IMBS3 - ISCID-CO, site de Dunkerque, 2015/2016

Mathématiques appliquées à la gestion - TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES

Fiche de Mathématiques 1 - Prérequis - Exercices de révision.

Exercice 1 *Correction* :

1. On considère l'événement

$$\mathcal{A} : \text{“la pompe à chaleur tombe en panne durant le mois”}.$$

Soit X_i la variable aléatoire qui vérifie $X_i(\mathcal{A}) = 1$ et $X_i(\overline{\mathcal{A}}) = 0$. On a d'après l'énoncé la distribution de probabilité suivante :

x_j	0	1	Total
$p_j = p(\{X_i = x_j\})$	$0,875$	$0,125$	1

donc $X_i \rightsquigarrow \mathcal{B}(0,125)$. X est la variable aléatoire qui, à chaque année, associe le nombre de pannes survenues donc $X = X_1 + X_2 + \ldots + X_{12}$. On en déduit que

$$X \rightsquigarrow \mathcal{B}(12; 0,125)$$

2. On a la relation $p(\{X = k\}) = C_{12}^k (0,125)^k (0,875)^{12-k}$ pour $k \in \{0, 1, \ldots, 12\}$.
   (a) $p(\{X = 0\}) = C_{12}^0 (0,125)^0 (0,875)^{12-0} \simeq 0,201$.
   (b) $$p(\{X \leq 2\}) = p(\{X = 0\}) + p(\{X = 1\}) + p(\{X = 2\})$$
   $$\Leftrightarrow p(\{X \leq 2\}) = C_{12}^0 (0,875)^{12} + C_{12}^1 (0,125)(0,875)^{11} + C_{12}^2 (0,125)^2 (0,875)^{10}$$
   $$\Leftrightarrow p(\{X \leq 2\}) \simeq 0,201 + 0,345 + 0,271 \simeq 0,817.$$

3. $E(X) = \displaystyle\sum_{i=1}^{12} E(X_i)$ car les variables X_i sont indépendantes deux à deux. On en déduit que $E(X) = 12 \times 0,125 = 1,5$ qui représente le nombre moyen de pannes annuel sur un grand nombre d'années.

4. (a) On a la relation $Y = 320X$ puisque 320 euros sont nécessaires pour réparer une panne.
   (b) Par linéarité de l'espérance, $E(Y) = 320E(X) = 320 \times 1,5 = 480$. Cette espérance correspond à la somme moyenne annuelle dépensée pour réparer les pannes sur un grand nombre d'années.
   (c) En comparant les sommes, il vaut mieux ne pas souscrire au contrat de maintenance et réparer les pannes ponctuellement.

5. On suppose que $\mathcal{B}(n, p) \sim \mathcal{P}(\lambda)$ avec $\lambda = np = 12 \times 0,125 = 1,5$.
   - $p(\{X = 0\}) = \dfrac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} \simeq 0,223$.
   - $p(\{X \leq 2\}) = p(\{X = 0\}) + p(\{X = 1\}) + p(\{X = 2\})$
   $$\Leftrightarrow p(\{X \leq 2\}) = \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!} \simeq 0,223 + 0,335 + 0,251 \simeq 0,809.$$

6. - $\left|\dfrac{0,223 - 0,201}{0,223}\right| \simeq 0,099 < 0,1$ donc l'approximation est justifiée.
   - $\left|\dfrac{0,809 - 0,817}{0,809}\right| \simeq 9,889 \times 10^{-3} < 0,1$ donc l'approximation est également justifiée.

Exercice 2 *Correction* :

1. . $p_1 = p(\{8,45 \leq X \leq 8,70\})$: pour pouvoir se référer aux tables de la loi normale, on pose
$$T = \frac{X - m_X}{\sigma_X} = \frac{X - 8,55}{0,05}.$$
Donc $p_1 = p(\{-2 \leq T \leq 3\}) = \Pi(3) - \Pi(-2) = \Pi(3) - (1 - \Pi(2))$ de par la symétrie de la fonction de densité. On utilise ensuite la table de la loi normale centrée réduite et on récupère les valeurs $\Pi(3) = 0,99865$ et $\Pi(2) = 0,9772$. Par conséquent, $p_1 = 0,99865 - (1 - 0,9772) = 0,97585$.

   . $p_2 = p(\{5,07 \leq Y \leq 5,33\})$: on pose
$$W = \frac{Y - m_Y}{\sigma_Y} = \frac{Y - 5,2}{0,05}.$$
Donc $p_2 = p(\{-2,6 \leq W \leq 2,6\}) = \Pi(2,6) - \Pi(-2,6) = \Pi(2,6) - (1 - \Pi(2,6)) = 2\Pi(2,6) - 1$. On utilise ensuite la table de la loi normale centrée réduite et on récupère la valeur $\Pi(2,6) = 0,9953$. Par conséquent, $p_2 = 2 \times 0,9953 - 1 = 0,9906$.

2. (a) Calculons le pourcentage de pilules conformes à la sortie de la chaîne de fabrication. On cherche donc la probabilité de l'intersection des deux événements $8,45 \leq X \leq 8,70$ et $5,07 \leq Y \leq 5,33$ soit
$$p = p(\{(8,45 \leq X \leq 8,70) \cap (5,07 \leq Y \leq 5,33)\}).$$
Comme les événements X et Y sont indépendants,
$$p = p(\{8,45 \leq X \leq 8,70\}) \times p(\{5,07 \leq Y \leq 5,33\}) = 0,97585 \times 0,9906 = 0,966677$$
soit un pourcentage de $96,6677\%$. Le pourcentage de pilules qui seront hors normes à la sortie de la chaîne de fabrication vaut par conséquent $100 - 96,6677 \simeq 3,33\%$.

   (b) Si on veut que le pourcentage de pilules défectueuses ne dépasse pas 3%, le procédé ne peut être retenu puisque $3,33\% > 3\%$. On cherche σ_Y pour que le pourcentage de pilules défectueuses soit inférieur à 3%. On veut déterminer la valeur de $\tilde{p}$ telle que
$$.\ 1 - 0,97585 \times \tilde{p} < 0,03 \Leftrightarrow \tilde{p} > 0,9940052$$
ce qui revient à déterminer a tel que
$$p(\{5,07 \leq Y \leq 5,33\}) = 2\Pi(a) - 1 > 0,9940052 \Leftrightarrow \Pi(a) > 0,9970026.$$
À l'aide de la table de la loi normale centrée réduite, on trouve
$$2,75 < a < 2,76$$
On utilise ensuite l'interpolation linéaire, on a les encadrements

2,75	a	2,76
0,9970	0,9970026	0,9971

ce qui nous permet d'écrire que
$$\frac{a - 2,75}{0,9970026 - 0,9970} = \frac{2,76 - 2,75}{0,9971 - 0,9970} \Leftrightarrow a = 2,75026.$$
Finalement comme $a = \dfrac{5,33 - 5,2}{\sigma_Y}$, on trouve $\sigma_Y \simeq 0,047$.

3. (a) On a $S = X + Y$. Par conséquent,
   . $E(S) = E(X+Y) = E(X) + E(Y) = 8,55 + 5,20 = 13,75$ car X et Y sont des variables indépendantes.
   . $V(S) = V(X+Y) = V(X) + V(Y)$ car X et Y sont des variables indépendantes. On en déduit que $V(S) = 2 \times (0,05)^2 = 0,005$. L'écart-type de S vaut alors $\sigma_S = 0,0707$.

   (b) $S \rightsquigarrow \mathcal{N}(13,75; 0,0707)$. On pose
$$V = \frac{S - 13,75}{0,0707}$$
ce qui implique que $p = p(\{13,6 \leq S \leq 13,8\}) = p(\{-2,122 \leq V \leq 0,707\}) = \Pi(0,707) - \Pi(-2,122)$. Toujours en utilisant la symétrie de la fonction de densité, on obtient

$$p = \Pi(0,707) + \Pi(2,122) - 1.$$

Déterminons $\Pi(0,707)$ et $\Pi(2,122)$ par interpolation linéaire.

. Pour $\Pi(0,707)$ on a les encadrements

0,70	0,707	0,71
0,7580	$\Pi(0,707)$	0,7611

ce qui nous permet d'écrire que

$$\frac{\Pi(0,707) - 0,7580}{0,7611 - 0,7580} = \frac{0,707 - 0,70}{0,71 - 0,70} \Leftrightarrow \Pi(0,707) = 0,7582.$$

. Pour $\Pi(2,122)$ on a les encadrements

2,12	2,122	2,13
0,9830	$\Pi(2,122)$	0,9834

ce qui nous permet d'écrire que

$$\frac{\Pi(2,122) - 0,9830}{0,9834 - 0,9830} = \frac{2,122 - 2,12}{2,13 - 2,12} \Leftrightarrow \Pi(2,122) = 0,9831.$$

Finalement, $p = 0,7582 + 0,9831 - 1 = 0,7413$.

4. Soit $p_{HN} = 0,01$.

(a) On tire sans remise 100 pilules dans un ensemble où chaque tirage est un tirage de Bernoulli (soit pilule hors norme de probabilité p, soit pilule conforme de probabilité $q - 1 = p$). Le nombre de pilules qui sont hors-norme est une variable binomiale de paramètres $n = 100$ et $p = 0,01$.

(b) On a $n \geq 30$, $p \leq 0,1$ et $np = 1 \leq 10$ qui justifient l'approximation de Poisson, le paramètre de cette variable étant $np = 1$. Le complémentaire de l'événement $Z \geq 5$ est $Z \in \{0,1,2,3,4\}$. On approche la probabilité de cet événement par la somme

$$e^{-1}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) = \frac{65}{24e} = 0,9963$$

puisque pour la loi de Poisson, $p(\{X = k\}) = e^{-\lambda}\dfrac{\lambda^k}{k!}$. On récupère finalement la valeur approximative cherchée en calculant $p(\{Z \geq 5\}) = 1 - 0,9963 = 0,0037$.

5. (a) On a $U \rightsquigarrow \mathcal{B}(n,p)$. On cherche $p = p(\{X + Y < 13,8\}) = p(\{S < 13,8\})$. On a vu précédemment que $p(\{S < 13,8\}) = p(\{V \leq 0,707\}) = \Pi(0,707) = 0,7582$ donc $p = 0,7582$.

(b) On a ici $n = 100 \geq 20$, $np = 75,82 \geq 10$ et $nq = 24,18 \geq 10$ et l'approximation par une variable normale de paramètres $m = np = 100 \times 0,7582 = 75,82$ et $\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0,7582 \times 0,2418} = 4,28$ est donc justifiée. Ce ne serait pas le cas pour la loi de Poisson car, par exemple, $p = 0,75 \geq 0,1$. La probabilité considérée a pour valeur approchée

$$\begin{aligned}\Pi\left(\tfrac{85-75,82}{4,28}\right) - \Pi\left(\tfrac{70-75,82}{4,28}\right) &= \Pi(2,14) - \Pi(-1,36) = \Pi(2,14) - (1 - \Pi(1,36)) \\ &= 0,9838 - (1 - 0,9131) = 0,8969.\end{aligned}$$

Exercice 3 *Correction* :

1. (a) On a : $p(A) = p(\{X = 0\}) = e^{-0,28}\dfrac{(0,28)^0}{0!} \simeq 0,756$.

(b) L'évènement B est l'union des évènements deux à deux incompatibles "un véhicule tiré au hasard dans le parc a eu exactement 0 sinistre pendant l'année considérée", "un véhicule tiré au hasard dans le parc a eu exactement 1 sinistre pendant l'année considérée" et "un véhicule tiré au hasard dans le parc a eu exactement 2 sinistres pendant l'année considérée". Donc

$$p(B) = p(\{X=0\}) + p(\{X=1\}) + p(\{X=2\}) = e^{-0,28}\frac{(0,28)^0}{0!} + e^{-0,28}\frac{(0,28)^1}{1!} + e^{-0,28}\frac{(0,28)^2}{2!} \simeq 0,997.$$

2. (a) Nous avons une suite de 15 épreuves de Bernoulli indépendantes. Chaque épreuve n'a que deux issues possibles : soit l'évènement E est réalisé avec une probabilité de $0,6$, soit il n'est pas réalisé, avec une probabilité de $0,4$. Ceci prouve que la variable aléatoire Y suit une loi binomiale de paramètres 15 et $0,6$.

(b) On veut calculer la probabilité de l'événement : $p(\{Y = 10\}) = C_{15}^{10} \times (0,6)^{10}(0,4)^5 \simeq 0,186$.

3. Puisque C suit la loi normale de moyenne 1200 et d'écart type 200, la variable aléatoire D définie par $D = \dfrac{C-1200}{200}$ suit la loi normale centrée réduite. On trouve alors $C = 1200 + 200D$ ce qui permet facilement d'en déduire : $1000 \leq C \leq 1500 \Leftrightarrow -1 \leq D \leq \dfrac{3}{2}$. Il vient alors, en utilisant le formulaire :

$$p(\{1000 \leq C \leq 1500\}) = p(\{-1 \leq D \leq \frac{3}{2}\}) = \Pi(\frac{3}{2}) - \Pi(-1) = \Pi(\frac{3}{2}) - (1 - \Pi(1)) \simeq 0,775.$$

4. (a) L'estimation ponctuelle est donnée par le pourcentage obtenu sur l'échantillon prélevé : $p = 91\% = 0,91$.

   (b) On pose : $T = \dfrac{F-p}{\sqrt{\frac{p(1-p)}{100}}}$. T suit la loi normale centrée réduite. L'intervalle de confiance pour T à 95% est défini par : $p(-t \leq T \leq t) = 0,95 \Leftrightarrow 2\Pi(t) - 1 = 0,95 \Leftrightarrow \Pi(t) = 0,975 \Leftrightarrow t \simeq 1,96$ en utilisant le formulaire. On en déduit : $p(-1,96 \leq T \leq 1,96) = 0,95 \Leftrightarrow p\left(-1,96 \leq \frac{F-p}{\sqrt{\frac{p(1-p)}{100}}} \leq 1,96\right) = 0,95$. Afin de pouvoir obtenir un intervalle de confiance, il nous faut dans ce cas approximer l'écart-type $\sqrt{\dfrac{p(1-p)}{100}}$ de F par σ' qui d'après la théorie vaut

$$\sigma' = \sqrt{\frac{n}{n-1}} \times \sqrt{\frac{f(1-f)}{n}} = \sqrt{\frac{f(1-f)}{n-1}} = \sqrt{\frac{0,91(1-0,91)}{100-1}} \simeq 2,8762 \times 10^{-2}.$$

   On en déduit ensuite : $p\left(-1,96 \leq \dfrac{F-p}{\sigma'} \leq 1,96\right) = 0,95 \Leftrightarrow p(p - 1,96 \times \sigma' \leq F \leq p + 1,96 \times \sigma') = 0,95$. On utilise l'estimation ponctuelle de p à savoir que $p \simeq 0,91$ pour trouver que $p - 1,96 \times \sigma' = 0,91 - 1,96 \times 2,8762 \times 10^{-2} \simeq 0,85363$ et $p + 1,96 \times \sigma' = 0,91 + 1,96 \times 2,8762 \times 10^{-2} \simeq 0,96637$. En conclusion, l'intervalle de confiance cherché est : $[0,853; 0,967]$

   (c) Cette affirmation est fausse d'après la théorie.

Exercice 4 *Correction* :

1. On a le tableau suivant :

Tâche	Antériorités	Niveaux			n_i	p_i	O_i	$E(t)$	$V(t)$
		0	1	2					
A		A			8	11	5	8	1
B	A		B		12	14	10	12	4/9
C	A		C		9	16	8	10	16/9
D	A,B,C			D	16	20	12	16	16/9
E	A		E		8	10	6	8	4/9

   La durée du chemin critique (A, C, D) (graphe MPM) vaut 36.

2. (a) Soient X, Y, T les variables aléatoires durées des tâches A, B, D, $X \rightsquigarrow \mathcal{N}(8, 1)$, $Y \rightsquigarrow \mathcal{N}(12, \frac{2}{3})$, $T \rightsquigarrow \mathcal{N}(16, \frac{4}{3})$.
   Soit U la variable aléatoire durée du chemin (A, C, D), $U = X + Y + Z$ est donc la somme de 3 variables aléatoires normales indépendantes. $E(U) = 8 + 1216 = 36$, $V(U) = V(X) + V(Y) + V(T) = 1 + \frac{4}{9} + \frac{16}{9} = \frac{29}{9}$.
   Donc $U \rightsquigarrow \mathcal{N}(36, \frac{\sqrt{29}}{3})$.

   (b) $P(U < 31) = P(Z < \frac{31-36}{\frac{\sqrt{29}}{3}}) = \Pi(-2,785) = 0,00265$.
   $P(U < 36) = P(Z < 0) = \Pi(0) = 0,5$.
   $P(U < 40) = P(Z < \frac{4}{\frac{\sqrt{29}}{3}}) = \Pi(2,228) = 0,9870$.

3. Le risque de l'entrepreneur est évalué par la probabilité que les travaux soit une durée supérieure à 33 jours. $P(U > 33) = P(Z > -\frac{3}{\frac{\sqrt{29}}{3}}) = \Pi(+1,671) = 0,9526$. Ce risque est très important. L'entrepreneur a 95,26% de chances de dépasser ce délai.

4. Le risque est représenté par la perte d'argent lors de la réalisation. Il ne faut pas que le chef d'entreprise verse un nombre de pénalités tel que le bénéfice soit absorbé. Le nombre de pénalités à ne pas dépasser est $\frac{200000}{40000} = 5$ jours. La durée du chantier ne doit donc pas dépasser $33 + 5 = 38$ jours.
   $P(U > 38) = P(Z > \frac{2}{\frac{\sqrt{29}}{3}}) = P(\frac{6}{\sqrt{29}}) = P(Z > 1,114) = 1 - \Pi(1,114) = 1 - 0,8673 = 0,1327$. Le risque que l'entrepreneur perde de l'argent sur le chantier est de 13,27%, risque non négligeable.

$\boxed{\textbf{Exercice 5}}$ *Correction* :

1. Une estimation de μ est $\hat{\mu} = \overline{x} = \dfrac{\sum_{i=1}^{12} x_i}{12} = 13,31$. Puisque σ est inconnue alors l'intervalle de confiance à 95% s'écrit comme $\left[x \pm \dfrac{\sigma}{\sqrt{12}} t_{11}(0,975)\right]$ avec $\sigma^2 = \dfrac{1}{11}\displaystyle\sum_{i=1}^{11}(x_i - \overline{x})^2 = (1,92)^2$ et $t_{11}(0,975) = 2,201$ est le quantile de la loi de student à 11 degrés correspondant à $0,975$. Donc l'intervalle de confiance à 95% pour μ est

$$\left[\overline{x} - \frac{\sigma}{\sqrt{12}} t_{11}(0,975), \frac{\sigma}{\sqrt{12}} t_{11}(0,975)\right] = [12,094; 14,534]$$

2. Si $\sigma^2 = 3,72$ est connue alors l'intervalle de confiance devient $\left[\overline{x} \pm \dfrac{\sigma}{\sqrt{12}} z(0,975)\right]$ avec $z(0,975) = 1,96$ est le quantile de la loi normale $\mathcal{N}(0,1)$ correspondant à $0,975$. Donc l'intervalle de confiance à 95% pour μ devient $[12,223; 14,405]$. Cet intervalle est plus petit, et donc plus précis.

$\boxed{\textbf{Exercice 6}}$ *Correction* :

1. Les variables sont i.i.d. (indépendantes et identiquement distribuées, c'est-à-dire qui ont toutes la même loi de probabilité) de même loi que X où X suit une loi de Bernoulli $\mathcal{B}(p)$ à chaque individu sondé une variable aléatoire X_i pour $i = 1, \ldots, N$ telle que

$$X_i = \begin{cases} 1, & p(\{X_i = 1\}) = p \\ 0, & p(\{X_i = 0\}) = 1 - p \end{cases}$$

ou de façon générale $p(\{X = x\}) = p^x(1-p)^{1-x}$.

2. On montre que $E(\hat{p}) = \dfrac{1}{N} E\left(\displaystyle\sum_{i=1}^{N} X_i\right) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} E(X_i)$. Puisque les X_i sont i.i.d. de même loi que X, $E(\hat{p}) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} E(X) = E(X) = p$ donc l'estimateur est sans biais. De plus on sait que $V(\hat{p}) = \dfrac{1}{N^2} V\left(\displaystyle\sum_{i=1}^{N} X_i\right)$. Les variables X_i étant indépendantes et de même loi que X, on a donc :

$$V(\hat{p}) = \frac{1}{N^2}\left(\sum_{i=1}^{N} V(X_i)\right) = \frac{1}{N^2}\left(\sum_{i=1}^{N} V(X)\right) = \frac{V(X)}{N} = \frac{p(1-p)}{N}.$$

Ainsi on a $E(\hat{p}) = p$ et $\displaystyle\lim_{N\to\infty} V(\hat{p}) = \lim_{N\to\infty} \frac{p(1-p)}{N} = 0$. On peut donc conclure que l'estimateur est convergent : $\hat{p} \xrightarrow{p} p$ quand $N \to \infty$.

3. Déterminons la quantité d'information de Fisher associée à l'échantillon $\{X_1, \ldots, X_N\}$:

$$I_N(p) = -E\left[\frac{\partial^2 \text{log} L(X_1, \ldots, X_N, p)}{\partial p^2}\right] = E\left[\frac{\sum X_i}{p^2} + \frac{(N - \sum X_i)}{(1-p)^2}\right] = \frac{1}{p^2}\sum E(X_i) + \frac{N - \sum E(X_i)}{(1-p)^2} = \frac{Np}{p^2} + \frac{N - Np}{(1-p)^2} = \frac{N}{p(1-p)}.$$

Par conséquent, on a $V(\hat{p}) = \dfrac{1}{I_N(p)}$ et l'estimateur est efficace au sens de la borne FDCR (Fréchet-Darmois-Cramér-Rao).

4. On sait que $\sqrt{N}(\hat{p} - p) \underset{N\to\infty}{\xrightarrow{d}} \mathcal{N}\left(0, \dfrac{1}{I_n(p)}\right)$ donc $\sqrt{N}(\hat{p} - p) \underset{N\to\infty}{\xrightarrow{d}} \mathcal{N}\left(0, \dfrac{p(1-p)}{N}\right)$.
Remarque : on retrouverait le même résultat par approximation de la loi binomiale.

5. Estimation ponctuelle : $\hat{p} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} X_i = \dfrac{98}{200} = 0,49$.

6. Si $N > 30$, on peut approximer la loi de la fréquence empirique $\hat{p} = \sum_{i=1}^{N} \frac{X_i}{N}$ par une loi normale : $\hat{p} \rightsquigarrow \mathcal{N}\left(0, \frac{p(1-p)}{N}\right)$. On en déduit l'intervalle de confiance au seuil de $\alpha\%$ suivant la formule :

$$p\left(\hat{p} - \phi^{-1}(1-\frac{\alpha}{2})\sqrt{\frac{p(1-p)}{N}}; \hat{p} + \phi^{-1}(1-\frac{\alpha}{2})\sqrt{\frac{p(1-p)}{N}}\right) = 1-\alpha$$

où ϕ désigne la fonction de répartition de la loi normale centrée réduite. On a donc

$$\alpha = 5\%,\ IC_5 = [0,42; 0,56]$$
$$\alpha = 2\%,\ IC_2 = [0,41; 0,57]$$

Le deuxième intervalle est moins précis puisque le niveau de risque est plus élevé.

7. Le candidat est élu (en prévision) si la probabilité que l'on vote pour lui est supérieure à $0,5$. On cherche N tel que :

$$p(\hat{p} > 0,5) = 0,95 \Leftrightarrow p(\hat{p} \leq 0,5) = 0,05 \Leftrightarrow p\left(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{N}}} \leq \frac{0,5-p}{\sqrt{\frac{p(1-p)}{N}}}\right) = 0,05 \Leftrightarrow \phi\left(\frac{0,5-p}{\sqrt{\frac{p(1-p)}{N}}}\right) = 0,05 \Leftrightarrow \frac{0,5-p}{\sqrt{\frac{p(1-p)}{N}}} = \phi^{-1}(0,05).$$

Ainsi, pour $p = 0,52$, $N = 1689$.

Exercice 7 *Correction* :

1. Échantillon 1 : $\hat{p}_{obs,1} = \frac{21}{500} = 0,042 = 4,2\%$.
Échantillon 2 : $\hat{p}_{obs,2} = \frac{57}{1000} = 0,057 = 5,7\%$.

2. Échantillon 1 à 95% : $1,96 \times \sqrt{\frac{0,042 \times 0,958}{500}} = 0,018$
Échantillon 1 à 99% : $2,58 \times \sqrt{\frac{0,042 \times 0,958}{500}} = 0,023$
Échantillon 2 à 95% : $1,96 \times \sqrt{\frac{0,057 \times 0,943}{1000}} = 0,014$
Échantillon 2 à 99% : $2,58 \times \sqrt{\frac{0,057 \times 0,943}{1000}} = 0,019$

3. Lorsque la taille de l'échantillon augmente, la largeur diminue.
Lorsque le niveau de confiance augmente, la largeur augmente.

4. On a $IC_1 = [0,024; 0,06] = [2,4\%; 6\%]$ et $IC_2 = [0,043; 0,071] = [4,3\%; 7,1\%]$. Les deux intervalles se superposent, la probabilité que leur union recouvre la proportion réelle est supérieure à 95%.

5. Échantillon fusionné : $\hat{p}_{obs} = \frac{78}{1500} = 0,052 = 5,2\%$.

6. Échantillon fusionné à 95% : $1,96 \times \sqrt{\frac{0,052 \times 0,948}{1500}} = 0,011$
Échantillon fusionné à 99% : $2,58 \times \sqrt{\frac{0,052 \times 0,948}{1500}} = 0,015$.

7. En fusionnant les deux échantillons, on obtient un échantillon de taille plus important, permettant dès lors une estimation plus précise (un intervalle de confiance plus étroit).

8. La demi-largeur de l'IC au niveau de confiance de 95% est égale à $1,96 \times \sqrt{\dfrac{\hat{p}_{obs} \times (1-\hat{p}_{obs})}{n}} \leq 0,005 \Leftrightarrow$ $1,96 \times \sqrt{\dfrac{0,052 \times 0,948}{n}} \leq 0,005 \Leftrightarrow 1,96^2 \times 0,052 \times 0,948 \leq 0,005^2 \times n$ $\Leftrightarrow n \geq \dfrac{1,96^2 \times 0,052 \times 0,948}{0,005^2} = 7575,021$. L'échantillon devra compter au minimum 7576 sujets soit environ 7600 sujets.

9. La demi-largeur de l'IC au niveau de confiance de 99% est égale à $2,58 \times \sqrt{\dfrac{\hat{p}_{obs} \times (1-\hat{p}_{obs})}{n}} \leq 0,005 \Leftrightarrow$ $2,58 \times \sqrt{\dfrac{0,052 \times 0,948}{n}} \leq 0,005 \Leftrightarrow 2,58^2 \times 0,052 \times 0,948 \leq 0,005^2 \times n$ $\Leftrightarrow n \geq \dfrac{2,58^2 \times 0,052 \times 0,948}{0,005^2} = 13125,356$. L'échantillon devra compter au minimum 13126 sujets soit environ 13200 sujets.

Exercice 8 *Correction* : On se propose d'estimer $E(X)$ par un intervalle de confiance au niveau 98%. Ici la loi des v.a. X_i est inconnue. Il est clair pour des raisons physiques que les X_i sont des v.a. bornées, donc à fortiori de carré intégrable. Un estimateur naturel de $E(X)$ est la moyenne empirique X en raison de la loi forte des grands nombres. On va chercher pour $E(X)$ un intervalle de confiance centré sur $\overline{X}(\omega) = 35200/400 = 88$ km/h. Pour construire cet intervalle de confiance, on ne peut pas utiliser ici le théorème limite central classique car l'écart-type σ des x_i est inconnu. On va utiliser le TLC avec autonormalisation où l'on remplace σ par S, la racine carrée de la variance empirique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2.$$

Les X_i étant de carré intégrable, le TLC avec autonormalisation nous dit que

$$\sqrt{n}\frac{\overline{X} - E(X)}{S} \text{ converge en loi vers } Z \text{ où } Z \rightsquigarrow \mathcal{N}(0,1).$$

Cette convergence est légitime pour n grand donc l'approximation

$$p\left(\left|\sqrt{n}\frac{\overline{X} - E(X)}{S}\right| \leq t\right) \simeq p(|Z| \leq t) = 2\phi(t) - 1$$

où ϕ est le fonction de répartition de la loi $\mathcal{N}(0,1)$. La résolution d'inégalité :

$$\left|\sqrt{n}\frac{\overline{X} - E(X)}{S}\right| \leq t \Leftrightarrow \overline{X} - \frac{tS}{\sqrt{n}} \leq E(X) \leq \overline{X} + \frac{tS}{\sqrt{n}}$$

nous donne alors un intervalle de confiance pour $E(X)$ au niveau $2\phi(t) - 1$. Il s'agit bien d'un intervalle de confiance puisque les bornes $\overline{X} \pm tSn^{-1/2}$ sont calculables à partir des observations sans connaissance de la loi des X_i. Pour terminer les calculs, on détermine t en résolvant $2\phi(t) - 1 = 0,98$, ce qui équivaut à $\phi(t) = 0,99$, d'où $t = 2,33$ (approximativement). On calcule ensuite $\overline{X}(\omega)$ et $S(\omega)$:

$$\overline{X}(\omega) = \overline{x} = \frac{1}{400}\sum_{i=1}^{400} x_i = \frac{35200}{400} = 99 \text{ km/h}.$$

$$S^2(\omega) = s^2 = \frac{1}{400}\sum_{i=1}^{400} x_i^2 - (\overline{x})^2 = \frac{3107600}{400} - 88^2 = 7769 - 7744 = 25 \text{ km/h}^2$$

d'où $S(\omega) = s = 5$ km/h. Un intervalle de confiance I au niveau 98% pour $E(X)$ en km/h est donc

$$I = \left[88 - \frac{2,33 \times 5}{20}; 88 + \frac{2,33 \times 5}{20}\right] = [87,41; 88,59].$$