Veuillez rendre ce sujet avec votre copie.

Le formulaire et la calculatrice sont autorisés. Vous êtes invités à répondre directement sur l'énoncé mais si vous avez besoin de place supplémentaire, vous pouvez détailler certaines questions sur la copie que vous rendrez avec l'énoncé. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où les détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.

Exercice 1. Workaholism.

Dans une société à la recherche de plus en plus de productivité, l'implication au travail est socialement valorisée. L'addiction au travail appelée *workaholism* souvent favorisée par l'obtention de primes, de récompenses de la part de l'employeur est alors considérée comme une addiction "positive". Elle se caractérise par la compulsion à consacrer toujours plus de temps et d'énergie au travail en dépit des conséquences négatives sur la santé et/ou la vie personnelle affective et familiale. Un des derniers stades le plus connu de cette addiction est le syndrome du *Burn Out*. Parmi les 440 000 salariés du secteur automobile en France, on compte 92 400 personnes présentant des signes d'addiction au travail.

1. On choisit au hasard un échantillon $\mathcal{E}_1$ de 20 salariés du secteur automobile français. Expliquer pourquoi ce choix suit un modèle de tirage sans remise. Justifier que l'on puisse se ramener aux calculs d'un modèle de tirage avec remise.

   Les salariés qui sont choisis au hasard ne peuvent participer à l'évaluation qu'une seule fois : il s'agit donc bien d'un tirage sans remise. Mais la taille de la population totale 440 000 est plus de 10 fois plus grande que celle de l'échantillon, on peut remplacer le modèle du tirage sans remise par celui du tirage avec remise.

2. On note S le nombre de salariés au sein de l'échantillon qui présentent des signes d'addiction au travail. Quelle est la loi de probabilité de S et quelle est sa moyenne ?

   Comme il s'agit d'un modèle de tirage et qu'on peut utiliser ici le tirage avec remise, S suit une loi binomiale. La proportion de personne sujettes au workaholism est $\frac{92400}{440000} \approx 0,21$

   $S \sim \mathcal{B}(20; 0,21)$ $\qquad$ $m(S) = 4,2$

3. Calculer la probabilité qu'il y ait au plus 3 addictes au travail dans cet échantillon de 20 salariés. *Détailler les calculs.*

$$\begin{aligned}\mathbb{P}[S \leq 3] &= \mathbb{P}[S=0] + \mathbb{P}[S=1] + \mathbb{P}[S=2] + \mathbb{P}[S=3] \\ &= \binom{20}{0}(0,21)^0(0,79)^{20} + \binom{20}{1}(0,21)^1(0,79)^{19} + \binom{20}{2}(0,21)^2(0,79)^{18} \\ &+ \binom{20}{3}(0,21)^3(0,79)^{17} \approx 0,369\end{aligned}$$

4. On considère maintenant un nouvel échantillon $\mathcal{E}_2$ de 55 salariés du secteur automobile français et S désigne encore le nombre de salariés présentant des signes d'addiction au travail. Justifier que l'on peut approcher S par la loi normale $X \sim \mathcal{N}(11,55; 3,02)$.

$n = 55 > 30$, $np = 0,21 * 55 = 11,55 > 5$ et $n(1-p) = 0,79 * 55 = 43,45 > 5$.
On peut donc approcher la loi binomiale par une loi normale, de moyenne $np = 11,55$ et d'écart-type $\sqrt{np(1-p)} = 3,02$.

5. Quelle est la probabilité qu'il y ait au moins 17 addictes au travail au sein de l'échantillon $\mathcal{E}_2$?

On demande d'effectuer une correction de continuité et de justifier brièvement la réponse.

$$\mathbb{P}[S \geq 17] \approx \mathbb{P}[X \geq 16,5] = \mathbb{P}\Big[Z \geq \frac{16,5 - 11,55}{3,02}\Big] = \mathbb{P}[Z \geq 1,64]$$
$$= 1 - F(1,64) = 1 - 0,9495 = 0,0505$$

6. On choisit au hasard un échantillon de 50 salariés du secteur automobile japonais, 13 salariés présentent des signes d'addiction au travail. Donnez une estimation de la proportion de personnes dépendantes au travail dans la population des salariés du secteur automobile japonais. (on choisira une confiance de 95%).

Répondre en calculant un intervalle de confiance.
Justifier soigneusement les calculs.

On calcule $p_e = \frac{13}{50} = 0,26$.
On remarque que $n = 50 \geq 30$, que $np_e = 13 \geq 5$ et $n(1-p_e) = 37 \geq 5$.
Comme il s'agit d'un grand échantillon on peut utiliser la loi normale.
On lit $z_{0.05} = 1,96$ dans la table. La marge est alors:

$$a_{0,05} = 1,96\sqrt{\frac{0,26(1-0,26)}{50}} = 0,122$$

L'intervalle de confiance à 95% est donc

$$[0,26 - 0,122;\ 0,26 + 0,122] = [0,138;\ 0,382]$$

7. Peut-on affirmer avec une confiance de 95% que les salariés du secteur automobile sont plus sujets à l'addiction au travail au Japon qu'en France ? *Justifier votre réponse.*

Comme la proportion au sein de la population française appartient à l'intervalle de confiance de l'estimation de la proportion au Japon, on ne peut pas affirmer avec une confiance de 95% que les japonais sont plus sujets à l'addiction au travail.
Il faudrait un échantillon plus important pour avoir une estimation plus précise.

Exercice 2. Nombre de pas quotidiens. Une association souhaite mener une étude sur l'activité physique et sportive de la population française et comprendre le lien entre l'activité de chaque individu et son indice de masse corporelle (IMC). Ainsi pour chaque individu, l'association enregistre l'IMC et le nombre de pas quotidiens. Sur un échantillon de 12 personnes, on obtient les données suivantes:

Nunéro	1	2	3	4	5	6	7	8	9	10	11	12		
IMC	23,0	32,1	18,7	18,2	40,4	34,2	21,5	28,0	17,2	20,0	24,3	25,3		
Nbre de pas	8 504	3 732	9 290	7 533	3 430	9 268	10 230	6 560	7 400	9 307	12 468	7 890		

Pour la suite, on note X l'IMC et Y le nombre de pas quotidiens.

1. Déterminer la moyenne et l'écart-type des variables statistiques X et Y.

$m(X) = 25,24$ $s(X) = 6,89$

$m(Y) = 7\,967,7$ $s(Y) = 2\,448,2$

2. L'association pense qu'il y a un lien entre l'IMC d'un individu et son nombre de pas quotidiens. Calculer le coefficient de corrélation linéaire des variables X et Y et conclure.

Calcul du coefficient de corrélation linéaire des variables X et Y: $-0,55$

Interprétation : Le coefficient de corrélation est éloigné de 0 donc il existe un lien significatif entre les variable X et Y. Ce coefficient est négatif: plus l'IMC est important plus le nombre de pas est petit et vice et versa.

3. Donner une estimation du nombre moyen de pas quotidiens avec une confiance de 98% (par hypothèse, le nombre de pas quotidiens d'un individu choisi au hasard suit une loi normale). *Répondre en calculant un intervalle de confiance.*

Estmation d'un moyenne avec un petit échantillon : loi de Student.
$m_e = 7\,967,7$ et $s_e = 2\,448,2$. On cherche dans la table $t_{0.02}$ pour une loi de Student à 11 ddl. $t_{0.02} = 2,7181$. On calcule la marge:

$$a_{0.05} = 2,7181 \times \frac{2\,448,2}{\sqrt{11}} = 2\,006,4$$

L'intervalle de confiance à 98% est donc

$$[7\,967,7 - 2\,006,4; 7\,967,7 + 2\,006,4] = [5\,961,3;\ 9\,974,1]$$

Exercice 3. Recrutement.
Une entreprise de production vient d'apprendre la démission de quatre salariés. Elle organise dans l'urgence une campagne de recrutement et 26 ouvriers se présentent, tous avec des expériences d'emploi très différentes. Le responsable des ressources humaines décide alors de choisir les 4 nouveaux salariés au hasard parmi les ouvriers demandeurs d'emploi.

On souhaite calculer la probabilité que l'ouvrier ayant le plus d'expérience soit sélectionné. Pourquoi ne peut-on pas utiliser le modèle binomial pour calculer cette probabilité ?

Car il ne s'agit pas d'un modèle de tirage avec remise, un ouvrier ne peut pas etre séléctionné plusieurs fois.

Combien de groupes de 4 ouvriers contenant l'ouvrier le plus expérimenté peut-on constituer ?
On demande de justifier la réponse.

Pour constituer un groupe de 4 ouvriers contenant l'ouvier le plus expérimenter, il suffit de choisir cet ouvrier en particulier et d'ajouter 3 autres ouvriers. Ainsi le nombre de choix possibles correspond au nombre de choix de 3 personnes dans un groupe de 25 ouvriers: $\binom{25}{3}$

Avec quelle probabilité, l'ouvrier ayant le plus d'expérience se trouve-il parmi les 4 ouvriers recrutés ? *On demande de justifier la réponse.*

Ici on peut utiliser pour le calcul des probabilités, le rapport entre le nombre de cas favorables sur le nombre de cas possibles. On compte donc tous les groupes différents de 4 personnes qui contiennent l'ouvrier ayant le plus d'expérience : $\binom{25}{3}$ (cas favorables) puis le nombre de groupes de 4 personnes quelconques.
$\frac{\binom{25}{3}}{\binom{26}{4}} = \frac{4}{26} = \frac{2}{13} \approx 0,154.$

Parmi les 26 ouvriers candidats, 10 n'ont jamais utilisé la machine outil présente sur la chaîne de production. Si l'entreprise les recrute, elle devra leur payer une formation spécifique. Quelle est la probabilité que parmi les quatre ouvriers recrutés, deux doivent suivre cette formation spécifique ? *On demande de justifier la réponse.*

$\frac{\binom{10}{2}\binom{16}{2}}{\binom{26}{4}} \approx 0,361$

Exercice 4. Nombre d'amis dans un réseau social. On considère un réseau social dans lequel chaque abonné a un certain nombre d'amis (personnes qui sont en relation avec l'abonné). Pour décrire la variable statistique X qui représente le nombre d'amis d'un abonné choisi au hasard, un groupe de 510 abonnés est sélectionné et les données recueillies sont présentées dans le tableau ci-dessous:

Nbre d'amis	$[0;40[$	$[40;80[$	$[80;120[$	$[120;160[$	$[160;200[$	$[200;240[$
Effectif	42	123	221	88	22	14

1. Calculer la moyenne et l'écart-type du nombre d'amis.
Justifier votre réponse en présentant le calcul effectué.

$$m(X) = \frac{1}{510}\Big(20 \times 42 + 60 \times 123 + \dots\Big) \approx 97,41$$

$$m(X^2) = \frac{1}{510}\Big(20^2 \times 42 + 60^2 \times 123 + \dots\Big) \approx 11\,342,75$$

$\mathrm{Var}(X) = 11\,342,75 - 97,41^2 = 1854,04$ et $s(X) = 43,06$.

2. Quelle est la proportion d'abonnés ayant plus de 120 amis ?

$\frac{88+22+14}{510} \approx 0,243$

3. Calculer le premier quartile du nombre d'amis. *Présenter le calcul effectué.*

$40 + \frac{80-40}{0,32353-0,08235} \times (0,25 - 0,08235) = 67,8$

4. Supposons que X suit une loi normale de moyenne $\mu = 97,4$ et d'écart-type $43,06$. Calculer l'effectif théorique de la seconde classe $[40;80[$.
(Pas de correction de continuité pour cette question).

$$\begin{aligned}\mathbb{P}[40 < X \leq 80] = \mathbb{P}\Big[\frac{40-97,4}{43,06} < Z \leq \frac{80-97,4}{43,06}\Big] &= \mathbb{P}[-1,33 < Z \leq -0,40] \\ &= F(-0,40) - F(-1,33) = F(1,33) - F(0,40) \\ &= 0,9082 - 0,6554 \approx 0,2528\end{aligned}$$

L'effectif théorique est donné par $0,2528 \times 510 = 128,9$.

5. A partir des valeurs observées sur l'échantillon (tableau), donner une estimation du nombre moyen d'amis et donner la précision (ou marge) de l'estimation.
On déterminera par un court calcul l'intervalle de confiance à 98%.

il s'agit d'une estimation de moyenne pour un grand échantillon.
On utilise donc la loi normale.
$m_e = 97,4$ et $s_e = 43,06$. On lit dans la table $z_{0,02} = 2,326$.
La marge est égale à

$$a_{0,02} = 2,326 \times \frac{43,06}{\sqrt{509}} = 4,4$$

$I_\alpha = [93,0;\ 101,8]$ marge $= 4,4$