023\,82\,X + 0{,}756\,2$
Donc pour $x = 50$, on s'attend à $y = 0{,}023\,82 \times 50 + 0{,}756\,2 \simeq 1{,}947$.

Exercice 3 : Troubles du comportement

On considère dans cet exercice une personne qui souffre de troubles du comportement. On considère que chaque jour (indépendemment de ce qu'il s'est passé les jours précédents), cette personne a 45% de chances d'avoir besoin de l'aide d'une personne extérieure pour contrôler ses troubles du comportement.

On considère une période de 4 jours, et on désigne par X les nombre de jours (parmi ces 4 jours) au cours desquelles elle a eu besoin de l'aide d'une personne extérieure.

1. Calculer la probabilité $\mathbb{P}[X \leqslant 2]$. *Vous pourrez soit lister toutes les possibilités pour ces jours et préciser leur probabilité, soit constater directement qu'il s'agit d'une loi de probabilité vue en cours.*

Le plus simple de constater qu'on a la loi $\mathcal{B}(4; 0{,}45)$, car on compte le nombre de fois qu'un résultat (avoir besoin d'aide) a eu lieu au cours de 4 répétitions de la même expérience qui avait à chaque fois une probabilité 0,45.
La calculatrice indique alors que $\mathbb{P}\left[X \leqslant 2\right] \simeq 0{,}758\,5$.

Une autre méthode est de lister tous les résultats possibles : Si on note B quand cette personne a besoin d'aide et P quand elle n'a pas besoin, la situation « $X \leqslant 2$ » correspond aux possibilités FFFF (qui a probabilité $(0{,}55)^4$), BFFF, FBFF, FFBF, FFFB (qui ont probabilité $0{,}45 \times (0{,}55)^3$) et BBFF, BFBF, BFFB, FBBF, FBFB, FFBB (qui ont probabilité $(0{,}45)^2 \times (0{,}55)^2$). Cela donne en tout une probabilité $0{,}55^4 + 4 \times 0{,}45 \times 0{,}55^3 + 6 \times 0{,}45^2 \times 0{,}55^2 \simeq 0{,}758\,5$.

2. Calculer de même la probabilité $\mathbb{P}[2 \leqslant X \leqslant 3]$.

Avec la calculatrice, on a
$\mathbb{P}\left[2 \leqslant S \leqslant 3\right] = \mathbb{P}\left[S \leqslant 3\right] - \mathbb{P}\left[S \leqslant 1\right] \simeq 0{,}959 - 0{,}391 = 0{,}568$.

Si au contraire on listait les résultats possibles, la situation « $2 \leqslant X \leqslant 3$ » correspond aux possibilités BBFF, BFBF, BFFB, FBBF, FBFB, FFBB (qui ont probabilité $(0{,}45)^2 \times (0{,}55)^2$) et BBBF, BBFB, BFBB, FBBB (qui ont probabilité $(0{,}45)^3 \times 0{,}55$). Cela donne en tout une probabilité $6 \times 0{,}45^2 \times 0{,}55^2 + 4 \times 0{,}45^3 \times 0{,}55 \simeq 0{,}568$.