MAT2777 – Exercices et questions à choix multiples I

(avec réponses)

Q1. Deux événements indépendants ont chacun une probabilité de 0.2 de se produire. La probabilité qu'aucun ne se produise est :

a) 0.64 b) 0.04 c) 0.2 d) 0.4 e) N/A

Solution : puisque A et B sont des événements, alors

$$P(\text{ni l'un, ni l'autre}) = P((A \cup B)^c) = P(A^c \cap B^c).$$

Puisque A et B sont indépendants, A^c et B^c le sont aussi (pourquoi ?). Ainsi,

$$P(\text{ni l'un, ni l'autre}) = P(A^c)P(B^c) = (1 - P(A))(1 - P(B)) = (1 - 0.2)(1 - 0.2) = 0.64.$$

Q2. Deux événements mutuellement exclusifs ont chacun une probabilité de 0.2. La probabilité qu'aucun ne se produise est :

a) 0.36 b) 0.04 c) 0.2 d) 0.6 e) N/A

Solution : puisque A et B sont des événements, alors

$$P(\text{ni l'un, ni l'autre}) = P((A \cup B)^c) = 1 - P(A \cup B).$$

Mais $P(A \cup B) = P(A) + P(B)$ puisque A et B sont mutuellement exclusifs, d'où

$$P(\text{ni l'un, ni l'autre}) = 1 - P(A \cup B) = 1 - P(A) - P(B) = 1 - 0.2 - 0.2 = 0.6.$$

Q3. Un système de détection de fumée se compose de deux éléments A et B. L'élément A détecte la fumée avec une probabilité de 0.95, l'élément B avec une probabilité de 0.98, et les deux éléments avec une probabilité de 0.94. Quelle est la probabilité que la fumée ne soit pas détectée ?

a) 0.01 b) 0.99 c) 0.04 d) 0.96 e) N/A

Solution : Soit A l'événement où l'élément A détecte la fumée, et de même pour B. Nous avons $P(A) = 0.95$, $P(B) = 0.98$, et $P(A \cap B) = 0.94$. Alors

$$\begin{aligned} P(\text{fumée non détectée}) &= 1 - P(\text{fumée détectée}) \\ &= 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) \\ &= 1 - (0.95 + 0.98 - 0.94) = 0.01. \end{aligned}$$

Q4. Trois joueurs de football tentent un placement. Soient A_1, A_2, A_3 les événements que le placement est fait par les joueura $1, 2, 3$, respectivement. Supposons que A_1, A_2, A_3 sont indépendants et que $P(A_1) = 0.5$, $P(A_2) = 0.7$, $P(A_3) = 0.6$. Calculez la probabilité qu'un seul joueur réussisse son placement.

a) 0.29 b) 0.21 c) 0.71 d) 0.79 e) N/A

Solution : nous avons

$$
\begin{aligned}
P(\text{seul le 1er réussit}) &= P(A_1 \cap A_2^c \cap A_3^c) = P(A_1)P(A_2^c)P(A_3^c) = 0.5 \times 0.3 \times 0.4 = 0.06 \\
P(\text{seul le 2e réussit}) &= P(A_1^c \cap A_2 \cap A_3^c) = P(A_1^c)P(A_2)P(A_3^c) = 0.5 \times 0.7 \times 0.4 = 0.14 \\
P(\text{seul le 3e réussit}) &= P(A_1^c \cap A_2^c \cap A_3) = P(A_1^c)P(A_2^c)P(A_3) = 0.5 \times 0.3 \times 0.6 = 0.09
\end{aligned}
$$

Ces 3 événements sont mutuellement exclusifs, d'où

$$P(\text{un seul succès}) = P(1) \cup P(2) \cup P(3) = P(1) + P(2) + P(3) = 0.29.$$

Q5. Dans un groupe de 16 candidats à des postes de recherche en laboratoire, 7 sont des chimistes et 9 des physiciens. De combien de façons peut-on choisir un groupe de 5 candidats contenant 2 chimistes et 3 physiciens ?

Solution : il s'agit d'une procédure en deux étapes. Il existe $\binom{7}{2}$ façons de sélectionner 2 chimistes parmi les 7 (la première étape), et $\binom{9}{3}$ façons de sélectionner 3 physiciens parmi les 9 (la deuxième étape).

Ainsi, il y a

$$\binom{7}{2}\binom{9}{3} = \frac{7!}{5!2!} \times \frac{9!}{6!3!} = \frac{7 \cdot 6}{2} \times \frac{9 \cdot 8 \cdot 7}{3 \cdot 2} = 21 \times 84 = 1764$$

façons de sélectionner un groupe de candidats avec les contraintes requises.

Q6. Un théorème de l'analyse combinatoire stipule que le nombre de permutations de n objets dont n_1 sont de même nature 1, n_2 sont de même nature 2, ..., et n_r sont de même nature r (c'est-à-dire, $n = n_1 + n_2 + \cdots + n_r$) est

$$\frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_r!}.$$

Trouvez le nombre de mots différents qui peuvent être formés en réarrangeant les lettres dans les mots suivants (incluez le mot donné dans le compte) :

a) NORMAL b) HHTTTT c) ILLINI d) MISSISSIPPI

Solution :

- NORMAL: chaque lettre est différente, alors $\frac{6!}{1!1!1!1!1!1!} = 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$ (NORMAL, NROMAL, NOMRAL, etc.)
- HHTTTT: 2×H et 4×T, d'où $\frac{6!}{2!4!} = \frac{6 \cdot 5}{2} = 15$ (HHTTTT, HTHTTT, HTTHTT, HTTTHT, etc.)
- ILLINI: 3×I, 2×L, et 1×N, d'où $\frac{6!}{3!2!1!} = \frac{6 \cdot 5 \cdot 4}{2} = 60$ (ILLINI, ILILNI, ILLNII, etc.)
- MISSISSIPPI: 4×I, 1×M, 4×S, et 2×P, d'où $\frac{11!}{4!1!4!2!} = \frac{39{,}916{,}800}{1152} = 34,650.$

Q7. Une classe se compose de 49 élèves ingénieurs et de 51 élèves en sciences. Les étudiants sont répartis en fonction de leurs notes :

	Réussi	Échoué
Ing.	43	6
Sci.	41	10

Si l'on choisit une personne au hasard, quelle est la probabilité qu'elle échoue s'il s'agit d'un étudiant en ingénierie ?

a) 0.06 b) 0.12 c) 0.41 d) 0.81 e) N/A

Solution : Il y a 100 étudiants en tout. Soient A et I représentant les événements que l'étudiant a réussi et que l'étudiant est un ingénieur, respectivement. Alors

$$P(A^c|I) = \frac{P(A^c \cap I)}{P(I)} = \frac{6/10}{49/100} = 0.12.$$

Q8. Une entreprise qui produit un médicament particulier possède deux usines, A et B. 70% des médicaments sont fabriqués dans l'usine A, 30% dans l'usine B. Supposons que 95% des médicaments produits par l'usine A répondent aux spécifications tandis que seulement 75% le font pour l'usine B. Si j'achète une dose du médicament de l'entreprise, quelle est la probabilité qu'elle réponde aux spécifications ?

a) 0.89 b) 0.95 c) 0.75 d) 0.7 e) N/A

Solution : Soient M les événements que le médicament répond aux spécifications, A qu'il est produit par l'usine A, B qu'il est produit par l'usine B.

Nous avons $P(A) = 0.7$, $P(B) = 0.3$, $P(M|A) = 0.95$, et $P(M|B) = 0.75$. Selon la loi de la probabilité totale, nous avons

$$P(M) = P(M|A)P(A) + P(M|B)P(B) = 0.95 \cdot 0.7 + 0.75 \cdot 0.3 = 0.89.$$

Q9. Une équipe de recherche médicale souhaite évaluer un test de dépistage proposé pour la maladie d'Alzheimer. Le test a été administré à un échantillon aléatoire de 450 patients atteints de la maladie d'Alzheimer ; dans 436 des cas, le résultat du test était positif. Le test a également été administré à un échantillon aléatoire de 500 patients non atteints de la maladie ; le résultat n'a été positif que dans 5 des cas. On sait qu'au Canada, 11.3% de la population âgée de 65 ans et plus est atteint d'Alzheimer. Trouvez la probabilité qu'une personne soit atteinte étant donné que son test était positif (choisissez la réponse la plus proche).

a) 0.97 b) 0.93 c) 0.99 d) 0.07 e) N/A

Solution : Soient A et D les événements que le test est positif et que la personne soit atteinte, respectivement. D'après l'énoncé du problème, on sait que

$$P(A|D) = \frac{436}{450}, P(A|D^c) = \frac{5}{500}, P(D) = 0.113.$$

Selon le théorème de Bayes,

$$P(D|A) = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} = \frac{436/450 \cdot 0.113}{436/450 \cdot 0.113 + 5/500 \cdot (1 - 0.113)} = 0.925.$$

Q10. Douze articles sont échantillonnés indépendamment sur une chaîne de production. Si la probabilité qu'un article donné soit défectueux est de 0.1, la probabilité qu'il y ait au plus deux articles défectueux dans l'échantillon est ...

a) 0.38748 b) 0.9872 c) 0.7361 d) 0.8891 e) N/A

Solution :soit $p = 0.1$ la probabilité qu'un article soit défectueux. Alors $1 - p = 0.9$ est la probabilité qu'un article ne soit pas défectueux. Soit X le nombre d'articles défectueux dans l'échantillon.

La probabilité qu'aucun des articles ne soit défectueux est

$$\begin{aligned} P(X = 0) &= P(\text{article 1 n'est pas défectueux}) \times \cdots \times P(\text{article 12 n'est pas défectueux}) \\ &= (1 - p)^{12} = 0.9^{12} \approx 0.2824 \end{aligned}$$

La probabilité qu'un seul des articles soit défectueux est

$$P(X = 1) = P(\text{seul l'article 1 est défectueux}) + \cdots + P(\text{seul l'article 12 est défectueux}).$$

L'événement selon lequel seul l'élément j est défectueux (pour $j = 1, \cdots, 12$) se produit lorsque l'élément j est défectueux (avec une probabilité p) ET que les 11 éléments restants ne sont pas défectueux (chacun avec une probabilité $1 - p$). Puisque les articles sont échantillonnés indépendamment,

$$P(\text{seul l'article } j \text{ est défectueux}) = p(1 - p)^{11} \approx 0.0314.$$

Mais il existe $\frac{12}{1} = \frac{12!}{1!11!} = 12$ façons de choisir lequel des 12 articles sera défectueux (échantillonnage sans remplacement), donc

$$P(X = 1) = \underbrace{p(1 - p)^{11} + \cdots + p(1 - p)^{11}}_{\binom{12}{1} = 12 \text{ fois}} = \binom{12}{1} p(1 - p)^{11} \approx 12(0.0314) \approx 0.3766.$$

La probabilité qu'exactement deux des articles soient défectueux est

$$\begin{aligned} P(X = 2) = P(&\text{seulement les articles } 1, 2 \text{ sont défectueux}) + \cdots \\ &+ P(\text{seulement les articles } 11, 12 \text{ sont défectueux}). \end{aligned}$$

L'événement selon lequel seuls les articles j, k sont défectueux (pour $j, k = 1, \ldots, 12$, $j \neq k$) se produit lorsque les articles $j \neq k$ sont défectueux (chacun avec une probabilité p) ET les 10 articles restants ne sont pas défectueux (chacun avec une probabilité $1 - p$). Puisque les éléments sont échantillonnés indépendamment,

$$P(\text{seulement les éléments } j, k \text{ sont défectueux}) = p^2(1 - p)^{10} \approx 0.0031.$$

Mais il existe $\frac{12}{2} = \frac{12!}{2!10!} = 66$ façons de choisir lesquels des 2 éléments seront défectueux, donc

$$P(X = 2) = \underbrace{p^2(1 - p)^{10} + \cdots + p^2(1 - p)^{10}}_{\binom{12}{2} = 66 \text{ fois}} = \binom{12}{2} p^2(1 - p)^{10} \approx 66(0.0031) = 0.2301.$$

La probabilité qu'il y ait au plus deux articles défectueux dans l'échantillon est donc

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= \binom{12}{1} p^0 (1-p)^{12} + \binom{12}{2} p^1 (1-p)^{11} + \binom{12}{2} p^2 (1-p)^{10} \\ &\approx 0.2824 + 0.3766 + 0.2071 = 0.8891 \end{aligned}$$

Q11. Une étudiante peut résoudre 6 problèmes à partir d'une liste de 10. Pour un examen, 8 questions sont choisies au hasard de la liste. Quelle est la probabilité que l'étudiante résolve exactement 5 problèmes ?

a) 0.98 b) 0.02 c) 0.28 d) 0.53 e) N/A

Solution : Soient X et $8-X$ le nombre de questions de l'examen que l'étudiant peut et ne peut pas résoudre, respectivement.

Il existe $\binom{6}{X}$ façons de choisir au hasard X questions d'examen que l'étudiant peut résoudre, et $\binom{10-6}{8-X} = \binom{4}{8-X}$ façons de choisir au hasard $8-X$ questions que l'étudiant ne peut pas résoudre : il existe donc $\binom{6}{X}\binom{4}{8-X}$ façons de choisir au hasard X questions que l'étudiant peut résoudre ET $8-X$ questions que l'étudiant ne peut pas résoudre.

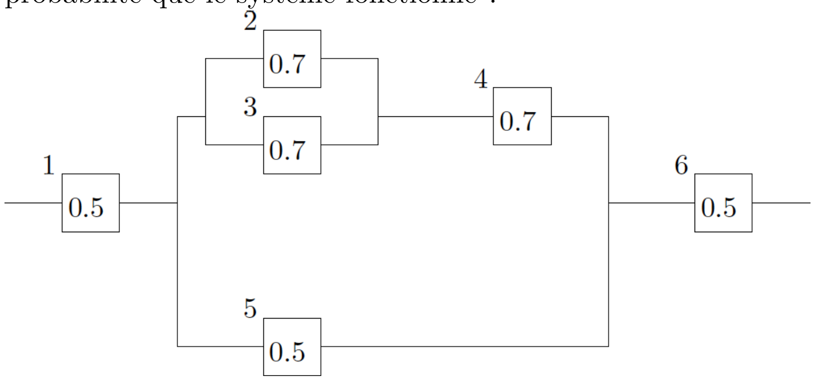
Puisqu'il existe au total $\binom{10}{8}$ possibilités de choisir au hasard 8 questions d'examen parmi les 10 problèmes,

$$P(X) = \frac{\binom{6}{X}\binom{4}{8-X}}{\binom{10}{8}}.$$

Pour $X = 5$,

$$P(X=5) = \frac{\binom{6}{5}\binom{4}{3}}{\binom{10}{8}} = \frac{6 \cdot 4}{45} \approx 0.53.$$

Q12. Considérons le système suivant à six composantes. On dit qu'il est fonctionnel s'il existe un chemin de composantes fonctionnelles de gauche à droite. La probabilité que chaque composante fonctionne est indiquée. Supposons que les composantes fonctionnent ou échouent indépendamment. Quelle est la probabilité que le système fonctionne ?



a) 0.1815 b) 0.8185 c) 0.6370 d) 0.2046 e) N/A

Solution : soient la boîte A constituée des composantes $2,3,4,5$, la boîte B constituée des composantes $2,3,4$, et la boîte C constituée des composantes $2,3$. Nous désignerons la probabilité que la boîte j fonctionne par $P(j)$, $j \in \{A,B,C\}$.

Nous sommes intéressés par la probabilité $P(S)$ que le système fonctionne. Puisque les composantes fonctionnent ou échouent indépendamment,

$$\begin{aligned}
P(S) &= P(\text{la composante 1 et la boîte } A \text{ et la composante 6 fonctionnent}) \\
&= P(1 \text{ fonctionne}) \times P(A) \times P(6 \text{ fonctionne}) \\
&= 0.5 \cdot P(A) \cdot 0.5 = 0.5^2 P(A).
\end{aligned}$$

Il y a deux façons pour la boîte A de fonctionner : soit la composante 5 fonctionne (avec une probabilité de 0.5), soit la boîte B fonctionne :

$$\begin{aligned}
P(A) &= P(\text{composante 5 fonctionne ou boîte } B \text{ fonctionne}) \\
&= P(5 \text{ fonctionne}) + P(B) - P(\text{composante 5 fonctionne et boîte } B \text{ fonctionne}) \\
&= P(5 \text{ fonctionne}) + P(B) - P(5 \text{ fonctionne})P(B) \\
&= 0.5 + P(B) - 0.5P(B) = 0.5(1 + P(B)).
\end{aligned}$$

Ainsi,

$$P(S) = 0.5^2 P(A) = 0.5^2 \cdot 0.5(1 + P(B) = 0.5^3(1 + P(B)).$$

Pour que la boîte B fonctionne, nous avons besoin que la boîte C et la composante 4 fonctionnent tous les deux, de sorte que

$$\begin{aligned}
P(B) &= P(\text{boîte } C \text{ fonctionne et la composante 4 fonctionne}) \\
&= P(4 \text{ fonctionne})P(C) = 0.7P(C).
\end{aligned}$$

Ainsi,

$$P(S) = 0.5^3(1 + P(B)) = 0.5^3(1 + 0.7P(C)).$$

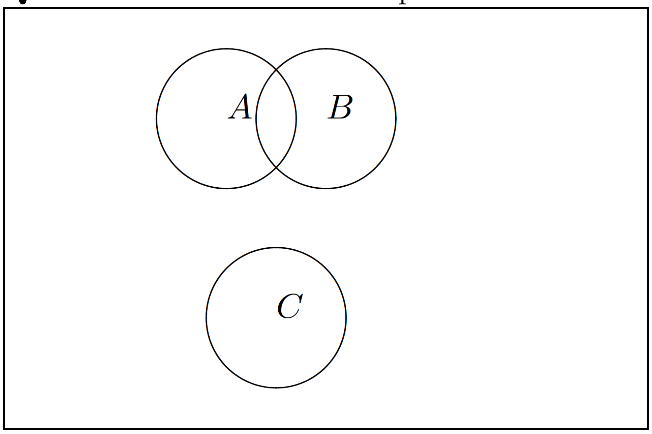
Enfin, il y a deux façons pour la boîte C de fonctionner : soit la composante 2 fonctionne (avec une probabilité de 0.7), soit la composante 3 fonctionne (également avec une probabilité de 0.7) :

$$\begin{aligned}
P(C) &= P(\text{composante 2 fonctionne ou composante 3 fonctionne}) \\
&= P(2 \text{ fonctionne}) + P(3 \text{ fonctionne}) - P(\text{composante 2 fonctionne et composante 3 fonctionne}) \\
&= P(2 \text{ fonctionne}) + P(3 \text{ fonctionne}) - P(2 \text{ fonctionne})P(3 \text{ fonctionne}) \\
&= 0.7 + 0.7 - 0.7 \cdot 0.7 = 0.7(2 - 0.7) = 0.7 \cdot 1.3.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
P(S) &= 0.5^3(1 + 0.7P(C)) = 0.5^3(1 + 0.7 \cdot 0.7 \cdot 1.3) \\
&= 0.5^3(1 + 0.7^2 \cdot 1.3) = 0.24046.
\end{aligned}$$

Q13. Trois événements sont représentés dans le diagramme de Venn ci-dessous.



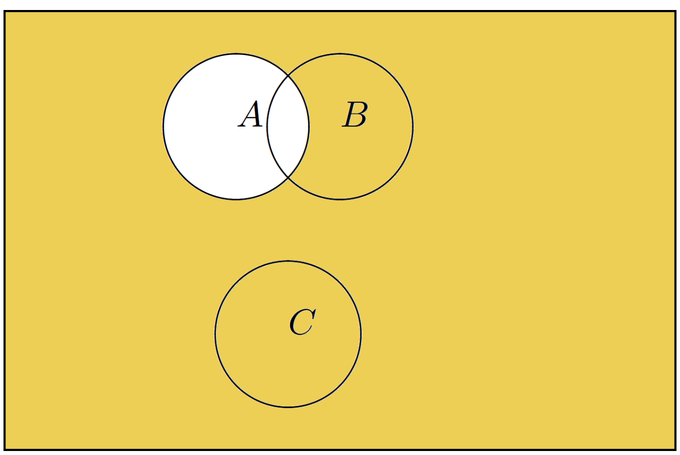
Ombragez la région correspondant aux événements suivants :

a) A^c b) $(A \cap B) \cup (A \cap B^c)$ c) $(A \cap B) \cup C$ d) $(B \cup C)^c$

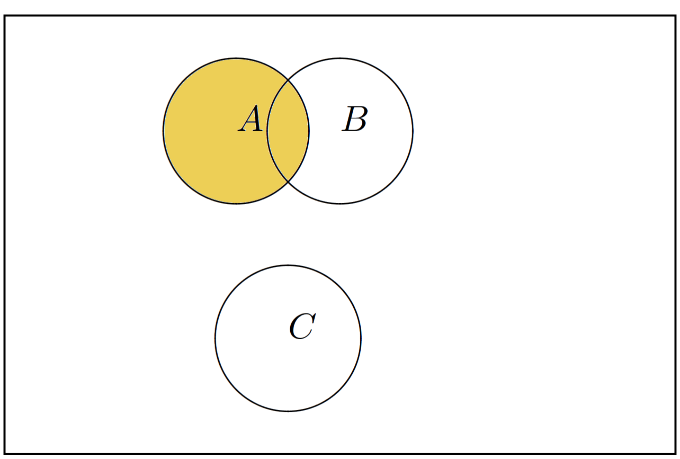
e) $(A \cap B)^c \cup C$

Solution :

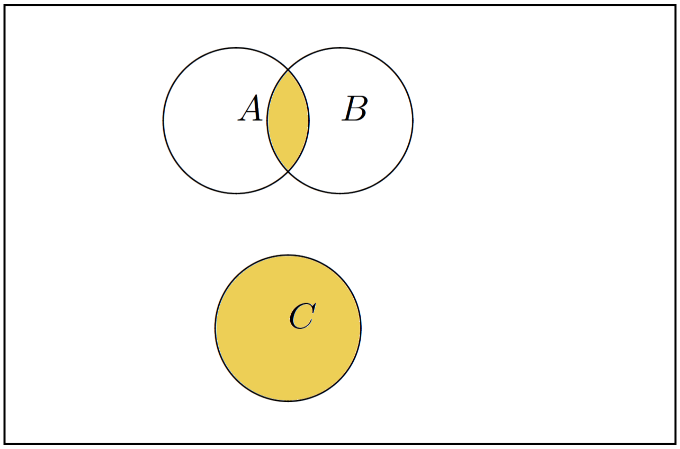
a)



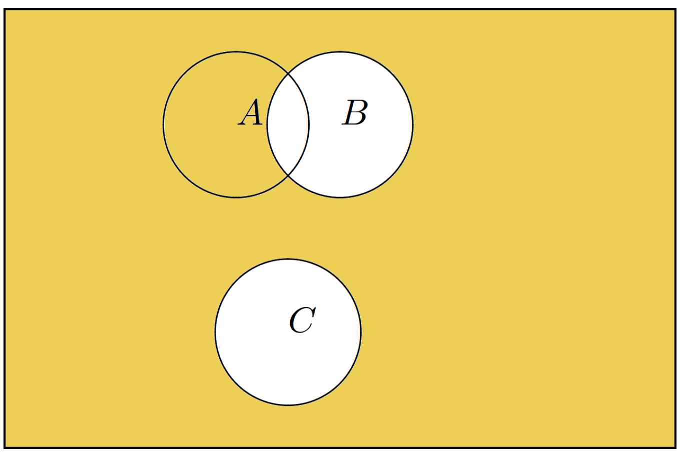
b)



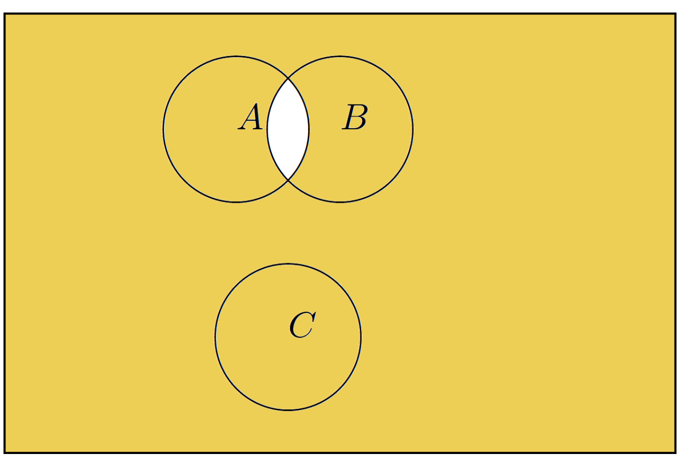
c)



d)



e)



Q14. On classe des pièces d'aluminium en fonction de la finition de la surface et celle du bord. Les résultats de 85 échantillons se résument comme suit :

	Bord	
Surface	excellente	bonne
excellente	60	5
bonne	16	4

Soient A, B l'événement qu'une pièce sélectionnée ait une finition de surface "excellente" et de bord "excellente", respectivement. Si les échantillons sont choisis au hasard, déterminez les probabilités suivantes :

a) $P(A)$ b) $P(B)$ c) $P(A^c)$ d) $P(A \cap B)$ e) $P(A \cup B)$ f) $P(A^c \cup B)$

g) Si la pièce a une excellente finition des bords, quelle est la probabilité qu'elle a une excellente surface ?

h) Si la pièce n'a pas une excellente finition de surface, quelle est la probabilité qu'elle a d'excellents bords ?

i) Est-ce que A et B sont indépendants ?

Solution :

a) $P(A) = 65/85$
b) $P(B) = 76/85$
c) $P(A^c) = 1 - 65/85 = 20/85$
d) $P(A \cap B) = 60/85$
e) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 65/85 + 76/85 - 60/85 = 81/85$
f) $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = 20/85 + 76/85 - 16/85 = 80/85$
g) $P(A|B) = P(A \cap B)/P(B) = 60/76$
h) $P(B|A^c) = P(B \cap A^c)/P(A^c) = 16/20$
i) $P(A|B) \neq P(A)$ – événements indépendants.

Q15. Si $P(A) = 0.1$, $P(B) = 0.3$, $P(C) = 0.3$, et si les événements A, B, C sont mutuellement exclusifs, déterminez les probabilités suivantes :

a) $P(A \cup B \cup C)$ b) $P(A \cap B \cap C)$ c) $P(A \cap B)$

d) $P((A \cup B) \cap C)$ e) $P(A^c \cap B^c \cap C^c)$ f) $P[(A \cup B \cup C)^c]$

Solution :

a) $P(A \cup B \cup C) = P(A) + P(B) + P(C) = 0.7$, puisque les événements sont mutuellement exclusifs
b) $P(A \cap B \cap C) = 0$
c) $P(A \cap B) = 0$
d) $P((A \cup B) \cap C) = 0$
e) $P(A^c \cap B^c \cap C^c) = P[(A \cup B \cup C)^c] = 1 - P(A \cup B \cup C) = 0.3$ (diagramme de Venn)
f) $P[(A \cup B \cup C)^c] = 1 - P(A \cup B \cup C) = 0.3$

Q16. La probabilité qu'un interrupteur électrique, qui est conservé au sec, tombe en panne pendant la période de garantie, est de 1%. Si l'interrupteur est humide, la probabilité de défaillance est de 8%. Supposons que 90% des interrupteurs sont conservés au sec, tandis que les 10% restants sont conservés dans des conditions humides.

a) Quelle est la probabilité que le commutateur tombe en panne pendant la période de garantie ?

b) Si le commutateur est tombé en panne pendant la période de garantie, quelle est la probabilité qu'il ait été conservé dans des conditions humides ?

Solution : Soient F, H, D représentant les événement que le commutateur tombe en panne, qu'il soit humide, et qu'il soit sec, respectivement. Notez que H et D sont mutuellement exclusifs et exhaustifs. Nous avons $P(F|D) = 0.01$, $P(F|H) = 0.08$, $P(D) = 0.9$, $P(H) = 0.1$.

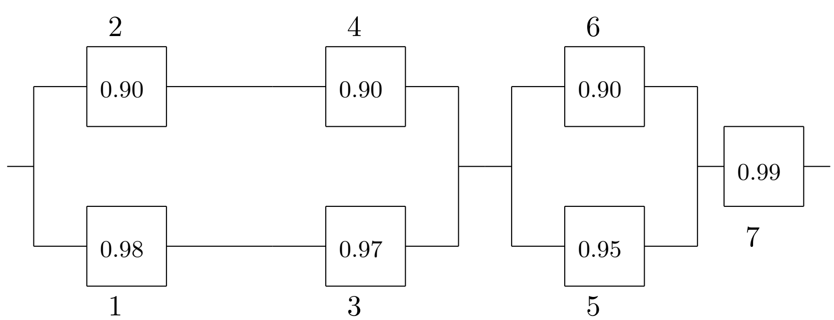
a) Selon la loi de la probabilité totale,

$$P(F) = P(F|D)P(D) + P(F|H)P(H) = 0.01 \cdot 0.9 + 0.08 \cdot 0.1 = 0.009 + 0.008 = 0.017.$$

b) Selon le théorème de Bayes,

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{P(F|H)P(H)}{P(F)} = 0.4706.$$

Q17. Le système suivant ne fonctionne que s'il y a un chemin de dispositif fonctionnel de gauche à droite. La probabilité que chaque dispositif fonctionne est telle qu'indiquée. Quelle est la probabilité que le circuit fonctionne ? Supposez l'indépendance des dispositifs.



Solution : Soient la boîte A constituée des composantes $1, 2, 3, 4$; la boîte B, des composantes $5, 6$; la boîte C de la composante 7. Puisque toutes les composantes sont indépendantes,

$$P(\text{système fonctionne}) = P(A \text{ fonctionne})P(B \text{ fonctionne})P(7 \text{ fonctionne}).$$

B n'est qu'un système parallèle, de sorte qu'il fonctionne si l'une de ses deux composantes fonctionne :

$$P(B \text{ fonctionne}) = 0.9 + 0.95 - 0.9 \cdot 0.95 = 1,85 - 0.855 = 0.995.$$

De plus, puisque toutes les composantes sont indépendantes les unes des autres,

$$P(2 \text{ et } 4 \text{ fonctionnent}) = P(2 \text{ fonctionne})P(4 \text{ fonctionne}) = 0.9 \cdot 0.9 = 0.81$$
$$P(1 \text{ et } 3 \text{ fonctionnent}) = P(1 \text{ fonctionne})P(3 \text{ fonctionne}) = 0.98 \cdot 0.97 = 0.9506.$$

A est un système parallèle composé de deux sous-systèmes indépendants : $2, 4$ et $1, 3$, de sorte que

$$\begin{aligned} P(A \text{ fonctionne}) &= P(2,4 \text{ fonctionnent}) + P(1,3 \text{ fonctionnent}) - P(2,4 \text{ fonctionnent})P(1,3 \text{ fonctionnent}) \\ &= 0.81 + 0.9506 - 0.81 \cdot 0.9506 = 0.9906. \end{aligned}$$

Ainsi, $P(\text{système fonctionne}) = 0.9906 \cdot 0.995 \cdot 0.99 = 0.9758.$

Q18. Un inspecteur travaillant pour une entreprise de fabrication a 95% de chance d'identifier correctement les articles défectueux et 2% de chances de classer incorrectement un article conforme comme défectueux. L'entreprise sait que 1% des articles qu'elle produit sont non conformes (défectueux).

1. Quelle est la probabilité qu'un article sélectionné pour l'inspection soit classé comme défectueux ?
2. Si un article sélectionné au hasard est classé comme conforme, quelle est la probabilité qu'il soit effectivement conforme ?

Solution : Soit A l'événement qu'un article est classé comme défectueux et D l'événement qu'un article est défectueux ; de sorte que D^c est l'événement qu'un article est conforme. On sait que $P(D) = 0.01$, $P(A|D) = 0.95$, et $P(A|D^c) = 0.02$.

a) Selon la loi des probabilités totales,

$$P(A) = P(A \cap D) + P(A \cap D^c) = P(A|D)P(D) + P(A|D^c)P(D^c) \approx 0.0293.$$

b) Selon le théorème de Bayes,

$$P(D^c|A^c) = \frac{P(A^c|D^c)P(D^c)}{P(A^c)} = \frac{(1 - P(A|D^c))P(D^c)}{1 - P(A)} \approx 0.999.$$

Q19. Considérons un jeu de carte ordinaire de 52 cartes (4 couleurs, 13 cartes par couleur).

a) Combien y a-t-il de mains de poker de 5 cartes ?

b) Combien y a-t-il de mains de bridge de 13 cartes ?

c) Quelle est la probabilité d'une main de poker de 5 piques ?

d) Quelle est la probabilité d'une "flush" (5 cartes de la même couleur) ?

e) Quelle est la probabilité qu'une main de poker contienne exactement 3 rois et 2 dames ?

f) Quelle est la probabilité qu'une main de poker contienne exactement 2 rois, 2 reines, et 1 valet ?

Solution :

a) Le nombre de mains de poker possibles est

$$_{52}C_5 = \binom{52}{5} = \frac{52!}{5!47!} = 2,598,960.$$

b) Le nombre de mains de bridge possibles est

$$_{52}C_{13} = \binom{52}{13} = \frac{52!}{13!39!} = 635,013,559,600.$$

c) Le nombre de mains possibles de 5 cartes qui sont toutes des piques est $N = \binom{13}{5}\binom{39}{0}$: il y a $\binom{13}{5}$ façons de choisir 5 piques parmi les 13 piques du paquet de cartes, après quoi il y a $\binom{39}{0} = 1$ façcon de choisir zéro carte qui ne sont pas des piques parmi les 39 cartes qui ne sont pas des piques, d'où $N = \binom{13}{5} \cdot 1 = \frac{13!}{5!8!} = 1287$. La probabilité d'obtenir une telle main est donc

$$\frac{1287}{2,598,960} \approx 0.000495.$$

d) Pour n'importe laquelle des 4 couleurs, la probabilité d'obtenir une main de 5 cartes dont toutes les cartes sont de la mm̂eme couleur a été calculée ci-dessus comme étant 0.000495. Ainsi,

$$\begin{aligned} P(\text{“flush”}) &= P(\text{“flush” en piques}) + P(\text{“flush” en coeurs}) + P(\text{“flush” en carreaux}) + P(\text{“flush” en trèfles}) \\ &\approx 4(0.000495) \approx 0.00198. \end{aligned}$$

e) Soit A le résultat selon lequel la main est constituée d'exactement 3 R et 2 D. Nous pouvons choisir les 3 R de l'une de $\binom{4}{3}$ façons et les 2 D de l'une de $\binom{4}{2}$ façons, de sorte que le nombre de telles mains est

$$N_A = \binom{4}{3}\binom{4}{2} = \frac{4!}{1!3!} \cdot \frac{4!}{2!2!} = 4 \cdot (24/4) = 24,$$

et

$$P(A) = \frac{24}{2,598,960} \approx 0.0000092.$$

f) C'est la même idée :

$$P(C) = \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{2,598,960} = \frac{144}{2,598,960} \approx 0.000055.$$

Q20. Dans un bateau, des matelots envoient des messages à la côte en disposant sept drapeaux de couleur sur un mât vertical.

a) S'ils ont 4 drapeaux oranges et 3 drapeaux bleus, combien de messages peuvent-ils envoyer ?

b) S'ils ont 7 drapeaux de couleurs différentes, combien de messages peuvent-ils envoyer ?

c) S'ils ont 3 drapeaux violets, 2 drapeaux rouges et 4 drapeaux jaunes, combien de messages peuvent-ils envoyer ?

Solution :

a) La question est de trouver le nombre de permutations distinguables de 4 O et 3 B, soit

$${}_7C_4 = \binom{7}{4} = \frac{7!}{4!3!} = 35.$$

b) Si les drapeaux étaient de couleurs différentes, nous chercherions le nombre de permutations (distinguables) de 7 objets différents, c'est-à-dire $7! = 5040$.

c) Le nombre de combinaisons distinguables de 3 V, 2 R, et 4 J est

$$\frac{9!}{3!4!2!} = 1260.$$

Q21. Les finales de la Coupe Stanley au hockey ou les finales de la NBA au basket-ball se poursuivent jusqu'à ce que l'équipe représentative de la Conférence Ouest ou de la Conférence Est gagne 4 parties. Combien d'ordres différents sont possibles ($OOEEEE$ signifie que l'équipe de l'Est a gagné en 6 matchs) si la série se déroule comme suit :

a) 4 parties ? b) 5 parties ? c) 6 parties ? d) 7 parties ?

Solution :

a) Il n'y a que 2 ordres : $EEEE$ et $OOOO$.

b) Commençons par les finales remportées par l'équipe de l'Est, c'est-à-dire : 4 victoires en 5 parties remportées par E, avec la 5ème (et dernière) victoire par E. Notez que cela est égal à $\binom{3}{0}+\binom{3}{3}$.

En d'autres termes, seul l'ordre des 4 premières parties compte, et exactement 3 doivent être des victoires par E (pour que la victoire par E de la 5ème partie mette fin à la finale) : il y a $\binom{4}{3}=4$ façons de choisir 3 victoires par E lors des 4 premières parties.

Le même raisonnement s'applique aux finales remportées par l'équipe de l'Ouest. Il y a donc 8 façons de terminer les Finales en 5 parties. Notez que cela est égal à $\binom{4}{1}+\binom{4}{3}$.

c) Un raisonnement similaire montre qu'il y a $\binom{5}{3}=\frac{5\cdot 4}{2}=10$ façons pour l'équipe de l'Est ou de l'Ouest de gagner en 6 matchs, soit 20 façons au total. Notez que cela est égal à $\binom{5}{2}+\binom{5}{3}$.

d) Un raisonnement similaire montre qu'il y a $\binom{6}{3}=\frac{6\cdot 5\cdot 4}{6}=20$ façons pour l'équipe de l'Est ou de l'Ouest de gagner en 7 parties, soit 40 façons au total. Notez que cela est égal à $\binom{6}{3}+\binom{6}{3}$.

Q22. Considérons un jeu de carte ordinaire de 52 cartes (4 couleurs, 13 cartes par couleur), duquel on chosit des cartes au hasard, et sans remise, jusqu'à ce que nous obtenions 3 piques.

a) Quelle est la probabilité qu'il y ait 2 piques dans les 5 premiers tirages ?

b) Quelle est la probabilité qu'un pique soit tiré lors du 6ème tirage étant donné qu'il y avait 2 piques dans les premiers 5 tirages ?

c) Quelle est la probabilité qu'il faille tirer 6 cartes pour obtenir 3 piques ?

d) Toutes les cartes sont remises dans le paquet, qui est mélangé. On tire ensuite 4 cartes. Quelle est la probabilité d'avoir tiré un pique, un cœur, un carreau, et un trèfle, dans cet ordre ?

Solution :

a) Soit A l'événement où il y a 2 piques dans les premières 5 cartes tirées. Puisque les cartes sont tirées sans remise,

$$P(A)=\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}}=\frac{13!}{2!11!}\cdot\frac{39!}{3!36!}\cdot\frac{5!47!}{52!}$$
$$\frac{13\cdot 12}{2}\cdot\frac{39\cdot 38\cdot 37}{6}\cdot\frac{120}{52\cdot 51\cdot 50\cdot 49\cdot 48}=\frac{1,026,492,480}{3,742,502,400}\approx 0.274.$$

b) Soit B l'événement que le 6ème tirage était un pique. Si A s'est déjà produit, il ne reste que 47 cartes pour le 6ème tirage, dont seulement 11 sont des piques. Ainsi $P(B|A)=\frac{11}{47}\approx 0.234$.

c) L'événement $A\cap B$ est l'événement selon lequel il faut tirer 6 cartes pour obtenir 3 piques : d'abord, il doit y avoir 2 piques dans les 5 premiers tirages (A), et une fois que A s'est produit, la 6ième carte tirée est un pique (B) :

$$P(A\cap B)=P(B|A)P(A)\approx 0.234\cdot 0.274\approx 0.064.$$

d) Nous pouvons utiliser à plusieurs reprises la définition des probabilités conditionnelles :

$$\begin{aligned} P(S_1 \cap H_2 \cap D_3 \cap C_4) &= P(S_1)P(H_2|S_1)P(D_3|S_1 \cap H_2)P(C_4|S_1 \cap H_2 \cap D_3) \\ &= \frac{13}{52} \cdot \frac{13}{51} \cdot \frac{13}{50} \cdot \frac{13}{49} \approx 0.0044. \end{aligned}$$

Q23. Un élève a 5 billes bleues et 4 billes rouges dans sa poche gauche, et 4 billes bleues et 5 billes rouges dans sa poche droite. S'il transfère une bille au hasard de sa poche gauche à sa poche droite, quelle est la probabilité qu'il tire ensuite une bille bleue de sa poche droite ?

Solution : soient BG, BD, et RG les évènements où on tire une bille bleue de la poche gauche, une bille bleue de la poche droite, et une bille rouge de la poche gauche. Par la loi de probabilité totale,

$$\begin{aligned} P(BD) &= P(BG \cap BD) + P(RL \cap BD) \\ &= P(BG)P(BD|BG) + P(RG)P(BD|RG) = \frac{5}{9} \cdot \frac{5}{10} + \frac{4}{9} \cdot \frac{4}{10} = \frac{41}{90} \approx 0.456. \end{aligned}$$

Q24. Une compagnie vend des polices d'assurances ; parmi celles-ci, 60% sont pour les voitures, 40% sont pour les maisons, et 20% sont pour les deux. Soient A_1, A_2, A_3, A_4 les personnes ayant respectivement une police d'assurance automobile uniquement, une police d'assurance habitation uniquement, les deux, ou aucune. Soit B l'événement où un assuré renouvelle au moins une des polices auto ou habitation.

a) Calculez $P(A_1)$, $P(A_2)$, $P(A_3)$, et $P(A_4)$.

b) Disons que $P(B|A_1) = 0.6$, $P(B|A_2) = 0.7$, $P(B|A_3) = 0.8$. Si un client choisi au hasard a une police d'assurance automobile ou habitation, quelle est la probabilité qu'il renouvelle l'une de ces polices ?

Solution :

a) $P(A_1) = 0.4$, $P(A_2) = 0.2$, $P(A_3) = 0.2$. Comme les événements sont mutuellement exclusifs et exhaustifs, $P(A_4) = 1 - P(A_1) - P(A_2) - P(A_3) = 0.2$.

b) L'événement $A_1 \cup A_2 \cup A_3$ représente les clients qui ont une police d'assurance automobile, une police d'assurance habitation, ou les deux. Ainsi,

$$P(B|A_1 \cup A_2 \cup A_3) = \frac{P(B \cap (A_1 \cup A_2 \cup A_3))}{P(A_1 \cup A_2 \cup A_3)} = \frac{P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3))}{P(A_1 \cup A_2 \cup A_3)}.$$

Mais A_1, A_2, A_3 sont mutuellement exclusifs, d'où $B \cap A_1, B \cap A_2, B \cap A_3$ le sont aussi :

$$\begin{aligned} P(B|A_1 \cup A_2 \cup A_3) &= \frac{P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)}{P(A_1) + P(A_2) + P(A_3)} \\ &= \frac{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)}{P(A_1) + P(A_2) + P(A_3)} \\ &= \frac{0.6 \cdot 0.4 + 0.7 \cdot 0.2 + 0.8 \cdot 0.2}{0.4 + 0.2 + 0.2} = \frac{0.54}{0.80} = 0.675. \end{aligned}$$

Q25. Une urne contient quatre boules numérotées de 1 à 4. Les boules sont sélectionnées une par une, sans remplacement. Un "match" a lieu si la boule m est la $m$ème boule sélectionnée. Soit l'événement A_i qui désigne un "match" lors du $i$ème tirage, $i = 1, 2, 3, 4$.

a) Calculez $P(A_i)$, $i = 1, 2, 3, 4$.

b) Calculez $P(A_i \cap A_j)$, $i, j = 1, 2, 3, 4$, $i \neq j$.

c) Calculez $P(A_i \cap A_j \cap A_k)$, $i, j, k = 1, 2, 3, 4$, $i \neq j, i \neq k, j \neq k$.

d) Quelle est la probabilité d'une correspondance d'au moins 1 ?

Solution : (ce problème est très difficile!)

a) $P(A_i) = \frac{3!}{4!}$.

b) $P(A_i \cap A_j) = \frac{2!}{4!}$.

c) $P(A_i \cap A_j \cap A_k) = \frac{1!}{4!}$.

d) $P(A_1 \cup A_2 \cup A_3 \cup A_4) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}$.

Q26. La probabilité que l'effectif d'une entreprise ait au moins un accident au cours d'un mois donné est de $(0.01)k$, où k est le nombre de jours du mois. Supposons que le nombre d'accidents soit indépendant d'un mois à l'autre. Si l'année de l'entreprise commence le 1er janvier, quelle est la probabilité que le premier accident se produise en avril ?

Solution : pour tout mois X, soit X l'événement qu'un accident a lieu pendant le mois X. Si les probabilités mensuelles sont indépendantes les unes des autres, alors

$$\begin{aligned} P(J^c \cap F^c \cap M^c \cap A) &= P(J^c)P(F^c)P(M^c)P(A) \\ &= (1 - 31(0.01)) \cdot (1 - 28(0.01)) \cdot (1 - 31(0.01)) \cdot 30(0.01) = 0.69 \cdot 0.72 \cdot 0.69 \cdot 0.30 \approx 0.103. \end{aligned}$$

Q27. Un frottis est une procédure de dépistage utilisée pour détecter le cancer du col de l'utérus. Soient T^- et T^+ représentant les événements que le test est négatif et positif, respectivement, et soit C représentant l'événement que la personne testée a le cancer. Le taux de faux négatifs pour ce test lorsque la patiente a le cancer est de 16% ; le taux de faux positifs pour ce test lorsque la patiente n'a pas le cancer est de 19%. En Amérique du Nord, le taux d'incidence de ce cancer est d'environ 8 sur 100 000 femmes. Sur la base de ces chiffres, pensez-vous que le frottis est une procédure efficace ? Quels facteurs influencent votre conclusion ?

Solution : d'après l'énoncé du problème, nous avons

$$P(T^-|C) = 0.16,\ P(T^+|C) = 0.84,\ P(T^+|C^c) = 0.19,\ P(T^-|C^c) = 0.81,\ P(C) = 0.00008,\ P(C^c) = 0.99992.$$

Selon le théorème de Bayes et la loi de la probabilité totale,

$$\begin{aligned} P(C|T^+) &= \frac{P(T^+|C)P(C)}{P(T^+)} = \frac{P(T^+|C)P(C)}{P(T^+|C)P(C) + P(T^+|C^c)P(C^c)} \\ &= \frac{0.84 \cdot 0.00008}{0.84 \cdot 0.00008 + 0.19 \cdot 0.99992} \approx 0.0000354. \end{aligned}$$

Pour chaque million de frottis positifs, seuls 354 représentent de véritables cas de cancer du col de l'utérus. La procédure est inefficace parce que le taux de cancer est faible, et parce que les taux d'erreur de la procédure sont relativement élevés.

Q28. Des dés équitables différents sont donnés à Elowyn, Llewellyn, et Gwynneth. Ils lancent chacun le dé qu'ils ont reçu. Soit $E = \{\text{Elowyn lance un 1 ou un 2}\}$, $LL = \{\text{Llewellyn lance un 3 ou un 4}\}$, et $G = \{\text{Gwynneth lance un 5 ou un 6}\}$ trois événements d'intérêt.

a) Quelles sont les probabilités d'occurrence de chacun de E, LL et G ?

b) Quelles sont les probabilités que deux événements de E, LL et G se produisent simultanément ?

c) Quelles sont les probabilités que tous les événements se produisent simultanément ?

d) Quelle est la probabilité qu'au moins un des événements E, LL, ou G se produise ?

Solution :

a) $P(E) = P(LL) = P(G) = 1/3$.

b) $P(E \cap LL) = P(LL \cap G) = P(G \cap E) = 1/9$.

c) $P(E \cap LL \cap G) = 1/27$.

d) En utilisant la loi de De Morgan et en supposant que les événements sont indépendants,

$$\begin{aligned} P(E \cup LL \cup G) &= 1 - P((E \cup LL \cup G)^c) = 1 - P(E^c \cap LL^c \cap G^c) \\ &= 1 - P(E^c)P(LL^c)P(G^c) \\ &= 1 - (1 - P(E))(1 - P(LL))(1 - P(G)) \\ &= 1 - (1 - 1/3)^3 = 1 - 8/27 \approx 0.704. \end{aligned}$$

Q29. Au cours de deux saisons de baseball, le joueur A a obtenu 126 coups sûrs en 500 présences au bâton lors de la 1e saison, et 90 coups sûrs en 300 présences au bâton lors de la 2e saison ; le joueur B, quant à lui, a obtenu 75 coups sûrs en 300 présences au bâton lors de la 1e saison, et 145 coups sûrs en 500 présences au bâton lors de la 2e saison. La moyenne au bâton d'un joueur est le nombre de coup sûrs qu'il obtient, divisé par le nombre de présences au bâton.

a) Quel joueur a la meilleure moyenne au bâton lors de la 1e saison ? Lors de la 2e saison ?

b) Quel joueur a la meilleure moyenne au bâton lors des deux saisons ?

c) ????

Solution :

a) Lors de la 1e saison, la moyenne au bâton de A est $\frac{126}{500} = 0.252$; celle de B est $\frac{75}{300} = 0.250$. Lors de la 2e saison, la moyenne au bâton de A est $\frac{90}{300} = 0.300$; celle de B est $\frac{145}{500} = 0.290$. À chaque saison, A à une moyenne au bâton plus élevée que B.

b) Sur une période de 2 saisons, la moyenne au bâton de A est $\frac{126+90}{500+300} = 0.270$; celle de B est $\frac{75+145}{300+500} = 0.275$; la moyenne au bâton de B est plus élevée.

c) C'est le paradoxe de Simpson !

Q30. Un inconnu vous montre ce qui semble être une pièce de monnaie normale, avec deux faces distinctes, Pile (P) et Face (F). Il lance la pièce 4 fois et enregistre la séquence de lancers suivante : $FFFF$.

a) Quelle est la probabilité d'obtenir cette séquence spécifique de tirs à pile ou face ? Quelles hypothèses faites-vous en cours de route pour calculer cette probabilité ? Quelle est la probabilité que le prochain lancer soit un P.

b) L'inconnu vous propose un pari : il tirera à pile ou face une autre fois ; si le lancer est P, il vous donnera 100\$, mais s'il est F, vous lui donnerez 10\$ Accepteriez-vous le pari (si vous n'êtes pas moralement opposé aux jeux d'argent) ?

c) Maintenant, l'étranger lance la pièce 60 fois et enregistre 60 F d'affilée : $F \cdots F$. Il vous propose le même pari. L'acceptez-vous ?

d) Et s'il vous proposait 1000\$ à la place ? $1,000.000$\$?

Solution :

a) $(1/2)^4 = 0.0625$; indépendance des lancers et équité de la pièce ; $1/2 = 0.5$.

b) Le pari est à votre avantage : la moitié du temps, vous recevrez 100\$, l'autre moitié, vous perdrez 10\$; en moyenne, vous recevez plus que vous perdez si vous prenez le pari.

c) Vous savez quoi ? Même s'il est théoriquement possible qu'une pièce de monnaie équitable soit lancée indépendamment 60 fois et qu'elle tombe sur F 60 fois de suite, la probabilité que cela se produise est si faible, soit $(1/2)^{60} = 8.7 \times 10^{-19}$, que je ne crois pas vraiment que mes hypothèses étaient justifiées : soit la pièce n'est pas équitable, soit les lancers ne sont pas indépendants. Quoi qu'il en soit, j'ai mes doutes et je ne prends pas le pari.

d) Qu'en pensez-vous ?

Q31. L'espace d'échantillonnage d'une expérience aléatoire est $\{a, b, c, d, e, f\}$ et chaque résultat a la même probabilité d'être obtenu. Une v.a. est définie comme suit

résultat	a	b	c	d	e	f
X	0	0	1.5	1.5	2	3

Déterminez la f.m.p. de X. Déterminez les probabilités suivantes :

a) $P(X = 1.5)$

b) $P(0.5 < X < 2.7)$

c) $P(X > 3)$

d) $P(0 \leq X < 2)$

e) $P(X = 0 \text{ or } 2)$

Solution : la fonction de masse de probabilité est

$$P(X=0) = P(\{a,b\}) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{3},\ P(X=1.5) = P(\{c,d\}) = \tfrac{1}{3},$$
$$P(X=2) = P(\{e\}) = \tfrac{1}{6},\ P(X=3) = P(\{f\}) = \tfrac{1}{6}.$$

a) $P(X = 1.5) = \frac{2}{6} = \frac{1}{3}$
b) $P(0.5 < X < 2.7) = P(X = 1.5) + P(X = 2) = \frac{3}{6} = \frac{1}{2}$
c) $P(X > 3) = 0$
d) $P(0 \leq X < 2) = P(X = 0) + P(X = 1.5) = \frac{4}{6} = \frac{2}{3}$
e) $P(X = 0 \text{ or } X = 2) = \frac{3}{6} = \frac{1}{2}$

Q32. Déterminez la moyenne et la variance de la variable aléatoire définie à la **Q31**.

Solution : nous avons

$$\begin{aligned}
\mathrm{E}[X] &= \sum_i X_i P(X = X_i) \\
&= 0 \cdot P(X=0) + 1.5 \cdot P(X=1.5) + 2 \cdot P(X=2) + 3 \cdot P(X=3) \\
&= 0 \cdot \tfrac{1}{3} + 1.5 \cdot \tfrac{1}{3} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} = \tfrac{4}{3} \approx 1.33 \\
\mathrm{V}[X] &= \sum_i (X_i - \mathrm{E}[X])^2 P(X = X_i) \\
&= (0 - \tfrac{4}{3})^2 P(0) + (1.5 - \tfrac{4}{3})^2 P(1.5) + (2 - \tfrac{4}{3})^2 P(2) + (3 - \tfrac{4}{3})^2 P(3) \\
&= (\tfrac{4}{3})^2 \cdot \tfrac{1}{3} + (\tfrac{1}{6})^2 \cdot \tfrac{1}{3} + (\tfrac{2}{3})^2 \cdot \tfrac{1}{6} + (\tfrac{5}{3})^2 \cdot \tfrac{1}{6} = 41/36 \approx 1.39.
\end{aligned}$$

Q33. Une v.a. discrète X suit une loi uniforme à support sur $\{X_1, \ldots, X_k\}$ si

$$P(X = X_i) = \frac{1}{k}, \qquad i = 1, \ldots, k.$$

Les mesures d'épaisseur d'un processus de revêtement sont **uniformément distribuées** avec comme valeurs 0.15, 0.16, 0.17, 0.18, 0.19. Déterminez la moyenne et la variance des mesures d'épaisseur. Ce résultat est-il compatible avec une distribution uniforme ?

Solution : On note que pour $X_i \in \{0.15, 0.16, 0.17, 0.18, 0.18.0.19\}$ nous avons $P(X_i) = P(X = X_i) = 1/5$. La moyenne est

$$\mathrm{E}[X] = \sum_{i=1}^{5} X_i P(X_i) = \frac{1}{5}(0.15 + 0.16 + 0.17 + 0.18 + 0.19) = 0.17.$$

Dans ce cas, la variance est

$$\mathrm{V}[X] = \sum_{i=1}^{5} (X_i - \mathrm{E}[X])^2 P(X_i) = \frac{1}{5}\left(0.02^2 + 0.01^2 + 0^2 + 0.01^2 + 0.02^2\right) = 0.0002$$

Q34. 1% des échantillons de mitochondries rajeunies ont une mutation. On étudie 15 échantillons indépendants (du point de vue de la mutation). Déterminez les probabilités suivantes :

a) aucun échantillon n'a subit de mutation ;
b) au plus un échantillon a subit une mutation, et
c) plus de la moitié des échantillons ont subit une mutation.

On peut se servir de la table de f.r.c. de $\mathcal{B}(n, p)$, avec $n = 15$ et $p = 0.99$:

r	0	1	2	3	4	5	6	7
$P(X \le r)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
r	8	9	10	11	12	13	14	15
$P(X \le r)$	0.0000	0.0000	0.0000	0.0000	0.0004	0.0096	0.1399	1.0000

Solution : Soit X le nombre d'échantillons n'ayant pas subit de mutation ; X suit une loi binomiale avec $n = 15$ et $p = 0.99$ (succès : aucune mutation ; notez que si nous choisissons la mutation comme succès de l'épreuve, alors $p = 0.01$ et nous ne pouvons pas utiliser le tableau) :

a) 0 échantillon ayant subit une mutation = 15 échantillons n'ayant pas subit de mutation, nous devons donc évaluer $P(X = 15) = P(X \leq 15) - P(X \leq 14) = 1 - 0.1399 \approx 0.8601$;

b) au plus 1 échantillon ayant subit une mutation = au moins 14 échantillons n'ayant pas subit de mutation, donc nous devons évaluer

$$P(X \geq 14) = 1 - P(X < 14) = 1 - P(X \leq 13) = 1 - 0.0096 = 0.9904;$$

c) plus de la moitié des échantillons ayant subit une mutation = moins de (ou exactement) la moitié des échantillons n'ayant pas subit de mutation ; $P(X \leq 7,5) = P(X \leq 7) = 0.0000$.

Dans R, la f.r.c. binomiale pour $\leq r$ succès parmi n épreuves (chacun avec une probabilité p) est donnée par la fonction `pbinom(r,size=n,prob=p)`. La probabilité d'obtenir exactement r succès est donnée par la fonction `dbinom(r,size=n,prob=p)`.

a)
```
> pbinom(15,15,0.99) - pbinom(14,15,0.99) = 0.8600584
> dbinom(15,15,0.99) = 0.8600584
```

b)
```
> pbinom(15,15,0.99) - pbinom(13,15,0.99) = 0.9903702
> dbinom(14,15,0.99) + dbinom(15,15,0.99) = 0.9903702
```

c)
```
> pbinom(7.5,15,0.99) = 6.045248e-13
> pbinom(7,15,0.99) = 6.045248e-13
```

Q35. Des échantillons de 20 pièces provenant d'un processus de poinçonnage de métal sont sélectionnés à chaque heure. Typiquement, 1% des pièces doivent être retravaillées. Soit X le nombre de pièces de l'échantillon qui doivent être retravaillées. On soupçonne un problème de préparation si X dépasse sa moyenne de plus de trois écarts types.

a) Quelle est la probabilité qu'il y ait un problème de préparation ?

b) Si le proportion des pièces à retravailler augmente à 4%, quelle est la probabilité que X dépasse 1 ?

c) Si le proportion des pièces à retravailler augmente à 4%, quelle est la probabilité que X dépasse 1 dans au moins une des cinq prochaines heures d'échantillonnage ?

Solution :

a) Nous avons $X \sim \mathcal{B}(n, p)$, $n = 20$, $p = 0.01$ (où un succès = une pièce doit être retravaillée). Selon la définition de la loi binomiale, $\mathrm{E}[X] = np = 0.2$, $\mathrm{V}[X] = np(1-p) = 0.2 \times 0.99 = 0.198$, et $\mathrm{ET}[X] = \sqrt{\mathrm{V}[X]} \approx 0.44$. On cherche à calculer $P(X > \mathrm{E}[X] + 3 \cdot \mathrm{ET}[X])$. Ainsi,

$$\begin{aligned} P(X > 0.2 + 3 \cdot 0.44) = P(X > 1.535) &= P(X \geq 2) \\ &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \binom{20}{0} 0.01^0 \cdot 0.99^{20} - \binom{20}{1} 0.01^1 \cdot 0.99^{19} \\ &(= 1 - \mathtt{pbinom}(1, 20, 0.01)) \approx 0.017. \end{aligned}$$

Autrement, on peut travailler avec $Y \sim \mathcal{B}(n, 0.99)$, où Y est le # d'échantillons qui ne nécessitent pas de travail supplémentaire.

b) Nous avons la même configuration, mais avec $p = 0.04$. Nous sommes toujours intéressés par

$$\begin{aligned}P(X > 1) &= P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \binom{20}{0} 0.04^0 \cdot 0.96^{20} - \binom{20}{1} 0.04^1 \cdot 0.96^{19} \\ &(= 1 - \texttt{pbinom}(1, 20, 0.04)) \approx 0.19.\end{aligned}$$

c) Nous nous retrouvons avec 5 échantillons, et chacun d'eux est composé de 20 pièces. Soit W le nombre d'échantillons où le nombre d'articles qui nécessitent une retouche est supérieur à 1 (où un succès = plus de 1 d'article défectueux, c'est-à-dire $X > 1$).

Nous avons $W \sim \mathcal{B}(5, p_0)$, où $p_0 = P(X > 1) = 0.19$ est la probabilityé d'un succès. Nous devons ainsi évaluer

$$P(W \geq 1) = 1 - P(W = 0) = 1 - \binom{5}{0} 0.19^0 \cdot 0.81^5 \approx 0.651.$$

Cet exemple met en évidence la procédure à suivre pour les problèmes de cette nature : nous identifions la loi de probabilité appropriée ; nous évaluons ses paramètres ; nous identifions la probabilité appropriée à évaluer, et nous calculons sa valeur.

Q36. Dans le cadre d'une étude clinique, on teste la présence d'un gène qui augmente le risque de contracter une maladie particulière. La probabilité qu'une personne soit porteuse du gène est de 0.1.

a) Quelle est la probabilité qu'il faille tester ≥ 4 personnes afin de détecter 1 personne porteuse du gène ?

b) Combien de personnes devront être testées afin de détecter 1 personne porteuse du gène, en moyenne ?

c) Combien de personnes doivent être testées afin de détecter 2 personnes porteuses du gène, en moyenne ?

Solution : si X est le nombre de tests, le dernier étant le 1^{ier} succès (détection du gène), alors X suit une loi géométrique avec $p = 0.1$.

a) Dans ce cas, nous voulons évaluer

$$P(X \geq 4) = \sum_{k=4}^{\infty} (1-p)^{k-1} p = p \frac{(1-p)^3}{1-(1-p)} = (1-p)^3 = 0.729.$$

b) Selon la définition, $\mathrm{E}[X] = 1/p = 1/0.1 = 10$.

c) On peut considérer que cette procédure consiste à diviser les patients en 2 groupes, au hasard, et à tester chacun des groupes un par un. Il faut en moyenne 10 tests afin de détecter 1 gène dans l'un ou l'autre des groupes, donc 20 pour les deux groupes combinés.

Q37. Le nombre de pannes d'un instrument dues à des particules contaminées est une v.a. de Poisson avec une moyenne de 0.02 pannes par heure.

a) Quelle est la probabilité que l'instrument ne tombe pas en panne au cours d'un période de travail de 8 heures ?

b) Quelle est la probabilité qu'il y ait au moins 1 panne dans une journée de 24 heures ?

Solution:

a) Soit X le nombre de pannes par période de 8 heures. Le taux de pannes par 8 heures est de $8 \times 0.02 = 0.16$.

   Si X est une v.a. de Poisson avec $\lambda = 0.16$, nous voulons évaluer $P(X = 0) = \exp(-0.16) =$ $\texttt{ppois}(0.0.16) \approx 0.85$.

b) Dans ce cas, Y est le nombre de pannes lors d'une journée de 24 heures. Le taux d'échec lors d'une journée complète est $24 \times 0.02 = 0.48$.

   Si $Y \sim \mathcal{P}(0.48)$, nous cherchons à évaluer

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) = 1 - \exp(-0.48) \\ &= 1 - \texttt{ppois}(0.0.48) \approx 0.38. \end{aligned}$$

Q38. Utilisez `R` pour générer un échantillon à partir d'une loi binomiale et d'une loi de Poisson (sélectionnez les paramètres comme vous le souhaitez). Utilisez `R` afin de calculer les moyennes et les variances de l'échantillon. Comparez ces valeurs aux moyennes et aux variances de la population.

Solution: on peut générer des échantillons suivant les lois souhaitées sont générés par les fonctions `rbinom(n,size,prob)` et `rpois(n,lambda)`.

Nous simulons `n=1000` valeurs pour chaque échantillon. **ATTENTION:** dans le cours, on utilise n pour représenter le nombre d'épreuves ; dans `R`, n est utilisé pour représenter la taille de l'échantillon, et le nombre d'épreuves est représenté par le paramètre `size` (... ce n'est pas ce que j'aurais choisi).

Pour la loi binomiale X, nous utilisons `size=20` et `prob=0.2` ; pour la loi de Poisson Y, nous utilisons `lambda=8.5`.

Les véritables moyennes et variances sous-jacentes sont

$$\begin{aligned} &\mathrm{E}[X] = 20(0.2) = 4, \quad \mathrm{V}[X] = 20(0.2)(0.8) = 3.2 \\ &\mathrm{E}[Y] = \mathrm{V}[Y] = 8.5. \end{aligned}$$

Les échantillons et les valeurs estimées sont en accord avec les théoriques :

```
> X=rbinom(1000,20,0.2);
> mean(X); var(X);
[1] 4.022 [1] 3.0745901

> Y=rpois(1000,8.5);
> mean(Y); var(Y);
[1] 8.388 [1] 9.090547
```

Q39. Un récipient de 100 ampoules contient 5 mauvaises ampoules. On prélève 10 ampoules sans les remplacer. Trouvez la probabilité de tirer au moins 1 ampoule défectueuse.

a) 0.4164 b) 0.584 c) 0.1 d) 0.9 e) N/A

Solution : bien que cela ressemble à une expérience binomiale, il s'agit en fait d'une simple question de probabilité classique ; nous ne travaillons pas avec des essais indépendants car les ampoules sont échantillonnées SANS remplacement.

Soit X le nombre d'ampoules défectueuses. Nous cherchons à évaluer

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0).$$

Mais

$$P(X = 0) = \frac{\binom{95}{10}\binom{5}{0}}{\binom{100}{10}} = 0.584,$$

d'où $P(X \geq 1) = 1 - 0.584 = 0.416$.

Q40. Soit X une v.a. discrète avec un domaine $\{0, 1, 2\}$ et une f.m.p. donnée par $f(0) = 0.5$, $f(1) = 0.3$, et $f(2) = 0.2$. L'espérance et la variance de X sont, respectivement,

a) 0.7, 0.61 b) 0.7, 1.1 c) 0.5, 0.61 d) 0.5, 1.1 e) N/A

Solution : par définition,

$$\mathrm{E}[X] = \sum_i X_i f(X_i) = 0 \cdot 1.5 + 1 \cdot 0.3 + 2 \cdot 0.2 = 0.7$$

et

$$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum_i X_i^2 f(X_i) - (\mathrm{E}[X])^2 = 0^2 \cdot 0.5 + 1^2 \cdot 0.3 + 2^2 \cdot 0.2 - 0.7^2 = 0.61.$$

Q41. Une usine emploie plusieurs milliers de travailleurs, dont 30% ne sont pas d'origine anglophone. Si 15 membres du comité exécutif du syndicat ont été choisis au hasard parmi les travailleurs, évaluez la probabilité qu'exactement 3 membres du comité ne soient pas d'origine anglophone.

a) 0.17 b) 0.83 c) 0.98 d) 0.51 e) N/A

On peut se servir de la table de f.r.c. de $\mathcal{B}(n, p)$, avec $n = 15$ et $p = 0.30$:

r	0	1	2	3	4	5	6	7
$P(X \leq r)$	0.0047	0.0353	0.1268	0.2969	0.5155	0.7216	0.8689	0.9500
r	8	9	10	11	12	13	14	15
$P(X \leq r)$	0.9848	0.9963	0.9993	0.9999	1.0000	1.0000	1.0000	1.0000

Solution : Soit X le nombre de membres du comité exécutif qui ne sont pas d'origine anglophone. Comme le nombre d'employés est élevé et que le nombre de membres du comité est relativement faible, nous pouvons traiter l'acte de sélection des membres du comité comme des essais (approximativement) indépendants.

Alors $X \sim \mathcal{B}(15, 0.3)$ et nous voulons évaluer

$$P(X = 3) = \binom{15}{3} 0.3^3 \cdot 0.7^{12} \approx 0.17.$$

Alternativement, nous pourrions utiliser le tableau et calculer :

$$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.2969 - 0.1268 = 0.1701.$$

Q42. En supposant le contexte de **Q41**, quelle est la probabilité qu'une majorité des membres du comité ne soit pas d'origine anglophone ?

Solution : Soit X comme dans **Q41**. Nous voulons évaluer

$$P(X > 7.5) = P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9500 = 0.0500.$$

Q43. Dans un jeu vidéo, un joueur est confronté à une série d'adversaires et a une probabilité de 80% de vaincre chacun d'entre eux. La réussite contre un adversaire quelconque (c'est-à-dire la défaite de l'adversaire) est indépendante des rencontres précédentes. Le joueur continue jusqu'à ce qu'il soit vaincu. Quelle est la probabilité que le joueur rencontre au moins trois adversaires ?

a) 0.8 b) 0.64 c) 0.5 d) 0.36 e) N/A

Vous devrez peut-être utiliser la formule suivante :

$$\sum_{x=k}^{\infty} q^x = \frac{q^k}{1-q}, \quad 0 < q < 1.$$

Solution : soit X le nombre d'adversaires rencontrés (y compris le dernier, une perte par le joueur) ; X suit une loi géométrique, où un succès à l'essai est une perte contre un adversaire. La probabilité de succès est $p = 0.2$. Puisque X est géométrique, et que le joueur fait face à au moins 1 adversaire,

$$P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots.$$

Nous voulons évaluer

$$P(X \geq 3) = \sum_{x=3}^{\infty}(1-p)^{x-1}p = p\frac{(1-p)^2}{p} = (1-p)^2 = 0.64.$$

Q44. En supposant le contexte de **Q43**, combien de rencontres le joueur est-il censé faire ?

a) 5 b) 4 c) 8 d) 10 e) N/A

Solution : soit X comme dans **Q43**. Selon la formule de l'espérance pour une loi géometrique, $\mathrm{E}[X] = 1/p = 1/0.2 = 5$.

Q45. Par expérience, on sait que 3% des comptes d'une grande société comptable sont "en erreur". La probabilité qu'exactement 5 comptes soient vérifiés avant qu'un compte en erreur soit trouvé, est :

a) 0.242 b) 0.011 c) 0.030 d) 0.026 e) N/A

Solution : soit X le nombre de comptes qui doivent être audités avant qu'un compte en erreur soit trouvé. Alors X suit une loi géométrique avec probabilité de succès $p = 0.03$. Nous cherchons à évaluer

$$P(X = 5) = P(4 \text{ premiers comptes pas en erreur})P(5\text{ième compte en erreur}) = 0.97^4 \cdot 0.03 \approx 0.026.$$

Q46. La réception d'une compagnie reçoit en moyenne 2 appels téléphoniques par minute. Supposons que le nombre d'appels puisse être modélisé par une v.a. de Poisson. Quelle est la probabilité que la réception ne reçoive pas d'appel dans un intervalle de 3 minutes ?

a) e^{-2} b) $e^{-1/2}$ c) e^{-6} d) e^{-1} e) N/A

Solution : soit X le nombre d'appels reçus dans un intervalle de 3−minutes. Le taux d'appels par intervalle de 3−minutes est $3 \times 2 = 6$. Si X est une v.a. de Poisson avec $\lambda = 6$, nous cherchons à évaluer

$$P(X = 0) = \exp(-6) = e^{-6} \approx 0.0025$$

Q47. Considérons une v.a. X dont la f.d.p. est donnée par

$$f(x) = \begin{cases} 0 & \text{si } x \leq -1 \\ 0.75(1 - x^2) & \text{si } -1 \leq x < 1 \\ 0 & \text{si } x \geq 1 \end{cases}$$

Que sont l'espérance et l'écart-type de X ?

a) 0, 3 b) 0, 0.447 c) 1, 0.2 d) 1, 3 e) N/A

Solution : l'espérance de X est

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = 0.75 \int_{-1}^{1} x(1 - x^2)\, dx = 0.75 \int_{-1}^{1} (x - x^3)\, dx = 0.75 \left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{1} = 0$$

Son écart-type est

$$\mathrm{ET}[X] = \sqrt{\int_{-\infty}^{\infty} (x - E[X])^2 f(x)\, dx} = \sqrt{0.75 \int_{-1}^{1} (x^2 - x^4)\, dx} = \sqrt{0.75 \left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1}} = \sqrt{2(0.75)\left[\frac{1}{3} - \frac{1}{5}\right]} = \sqrt{0.2}.$$

Q48. Une v.a. X a une f.r.c. donnée par

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x/2 & \text{si } 0 < x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

Quelle est la valeur moyenne de X ?

a) 1 b) 2 c) 0 d) 0.5 e) N/A

Solution : la f.d.p. correspondante est

$$f(x) = F'(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1/2 & \text{si } 0 < x < 2 \\ 0 & \text{si } x \geq 2 \end{cases}$$

Ainsi

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = 0.5 \int_0^2 x\, dx = 0.5 \left[\frac{x^2}{2}\right]_0^2 = 0.5 \left[\frac{2^2}{2} - \frac{0^2}{2}\right] = 1.$$

Q49. Soit X une v.a. avec une f.d.p. donnée par $f(x) = \frac{3}{2}x^2$ pour $-1 \leq x \leq 1$, et $f(x) = 0$ autrement. Trouvez $P(X^2 \leq 0.25)$.

a) 0.250 b) 0.125 c) 0.500 d) 0.061 e) N/A

Solution : la f.r.c. correspondante est

$$F(x) = \int_{-1}^{x} f(t)\, dt = \begin{cases} 0 & \text{si } x \leq -1 \\ \frac{x^3}{2} + \frac{1}{2} & \text{si } -1 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

(vous devriez vérifier que $F(x)$ est continue). Dans ce cas,

$$\begin{aligned} P(X^2 \leq 0.25) &= P(-0.5 \leq X \leq 0.5) = F(0.5) - F(-0.5) \\ &= \left(\frac{0.5^3}{2} + \frac{1}{2}\right) - \left(\frac{(-0.5)^3}{2} + \frac{1}{2}\right) = 0.125. \end{aligned}$$

Q50. Lors de l'inspection de fer blanc produit par un procédé électrolytique continu, 0.2 imperfections sont repérées par minute, en moyenne. Trouvez la probabilité de repérer au moins 2 imperfections en 5 minutes (supposons que l'on puisse modéliser les occurrences d'imperfections selon une loi de Poisson).

a) 0.736 b) 0.264 c) 0.632 d) 0.368 e) N/A

Solution : soit X le nombre d'imperfections trouvées en 5 minutes. Le taux de repérage des imperfections par 5 minutes est de $0.2 \times 5 = 1$. Si X est une v.a. de Poisson avec $\lambda = 1$, nous cherchons à évaluer

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \exp(-\lambda) - \lambda \exp(-\lambda) = 1 - 2\exp(-1) \approx 0.264. \end{aligned}$$

Q51. Si $X \sim \mathcal{N}(0.4)$, la valeur de $P(|X|$
$ge2.2)$ est (en utilisant la table normale) :

a) 0.2321 b) 0.8438 c) 0.2527 d) 0.2713 e) 0.7286

f) N/A

Solution : soit $Z = \frac{X-0}{\sqrt{4}}$. Alors $Z \sim \mathcal{N}(0.1)$ et nous avons

$$P(|X| \geq 2.2) = 1 - P(|X| \leq 2.2) = 1 - P(-2.2 \leq X \leq 2.2) = 1 - P\left(\frac{-2.2-0}{\sqrt{4}} \leq \frac{X-0}{\sqrt{4}} \leq \frac{2.2-0}{\sqrt{4}}\right)$$
$$= 1 - P(-1.1 \leq Z \leq 1.1) = 1 - (\Phi(1.1) - \Phi(-1.1))$$
$$= 1 - (\texttt{pnorm}(1.1, 0, 1) - \texttt{pnorm}(-1.1, 0, 1)) \approx 0.2713.$$

On aurait également pu le calculer directement comme suit :

$$1 - P(-2.2 \leq X \leq 2.2) = 1 - (\texttt{pnorm}(2.2, 0, 2) - \texttt{pnorm}(-2.2, 0, 2)).$$

Q52. Si $X \sim \mathcal{N}(10.1)$, pour quelle valeur de k a-t-on $P(X \leq k) = 0.701944$?

a) 0.59 b) 0.30 c) 0.53 d) 10.53 e) 10.30

f) 10.59

Solution : soit $Z = \frac{X-10}{\sqrt{1}} = X - 10$. Alors $Z \sim \mathcal{N}(0.1)$ et nous avons

$$P(X \leq k) = P\left(\frac{X-10}{1} \leq \frac{k-10}{1}\right) = P(Z \leq k-10) = 0.701944.$$

D'après le tableau (ou `R`),

$$k - 10 = \Phi^{-1}(0.701944) = \texttt{qnorm}(0.701944, 0, 1) \approx 0.53,$$

d'où $k \approx 10.53$.

Q53. Le temps requis par un supercalculateur pour accomplir une tâche suit une loi normalede moyenne 10 ms et d'écart-type 4 ms. Quelle est la probabilité qu'il faille plus de 18.2 ms pour accomplir la tâche ? (utilisez le tableau normal ou `R`).

a) 0.9798 b) 0.8456 c) 0.0202 d) 0.2236 e) 0.5456

f) N/A

Solution : Soit $X \sim \mathcal{N}(10.4^2)$ et $Z = \frac{X-10}{4}$. Alors $Z \sim \mathcal{N}(0.1)$ et

$$P(X \geq 18.2) = 1 - P(X \leq 18.2) = 1 - P\left(\frac{X-10}{4} \leq \frac{18.2-10}{4}\right)$$
$$= 1 - P(Z \leq 2.05) \approx 0.0202.$$

Q54. Lancez deux fois un dé équitable à 4 faces. Soit X le plus grand des deux résultats s'ils sont différents ou la valeur commune s'ils sont identiques. Trouvez la f.p.m. et la f.r.c. de X.

Solution : l'espace d'échantillonnage de cette épreuve est

$$\mathcal{S} = \{(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}.$$

On suppose que chacun de ces 16 résultats a une probabilité égale, 1/16.

Par définition, $X(d_1, d_2) = \max\{d_1, d_2\}$ for any $(d_1, d_2) \in \mathcal{S}$. Ainsi,

$$P(X=1) = P[(1,1)] = \tfrac{1}{16},\ P(X=2) = P[\{((1,2),(2,1),(2,2)\}] = \tfrac{3}{16},$$
$$P(X=3) = P[\{(1,3),(2,3),(3,1),(3,2),(3,3)\}] = \tfrac{5}{16}$$
$$P(X=4) = P[\{(1,4),(2,4),(3,4),(4,1),(4,2),(4,3),(4,4)\}] = \tfrac{7}{16},$$

et la f.m.p. peut être simplement réécrite sous la forme

$$f(x) = P(X=x) = \tfrac{1}{16}(2x-1),\ \text{si } x = 1,2,3,4,\ \text{et } f(x) = 0 \text{ autrement.}$$

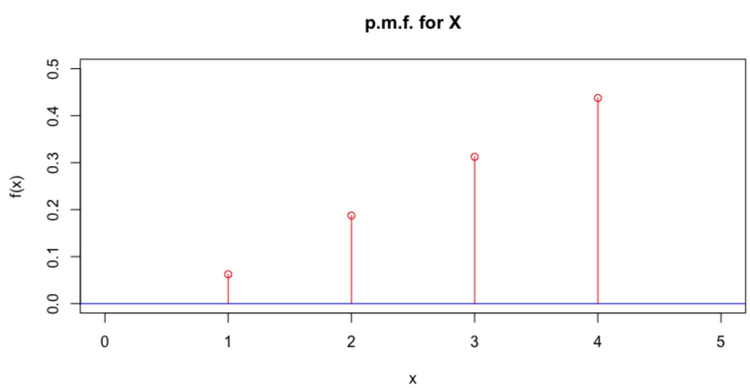
Nous pouvons vérifier que

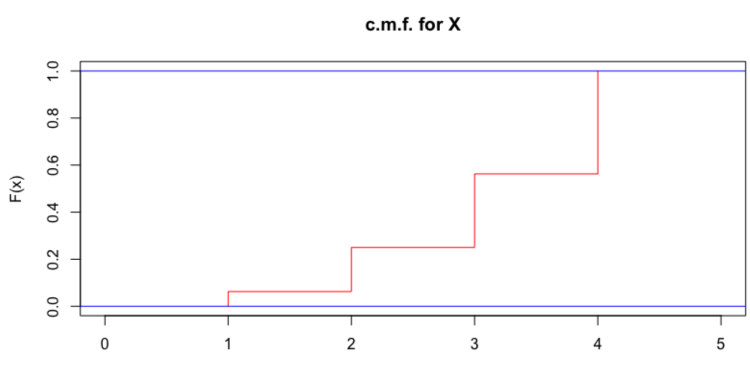
$$\sum_{x=1}^{4} f(x) = \frac{1}{16}\sum_{x=1}^{4}(2x-1) = \frac{2}{16}\sum_{x=1}^{4} x - \frac{1}{16}\sum_{x=1}^{4} 1$$
$$= \frac{2}{16}\cdot\frac{4(5)}{2} - \frac{1}{16}\cdot 4 = \frac{5}{4} - \frac{1}{4} = 1.$$

Puisque $2x > 1$ lorsque $x = 1,2,3,4,$ $f(x) \geq 0$ pour tout x et f est bien un f.m.p.

La f.r.c. correspondante est alors

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{(\lfloor x \rfloor)^2}{16} & \text{si } 1 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$





Le graphique de la page précédente a été produit avec le code suivant R (non commenté) :

```
> X <- 1:4
> P <- c(1/16,3/16,5/16,7/16)
> require(graphics)
> par(mfrow=c(2,1))
> plot(X,P,type="h",col=2,main="PMF",xlim=c(0,5),ylim=c(0,0.5),xlab="x", ylab="f(x)")
> points(X,P,col=2)
> abline(h=0,col=4)
> F <- cumsum(P)
> plot(c(1,X),c(0,F),type="s",main="CMF",xlim=c(0,5),ylim=c(0,1),col=2,xlab="x",ylab="F(x)")
> abline(h=0:1,col=4)
```

Q55. Calculez la moyenne et la variance de la v.a. X définie à la question **Q54**, tout comme $\mathrm{E}[X(5-X)]$.

Solution : soit X tel que définie à la question **Q54**. Alors

$$\mathrm{E}[X] = \sum_{x=1}^{4} xP(X=x) = 1\cdot\frac{1}{16} + 2\cdot\frac{3}{16} + 3\cdot\frac{5}{16} + 4\cdot\frac{7}{16} = \frac{25}{8} = 3.125$$

$$\begin{aligned}\mathrm{V}[X] &= \mathrm{E}[X^2] - \mathrm{E}^2[X] = \sum_{x=1}^{4} x^2P(X=x) - \left(\frac{25}{8}\right)^2 \\ &= \left(1^2\cdot\tfrac{1}{16} + 2^2\cdot\tfrac{3}{16} + 3^2\cdot\tfrac{5}{16} + 4^2\cdot\tfrac{7}{16}\right) - \tfrac{625}{64} = \tfrac{170}{16} - \tfrac{625}{64} = \tfrac{55}{64} \approx 0.859,\end{aligned}$$

tandis que $\mathrm{E}[X(5-X)] = \mathrm{E}[5X - X^2] = 5\mathrm{E}[X] - \mathrm{E}[X^2]$. Les deux premiers moments ont été calculés ci-dessus : $\mathrm{E}[X] = \frac{25}{8}$ and $\mathrm{E}[X^2] = \frac{170}{16}$, d'où

$$\mathrm{E}[X(5-X)] = 5\cdot\frac{25}{8} - \frac{170}{16} = 5.$$

Q56. Dans 80% des cas, lorsqu'un joueur de basket-ball tente un lancer franc, il le réussit. Supposons que chacune des tentatives de lancer franc soit indépendante. Soit X le nombre minimum de tentatives pour réussir 10 fois. Trouvez la f.m.p. de X et la probabilité que $X = 12$.

Solution : le joueur doit réussir au moins $x = 10$ lancers sur 10 tentatives distinctes.

Il doit y avoir 10 réussites au total (avec une probabilité de 0.8^{10}), et $x - 10$ échecs (avec une probabilité de 0.2^{10-x}).

De plus, le joueur doit réussir son $x$ème lancer (sinon x ne serait pas le nombre minimum de tentatives pour réussir 10 fois) ; dans les $x-1$ premiers lancers, exactement 9 doivent être réussis, et il y a ${}_{x-1}C_9 = \binom{x-1}{9}$ façons de les ordonner.

Ainsi, la f.m.p. de X est

$$f(x) = P(X=x) = \binom{x-1}{9}(0.8)^{10}(0.2)^{x-10}, \quad x = 10.11, 12, \ldots$$

et $P(X=12) = f(12) = \binom{11}{9}(0.80)^{10}(0.20)^2 \approx 0.2362$.

Q57. Soit X le nombre minimum d'essais indépendants (chacun avec une probabilité de succès p) qui sont nécessaires pour observer r succès. La f.m.p. de X est

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-1}, \quad x = r, r+1, \ldots$$

La moyenne et la variance de X sont $\mathrm{E}[X] = \frac{r}{p}$ et $\mathrm{V}[X] = \frac{r(1-p)}{p^2}$. Calculez le nombre minimum moyen de tentatives indépendantes de lancer franc nécessaires pour observer 10 lancers francs réussis si la probabilité de réussite à la ligne de lancer franc est de 80%. Qu'en est-il de l'écart-type de X ?

Solution : Soit X comme dans la question **Q56**. Nous avons

$$\mathrm{E}[X] = \frac{r}{p} = \frac{10}{0.80} = 12.5 \text{ et } \mathrm{V}[X] = \frac{r(1-p)}{p^2} = \frac{10(0.20)}{0.80^2} \approx 3.125,$$

d'où l'on conclut que $\mathrm{ET}[X] \approx \sqrt{3.125} \approx 1.768$.

Q58. Si $n \geq 20$ et $p \leq 0.05$, on peut montrer que la distribution binomiale avec n essais et une probabilité de succès indépendante p peut être approximée par une distribution de Poisson avec paramètre $\lambda = np$ (c'est l'approximation de Poisson) :

$$\frac{(np)^x e^{-np}}{x!} \approx \binom{n}{x} p^x (1-p)^{n-x}.$$

Un fabricant d'ampoules sait que 2% de ses ampoules sont défectueuses. Quelle est la probabilité qu'une boîte de 100 ampoules contienne au plus 3 ampoules défectueuses ? Utilisez l'approximation de Poisson pour estimer cette probabilité.

Solution : Soit X le nombre d'ampoules défectueuses dans la boîte de 100. Puisque $X \sim \mathcal{B}(100.0.02)$, nous avons

$$P(X = x) = \binom{100}{x} (0.02)^x (0.98)^{100-x}$$

et

$$P(X \leq 3) = \sum_{x=0}^{3} \binom{100}{x} (0.02)^x (0.98)^{100-x}.$$

Mais $n = 100 \geq 20$ et $p = 0.02 \leq 0.05$, d'où l'approximation de Poisson s'applique. Si $\lambda = np = 2$, alors

$$\frac{2^x e^{-2}}{x!} \approx \binom{100}{x} (0.02)^x (0.98)^{100-x},$$

donc

$$P(X \leq 3) \approx \sum_{x=0}^{3} \frac{2^x e^{-2}}{x!}$$

$$= e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} \right) \approx 0.857.$$

Q59. Considérons une v.a. discrète X qui a une distribution uniforme sur les m premiers entiers positifs, c'est-à-dire que

$$f(x) = P(X = x) = \frac{1}{m}, \quad x = 1, \dots, m,$$

et $f(x) = 0$ autrement. Calculez la moyenne et la variance de X. Pour quelles valeurs de m est-ce que $\mathrm{E}[X] > \mathrm{V}[X]$?

Solution : nous avons

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x=1}^{m} x f(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m} \sum_{x=1}^{m} x = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2}, \\
\mathrm{V}[X] &= E[X^2] - E^2[X] = \sum_{x=1}^{m} x^2 f(x) - \frac{(m+1)^2}{4} = \frac{1}{m} \sum_{x=1}^{m} x^2 - \frac{(m+1)^2}{4} \\
&= \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} - \frac{(m+1)^2}{4} = \frac{m^2-1}{12}.
\end{aligned}$$

La moyenne est supérieure à la variance lorsque

$$\frac{m+1}{2} > \frac{m^2-1}{12} \Longleftrightarrow 6(m+1) > m^2 - 1 \leftrightarrow m^2 - 6m - 7 < 0 \Longleftrightarrow 1 \leq m < 7.$$

Q60. Soit X une variable aléatoire. Quelle valeur de b (où b n'est pas une fonction de X) minimise $\mathrm{E}[(X-b)^2]$?

Solution : il suffit d'écrire

$$g(b) = \mathrm{E}[(X-b)^2] = \mathrm{E}[X^2 - 2bX + b^2] = \mathrm{E}[X^2] - 2b\mathrm{E}[X] + \mathrm{E}[b^2] = \mathrm{E}[X^2] - 2b\mathrm{E}[X] + b^2.$$

Pour trouver le minimum de $g(b)$ par rapport à b, on pose $g'(b) = 0$ et on résoud pour b :

$$g'(b) = -2\mathrm{E}[X] = 2b = 0 \Longrightarrow b = \mathrm{E}[X].$$

Puisque $g''(b) = 2 > 0$, $\mathrm{E}[X]$ est la valeur de b qui minimise $\mathrm{E}[(X-b)^2]$.

Q61. Une expérience consiste à choisir un bol, puis à tirer une boule de ce bol. Le bol B_1 contient deux boules rouges et quatre boules blanches ; le bol B_2 contient une boule rouge et deux boules blanches ; et le bol B_3 contient cinq boules rouges et quatre boules blanches. Les probabilités de sélection des bols ne sont pas uniformes : $P(B_1) = 1/3$, $P(B_2) = 1/6$, et $P(B_3) = 1/2$, respectivement.

a) Quelle est la probabilité de choisir une boule rouge $P(R)$?

b) Si l'expérience est réalisée et une boule rouge est tirée, de quel bol vient-elle ?

Solution :

a) Puisque B_1, B_2 et B_3 sont mutuellement exclusifs, nous pouvons appliquer la loi de la probabilité totale pour obtenir

$$\begin{aligned}
P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\
&= P(R|B_1)(1/3) + P(R|B_2)(1/6) + P(R|B_3)(1/2).
\end{aligned}$$

Mais selon l'énoncé du problème, nous avons $P(R|B_1) = \frac{2}{6}$, $P(R|B_2) = \frac{1}{3}$, $P(R|B_3) = \frac{5}{9}$, d'où

$$P(R) = (2/6)(1/3) + (1/3)(1/6) + (5/9)(1/2) = 4/9 \approx 0.44.$$

b) Nous cherchons les probabilités conditionnelles $P(R|B_1)$, $P(R|B_2)$, et $P(R|B_3)$. Selon le théorème de Bayes,

$$\begin{aligned} P(R|B_1) &= \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(2/6)(1/3)}{4/9} = \frac{2}{8} = \frac{1}{4} \\ P(B_2|R) &= \frac{(1/3)(1/6)}{4/9} = \frac{1}{8} \\ P(B_3|R) &= \frac{(5/9)(1/2)}{4/9} = \frac{5}{8}. \end{aligned}$$

Une fois que la boule rouge a été choisie, la probabilité concernant B_3 semble plus favorable car B_3 a un plus grand pourcentage de boules rouges que B_1 et B_2.

Q62. Le temps de réaction après un signal visuel suit une distribution normale avec une moyenne de 0.5 seconde et un écart-type de 0.035 seconde.

a) Quelle est la probabilité que le temps de réaction dépasse 1 seconde ?

b) Quelle est la probabilité que le temps de réaction soit compris entre 0.4 et 0.5 seconde ?

c) Quel est le temps de réaction qui est dépassé avec une probabilité de 0.9 ?

Solution : Soit X le temps de réaction. Alors $X \sim \mathcal{N}(0.5, (0.035)^2)$.

a) Nous cherchons à évaluer

$$P(X > 1) = 1 - \Phi\left(\tfrac{1-0.5}{0.035}\right) = 1 - \Phi(14.29) \approx 0.$$

b) Nous devons évaluer

$$P(0.4 < X < 0.5) = \Phi\left(\tfrac{0.5-0.5}{0.035}\right) - \Phi\left(\tfrac{0.4-0.5}{0.035}\right) = \Phi(0) - \Phi(-2.86) \approx 0.4979.$$

c) Nous voulons trouver x tel que $0.90 = P(X > x) = 1 - \Phi\left(\tfrac{x-0.5}{0.035}\right)$. Ainsi,

$$\Phi\left(\frac{x-0.5}{0.035}\right) = 0.10 \implies \frac{x-0.5}{0.035} = -1.28 \implies x = 0.4552.$$

Q63. Supposons que la v.a. X ait la f.r.c. suivante :

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ x^3, & 0 \leq x \leq 1 \\ 1, & x \geq 1. \end{cases}$$

a) Calculez $P(X > 0.5)$.

b) Calculez $P(0.2 < X < 0.8)$.

c) Trouvez la fonction de densité de probabilité de X.

d) Trouvez $\mathrm{E}[X]$ et $\mathrm{V}[X]$.

Solution :

a) $P(X > 0.5) = 1 - F(0.5) = 1 - 0.5^3 = 0.875$.

b) $P(0.2 < X < 0.8) = F(0.8) - F(0.2) = (0.8)^3 - (0.2)^3 = 0.504$.

c) $f_X(x) = F'(x) = 3x^2, \quad 0 < x < 1$.

d) Nous avons

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 3\, x^3\, dx = \frac{3}{4}x^4\Big|_0^1 = \frac{3}{4}$$

$$\mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2\, f(x)\, dx = \int_0^1 3\, x^4\, dx = \frac{3}{5}x^5\Big|_0^1 = \frac{3}{5}.$$

Ainsi, $\mathrm{V}[X] = \mathrm{E}[X^2] - \mu_X^2 = (3/5) - (3/4)^2 = 3/80 = 0.0375$.

Q64. Supposons que les arrivées de petits avions dans un aéroport puissent être modélisées par une v.a. de Poisson avec une moyenne de 1 avion par heure.

a) Quelle est la probabilité que plus de 3 avions arrivent en une heure ?

b) Considérez 15 intervalles consécutifs et disjoints de 1 heure. Quelle est la probabilité qu'aucun de ces intervalles ne voit l'arrivée de plus de 3 avions ?

c) Quelle est la probabilité qu'exactement 3 avions arrivent dans une période de 2 heures ?

Solution :

a) Soit X le nombre d'avions qui arrivent à l'aéroport en une heure. Ainsi, $X \sim \mathcal{P}(\lambda)$, avec $\lambda = 1$, et

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - \left[e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} + e^{-1}\frac{1^2}{2!} + e^{-1}\frac{1^3}{3!}\right] \approx 0.01899. \end{aligned}$$

Nous pouvons également calculer cette probabilité avec `R` :

$$1 - \texttt{ppois}(3, 1) \approx 0.01899.$$

b) Soit Y le nombre d'intervalles d'une heure avec plus de 3 arrivées. Si les arrivées sont indépendantes, nous pouvons considérer Y comme une épreuve binomiale avec $p = 0.01899$. Ainsi, $Y \sim B(15, 0.01899)$ et

$$P(Y = 0) = \binom{15}{0}(0.01899)^0(1 - 0.01899)^{15} = \texttt{dbinom}(0.15, 0.01899) \approx 0.7501.$$

c) Soit W le nombre d'arrivées dans une période de 2 heures. Ainsi, $W \sim \mathcal{P}(\lambda^*)$ avec $\lambda^* = 2$ et

$$P(W = 3) = e^{-2} \cdot \frac{2^3}{3!} = \texttt{dpois}(3, 2) \approx 0.1804.$$

Q65. Reportez-vous à la situation décrite dans la question **Q64**.

a) Quelle est la longueur de l'intervalle pour lequel la probabilité de n'avoir aucune arrivée est de 0.1 ?

b) Quelle est la probabilité que l'on doive attendre au moins 3 heures pour l'arrivée de 3 avions ?

c) Déterminez la moyenne et la variance du temps d'attente pour 3 avions ?

Solution :

a) Soit T le temps entre deux arrivées consécutives (en heures). Ainsi, $T \sim \text{Exp}(\lambda)$, pour $\lambda = 1$. Nous voulons trouver t tel que

$$0.1 = P(\text{aucune arrivée dans } [0.t]) = P(T > t) = e^{-\lambda t}.$$

Ainsi, $t \approx -\ln(0.1)/\lambda = 2,3026$ heures.

b) Soit S le nombre d'arrivées dans les 3 heures. Ainsi, $S \sim \mathcal{P}(\lambda^{**})$, avec $\lambda^{**} = 3$. La probabilité que l'on doive attendre au moins 3 heures pour l'arrivée de 3 avions est la probabilité qu'au plus 2 avions arrivent dans les 3 heures.Ainsi, nous calculons

$$\begin{aligned} P(S \leq 2) = P(S = 0) + P(S = 1) + P(S = 2) &= e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right] \\ &= \texttt{ppois}(2, 3) \approx 0.4232. \end{aligned}$$

c) Soit T le temps d'attente (en heures) pour 3 d'arrivées. Ainsi, $T \sim \Gamma(\lambda, r)$ avec $\lambda = 1$ et $r = 3$; alors

$$\text{E}[T] = \frac{r}{\lambda} = 3 \quad \text{et} \quad \text{V}[T] = \frac{r}{\lambda^2} = 3.$$

Q66. Supposons que X suit une loi normale avec une moyenne de 10 et un écart-type de 3. Dans chaque cas, trouvez la valeur x telle que :

a) $P(X > x) = 0.5$

b) $P(X > x) = 0.95$

c) $P(x < X < 10) = 0.2$

d) $P(-x < X - 10 < x) = 0.95$

e) $P(-x < X - 10 < x) = 0.99$

Solution : si $X \sim \mathcal{N}(10, 3^2)$, alors $Z = \frac{X-10}{3} \sim \mathcal{N}(0, 1)$. En utilisant les tables, nous obtenons :

a) $0.5 = P(X > x) = 1 - \Phi(\frac{x-10}{3}) \implies \Phi(\frac{x-10}{3}) = 0.5 \implies \frac{x-10}{3} = 0 \implies x = 10.$

b) $0.95 = P(X > x) = 1 - \Phi(\frac{x-10}{3}) \implies \Phi(\frac{x-10}{3}) = 0.05 \implies \frac{x-10}{3} \approx -1.64 \implies x \approx 5.08.$

c) $0.2 = P(x < X < 10) = \Phi(\frac{10-10}{3}) - \Phi(\frac{x-10}{3}) \implies \Phi(\frac{x-10}{3}) = \Phi(0) - 0.2 = 0.3 \implies \frac{x-10}{3} \approx -0.52 \implies x = 8.44.$

d) $0.95 = P(-x < X - 10 < x) = \Phi(x/3) - \Phi(-x/3) = \Phi(x/3) - [1 - \Phi(x/3)] \implies \Phi(x/3) = (0.95 + 1)/2 = 0.975 \implies x/3 \approx 1.96 \implies x \approx 5.88.$

e) $0.99 = P(-x < X - 10 < x) = \Phi(x/3) - \Phi(-x/3) = \Phi(x/3) - [1 - \Phi(x/3)] \implies \Phi(x/3) = (0.99 + 1)/2 = 0.995 \implies x/3 \approx 2.58 \implies x \approx 7.74.$

Q67. Soit $X \sim \text{Exp}(\lambda)$, avec une moyenne de 10. Quelle est la valeur de $P(X > 30|X > 10)$?

a) $1 - \exp(-2)$ b) $\exp(-2)$ c) $\exp(-3)$

d) $1/10$ e) $\exp(-200)$ f) N/A

Solution : nous avons $10 = \text{E}[X] = \frac{1}{\lambda}$, d'où $\lambda = \frac{1}{10}$. Selon la propriété de la perte de mémoire des lois exponentielles,

$$P(X > 30|X > 10) = P(X > 20) = \exp(-\lambda \cdot 20) = \exp(-2) \approx 0.1353.$$

Q68. Soit X le nombre de pannes d'une machine particulière au cours d'un mois. Sa f.m.p. est donnée par

x	0	1	2	3	4	5
$P(X = x)$	0.17	0.23	0.19	0.13	0.08	0.2

La probabilité qu'il y ait moins de 3 pannes en un mois, et l'espérance du nombre de pannes en un mois sont, respectivement,

a) $0.28; 2.50$ b) $0.72; 2.32$ c) $0.59; 2.32$ d) $0.80; 2.50$ e) N/A

Solution : Nous avons

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0.17 + 0.23 + 0.19 = 0.59$$

et

$$\text{E}[X] = \sum_{x=0}^{5} xP(X = 5) = 0 \cdot 0.17 + 1 \cdot 0.23 + 2 \cdot 0.19 + 3 \cdot 0.13 + 4 \cdot 0.08 + 5 \cdot 0.2 = 2.32$$

Q69. Le document de garantie d'une entreprise indique que la probabilité qu'une nouvelle piscine nécessite quelques réparations au cours de la première année est de 20%. Quelle est la probabilité que la sixième piscine vendue soit la première à nécessiter des réparations au cours de la première année ?

a) 0.6068 b) 0.3932 c) 0.9345 d) 0.0655 e) N/A

Solution : il s'agit d'une question sur la loi géométrique, le “succès” étant une piscine nécessitant quelques réparations au cours de la première année. Soit X le rang de la première piscine vendue qui nécessite des réparations. Nous avons $p = 0.2$ et $x = 6$, et

$$P(X = 6) = (1 - 0.2)^{6-1}(0.2) = 0.8^5 \cdot 0.2 \approx 0.0655,$$

en supposant que le besoin de réparations au cours de la première année est indépendant d'une piscine à l'autre.

Q70. Dans un groupe de dix étudiants, chaque étudiant a une probabilité de 0.7 de réussir l'examen. Quelle est la probabilité qu'exactement 7 d'entre eux réussissent l'examen ?

a) 0.9829 b) 0.2668 c) 0.0480 d) 0.9520 e) N/A

Solution : soit X le nombre d'étudiants qui réussissent l'examen. Nous supposons que $X \sim \mathcal{B}(10.0.7)$. Alors

$$P(X=7) = \binom{10}{7}(0.7)^7(0.3)^3 = \frac{10!}{7!3!}(0.7)^7(0.3)^3 \approx 0.2668.$$

Q71. Deux entreprises A et B envisagent de faire une proposition pour la construction d'une route. L'entreprise A soumet une proposition. La probabilité que B soumette une proposition est de 1/3. Si B ne soumet pas de proposition, la probabilité que A obtienne le contrat est de 3/5. Si B soumet une proposition, la probabilité que A l'obtienne est de 1/3. Quelle est la probabilité que A obtienne le contrat ?

a) 0.6667 b) 0.5111 c) 0.7500 d) 0.3333 e) N/A

Solution : soit A l'événement où l'entreprise A obtient le contrat, et B l'événement où l'entreprise B soumet une proposition.

Selon la loi de la probabilité totale,

$$\begin{aligned} P(A) &= P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B}) = P(A \mid B) \cdot 1/3 + P(A \mid \overline{B}) \cdot (1-1/3) \\ &= 1/3 \cdot 1/3 + 3/5 \cdot 2/3 = 1/9 + 2/5 \approx 0.5111 \end{aligned}$$

Q72. Dans une boîte de 50 fusibles, il y a 8 fusibles défectueux. On choisit 5 fusibles au hasard (sans remise). Quelle est la probabilité que les 5 fusibles soient tous non défectueux ?

a) 0.4015 b) 0.84 c) 0.3725 d) 0.4275 e) N/A

Solution : si nous supposons que les fusibles défectueux sont trouvés indépendamment dans la boîte, alors la probabilité qu'un fusible soit défectueux est $p = 8/50 = 0.16$. Supposons que nous échantillonnions $n = 5$ fusibles au hasard. Soit X le nombre de fusibles défectueux dans l'échantillon. Alors $X \sim \mathcal{B}(5, 0.16)$.

Nous cherchons

$$P(X=0) = \binom{5}{0}(0.16)^0(1-0.16)^{5-0} = 0.84^5 \approx 0.4182.$$

Q73. Considérez une v.a. X avec f.d.p. :

$$f(x) = \begin{cases} 0 & \text{si } x \leq -1 \\ \frac{3}{4}(1-x^2) & \text{si } -1 < x < 1 \\ 0 & \text{si } x \geq 1 \end{cases}$$

Quelle est la valeur de $P(X \leq 0.5)$?

a) 11/32 b) 27/32 c) 16/32 d) 1 e) N/A

Solution : on doit calculer

$$P(X \leq 0.5) = \int_{-1}^{0.5} \frac{3}{4}(1 - x^2)\, dx = \left[\frac{3}{4}x\right]_{-1}^{0.5} - \left[\frac{1}{4}x^3\right]_{-1}^{0.5} = 27/32.$$

Q74. La réception reçoit en moyenne 2 appels téléphoniques par minute. Si le nombre d'appels suit un processus de Poisson, quelle est la probabilité que le temps d'attente pour un appel soit supérieur à 1 minute ?

a) $e^{-1/15}$ b) $e^{-1/30}$ c) e^{-2} d) e^{-1} e) N/A

Solution : Nous avons un processus de Poisson avec $\lambda = 2$. Le temps d'attente dans un processus de Poisson suit un loi exponentielle. Soit W une v.a. exponentielle avec le paramètre $\lambda = 2$. Alors

$$P(W > 1) = 1 - P(W \leq 1) = 1 - (1 - \exp(-2 \cdot 1)) = \exp(-2).$$

Q75. Une entreprise fabrique des rondelles de hockey. On sait que leur poids suit une loi normale avec une moyenne de 1 et un écart-type de 0.05. Les rondelles utilisées par la LNH doivent peser entre 0.9 et 1.1. Quelle est la probabilité qu'une rondelle choisi au hasard soit conforme aux règles de la LNH ?

a) 1 b) 0.9545 c) 0.4560 d) 0.9772 e) N/A

Solution : on cherche à évaluer

$$P(0.9 < X < 1.1) = P\left(\frac{0.9 - 1.0}{0.05} < Z < \frac{1.1 - 1.0}{0.05}\right) = \Phi(2) - \Phi(-2) = 0.977250 - 0.022750 = 0.9545.$$

Q76. Considérez l'ensemble de données suivant :

`12 14 6 10 1 20 4 8`

Une médiane et une valeur de premier quartile de l'ensemble de données sont, respectivement :

a) 9, 5 b) 5.5, 6 c) 10, 5 d) 5, 10 e) N/A

Solution : commençons par ordonner l'ensemble des données :

`1 4 6 8 10 12 14 20`

Il y a $n = 8$ observations, donc Q_2 est le point médian des observations $n/2 = 4$ et $(n/2) + 1 = 5$, ce qui revient à dire :

$$\frac{x_4 + x_5}{2} = \frac{8 + 10}{2} = 9,$$

tandis que Q_1 est le point médian des observations $n/4 = 2$ et $(n/4) + 1 = 3$, ce qui revient à dire :

$$\frac{x_2 + x_3}{2} = \frac{4 + 6}{2} = 5.$$

Il existe des définitions informatiques concurrentes – souvenez-vous des définitions conceptuelles !

Q78. Considérons le code `R` suivant :

```
> pbinom(15,100,0.25)        > pbinom(16,100,0.25)        > pbinom(31,100,0.25)
[1] 0.01108327               [1] 0.02111062               [1] 0.9306511
> pbinom(17,100,0.25)        > pbinom(30,100,0.25)        > pbinom(32,100,0.25)
[1] 0.03762626               [1] 0.8962128                [1] 0.9554037
```

Soit X une v.a. binomiale avec $n = 100$ et $p = 0.25$. Quelle est la valeur de $P(16 \leq X \leq 31)$?

a) 0.9196 b) 0.9095 c) 0.9348 d) 0.9443 e) N/A

Solution : nous avons

$$
\begin{aligned}
P(16 \leq X \leq 31) &= P(X \leq 31) - P(X < 16) = P(X \leq 31) - P(X \leq 15) \\
&= \texttt{pbinom}(31, 100, 0.25) - \texttt{pbinom}(15, 100, 0.25) \approx 0.9307 - 0.0111 \approx 0.9196
\end{aligned}
$$

Q79. Supposons que des échantillons de taille $n = 25$ soient choisis au hasard dans une population normale avec une moyenne de 100 et un écart-type de 10. Quelle est la probabilité que $\overline{x}$ se situe dans l'intervalle

$$(\mu_{\overline{X}} - 1.8\sigma_{\overline{X}}, \mu_{\overline{X}} + 1.0\sigma_{\overline{X}})?$$

Solution : rappelons que $\overline{X} = \mu_X = 100$, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{5} = 2$ (notez cependant que cette information n'est absolument pas pertinente). À la place, nous utilisons le fait que $(\overline{X} - \mu_{\overline{X}})/\sigma_{\overline{X}} \sim \mathcal{N}(0.1)$, d'où

$$
\begin{aligned}
P\left(\mu_{\overline{X}} - 1.8\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} + 1.0\sigma_{\overline{X}}\right) = P\left(-1.8 < \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < 1.0\right) &= \Phi(1.0) - \Phi(-1.8) \\
&= 0.8413 - 0.0359 = 0.8054.
\end{aligned}
$$

Q80. La résistance à la compression du béton suit une loi normale avec $\mu = 2500$ et $\sigma = 50$. On prélève un échantillon aléatoire de taille 5. Quelle est l'erreur standard de la moyenne de l'échantillon ?

Solution : selon la définition, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{5}} = 22.3607.$

Q81. Supposons que $X_1 \sim \mathcal{N}(3, 4)$ et $X_2 \sim \mathcal{N}(3, 45)$. Sachant que X_1 et X_2 sont des variables aléatoires indépendantes, quelle est une bonne approximation de $P(X_1 + X_2 > 9.5)$?

a) 0.3085 b) 0.6915 c) 0.5279 d) 0.4271 e) N/A

Solution : puisque X_1 et X_2 sont indépendants,

$$X_1 + X_2 \sim \mathcal{N}(3 + 3, 4 + 45) = \mathcal{N}(6, 49).$$

Si $Y = X_1 + X_2$, alors

$$P(X_1 + X_2 > 9.5) = P(Y > 9.5) = P\left(\frac{Y - 6}{7} > \frac{9.5 - 6}{7}\right) = P(Z > 0.5) \approx 1 - 0.6915 = 0.3085$$

Q82. Le temps d'attente d'un client au comptoir d'enregistrement d'un aéroport est une v.a. avec une moyenne de $\mu = 8.2$ minutes et un écart-type de $\sigma = 1.5$ minutes. Supposons qu'un échantillon aléatoire de $n = 49$ clients soit prélevé. Calculez la probabilité approximative que le temps d'attente moyen de ces clients soit de :

a) moins de 10 min ; b) entre 5 et 10 min ; c) moins de 6 min.

Solution : soit $\overline{X}$ le temps d'attente moyen pour 49 clients. Selon le théorème de la limite centrale, $\overline{X} \sim \mathcal{N}\left(8.2, \frac{1.5^2}{49}\right)$.

a) $P(\overline{X} < 10) = \Phi\left(\frac{10-8.2}{1.5/\sqrt{49}}\right) = \Phi(8.4) \approx 1$.

b) $P(5 < \overline{X} < 10) = \Phi\left(\frac{10-8.2}{1.5/\sqrt{49}}\right) - \Phi\left(\frac{5-8.2}{1.5/\sqrt{49}}\right) = \Phi(8.4) - \Phi(-14.93) \approx 1 - 0 = 1$.

c) $P(\overline{X} < 6) = \Phi\left(\frac{6-8.2}{1.5/\sqrt{49}}\right) = \Phi(-10.26) \approx 0$.

Q83. On prélève un échantillon aléatoire de taille $n_1 = 16$ d'une population normale de moyenne 75 et d'écart-type de 8. On prélève un deuxième échantillon aléatoire de taille $n_2 = 9$ indépendamment d'une autre population normale de moyenne 70 et d'écart-type 12. Soient $\overline{X}_1$ et $\overline{X}_2$ les moyennes des deux échantillons. Trouvez

a) la probabilité que $\overline{X}_1 - \overline{X}_2$ dépasse 4, et

b) la probabilité que $3.5 < \overline{X}_1 - \overline{X}_2 < 5.5$.

Solution : nous avons $\overline{X}_1 - \overline{X}_2 \sim \mathcal{N}(75 - 70, 8^2/16 + 12^2/9) = \mathcal{N}(5, 20)$.

a) Ainsi,

$$P(\overline{X}_1 - \overline{X}_2 > 4) = P\left(Z > \frac{4-5}{\sqrt{20}}\right) = 1 - \Phi\left(\frac{4-5}{\sqrt{20}}\right) = 1 - \Phi(-0.22) = 0.5871.$$

b) De plus,

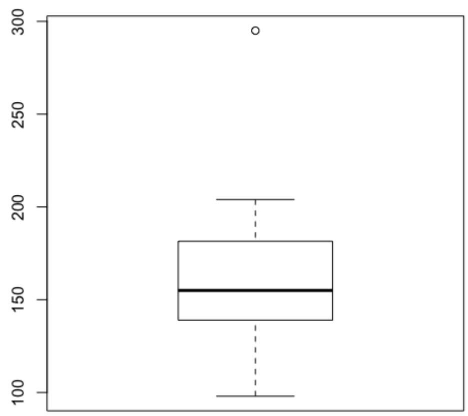
$$P(3.5 \leq \overline{X}_1 - \overline{X}_2 \leq 5.5) = \Phi\left(\frac{5.5-5}{\sqrt{20}}\right) - \Phi\left(\frac{3.5-5}{\sqrt{20}}\right) = \Phi(0.11) - \Phi(-0.34) = 0.5438 - 0.3669 = 0.1769$$

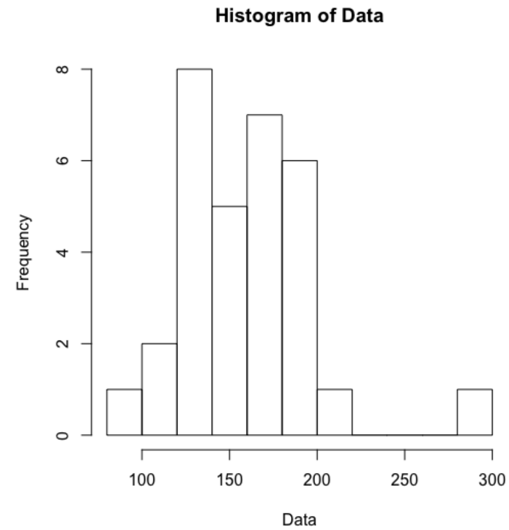
Q84. Discutez de la normalité de l'ensemble de données suivant :

```
170,295,200,165,140,190,195,142,138,148,110,140,103,176,125,
126,204,196,98,123,124,152,177,168,175,186,140,147,174,155,195
```

Solution : le code `R` suivant produira un boxplot et un histogramme pour les données.

```
> Data=c(170,295,200,165,140,190,195,142,138,148,110,140,103,
        176,125,126,204,196,98,123,124,152,177,168,175,186,140,147,174,155,195);
> par(mfrow=c(1,2)); boxplot(Data);hist(Data,breaks=10)
```





Les observations ne semblent pas être symétriques ; il y a une valeur aberrante ; les données ne semblent pas être normales.

Q85. En utilisant R, illustrez le TLC en générant $M = 300$ échantillons de taille $n = 30$ à partir de :

a) une v.a. normale de moyenne 10 et de variance 0.75 ;

b) une v.a. binomiale avec 3 d'essais et une probabilité de succès 0.3.

Répétez la même procédure pour des échantillons de taille $n = 200$. Qu'observez-vous ? **Indice :** Dans chaque cas, évaluez la normalité à l'aide d'un histogramme et d'un graphique QQ.

Solution : essayons le code suivant

```
> set.seed(1234); n=30; M=300
> x <- rnorm(n, mean=10, sd=sqrt(0.75)); hist(x) # or rbinom(n,3,0.3)
> mean(x); sd(x)
> means.x=c()
> for(m in 1:M){
    x <- rnorm(n, mean=10, sd=sqrt(0.75))
    means.x[m] = mean(x)
  }
> hist(means.x)
> mean(means.x); sd(means.x); sqrt(0.75)/sqrt(n)
> qqnorm(means.x, pch = 1, frame = FALSE)
> qqline(means.x, col = "steelblue", lwd = 2)
```

Q86. Supposons que le poids en livres d'un adulte nord-américain puisse être représenté par une v.a. normale avec une moyenne de 150 lbs et une variance de 900 lbs^2. Un ascenseur contenant un panneau "Maximum 12 personnes" peut transporter 2000 lbs en toute sécurité. La probabilité que 12 adultes nord-américains ne surchargent pas l'ascenseur est :

a) 0.9729 b) 0.4501 c) 0.0271 d) 0.0001 e) N/A

Solution : soit S le poids total de 12 adultes. Alors

$$S \sim \mathcal{N}(12 \times 150.12 \times 900) = \mathcal{N}(1800.10800),$$

et la probabilité qu'ils ne surchargent pas l'ascenseur est de

$$P(S < 2000) = P\left(\frac{S - 1800}{\sqrt{10800}} < \frac{2000 - 1800}{\sqrt{10800}}\right) \qquad = P(Z < 1.924) = \texttt{pnorm}(\texttt{1.924501}, \texttt{0.1}) \approx 0.9729.$$

Q87. Soient $X_1, \ldots, X_{50}$ un échantillon aléatoire indépendant issu d'une loi de Poisson de moyenne 1 et $Y = X_1 + \cdots + X_{50}$. La probabilité approximative $P(48 \leq Y \leq 52)$ est :

a) 0.6368 b) 0.4534 c) 0.2227 d) 0.9988 e) N/A

Solution : pour chaque variable de Poisson X_i, on a $\mathrm{E}[X_i] = \mathrm{V}[X_i] = 1$. Ainsi $\mathrm{E}[Y] = \mathrm{V}[Y] = 50$, et, selon le TLC, nous avons

$$\begin{aligned} P(48 \leq Y \leq 52) &= P\left(\frac{48 - 50}{\sqrt{50}} \leq \frac{Y - 50}{\sqrt{50}} \leq \frac{52 - 50}{\sqrt{50}}\right) \\ &\approx P(-0.2828 < Z < 0.2828) \\ &= \Phi(0.2828) - \Phi(-0.2828) \approx 0.2227. \end{aligned}$$

Q88. Un nouveau type de flash électronique pour appareils photo dure en moyenne 5000 heures avec un écart-type de 500 heures. Un ingénieur de contrôle de la qualité a l'intention de sélectionner un échantillon aléatoire de 100 de ces flashs et de les utiliser jusqu'à ce qu'ils tombent en panne. Quelle est la probabilité que la durée de vie moyenne de l'échantillon de 100 flashs soit inférieure à 4928 heures ?

a) 0.0749 b) 0.9251 c) 0.0002 d) 0.4532 e) N/A

Solution : nous avons $\mu = 5000$ et $\sigma = 500$. Soit $\overline{X}$ la durée de vie moyenne de $n = 100$ flashs. Alors $\mu_{\overline{X}} = 5000$ et $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{10}$. Nous normalisons pour obtenir

$$\begin{aligned} P(\overline{X} < 4298) &= P\left(\frac{\overline{X} - 5000}{500/10} < \frac{4928 - 5000}{500/10}\right) \\ &\approx P(Z < -1.44) = \Phi(-1.44) \approx 0.0749. \end{aligned}$$

Q89. Un fabricant de dentifrice fluoré mesure régulièrement la concentration de fluorure dans le dentifrice pour s'assurer qu'elle est conforme aux spécifications de 0.85 à 1.10 mg/g. Le tableau de la page suivante répertorie 100 mesures de ce type. Construisez un histogramme de fréquence relative des données (un histogramme d'aire = 1).

Table 6.1-3 Concentrations of fluoride in mg/g in toothpaste									
0.98	0.92	0.89	0.90	0.94	0.99	0.86	0.85	1.06	1.01
1.03	0.85	0.95	0.90	1.03	0.87	1.02	0.88	0.92	0.88
0.88	0.90	0.98	0.96	0.98	0.93	0.98	0.92	1.00	0.95
0.88	0.90	1.01	0.98	0.85	0.91	0.95	1.01	0.88	0.89
0.99	0.95	0.90	0.88	0.92	0.89	0.90	0.95	0.93	0.96
0.93	0.91	0.92	0.86	0.87	0.91	0.89	0.93	0.93	0.95
0.92	0.88	0.87	0.98	0.98	0.91	0.93	1.00	0.90	0.93
0.89	0.97	0.98	0.91	0.88	0.89	1.00	0.93	0.92	0.97
0.97	0.91	0.85	0.92	0.87	0.86	0.91	0.92	0.95	0.97
0.88	1.05	0.91	0.89	0.92	0.94	0.90	1.00	0.90	0.93

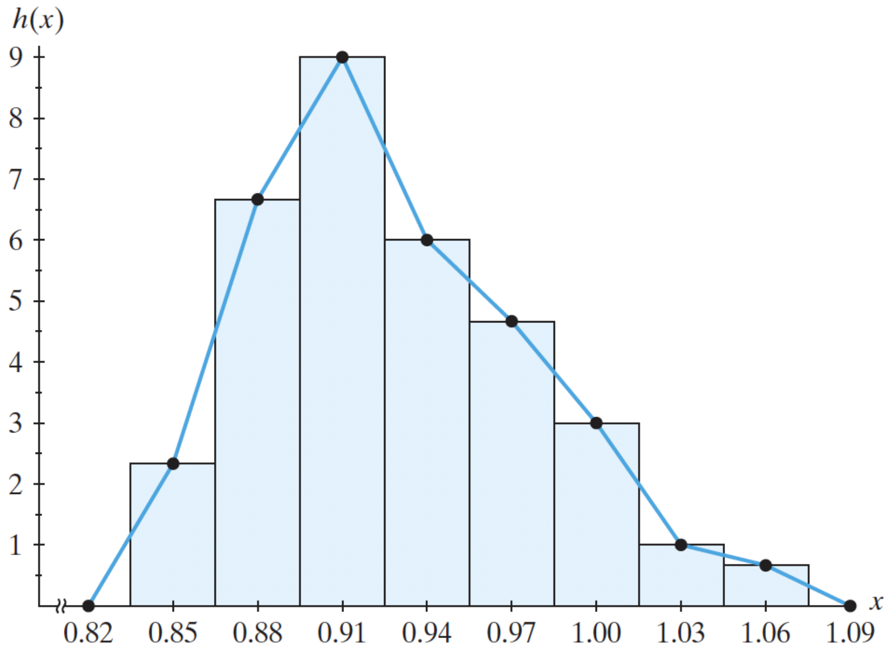
Solution : le minimum de ces mesures est de 0.85 et le maximum est de 1.06. L'intervalle est de $1.06 - 0.85 = 0.21$. Nous utilisons $k = 8$ classes de longueur 0.03. Notez que $8(0.03) = 0.24 > 0.21$. Nous commençons à 0.835 et finissons à 1.075. Ces limites sont à la même distance en dessous du minimum et au-dessus du maximum. Dans le tableau suivant, nous donnons les valeurs de la hauteur de chacun des rectangles de l'histogramme (la superficie totale doit être égale à 1.) Les hauteurs sont

$$h(x) = \frac{f_i}{(0.03)(100)} = \frac{f_i}{3},$$

où f_i est la fréquence des concentrations dans la classe i. Le tracé de l'histogramme est présenté à la page suivante.

Table 6.1-4 Frequency table of fluoride concentrations

Class Interval	Class Mark (u_i)	Tabulation	Frequency (f_i)	$h(x) = f_i/3$
$(0.835, 0.865)$	0.85	~~||||~~ ||	7	7/3
$(0.865, 0.895)$	0.88	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	20	20/3
$(0.895, 0.925)$	0.91	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ||	27	27/3
$(0.925, 0.955)$	0.94	~~||||~~ ~~||||~~ ~~||||~~ |||	18	18/3
$(0.955, 0.985)$	0.97	~~||||~~ ~~||||~~ ||||	14	14/3
$(0.985, 1.015)$	1.00	~~||||~~ ||||	9	9/3
$(1.015, 1.045)$	1.03	|||	3	3/3
$(1.045, 1.075)$	1.06	||	2	2/3



Q90. Utilisez les données de la question **Q89**.

a) Calculez la moyenne des données $\overline{x}$ et son écart-type s_x (utilisez un programme informatique).

b) À l'aide du tableau de fréquence des concentrations de fluorure (tableau 6.1-4), vous pouvez également calculer approximativement la moyenne et la variance. Soient u_i la **marque de classe** pour chacune des 8 classes de l'histogramme (le point médian sur la largeur des rectangles), n le nombre total d'observations et k le nombre de classes. Alors

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} f_i u_i \quad \text{et} \quad s_u^2 = \frac{1}{n-1} f_i (u_i - \overline{u})^2.$$

Calculez $\overline{u}$ et s_u. Comment se comparent-ils à $\overline{x}$ et s_x ?

Solution :

a) Le code R suivant fera l'affaire :

```
> CCs_Fl<-c(0.98,0.92,0.89,0.90,0.94,0.99,0.86,0.85,1.06,1.01,
      1.03,0.85,0.95,0.90,1.03,0.87,1.02,0.88,0.92,0.88,
      0.88,0.90,0.98,0.96,0.98,0.93,0.98,0.92,1.00,0.95,
      0.88,0.90,1.01,0.98,0.85,0.91,0.95,1.01,0.88,0.89,
      0.99,0.95,0.90,0.88,0.92,0.89,0.90,0.95,0.93,0.96,
      0.93,0.91,0.92,0.86,0.87,0.91,0.89,0.93,0.93,0.95,
      0.92,0.88,0.87,0.98,0.98,0.91,0.93,1.00,0.90,0.93,
      0.89,0.97,0.98,0.91,0.88,0.89,1.00,0.93,0.92,0.97,
      0.97,0.91,0.85,0.92,0.87,0.86,0.91,0.92,0.95,0.97,
      0.88,1.05,0.91,0.89,0.92,0.94,0.90,1.00,0.90,0.93)
> mean(CCs_Fl)
[1] 0.9293
> sd(CCs_Fl)
[1] 0.0489538
```

b) En utilisant les fréquences relatives f_i et les marques de classe u_i des tables de fréquence, nous obtenons

$$\overline{u} = \frac{1}{100}\sum_{i=1}^{8} f_i u_i = \frac{92.83}{100} = 0.9283,$$

$$s_u^2 = \frac{1}{100-1}\sum_{i=1}^{8} f_i(u_i-\overline{u})^2 = \frac{1}{100-1}\left(\sum_{i=1}^{8} f_i u_i^2 - \frac{1}{100}\left(\sum_{i=1}^{8} f_i u_i\right)^2\right) = \frac{0.237411}{99} \approx 0.002398,$$

d'où $s_u = \sqrt{0.002398} = 0.04897$.

Q91. Utilisez les données de la question **Q89**.

a) Fournissez un résumé en 5 chiffres des données $(q_0.q_1, q_2, q_3, q_4)$, ainsi que de son étendue interquartile.

b) Affichez le résumé numérique à 5 sous forme d'un graphique boxplot.

Solution :

a) Pour faciliter le calcul des quartiles, nous fournissons un **diagramme à tiges et à feuilles ordonné** des concentrations sur la diapositive suivante. Il y avait 100 observations, donc

$$Q_0 = 0.85, Q_1 = 0.89, Q_2 = 0.92, Q_3 = 0.97, Q_4 = 1.06.$$

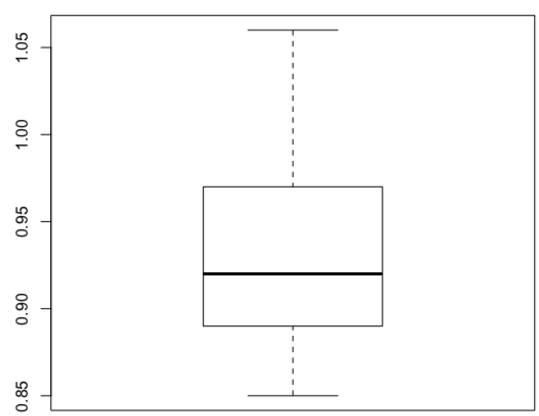
Le EIQ, quant à lui, est $Q_3 - Q_1 = 0.08$.

b) Il n'y a pas de valeurs aberrantes puisque

$$Q_1 - 1.5 \times \text{EIQ} = 0.85 - 1.5(0.08) < Q_0, \quad \text{et}$$
$$Q_3 + 1.5 \times \text{EIQ} = 0.97 + 1.5(0.08) > Q_4.$$

Table 6.2-6 Ordered stem-and-leaf diagram of fluoride concentrations

Stems	Leaves	Frequency
0.8*f*	5 5 5 5	4
0.8*s*	6 6 6 7 7 7 7	7
0.8•	8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9	16
0.9∗	0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1	17
0.9*t*	2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3	19
0.9*f*	4 4 5 5 5 5 5 5 5	9
0.9*s*	6 6 7 7 7 7	6
0.9•	8 8 8 8 8 8 8 8 9 9	10
1.0∗	0 0 0 0 1 1 1	7
1.0*t*	2 3 3	3
1.0*f*	5	1
1.0*s*	6	1



Q92. Utilisez les données de la question **Q89**. Calculez le **milieu de gamme** $\frac{1}{2}(Q_0 + Q_4)$, la **moyenne tiercée** $\frac{1}{4}(Q_1 + 2Q_2 + Q_3)$, et l'**étendue** $Q_4 - Q_0$ pour les données de fluorure.

Solution : le milieu de gamme est $\frac{1}{2}(Q_0 + Q_4) = \frac{1}{2}(0.85 + 1.06) = 0.955$, la moyenne tiercée est $\frac{1}{4}(Q_1 + 2Q_2 + Q_3) = \frac{1}{4}(0.89 + 2(0.92) + 0.97) = 0.925$, et l'étendue est $Q_4 - Q_0 = 1.06 - 0.85 = 0.21$.

Q93. Une nouvelle cure a été développée pour un certain type de ciment qui devrait modifier sa résistance moyenne à la compression. On sait que l'écart-type de la résistance à la compression est de 130 kg/cm^2 et que l'on peut supposer qu'il suit une distribution normale. 9 morceaux de ciment ont été testés et la moyenne observée de l'échantillon est $\overline{X} = 4970$. Trouvez l'intervalle de confiance à environ 95% pour la moyenne de la résistance à la compression.

a) $[4858.37, 5081.63]$ b) $[4885.07, 5054.93]$ c) $[4858.37, 5054.93]$

d) $[4944.52, 4995.48]$ e) N/A

Solution : pour une population normale dont l'écart-type σ est connu, l'IC à 95% est

$$\overline{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 4970 \pm 1.96\left(\frac{130}{\sqrt{9}}\right) \approx [4885.07, 5054.93].$$

Q94. Considérons la même situation qu'à la question **Q93**, mais 100 morceaux de ciments on été testés avec $\overline{X} = 4970$. Déterminez l'IC à environ 95% pour la moyenne de la résistance à la compression.

a) $[4858.37, 5081.63]$ b) $[4885.07, 5054.93]$ c) $[4858.37, 5054.93]$

d) $[4944.52, 4995.48]$ e) N/A

Solution : pour une population normale dont l'écart-type σ est connu, l'IC de la moyenne à 95% est

$$\overline{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 4970 \pm 1.96\left(\frac{130}{\sqrt{100}}\right) \approx [4944.52, 4995.48].$$

Comparez avec les résultats de la question précédente, et observez l'effet de la plus grande taille de l'échantillon.

Q95. Considérons la même configuration qu'à la question **Q93**, mais maintenant nous ne connaissons pas l'écart-type de la distribution normale. 9 morceaux de ciment ont été testés, et les mesures sont

$$5001, 4945, 5008, 5018, 4991, 4990.4968, 5020.5003.$$

Trouvez l'intervalle de confiance à environ 95% de la moyenne de la résistance à la compression.

a) $[4858.37, 5081.63]$ b) $[4885.07, 5054.93]$ c) $[4858.37, 5054.93]$

d) $[4944.52, 4995.48]$ e) N/A

Solution : nous avons $\overline{X} \approx 4993.78$ et $s \approx 24.18$. Pour une population normale dont l'écart-type σ est inconnu, l'IC à 95% est

$$\overline{X} \pm t_{0.025,8} \frac{s}{\sqrt{n}} = 4993.78 \pm 2.306 \left(\frac{24.18}{\sqrt{9}} \right) \approx [4975.19, 5012.37].$$

Q96. Une barre d'acier est mesurée avec un appareil dont la précision connue est de $\sigma = 0.5$mm. Supposons que nous voulions estimer la mesure moyenne avec une erreur d'au plus 0.2mm à un niveau de signification $\alpha = 0.05$. Quelle taille d'échantillon est nécessaire (supposez que la population est normale).

a) 25 b) 24 c) 6 d) 7 e) N/A

Solution : si nous supposons la normalité, l'erreur est

$$E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \quad \Longrightarrow n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2.$$

Puisque $E = 0.2$, $\alpha = 0.05$, et puisque $\sigma = 0.5$, alors

$$n = \left(\frac{z_{0.025}\sigma}{E} \right)^2 = \left(\frac{1.96 \times 0.5}{0.2} \right)^2 = 24.01,$$

d'où $n \geq 25$.

Q97. Dans un échantillon aléatoire de 1000 maisons, on trouve que 228 sont chauffées au mazout. Trouvez un IC à environ 99% de la proportion de maisons qui sont chauffées au mazout.

a) $[0.202, 0.254]$ b) $[0.197, 0.259]$ c) $[0.194, 0.262]$

d) $[0.185, 0.247]$ e) N/A

Solution : selon la formule de l'intervalle de confiance d'une proportion, nous avons

$$\begin{aligned} \hat{p} \pm z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= \frac{228}{1000} \pm 2.576 \sqrt{\frac{228/1000(1-228/1000)}{1000}} \\ &\approx [0.194, 0.262]. \end{aligned}$$

Q98. Historiquement, la résistance à la rupture des fils utilisés dans la fabrication des tissus pour rideaux suit une loi normale avec $\sigma = 2$ psi. Un échantillon aléatoire de 15 spécimens est testé et la force de rupture moyenne s'avère être $\overline{x} = 97.5$ psi.

a) Trouvez un intervalle de confiance de la véritable force de rupture moyenne, à environ 95%.

b) Trouvez un intervalle de confiance de la véritable force de rupture moyenne, à environ 95%.

Solution : dans les deux cas, nous avons une population normale avec une variance σ connue.

a) L'intervalle de confiance est donné par

$$\overline{x} \pm z_{.025} \frac{\sigma}{\sqrt{n}} = 97.5 \pm 1.96 \left(\frac{2}{\sqrt{15}} \right) = 97.5 \pm 1.012140 \approx [96.49, 98.51].$$

b) L'intervalle de confiance est donné par

$$\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = 97.5 \pm 2.576\left(\frac{2}{\sqrt{15}}\right) = 97.5 \pm 1.330241 \approx [96.17, 98.83].$$

Q99. Le diamètre des trous d'un faisceau de câbles suit une loi normale avec $\sigma = 0.01$ pouce. Pour un échantillon de taille 10, le diamètre moyen est de 1.5045 pouces.

a) Trouvez un intervalle de confiance du véritable diamètre moyen du trou, à environ 99%.

b) Répétez cette opération en supposant que $n = 100$.

Solution : dans les deux cas, nous avons une population normale avec une variance σ connue.

a) Si $n = 10$, l'intervalle de confiance de la moyenne à environ 99% est

$$\begin{aligned}\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} &= 1.5045 \pm 2.576\left(\frac{0.01}{\sqrt{10}}\right) = 1.5045 \pm 0.008146027\\ &\approx [1.496354, 1.512646].\end{aligned}$$

b) Si $n = 100$, l'intervalle de confiance de la moyenne à environ 99% est

$$\begin{aligned}\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} &= 1.5045 \pm 2.576\left(\frac{0.01}{\sqrt{100}}\right) = 1.5045 \pm 0.002576\\ &\approx [1.501924, 1.507076].\end{aligned}$$

Q100. Un article de journal décrit l'effet de la délamination sur la fréquence naturelle des poutres fabriquées à partir de composés laminés. Les observations sont les suivantes :

$$230.66, 233.05, 232.58, 229.48, 232.58, 235.22.$$

En supposant que la population est normale, trouvez un intervalle de confiance à environ 95% de la fréquence naturelle moyenne.

Solution : nous avons $n = 6$, $\overline{x} = 232.2617$, $s = 1.993935$. Comme nous ne connaissons pas le véritable écart-type, on obtient l'intervalle de confiance recherché à l'aide de

$$\overline{x} \pm t_{0.025,5}\frac{s}{\sqrt{n}} = 232.2617 \pm 2.571\left(\frac{1.993935}{\sqrt{6}}\right) = [230.169, 234.355].$$

Q101. Un fabricant de fibres textiles étudie un nouveau fil de draperie qui, selon le vendeur, présente un allongement moyen du fil de $\mu = 12$ kilogrammes avec un écart-type de $\sigma = 0.5$ kilogrammes.

a) Quelle doit être la taille de l'échantillon pour qu'avec une probabilité de 0.95 nous estimions l'allongement moyen du fil avec une erreur d'au plus 0.15 kg ?

b) Quelle devrait être la taille de l'échantillon pour qu'avec une probabilité de 0.95, nous estimions l'allongement moyen du fil avec une erreur maximale de 0.05 kg ?

Solution :

a) On doit avoir

$$n \geq \left(\frac{z_{.025}\,\sigma}{E}\right)^2 = \left(\frac{(1.96)(0.5)}{0.15}\right)^2 = 42.68.$$

b) On doit avoir

$$n \geq \left(\frac{z_{.025}\,\sigma}{E}\right)^2 = \left(\frac{(1.96)(0.5)}{0.05}\right)^2 = 384.16.$$

Q102. La luminosité d'un tube image de télévision peut être évaluée en mesurant la quantité de courant nécessaire pour atteindre un niveau de luminosité particulier. Une ingénieure pense qu'il faut utiliser 300 microampères de courant pour atteindre le niveau de luminosité requis.

a) Formulez l'hypothèse nulle et l'hypothèse alternative. Utilisez un test bilatéral.

b) Pour un échantillon aléatoire de taille $n = 20$ on obtient $\overline{x} = 319.2$ et $s = 18.6$. Testez les hypothèses de la partie a) avec $\alpha = 5\%$ en calculant une région critique. Calculez la valeur de p associée.

c) Utilisez les données de la partie b) pour construire un I.C. à environ 95% du courant moyen requis.

Solution :

a) On cherche à vérifier si $\mu = 300$, on test alors $H_0 : \mu = 300$ envers l'alternative $H_1 : \mu \neq 300$.

b) La valeur observée de la statistique du test

$$t_0 = \frac{\overline{x} - 300}{s/\sqrt{n}} = \frac{319.2 - 300}{18.6/\sqrt{20}} = 4.61.$$

Région critique de réjection : on rejette H_0 si $|t_0| > t_{0.025,19} = 2.093$.

Conclusion: Puisque $|t_0| = 4.61$, on rejette H_0. À un niveau de signification de 5%, on peut conclure que la moyenne n'est pas 300.

Approche de la valeur$-p$:

$$2\,P(\overline{X} > 319.2) = 2 \times P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > \frac{319.2 - 300}{18.6/\sqrt{20}}\right) = 2 \times P(t_{19} > 4.61).$$

D'après la table des valeurs de T, on constate que $P(T > 4.61) < 0.0005$ et donc que la valeur$-p$ est $< 2(0.0005) = 0.001$, avec la même conclusion.

c) Si la population est normale, l'I.C. de la moyenne à environ 95% est

$$\overline{x} \pm t_{0.025,19}\frac{s}{\sqrt{n}} = 319.2 \pm (2.093)\frac{18.6}{\sqrt{20}} = [310.495, 327.905].$$

En particulier, puisque 300 ne se retrouve pas dans l'I.C., on rejette H_0.

Q103. Un procédé de production particulier est **stable** s'il produit au plus 2% d'articles défectueux. Soit p la proportion réelle d'articles défectueux.

a) On échantillonne $n = 200$ articles au hasard et on envisage de tester des hypothèses sur p. Formulez les hypothèses nulles et alternatives.

b) Quelle est votre conclusion du test ci-dessus, si l'on observe 3 d'articles défectueux sur 200 ? Vous devez choisir un niveau approprié α.

Solution :

a) $H_0 : p = 0.02$ vs $H_1 : p > 0.02$

b) **Approche de la valeur$-p$:** soit X le nombre d'items déféctueux. Alors

$$2 \cdot P(X \leq 3) = 2 \cdot P\left(\frac{X - np + 0.5}{\sqrt{np(1-p)}} \leq \frac{3 - 4 + 0.5}{\sqrt{200(0.02)(0.98)}}\right) = 2 \cdot P(Z < -0.2525).$$

Notez que nous évaluons $P(X \leq 3)$ puisque le pourcentage observé (1.5%) est inférieur au taux de défectuosité revendiqué. Nous ne rejetons pas H_0 à α puisque la valeur$-p$ est

$$2P(Z < -2525) = 0.794.$$

Q104. Les connaissances de dix ingénieurs sur les concepts statistiques de base ont été mesurées sur une échelle de 0 à 100, avant et après un court cours sur le contrôle statistique de la qualité. Les résultats sont les suivants :

Individu	1	2	3	4	5	6	7	8	9	10
Avant X_{1i}	43	82	77	39	51	66	55	61	79	43
Après X_{2i}	51	84	74	48	53	61	59	75	82	53

Soit μ_1 et μ_2 la note moyenne avant et après le cours. Effectuez le test $H_0 : \mu_1 = \mu_2$ envers $H_1 : \mu_1 < \mu_2$. Utilisez $\alpha = 0.05$. **Solution :** Les différences $D_i = X_{1i} - X_{2i}$ sont :

Individu	1	2	3	4	5	6	7	8	9	10
Avant X_{1i}	43	82	77	39	51	66	55	61	79	43
Après X_{2i}	51	84	74	48	53	61	59	75	82	53
Différence D_i	-8	-2	3	-9	-2	5	-4	-14	-3	-10

de sorte que $\overline{D} = -4.4$ et $S_D = 5.91$. Nous calculons la valeur$-p$ selon

$$P(\overline{D} \leq -4.4) = P\left(\frac{\overline{D}}{S_D/\sqrt{n}} \leq \frac{-4.4}{\sqrt{34.92/10}}\right) = P(t_9 \leq -2.35) = P(t_9 > 2.35) \in (0.02, 0.05).$$

Puisque la valeur$-p$ est plus petite que 0.05, nous rejetons H_0 en faveur de H_1 lorsque $\alpha = 0.05$, c-à-d que la note moyenne s'améliore après le cours.

Q105. Une entreprise utilise actuellement des tiges en alliage de titane qu'elle achète au fournisseur A. Un nouveau fournisseur B approche l'entreprise et propose des tiges de même qualité (du moins selon les dires du fournisseur) à un prix inférieur. L'entreprise est certainement intéressée par cette offre. En même temps, l'entreprise veut s'assurer que la sécurité de son produit n'est pas compromise. L'entreprise sélectionne au hasard dix tiges dans chacun des lots expédiés par les fournisseurs A et B et mesure les limites d'élasticité des tiges sélectionnées. La moyenne de l'échantillon et l'écart-type de l'échantillon observés sont respectivement de 651 MPa et 2 MPa pour les tiges du fournisseur A, et les mêmes paramètres sont de 657 MPa et 3 MPa pour les tiges du fournisseur B. Effectuez le test $H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$. Utilisez $\alpha = 0.05$. Supposez que les variances sont égales mais inconnues.

Solution : Il s'agit d'un test à deux échantillons : $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A \neq \mu_B$. Nous avons $\overline{x}_1 = 651$, $\overline{x}_2 = 657$, $s_1 = 2$, $s_2 = 3$. La différence observée entre les moyennes est $\overline{x}_1 - \overline{x}_2 = -6$. La statistique de test est

$$T_0 = \frac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2},$$

où S_p^2 est la **variance groupée** qui est calculée comme suit :

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 6.5.$$

Nous calculons la valeur de p comme suit

$$2P(\overline{x}_1 - \overline{x}_2 \leq -6) = 2P\left(\frac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq -5.26\right) = 2P(t_{18} < -5.26) = 2P(t_{18} > 5.26) < 2(0.0005) = 0.001.$$

Ce résultat est inférieur à $\alpha = 0.05$, nous rejetons donc H_0 en faveur de H_1 lorsque $\alpha = 0.05$.

---

106. On étudie la température de déflexion sous charge pour deux types différents de tuyaux en plastique. Deux échantillons aléatoires de 15 spécimens de tuyaux sont testés :

Type 1**:** 206, 188, 205, 187, 194, 193, 207, 185, 189, 213, 192, 210, 194, 178, 205.

Type 2**:** 177, 197, 206, 201, 180, 176, 185, 200, 197, 192, 198, 188, 189, 203, 192.

Les données confirment-elles l'affirmation selon laquelle la température de déflexion sous charge des tuyaux de type 1 dépasse celle des tuyaux de type 2 ? Calculez la valeur$-p$, en utilisant $\alpha = 0.05$, et énoncez votre conclusion.

Solution : nous testons $H_0 : \mu_1 = \mu_2$ vs. $H_0 : \mu_1 > \mu_2$. Nous avons $\overline{x}_1 = 196.4$, $\overline{x}_2 = 192.0667$, $s_1^2 = 109.8286$, $s_2^2 = 89.06667$, $n_1 = n_2 = 15$, et

$$s_p^2 = \frac{(15-1)109.8286 + (15-1)89.06667}{15+15-2} = 99.44762.$$

On se retrouve dans le 2e cas (σ_1^2, σ_2^2 inconnues, petites tailles) ; la statistique de test est

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1+n_2-2).$$

Sa valeur observée

$$t_0 = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = 1.19.$$

D'après la table T, nous obtenons $t_{0.05}(28) = 1.701$, d'où $t_0 < t_{0.05}(28)$ et nous ne rejetons pas H_0; il n'y a pas d'évidence que la température de déflexion des tuyaux de type 1 dépasse celle des tuyaux de type 2. La valeur$-p$ est

$$P(t(28) > 1.19) \in (0.1, 0.25),$$

puisque $P(t(28) > 1.313) = 0.1$ et $P(t(28) > 0.683) = 0.25$.

Q107. On prétend que 15% d'une certaine population est gauchère, mais un chercheur doute de cette affirmation. Il décide d'échantillonner au hasard 200 personnes et d'utiliser le petit nombre prévu pour fournir des preuves contre l'affirmation de 15%. Supposons que 22 des 200 soient gauchers. Calculez la valeur$-p$ associée à l'hypothèse (en supposant une loi binomiale), et donnez une interprétation.

Solution : nous supposons que la loi binomiale est appropriée. Soit X le nombre aléatoire (c-à-d avant observation) de gauchers dans l'échantillon, et soit p la proportion réelle de gauchers dans la population. Nous pouvons établir le test d'hypothèse formel comme suit :

- **modèle** – $X \sim \mathcal{B}(200.p)$, où p est la vraie proportion.
- $H_0 \colon p = 0.15$ (**affirmation**) vs. $H_1 \colon p < 0.15$ (**suspicion** : $12/200 = 11\%$)
- **preuve contre** H_0 – petites valeurs de X ; **valeur observée** : 22.
- **valeur**$-p$: $P(X \leq 22)$ si $X \sim \mathcal{B}(200.0.15)$ (c-à-d quand H_0 est valide) ; mais $P(X \leq 22) =$ $\texttt{pbinom}(22, 200.0.15) \approx 0.0645$. La petite valeur$-p$ fournit une preuve contre l'affirmation de 15%.

Q108. Une pédo-psychologue pense que la fréquentation de l'école maternelle améliore la perceptivité sociale (PS) des enfants. Elle utilise 8 paires de jumeaux, choisissant au hasard l'un d'entre eux pour aller à l'école maternelle et l'autre pour rester à la maison, et obtient ensuite des scores pour les 16 enfants. Dans 6 des 8 paires, le jumeau allant à l'école maternelle a obtenu de meilleurs résultats au test PS. Calculez la valeur de $p-$ associée à l'hypothèse (en supposant une loi binomiale), et fournissez une interprétation.

Solution :

- le modèle est $X \sim \mathcal{B}(8, p)$, où X est le nombre de paires pour lesquelles le jumeau scolarisé à l'école maternelle a obtenu de meilleurs résultats, et p est la probabilité réelle qu'un jumeau scolarisé à l'école maternelle obtienne de meilleurs résultats ;
- H_0: "la fréquentation de l'école maternelle n'a pas d'effet sur la PS" ($H_0 : p = 0.5$) vs. H_1: "la fréquentation de l'école maternelle améliore la PS" ($H_1 : p > 0.5$) ;
- si H_0 est valide, $X \sim \mathcal{B}(8, 0.5)$; si H_1 est valide, X aurait tendance à prendre **de plus grandes valeurs** que si H_0 ne l'était ; ainsi, des valeurs élevées de X fournissent de l'évidence contre H_0 (en faveur de H_1) ;
- la valeur$-p$ si H_0 est valide est $P(X \geq 6)$, $\quad X \sim \mathcal{B}(8, 0.5)$; la valeur$-p$ est
$$P(X \geq 6) = 1 - P(X \leq 5) = \texttt{pbinom}(6, 8, 0.5) = 0.1445.$$

Interprétation: s'il n'y avait pas d'effet réel, nous verrions 6+ améliorations sur 8 environ 14% du temps, simplement par hasard (ce qui est assez élevé, tout bien considéré). Les données ne fournissent pas de preuves convaincantes contre H_0 (aucun effet). Par conséquent, la pédo-psychiatre ne peut pas nous convaincre que la fréquentation de l'école maternelle améliore la PS.

Q109. On affirme que la résistance à la rupture du fil utilisé dans la fabrication des tissus pour rideaux suit une loi normale avec $\mu = 97$ psi et $\sigma = 2$ psi. Un échantillon aléatoire de neuf spécimens est testé et la force de rupture moyenne s'avère être $\overline{X} = 98$ psi. Formulez un test pour cette situation. Doit-il être unilatéral ou bilatéral ? Quelle valeur de α devriez-vous utiliser ? Quelle conclusion tirez-vous ?

Q110. A civil engineer is analyzing the compressive strength of concrete. It is claimed that its mean is 80 and variance is known to be 2. A random sample of size 60 yields the sample mean 59. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Q111. On prétend que la teneur en sucre du sirop des pêches en conserve suit une loi normale de moyenne 10 et de variance 2. Un échantillon aléatoire de $n = 10$ boîtes de conserve donne une moyenne de $\overline{x} = 11$. Un autre échantillon aléatoire de $n = 10$ boîtes de conserve donne une moyenne de $\overline{y} = 9$. Formulez un test pour cette situation. Doit-il être unilatéral ou bilatéral ? Quelle valeur de α devez-vous utiliser ? Quelle conclusion le test et l'échantillon permettent-ils de tirer ?

Q112. Une certaine alimentation électrique est censée fournir une tension de sortie constante de 10kV. Dix mesures sont effectuées et donnent une moyenne d'échantillon de 11kV. Formulez un test pour cette situation. Doit-il être unilatéral ou bilatéral ? Quelle valeur de α devez-vous utiliser ? Quelle conclusion le test et l'échantillon permettent-ils de tirer ?

Q113. La température moyenne de l'eau en aval du tuyau de décharge d'une tour de refroidissement d'une centrale électrique ne devrait pas dépasser 100F. L'expérience passée indique que l'écart-type est de 2F. La température de l'eau est mesurée pendant neuf jours choisis au hasard, et on constate que la température moyenne est de 98F. Formulez un test pour cette situation. Doit-il être unilatéral ou bilatéral ? Quelle valeur de α devez-vous utiliser ? Quelle conclusion le test et l'échantillon permettent-ils de tirer ?

Q114. Nous nous intéressons à la vitesse de combustion moyenne d'un propergol solide utilisé pour alimenter les systèmes d'évacuation des équipages d'avion. Nous voulons déterminer si le taux de combustion moyen est ou non de 50 cm/seconde. Un échantillon de 16 spécimens est testé et nous observons $\overline{X} = 48.5$. Supposez la normalité avec $\sigma = 2.5$.

Solution : nous testons pour $H_0 : \mu = 50$ vs. $H_1 : \mu \neq 50$. La valeur$-p$ est

$$2\cdot \min\left(P\left(Z \geq \frac{48.5-\mu_0}{\sigma/\sqrt{n}}\right), P\left(Z \leq \frac{48.5-\mu_0}{\sigma/\sqrt{n}}\right)\right) = 2\cdot P(Z \leq -2.4) \approx 2\cdot 0.0082 = 0.0164.$$

Nous ne rejetons pas H_0 lorsque $\alpha = 0.01$, mais nous le faisons lorsque $\alpha = 0.05$.

Q115. Dix personnes ont participé à un programme de modification de leur régime alimentaire afin de stimuler la perte de poids. Leur poids avant et après leur participation au programme est indiqué ci-dessous :

Avant	$195, 213, 247, 201, 187, 210, 215, 246, 294, 310$
Après	$187, 195, 221, 190, 175, 197, 199, 221, 278, 285$

Y a-t-il des preuves pour soutenir l'affirmation selon laquelle ce programme particulier de modification du régime alimentaire est efficace pour produire une réduction moyenne du poids ? Utilisez $\alpha = 0.05$. Calculez la valeur$-p$ associée.

Solution : il s'agit d'un test t apparié, et non d'un test à 2 échantillons. Nous calculons la différence après-avant :

D_i	$-8, -18, -26, -11, -12, -13, -16, -25, -16, -25$

Nous testons $H_0 : \mu_D = 0$ vs. $H_0 : \mu_D < 0$; nous avons $\overline{d} = -17$ et $s_D^2 = 41.11$. La statistique de test T_0 suit une loi $t(10-1)$ sous H_0. Sa valeur observée est

$$t_0 = \frac{\overline{d}}{s_D/\sqrt{10}} = -8.38.$$

La valeur$-p$ associée est $P(t(9) < -8.38) = P(t(9) > 8.38) < 0.0005$, et donc il y a suffisamment de preuves pour rejeter l'hypothèse (à $\alpha = 0.05$) que le régime ne réduit pas le poids.

Q116. Nous voulons tester l'hypothèse selon laquelle le contenu moyen des conteneurs d'un lubrifiant particulier est égal à 10L vs. l'alternative bilatérale. Le contenu d'un échantillon aléatoire de 10 conteneurs est de

10.2	9.7	10.1	10.3	10.1
9.8	9.9	10.4	10.3	9.5

Trouvez la valeur$-p$ de ce test bilatéral. Supposez que le volume suit une loi normale. Notez que $\sum_{i=1}^{10} x_i^2 = 1006,79$, si x_i représentent les mesures.

a) $0.05 < p < 0.10$ b) $0.10 < p < 0.20$ c) $0.25 < p < 0.40$

d) $0.50 < p < 0.80$ e) N/A

Solution : nous testons pour $H_0 : \mu = 10$ vs. $H_1 : \mu \neq 10$. Nous avons $\overline{x} = 10.03$ et $s^2 = 0.08678$. La valeur observée de la statistique de test est

$$t_0 = \frac{\overline{x} - 10}{s/\sqrt{n}} = \frac{10.03 - 10}{\sqrt{0.08678}/\sqrt{10}} = 0.322.$$

Il y a 10 observations, donc nous utilisons $\nu = 10 - 1 = 9$ degrés de liberté pour le test bilatéral (cf. le tableau t). La valeur$-p$ est donc $P(T(9) > 0.322)$, qui se situe entre 0.25 et 0.40 puisque $P(T(9) > 0.703) = 0.25$ et $P(T(9) > 0.261) = 0.40$. Il suffit de multipliez par deux pour obtenir la réponse.

Q117. Un ingénieur mesure le poids de $n = 25$ pièces d'acier, qui suit une loi normale avec une variance de 16. Le poids moyen de l'échantillon est $\overline{X} = 6$. Il veut tester $H_0 : \mu = 5$ vs. $H_1 : \mu > 5$. Quelle est la valeur$-p$ pour ce test ?

a) 0.05000 b) 0.10565 c) 0.89435 d) 1.0000 e) N/A

Solution : sous H_0, nous avons

$$P(\overline{X} > 6|\mu = 5) = P\left(Z > \frac{6-5}{4/5}\right) = P(Z > 1.25)$$
$$= 1 - 0.89435 = 0.10565.$$

Q118. On pense que l'épaisseur d'un film plastique (en mm) sur un matériau est influencée par la température à laquelle le revêtement est appliqué. Une expérience aléatoire est réalisée ; 11 items sont enduits à 125F, ce qui donne une épaisseur moyenne de $\overline{x}_1 = 103.5$ et un écart-type de $s_1 = 10.2$. Un autre 11 items sont revêtus à 150F, pour lesquels on observe $\overline{x}_2 = 99.7$ et $s_2 = 11.7$. Supposez que les variances sont inconnues mais égales. Nous voulons tester l'égalité des moyennes par rapport à l'alternative bilatérale. La valeur des statistiques de test appropriées et la décision sont les suivantes (for $\alpha = 0.05$) :

a) 0.81; nous rejetons H_0.

b) 0.81; nous ne rejetons pas H_0.

c) 1.81; nous rejetons H_0.

d) 1.81; nous ne rejetons pas H_0.

e) N/A

Solution : c'est un test à 2 échantillons, de petites tailles et variances inconnues, d'où le besoin de calculer la variance mixte

$$s_p^2 = \frac{(11-1)10.2^2 + (11-1)11.7^2}{11+11-2} = 120.465,$$

or $s_p = 10.97$. La valeur observée du test de la statistique est

$$t_0 = \frac{\overline{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{103.5 - 99.7}{10.97\sqrt{1/101 + 1/11}} = 0.81.$$

Puisque $t_{0.05/2}(11+11-2) = 2.086 > t_0$, nous ne rejetons H_0.

Q119. On produit ce qui suit avec la commande `t.test` de `R`.

```
One Sample t-test
data:  x
t = 2.0128, df = 99, p-value = 0.02342
alternative hypothesis: true mean is greater than 0
```

Sur la base de ce résultat, quelle affirmation est correcte ?

a) Si l'erreur de type I est de 0.05, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu > 0$. **(c'est celle-ci !)**

b) Si l'erreur de type I est de 0.05, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu \neq 0$.

c) Si l'erreur de type I est de 0.01, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu > 0$.

d) Si l'erreur de type I est de 0.01, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu < 0$.

e) L'erreur de type I est de 0.02342.

Q120. Une société pharmaceutique prétend qu'un médicament diminue la pression artérielle. Un médecin doute de cette affirmation. Il teste 10 patients et enregistre les résultats avant et après le traitement médicamenteux :

```
> Before=c(140,135,122,150,126,138,141,155,128,130)
> After=c(135,136,120,148,122,136,140,153,120,128)
```

Il utilise `R` :

```
> test.t(Before,After,alternative="greater")
    data:  Before and After
    t = 0.5499, p-value = 0.2946
    alternative hypothesis: true difference in means is greater than 0
    sample estimates: mean of x mean of y
        136.5     133.8
```

Son assistant affirme que la commande devrait plutôt être :

```
> test.t(Before,After,paired=TRUE,alternative="greater")

    data: Before and After t = 3.4825, df = 9, p-value = 0.003456
    alternative hypothesis: true difference in means is greater than 0
    sample estimates: mean of the differences
        2.7
```

Quelle est la meilleure réponse ?

a) L'assistant utilise la bonne commande. Il n'y a **pas suffisamment** de preuves afin de justifier l'affirmation que le nouveau médicament diminue la pression artérielle.
b) L'assistant utilise la bonne correcte. Il y a des preuves **suffisantes** afin de justifier l'affirmation que le nouveau médicament diminue la pression artérielle pour tout choix raisonnable de α. **(c'est celle-ci)**
c) Le médecin utilise la bonne commande. Il n'y a **pas suffisamment** de preuves afin de justifier l'affirmation que le nouveau médicament diminue la pression artérielle.
d) Le médecin utilise la bonne commande. Il y a des preuves **suffisantes** afin de justifier l'affirmation que le nouveau médicament diminue la pression artérielle pour tout choix raisonnable de α.
e) Personne n'a raison ; on ne devrait pas utiliser de tests t dans ce contexte.

Q121. Une entreprise affirme que la déviation moyenne d'une pièce d'acier de 10 cm de long est égale à 0.012 cm. Un acheteur soupçonne qu'elle est supérieure à 0.012 cm. Les données suivantes x_i ont été recueillies :

```
0.0132 0.0138 0.0108 0.0126 0.0136 0.0112 0.0124 0.0116 0.0127 0.0131
```

En supposant la normalité et que $\sum_{i=1}^{10} x_i^2 = 0.0016$, quelles sont la valeur$-p$ pour le test unilatéral approprié et la décision correspondante ?

a) $p \in (0.05, 0.1)$ et H_0 rejeté à $\alpha = 0.05$.
b) $p \in (0.05, 0.1)$ et H_0 pas rejeté à $\alpha = 0.05$.
c) $p \in (0.1, 0.25)$ et H_0 rejeté à $\alpha = 0.05$.
d) $p \in (0.1, 0.25)$ et H_0 pas rejeté à $\alpha = 0.05$.
e) N/A

Solution : nous testons $H_0 : \mu = 0.012$ vs. $H_1 : \mu > 0.012$. Comme la variance de la population sous-jacente est inconnue, nous utiliserons le test $t-$ unilatéral. La variance estimée de l'échantillon est

$$S^2 = \frac{1}{10-1}\left(\sum_{i=1}^{10} x_i^2 - \frac{1}{10}\left(\sum_{i=1}^{10} x_i\right)^2\right) = 0.00000102.$$

La moyenne observée est $\overline{x} = 0.0125$. Nous calculons la valeur$-p$ correspondante :

$$P(\overline{X} > \overline{x}) = P\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > \frac{\overline{x}-\mu_0}{S/\sqrt{n}}\right) = P\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > \frac{0.0125-0.012}{\sqrt{0.00000102/10}}\right)$$
$$= P(t(10-1) > 1.5638) = P(t(9) > 1.5638) \in (0.05, 0.1)$$

et nous ne rejetons pas H_0 à $\alpha = 0.05$.

Q122. Afin de comparer la durabilité de deux types différents de papier sablé, 10 spécimens de type A ont été soumis à un traitement par une machine qui mesure l'usure abrasive ; 11 spécimnes de type B ont été soumis au même traitement. On obtient les observations suivantes :

```
xA 27 26 24 29 30 26 27 23 28 27
xB 24 23 22 27 24 21 24 25 24 23 20
```

Notez que $\sum x_{A,i} = 267$, $\sum x_{B,i} = 257$, $\sum x^2_{A,i} = 7169$, $\sum x^2_{B,i} = 6041$. En supposant la normalité et l'égalité des variances de l'usure abrasive, nous voulons tester l'égalité de l'usure abrasive moyenne pour les deux types. La valeur$-p$ appropriée est

a) $p < 0.01$ b) $p > 0.2$ c) $p \in (0.01, 0.05)$

d) $p \in (0.1, 0.2)$ e) $p \in (0.05, 0.1)$ f) N/A

Solution : il s'agit d'un test à deux échantillons. Nous testons $H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$. Nous calculons $s^2_A = 4,45$, $s^2_B = 3,65$, $s^2_p = 4,03$, $\overline{x}_A = 26,71$, $\overline{x}_B = 23,26$. La valeur$-p$ est

$$\begin{aligned}
2P(\overline{X}_A - \overline{X}_B > \overline{x}_A - \overline{x}_B) &= 2P\left(\frac{\overline{X}_A - \overline{X}_B}{S_p\sqrt{1/n_A + 1/n_B}} > \frac{3.34}{\sqrt{4.03}\sqrt{1/10 + 1/11}}\right) \\
&= 2P(t(10 + 11 - 2) > 3.8037) \\
&= 2P(t(19) > 3.8037) < 0.01,
\end{aligned}$$

puisque $P(t(19) > 3.8037) < 0.005$.

Q123. Le résultat suivant a été produit avec la commande `t.test` de `R`.

```
One Sample t-test
data:  x
t = 32.9198, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
```

D'après ce résultat, quelle affirmation est correcte ?

a) Si l'erreur de type I est de 0.05, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu > 0$.

b) Si l'erreur de type I est de 0.05, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu \neq 0$. **(c'est celle-ci !)**

c) Si l'erreur de type I est de 0.01, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu > 0$.

d) Si l'erreur de type I est de 0.01, nous rejetons $H_0 : \mu = 0$ en faveur de $H_1 : \mu < 0$.

e) N/A

Q124. Considérons un échantillon $\{X_1, \ldots, X_{10}\}$ provenant d'une population normale $X_i \sim \mathcal{N}(4, 9)$. Soient $\overline{X}$ et S^2 la moyenne de l'échantillon et la variance de l'échantillon, respectivement. Trouvez la valeur de c telle que

$$P\left(\frac{\overline{X} - 4}{S/\sqrt{10}} \leq c\right) = 0.99$$

a) 1.833 b) 2.326 c) 1.645 d) 2.821 e) N/A

Solution : c'est équivalent à trouver c tel que

$$P\left(\frac{\overline{X}-4}{S/\sqrt{10}} \geq c\right) = 0.01.$$

On sait que $\frac{\overline{X}-4}{S/\sqrt{10}} \sim t(10-1)$. D'après le tableau, nous avons

$$P(t(9) > 2.821) = 0.01,$$

d'où $c = 2.821$.

Q125. Considérez l'ensemble de données suivant :

```
2.6  3.7  0.8  9.6  5.8 -0.8  0.7 0.6
4.8  1.2  3.3  5.0  3.7  0.1 -3.1 0.3
```

La médiane et l'écart interquartile de l'échantillon sont, respectivement :

a) 2.4, 3.3 b) 1.9, 3.8 (*) c) 1.9, 1.8 d) 2.9, 12.2 e) N/A

Q126. Un article paru dans *Computers and Electrical Engineering* s'est penché sur l'accélération des réseaux neuronaux cellulaires (CNN) pour une architecture de calcul parallèle polyvalente. Diverses accélérations sont observées :

```
3.77  3.35  4.21  4.03  4.03  4.63
4.63  4.13  4.39  4.84  4.26  4.60
```

Supposons que la population soit normalement distribuée. Quel est le C.I. de l'accélération moyenne à 99% ?

a) $[4.155, 4.323]$ b) $[3.863, 4.615]$ c) $[4.040, 4.438]$ d) $[3.77, 4.60]$ e) N/A

Solution : on utilise `R`.

```
> x=c(3.77,3.35,4.21,4.03,4.03,4.63,4.63,4.13,4.39,4.84,4.26,4.60);
> alpha=0.01
> n=length(x)
> mean(x)-qt(1-alpha/2,n-1)*sd(x)/sqrt(n)
    3.863531
> mean(x)+qt(1-alpha/2,n-1)*sd(x)/sqrt(n)
    4.614802
```

La fonction `qt(beta,nu)` trouve le quantile β de la loi de Student avec ν degrés de liberté ; elle joue un rôle semblable à celui de la fonction `qnorm()`.

Q127. Un ingénieur mesure le poids de $n = 25$ pièces d'acier, qui suivent une loi normale avec une variance de 16. Le poids moyen observé pour l'échantillon est $\overline{x} = 6$. Donnez un I.C. à environ 95% de la moyenne μ.

a) $[-0.272, 12.272]$ b) $[4.432, 7.568]$ c) $[3.250, 8.750]$ d) $[4.120, 7.522]$ e) N/A

Solution : puisque les poids suivent une loi normale avec une variance connue, l'I.C. est

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6 \pm 1.96\frac{4}{5} = [4.432, 7.568].$$

Q128. Supposons que les v.a. $\{X_1, \ldots, X_8\}$ suivent une loi normale de moyenne 2 et de variance 24. Indépendamment, supposez que les v.a. $\{Y_1, \ldots, X_{16}\}$ suivent une loi normale de moyenne 1 et de variance 16. Soient $\overline{X}$ et $\overline{Y}$ les moyennes d'échantillon correspondantes. Quelle valeur prend $P(\overline{X} + \overline{Y} > 4)$?

a) 0.7721 b) 0.30855 c) 0.69165 d) 0.9883 e) N/A

Solution : puisque les X_i et les Y_j sont indépendants, nous avons

$$\overline{X} + \overline{Y} \sim \mathcal{N}\left(2+1, \frac{24}{8} + \frac{16}{16}\right) = \mathcal{N}(3, 4).$$

Ainsi,

$$P(\overline{X} + \overline{Y} > 4) = P\left(\frac{\overline{X} + \overline{Y} - 3}{\sqrt{4}} > \frac{4-3}{\sqrt{4}}\right) = P(Z > 0.5) = 1 - P(Z < 0.5) = 1 - \Phi(0.5) \approx 1 - 0.6915 = 0.3085.$$

Q129. Une équipe médicale veut tester si un médicament particulier diminue la pression artérielle diastolique. Neuf personnes ont été testées. L'équipe a mesuré la pression artérielle avant (X) et après (Y) l'application du médicament. Les moyennes correspondantes étaient $\overline{X} = 91$, $\overline{Y} = 87$. La variance d'échantillon des différences était $S_D^2 = 25$. La valeur$-p$ pour le test unilatéral approprié se situe entre :

a) 0 et 0.025 b) 0.025 et 0.05 c) 0.05 et 0.1

d) 0.1 et 0.25 e) 0.25 et 1 f) N/A

Solution : il s'agit d'un test t unilatéral apparié, $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X > \mu_Y$. La différence observée des moyennes est $\overline{d} = 4$. La valeur$-p$ associée est la suivante :

$$P(\overline{D} \geq \overline{d}) = P(\overline{D} \geq 4) = P\left(\frac{\overline{D}}{S_D/\sqrt{n}} \geq \frac{4}{5/3}\right) = P(t(n-1) > 2.4) = P(t(8) > 2.4) < 0.025,$$

puisque $P(t(8) > 2.4) \in (0.01, 0.025)$ selon la table.

Q130. Une chercheuse étudie une différence entre deux langages de programmation. Elle a demandé à douze experts connaissant les deux langages d'écrire un code pour une fonction particulière en utilisant les deux langages et le temps d'écriture de ces codes a été enregistré. Les observations sont les suivantes.

Expert	01	02	03	04	05	06	07	08	09	10	11	12
Lang 1	17	16	21	14	18	24	16	14	21	23	13	18
Lang 2	18	14	19	11	23	21	10	13	19	24	15	29

Construisez un I.C. à environ 95 % de la différence moyenne entre la première et la deuxième langue. Avons-nous la preuve qu'une des langues est préférable à l'autre (c'est-à-dire que le temps moyen pour écrire une fonction est plus court) ?

Solution :

Q130 : $n=12$

L1:	17	16	21	14	18	24	16	14	21	23	13	18
L2:	18	14	19	11	23	21	10	13	19	24	15	29
D:	-1	+2	+2	3	-5	3	6	1	2	-1	-2	-11

Because Both L1 & L2 are not independent of on each other so, we use two-Sample test (paired)

We test for $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{12} D_i = \frac{1}{12}[-1 + \cdots + (-11)] = -\frac{1}{12}$$

$$S_D^2 = \frac{1}{n-1}\sum (D_i - \bar{D})^2 = \frac{1}{11}\left[\sum_{i=1}^{12} D_i^2 - n\bar{D}^2\right] = \frac{1}{11}\left[219 - 12 \cdot \frac{1}{144}\right] = 19.9$$

$$S_D = \sqrt{19.9} = 4.46$$

$$\therefore\ t_0 = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{-\frac{1}{12} \cdot \sqrt{12}}{4.46} = 0.065 \sim t(\overset{11}{n-1})$$

$$2P(T_{11} > 0.065) \doteq 2 \cdot (0.20, 0.30) \doteq [0.40, 0.60] > \cancel{T_{0.05/2}(11)}\ \alpha = 0.05$$

$\alpha = 0.05$

$\therefore$ we fail to reject to H_0.

$$T_{\alpha/2}(11) = T_{0.025}(11) = 2.201$$

$$\cancel{\bar{D} \pm 2.201 \frac{S_D}{S_D/\sqrt{n}} = -\frac{1}{12} \pm 2.201 \times}$$

$$\bar{D} \pm 2.201 \frac{S_D}{\sqrt{n}} = -\frac{1}{12} \pm 2.201 \times \frac{4.46}{\sqrt{12}} = -\frac{1}{12} \pm 2.8249 = [-2.9082,\ 2.7415]$$

no evidence that any of them is better.

Q131. Dans un ensemble de 12 paires d'observations (x_i, y_i) provenant d'une expérience, on obtient les sommes suivantes :

$$\sum_{i=1}^{12} x_i = 25,\ \sum_{i=1}^{12} y_i = 432,\ \sum_{i=1}^{12} x_i^2 = 59,\ \sum_{i=1}^{12} x_i y_i = 880.5,\ \sum_{i=1}^{12} y_i^2 = 15648.$$

Quelle est la valeur estimée de y obtenue à partir de l'approche des moindres carrés lorsque $x = 5$?

a) 27.78 b) 47.77 c) 41.87 d) 55.97 e) N/A

Solution : en supposant que le modèle de régression linéaire est justifié, la valeur estimée à $x = 5$ est donnée par

$$\hat{y}(5) = b_0 + b_1(5).$$

Nous avons

$$\begin{aligned}
\overline{x} &= \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{25}{12}, \quad \overline{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = 36 \\
S_{xx} &= \sum_{i=1}^{12} x_i^2 - 12\overline{x}^2 = \frac{83}{12}, \quad S_{xy} = \sum_{i=1}^{12} x_i y_i - 12\overline{xy} = -19.5, \\
b_1 &= -\frac{19.5}{83/12} = -2.82, \quad b_0 = 36 - (-2.82)(25/12) = 41.87, \\
\hat{y}(5) &= 41.87 - 2.82(5) = 27.78.
\end{aligned}$$

Q132. En supposant que le modèle de régression linéaire simple $y = \beta_0 + \beta_1 x + \varepsilon$ est approprié pour $n = 14$ observations, la droite de régression estimée est

$$\hat{y} = 0.66490 + 0.83075x.$$

Étant donné que $S_{yy} = 4.1289$ et $S_{xy} = 4.49094$, calculez l'erreur-type de la pente de régression.

a) 0.3176 b) 0.0783 c) 0.0855 d) 0.0073 e) N/A

Solution : l'erreur-type de la pente de la régression est

$$\text{se}(b_1) = \sqrt{\hat{\sigma}^2/S_{xx}}.$$

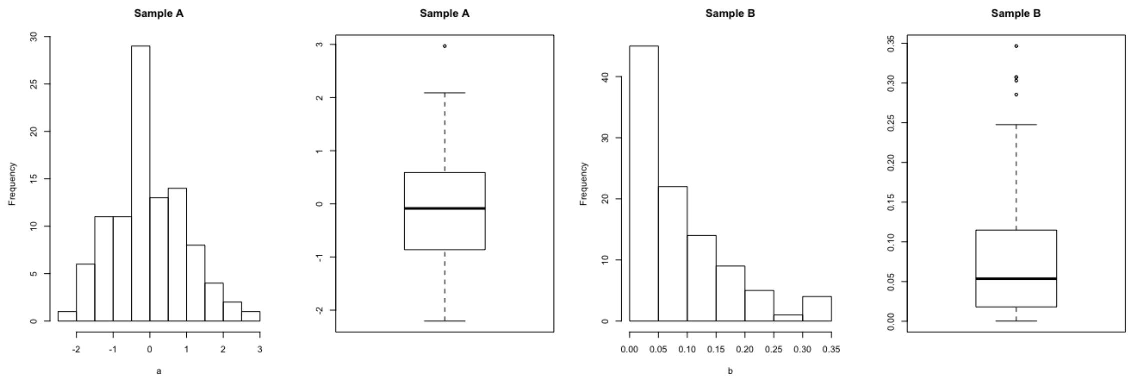
Mais

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} \quad \text{et} \quad S_{xx} = \frac{S_{xy}}{b_1},$$

de sorte que

$$\begin{aligned}
\text{se}(b_1) &= \sqrt{\frac{b_1(S_{yy} - b_1 S_{xy})}{(n-2)S_{xy}}} = \sqrt{\frac{0.83075(4.1289 - 0.83075 \cdot 4.49094)}{(14-2)4.49094}} \\
&= 0.07833.
\end{aligned}$$

Q136. Les graphiques suivants présentent un histogramme et un boxplot pour deux échantillons, A et B.



Sur la base de ces graphiques, nous pouvons conclure que :

a) seule A provient d'une population normale

b) seule B provient d'une population normale

c) et A, et B proviennent de populations normales

Solution : il serait raisonnable de s'attendre à ce que A provienne d'une population normale, mais l'asymétrie de B signifie qu'il ne provient sans doute pas d'une telle population.

Q137. Considérons un ensemble de données de $n = 25$ observations apariées (x_i, y_i) pour lequel

$$\sum_{i=1}^{n} x_i = 325.000, \; \sum_{i=1}^{n} y_i = 658.972,$$
$$\sum_{i=1}^{n} x_i^2 = 5525.000, \; \sum_{i=1}^{n} x_i y_i = 11153.588, \; \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Notons que $t_{0.05/2}(23) = 2.069$. Quel est l'estimation de la pente de la droite de régression ?

a) 1.99 b) -1.99 c) 0.49 d) 0.59 e) N/A

Solution : nous avons $b_1 = \frac{S_{xy}}{S_{xx}}$ et

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 13, \; \overline{y} = \frac{1}{25}\sum_{i=1}^{25} y_i = 26.359, \; S_{xx} = \sum_{i=1}^{25} x_i^2 - 25\overline{x}^2 = 1300$$
$$S_{xy} = \sum_{i=1}^{25} x_i y_i - 25\overline{xy} = 2586.952, \; S_{yy} = \sum_{i=1}^{25} y_i^2 - 25\overline{y}^2 = 5261.613,$$
$$b_1 = \frac{2586.952}{1300} = 1.99, \quad b_0 = 26.359 - (1.990)(13) = 0.49.$$
$$\hat{y}(30) = 0.49 + 1.99(30) = 60.19.$$

Q138. Considérons un ensemble de données de $n = 25$ observations apariées (x_i, y_i) pour lequel

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$
$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Notons que $t_{0.05/2}(23) = 2.069$. Quel est l'estimation de l'ordonnée à l'origine de la droite de régression ?

a) 1.99 b) -1.99 c) 0.49 d) 0.59 e) N/A

Solution : consulter la réponse à la question précé dente.

Q139. Considérons un ensemble de données de $n = 25$ observations apariées (x_i, y_i) pour lequel

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$
$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Notons que $t_{0.05/2}(23) = 2.069$. Quel est la prédiction de la réponse y lorsque $x = 30$?

a) 60.19 b) 16.67 c) 30 d) 30.54 e) N/A

Solution : onsulter la réponse à la question précé dente.

Q140. Considérons un ensemble de données de $n = 25$ observations apariées (x_i, y_i) pour lequel

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$
$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Notons que $t_{0.05/2}(23) = 2.069$. Is the La régression linéaire est-elle significative ?

Solution : nous testons $H_0 : \beta_1 = 0$ versus $H_0 : \beta_1 \neq 0$; utilisons $\alpha = 0.05$. Si nous rejetons H_0 en faveur de H_1, alors l'évidence suggère qu'il y a une relation linéaire entre X et Y.

Si H_0 est valide, la statistique de test $T_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \sim t_{0.05/2}(23)$. À l'aide des résultats obtenus à la réponse de **Q137**, nous obtenons

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{5261.613 - 1.99 \cdot 2586.952}{23} = 4.94;$$

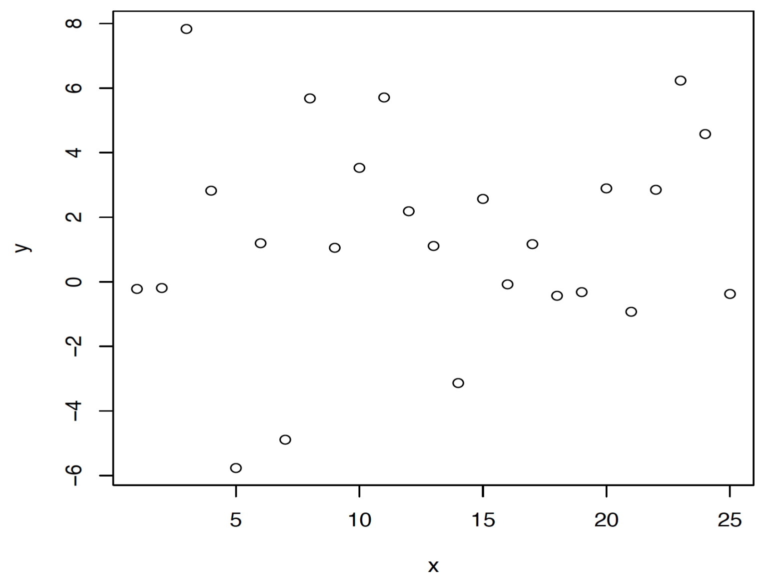
la valeur observée est ainsi

$$t_0 = \frac{1.99}{\sqrt{4.94/1300}} = 32.27 < t_{0.05/2}(23) = 2.069.$$

Nous rejetons alors H_0 en faveur d'une régression significative.

Q141. Pour les données suivantes, le coefficient de corrélation est le plus susceptible d'être ...

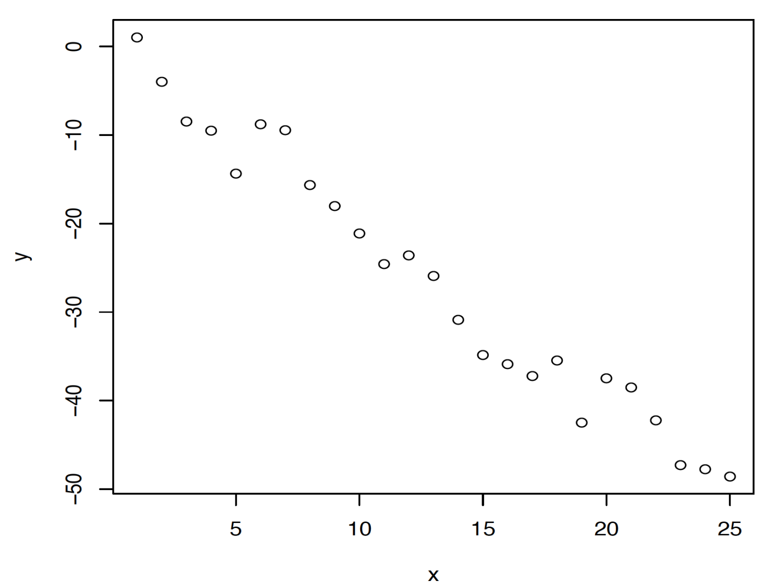
a) 0.01 b) 0.98 c) -0.5 d) -0.98



Solution : le nuage de points ne montre aucune structure ou relation réelle entre x et y. La réponse la plus probable est $\rho = 0.01$.

Q142. Pour les données suivantes, le coefficient de corrélation est le plus susceptible d'être ...

a) 0.01 b) 0.98 c) -0.5 d) -0.98



Solution : le nuage de points montre un schéma clair d'anti-corrélation entre x et y – lorsque x augmente, y diminue et vice-versa. La valeur la plus probable est $\rho = -0.98$.

Q143. Une entreprise emploie 10 chauffeurs à temps partiel pour sa flotte de camions. La directrice souhaite trouver une relation entre le nombre de km parcourus (X) et le nombre de jours de travail (Y) dans une semaine typique. Les chauffeurs sont engagés pour conduire des demi-journées, de sorte que $X = 3.5$ correspond à 7 demi-journées. La directrice veut utiliser le modèle de régression linéaire $Y = \beta_0 + \beta_1 x + \epsilon$ avec les données suivantes :

	1	2	3	4	5	6	7	8	9	10
x	825	215	1070	550	480	920	1350	325	670	1215
y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

Notez que $\sum x_i^2 = 7104300$, $\sum y_i^2 = 99.75$, et $\sum x_i y_i = 26370$. Déterminez la droite de meilleur ajustement.

Solution : nous avons

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 1297860 \quad \text{et} \quad S_{xy} = 4653,$$

de sorte que

$$b_1 = S_{xy}/S_{xx} = 0.0036$$

et

$$b_0 = \sum_{i=1}^{n} y_i/n - b_1 \sum_{i=1}^{n} x_i = 0.1181;$$

la droite recherchée est ainsi $\hat{y} = 0.1181 + 0.0036x$.

Q144. En utilisant les données de la question **Q143**, de quelle valeur le coefficient de corrélation entre x et y est-il le plus proche ?

a) 0.437 b) 0.949 c) 0.113 d) 1.123 e) N/A

Solution : comme à la question **Q143**, nous avons $S_{xx} = 12978600$ et $S_{xy} = 4653$. Nous avons de plus

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 18.525,$$

d'où le coefficient de corrélation est

$$\rho_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{4653}{\sqrt{18.525 \cdot 1297860}} \approx 0.949.$$

Q145. Nous cherchons à déterminer si la régression de la question **Q143** est significative ou non en testant $H_0 : \beta_1 = 0$ par rapport à $H_1 : \beta_1 \neq 0$. La valeur de la statistique appropriée et la décision correspondante lorsque $\alpha = 0.05$ sont :

a) 8.55; nous ne rejetons pas H_0

b) 2.31; nous rejetons H_0

c) 8.55; nous rejetons H_0

d) 2.31; nous ne rejetons pas H_0

e) N/A

Solution : l'estimation de la variance est

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{1.8434}{8} = 0.23.$$

Par conséquent, la statistique de test observée est

$$t_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{0.0036}{\sqrt{0.23/1297860}} = 8.551701.$$

Puisque $t_{0.05/2}(n-2) = t_{0.025}(8) = 2.306$, nous rejetons H_0.

Q146. On utilise des méthodes de régression afin d'analyser les données d'une étude portant sur la relation entre la température de surface de la chaussée en F (x) et la déviation de la chaussée (y). Les quantités probantes résumées sont : $n = 20$,

$$\sum y_i = 12.75, \sum y_i^2 = 8.86, \sum x_i = 1478 \sum x_i^2 = 143,215.8 \sum x_i y_i = 1083.67.$$

a) Calculez les estimations par les moindres carrés de la pente et de l'ordonnée à l'origine. Donnez un estimé de σ^2.

b) Utilisez l'équation de la droite ajustée afin de prédire la déviation de la chaussée qui serait observée si la température de surface est de 90F.

c) Donnez une estimation ponctuelle de la déviation moyenne de la chaussée lorsque la température de la surface est de 85F.

d) Quel changement dans la déflexion moyenne de la chaussée pourrait-on attendre d'un changement de 1F dans la température de surface ?

Solution :

a) Nous avons

$$b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \overline{y} - b_1 \overline{x}, \quad \hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2},$$

où

$$\begin{aligned} S_{xy} &= \sum x_i y_i - \tfrac{1}{n}(\sum x_i)(\sum y_i) = 141.445 \\ S_{xx} &= \sum x_i^2 - \tfrac{1}{n}(\sum x_i)^2 = 33991.6 \\ S_{yy} &= \sum y_i^2 - \tfrac{1}{n}(\sum y_i)^2 = 0.731875, \end{aligned}$$

de sorte que $b_1 = 0.00416$, $b_0 = 0.32999$, et $\hat{\sigma}^2 = 0.00797$

b) $\hat{y}(90) = b_0 + b_1 \cdot 90 = 0.70$

c) La question peut être reformulée comme suit : "Utilisez l'équation de la ligne ajustée pour prédire quelle déviation de la chaussée serait observée lorsque la température de surface est de 85F", c'est-à-dire : $\hat{y}(85) = b_0 + b_1 \cdot 85 = 0.68$.

d) C'est tout simplement la pente $b_1 = 0.00416$

Q147. Considérez les données de la question **Q146**.

a) Testez pour la signification de la régression lorsque $\alpha = 0.05$. Trouvez la valeur$-p$ de ce test. Quelle conclusion pouvez-vous en tirer ?

b) Donnez une estimation de l'erreur-type de la pente et de l'ordonnée à l'origine.

Solution :

a) Nous testons $H_0 : \beta_1 = 0$ par rapport à $H_1 : \beta_1 \neq 0$. La statistique de test est $T_0 = \frac{b_1 - 0}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}}$; sa valeur observée est $t_0 = \frac{b_1-0}{\sqrt{\hat{\sigma}^2/S_{xx}}} = 8.6$. La valeur$-p$ correspondante est $2P(t_{18} > 8.6) < 0.001$ (à l'aide du tableau), et nous rejetons alors H_0 en faveur d'une relation linéaire entre x et y.

b) Les erreurs-types sont

$$\mathrm{se}(b_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \quad \mathrm{se}(b_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}.$$

Ainsi, $\mathrm{se}(b_1) = 0.00048$, $\mathrm{se}(b_0) = 0.04098$.

Q148. Résolvez cette question en utilisant `R`.

a) Obtenez un échantillon `x` de taille $n = 100$ à partir d'une loi normale.

b) Définissez `y=1+2*x+rnorm(100)`.

c) Tracez un diagramme de dispersion (nuage de points).

d) Trouvez les estimateurs des paramètres de régression et ajoutez la ligne au nuage de points.

f) Calculez le coefficient de corrélation.

g) Tracez les résidus.

h) Offrez un commentaire au sujet de vos résultats.

Solution : le code suivant fera l'affaire.

```
> library(ggplot2) ## required for plotting
> set.seed(1234) ## so we all get the same results

# a), b), c)
> x = rnorm(100, mean = -10,sd=3)
> y = 1 + 2*x + rnorm(100)
> data.Q148 = data.frame(x,y)
> ggplot(data.Q148) + geom_point(aes(x=x, y=y)) +
    theme_bw()

# d)
> model <- lm(y ~ x, data=data.Q148)
> summary(model)
    Call:
    lm(formula = y ~ x, data = data.Q148)
```

```
    Residuals:
         Min       1Q   Median       3Q      Max
    -2.88626 -0.61401  0.00236  0.58645  2.98774

    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  0.95020    0.37674   2.522   0.0133 *
    x            1.99131    0.03459  57.566   <2e-16 ***
    ---
    Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Residual standard error: 1.037 on 98 degrees of freedom
    Multiple R-squared:  0.9713,Adjusted R-squared:  0.971
    F-statistic:  3314 on 1 and 98 DF,  p-value: < 2.2e-16

> ggplot(model) + geom_point(aes(x=x, y=y)) +
    geom_line(aes(x=x, y=.fitted), color="blue" ) + theme_bw()

# e)
> Sxy=sum((x-mean(x))*(y-mean(y)))
> Sxx=sum((x-mean(x))^2)
> Syy=sum((y-mean(y))^2)
> rho=Sxy/(sqrt(Sxx*Syy))
> rho
    0.9855334

# f)
> ggplot(model) + geom_point(aes(x=x, y=y)) +   ### plotting residuals
    geom_line(aes(x=x, y=.fitted), color="blue" ) +
    geom_linerange(aes(x=x, ymin=.fitted, ymax=y), color="red") +
    theme_bw()

> ggplot(model) +
    geom_point(aes(x=.fitted, y=.resid)) + theme_bw()
```

Les graphiques correspondants sont présentés aux pages suivantes. Vous pourriez obtenir des résultats légèrement différents si vous utilisez une loi normale différente pour générer x.

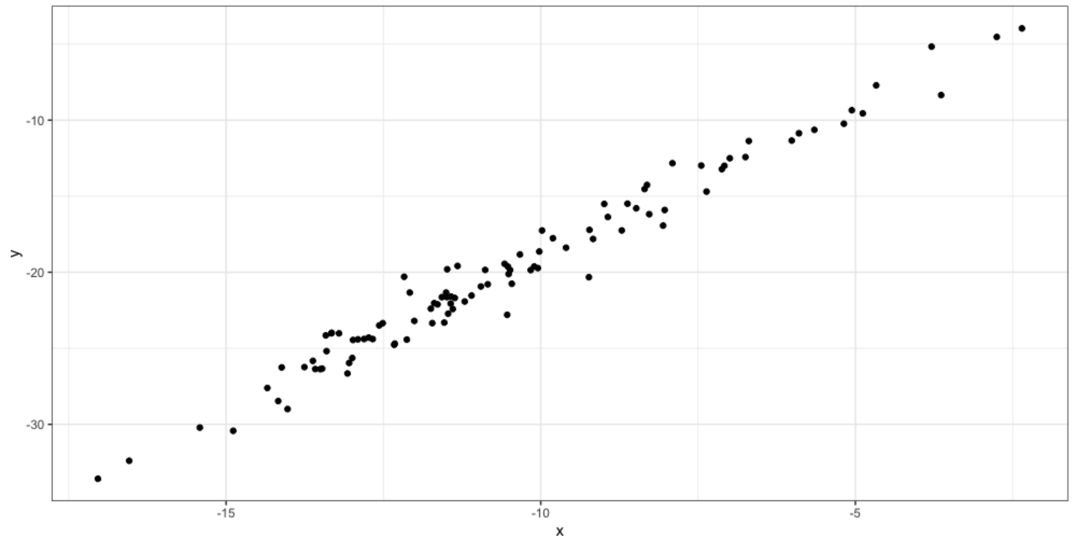
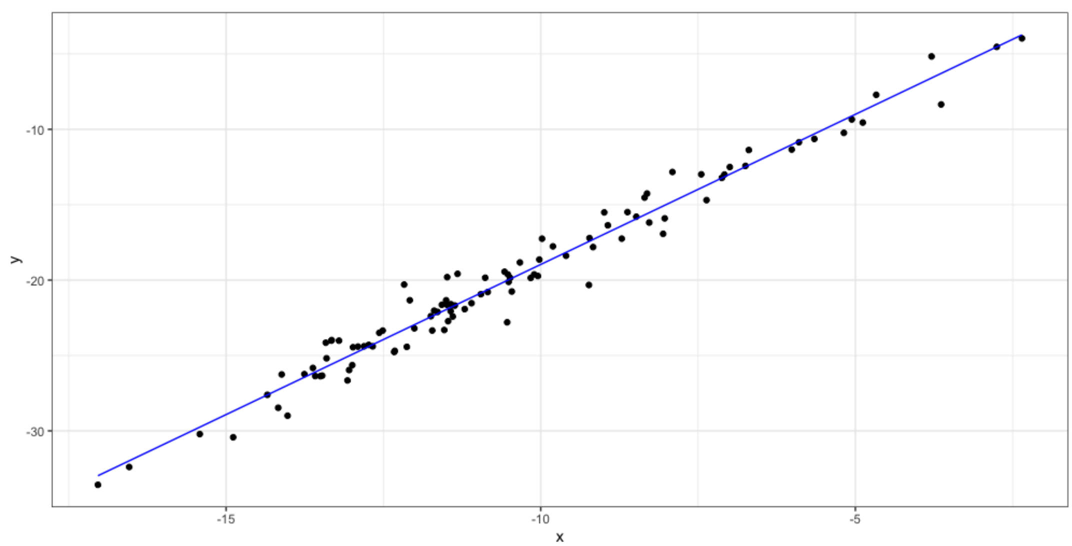
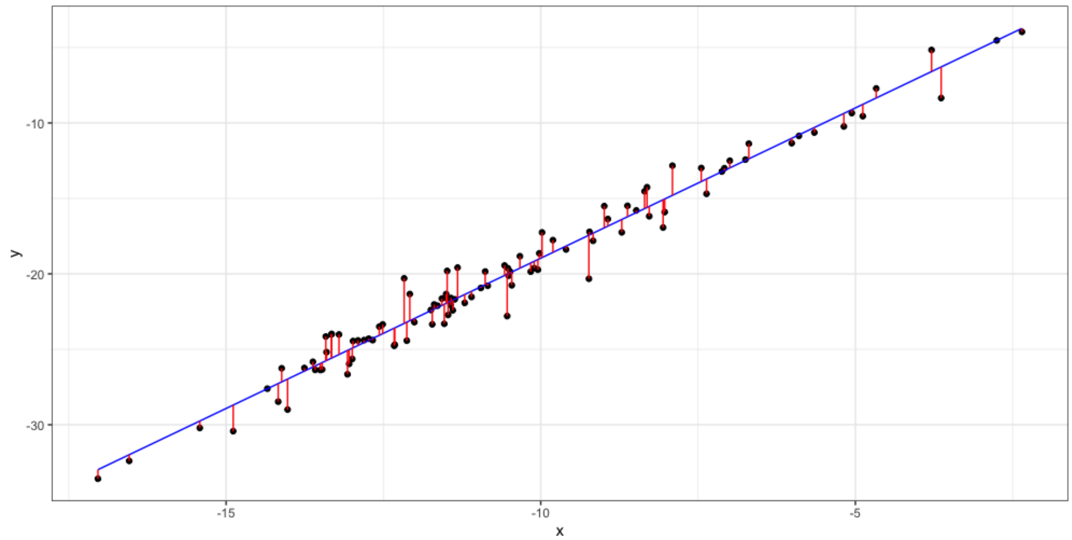


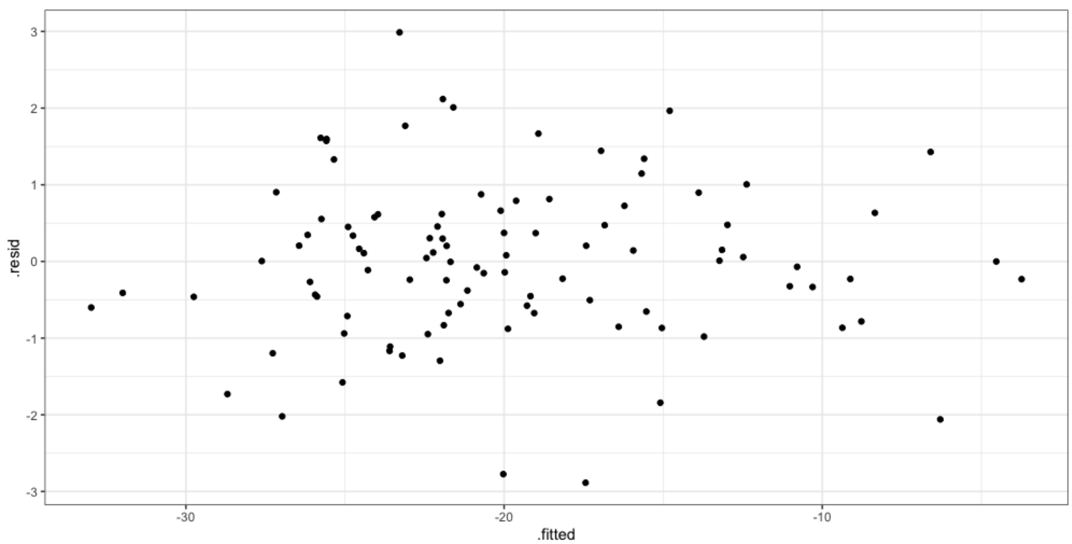
diagramme de dispersion



droite de meilleur ajustement : $\hat{y} = 0.95020 + 1.99131x$
très près de la relation réelle
coefficient de corrélation: $\rho = 0.986$



les résidus par rapport aux valeurs ajustées
ils sont relativement petits



les résidus par rapport aux valeurs ajustées (rotation)
il n'y a pas de structure spécifique, il semble que le modèle linéaire soit justifié