Remarque : En 2020-2021, les examens ont eu lieu en ligne. Ce document retranscrit un exemple d'énoncé du contrôle de deuxième session du 03 juillet.

Exercice 1 : Habiletés spatiales de sujets atteints (ou pas) de Schizophrénie

Mme Olivier, chercheuse en neurologie, étudie les habiletés spatiales de sujets atteints, ou pas, de Schizophrénie. Chaque sujet doit reconstituer une figure proposée à partir de morceaux de figures. Pour chaque sujet une trentaine de figures sont proposées. Chaque sujet obtient un score de performance qui tient compte du nombre de figures correctement reconstituées et du temps passé.

Elle fait passer ce test à un échantillon de 9 patients atteints de schizophrénie ("groupe cible"), et obtient les notes ci-dessous :

123	129	83	102	97	113	76	120	105

On supposera dans la suite que les notes obtenues à ce test suivent une loi normale.

1. (a) Calculer le score moyen et l'écart type au sein ce "groupe cible".
   moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{123+129+83+\cdots+105}{9} = \frac{948}{9} \simeq 105{,}33$
   $m(X^2) = \frac{\sum x_i^2}{n} = \frac{123^2+129^2+83^2+\cdots+105^2}{9} = \frac{102\,442}{9}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{102\,442}{9} - \left(\frac{948}{9}\right)^2 \simeq 287{,}33$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 16{,}95$

   (b) Estimer le score moyen qu'obtiendrait à ce test l'ensemble des patients atteints de schizophrénie. *On demande de déterminer un intervalle de confiance, avec la confiance* 98%.
   Comme $n = 9 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}01$ et $n-1 = 8$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 2{,}8965$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 17{,}358$.
   On estime donc que le score moyen des patients atteints de schizophrénie est dans l'intervalle $[105{,}33 - 17{,}358\,;\, 105{,}33 + 17{,}358] \simeq [87{,}97\,;\, 122{,}69]$ avec la confiance $c = 0{,}98$

   (c) Quelle taille d'échantillon serait nécessaire pour estimer ce score moyen à 0,05 point près, avec la confiance 98% ?
   Pour avoir une précision 0,05 avec la confiance 0,98, il faut $n > z_\alpha^2 \frac{(s_e)^2}{h^2} \simeq 2{,}326^2 \frac{16{,}95^2}{0{,}05^2} \simeq 621754$

2. À titre de comparaison, on fait passer le même test à un groupe de personnes qui n'ont jamais été diagnostiquées comme schizophrènes, ("groupe témoin"). On obtient les scores suivants :

Score	[65 ; 77[	[77 ; 89[	[89 ; 95[	[95 ; 101[	[101 ; 107[	[107 ; 113[	[113 ; 125[	[125 ; 137[
Effectif	4	6	10	11	6	6	5	2

(a) Combien y a-t-il d'individus dans ce groupe témoin ?
Il y en a $4+6+10+11+6+6+5+2=50$.

(b) Calculer le score moyen et l'écart type au sein de cet échantillon.
moyenne : $m(X)=\frac{\sum c_i\, n_i}{n}=\frac{71\times4+83\times6+92\times10+\cdots+131\times2}{50}=\frac{4\,921}{50}=98{,}42$
$m(X^2)=\frac{\sum c_i^2\, n_i}{n}=\frac{71^2\times4+83^2\times6+92^2\times10+\cdots+131^2\times2}{50}=\frac{494\,405}{50}$
$Var(X)=m(X^2)-m(X)^2=\frac{494\,405}{50}-\left(\frac{4\,921}{50}\right)^2\simeq 201{,}6$
Écart-type : $s(X)=\sqrt{Var(X)}\simeq 14{,}2$

(c) Estimer le score moyen qu'obtiendrait à ce test l'ensemble des personnes non diagnostiquées schizophrènes. *On demande de déterminer un intervalle de confiance, avec la confiance 98%.*
Comme n= 50 >30, on cherche z_α tel que $F(z_\alpha)=\frac{0{,}98+1}{2}=0{,}99$
On a $F(2{,}326)\simeq 0{,}99$ d'où $z_\alpha\simeq 2{,}326$ d'où $a_\alpha=z_\alpha\frac{s_e}{\sqrt{n-1}}\simeq 4{,}718$
On estime donc que le score moyen des personnes non diagnostiquées pour schizophrénie est dans l'intervalle $[98{,}42-4{,}718\,;\,98{,}42+4{,}718]\simeq[93{,}7\,;\,103{,}14]$, avec la confiance $c=0{,}98$.

3. Peut-on affirmer, avec la confiance 98% que les patients atteints de schizophrénie obtiennent en moyenne des scores moins bons à ce test que les personnes qui n'ont jamais été diagnostiquées schizophrènes ? *Vous conclurez en vous référant aux réponses des questions précédentes.*
Les intervalles que l'on trouvé aux questions 1.b. et 2.b. se chevauchent, donc on ne peut pas en déduire dans laquelle des deux situations la moyenne est la plus élevée.

Exercice 2 : Utilisation de la musique pour l'apprentissage

On étudie l'effet de la musique comme support d'apprentissage sur le fonctionnement mnésique de sujets atteints de la maladie d'Alzheimer.

On fait appendre à quinze patients des séries de gestes et de mouvements, synchronisés avec une musique spécifique. On attribue à chaque sujet une note (X) reflétant la qualité de restitution des gestes appris. Un échantillon de 425 individus donnes les notes ci-dessous :

Note X	[4 ; 6[	[6 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 14[	[14 ; 16[	[16 ; 18[	[18 ; 20[
Effectif	1	8	71	138	133	66	7	1

1. Sur ces données, combien vallent la moyenne et l'écart type de la note X ?
moyenne : $m(X)=\frac{\sum c_i\, n_i}{n}=\frac{5\times1+7\times8+9\times71+\cdots+19\times1}{425}=\frac{5\,075}{425}\simeq 11{,}94$
$m(X^2)=\frac{\sum c_i^2\, n_i}{n}=\frac{5^2\times1+7^2\times8+9^2\times71+\cdots+19^2\times1}{425}=\frac{62\,577}{425}$
$Var(X)=m(X^2)-m(X)^2=\frac{62\,577}{425}-\left(\frac{5\,075}{425}\right)^2\simeq 4{,}65$
Écart-type : $s(X)=\sqrt{Var(X)}\simeq 2{,}16$

2. À titre de comparaison, quels auraient été les effectifs théoriques si cette note suivait la loi $\mathcal{N}(12\,;\,2)$?

On obtient les probabilités suivantes et les effectifs théoriques suivants :

Note X	$[4\,;\,6[$	$[6\,;\,8[$	$[8\,;\,10[$	$[10\,;\,12[$	$[12\,;\,14[$	$[14\,;\,16[$	$[16\,;\,18[$	$[18\,;\,20[$
Probabilité	0,0013	0,0214	0,1359	0,3413	0,3413	0,1359	0,0214	0,0013
Effectif théorique	0,6	9,1	57,8	145,1	145,1	57,8	9,1	0,6

- On a d'abord calculé la probabilité associée à chaque intervalle, en remplaçant le premier intervalle par $[-\infty\,;\,6[$, et le dernier intervalle par $[18\,;\,+\infty]$.
- On a ensuite obtenu les effectifs théoriques en multipliant ces probabilités par l'effectif total qui est 425.

3. On décide désormais de vérifier si la musique aide à l'apprentissage de ces gestes. On fait donc apprendre deux séries de geste à des sujets atteints de la maladie d'Alzheimer : une série de gestes apprise en musique (aboutissant à la note X) et une autre série apprise avec uniquement la pulsation d'un métronome (aboutissant à la note Y). On obtient les données ci-dessous :

Sujet	1	2	3	4	5	6	7	8	9	10	11	12
Avec musique (X)	15,6	13,3	15,7	11,2	11,7	15,2	10,1	9,4	13,6	16,2	12	11,7
Sans musique (Y)	14,8	14,1	15,7	14,9	11,4	13,9	9	11,9	15,9	17,4	12,6	9,9

(a) Calculer le coefficient de corrélation linéaire entre ces deux variables. *On demande dans cette question d'indiquer les calculs effectués.*

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{15{,}6+13{,}3+15{,}7+\cdots+11{,}7}{12} = \frac{155{,}7}{12} \simeq 12{,}97$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{15{,}6^2+13{,}3^2+15{,}7^2+\cdots+11{,}7^2}{12} = \frac{2\,078{,}77}{12}$

$Var(X) = m(X^2) - m(X)^2 = \frac{2\,078{,}77}{12} - \left(\frac{155{,}7}{12}\right)^2 \simeq 4{,}88$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}21$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{14{,}8+14{,}1+15{,}7+\cdots+9{,}9}{12} = \frac{161{,}5}{12} \simeq 13{,}46$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{14{,}8^2+14{,}1^2+15{,}7^2+\cdots+9{,}9^2}{12} = \frac{2\,244{,}47}{12}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{2\,244{,}47}{12} - \left(\frac{161{,}5}{12}\right)^2 \simeq 5{,}91$

Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 2{,}43$

$m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{15{,}6\times14{,}8+13{,}3\times14{,}1+\ldots+11{,}7\times9{,}9}{12} = \frac{2\,144{,}35}{12} \simeq 178{,}696$

$Cov(X{,}Y) = m(X\,Y) - m(X)\,m(Y) = \frac{2\,144{,}35}{12} - \frac{155{,}7}{12}\frac{161{,}5}{12} \simeq 4{,}074$

$r(X{,}Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{4{,}074}{\sqrt{4{,}88\times5{,}912}} \simeq 0{,}758\,5$

(b) Si un·e patient·e obtient la note 11 sans musique, quelle note s'attend-on à ce qu'il/elle obtienne avec la musique ?

on pose $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{4{,}074}{5{,}91} \simeq 0{,}6893$ et $b' = m(X) - a'\,m(Y) \simeq 12{,}975 - 0{,}689\,3 \times 13{,}458 \simeq 3{,}698$
D'où l'équation de la droite $D_{X|Y}$: $X = 0{,}689\,3\,Y + 3{,}698$
Donc pour $y = 11$, on s'attend à $x = 0{,}689\,3 \times 11 + 3{,}698 \simeq 11{,}28$.

Exercice 3 : Estime de soi et dépression

Dans cet exercice, on cherche à déterminer si le déficit de confiance en soi touche plus les patients atteints de dépression que le reste de la population.

Pour cela on remarque tout d'abord qu'au sein de la population française adulte, il y a 35% d'individus qui ont une estime de soi faible (score inférieur à 30 dans l'échelle de Rosenberg). On va comparer cette proportion à celle que l'on observe chez un échantillon de patients souffrant de dépression chronique.

1. On considère un échantillon de 9 adultes français·e·s, et on désigne par X le nombre de personnes de l'échantillon qui ont une estime de soi faible.
   (a) Dans ces conditions, X suit une loi binomiale. Indiquez les paramètres le cette loi binomiale, ainsi que sa moyenne.
   $X \sim \mathcal{B}(9\,;\,0{,}35)$ donc $m(X) = 9 \times 0{,}35 = 3{,}15$.
   (b) Calculer la probabilité $\mathbb{P}[2 \leqslant X \leqslant 4]$. *Vous indiquerez quels calculs vous avez effectués (et si c'est la calculatrice qui a effectué des calculs, vous indiquerez sommairement ce que vous avez demandé à la calculatrice).*
   $\mathbb{P}[2 \leqslant X \leqslant 4] = \mathbb{P}[X \leqslant 4] - \mathbb{P}[X \leqslant 1] \simeq 0{,}828\,28 - 0{,}121\,1 \simeq 0{,}707\,2$, où l'on a utilisé la calculatrice pour calculer $\mathbb{P}[X \leqslant 4] \simeq 0{,}828\,28$ et $\mathbb{P}[X \leqslant 1] \simeq 0{,}121\,09$
2. On considère désormais un échantillon de 95 adultes français·e·s, et on désigne à nouveau par X le nombre de personnes de l'échantillon dont le niveau d'estime de soi est faible.
   (a) X suit à nouveau une loi binomiale. Justifier que l'on peut l'approximer par une loi normale, et précisez sa moyenne et son écart type.
   On a $n = 95 > 30$, et $n\,p \simeq 33{,}2 > 5$, et $n(1-p) \simeq 61{,}8 > 5$ donc on peut approximer X par $\mathcal{N}(33{,}2; 4{,}65)$ où $4{,}65 \simeq \sqrt{n\,p(1-p)}$
   (b) En utilisant la loi normale de la question précédente, déterminez $\mathbb{P}[28 \leqslant X \leqslant 36]$. *On vous demande de faire une correction de continuité, et d'indiquer votre calcul en utilisant la table du formulaire.*

$$\begin{aligned}\mathbb{P}[28 \leqslant X \leqslant 36] &= \mathbb{P}[27{,}5 \leqslant X \leqslant 36{,}5] = \mathbb{P}[\tfrac{27{,}5-33{,}2}{4{,}65} < \tfrac{X-33{,}2}{4{,}65} < \tfrac{36{,}5-33{,}2}{4{,}65}] \\ &\simeq \mathbb{P}[-1{,}23 < Z < 0{,}71] \simeq F(0{,}71) - F(-1{,}23) \\ &\simeq 0{,}7611 - (1 - 0{,}890\,7) \simeq 0{,}761\,1 - 0{,}109\,3 \\ &\simeq 0{,}652\end{aligned}$$

3. On considère un échantillon de 95 patients adultes souffrant de dépression chronique. On observe qu'au sein de cet échantillon il y en a 44 qui ont une faible estime de soi.

   (a) À partir de cet échantillon, estimer avec la confiance 95% la proportion de personnes qui ont une faible estime de soi parmi l'ensemble des patients soufrant de dépression chronique. *Vous établirez un intervalle de confiance, avec la confiance 95%.*

   On a $n = 95 > 30$, et $p_e = \frac{44}{95} \simeq 0{,}463\,2$, d'où $n\,p_e = 44 > 5$, et $n(1-p_e) = 51 > 5$ donc on peut utiliser la procédure du formulaire pour estimer cette proportion .

   On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$

   d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}1003$.

   On estime donc que est dans l'intervalle $[0{,}3629; 0{,}5635]$ avec la confiance $c = 0{,}95$.

   (b) Conclure : Avec la confiance 95%, peut-on affirmer qu'il y a une plus grande proportion de personnes qui ont une faible estime de soi parmi les patients souffrant de dépression chronique que parmi l'ensemble des adultes ?

   Oui, car toutes les valeurs de l'intervalle trouvé en question 3.a sont plus grandes que 35%.

Exercice 4 : Pratique du sport

On considère les 5 habitants d'un même lotissement, et on leur demande la fréquence à laquelle il pratiquent une activité sportive, que l'on résume par le mot-clé "fréquent" (plus de deux fois par semaine) "occasionnel" (entre deux fois par moi et deux fois par semaine) et "rare" (moins de deux fois par mois). On obtient les réponses ci-dessous :

Nom	Mme Maurin	Mme Dos Santos	Mr Hebert	Mr Lemonnier	Mme Fournier
Pratique sportive	occasionnel	occasionnel	fréquent	fréquent	rare

On choisit 3 personnes au hasard parmi ces 5 habitants du lotissement.

1. Lister tous les choix possibles pour ces trois personnes, et conclure que chaque cas a une probabilité de 10%.

   En notant M pour Mme Maurin, D pour Mme Dos Santos, H pour Mr Hebert, L pour Mr Lemonnier et F pour Mme Fournier, on a les 10 choix ci-dessous :
   MDH, MDL, MDF, MHL, MHF, MLF, DHL, DHF, DLF et HLF
   Chacun a autant de chance d'être choisi, soit une probabilité $\frac{1}{10} = 10\%$.

2. Quelle est la probabilité d'avoir

   (a) exactement 2 femmes dont une pratique "occasionnellement" le sport ?

Cela correspond aux 4 cas suivants : MHF, MLF, DHF et DLF.
La probabilité est donc $\frac{4}{10} = 0{,}4$.

(b) moins de 2 hommes ?
Cela correspond aux 7 cas suivants : MDH, MDL, MDF, MHF, MLF, DHF et DLF.
La probabilité est donc $\frac{7}{10} = 0{,}7$.

(c) deux sportif/sportives occasionnel·le·s ?
Cela correspond aux 3 cas suivants : MDH, MDL et MDF.
La probabilité est donc $\frac{3}{10} = 0{,}3$.

3. On note X le nombre de femmes parmi les trois personnes choisies au hasard. Déterminer la loi de X (c'est à dire, calculer chaque probabilité $\mathbb{P}[X = \dots]$).
On remarque que X prend soit la valeur 1 (cas MHL, DHL et HLF), soit la valeur 2 (cas MDH, MDL, MHF, MLF, DHF et DLF), soit la valeur 3 (cas MDF).
On obtient donc les probabilités

k	1	2	3
$\mathbb{P}[X = k]$	$\frac{3}{10} = 0{,}3$	$\frac{6}{10} = 0{,}6$	$\frac{1}{10} = 0{,}1$

.