

STT-1900 – Méthodes statistiques pour l'ingénierie

Recueil d'exercices

Le niveau de difficulté de chaque exercice est indiqué par des étoiles, de ★ (le plus facile) à ★★★★★ (le plus difficile). Cette échelle est estimée à partir des résultats obtenus par les étudiants des sessions précédentes ; elle donne une indication utile, mais n'est pas une mesure exacte.

Module 1 - Introduction aux probabilités

Exercice

★ On vous dit que 20 % des nouvelles publiées sur un réseau social sont publiées par des trolls (T), 50 % sont publiées par de véritables journalistes (V) et 30 % sont publiées par des intelligences artificielles (I). Une étude démontre que 3/4 des nouvelles publiées par les trolls sont fausses, alors que la proportion de fausses nouvelles est de 1/10 chez les véritables journalistes et de 2/3 pour les intelligences artificielles. Utilisez F pour dénoter l'événement «la nouvelle est fausse».

- (a) Quelle proportion des nouvelles publiées sur ce réseau social sont fausses ?
- (b) Une nouvelle que vous trouvez particulièrement choquante est publiée sur ce réseau, mais vous ignorez par qui/quoi. Il s'avère que cette nouvelle n'est PAS fausse. Quelle est la probabilité qu'elle ait été publiée par une intelligence artificielle ?

Exercice

★★ Soit A , B et C trois événements avec $P(A) = 0,5$, $P(B) = 0,5$, $P(C) = 0,2$, $P(A \cap B) = 0,25$. Nous avons aussi que C est indépendant de l'événement A et de l'événement B . Dites si chacun des énoncés ci-dessous est vrai ou faux. Votre réponse doit être justifiée en mots ou par des calculs. Un vrai ou faux sans justification n'aura aucun point.

- (a) Les événements A et B sont indépendants.

☐ VRAI ☐ FAUX

Justification :

- (b) $P(A \cup C) = 0,6$.

☐ VRAI ☐ FAUX

Justification :

- (c) Les événements A et B forment une partition de l'espace échantillonnal.

☐ VRAI ☐ FAUX

Justification :

Exercice

★★ On nous donne les probabilités suivantes :

$$P(A) = 0,45$$

$$P(B) = 0,25$$

$$P(C) = 0,30$$

$$P(C|A) = 0,30$$

$$P(C|B) = 0,30$$

$$P(A \cap B) = 0,05$$

- (a) Cochez tous les énoncés qui sont **vrais**.

☐ La probabilité de l'événement complémentaire de A , $P(\bar{A}) = 0,55$

☐ $P(A \cup B) = 0,70$

☐ $P(A \cap \bar{B}) = 0,40$

☐ $P(A \cap C) = 0,1575$

☐ $P(A \cup C) = 0,75$

- (b) A et B sont indépendants.

☐ VRAI ☐ FAUX

Justification :

Exercice

★★ Dans une usine, deux machines, A et B , produisent des composants. La machine A produit 45% des composants et la machine B produit le reste. 15% des composants produits par A sont défectueux, tandis que 10% des composants

produits par B sont défectueux. Si un composant est défectueux, quelle est la probabilité qu'il ait été produit par la machine A?

Exercice

★★ On nous donne les probabilités suivantes :

$$P(A) = 0,45$$

$$P(B) = 0,55$$

$$P(C) = 0,1$$

$$P(A \cap C) = 0$$

$$P(B \cap C) = 0$$

$$P(A|B) = 0,45$$

(a) Cochez tous les énoncés qui sont **vrais**. Chaque énoncé qui est coché alors qu'il ne devrait pas l'être annule les points d'un élément coché qui devrait l'être.

☐ La probabilité du complément de A est égale à 0,55.

☐ $P(A \cup B) = 1$.

☐ A, B et C sont mutuellement exclusifs.

☐ A et C sont indépendants.

☐ Comme $P(A) + P(B) = 1$, l'intersection de A et B est vide, c'est-à-dire $A \cap B = \emptyset$.

☐ Il y a certainement une erreur dans les probabilités données, car $P(A) + P(B) + P(C) > 1$.

(b) A et B sont indépendants.

☐ VRAI ☐ FAUX

Justification :

Exercice

★★

Un sac contient 40 billes rouges, 40 billes vertes et 20 billes bleues. Sur certaines de ces billes, une étoile a été dessinée. Nous savons que la probabilité de choisir une bille sans étoile parmi les billes rouges est de 0,875, la probabilité de choisir une bille qui est verte et qui a une étoile est de 0,1 et la probabilité de choisir une bille qui a une étoile parmi les billes bleue est de 0,5.

En utilisant les événements suivants :

— R : la bille est rouge

— V : la bille est verte

— B : la bille est bleue

— E : une étoile est dessinée sur la bille

(a) Tracez le diagramme de Venn qui représente cette situation. Vous n'avez qu'à identifier les événements et **l'espace échantillonnal**. Il n'est pas nécessaire d'inscrire les probabilités.

(b) Si vous avez pigé une bille avec une étoile, quelle est la probabilité que celle-ci soit bleue?

Exercice

★★ On nous donne les probabilités suivantes :

$$P(A) = 0,45$$

$$P(B) = 0,55$$

$$P(C) = 0,1$$

$$P(A \cap C) = 0$$

$$P(B \cap C) = 0$$

$$P(A|B) = 0,45$$

(a) (3 points) Cochez tous les énoncés qui sont **vrais**. Chaque énoncé qui est coché alors qu'il ne devrait pas l'être annule les points d'un élément coché qui devrait l'être.

☐ La probabilité du complément de A est égale à 0,55.

☐ $P(A \cup B) = 1$.

☐ A, B et C sont mutuellement exclusifs.

☐ A et C sont indépendants.

☐ Comme $P(A) + P(B) = 1$, l'intersection de A et B est vide, c'est-à-dire $A \cap B = \emptyset$.

☐ Il y a certainement une erreur dans les probabilités données, car $P(A) + P(B) + P(C) > 1$.

(b) (3 points) A et B sont indépendants.

☐ VRAI ☐ FAUX

Justification :

Exercice

★★

Un sac contient 40 billes rouges, 40 billes vertes et 20 billes bleues. Sur certaines de ces billes, une étoile a été dessinée. Nous savons que la probabilité de choisir une bille sans étoile parmi les billes rouges est de 0,875, la probabilité de choisir une bille qui est verte et qui a une étoile est de 0,1 et la probabilité de choisir une bille qui a une étoile parmi les billes bleues est de 0,5.

En utilisant les événements suivants :

- R : la bille est rouge
- V : la bille est verte
- B : la bille est bleue
- E : une étoile est dessinée sur la bille

(a) (4 points) Tracez le diagramme de Venn qui représente cette situation. *Vous n'avez qu'à identifier les événements et l'espace échantillonnal. Il n'est pas nécessaire d'inscrire les probabilités.*

(b) (6 points) Si vous avez pigé une bille avec une étoile, quelle est la probabilité que celle-ci soit bleue ?

Exercice

★★ Dites si chaque énoncé est vrai ou faux.

(a) (1 point) Si A et B sont deux événements mutuellement exclusifs, alors $P(A \cup B) = P(A) + P(B)$.

☐ Vrai ☐ Faux

(b) (1 point) Si $P(A) = 0,6$ et $P(B) = 0,5$, alors A et B peuvent être mutuellement exclusifs.

☐ Vrai ☐ Faux

(c) (1 point) Si A et B sont deux événements indépendants, alors $P(A \cap B) = 0$.

☐ Vrai ☐ Faux

(d) (1 point) Si $P(A \cup B) = 1$, alors A est le complément de B .

☐ Vrai ☐ Faux

Exercice

★★ Soit $W \sim \mathcal{N}(1275; 49)$. Laquelle des probabilités suivantes est la plus grande ?

☐ $P(W > 1282)$

☐ $P(W < 1268)$

☐ $P\left(\frac{W - 1275}{7} > 0,5\right)$

☐ $P\left(\frac{W - 1275}{7} > 1\right)$

Exercice

★★★ On définit les événements A , B et C et les probabilités $P(A)$, $P(B)$ et $P(C)$. Que vaut $P(A)$ dans les cas suivants ?

(a) A et B sont indépendants, $P(B) = 0,3$ et $P(A \cup B) = 0,7$?

(b) $A \subset B$, $P(A \cup B) = 0,5$ et $P(A \cap B) = 0,1$?

(c) $B \subset A$, $P(A \cup B) = 0,5$ et $P(A \cap B) = 0,1$?

(d) $P(A|B) = 0,3$, $P(A \cup B) = 0,4$ et $P(B) = 0,2$?

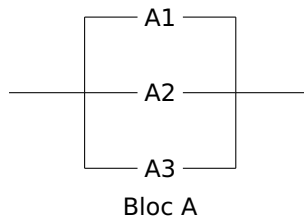
(e) B et C forment une partition de l'espace échantillonnal, et $P(A \cap C) = 0,1$, $P(C) = 0,6$ et $P(A|B) = 0,25$.

Exercice

★★★ Considérons le montage électronique représenté ci-dessous.

Le bloc A contient trois éléments. Nous avons que

- L'élément $A1$ fonctionne avec une probabilité de 0,60.
- L'élément $A2$ fonctionne avec une probabilité de 0,50.
- L'élément $A3$ fonctionne avec une probabilité de 0,60.



Pour que le montage fonctionne, au moins un élément du bloc A doit fonctionner. Les éléments d'un bloc fonctionnent de façon indépendante.

Quelle est la probabilité que le montage ne fonctionne pas ?

☐ 0,08 ☐ 0,92 ☐ 0,18 ☐ 0,82 ☐ 0,63

Exercice

★★★ On nous donne les probabilités suivantes :

$$P(A) = 0,45$$

$$P(B) = 0,25$$

$$P(C) = 0,30$$

$$P(C|A) = 0,30$$

$$P(C|B) = 0,30$$

$$P(A \cap B) = 0,05$$

Cochez tous les énoncés qui sont **vrais**, sachant que trois énoncés sont vrais. *Chaque énoncé qui est coché alors qu'il ne devrait pas l'être annule les points d'un élément coché qui devrait l'être.*

- ☐ La probabilité de l'événement complémentaire de A est $P(\bar{A}) = 0,55$
- ☐ $P(A \cap \bar{B}) = 0,40$
- ☐ $P(A \cup C) = 0,75$
- ☐ A et C sont indépendants
- ☐ Comme B et C sont indépendants, ils sont mutuellement exclusifs

Exercice

★★★ Un sac contient 40 dés rouges, 40 dés verts et 20 dés bleus. Les faces des dés rouges et les dés verts sont numérotées de «1» à «6». Les faces des dés bleus ont 6 fois la face «1».

L'expérience aléatoire consiste à choisir un dé dans le sac, à le lancer et à noter la face du dessus.

Pour répondre aux questions, utilisez $\mathcal{S}$ pour désigner l'espace échantillonnal et les événements suivants :

- R : le dé choisi est rouge,
- V : le dé choisi est vert,
- B : le dé choisi est bleu,
- N : la face sur le dessus du dé lancé est «1»,
- D : la face sur le dessus du dé lancé est «2».

(a) La face sur le dessus du dé lancé est «1». Quelle est la probabilité que le dé lancé soit rouge ?

(b) Dites si les énoncés suivants sont vrais ou faux.

i. Les événements R et B sont indépendants.

☐ Vrai ☐ Faux

ii. Les événements R et B sont mutuellement exclusifs.

☐ Vrai ☐ Faux

iii. Les événements B et N sont indépendants.

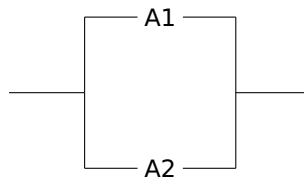
☐ Vrai ☐ Faux

iv. Les événements B et D sont mutuellement exclusifs.

☐ Vrai ☐ Faux

v. Les événements R et D sont indépendants.

☐ Vrai ☐ Faux



Bloc A

Exercice

★★★ Considérons le montage électronique représenté ci-dessous.

Le bloc A contient deux éléments. Nous avons que

- L'élément A1 fonctionne avec une probabilité de 0,90.
- L'élément A2 fonctionne avec une probabilité de 0,70.

Pour que le montage fonctionne, au moins un élément du bloc A doit fonctionner. Les éléments d'un bloc fonctionnent de façon indépendante.

Quelle est la probabilité que le montage ne fonctionne pas ?

- ☐ 0,40
- ☐ 0,03
- ☐ 0,97
- ☐ 0,37
- ☐ 0,63

Exercice

★★★

La population étudiante d'une certaine école secondaire est répartie dans les proportions suivantes : 45% de la population porte des lunettes et 55% ne porte pas de lunettes. De plus, une analyse plus détaillée des données permet de préciser que 36% des personnes sans lunettes sont gauchères alors que 14,2% des personnes avec lunettes le sont. On choisit au hasard une personne du secondaire.

- (a) (4 points) Quelle est la probabilité que la personne choisie ne porte pas de lunettes et soit gauchère ? Justifiez votre réponse.
- (b) (5 points) Quelle est la probabilité que la personne choisie soit gauchère ? Justifiez votre réponse.
- (c) (6 points) On vient de classer la personne choisie comme droitère. Quelle est la probabilité que la personne porte des lunettes ? Justifiez votre réponse.

Exercice

★★★ Dites si chaque énoncé est vrai ou faux.

- (a) (1 point) La somme des probabilités de deux événements mutuellement exclusifs peut être plus grande que 1.
☐ Vrai ☐ Faux
- (b) (1 point) La somme des probabilités de deux événements qui ne sont pas mutuellement exclusifs peut être plus grande que 1.
☐ Vrai ☐ Faux
- (c) (1 point) Si la somme des probabilités de deux événements égale 1, nous savons qu'il n'y a aucune autre possibilité que ces deux événements.
☐ Vrai ☐ Faux
- (d) (1 point) Un événement n'est pas nécessairement mutuellement exclusif avec son complément.
☐ Vrai ☐ Faux

Exercice

★★★ Un sac contient 30 dés rouges, 50 dés verts et 20 dés bleus. Les dés rouges et les dés verts sont numérotés de «1» à «6». Les dés bleus ont 6 fois la face «6».

L'expérience aléatoire consiste à choisir un dé dans le sac, à le lancer et à noter la face du dessus.

Pour répondre aux questions, utilisez $\mathcal{S}$ pour désigner l'espace échantillonnal et les événements suivants :

- R : le dé choisi est rouge,
- V : le dé choisi est vert,
- B : le dé choisi est bleu,
- S : la face sur le dessus du dé lancé est «6»,
- D : la face sur le dessus du dé lancé est «2».

(a) (7 points) La face sur le dessus du dé lancé est «6». Quelle est la probabilité que le dé lancé soit bleu ?

(b) (3 points) Quel énoncé est vrai parmi les suivants ?

- ☐ Les événements R et B sont indépendants.
- ☐ Les événements R et B sont mutuellement exclusifs.
- ☐ Les événements B et S sont indépendants.
- ☐ Les événements B et S sont mutuellement exclusifs.

Exercice

★★★★ La lune suscite un grand intérêt pour les pêcheurs car la qualité de la pêche dépend directement des différentes phases lunaires. Un pêcheur a remarqué les faits suivants pour chaque période d'une heure de pêche :

- Lorsque la lune est croissante, la probabilité de pêcher au moins un poisson est de cinq sur sept ;
- Lorsque la lune est décroissante, la probabilité de ne pêcher aucun poisson est de deux sur cinq.

La lune est considérée la moitié du temps comme croissante, et l'autre moitié comme décroissante.

- (a) Quelle est la probabilité qu'en pêchant une heure, le pêcheur pêche au moins un poisson ?
- (b) Si le pêcheur ne pêche aucun poisson après une heure de pêche, quelle est la probabilité que la lune soit croissante ?
- (c) Si, en période de lune croissante, il pêche trois heures de suite, quelle est la probabilité qu'il ne pêche aucun poisson ?
- (d) Si, en période de lune décroissante, il pêche aussi trois heures de suite, quelle est la probabilité qu'il pêche au moins un poisson ?

Exercice

★★★★ Dans une usine, deux machines, A et B, produisent des composants. La machine A produit les trois quarts des composants et la machine B produit le reste. Un quart des composants produits dans l'usine sont défectueux. Lorsqu'un composant est choisi au hasard, sa probabilité d'avoir été produit par la machine A et d'être défectueux est d'un sixième.

- (a) (3 points) Tracez le diagramme de Venn qui représente cette situation. *Vous n'avez qu'à identifier les événements. Il n'est pas nécessaire d'inscrire les probabilités.*
- (b) (4 points) Si un composant est défectueux, quelle est la probabilité qu'il ait été produit par la machine B ?

Module 2 - Variables aléatoires à une dimension

Exercice

★★ Laquelle des fonctions suivantes est une fonction de répartition ?

- ☐ $F(a) = a$, pour $0 < a$
- ☐ $F(b) = \begin{cases} 0 & \text{si } b < 0, \\ b & \text{si } 0 \leq b < 1, \\ 1 & \text{si } 1 \leq b \end{cases}$
- ☐ $F(c) = \begin{cases} \sqrt{c} & \text{si } 0 \leq c \leq 1, \\ 0 & \text{sinon} \end{cases}$
- ☐ $F(d) = \begin{cases} 0 & \text{si } d < -1, \\ |d| & \text{si } -1 \leq d \leq 1, \\ 1 & \text{si } 1 < d \end{cases}$

Exercice

★★ Laquelle des fonctions suivantes est une fonction de densité ?

- ☐ $f(a) = a$, pour $0 < a$
- ☐ $f(b) = \begin{cases} 0 & \text{si } b < 0, \\ b & \text{si } 0 \leq b < 2, \\ 0 & \text{si } 2 \leq b \end{cases}$
- ☐ $f(c) = \begin{cases} \sqrt{c} & \text{si } 0 \leq c \leq 1, \\ 0 & \text{sinon} \end{cases}$
- ☐ $f(d) = \begin{cases} 0 & \text{si } d < 0, \\ d & \text{si } 0 \leq d \leq 1, \\ 1 & \text{si } 1 < d \end{cases}$

Exercice

★★ Laquelle des fonctions suivantes est une fonction de répartition ?

☐ $F(a) = a$, si $0 < a$

☐ $F(b) = \begin{cases} 0 & \text{si } b < 0, \\ b & \text{si } 0 \leq b < 1, \\ 1 & \text{si } 1 \leq b \end{cases}$

☐ $F(c) = \begin{cases} \sqrt{c} & \text{si } 0 \leq c \leq 1, \\ 0 & \text{sinon} \end{cases}$

☐ $F(d) = \begin{cases} 0 & \text{si } d < -1, \\ |d| & \text{si } -1 \leq d \leq 1, \\ 1 & \text{si } 1 < d \end{cases}$

Exercice

★★★ La variable aléatoire X a la fonction de répartition continue suivante :

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{\sqrt{x}}{5} & \text{si } 0 \leq x \leq u, \\ 1 & \text{si } u < x. \end{cases}$$

- (a) Que vaut u ?
(b) Quelle est la fonction de densité de X ? Si vous n'avez pas trouvé « u » à la question précédente, vous pouvez exprimer la densité avec « u ».
(c) Quelle est la probabilité que X soit supérieur à 9, sachant que X est entre 4 et 16 ?
(d) Soit les quantités suivantes :

$$\int_{-\infty}^{+\infty} xf(x)dx = \frac{25}{3},$$

$$\int_{-\infty}^{+\infty} x^2 f(x)dx = 125,$$

où $f(x)$ est la fonction de densité de X .

Quel est l'écart-type de X ?

Exercice

★★★ Soit la fonction f définie sur $\mathbb{R}$ par

$$f(x) = \begin{cases} ce^x, & \text{si } x < 0, \\ 0, & \text{sinon,} \end{cases}$$

où $c \in \mathbb{R}$ est une constante donnée.

- (a) Quelle doit être la valeur de la constante c pour que f soit une densité de probabilité d'une certaine variable aléatoire X ?
(b) Déterminer la fonction de répartition de X . Si vous n'avez pas trouvé « c » à la question (a), vous pouvez exprimer la répartition avec « c ».
(c) Calculer l'espérance de X . Si vous n'avez pas trouvé « c » à la question (a), vous pouvez exprimer l'espérance avec « c ».

Exercice

★★★ X est une variable aléatoire continue dont la densité est donnée par :

$$f(x) = \begin{cases} \frac{6}{5}(1-2x^2) & \text{si } -0.5 < x < 0.5 \\ 0 & \text{sinon.} \end{cases}$$

- (a) Cochez tous les énoncés qui sont **vrais**.

☐ $F(x) = \frac{6}{5}(x - \frac{2}{3}x^3 + \frac{5}{12})$

☐ $E(3X + 20) = 3E(X)$

☐ $V(2X) = 2V(X)$

☐ $V(X + 10) > V(X)$

- (b) $V(X) = E(X^2) - 2E(X)^2$.

☐ VRAI ☐ FAUX

Justification :

(c) Calculez $P(X < 0 \mid X > -0.25)$:

Exercice

★★★ Laquelle des fonctions suivantes est une fonction de répartition ?

☐ $F(x) = \begin{cases} x & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$

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

☐ $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sqrt{x} & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1 \end{cases}$

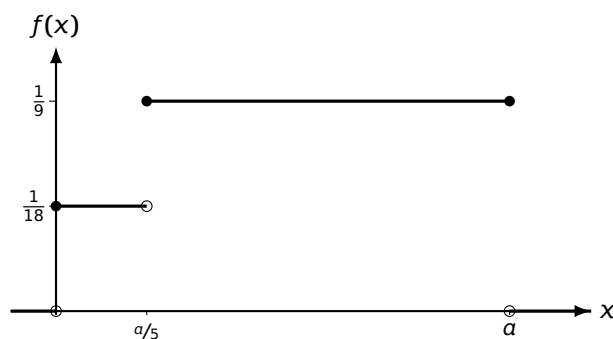
☐ $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x + 1 & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1. \end{cases}$

☐ $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 \leq x \leq 1, \\ 1 - x & \text{si } 1 < x \leq 2, \\ 1 & \text{si } x > 2. \end{cases}$

Exercice

★★★ (H26) Soit X une variable aléatoire continue de densité f_X définie ci-dessous.

$$f_X(x) = \begin{cases} \frac{1}{18} & \text{si } 0 \leq x < \frac{a}{5}, \\ \frac{a}{9} & \text{si } \frac{a}{5} \leq x \leq a, \\ 0 & \text{sinon.} \end{cases}$$



(a) Que vaut a ?

(b) Quelle est la probabilité que X prenne une valeur entre $\frac{a}{10}$ et $\frac{a}{2}$? Si vous n'avez pas trouvé la valeur de a , vous pouvez donner votre réponse en fonction de a .

(c) Quelle est l'espérance de X . Si vous n'avez pas trouvé la valeur de a , vous pouvez donner votre réponse en fonction de a .

Exercice

★★★ Laquelle des fonctions suivantes est une fonction de répartition ?

☐ $F(x) = \begin{cases} x & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$

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

☐ $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sqrt{x} & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1 \end{cases}$

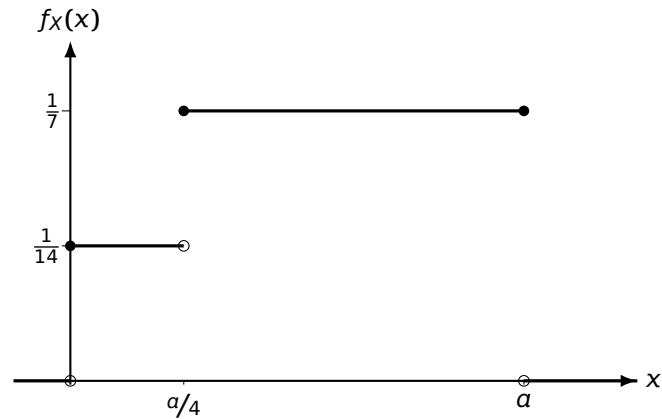
☐ $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x + 1 & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1. \end{cases}$

☐ $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 \leq x \leq 1, \\ 1 - x & \text{si } 1 < x \leq 2, \\ 1 & \text{si } x > 2. \end{cases}$

Exercice

★★★ Soit X une variable aléatoire continue de densité f_X définie ci-dessous.

$$f_X(x) = \begin{cases} \frac{1}{14} & \text{si } 0 \leq x < \frac{a}{4}, \\ \frac{1}{7} & \text{si } \frac{a}{4} \leq x \leq a, \\ 0 & \text{sinon.} \end{cases}$$



(a) (4 points) Que vaut a ?

(b) (6 points) Quelle est la probabilité que X prenne une valeur entre $\frac{a}{8}$ et $\frac{a}{2}$? *Si vous n'avez pas trouvé la valeur de a , vous pouvez donner votre réponse en fonction de a .*

(c) (4 points) Quelle est l'espérance de X ? *Si vous n'avez pas trouvé la valeur de a , vous pouvez donner votre réponse en fonction de a .*

Exercice

★★★★ Soit la variable aléatoire X et sa fonction de répartition

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & \text{si } x \geq 0, \\ 0 & \text{sinon,} \end{cases}$$

où α et β sont des constantes.

En posant $\alpha = 0,5$ et $\beta = 1$, calculez $P(16 \leq X \leq 25 | X \geq 9)$.

Exercice

★★★★ (A25)

La densité d'une variable aléatoire X est donnée par

$$f(x) = \begin{cases} 1, & 0 < x < a \\ 2, & a < x < \frac{2}{3} \\ 0, & \text{ailleurs.} \end{cases}$$

Parmi les fonctions F_1 à F_5 suivantes, laquelle est la fonction de répartition de X ?

☐ $F_1(x) = \begin{cases} a, & 0 \leq x < a, \\ 2a, & x \leq a < \frac{2}{3}, \\ 0, & \text{ailleurs.} \end{cases}$

☐ $F_2(x) = \begin{cases} x, & 0 \leq x < a, \\ 2x, & a \leq x < \frac{2}{3}, \\ 0, & \text{ailleurs.} \end{cases}$

☐ $F_3(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < a \\ 2x, & a \leq x < \frac{2}{3}, \\ 1, & x \geq \frac{2}{3}. \end{cases}$

☐ $F_4(x) = \begin{cases} x, & 0 \leq x < a, \\ a + 2(x - a), & a \leq x < \frac{2}{3}, \\ 0, & \text{ailleurs.} \end{cases}$

☐ $F_5(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < a \\ a + 2(x - a), & a \leq x < \frac{2}{3}, \\ 1, & x \geq \frac{2}{3}. \end{cases}$

☐ Aucune des fonctions F_1 à F_5 n'est la fonction de répartition de X .

Exercice

★★★★ Soit la variable aléatoire X dont la fonction de répartition est donnée par :

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$$

- (a) Quelle est la valeur de la médiane de X ?
- (b) Calculez la probabilité que X soit compris entre 0,2 et 0,3, sachant que X est supérieur à 0,1.
- (c) Donnez la fonction de densité de X .

Exercice

★★★★

Une variable aléatoire continue X est distribuée selon la fonction de densité suivante :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{9} & \text{si } 0 \leq x < 3 \\ \frac{2}{3} - \frac{x}{9} & \text{si } 3 \leq x < 6 \\ 0 & \text{si } x \geq 6. \end{cases}$$

- (a) (3 points) En utilisant la fonction de densité ci-dessus, calculez $P(X \geq 4)$. Justifiez votre réponse.
- (b) (3 points) Calculez l'espérance de la variable aléatoire X . Justifiez votre réponse.
- (c) (6 points) Déterminez la fonction de répartition $F(x)$ de la variable aléatoire X . Justifiez votre réponse.

Exercice

★★★★ Soit la variable aléatoire X et sa fonction de répartition

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & \text{si } x \geq 0, \\ 0 & \text{sinon,} \end{cases}$$

où α et β sont des constantes.

En posant $\alpha = 0,5$ et $\beta = 1$, calculez $P(16 \leq X \leq 25 | X \geq 9)$.

Exercice

★★★★ En analyse de survie, nous définissons la fonction de survie $S(x)$ comme étant la probabilité de survivre plus de x unités de temps, c'est-à-dire $S(x) = P(X > x)$, où X est une variable aléatoire continue qui mesure la durée de vie. Soit la variable aléatoire X , dont la fonction de survie est donnée par

$$S(x) = \begin{cases} 1 & \text{si } x < 1 \\ -x^2 + 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{si } x > 2. \end{cases}$$

Laquelle des fonctions suivantes est la fonction de densité de X ?

☐ $f(x) = \begin{cases} 1 - x^2 + 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} 1 + x^2 - 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} -2x + 2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} 2x - 2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} \frac{-x^3}{3} + x^2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ Nous ne pouvons pas déterminer la fonction de densité à partir des informations données.

Exercice

★★★★ En analyse de survie, nous définissons la fonction de survie $S(x)$ comme étant la probabilité de survivre plus

de x unités de temps, c'est-à-dire $S(x) = P(X > x)$, où X est une variable aléatoire continue qui mesure la durée de vie. Soit la variable aléatoire X , dont la fonction de survie est donnée par

$$S(x) = \begin{cases} 1 & \text{si } x < 1 \\ -x^2 + 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{si } x > 2. \end{cases}$$

Laquelle des fonctions suivantes est la fonction de densité de X ?

☐ $f(x) = \begin{cases} 1 - x^2 + 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} 1 + x^2 - 2x & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} -2x + 2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} 2x - 2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ $f(x) = \begin{cases} \frac{-x^3}{3} + x^2 & \text{si } 1 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$

☐ Nous ne pouvons pas déterminer la fonction de densité à partir des informations données.

Exercice

★★★★ On définit la fonction de survie S_X d'une variable aléatoire X par $S_X(x) = P(X > x)$ et le taux de défaillance h_X par $h_X(x) = \frac{f_X(x)}{S_X(x)}$, où f_X est la fonction de densité de X . Soit la variable aléatoire X dont la fonction de répartition est

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-3x} & \text{si } x > 0. \end{cases}$$

Que vaut h_X ?

☐ $h_X(x) = 3e^{-3x}$ pour $x > 0$ et 0 sinon.

☐ $h_X(x) = e^{-3x}$ pour $x > 0$ et 0 sinon.

☐ $h_X(x) = 3$ pour $x > 0$ et 0 sinon.

☐ $h_X(x) = 1$ pour $x > 0$ et 0 sinon.

Module 3 - Loi normale et théorème central limite

Exercice

★★ L'examen A et l'examen B servent pour la classification à un même emploi. Nous savons que les notes de ces deux examens suivent des lois normales d'espérance $\mu_A = 1100$ et $\mu_B = 21$, respectivement, de variance $\sigma_A^2 = 40\,000$ et $\sigma_B^2 = 36$, respectivement et de fonctions de répartition F_A et F_B , respectivement. Nous établissons le pointage d'un candidat par la probabilité d'obtenir une note inférieure ou égale à la note du candidat en question (donc en évaluant $F_A(a)$ ou $F_B(b)$, pour un candidat ayant obtenu a à l'examen A ou b à l'examen B).

(a) Sophie a eu une note de 1300 à l'examen A alors que Maxime a eu une note de 24 à l'examen B. Lequel des deux a un meilleur pointage ?

(b) Quelle note aurait dû avoir Sophie afin d'obtenir le même pointage que Maxime ?

Exercice

★★ Vous vendez 100 véhicules. Pour chacun, le prix inclut 2 000 \$ pour couvrir les frais de garantie ; au total, vous avez donc 200 000 \$ de provision. Soit $C_1, \dots, C_{100}$ les coûts que vous devez rembourser aux clients pour des problèmes sous garantie pour chacun de ces 100 véhicules vendus.

On vous dit que $C_1, \dots, C_{100}$ est un échantillon aléatoire d'une loi inconnue d'espérance 1 950 \$ et d'écart-type 400 \$. Quelle est la probabilité approximative que la provision ne soit pas suffisante pour couvrir les coûts de la garantie ? Autrement dit, calculez la valeur approximative de

$$P\left(\sum_{i=1}^{100} C_i > 200\,000\right).$$

Exercice

★★ Soit un système électronique constitué d'un composant de type 1 et d'un composant de type 2. Considérons X_1 et X_2 deux variables aléatoires qui représentent respectivement la durée de vie (en mois) du composant de type 1 et celle du composant de type 2. On suppose que la distribution conjointe de (X_1, X_2) est une loi normale bidimensionnelle de paramètres $\mu_1 = 92$, $\mu_2 = 84$, $\sigma_1^2 = 9$, $\sigma_2^2 = 25$ et $\rho = 0,8$.

Sachant que la durée de vie du composant de type 1 vaut 101 mois, identifiez la loi de la durée de vie du composant de type 2 avec sa moyenne et sa variance.

- ☐ $N(92, 9)$
☐ $N(96, 9)$
☐ $N(84, 25)$
☐ $N(100, 9)$
☐ $N(144, 9)$

Exercice

★★ Soit un système électronique constitué d'un composant de type 1 et d'un composant de type 2. Considérons X_1 et X_2 , deux variables aléatoires qui représentent respectivement la durée de vie (en mois) du composant de type 1 et celle du composant de type 2. On suppose que la loi conjointe de (X_1, X_2) est une loi normale bidimensionnelle de paramètres $\mu_1 = 92$, $\mu_2 = 84$, $\sigma_1^2 = 9$, $\sigma_2^2 = 25$ et $\rho = 0,8$.

Sachant que la durée de vie du composant de type 1 vaut 101 mois, identifiez la loi de la durée de vie du composant de type 2.

- ☐ $\mathcal{N}(92, 9)$ ☐ $\mathcal{N}(96, 9)$ ☐ $\mathcal{N}(84, 25)$ ☐ $\mathcal{N}(100, 9)$ ☐ $\mathcal{N}(144, 9)$

Exercice

★★ La résistance au cisaillement Y et le diamètre X de soudures par points ont fait l'objet d'une étude visant à remplacer un essai destructif par un essai non destructif. Voici les conclusions de cette étude.

1. Le vecteur $[X, Y]$ obéit à une loi normale bivariable.
2. $\mu_X = 0,20$ cm, $\mu_Y = 1100$ kg, $\sigma_X^2 = 0,02$ cm², $\sigma_Y^2 = 525$ kg² et $\rho = 0,9$.

En supposant que le diamètre vaut d cm, quelle est l'expression de l'espérance de Y , sachant que $X = d$?

- ☐ $E[Y|X = d] = 1100$
☐ $E[Y|X = d] = 145,8d + 1070,84$ kg
☐ $E[Y|X = d] = 1100,20$ kg
☐ $E[Y|X = d] = 1100 + d$ kg
☐ $E[Y|X = d] = 145,8d + 1100$ kg

Exercice

★★ L'examen A et l'examen B servent pour la classification à un même emploi. Nous savons que les notes de ces deux examens suivent des lois normales d'espérance $\mu_A = 1100$ et $\mu_B = 21$, respectivement, de variance $\sigma_A^2 = 40\,000$ et $\sigma_B^2 = 36$, respectivement et de fonctions de répartition F_A et F_B , respectivement. Nous établissons le pointage d'un candidat par la probabilité d'obtenir une note inférieure ou égale à la note du candidat en question (donc en évaluant $F_A(a)$ ou $F_B(b)$, pour un candidat ayant obtenu a à l'examen A ou b à l'examen B).

- (a) (7 points) Lucie a eu une note de 1300 à l'examen A alors que Camille a eu une note de 24 à l'examen B. Laquelle des deux a un meilleur pointage ?
- (b) (5 points) Quelle note aurait dû avoir Lucie afin d'obtenir le même pointage que Camille ?
- (c) (6 points) On tente de faire accréditer l'examen C pour le même emploi. On sait que la note à l'examen C est de loi quelconque, d'espérance 50 et de variance 25. On demande à 64 personnes de passer l'examen C. Quelle est la probabilité que la note moyenne de ces 64 personnes soit inférieure à 51,225 ?

Exercice

★★★ La variable aléatoire X suit une loi normale d'espérance μ et de variance μ^4 .

Sachant que $P(X \leq 0) = 0,84134$ et que $P(X > 1) = 0,02775$, que vaut μ ?

Exercice

★★★ Soit les variables aléatoires $X \sim \mathcal{N}(70, 100)$ et $Y \sim \mathcal{N}(80, 64)$. Calculez $P(X > Y)$.

Exercice

★★★ Un producteur de Gruyère sait que le poids d'une de ses meules de fromage, X , est une variable aléatoire d'espérance $\mu = 30$ kg et de variance $\sigma^2 = 9$ kg². Il doit envoyer 144 meules à un festival. Le transporteur lui chargera 1500\$, à condition que le poids total des 144 meules soit inférieur ou égal à 4431,24 kg. Dans le cas contraire, il lui chargera 2000\$.

Quelle est la probabilité approximative que le transporteur lui charge 1500\$?

Exercice

★★★ Un train circule de façon régulière entre Neuchâtel et Yverdon. Le temps de trajet entre Neuchâtel et Yverdon suit une loi normale avec une moyenne de $\mu_1 = 18$ minutes et une variance $\sigma_1^2 = 9$.

- (a) (4 points) Quelle est la probabilité que le train prenne entre 18 et 21 minutes pour effectuer le trajet de Neuchâtel à Yverdon ? Justifiez votre réponse.
- (b) (5 points) Nous savons que 2,5% des fois, le train prend plus de k minutes pour effectuer le trajet entre Neuchâtel et Yverdon. Quelle est la valeur de k ? Justifiez votre réponse.
- (c) (6 points) Après un arrêt d'une durée négligable (d'un temps nul), le train continue jusqu'à Lausanne. Le temps de trajet entre Yverdon et Lausanne suit une loi normale avec une moyenne μ_2 inconnue et une variance $\sigma_2^2 = 16$. Nous savons 94,95% du temps, la somme du temps de trajet entre Neuchâtel et Yverdon et du temps de trajet entre Yverdon et Lausanne est inférieure à 45 minutes. Quelle est la valeur de μ_2 , le temps moyen de trajet entre Yverdon et Lausanne ? Justifiez votre réponse.

Exercice

★★★ Soit les variables aléatoires $X \sim \mathcal{N}(70, 100)$ et $Y \sim \mathcal{N}(80, 64)$. Calculez $P(X > Y)$.

Exercice

★★★ Un producteur de Gruyère sait que le poids d'une de ses meules de fromage, X , est une variable aléatoire d'espérance $\mu = 30$ kg et de variance $\sigma^2 = 9$ kg². Il doit envoyer 144 meules à un festival. Le transporteur lui chargera 1500\$, à condition que le poids total des 144 meules soit inférieur ou égal à 4431,24 kg. Dans le cas contraire, il lui chargera 2000\$.

Quelle est la probabilité approximative que le transporteur lui charge 1500\$?

Exercice

★★★ L'examen A et l'examen B servent pour la classification à un même emploi. Nous savons que les notes de ces deux examens suivent des lois normales d'espérance $\mu_A = 500$ et $\mu_B = 30$, respectivement, de variance $\sigma_A^2 = 10\,000$ et $\sigma_B^2 = 25$, respectivement et de fonctions de répartition F_A et F_B , respectivement. Nous établissons le pointage d'un candidat par la probabilité d'obtenir une note inférieure ou égale à la note du candidat en question (donc en évaluant $F_A(a)$ ou $F_B(b)$, pour un candidat ayant obtenu a à l'examen A ou b à l'examen B).

- (a) (7 points) Marc a eu une note de 650 à l'examen A alors que Sophie a eu une note de 33 à l'examen B. Lequel des deux a un meilleur pointage ?
- (b) (5 points) Quelle note aurait dû avoir Marc afin d'obtenir le même pointage que Sophie ?
- (c) (6 points) On tente de faire accréditer l'examen C pour le même emploi. On sait que la note à l'examen C est de loi quelconque, d'espérance 60 et de variance 36. On demande à 49 personnes de passer l'examen C. Quelle est la probabilité que la note moyenne de ces 49 personnes soit inférieure à 61,68 ?

Exercice

★★★★ Supposons qu'un ingénieur travaille sur le développement de deux capteurs de température pour un processus industriel. On peut modéliser les mesures des deux capteurs X et Y comme deux variables aléatoires normales indépendantes, avec des moyennes respectives $\mu_X = 20$ et $\mu_Y = 25$ et des variances $\sigma_X^2 = 4$ et $\sigma_Y^2 = 9$.

- (a) Cochez tous les énoncés qui sont **vrais**.

☐ $X - Y$ est également une variable aléatoire qui suit une loi normale.

☐ $\frac{X - 20}{4} \sim \mathcal{N}(0, 1)$

☐ $P(X < 18.5) > P(X > 21.5)$.

☐ $P(Y < 20) > P(Y > 29)$.

☐ $P(X < 18) < P(Y < 18)$

- (b) $\text{Var}[X - Y] < \text{Var}[X + Y]$.

☐ VRAI ☐ FAUX

Justification :

- (c) Calculez $P(5Y > 4X + 62)$.

Exercice

★★★★ Soit X , le volume d'eau qui arrive dans un bassin dans une journée. Soit Y , le volume d'eau puisée de ce même bassin dans une journée. Les volumes d'eau sont en milliers de mètres cubes. On vous dit que X et Y sont des variables aléatoires indépendantes qui suivent toutes deux des lois normales, avec $X \sim \mathcal{N}(8, 4)$ et $Y \sim \mathcal{N}(12, 9)$.

- (a) On souhaite redimensionner le bassin de sorte que son nouveau volume devra être égal au 0,985^e quantile de la distribution de la quantité d'eau y arrivant dans une journée. Quel devra être ce nouveau volume ?
- (b) Dans une journée donnée, quelle est la probabilité que la quantité d'eau puisée du bassin soit supérieure au double de la quantité d'eau qui arrive dans le bassin ?

Exercice

★★★★ Un producteur de Vacherin Fribourgeois sait que le poids d'une de ses meules de fromage, X , est une variable aléatoire d'espérance $\mu = 7$ kg et de variance $\sigma^2 = 1$ kg². Il doit envoyer 100 meules à un festival. Le transporteur lui chargera 1500\$, à condition que le poids total des 100 meules soit inférieur ou égal à 720 kg. Dans le cas contraire, il lui chargera 2000\$.

Quelle est la probabilité approximative que le transporteur lui charge 2000\$?

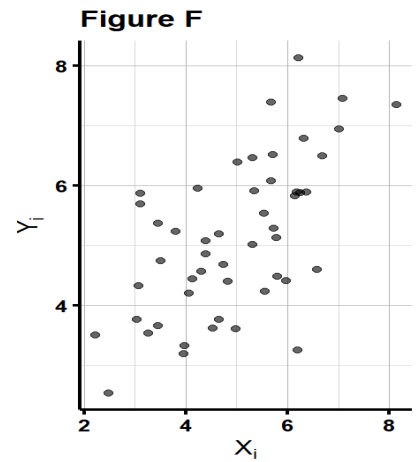
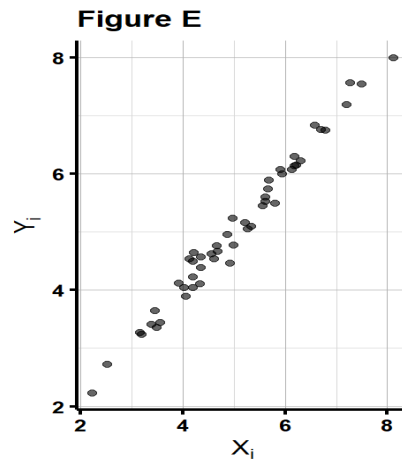
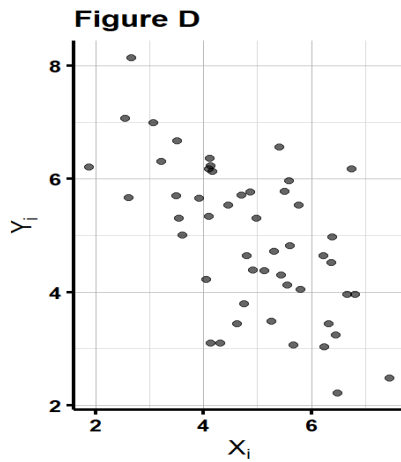
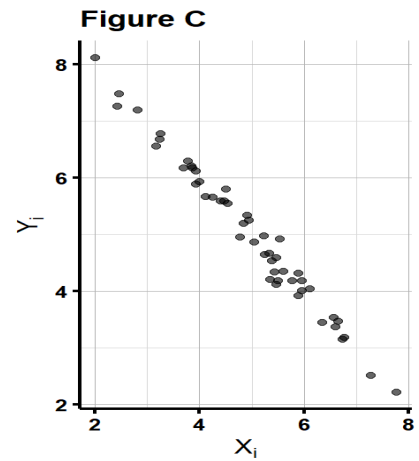
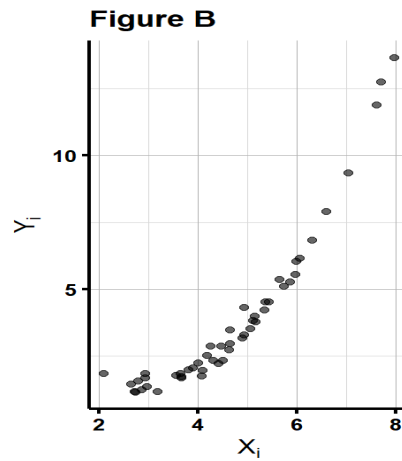
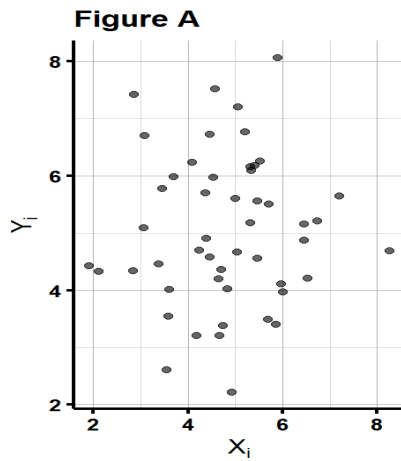
Module 4 - Covariance et corrélation

Exercice

★ (A25)

Les figures A à F ci-dessous montrent toutes un diagramme de dispersion de 50 paires (X_i, Y_i) indépendantes. Chaque échantillon qui a servi à produire ces figures a été généré à l'aide de l'un des modèles 1 à 6 ci-dessous. Dans la boîte à droite de chaque modèle, écrivez la lettre de la figure qu'il a générée.

1. $X_i \sim N(5, 2)$, $Y_i \sim N(5, 2)$ et la corrélation entre X_i et Y_i est de -0,99
2. $X_i \sim N(5, 2)$, $Y_i \sim N(5, 2)$ et la corrélation entre X_i et Y_i est de -0,60
3. $X_i \sim N(5, 2)$, $Y_i \sim N(5, 2)$ et la corrélation entre X_i et Y_i est de 0
4. $X_i \sim N(5, 2)$, $Y_i \sim N(5, 2)$ et la corrélation entre X_i et Y_i est de 0,60
5. $X_i \sim N(5, 2)$, $Y_i \sim N(5, 2)$ et la corrélation entre X_i et Y_i est de 0,99
6. $X_i \sim N(5, 2)$ et $Y_i = 6 - 3X_i + X_i^2/2 + E_i$, où $E_i \sim N(0, 0,09)$



Exercice

★ (H26) Soit X , Y et W trois variables aléatoires avec

- $E(X) = 1$, $E(Y) = 0$, $E(W) = 0$
- $V(X) = 16$, $V(Y) = 4$, $V(W) = 25$
- $Cov(X, Y) = -2$, $Cov(X, W) = 2$ et W est indépendante de Y .

(a) Calculez $E(3 + 4X - Y + W)$.

(b) Calculez $V(3 + 4X - Y + W)$.

Exercice

★★ (A25)

Vous lisez toujours les conseils de Jack et d'Averell avant d'acheter l'action d'une compagnie, mais les prix d'achat suggérés par ces deux experts varient beaucoup. Soit J , le prix d'achat de l'action qui vous intéresse suggéré par Jack et A , le prix d'achat pour cette même action suggéré par Averell. On vous dit que $J \sim N(10 ; 3)$, $A \sim N(10 ; 1)$ et que $Cov(J, A) = -2$.

Vous décidez d'acheter à la moyenne des deux prix suggérés par les experts, soit au prix

$$M = \frac{J + A}{2}.$$

Calculez la variance de M .

Exercice

★★

Soit X , Y et W trois variables aléatoires avec

- $E(X) = 2$, $E(Y) = 1$, $E(W) = -1$
- $V(X) = 9$, $V(Y) = 16$, $V(W) = 4$
- $Cov(X, Y) = 3$, $Cov(X, W) = -1$ et W est indépendante de Y .

(a) (3 points) Calculez $E(2 + 3X - 2Y + W)$.

(b) (7 points) Calculez $V(2 + 3X - 2Y + W)$.

Exercice

★★★★ Parmi les quatre façons de générer X et Y , laquelle donnera $\rho_{X,Y} = 1$?

- ☐ $X \sim \mathcal{N}(0,1)$ et $Y \sim \mathcal{N}(0,1)$
☐ $X \sim \mathcal{N}(0,1)$ et $Y = X$
☐ $X \sim \mathcal{N}(0,1)$ et $Y = X^2$
☐ $X \sim \mathcal{N}(0,1)$ et $Y = X + \epsilon$ où $\epsilon \sim \mathcal{N}(0,1)$

Module 5 - Statistiques descriptives

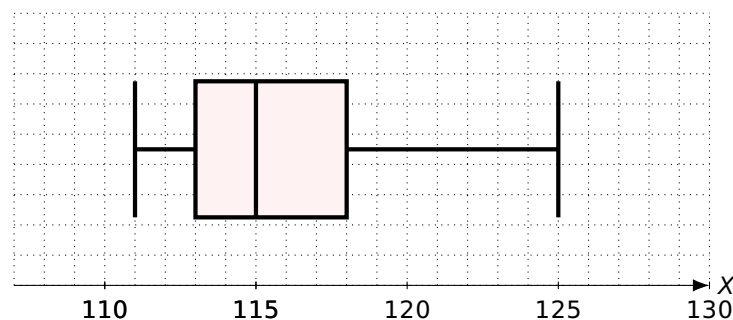
Exercice

★ Dans un laboratoire de microtechnique, Audrey et Vincent ont étudié le temps de réaction (en millisecondes) d'un certain type de détecteur de mouvements. Les 21 valeurs suivantes sont des temps de réaction collectés au cours d'une journée :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	125	128

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats. Il manque toutefois quelque chose à ce diagramme (il ne s'agit pas d'un titre ou de noms d'axes).

- (a) Ajoutez **sur le diagramme** l'élément manquant et dites ce que vous avez fait (ceci *doit* inclure des calculs) sous le diagramme. Un ajout sur le diagramme sans justification n'aura aucun point.



- (b) Nous apprenons que la plus grande observation (128) est erronée. Nous ne connaissons pas sa valeur, mais nous savons que celle-ci est plus grande que 150. Nous notons cette valeur XXX et nous savons que $XXX > 150$. Le nouveau jeu de données est donc :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	127	XXX

À partir de ce nouveau jeu de données, répondez aux questions suivantes. Nous nommons «jeu de données original» celui qui contient l'observation 128 et «nouveau jeu de données» celui qui contient l'observation XXX. Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

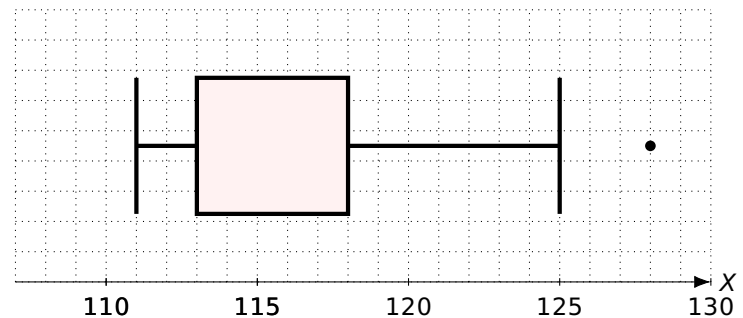
- i. La médiane échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- ii. La moyenne échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- iii. La variance échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux

Exercice

★★ Dans un laboratoire de microtechnique, Benjamin et Guillaume ont étudié le temps de réaction (en millisecondes) d'un certain type de détecteur de mouvements. Les 21 valeurs suivantes sont des temps de réaction collectés au cours d'une journée :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	125	128

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats. Il manque toutefois quelque chose à ce diagramme (il ne s'agit pas d'un titre ou de noms d'axes). Ajoutez **sur le diagramme** l'élément manquant et dites ce que vous avez fait sous le diagramme. Un ajout sur le diagramme sans justification n'aura aucun point.

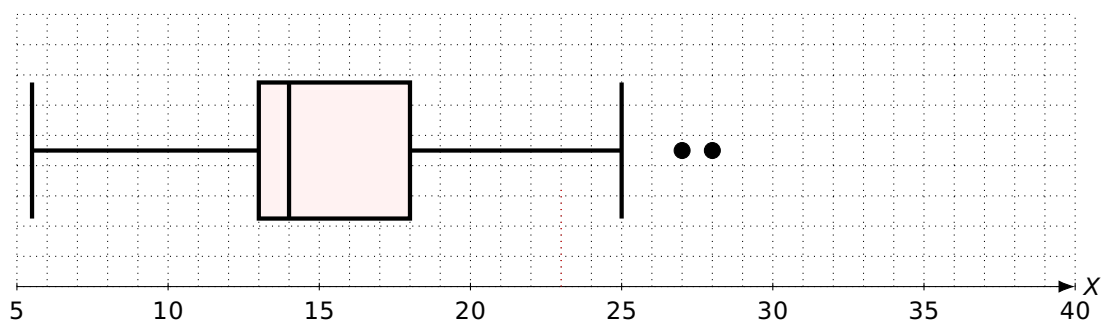


Exercice

★★ Dans une boulangerie, un pâtissier a étudié le temps de cuisson (en minutes) d'un croissant au chocolat. Les 21 valeurs suivantes sont des temps de cuisson collectés au cours d'une journée par ordre croissant :

11	11	11	12	13	13	13
13	14	14	14	14	15	16
16	18	18	19	25	27	28

- (a) (3 points) Trouvez la médiane échantillonnale des observations ci-dessus. Justifiez votre réponse.
- (b) (4 points) Pour les observations ci-dessus, le premier quartile ($q_{0,25}$) et le troisième quartile ($q_{0,75}$) sont respectivement égaux à 13 et 18. Voici un diagramme en boîte pour ces observations. Vous savez qu'il y a **une** erreur dans celui-ci. Quelle est-elle ? Justifiez votre réponse.



- (c) Vous apprenez que les trois plus grandes observations (25, 27, 28) sont erronées. Le nouveau jeu de données est donc :

11	11	11	12	13	13	13
13	14	14	14	14	15	16
16	18	18	19			

À partir de ce nouveau jeu de données, répondez aux questions suivantes. Nous nommons «jeu de données original» celui qui inclut les observations 25, 27 et 28.

Inscrivez «Vrai» ou «Faux» dans les cases appropriées.

- i. (1 point) La médiane échantillonnale du nouveau jeu de données est inférieure ou égale à celle du jeu de données original.
- ii. (1 point) La moyenne échantillonnale du nouveau jeu de données est égale à celle du jeu de données original.
- iii. (1 point) La variance échantillonnale du nouveau jeu de données est inférieure à celle du jeu de données original.

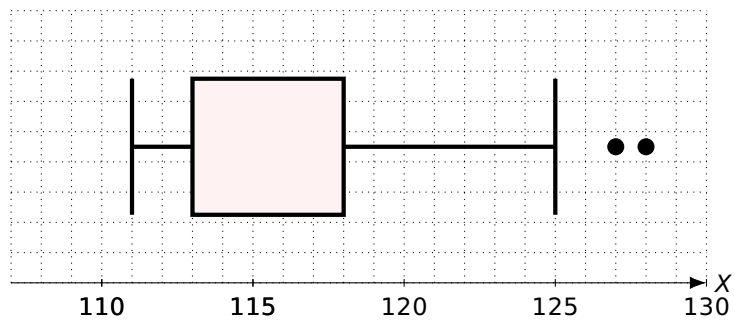
Exercice

★★ Dans un laboratoire de microtechnique, un technicien a étudié le temps de réaction (en millisecondes) d'un certain type de détecteur de mouvements. Les 21 valeurs suivantes sont des temps de réaction collectés au cours d'une journée :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	127	128

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats. Il manque toutefois quelque chose à ce diagramme (il ne s'agit pas d'un titre ou de noms d'axes).

(a) (3 points) Ajoutez **sur le diagramme** l'élément manquant et dites ce que vous avez fait (ceci *doit* inclure des calculs) sous le diagramme.



(b) (6 points) Nous apprenons que la plus grande observation (128) est erronée. Nous ne connaissons pas sa valeur, mais nous savons que celle-ci est plus grande que 150. Nous notons cette valeur XXX et nous savons que $XXX > 150$. Le nouveau jeu de données est donc :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	127	XXX

À partir de ce nouveau jeu de données, répondez aux questions suivantes. Nous nommons «jeu de données original» celui qui contient l'observation 128 et «nouveau jeu de données» celui qui contient l'observation XXX. Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

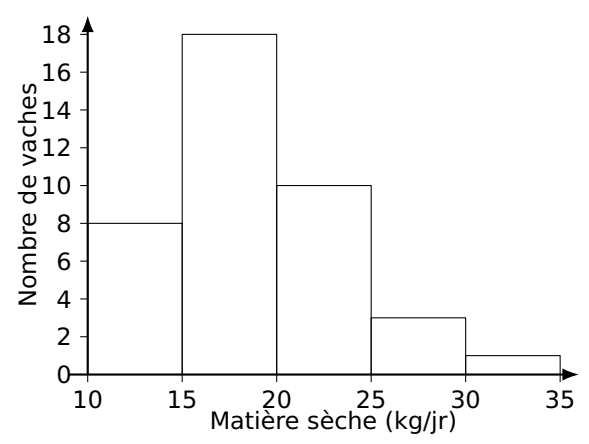
- i. La médiane échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- ii. La moyenne échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- iii. La variance échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux

Exercice

★★ Soit l'histogramme ci-contre.

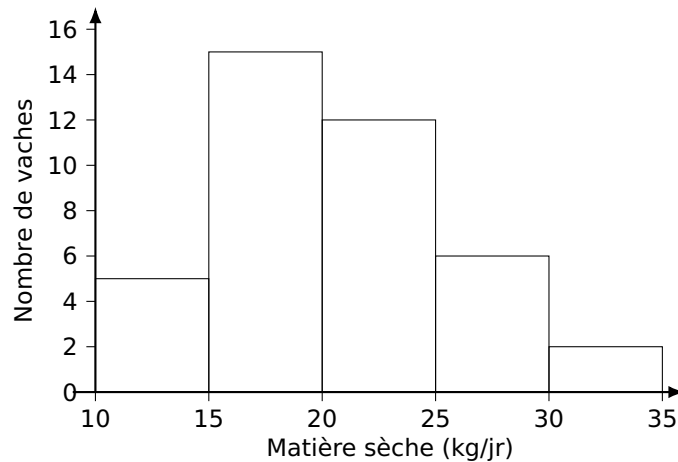
Dans quelle classe de l'histogramme se situe le quantile d'ordre 0,30 ?

- ☐ Dans la classe [10,15[
- ☐ Dans la classe [15,20[
- ☐ Dans la classe [20,25[
- ☐ Dans la classe [25,30[
- ☐ Dans la classe [30,35[



Exercice

★★ Soit l'histogramme ci-dessous.



Dans quelle classe de l'histogramme se situe le quantile d'ordre 0,60 ?

- ☐ Dans la classe [10,15[
- ☐ Dans la classe [15,20[
- ☐ Dans la classe [20,25[
- ☐ Dans la classe [25,30[
- ☐ Dans la classe [30,35]

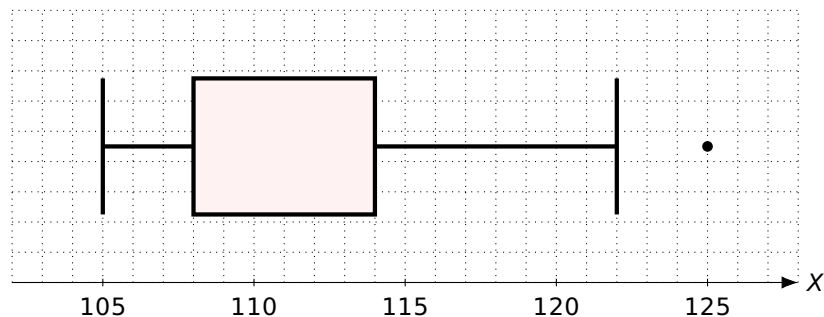
Exercice

★★ Dans un laboratoire d'ergonomie, Alexis et Émilie ont étudié le temps de réponse (en millisecondes) d'un certain type de capteur de proximité. Les 21 valeurs suivantes sont des temps de réponse collectés au cours d'une journée :

105	106	106	107	108	108	109
109	110	110	111	111	112	112
113	114	114	115	120	122	125

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats. Il manque toutefois quelque chose à ce diagramme (il ne s'agit pas d'un titre ou de noms d'axes).

Ajoutez **sur le diagramme** l'élément manquant et dites ce que vous avez fait sous le diagramme. Un ajout sur le diagramme sans justification n'aura aucun point.

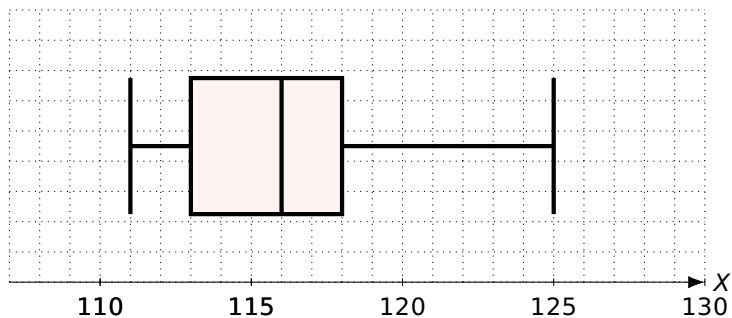


Exercice

★★ Dans un laboratoire de microtechnique, Audrey et Vincent ont étudié le temps de réaction (en millisecondes) d'un certain type de détecteur de mouvements. Les 21 valeurs suivantes sont des temps de réaction collectés au cours d'une journée :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
118	118	118	119	125	125	125

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats. Il y a toutefois une erreur (il ne s'agit pas d'un titre ou de noms d'axes).



(a) (3 points) Quelle est cette erreur ?

- ☐ La médiane n'est pas au bon endroit.
- ☐ La moyenne échantillonnale n'est pas au bon endroit.
- ☐ Il devrait y avoir deux points : un à 105,5 et un à 125,5.
- ☐ La moustache de gauche devrait être à 105,5
- ☐ La moustache de droite devrait être à 125,5

(b) (6 points) Nous apprenons qu'une observation (125) est erronée. Nous ne connaissons pas sa valeur, mais nous savons que celle-ci est plus grande que 150. Nous notons cette valeur XXX et nous savons que $XXX > 150$. Le nouveau jeu de données est donc :

111	111	111	112	113	113	113
113	115	115	115	115	115	116
116	118	118	119	125	125	XXX

À partir de ce nouveau jeu de données, répondez aux questions suivantes. Nous nommons «jeu de données original» celui qui contient l'observation 125 et «nouveau jeu de données» celui qui contient l'observation XXX. Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- i. La médiane échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- ii. La moyenne échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux
- iii. La variance échantillonnale du nouveau jeu de données est supérieure à celle du jeu de données original.
☐ Vrai ☐ Faux

Exercice

★★★ Soit le jeu de données $x_1, \dots, x_n$, où n est une nombre impair. Nous supposons que toutes les valeurs sont différentes. La moyenne échantillonnale de ce jeu de données est $\bar{x}$ et sa médiane est $\tilde{x}$. Dans chacune des situations suivantes, cochez toutes les réponses vraies.

(a) (2 points) Que pouvons-nous faire pour augmenter $\bar{x}$ sans augmenter $\tilde{x}$?

- ☐ Changer la valeur de $x_{(1)}$ pour une valeur supérieure à $x_{(n)}$
- ☐ Changer la valeur de $x_{(n)}$ pour une valeur inférieure à $x_{(1)}$
- ☐ Changer la valeur de $x_{(n)}$ pour une valeur supérieure à $x_{(n)}$
- ☐ Changer la valeur de $x_{(1)}$ pour une valeur supérieure à $x_{(1)}$, mais inférieure à $x_{(0,5n+0,5)}$

(b) (2 points) Que pouvons-nous faire pour augmenter $\bar{x}$ et $\tilde{x}$?

- ☐ Changer la valeur de $x_{(1)}$ pour une valeur supérieure à $x_{(n)}$
- ☐ Changer la valeur de $x_{(n)}$ pour une valeur inférieure à $x_{(1)}$
- ☐ Changer la valeur de $x_{(n)}$ pour une valeur supérieure à $x_{(n)}$
- ☐ Changer la valeur de $x_{(1)}$ pour une valeur supérieure à $x_{(1)}$, mais inférieure à $x_{(0,5n+0,5)}$

Exercice

★★★ Dites si les énoncés suivants sont vrais ou faux.

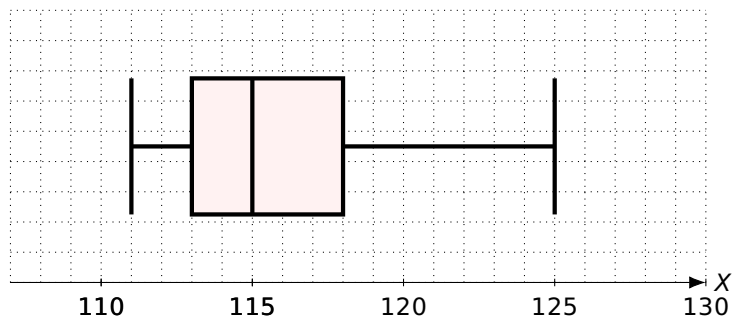
- (a) (1 point) Soit $x_1, \dots, x_n$. Si la médiane échantillonnale est 25, nous sommes certains que la valeur de l'observation du milieu est 25.
☐ Vrai ☐ Faux
- (b) (1 point) Soit $x_1, \dots, x_n$. Si la moyenne échantillonnale est 25, environ 50% des observations sont plus grandes que 25 et environ 50% sont plus petites que 25.
☐ Vrai ☐ Faux
- (c) (1 point) Soit $x_1, \dots, x_n$. Si $Q_1 = 12$, nous savons que le minimum de l'échantillon est 12.
☐ Vrai ☐ Faux
- (d) (1 point) Soit $x_1, \dots, x_n$. Nous savons que $Q_3 > 0$.
☐ Vrai ☐ Faux

Exercice

★★★★ Soit les 21 valeurs suivantes :

111	112	112	112	113	113	113
113	115	115	115	115	115	116
118	118	118	119	124	124	125

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats.

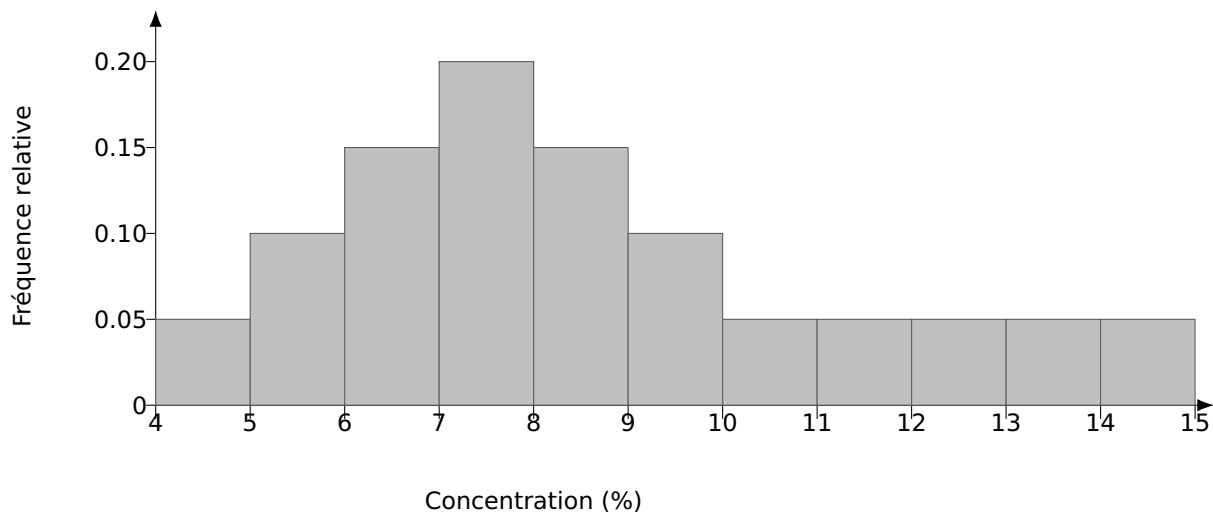


Lequel des énoncés suivants est vrai ?

- ☐ La médiane n'est pas au bon endroit.
- ☐ La moyenne échantillonnale n'est pas au bon endroit.
- ☐ Il devrait y avoir deux points : un à 105,5 et un à 125,5.
- ☐ La moustache de gauche devrait être à 105,5.
- ☐ La moustache de droite devrait être à 125,5.
- ☐ Il y a plus qu'une erreur.
- ☐ Il n'y a aucune erreur.

Exercice

★★★★ Soit l'histogramme suivant :



Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

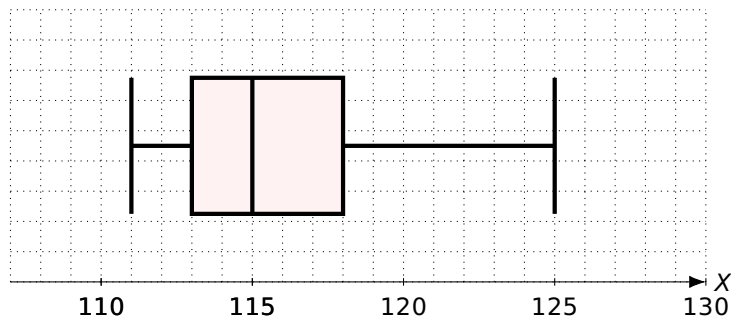
- (a) Il est impossible de donner la taille de l'échantillon qui a servi à tracer cet histogramme avec l'information donnée.
☐ Vrai ☐ Faux
- (b) Cet histogramme est typique de données provenant d'une distribution où la moyenne est supérieure à la médiane.
☐ Vrai ☐ Faux
- (c) L'écart interquartile vaut au moins 5,5.
☐ Vrai ☐ Faux
- (d) La médiane échantillonnale est entre 7% et 9%, inclusivement.
☐ Vrai ☐ Faux
- (e) L'écart-type échantillonnal est de 11%.
☐ Vrai ☐ Faux

Exercice

★★★★ Soit les 21 valeurs suivantes :

111	112	112	112	113	113	113
113	115	115	115	115	115	116
118	118	118	119	124	124	125

Le diagramme en boîte ci-dessous a été tracé à partir de ces résultats.



Lequel des énoncés suivants est vrai ?

- ☐ La médiane n'est pas au bon endroit.
- ☐ La moyenne échantillonnale n'est pas au bon endroit.
- ☐ Il devrait y avoir deux points : un à 105,5 et un à 125,5.
- ☐ La moustache de gauche devrait être à 105,5.
- ☐ La moustache de droite devrait être à 125,5.
- ☐ Il y a plus qu'une erreur.
- ☐ Il n'y a aucune erreur.

Module 6 - Échantillons aléatoires, lois échantillonnales et estimation ponctuelle

Exercice

★★ Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) Si $X_1, \dots, X_{100}$ constituent un échantillon aléatoire simple issu d'une loi normale d'espérance $\mu = 16$ et de variance $\sigma^2 = 100$, alors la moyenne échantillonnale de ces 100 valeurs suit approximativement une loi normale centrée-réduite.
☐ Vrai ☐ Faux

- (b) Si $X_1, \dots, X_{17}$ constituent un échantillon aléatoire simple issu d'une loi $\mathcal{N}(5; 4)$, alors la variable $4S^2$ a une espérance de 16 et une variance de 32.
☐ Vrai ☐ Faux
- (c) Si W suit une χ_4^2 (khi-carré avec 4 degrés de liberté), alors $P(W > 4) = P(W < 4)$.
☐ Vrai ☐ Faux
- (d) Le poids des tomates d'un certain agriculteur suit une loi normale de moyenne 300 grammes et de variance 100 grammes². On prend des mesures sur 25 tomates choisies aléatoirement et on obtient une moyenne échantillonnale de 309,12 grammes et une variance échantillonnale de 118,3 grammes². Dans cette situation, les nombres 309,12 et 118,3 sont des paramètres et les nombres 300 et 100 sont des statistiques.
☐ Vrai ☐ Faux
- (e) Soit deux échantillons aléatoires indépendants de taille $n = 26$ et $m = 26$ qui proviennent de lois normales $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Les moyennes échantillonnales sont respectivement $\bar{X}$ et $\bar{Y}$, alors que les variances échantillonnales sont respectivement S_X^2 et S_Y^2 . La statistique S_X^2/S_Y^2 suit une $F_{25,25}$.
☐ Vrai ☐ Faux

Exercice

★★ Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) Si $X_1, \dots, X_{100}$ constituent un échantillon aléatoire simple issu d'une loi normale d'espérance $\mu = 16$ et de variance $\sigma^2 = 100$, alors la moyenne échantillonnale de ces 100 valeurs suit approximativement une loi normale centrée-réduite.
☐ Vrai ☐ Faux
- (b) Si $X_1, \dots, X_{17}$ constituent un échantillon aléatoire simple issu d'une loi $\mathcal{N}(5; 4)$, alors la variable $4S^2$ a une espérance de 16 et une variance de 32.
☐ Vrai ☐ Faux
- (c) Si W suit une χ_4^2 (khi-carré avec 4 degrés de liberté), alors $P(W > 4) = P(W < 4)$.
☐ Vrai ☐ Faux
- (d) Le poids des tomates d'un certain agriculteur suit une loi normale de moyenne 300 grammes et de variance 100 grammes². On prend des mesures sur 25 tomates choisies aléatoirement et on obtient une moyenne échantillonnale de 309,12 grammes et une variance échantillonnale de 118,3 grammes². Dans cette situation, les nombres 309,12 et 118,3 sont des paramètres et les nombres 300 et 100 sont des statistiques.
☐ Vrai ☐ Faux
- (e) Soit deux échantillons aléatoires indépendants de taille $n = 26$ et $m = 26$ qui proviennent de lois normales $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Les moyennes échantillonnales sont respectivement $\bar{X}$ et $\bar{Y}$, alors que les variances échantillonnales sont respectivement S_X^2 et S_Y^2 . La statistique S_X^2/S_Y^2 suit une $F_{25,25}$.
☐ Vrai ☐ Faux

Exercice

★★★ (H23, A23) Soit $X_1, \dots, X_9$ un échantillon aléatoire d'une variable $X \sim \mathcal{N}(12, 4)$. Nous avons que $X_1, \dots, X_9$ sont indépendantes.

- (a) Calculez la probabilité que la moyenne échantillonnale soit supérieure à la vraie moyenne $\mu = 12$ par plus de 0,62 écarts-types échantillonnaux.
- (b) Calculez la probabilité que la variance échantillonnale soit inférieure à 1,09.

Exercice

★★★ Soit T , une variable aléatoire Student à 5 degrés de liberté. Nous notons μ_T et σ_T^2 , l'espérance et la variance de T , respectivement. Laquelle des affirmations suivantes est vraie ?

- ☐ $Z = \frac{T - \mu_T}{\sigma_T} \sim \mathcal{N}(0, 1)$
- ☐ Nous savons que $\mu_T = 5$ et $\sigma_T^2 = 10$
- ☐ La fonction de densité de T est symétrique
- ☐ La fonction de répartition de T est symétrique

Exercice

★★★ Soit K , S et F des variables aléatoires indépendantes qui suivent respectivement une khi-carré à 4 degrés de liberté, une Student à 35 degrés de liberté et une Fisher à 3 et 4 degrés de liberté.

Cochez tous les énoncés qui sont **faux** ?

- ☐ $P(S > 35) = 0,50$
- ☐ $P(K < 9,49) = 0,95$
- ☐ $P(F > 6,59) = 0,05$
- ☐ $P(S > a \cap K > b \cap F > c) = P(S > a) \times P(K > b) \times P(F > c)$

Exercice

★★★ Soit deux échantillons aléatoires indépendants provenant de populations normales, tels que :

$$X_1, \dots, X_{16} \sim \mathcal{N}(7, 4),$$

$$Y_1, \dots, Y_{100} \sim \mathcal{N}(8, 9).$$

(a) Quelle est la loi de $A = X_1 + X_2 - X_3$?

- ☐ $A \sim \mathcal{N}(7; 8)$
- ☐ $A \sim \mathcal{N}(7; 12)$
- ☐ $A \sim \mathcal{N}(14; 8)$
- ☐ $A \sim \mathcal{N}(14; 12)$
- ☐ $A \sim \mathcal{N}(21; 8)$
- ☐ $A \sim \mathcal{N}(21; 12)$
- ☐ Aucune de ces réponses

(b) Quelle est la loi de $B = \frac{1}{100} \sum_{i=1}^{100} Y_i$?

- ☐ $B \sim \mathcal{N}(0,08; 9)$
- ☐ $B \sim \mathcal{N}(0,08; 0,81)$
- ☐ $B \sim \mathcal{N}(0,08; 0,09)$
- ☐ $B \sim \mathcal{N}(8; 9)$
- ☐ $B \sim \mathcal{N}(8; 0,81)$
- ☐ $B \sim \mathcal{N}(8; 0,09)$
- ☐ Aucune de ces réponses

(c) Quelle est la loi de $C = X_1$?

- ☐ $C \sim \mathcal{N}(0,4375; 4)$
- ☐ $C \sim \mathcal{N}(0,4375; 0,25)$
- ☐ $C \sim \mathcal{N}(0,4375; 1)$
- ☐ $C \sim \mathcal{N}(7; 4)$
- ☐ $C \sim \mathcal{N}(7; 0,25)$
- ☐ $C \sim \mathcal{N}(7; 1)$
- ☐ Aucune de ces réponses

(d) Quelle est l'espérance de $D = \frac{\bar{Y} - 8}{S_Y/10}$, où S_Y est l'écart-type échantillonnal de Y ?

- ☐ $E(D) = 8$
- ☐ $E(D) = 100$
- ☐ $E(D) = 99$
- ☐ $E(D) = 0$
- ☐ $E(D) = 15$
- ☐ $E(D) = 800$
- ☐ Aucune de ces réponses

Exercice

★★★

Un échantillon aléatoire $X_1, \dots, X_n$ est obtenu à partir d'une loi dont la fonction de densité dépend d'un paramètre inconnu θ . Une prise de décision en ingénierie repose sur la valeur de la probabilité $P(X_i > 0,5)$. Or pour ce modèle, cette probabilité est égale à $e^{-0,5\theta}$.

On vous propose deux estimateurs $\hat{P}_1$ et $\hat{P}_2$ qui ont les propriétés suivantes :

$$\begin{aligned} E(\hat{P}_1) &= e^{-0,5\theta}, & V(\hat{P}_1) &= \frac{e^{-0,5\theta}(1 - e^{-0,5\theta})}{n}, \\ E(\hat{P}_2) &= \theta, & V(\hat{P}_2) &= \frac{\theta^2}{n}. \end{aligned}$$

Quels sont les **trois** énoncés qui sont **vrais** parmi les six énoncés ci-dessous ?

- ☐ $\hat{P}_1$ est un estimateur sans biais de $e^{-0,5\theta}$.
☐ $\hat{P}_2$ est un estimateur sans biais de $e^{-0,5\theta}$.
☐ L'erreur quadratique moyenne de $\hat{P}_1$ en tant qu'estimateur de $e^{-0,5\theta}$ est égale à

$$e^{-\theta} + \frac{e^{-0,5\theta}(1 - e^{-0,5\theta})}{n}.$$

- ☐ L'erreur quadratique moyenne de $\hat{P}_2$ en tant qu'estimateur de $e^{-0,5\theta}$ est égale à

$$(\theta - e^{-0,5\theta})^2 + \frac{\theta^2}{n}.$$

- ☐ L'erreur quadratique moyenne de $\hat{P}_1$ en tant qu'estimateur de $e^{-0,5\theta}$ est égale à $V(\hat{P}_1)$.
☐ L'erreur quadratique moyenne de $\hat{P}_2$ en tant qu'estimateur de $e^{-0,5\theta}$ est égale à $V(\hat{P}_2)$.

Exercice

★★★ Soit $X_1, \dots, X_{25}$, des réalisations indépendantes d'une variable aléatoire normale d'espérance $\mu = 176,45$ et d'écart-type $\sigma = \sqrt{48}$.

- (a) Quelle est la probabilité que la variance échantillonnale soit supérieure à 46,68 ?
(b) Quelle est la probabilité que la moyenne échantillonnale soit supérieure à $\mu + 0,749 S$, où S est l'écart-type échantillonnale ?

Exercice

★★★

Pour chacun des énoncés suivants, inscrivez «Vrai» ou «Faux» dans les cases appropriées.

- (a) Si $X_1, \dots, X_{17}$ constituent un échantillon aléatoire simple issu d'une distribution $\mathcal{N}(5; 4)$, alors la variable $4S^2$ a une espérance de 5 et une variance de 4.
☐ Vrai ☐ Faux
- (b) Si W suit une χ^2 avec 5 degrés de liberté, alors $P(W > 5) = 0,5$.
☐ Vrai ☐ Faux
- (c) Le poids des tomates d'un certain agriculteur suit une loi normale de moyenne 300 grammes et de variance 100 grammes². On prend des mesures sur 25 tomates choisies aléatoirement et on obtient une moyenne échantillonnale de 309,12 grammes et une variance échantillonnale de 118,3 grammes². Dans cette situation, les nombres 309,12 et 118,3 sont des paramètres et les nombres 300 et 100 sont des statistiques.
☐ Vrai ☐ Faux
- (d) Si $T \sim t_{14}$, alors $P(T > 14) = P(T < -14)$.
☐ Vrai ☐ Faux
- (e) Soit deux échantillons aléatoires indépendants de taille $n_1 = 26$ et $n_2 = 26$ qui proviennent de la même loi $\mathcal{N}(\mu, \sigma^2)$. Les moyennes échantillonnales sont respectivement $\bar{X}_1$ et $\bar{X}_2$, alors que les variances échantillonnales sont respectivement S_1^2 et S_2^2 . La statistique S_1^2/S_2^2 suit une $F_{25,25}$.
☐ Vrai ☐ Faux

Exercice

★★★

Soit $X_1, \dots, X_9$ un échantillon aléatoire d'une variable X de loi Normale $\mathcal{N}(12, 4)$. Nous avons que $X_1, \dots, X_9$ sont indépendantes.

- (a) (5 points) Calculez la probabilité que la moyenne échantillonnale soit supérieure à la vraie moyenne $\mu = 12$ par plus de 0,62 écarts-types échantillonnals. Justifiez votre réponse.
- (b) (4 points) Calculez la probabilité que la variance échantillonnale soit supérieure à 1,09. Justifiez votre réponse.

Exercice

★★★ Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) Si $X_1, \dots, X_{100}$ constituent un échantillon aléatoire simple issu d'une loi quelconque d'espérance $\mu = 0$ et d'écart-type $\sigma = 10$, alors la moyenne échantillonnale de ces 100 valeurs suit approximativement une loi normale centrée-réduite.
☐ Vrai ☐ Faux
- (b) Si $X_1, \dots, X_{17}$ constituent un échantillon aléatoire simple issu d'une loi $\mathcal{N}(5; 4)$, alors la variable $16S^2$ a une espérance de 16 et une variance de 32.
☐ Vrai ☐ Faux
- (c) Si W suit une χ_4^2 (khi-carré avec 4 degrés de liberté), alors la densité de W est symétrique.
☐ Vrai ☐ Faux
- (d) Le poids des tomates d'un certain agriculteur suit une loi normale de moyenne 300 grammes et de variance 100 grammes². On prend des mesures sur 25 tomates choisies aléatoirement et on obtient une moyenne échantillonnale de 309,12 grammes et une variance échantillonnale de 118,3 grammes². Dans cette situation, les nombres 309,12 et 118,3 sont des paramètres et les nombres 300 et 100 sont des statistiques.
☐ Vrai ☐ Faux
- (e) Soit deux échantillons aléatoires indépendants de taille $n_1 = 26$ et $n_2 = 26$ qui proviennent de lois normales $\mathcal{N}(\mu_1, \sigma^2)$ et $\mathcal{N}(\mu_2, \sigma^2)$. Les moyennes échantillonnales sont respectivement $\bar{X}_1$ et $\bar{X}_2$, alors que les variances échantillonnales sont respectivement S_1^2 et S_2^2 . La statistique S_1^2/S_2^2 suit une $F_{25,25}$.
☐ Vrai ☐ Faux

Exercice

★★★ Soit X_1 et X_2 deux réalisations d'une variable $X \sim \mathcal{N}(0, 4)$. Nous avons que X_1 et X_2 sont indépendantes et nous notons

$$\bar{X} = \frac{X_1 + X_2}{2},$$

$$S^2 = \sum_{i=1}^2 (X_i - \bar{X})^2.$$

- (a) (7 points) Calculez $P(\bar{X} > S^2)$.
- (b) (6 points) Calculez la probabilité que la variance échantillonnale soit inférieure à 15,36.

Exercice

★★★ Soit deux échantillons aléatoires indépendants provenant de populations normales, tels que :

$$X_1, \dots, X_{16} \sim \mathcal{N}(7, 4),$$

$$Y_1, \dots, Y_{100} \sim \mathcal{N}(8, 9).$$

- (a) (2 points) Quelle est la loi de $A = X_1 + X_2 - X_3$?

- ☐ $A \sim \mathcal{N}(7; 8)$
☐ $A \sim \mathcal{N}(7; 12)$
☐ $A \sim \mathcal{N}(14; 8)$
☐ $A \sim \mathcal{N}(14; 12)$
☐ $A \sim \mathcal{N}(21; 8)$
☐ $A \sim \mathcal{N}(21; 12)$
☐ Aucune de ces réponses

- (b) (2 points) Quelle est la loi de $B = \frac{1}{100} \sum_{i=1}^{100} Y_i$?

- ☐ $B \sim \mathcal{N}(0,08; 9)$
☐ $B \sim \mathcal{N}(0,08; 0,81)$

☐ $B \sim \mathcal{N}(0,08; 0,09)$

☐ $B \sim \mathcal{N}(8; 9)$

☐ $B \sim \mathcal{N}(8; 0,81)$

☐ $B \sim \mathcal{N}(8; 0,09)$

☐ Aucune de ces réponses

(c) (2 points) Quelle est la loi de $C = X_1$?

☐ $C \sim \mathcal{N}(0,4375; 4)$

☐ $C \sim \mathcal{N}(0,4375; 0,25)$

☐ $C \sim \mathcal{N}(0,4375; 1)$

☐ $C \sim \mathcal{N}(7; 4)$

☐ $C \sim \mathcal{N}(7; 0,25)$

☐ $C \sim \mathcal{N}(7; 1)$

☐ Aucune de ces réponses

(d) (2 points) Quelle est l'espérance de $D = \frac{\bar{Y} - 8}{S_Y/10}$, où S_Y est l'écart-type échantillonnal de Y ?

☐ $E(D) = 8$

☐ $E(D) = 100$

☐ $E(D) = 99$

☐ $E(D) = 0$

☐ $E(D) = 15$

☐ $E(D) = 800$

☐ Aucune de ces réponses

Exercice

★★★ Soit $X_1, \dots, X_{121}$, des variables aléatoires provenant d'une loi normale d'espérance $\mu = 0$ et de variance $\sigma^2 = 16$.

Quelle est la probabilité que $\frac{\bar{X}}{S}$ soit plus grand que 0,18, où S est l'écart-type échantillonnal ?

Exercice

★★★ Soit $X_1, \dots, X_{121}$, des variables aléatoires provenant d'une loi normale d'espérance $\mu = 0$ et de variance $\sigma^2 = 25$.

Quelle est la probabilité que $\left| \frac{\bar{X}}{S} \right|$ soit plus grand que 0,26, où S est l'écart-type échantillonnal ?

Exercice

★★★ Soit $X_1, \dots, X_9$ des réalisations indépendantes d'une variable $X \sim \mathcal{N}(\mu, 4)$.

(a) (5 points) Calculez $P(\bar{X} > \mu + 0,625)$.

(b) (5 points) Calculez la probabilité que la variance échantillonnale soit supérieure à 8,765.

Exercice

★★★★ Soit K , G et M des variables aléatoires indépendantes qui suivent respectivement une khi-carré à 4 degrés de liberté, une Student à 3 degrés de liberté et une Fisher à 3 et 4 degrés de liberté.

Que vaut $P(K > 1,06 \cap G > 0,277 \cap M < 16,69)$?

☐ 0,3564 ☐ 0,5346 ☐ 0,0036 ☐ 0,0054 ☐ 0,0396

Exercice

★★★★ Soit $X_1, \dots, X_9$ des réalisations indépendantes d'une variable $X \sim \mathcal{N}(12, 24)$ et $Y_1, \dots, Y_{18}$ des réalisations indépendantes d'une variable $Y \sim \mathcal{N}(21, 24)$. Notez S_X^2 et S_Y^2 , les variances des deux échantillons.

Quelle est la probabilité que S_X^2 soit deux fois plus grande que S_Y^2 ?

Exercice

★★★★ Soit $X_1, \dots, X_5$ un échantillon de taille $n = 5$ provenant d'une variable aléatoire de fonction de répartition F_X , d'espérance μ et de variance σ^2 .

Pour estimer μ , nous voulons comparer la performance de deux estimateurs :

$$\hat{\mu}_1 = \bar{X},$$

$$\hat{\mu}_2 = 0,1X_1 + 0,2X_2 + 0,2X_3 + 0,2X_4 + 0,1X_5.$$

- (a) Lequel des deux estimateurs a la plus faible variance ?
 (b) On sait que $\text{biais}(\hat{\mu}_2, \mu) = -0,2\mu$. En supposant que $\sigma^2 = 1$, pour quelles valeurs de μ a-t-on que l'estimateur $\hat{\mu}_2$ est meilleur que $\hat{\mu}_1$ au niveau du MSE ?

Exercice

★★★★

Soit les variables aléatoires indépendantes $X_1, \dots, X_{10}$. Ces 10 variables aléatoires suivent toutes la même loi normale de moyenne μ et de variance σ^2 . En utilisant

$$Y = \sum_{i=1}^{10} X_i,$$

$$W = \frac{1}{n} \sum_{i=1}^{10} X_i,$$

laquelle des affirmations suivantes est VRAIE ?

- ☐ $\frac{W-\mu}{\sigma} \sim \mathcal{N}(0,1)$
☐ $3Y + 10 \sim \mathcal{N}(30\mu + 10; 90\sigma^2)$
☐ La moyenne échantillonnale W est toujours égale à μ
☐ Comme $n < 30$, Y et W ne suivent pas des lois normales

Exercice

★★★★ Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 2 points, une mauvaise réponse vaut -1 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) Si $X_1, \dots, X_{25}$ constituent un échantillon aléatoire simple issu d'une distribution $\mathcal{N}(0; 1)$, alors la moyenne échantillonnale de ces 25 valeurs suit une loi $\mathcal{N}(0; 1)$.
☐ Vrai ☐ Faux
- (b) Si $X_1, \dots, X_{17}$ constituent un échantillon aléatoire simple issu d'une distribution $\mathcal{N}(5; 4)$, alors la variable $16S^2$ a une espérance de 16 et une variance de 32.
☐ Vrai ☐ Faux
- (c) Si W suit une χ_4^2 (khi-carré avec 4 degrés de liberté), alors la densité de W est symétrique.
☐ Vrai ☐ Faux
- (d) Le poids des tomates d'un certain agriculteur suit une loi normale de moyenne 300 grammes et de variance 100 grammes². On prend des mesures sur 25 tomates choisies aléatoirement et on obtient une moyenne échantillonnale de 309,12 grammes et une variance échantillonnale de 118,3 grammes². Dans cette situation, les nombres 309,12 et 118,3 sont des paramètres et les nombres 300 et 100 sont des statistiques.
☐ Vrai ☐ Faux
- (e) Si $T \sim t_3$ (Student avec 3 degrés de liberté), alors $P(T > 3) = P(T < 3) = 0,50$.
☐ Vrai ☐ Faux
- (f) Soit deux échantillons aléatoires indépendants de taille $n_1 = 26$ et $n_2 = 26$ qui proviennent de lois normales $\mathcal{N}(\mu_1, \sigma^2)$ et $\mathcal{N}(\mu_2, \sigma^2)$. Les moyennes échantillonnales sont respectivement $\bar{X}_1$ et $\bar{X}_2$, alors que les variances échantillonnales sont respectivement S_1^2 et S_2^2 . La statistique S_1^2/S_2^2 suit une $F_{25,25}$.
☐ Vrai ☐ Faux

Exercice

★★★★ Soit $X_1, \dots, X_9$ des réalisations indépendantes d'une variable $X \sim \mathcal{N}(12, 24)$ et $Y_1, \dots, Y_{18}$ des réalisations indépendantes d'une variable $Y \sim \mathcal{N}(21, 24)$. Notez S_X^2 et S_Y^2 , les variances des deux échantillons.

Quelle est la probabilité que S_X^2 soit deux fois plus grande que S_Y^2 ?

Exercice

★★★★ Soit $X_1, \dots, X_5$ un échantillon de taille $n = 5$ provenant d'une variable aléatoire de fonction de répartition F_X , d'espérance μ et de variance σ^2 .

Pour estimer μ , nous voulons comparer la performance de deux estimateurs :

$$\hat{\mu}_1 = \bar{X},$$

$$\hat{\mu}_2 = 0,1X_1 + 0,2X_2 + 0,3X_3 + 0,2X_4 + 0,1X_5.$$

- (a) (5 points) Lequel des deux estimateurs a la plus faible variance ?
 (b) (7 points) On sait que $\text{biais}(\hat{\mu}_2, \mu) = -0,1\mu$. En supposant que $\sigma^2 = 1$, pour quelles valeurs de μ a-t-on que l'estimateur $\hat{\mu}_2$ est meilleur que $\hat{\mu}_1$ au niveau du MSE ?

Exercice

★★★★ Soit $X_1, \dots, X_n$ qui sont indépendants et identiquement distribués selon une loi quelconque F_X . On veut estimer la médiane $\tilde{\mu}$ de F_X à l'aide de $\bar{X}$. Calculez le biais de $\bar{X}$ pour $\tilde{\mu}$.

Exercice

★★★★ Soit $X_1, \dots, X_9$ un échantillon aléatoire d'une variable $X \sim \mathcal{N}(12, 4)$. Nous avons que $X_1, \dots, X_9$ sont indépendantes.

- (a) (4 points) Calculez la probabilité que la moyenne échantillonnale soit supérieure à la vraie moyenne $\mu = 12$ par plus de 0,62 écarts-types échantillonnals.
- (b) (4 points) Calculez la probabilité que la variance échantillonnale soit inférieure à 1,09.

Exercice

★★★★ Laquelle des affirmations suivantes est toujours vraie au sujet de $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, où les X_i sont indépendants et identiquement distribués de loi quelconque d'espérance μ_X et de variance σ_X^2 ?

- ☐ $Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \sim \mathcal{N}(0, 1)$
- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ seulement si $n > 30$
- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ peu importe la loi des X_i
- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ seulement si les X_i sont de loi normale
- ☐ $\text{Var}[\bar{X}] = \sigma_X^2$

Exercice

★★★★ La quantité de pluie (en mm) tombée dans un bassin versant dans un mois suit une loi ayant comme fonction de répartition

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/\theta}, & x \geq 0, \end{cases}$$

où $\theta > 0$ est un paramètre de valeur inconnue. On peut montrer que $E(X) = \theta$ et $V(X) = \theta^2$.

La compagnie qui gère un barrage à la sortie de ce bassin versant a besoin de savoir l'espérance et la variance de la quantité de pluie tombée dans un mois afin d'optimiser la gestion du barrage. L'ingénieure de la compagnie a à sa disposition un échantillon aléatoire $X_1, \dots, X_{20}$ provenant de cette loi, dont elle souhaite se servir pour estimer les

quantités désirées. Soit $\bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$.

- (a) Calculez l'espérance de $\bar{X}$. Est-ce que $\bar{X}$ est un estimateur sans biais de θ ?
- (b) Calculez l'espérance de $\bar{X}^2$. Est-ce que $\bar{X}^2$ est un estimateur sans biais de θ^2 ?
- [Aide : Souvenez-vous que pour toute variable aléatoire Y , $V(Y) = E(Y^2) - \{E(Y)\}^2$.]

Exercice

★★★★ Laquelle des affirmations suivantes est toujours vraie au sujet de $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, où les X_i sont indépendants et identiquement distribués de loi quelconque d'espérance μ_X et de variance σ_X^2 ?

- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ seulement si $n > 30$
- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ peu importe la loi des X_i
- ☐ $\mathbb{E}[\bar{X}] = \mu_X$ seulement si les X_i sont de loi normale
- ☐ $\text{Var}[\bar{X}] = \sigma_X^2$

Module 7 - Intervalles de confiance

Exercice

★ Soit $X_1, \dots, X_{16}$, des variables aléatoires indépendantes de loi normale d'espérance μ et de variance σ^2 . Voici plusieurs intervalles :

1	$\bar{X} \pm t_{\alpha/2;15} \frac{S}{\sqrt{16}}$	6	$\bar{X} \pm z_{\alpha/2} \frac{\sigma^2}{\sqrt{16}}$
2	$\bar{X} \pm t_{\alpha/2;16} \frac{S}{\sqrt{16}}$	7	$\bar{X} \pm t_{\alpha/2;15} \frac{S}{\sqrt{15}}$
3	$\bar{X} \pm t_{\alpha/2;16} \frac{\sigma}{\sqrt{15}}$	8	$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{16}}$
4	$\bar{X} \pm t_{\alpha/2;16} \frac{\sigma}{\sqrt{16}}$	9	$\left[\frac{15S^2}{\chi_{\alpha/2;15}^2}; \frac{15S^2}{\chi_{1-\alpha/2;15}^2} \right]$
5	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{16}}$	10	$\left[\frac{16S^2}{\chi_{\alpha/2;16}^2}; \frac{16S^2}{\chi_{1-\alpha/2;16}^2} \right]$

Associez chacune des situations suivantes au numéro de l'intervalle de confiance de niveau $1 - \alpha$ qui correspond.

(a) (3 points) Intervalle de confiance pour μ lorsque σ^2 est connue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(b) (3 points) Intervalle de confiance pour σ^2 .

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(c) (3 points) Intervalle de confiance pour μ lorsque σ^2 est inconnue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Exercice

★★ Soit $X_1, \dots, X_{25}$, des variables aléatoires indépendantes de loi normale d'espérance μ et de variance σ^2 . Voici plusieurs intervalles :

1	$\bar{X} \pm t_{\alpha/2;24} \frac{S}{\sqrt{25}}$	6	$\bar{X} \pm z_{\alpha/2} \frac{\sigma^2}{\sqrt{25}}$
2	$\bar{X} \pm t_{\alpha/2;25} \frac{S}{\sqrt{25}}$	7	$\bar{X} \pm t_{\alpha/2;24} \frac{S}{\sqrt{24}}$
3	$\bar{X} \pm t_{\alpha/2;25} \frac{\sigma}{\sqrt{24}}$	8	$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{25}}$
4	$\bar{X} \pm t_{\alpha/2;25} \frac{\sigma}{\sqrt{25}}$	9	$\left[\frac{24S^2}{\chi_{\alpha/2;24}^2}; \frac{24S^2}{\chi_{1-\alpha/2;24}^2} \right]$
5	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{25}}$	10	$\left[\frac{25S^2}{\chi_{\alpha/2;25}^2}; \frac{25S^2}{\chi_{1-\alpha/2;25}^2} \right]$

Associez chacune des situations suivantes au numéro de l'intervalle de confiance de niveau $1 - \alpha$ qui correspond. Une seule réponse est bonne pour chaque situation.

(a) Intervalle de confiance pour μ lorsque σ^2 est connue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(b) Intervalle de confiance pour σ^2 .

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(c) Intervalle de confiance pour μ lorsque σ^2 est inconnue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Exercice

★★★ Un intervalle de confiance au niveau 95% pour l'espérance μ est $[10,11; 19,11]$. Pour chaque énoncé qui suit, indiquez si l'énoncé est vrai, faux ou si nous n'avons pas assez d'information pour le savoir.

- (a) Au seuil 5%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.
☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.
- (b) Au seuil 1%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.
☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.
- (c) Au seuil 10%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.
☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.
- (d) Au seuil 5%, on rejette $H_0 : \mu = 28$ pour $H_1 : \mu \neq 28$.
☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

Exercice

★★ Soit $X_1, \dots, X_{49}$, des variables aléatoires indépendantes d'espérance μ et de variance σ^2 . Voici plusieurs intervalles :

1	$\bar{X} \pm t_{\alpha/2;48} \frac{S}{\sqrt{49}}$	6	$\bar{X} \pm z_{\alpha/2} \frac{\sigma^2}{\sqrt{49}}$
2	$\bar{X} \pm t_{\alpha/2;49} \frac{S}{\sqrt{49}}$	7	$\bar{X} \pm t_{\alpha/2;48} \frac{S}{\sqrt{48}}$
3	$\bar{X} \pm t_{\alpha/2;49} \frac{\sigma}{\sqrt{48}}$	8	$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{49}}$
4	$\bar{X} \pm t_{\alpha/2;49} \frac{\sigma}{\sqrt{49}}$	9	$\left[\frac{48S^2}{\chi_{\alpha/2;48}^2}; \frac{48S^2}{\chi_{1-\alpha/2;48}^2} \right]$
5	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{49}}$	10	$\left[\frac{49S^2}{\chi_{\alpha/2;49}^2}; \frac{49S^2}{\chi_{1-\alpha/2;49}^2} \right]$

Associez chacune des situations suivantes au numéro de l'intervalle de confiance de niveau $1 - \alpha$ qui correspond.

- (a) Intervalle de confiance pour μ lorsque σ^2 est connue et que les X_i sont de loi normale.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

- (b) Intervalle de confiance pour σ^2 avec les X_i de loi normale.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

- (c) Intervalle de confiance pour μ lorsque σ^2 est inconnue et que les X_i sont d'une loi quelconque.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Exercice

★★ On a noté la fréquence cardiaque de neuf hommes. L'intervalle de confiance à 90% obtenu pour la fréquence cardiaque moyenne est [94; 112]. On suppose que la loi normale est adéquate pour cette variable aléatoire.

- (a) (1 point) On a confiance à 90% que cet intervalle contient la vraie valeur moyenne de la fréquence cardiaque.

☐ Vrai ☐ Faux

- (b) (1 point) La fréquence cardiaque moyenne des hommes ayant participé à l'étude est de 100 battements par minute.

☐ Vrai ☐ Faux

- (c) (1 point) Si on recommençait la même expérience avec neuf autres individus de la même population, il y aurait 90% de chances que la nouvelle moyenne échantillonnale soit comprise entre 94 et 112.

☐ Vrai ☐ Faux

- (d) (1 point) Environ 90% des hommes de cette population ont une fréquence cardiaque comprise entre 94 et 112.

☐ Vrai ☐ Faux

- (e) (1 point) Environ 90% des hommes de cet échantillon ont une fréquence cardiaque comprise entre 94 et 112.

☐ Vrai ☐ Faux

Exercice

★★ Soit $X_1, \dots, X_{25}$, des variables aléatoires indépendantes de loi normale d'espérance μ et de variance σ^2 . Voici plusieurs intervalles :

1	$\bar{X} \pm t_{\alpha/2;24} \frac{S}{\sqrt{25}}$	6	$\bar{X} \pm z_{\alpha/2} \frac{\sigma^2}{\sqrt{25}}$
2	$\bar{X} \pm t_{\alpha/2;25} \frac{S}{\sqrt{25}}$	7	$\bar{X} \pm t_{\alpha/2;24} \frac{S}{\sqrt{24}}$
3	$\bar{X} \pm t_{\alpha/2;25} \frac{\sigma}{\sqrt{24}}$	8	$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{25}}$
4	$\bar{X} \pm t_{\alpha/2;25} \frac{\sigma}{\sqrt{25}}$	9	$\left[\frac{24S^2}{\chi_{\alpha/2;24}^2}; \frac{24S^2}{\chi_{1-\alpha/2;24}^2} \right]$
5	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{25}}$	10	$\left[\frac{25S^2}{\chi_{\alpha/2;25}^2}; \frac{25S^2}{\chi_{1-\alpha/2;25}^2} \right]$

Associez chacune des situations suivantes au numéro de l'intervalle de confiance de niveau $1 - \alpha$ qui correspond. Une seule réponse est bonne pour chaque situation.

(a) (3 points) Intervalle de confiance pour μ lorsque σ^2 est connue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(b) (3 points) Intervalle de confiance pour σ^2 .

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

(c) (3 points) Intervalle de confiance pour μ lorsque σ^2 est inconnue.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Exercice

★★★ On souhaite obtenir un intervalle de confiance pour μ , l'espérance d'une variable aléatoire X normale de variance 0,81, à l'aide de l'échantillon suivant :

1,1,1,2,2,2,3,3,3,

pour lequel $\sum_{i=1}^n x_i = 18$ et $\sum_{i=1}^9 x_i^2 = 42$.

(a) (5 points) Donnez l'intervalle de confiance de niveau 95%.

(b) (5 points) Quel devrait-être le niveau de l'intervalle de confiance pour qu'il soit de longueur 0,708 ?

Exercice

★★★

Soit l'intervalle de confiance au niveau de confiance $(1 - \alpha)$ pour μ construit à partir d'un échantillon $Y_1, \dots, Y_n$:

$$\left[\bar{Y} - t_{\alpha/2;(n-1)} \sqrt{\frac{S^2}{n}}; \bar{Y} + t_{\alpha/2;(n-1)} \sqrt{\frac{S^2}{n}} \right]$$

où $\bar{Y}$ est la moyenne échantillonnale, $t_{\alpha/2;(n-1)} \sqrt{S^2/n}$ est la marge d'erreur et S^2 est la variance échantillonnale. Quel énoncé parmi les suivants est VRAI ?

- ☐ La probabilité que $\bar{Y}$ soit dans l'intervalle de confiance est $(1 - \alpha)$
- ☐ Nous savons que la proportion des Y se retrouvant dans l'intervalle de confiance est $(1 - \alpha)$
- ☐ Nous savons que la probabilité que μ soit dans l'intervalle de confiance est $(1 - \alpha)$
- ☐ Si n augmente, la longueur de l'intervalle de confiance augmente
- ☐ Si α augmente, la longueur de l'intervalle de confiance augmente
- ☐ Nous savons que l'intervalle de confiance est centré en μ

Exercice

★★★ On a mesuré le temps de réaction (en millisecondes) de seize étudiants. L'intervalle de confiance à 95% obtenu pour le temps de réaction moyen est [42; 54]. On suppose que la loi normale est adéquate pour cette variable aléatoire.

(a) (1 point) On a confiance à 95% que cet intervalle contient la vraie valeur moyenne du temps de réaction.

☐ Vrai ☐ Faux

(b) (1 point) Le temps de réaction moyen des étudiants ayant participé à l'étude est de 50 millisecondes.

☐ Vrai ☐ Faux

(c) (1 point) Si on recommençait la même expérience avec seize autres individus de la même population, il y aurait 95% de chances que la nouvelle moyenne échantillonnale soit comprise entre 42 et 54.

☐ Vrai ☐ Faux

(d) (1 point) Environ 95% des étudiants de cette population ont un temps de réaction compris entre 42 et 54.

☐ Vrai ☐ Faux

(e) (1 point) Environ 95% des étudiants de cet échantillon ont un temps de réaction compris entre 42 et 54.

☐ Vrai ☐ Faux

Exercice

★★★

Un intervalle de confiance au niveau 95% pour la moyenne théorique μ est [5,11; 9,11]. Pour chaque énoncé qui suit, indiquez si l'énoncé est vrai, faux ou si nous n'avons pas assez d'information pour le savoir.

(a) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(b) (1 point) Au seuil 1%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(c) (1 point) Au seuil 10%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(d) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 8$ pour $H_1 : \mu \neq 8$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

Exercice

★★★ Un intervalle de confiance au niveau 95% pour la moyenne théorique μ est [5,11; 9,11]. Pour chaque énoncé qui suit, indiquez si l'énoncé est vrai, faux ou si nous n'avons pas assez d'information pour le savoir.

(a) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(b) (1 point) Au seuil 1%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(c) (1 point) Au seuil 10%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(d) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 8$ pour $H_1 : \mu \neq 8$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

Exercice

★★★

Un intervalle de confiance au niveau 95% pour l'espérance μ est [5,11; 9,11]. Pour chaque énoncé qui suit, indiquez si l'énoncé est vrai, faux ou si nous n'avons pas assez d'information pour le savoir.

(a) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(b) (1 point) Au seuil 1%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(c) (1 point) Au seuil 10%, on rejette $H_0 : \mu = 5$ pour $H_1 : \mu \neq 5$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(d) (1 point) Au seuil 5%, on rejette $H_0 : \mu = 8$ pour $H_1 : \mu \neq 8$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

Exercice

★★★ Soit $X_1, \dots, X_{20}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance μ et de variance σ^2 . Un intervalle de confiance de niveau 95% pour μ est $[21,54; 28,46]$.

(a) (0,5 points) Au seuil 10%, nous rejetons $H_0 : \mu = 23$ pour $H_1 : \mu \neq 23$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

(b) (0,5 points) Au seuil 5%, nous rejetons $H_0 : \mu = 23$ pour $H_1 : \mu \neq 23$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

(c) (0,5 points) Au seuil 1%, nous rejetons $H_0 : \mu = 23$ pour $H_1 : \mu \neq 23$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

(d) (0,5 points) Au seuil 10%, nous rejetons $H_0 : \mu = 29$ pour $H_1 : \mu \neq 29$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

(e) (0,5 points) Au seuil 5%, nous rejetons $H_0 : \mu = 29$ pour $H_1 : \mu \neq 29$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

(f) (0,5 points) Au seuil 1%, nous rejetons $H_0 : \mu = 29$ pour $H_1 : \mu \neq 29$.

☐ Vrai ☐ Faux ☐ Il manque d'information.

Exercice

★★★

Un intervalle de confiance au niveau 95% pour l'espérance μ est $[10,11 ; 19,11]$. Pour chaque énoncé qui suit, indiquez si l'énoncé est vrai, faux ou si nous n'avons pas assez d'information pour le savoir.

(a) (2 points) Au seuil 5%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(b) (2 points) Au seuil 1%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(c) (2 points) Au seuil 10%, on rejette $H_0 : \mu = 10$ pour $H_1 : \mu \neq 10$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

(d) (2 points) Au seuil 5%, on rejette $H_0 : \mu = 18$ pour $H_1 : \mu \neq 18$.

☐ Vrai ☐ Faux ☐ Nous n'avons pas assez d'information pour le savoir.

Exercice

★★★★ Nous nous intéressons aux prix des loyers des étudiants de l'Université Laval. Avec un échantillon aléatoire de $n = 241$ étudiants, nous obtenons l'intervalle de confiance à 95% pour μ suivant : $[550; 1630]$. Cochez tous les énoncés qui sont vrais parmi les suivants. **Une réponse cochée alors qu'elle ne devrait pas l'être entraîne une pénalité de 1 point. Le minimum possible pour cette question est 0 point.**

- ☐ Il y a 95% des étudiants qui ont un loyer entre 550 \$ et 1630 \$
- ☐ Si on choisissait un échantillon de 100 étudiants pour construire un nouvel intervalle de confiance à 95% sur la moyenne, celui-ci aurait une plus faible probabilité de contenir la vraie valeur de μ que celui obtenu à partir de 241 étudiants.
- ☐ Si on choisissait un autre échantillon de 241 étudiants pour construire un intervalle de confiance à 95% sur la moyenne, l'intervalle obtenu aurait nécessairement la même longueur que celui-ci.
- ☐ Si on choisissait plusieurs échantillons de 241 étudiants dans cette population, et qu'on construisait avec chacun d'eux l'intervalle de confiance à 95% pour μ , alors environ 95% de ces intervalles contiendraient la vraie valeur moyenne des loyers des étudiants de l'Université Laval.
- ☐ Si on suppose que le prix des loyers ne suit pas une loi normale, l'intervalle de confiance est toujours valide, car $n \geq 30$, mais son niveau est approximatif.

Exercice

★★★★ On a noté la fréquence cardiaque de neuf hommes. L'intervalle de confiance à 90% obtenu pour la fréquence cardiaque moyenne est [94; 112]. On suppose que la loi normale est adéquate pour cette variable aléatoire.

Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 1 point, une mauvaise réponse vaut -0,5 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) La fréquence cardiaque moyenne des hommes ayant participé à l'étude est de 100 battements par minute.
☐ Vrai ☐ Faux
- (b) Si on recommençait la même expérience avec neuf autres individus de la même population, il y aurait 90% de chances que la nouvelle moyenne échantillonnale soit comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (c) Environ 90% des hommes de cette population ont une fréquence cardiaque comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (d) Environ 90% des hommes de cet échantillon ont une fréquence cardiaque comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (e) La probabilité que la moyenne échantillonnale soit comprise entre 94 et 112 est 90%.
☐ Vrai ☐ Faux
- (f) La probabilité que la moyenne de la population soit comprise entre 94 et 112 est 90%.
☐ Vrai ☐ Faux
- (g) Si on avait utilisé un échantillon de 50 individus, l'intervalle de confiance aurait probablement été plus court.
☐ Vrai ☐ Faux

Exercice

★★★★ On veut calibrer une certaine machine pour qu'elle soit le plus stable possible. On veut d'abord estimer la variation de la taille des pièces produites à l'aide d'un intervalle de confiance, en supposant que la loi normale modélise bien la variable aléatoire en question. Un échantillon de taille 25 a donné une variance échantillonnale de 0,03642 cm². La borne inférieure de l'intervalle de confiance pour la variance est 0,024.

- (a) Quel est le niveau de l'intervalle de confiance ?
- (b) Que vaut la borne supérieure de l'intervalle de confiance ?

Exercice

★★★★ Le vérificateur interne d'une entreprise a effectué une vérification du prix de 28 transactions choisies aléatoirement. Il a obtenu une moyenne échantillonnale de 225\$ avec un écart-type échantillonnal de 20\$. On suppose que le prix d'une transaction est distribué selon une loi normale.

- (a) (6 points) Calculez un intervalle de confiance pour la moyenne théorique du prix avec un niveau de confiance de 95%. Justifiez votre réponse.
- (b) (3 points) Quelle est la marge d'erreur dans l'estimation de la moyenne théorique du prix, pour un niveau de confiance de 99%. Justifiez votre réponse.

Exercice

★★★★ On a noté la fréquence cardiaque de neuf hommes. L'intervalle de confiance à 90% obtenu pour la fréquence cardiaque moyenne est [94; 112]. On suppose que la loi normale est adéquate pour cette variable aléatoire. Déterminez si les énoncés suivants sont vrais ou faux.

Une bonne réponse vaut 1 point, une mauvaise réponse vaut -0,5 point et une réponse vide vaut 0 point, pour un total minimum de 0 point.

- (a) La fréquence cardiaque moyenne des hommes ayant participé à l'étude est de 100 battements par minute.
☐ Vrai ☐ Faux
- (b) Si on recommençait la même expérience avec neuf autres individus de la même population, il y aurait 90% de chances que la nouvelle moyenne échantillonnale soit comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (c) Environ 90% des hommes de cette population ont une fréquence cardiaque comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (d) Environ 90% des hommes de cet échantillon ont une fréquence cardiaque comprise entre 94 et 112.
☐ Vrai ☐ Faux
- (e) La probabilité que la moyenne échantillonnale soit comprise entre 94 et 112 est 90%.
☐ Vrai ☐ Faux
- (f) La probabilité que la moyenne de la population soit comprise entre 94 et 112 est 90%.
☐ Vrai ☐ Faux
- (g) Si on avait utilisé un échantillon de 50 individus, l'intervalle de confiance aurait probablement été plus court.
☐ Vrai ☐ Faux

Exercice

★★★★ On veut calibrer une certaine machine pour qu'elle soit le plus stable possible. On veut d'abord estimer la variation de la taille des pièces produites à l'aide d'un intervalle de confiance, en supposant que la loi normale modélise bien la variable aléatoire en question. Un échantillon de taille 25 a donné une variance échantillonnale de 0,03642 cm². La borne inférieure de l'intervalle de confiance pour la variance est 0,024.

- (a) (6 points) Quel est le niveau de l'intervalle de confiance ?
 (b) (4 points) Que vaut la borne supérieure de l'intervalle de confiance ?

Exercice

★★★★ Soit $X_1, \dots, X_{100}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance μ et de variance $\sigma^2 = 400$. Nous désirons construire un intervalle de confiance approprié de niveau 95% pour μ . Nous notons cet intervalle $[C_1; C_2]$.

- (a) (3 points) Que vaut $\mathbb{E}[C_2]$?
 (b) (3 points) Que vaut $\text{Var}[C_2]$?
 (c) (4 points) Quelle est la probabilité que C_2 soit inférieure ou égale à μ ?

Exercice

★★★★ Soit $X_1, \dots, X_{100}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance μ et de variance $\sigma^2 = 400$. Nous désirons construire un intervalle de confiance approprié de niveau 90% pour μ . Nous notons cet intervalle $[C_1; C_2]$.

- (a) (3 points) Que vaut $\mathbb{E}[C_2]$?
 (b) (3 points) Que vaut $\text{Var}[C_2]$?
 (c) (4 points) Quelle est la probabilité que C_2 soit inférieure ou égale à μ ?

Module 8 - Tests d'hypothèses

Exercice

★ (A23, H24) On veut faire le test d'hypothèse suivant

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2,$$

où la variable 1 suit une loi normale de moyenne μ_1 et de variance σ_1^2 et la variable 2 suit une loi normale de moyenne μ_2 et de variance σ_2^2 . On sélectionne aléatoirement 30 observations de la variable 1 et 30 observations de la variable 2 de façon indépendante. On obtient la sortie Rcmdr suivante :

```
> var.test(x ~ obs, alternative='two.sided', conf.level=.95, data=donnees)
```

F test to compare two variances

data: x by obs

F = 0.15, num df = AAA, denom df = BBB, p-value = 0.00000204

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:
 0.07139472 0.3151494
 sample estimates:
 ratio of variances
 0.15

- (a) Donnez la valeur de AAA, soit le nombre de degrés de liberté associés au numérateur de la statistique de test. Justifiez votre réponse.
- (b) Donnez la valeur de BBB, soit le nombre de degrés de liberté associés au dénominateur de la statistique de test. Justifiez votre réponse.
- (c) On vous indique que la variance échantillonnale de la variable 1, s_1^2 , vaut 6,69. Quelle est la valeur de la variance échantillonnale de la variable 2, soit s_2^2 ? Justifiez votre réponse.
- (d) En utilisant l'information disponible dans la sortie Rcmdr, choisissez l'énoncé qui est vrai.
- ☐ La variance de la variable 1 est significativement différente de celle de la variable 2 au seuil 5%.
- ☐ La variance de la variable 1 n'est pas significativement différente de celle de la variable 2 au seuil 5%.
- ☐ Au seuil 5%, on ne peut rien conclure à partir de la sortie Rcmdr.

Exercice

★ Voici deux mises en situation :

- **Mise en situation 1** : On aimerait comparer les performances de coureurs automobiles avec les pneus A et les performances avec les pneus B. Un groupe de 14 coureurs choisis au hasard parcourront un même circuit deux fois : une fois avec les pneus A, et une fois avec les pneus B. En supposant que le temps de parcours du circuit suit une loi normale, pouvez-vous conclure au seuil $\alpha = 5\%$ que le temps de parcours moyen avec les pneus A est supérieur au temps de parcours moyen avec les pneus B?
- **Mise en situation 2** : Une entreprise fabrique des ampoules jaunes et des ampoules blanches. On aimerait savoir si les ampoules jaunes ont une durée de vie (en heures) supérieure à celle des ampoules blanches. On choisit un échantillon aléatoire de 12 ampoules jaunes et, de façon indépendante, un échantillon aléatoire de 13 ampoules blanches. On suppose que la durée de vie des deux types d'ampoules suivent des lois normales de même variance. Pouvez-vous conclure au seuil $\alpha = 5\%$ que la durée de vie des ampoules jaunes est supérieure à celle des ampoules blanches?

Pour un test d'hypothèse, on note l'hypothèse nulle par H_0 et on note la statistique observée du test par t_0 . Voici des règles de décision de tests d'hypothèse :

- **Règle de décision A** : On rejette H_0 si $t_0 > 1,771$.
- **Règle de décision B** : On rejette H_0 si $t_0 < 1,771$.
- **Règle de décision C** : On rejette H_0 si $|t_0| > 2,069$.
- **Règle de décision D** : On rejette H_0 si $t_0 > 2,069$.
- **Règle de décision E** : On rejette H_0 si $|t_0| > 2,160$.
- **Règle de décision F** : On rejette H_0 si $t_0 > 2,160$.
- **Règle de décision G** : On rejette H_0 si $t_0 > 1,714$.
- **Règle de décision H** : On rejette H_0 si $t_0 < 1,714$.

Associez chaque mise en situation à la règle de décision correspondante.

- (a) La règle de décision correspondant à la mise en situation 1 est la règle de décision

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

- (b) La règle de décision correspondant à la mise en situation 2 est la règle de décision

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

Exercice

★ (A25)

On veut faire le test d'hypothèse suivant

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

$$H_1 : \sigma_X^2 > \sigma_Y^2,$$

où la variable X suit une loi normale d'espérance μ_X et de variance σ_X^2 et la variable Y suit une loi normale d'espérance μ_Y et de variance σ_Y^2 . On sélectionne aléatoirement 10 observations de la variable X et 7 observations de la variable Y de façon indépendante.

La règle de décision du test d'hypothèse au seuil α est de rejeter H_0 si $F_0 > F_{\alpha; AAA; BBB}$, où AAA et BBB sont les degrés de liberté au numérateur et au dénominateur.

- (a) Donnez la valeur de AAA, soit le nombre de degrés de liberté associés au numérateur de la statistique de test.
- (b) Donnez la valeur de BBB, soit le nombre de degrés de liberté associés au dénominateur de la statistique de test.
- (c) On vous indique que $F_0 = 5,91$ et que la variance échantillonnale de la variable X , s_X^2 , vaut 11,82. Quelle est la valeur de la variance échantillonnale de la variable Y , soit s_Y^2 ? Justifiez votre réponse.
- (d) En utilisant $F_0 = 5,91$, choisissez l'énoncé qui est **vrai**.
- ☐ La variance de la variable X est significativement supérieure à celle de la variable Y au seuil 5%.
- ☐ La variance de la variable X n'est pas significativement supérieure à celle de la variable Y au seuil 5%.
- ☐ Au seuil 5%, on ne peut rien conclure.

Exercice



Voici deux mises en situation :

- **Mise en situation 1** : On aimerait comparer les performances de coureurs automobiles avec les pneus A et les performances avec les pneus B. Un groupe de 15 coureurs choisis au hasard parcourront un même circuit deux fois : une fois avec les pneus A, et une fois avec les pneus B. En supposant que le temps de parcours du circuit suit une loi normale, pouvez-vous conclure au seuil $\alpha = 5\%$ que le temps de parcours moyen avec les pneus A est supérieur au temps de parcours moyen avec les pneus B ?
- **Mise en situation 2** : Une entreprise fabrique des ampoules jaunes et des ampoules blanches. On aimerait savoir si les ampoules jaunes ont une durée de vie (en heures) supérieure à celle des ampoules blanches. On choisit un échantillon aléatoire de 15 ampoules jaunes et, de façon indépendante, un échantillon aléatoire de 15 ampoules blanches. On suppose que la durée de vie des deux types d'ampoules suivent des lois normales de même variance. Pouvés-vous conclure au seuil $\alpha = 5\%$ que la durée de vie des ampoules jaunes est supérieure à celle des ampoules blanches ?

Pour un test d'hypothèse, on note l'hypothèse nulle par H_0 et on note la statistique observée du test par t_0 . Voici des règles de décision de tests d'hypothèse :

- **Règle de décision A** : On rejette H_0 si $|t_0| > 2,145$.
- **Règle de décision B** : On rejette H_0 si $t_0 > 2,145$.
- **Règle de décision C** : On rejette H_0 si $|t_0| > 2,048$.
- **Règle de décision D** : On rejette H_0 si $t_0 > 2,048$.
- **Règle de décision E** : On rejette H_0 si $t_0 > 1,761$.
- **Règle de décision F** : On rejette H_0 si $t_0 < 1,761$.
- **Règle de décision G** : On rejette H_0 si $t_0 > 1,701$.
- **Règle de décision H** : On rejette H_0 si $t_0 < 1,701$.

Associez chaque mise en situation à la règle de décision correspondante.

- (a) (3 points) La règle de décision correspondante à la mise en situation 1 est la règle de décision

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

- (b) (3 points) La règle de décision correspondante à la mise en situation 2 est la règle de décision

☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

Exercice



On veut faire le test d'hypothèse suivant

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2,$$

où la variable 1 suit une loi normale de moyenne μ_1 et de variance σ_1^2 et la variable 2 suit une loi normale de moyenne μ_2 et de variance σ_2^2 . On sélectionne aléatoirement 30 observations de la variable 1 et 30 observations de la variable 2 de façon indépendante. On obtient la sortie Rcmdr suivante :

```
> var.test(x ~ obs, alternative='two.sided', conf.level=.95, data=donnees)
```

F test to compare two variances

```
data: x by obs
F = 0.15, num df = A, denom df = B, p-value = 0.00000204
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.07139472 0.3151494
sample estimates:
ratio of variances
0.15
```

- (a) (2 points) Donnez la valeur de A, soit le nombre de degrés de liberté associés au numérateur de la statistique de test.
- (b) (2 points) Donnez la valeur de B, soit le nombre de degrés de liberté associés au dénominateur de la statistique de test.
- (c) (3 points) On vous indique que la variance échantillonnale de la variable 1, s_1^2 , vaut 6,69. Quelle est la valeur de la variance échantillonnale de la variable 2, soit s_2^2 ? Justifiez votre réponse.
- (d) (2 points) En utilisant l'information disponible dans la sortie Rcmdr, choisissez l'énoncé qui est vrai.
- ☐ La variance de la variable 1 est significativement différente de celle de la variable 2 au seuil 5%.
- ☐ La variance de la variable 1 n'est pas significativement différente de celle de la variable 2 au seuil 5%.
- ☐ Au seuil 5%, on ne peut rien conclure à partir de la sortie Rcmdr.

Exercice



On veut faire le test d'hypothèse suivant

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2,$$

où la variable 1 suit une loi normale de moyenne μ_1 et de variance σ_1^2 et la variable 2 suit une loi normale de moyenne μ_2 et de variance σ_2^2 . On sélectionne aléatoirement 30 observations de la variable 1 et 30 observations de la variable 2 de façon indépendante. On obtient la sortie Rcmdr suivante où certaines valeurs ont été remplacées par AAAA et BBBB :

```
> var.test(x ~ obs, alternative='two.sided', conf.level=.95, data=donnees)
```

F test to compare two variances

```
data: x by obs
F = 0.15, num df = AAAA, denom df = BBBB, p-value = 0.00000204
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.07139472 0.3151494
sample estimates:
ratio of variances
0.15
```

- (a) (2 points) Donnez la valeur de AAAA, soit le nombre de degrés de liberté associés au numérateur de la statistique de test.
- (b) (2 points) Donnez la valeur de BBBB, soit le nombre de degrés de liberté associés au dénominateur de la statistique de test.
- (c) (3 points) On vous indique que la variance échantillonnale de la variable 1, s_1^2 , vaut 6,69. Quelle est la valeur de la variance échantillonnale de la variable 2, soit s_2^2 ? **Donnez les détails de votre démarche.**
- (d) (2 points) En utilisant l'information disponible dans la sortie Rcmdr, choisissez l'énoncé qui est vrai.
- ☐ La variance de la variable 1 est significativement différente de celle de la variable 2 au seuil 5%.
- ☐ La variance de la variable 1 n'est pas significativement différente de celle de la variable 2 au seuil 5%.
- ☐ Au seuil 5%, on ne peut rien conclure à partir de la sortie Rcmdr.

Exercice

★ Nous voulons faire le test d'hypothèse suivant

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2,$$

où la variable 1 suit une loi normale de moyenne μ_1 et de variance σ_1^2 et la variable 2 suit une loi normale de moyenne μ_2 et de variance σ_2^2 . Nous sélectionnons aléatoirement 30 observations de la variable 1 et 30 observations de la variable 2 de façon indépendante, et obtenons la sortie Rcmdr suivante où certaines valeurs ont été remplacées par AAA et BBB :

```
> var.test(x ~ obs, alternative='two.sided', conf.level=.95, data=donnees)
```

F test to compare two variances

data: x by obs

F = 0.15, num df = AAA, denom df = BBB, p-value = 0.00000204

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.07139472 0.3151494

sample estimates:

ratio of variances

0.15

(a) (2 points) Donnez la valeur de AAA, le nombre de degrés de liberté associés au numérateur de la statistique de test.

AAA=

(b) (2 points) Donnez la valeur de BBB, le nombre de degrés de liberté associés au dénominateur de la statistique de test.

BBB=

(c) (3 points) On nous indique que la variance échantillonnale de la variable 1, s_1^2 , vaut 6,69. Quelle est la valeur de la variance échantillonnale de la variable 2, soit s_2^2 ?

$s_2^2 =$

(d) (2 points) En utilisant l'information disponible dans la sortie Rcmdr, choisissez l'énoncé qui est vrai.

☐

La variance de la variable 1 est significativement différente de celle de la variable 2 au seuil 5%.

☐

La variance de la variable 1 n'est pas significativement différente de celle de la variable 2 au seuil 5%.

☐

Au seuil 5%, on ne peut rien conclure à partir de la sortie Rcmdr.

Exercice

★★ Nous voulons faire le test d'hypothèse suivant

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2,$$

où la variable 1 suit une loi normale d'espérance μ_1 et de variance σ_1^2 et la variable 2 suit une loi normale d'espérance μ_2 et de variance σ_2^2 . Nous sélectionnons aléatoirement 25 observations de la variable 1 et 30 observations de la variable 2 de façon indépendante et utilisons la statistique

$$F_0 = \frac{S_1^2}{S_2^2},$$

où S_1^2 et S_2^2 sont les variances échantillonnales des variables 1 et 2 respectivement.

(a) Donnez la loi échantillonnale de F_0 , sous H_0 .

(b) Nous obtenons $F_0 = 2,15$. Quelle est la valeur du seuil observé du test ?

(c) Lequel des énoncés suivant est vrai, au seuil $\alpha = 5\%$?

☐

Nous rejetons H_0 .

☐

Nous ne rejetons pas H_0 .

☐

Il nous manque d'information.

Exercice

★★ (A25) On considère que X , le temps que met un train pour effectuer le trajet entre Nyon et Morges, suit une loi normale d'espérance μ et de variance $\sigma^2 = 9$. On désire faire un test d'hypothèse au seuil $\alpha = 2,5\%$, avec les hypothèses suivantes :

$$H_0 : \mu = 15$$

$$H_1 : \mu > 15.$$

On mesure au hasard le temps que prennent 9 trains. Le temps moyen est de 17,8 minutes. Sachant qu'en réalité $\mu = 15,79$, quelle est la puissance du test ?

Exercice

★★

On aimerait comparer la performance de deux types de pneus d'hiver. Un groupe de 21 conducteurs choisis au hasard parcoureront le même trajet de montagne deux fois : une fois avec les pneus A, et une fois avec les pneus B. En supposant que le temps de parcours de la route suit une loi normale, on veut tester au seuil $\alpha = 0,01$ si le temps de parcours moyen avec les pneus A est supérieur au temps de parcours moyen avec les pneus B :

$$H_0 : \mu_A - \mu_B = 0$$

$$H_1 : \mu_A - \mu_B > 0$$

Voici des règles de décision de tests d'hypothèse, où t_0 est la statistique observée du test. Choisissez la règle de décision correspondant à la mise en situation ci-haut.

- ☐ On rejette H_0 si $t_0 < -2,423$.
- ☐ On rejette H_0 si $t_0 > 2,423$.
- ☐ On rejette H_0 si $|t_0| > 2,423$.
- ☐ On rejette H_0 si $t_0 < -2,704$.
- ☐ On rejette H_0 si $t_0 > 2,704$.
- ☐ On rejette H_0 si $|t_0| > 2,704$.
- ☐ On rejette H_0 si $t_0 < -2,528$.
- ☐ On rejette H_0 si $t_0 > 2,528$.
- ☐ On rejette H_0 si $|t_0| > 2,528$.
- ☐ On rejette H_0 si $t_0 < -2,845$.
- ☐ On rejette H_0 si $t_0 > 2,845$.
- ☐ On rejette H_0 si $|t_0| > 2,845$.

Exercice

★★ (A25)

Une entreprise prépare des sachets de fondue et s'intéresse à la variabilité de la quantité de fromage dans un sachet. Elle choisit au hasard 16 sachets. Dans l'échantillon, la quantité moyenne de fromage par sachet est de 398,4 g et la variance échantillonnale est de $25 g^2$. Sachant que X , la quantité de fromage dans un sachet, est une variable aléatoire qui suit une loi normale d'espérance $\mu = 400$ et de variance σ^2 , elle confronte les hypothèses suivantes, à l'aide du test approprié, au seuil $\alpha = 10\%$:

$$H_0 : \sigma^2 = 15$$

$$H_1 : \sigma^2 > 15.$$

(a) Quelle est la valeur de la statistique du test sous H_0 ?

- ☐ 15 ☐ 16 ☐ 25 ☐ -1,28 ☐ -1,652 ☐ 1,28 ☐ 1,652

(b) Le seuil observé du test est 0,05. Choisissez l'énoncé qui est **vrai**.

- ☐ Au seuil $\alpha = 10\%$, la variance de la quantité de fromage est significativement supérieure à $15 g^2$.
- ☐ Au seuil $\alpha = 10\%$, la variance de la quantité de fromage n'est pas significativement supérieure à $15 g^2$.
- ☐ Au seuil $\alpha = 10\%$, la quantité moyenne de fromage est significativement supérieure à $15 g^2$.

- ☐ Au seuil $\alpha = 10\%$, la variance de la quantité de fromage est significativement égale à $15 g^2$.
- ☐ Au seuil $\alpha = 10\%$, on ne peut pas dire si la variance de la quantité de fromage est significativement différente de $15 g^2$.

Exercice

★★ Soit $X_1, \dots, X_n$, des variables aléatoires indépendantes de loi normale d'espérance μ et de variance σ^2 . Voici plusieurs statistiques :

1	$\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	6	$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	11	$\frac{(n-1)S^2}{\sigma_0^2}$	16	z_α
2	$\frac{\bar{X} - \mu_0}{S^2/\sqrt{n}}$	7	$\frac{\bar{X} - \mu_0}{\sigma^2/n}$	12	$\frac{(n-1)S}{\sigma_0}$	17	$t_{\alpha;n-1}$
3	$\frac{\bar{X} - \mu_0}{S^2/n}$	8	$\frac{\bar{X} - \mu_0}{\sigma/n}$	13	$z_{\alpha/2}$	18	$\chi_{\alpha;n-1}^2$
4	$\frac{\bar{X} - \mu_0}{S/n}$	9	$\frac{(n-1)S^2}{\chi_{\alpha/2;n-1}^2}$	14	$t_{\alpha/2;n-1}$	19	$\chi_{1-\alpha/2;n-1}^2$
5	$\frac{\bar{X} - \mu_0}{\sigma^2/\sqrt{n}}$	10	$\frac{(n-1)S^2}{\chi_{1-\alpha/2;n-1}^2}$	15	$\chi_{\alpha/2;n-1}^2$	20	$\chi_{1-\alpha;n-1}^2$

Pour chacune des situations suivantes, inscrivez le numéro de la statistique observée associée dans la case.

- (a) ($1\frac{1}{2}$ points) On veut tester $H_0 : \mu = \mu_0$ contre $H_0 : \mu \neq \mu_0$, lorsque σ^2 est connue.
- (b) ($1\frac{1}{2}$ points) On veut tester $H_0 : \mu = \mu_0$ contre $H_0 : \mu < \mu_0$, lorsque σ^2 est connue.
- (c) ($1\frac{1}{2}$ points) On veut tester $H_0 : \sigma^2 = \sigma_0^2$ contre $H_0 : \sigma^2 > \sigma_0^2$.
- (d) ($1\frac{1}{2}$ points) On veut tester $H_0 : \mu = \mu_0$ contre $H_0 : \mu < \mu_0$, lorsque σ^2 est inconnue.

Exercice

★★ Soit $X_1, \dots, X_4$ et $Y_1, \dots, Y_3$ deux échantillons aléatoires provenant respectivement de lois normales d'espérances μ_X et μ_Y et de variance σ_X^2 et σ_Y^2 , supposées égales. On veut procéder au test de seuil 5% avec les hypothèses

$$H_0 : \mu_X - \mu_Y = 0,$$

$$H_1 : \mu_X - \mu_Y > 0.$$

On obtient $\bar{x} = 26,112$, $\bar{y} = 12$, $s_X^2 = 25$ et $s_Y^2 = 15$. Quel est le seuil observé du test ?

$\alpha^* =$

Exercice

★★★ Dans une manufacture, on considère que X , la variable aléatoire représentant le temps nécessaire pour fabriquer un clavier (en minutes), suit une loi normale de moyenne μ et de variance $\sigma^2 = 4$. On désire faire un test d'hypothèse au seuil $\alpha = 10\%$, avec les hypothèses suivantes :

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5.$$

On sélectionne au hasard un échantillon de 9 claviers. Le temps total pour fabriquer les 9 claviers est de 49,5 minutes. Dans ce qui suit, les quantiles z_α ont été arrondis à 3 décimales.

1	$P\left(\frac{\bar{X}-5}{2/3} > 1,280 \mid \mu = 5\right)$	9	$P\left(\frac{\bar{X}-5,5}{2/3} > 1,280 \mid \mu = 7\right)$
2	$P\left(\frac{\bar{X}-5}{4/9} > 1,280 \mid \mu = 5\right)$	10	$P\left(\frac{\bar{X}-5}{2/3} > 1,280 \mid \mu = 7\right)$
3	$P\left(\frac{\bar{X}-5}{2/3} > 1,645 \mid \mu = 5\right)$	11	$P\left(\frac{\bar{X}-5}{2/3} > 1,645 \mid \mu = 7\right)$
4	$P\left(\frac{\bar{X}-5}{4/9} > 1,645 \mid \mu = 5\right)$	12	$P\left(\frac{\bar{X}-7}{2/3} > 1,280 \mid \mu = 7\right)$
5	$P\left(\frac{\bar{X}-7}{2/3} > 1,280 \mid \mu = 5\right)$	13	$P\left(\frac{\bar{X}-5}{4/9} > 1,280 \mid \mu = 7\right)$
6	$P\left(Z > \frac{5,5-7}{2/3}\right)$	14	$P\left(Z > \frac{5,5-5}{4/9}\right)$
7	$P(\bar{X} > 5)$	15	$P(\bar{X} > 7)$
8	$P\left(Z > \frac{5,5-5}{2/3}\right)$	16	$P\left(Z > \frac{5-7}{2/3}\right)$

Sachant qu'en vérité $\mu = 7$, complétez correctement les énoncés suivants.

(a) La puissance de test est donnée par le numéro

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16

(b) Le seuil observé pour cet échantillon est donné par le numéro

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16

Exercice

★★★ Une entreprise fabrique des palettes de chocolat et s'intéresse à la quantité moyenne de sucre dans une palette. Elle choisit au hasard 16 palettes. Dans l'échantillon, la quantité moyenne de sucre par palette est de 5,4 g et la variance échantillonnale est de 5,29 g². Sachant que X , la quantité de sucre dans une palette, est une variable aléatoire qui suit une loi normale d'espérance μ et de variance $\sigma^2 = 4$, elle fait le test suivant, au seuil $\alpha = 10\%$,

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5.$$

(a) Quelle est la valeur de la statistique du test sous H_0 , arrondie à 3 décimales ?

☐ 1,341 ☐ 1,600 ☐ 0,800 ☐ 0,696 ☐ 1,210

(b) Le seuil observé du test est 0,212. Choisissez l'énoncé qui est vrai.

- ☐ Au seuil $\alpha = 5\%$, la quantité moyenne de sucre dans une palette est significativement supérieure à 5 g.
☐ Au seuil $\alpha = 5\%$, la quantité moyenne de sucre dans une palette n'est pas significativement supérieure à 5 g.
☐ Au seuil $\alpha = 5\%$, la variance de la quantité de sucre dans une palette est significativement supérieure à 5 g.
☐ Au seuil $\alpha = 5\%$, on ne peut pas dire si la quantité moyenne de sucre dans une palette est significativement différente de 5 g.

Exercice

★★★ Deux populations sont distribuées selon des lois normales de moyennes μ_X et μ_Y et de variances σ_X^2 et σ_Y^2 inconnues qui sont supposées inégales. Soit les échantillons

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu_1, \sigma_1^2),$$

$$Y_1, \dots, Y_m \stackrel{i.i.d.}{\sim} N(\mu_2, \sigma_2^2),$$

où $n = 8$ et $m = 8$, et les hypothèses

$$H_0 : \mu_X - \mu_Y = 0,$$

$$H_1 : \mu_X - \mu_Y > 0.$$

Sachant que $\alpha^* = 0,01$, $\bar{y} = 100$, $s_X = 12$ et $s_Y = 12$, que vaut $\bar{x}$?

Exercice

★★★ Amandine s'intéresse au poids moyen des citrons que produisent ses citronniers. Habituellement, le poids moyen des citrons est de 125 g. Elle veut savoir si son nouvel engrais a un impact positif sur le poids des citrons.

Les citrons qui poussent à l'aide du nouvel engrais ont une variance théorique de 196 g². Elle choisit au hasard 25 citrons ayant poussé à l'aide de son nouvel engrais. Dans l'échantillon, le poids moyen des citrons est de 132 g et la variance échantillonnale est de 169 g².

(a) Quelles hypothèses devra-t-elle confronter ?

☐ $H_0 : \mu = 125$ et $H_1 : \mu \neq 125$

☐ $H_0 : \mu = 125$ et $H_1 : \mu > 125$

☐ $H_0 : \mu = 125$ et $H_1 : \mu < 125$

☐ $H_0 : \mu = 132$ et $H_1 : \mu \neq 132$

☐ $H_0 : \mu = 132$ et $H_1 : \mu > 132$

☐ $H_0 : \mu = 132$ et $H_1 : \mu < 132$

☐ Nous ne pouvons pas connaître les hypothèses, car nous ne connaissons pas le seuil α .

(b) Supposons qu'elle désire faire un test unilatéral à droite au seuil $\alpha = 1,5\%$.

i. Quelle est la valeur critique du test, arrondie à 3 décimales ?

☐ 1,885 ☐ 2,170 ☐ 2,692 ☐ 2,500 ☐ 2,064 ☐ 1,711 ☐ 0,0150 ☐ 1,500

ii. Le seuil observé du test est 0,00621. Choisissez tous les énoncés qui sont vrais. **Une réponse cochée alors qu'elle ne devrait pas l'être entraîne une pénalité de 1 point. Le minimum possible pour cette question est 0 point.**

☐ Au seuil $\alpha = 1,5\%$, Amandine rejette H_0 .

☐ Au seuil $\alpha = 1,5\%$, Amandine ne rejette pas H_0 .

☐ Au seuil $\alpha = 1,5\%$, la variance du poids des citrons est significativement supérieure à 196 g.

☐ Au seuil $\alpha = 1,5\%$, nous ne pouvons pas dire si le poids moyen des citrons est influencé par le nouvel engrais.

☐ Amandine rejetterait H_0 pour tous les seuils supérieurs à 0,00621.

☐ Amandine rejetterait H_0 pour tous les seuils inférieurs à 0,00621.

Exercice

★★★

Une entreprise fabrique des palettes de chocolat et on s'intéresse à la quantité moyenne de sucre dans une palette. On choisit au hasard 16 palettes. Dans l'échantillon, la quantité moyenne de sucre par palette est de 5,4 g et la variance échantillonnale est de 4 g². On suppose que X , la quantité de sucre dans une palette, est une variable aléatoire qui suit une loi normale de moyenne μ . On fait le test suivant, au seuil $\alpha = 10\%$,

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5.$$

(a) (3 points) Quelle est la valeur de la statistique du test t_0 ?

☐ 1,645 ☐ 1,600 ☐ 0,800 ☐ 1,341 ☐ 1,753

(b) (2 points) On nous indique que le seuil observé du test est 0,218. Choisissez l'énoncé qui est vrai.

☐ Au seuil $\alpha = 10\%$, la quantité moyenne de sucre dans une palette est significativement supérieure à 5 g.

☐ Au seuil $\alpha = 10\%$, la quantité moyenne de sucre dans une palette n'est pas significativement supérieure à 5 g.

☐ Au seuil $\alpha = 10\%$, la variance de la quantité de sucre dans une palette est significativement supérieure à 5 g.

☐ Au seuil $\alpha = 10\%$, on ne peut pas dire si la quantité moyenne de sucre dans une palette est significativement différente de 5 g.

Exercice

★★★

Dans une manufacture, on considère que X , la variable aléatoire représentant le temps nécessaire pour fabriquer un clavier (en minutes), suit une loi normale de moyenne μ et de variance $\sigma^2 = 4$. On désire faire un test d'hypothèse au seuil $\alpha = 10\%$, avec les hypothèses suivantes :

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5.$$

On sélectionne au hasard un échantillon de 9 claviers. Le temps total pour fabriquer les 9 claviers est de 49,5 minutes. Dans ce qui suit, les quantiles z_α ont été arrondis à 3 décimales.

Expression 1	$P\left(\frac{\bar{X}-5}{2/3} > 1,280 \mid \mu = 5\right)$	Expression 9	$P\left(\frac{\bar{X}-5,5}{2/3} > 1,280 \mid \mu = 7\right)$
Expression 2	$P\left(\frac{\bar{X}-5}{4/9} > 1,280 \mid \mu = 5\right)$	Expression 10	$P\left(\frac{\bar{X}-5}{2/3} > 1,280 \mid \mu = 7\right)$
Expression 3	$P\left(\frac{\bar{X}-5}{2/3} > 1,645 \mid \mu = 5\right)$	Expression 11	$P\left(\frac{\bar{X}-5}{2/3} > 1,645 \mid \mu = 7\right)$
Expression 4	$P\left(\frac{\bar{X}-5}{4/9} > 1,645 \mid \mu = 5\right)$	Expression 12	$P\left(\frac{\bar{X}-7}{2/3} > 1,280 \mid \mu = 7\right)$
Expression 5	$P\left(\frac{\bar{X}-7}{2/3} > 1,280 \mid \mu = 5\right)$	Expression 13	$P\left(\frac{\bar{X}-5}{4/9} > 1,280 \mid \mu = 7\right)$
Expression 6	$P\left(Z > \frac{5,5-7}{2/3}\right)$	Expression 14	$P\left(Z > \frac{5,5-5}{4/9}\right)$
Expression 7	$P(\bar{X} > 5)$	Expression 15	$P(\bar{X} > 7)$
Expression 8	$P\left(Z > \frac{5,5-5}{2/3}\right)$	Expression 16	$P\left(Z > \frac{5-7}{2/3}\right)$

Sachant qu'en vérité $\mu = 7$, complétez correctement les énoncés suivants.

(a) (4 points) La puissance de test est donnée par l'expression

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16

(b) (4 points) Le seuil observé pour cet échantillon est donné par l'expression

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16

Exercice

★★★ Une entreprise fabrique des barres de chocolat et s'intéresse à la quantité moyenne de sucre dans une barre. Elle choisit au hasard 16 barres. Dans l'échantillon, la quantité moyenne de sucre par barre est de 5,05 g et la variance échantillonnaire est de 0,011 g². Sachant que X , la quantité de sucre dans une barre, est une variable aléatoire qui suit une loi normale d'espérance μ et de variance $\sigma^2 = 0,01$, elle effectue un test, au seuil $\alpha = 2,5\%$, avec les hypothèses suivantes :

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5.$$

(a) (2 points) Quelle est la valeur critique du test en grammes, c'est-à-dire, à partir de quelle valeur de $\bar{X}$ rejette-t-on H_0 ?

☐ 1,960 ☐ 2,131 ☐ 2,000 ☐ 5,049 ☐ 5,053

(b) (2 points) Quelle est la valeur de la statistique du test sous H_0 , arrondie à 3 décimales ?

☐ 1,960 ☐ 2,131 ☐ 2,000 ☐ 5,049 ☐ 5,053

(c) (2 points) Quel est le seuil observé du test ?

☐ $\alpha^* = 0,025$.

☐ $\alpha^* = 0,02275$.

☐ $\alpha^* \in]0,025; 0,05[$.

☐ $\alpha^* \in]0,05; 0,1[$.

(d) (4 points) Si en réalité $\mu = 5,1$ et que la règle de décision est de rejeter H_0 si $\frac{\bar{X} - 5}{0,025} > 1,96$, quelle est l'expression de la puissance du test ?

☐ $P\left(\frac{\bar{X} - 5,1}{0,025} > 1,96\right)$

☐ $P\left(\frac{\bar{X} - 5,1}{0,025} > -2,04\right)$

☐ $P\left(\frac{\bar{X} - 5}{0,025} > 1,96\right)$

☐ $P\left(\frac{\bar{X} - 5}{0,025} > -2,04\right)$

Exercice

★★★★ Amandine s'intéresse à la quantité moyenne de sel par kilogramme de gruyère. Elle choisit au hasard 16 morceaux de 1 kilogramme. Dans l'échantillon, la quantité moyenne de sel par kilogramme est de 3,4 g et la variance échantillonnale est de 4,41 g². Sachant que X , la quantité de sel, est une variable aléatoire qui suit une loi normale d'espérance μ et de variance $\sigma^2 = 4 \text{ g}^2$, elle confronte les hypothèses suivantes, au seuil $\alpha = 5\%$,

$$H_0 : \mu = 3$$

$$H_1 : \mu > 3.$$

(a) Quelle est la valeur critique du test, arrondie à 3 décimales ?

☐ 1,645 ☐ 1,960 ☐ 0,800 ☐ 0,762 ☐ 2,131 ☐ 1,753

☐ Comme $n < 30$, il est impossible de faire un test d'hypothèse valide.

(b) Le seuil observé du test est 0,212. Choisissez tous les énoncés qui sont vrais. **Une réponse cochée alors qu'elle ne devrait pas l'être entraîne une pénalité de 1 point. Le minimum possible pour cette question est 0 point.**

☐ Au seuil $\alpha = 5\%$, Amandine rejette H_0 .

☐ Au seuil $\alpha = 5\%$, Amandine ne rejette pas H_0 .

☐ Au seuil $\alpha = 5\%$, la variance de la quantité de sel dans un kilogramme est significativement supérieure à 3 g.

☐ Au seuil $\alpha = 5\%$, nous ne pouvons pas dire si la quantité moyenne de sel dans un kilogramme est significativement différente de 3 g.

☐ Amandine rejetterait H_0 pour tous les tests dont le seuil est supérieur ou égal à 0,212.

☐ Amandine rejetterait H_0 pour tous les tests dont le seuil est inférieur ou égal à 0,212.

Exercice

★★★★ Soit $X_1, \dots, X_{36}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance μ et de variance $\sigma^2 = 1296$. Nous désirons faire un **test unilatéral à droite** avec $H_0 : \mu = 10$ au seuil $\alpha = 0,025$. Nous avons les résultats suivants :

$$\bar{x} = 11,34$$

$$s^2 = 754.$$

Si nous savons qu'en réalité $\mu = \mu_1$ et que la puissance du test est 0,13567, que vaut μ_1 ?

Exercice

★★★★ (a24) On veut savoir si les élèves de 16 ans sont plus forts en mathématiques en juin qu'en septembre. On choisit 6 étudiants au hasard qui passent un examen similaire avant et après les vacances. Voici les résultats :

Examen	Étudiant						Moyenne
	1	2	3	4	5	6	
Examen 1	77	74	82	73	87	AAA	???
Examen 2	72	68	76	68	85	63	72

Répondez aux questions suivantes en sachant qu'il s'agit d'échantillons appariés et en utilisant les quantités suivantes :

$$s_D^2 = 2,8$$

$$\bar{d} = 4.$$

(a) Que vaut AAA ? Justifiez votre réponse.

(b) En confrontant les hypothèses $H_0 : \mu_1 = \mu_2$ et $H_1 : \mu_1 > \mu_2$ au seuil $\alpha = 5\%$, quelle est la loi de la statistique du test ?

- ☐ t_5
- ☐ t_6
- ☐ t_{11}
- ☐ t_{12}
- ☐ $t_{0,05;5}$
- ☐ $t_{0,025;5}$
- ☐ $t_{0,05;6}$
- ☐ $t_{0,025;6}$
- ☐ $t_{0,05;11}$
- ☐ $t_{0,025;11}$
- ☐ $t_{0,05;12}$
- ☐ $t_{0,025;12}$

(c) Pourquoi est-il probablement préférable d'utiliser des échantillons appariés dans ce contexte ? Une seule réponse est bonne.

- ☐ Car l'examen 1 est plus facile que l'examen 2.
- ☐ Car nous supposons que tous les élèves ont environ le même niveau en mathématiques.
- ☐ Car nous voulons nous protéger contre la possibilité que la difficulté des examens influence la décision du test.
- ☐ Car nous voulons nous protéger contre la possibilité que la force des élèves influence la décision du test.
- ☐ Car nous rejetons H_0 au seuil $\alpha = 5\%$.
- ☐ Car nous ne rejetons pas H_0 au seuil $\alpha = 5\%$.
- ☐ Car il s'agit d'observations prises dans le temps.

Exercice

★★★★ Soit $X_1, \dots, X_{16}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance $\mu = 36$ et de variance σ^2 . Nous désirons faire un **test unilatéral à droite** avec $H_0 : \sigma^2 = 33$ au seuil $\alpha = 0,025$. Nous avons les résultats suivants :

$$\bar{x} = 34,56$$

$$s^2 = 42,1.$$

Si nous savons qu'en réalité $\sigma^2 = \sigma_1^2$ et que la puissance du test est 0,05, que vaut σ_1^2 ? Justifiez toutes les étapes de votre démarche.

Exercice

★★★★ Aurélien suppose que le temps d'attente à l'hôpital suit une loi exponentielle de moyenne 240 minutes. Il veut construire un intervalle de confiance bilatéral à 95% pour le vérifier. Il procède donc comme suit. Il note le temps d'attente de 16 personnes choisies au hasard et obtient $\bar{x} = 250$ et $s^2 = 59'536$. Comme il construit un intervalle de

confiance, il choisit $z_{\alpha/2} = 1,96$, à partir de sa table de la loi normale. Son intervalle de confiance approximatif à 95% est donc $[130,44; 369,56]$. Comme celui-ci contient 240, il se dit que cette valeur est plausible pour μ . Après avoir vérifié les résultats d'Aurélien, vous constatez qu'il y a une erreur dans ce qu'il a fait. Quelle est cette erreur ?

Exercice

★★★★ On suppose deux populations distribuées selon des lois normales de moyennes μ_1 et μ_2 et de variances σ_1^2 et σ_2^2 inconnues qui sont supposées égales. Vous avez les échantillons

$$X_{11}, \dots, X_{1n_1} \stackrel{i.i.d.}{\sim} N(\mu_1, \sigma_1^2), \quad X_{21}, \dots, X_{2n_2} \stackrel{i.i.d.}{\sim} N(\mu_2, \sigma_2^2),$$

où $n_1 = 8$ et $n_2 = 8$. Vous avez construit un intervalle de confiance à 95% pour $\mu_1 - \mu_2$ et obtenu $[0,1; 4,39]$. Lors de la présentation des résultats, vous réalisez que vous avez pris en note que $S_1^2 = 4,52$, mais vous ne trouvez plus la valeur de S_2^2 . Quelle est la valeur de S_2^2 ? **Donnez les détails de votre démarche.**

Exercice

★★★★ Soit $X_1, \dots, X_{36}$, des variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance μ et de variance $\sigma^2 = 1296$. On désire faire un **test unilatéral à droite** avec $H_0 : \mu = 10$ au seuil $\alpha = 0,025$. On a les résultats suivants :

$$\bar{x} = 11,34 \quad s^2 = 754.$$

Si on sait qu'en réalité $\mu = \mu_1$ et que la puissance du test est 0,13567, que vaut μ_1 ? **Donnez les détails de votre démarche.**

Exercice

★★★★ Supposons que les variables aléatoires $X_1, \dots, X_{20}$ suivent une loi exponentielle décalée de paramètre $\theta \geq 0$, dont la **fonction de densité** est donnée par

$$f_X(x) = \begin{cases} e^{-(x-\theta)}, & \text{si } \theta \leq x, \\ 0, & \text{sinon.} \end{cases}$$

Nous aimerions faire un test statistique pour savoir si $\theta = 0$. Nous avons donc les hypothèses

$$H_0 : \theta = 0 \quad H_1 : \theta > 0.$$

Pour ce test, nous utiliserons la statistique

$$M = \min(X_1, \dots, X_{20}),$$

qui correspond à la valeur minimale de notre échantillon (nous pouvons aussi écrire $M = X_{(1)}$, dans la notation introduite au cours). Nous savons que sous H_0 , M suit une loi exponentielle de paramètre $n = 20$. Sa **fonction de répartition** est donnée par

$$F_M(m) = \begin{cases} 1 - e^{-20m}, & \text{si } 0 \leq m \\ 0, & \text{sinon.} \end{cases}$$

La table suivante donne quelques valeurs de cette fonction de répartition.

m	$F_M(m)$
0,0005	0,01
0,0025	0,05
0,0053	0,10
0,1151	0,90
0,1498	0,95
0,2303	0,99

Nous avons l'échantillon $x_1, \dots, x_{20}$ suivant :

0,1613	0,1933	0,2980	0,3331	0,3423	0,4392	0,4472	0,4580	0,5289	0,5530
0,5570	0,7737	0,8574	0,9376	1,2509	1,4475	1,9064	2,6019	2,8685	2,9829

- (a) (3 points) Au seuil $\alpha = 5\%$, quelle est la **décision** du test. **Une décision sans justification n'obtiendra aucun point, même si la décision est bonne.**
- (b) (3 points) **Calculez** la valeur-p du test en ne gardant que 4 chiffres après la virgule. **Expliquez votre réponse !**

Exercice

★★★★ Soit $X_1, \dots, X_n$ provenant d'une loi normale d'espérance μ_X et de variance σ_X^2 et $Y_1, \dots, Y_n$ provenant d'une loi normale d'espérance μ_Y et de variance σ_Y^2 . On procède à un test d'égalité des variances avec $H_0 : \sigma_X^2 = \sigma_Y^2$ et

$H_1 : \sigma_X^2 > \sigma_Y^2$, au seuil α , en utilisant la statistique $F_0 = \frac{S_Y^2}{S_X^2}$.

- ☐ Il s'agit d'un test bilatéral.
- ☐ Il s'agit d'un test unilatéral à droite.
- ☐ Il s'agit d'un test unilatéral à gauche.
- ☐ Il nous manque d'information.

Module 9 - ANOVA à un facteur

Exercice

★

On aimerait déterminer si la taille des pétales des iris diffère d'une espèce à une autre. Pour chacune des trois espèces d'iris (Iris setosa, versicolor, virginica), on prend un échantillon de 41 fleurs et on mesure la taille d'un pétale. On fait une analyse de la variance à un facteur en supposant que le modèle et les postulats sont adéquats. On obtient les résultats suivants.

```
> summary(aov(Petal.Taille ~ Espece, data = iris))
              Df    Sum Sq   Mean Sq    F value    Pr(>F)
Espece         2     348.1    174.04      1016     0.0006
Residuals    120      20.6      0.17
```

```
> numSummary(iris$Petal.Taille, groups = iris$Espece, statistics = c("mean", "sd"))
      mean      sd data:n
setosa  1.451220  0.1804466   41
versicolor 4.270732  0.4817904   41
virginica  5.463415  0.4993777   41
```

Calculez la valeur de la plus petite différence significative (PPDS) au seuil de 5%, et utilisez-la pour déterminer si la taille moyenne des pétales de l'espèce setosa est significativement différente de la taille moyenne des pétales de l'espèce versicolor.

Exercice

★ Un ingénieur veut vérifier s'il existe une différence de rigidité d'un certain type de disque, selon le pourcentage de *Pouliovachium* qui le compose.

Il compare donc la rigidité de 8 pourcentages différents (10%, 20%, 25%, 30%, 35%, 40%, 45% et 50%) en produisant 4 disques pour chaque pourcentage. La table d'ANOVA suivante est produite par le logiciel utilisé par l'ingénieur. Dans celle-ci, certaines quantités ont été remplacées par AAA, BBB et ????. Nous considérons que les postulats du modèle sont respectés.

ANOVA

	SS	ddl	MS	F0
pouliovachium	1154.3	???	???	AAA
Residuals	141.4	???	???	
Total	BBB	???		

(a) (3 points) Que vaut AAA ?

- ☐ 164,9
- ☐ 8,163
- ☐ 27,989
- ☐ 1154,3
- ☐ Il n'est pas possible de trouver la valeur de AAA avec les informations données.

(b) (3 points) Quelle est l'estimation de σ^2 ?

- ☐ 141,4

☐ 24☐ 5,89☐ 164,9☐ Il n'est pas possible de donner une estimation de σ^2 avec les informations données.

(c) (3 points) Que vaut BBB ?

☐ 8,163☐ 1295,7☐ 1012,9☐ 0,122☐ Il n'est pas possible de trouver la valeur de BBB avec les informations données.

(d) (3 points) Au seuil de 5%, quelle conclusion pouvons-nous tirer de la table d'ANOVA ?

☐ Il n'est pas possible de tirer de conclusion car nous ne connaissons pas les valeurs remplacées par ???.☐ Il faudrait répéter l'expérience plus de fois pour pouvoir conclure.☐ Toutes les moyennes sont significativement différentes.☐ La rigidité moyenne des disques d'au moins un pourcentage est significativement différente de celle d'au moins un autre pourcentage.☐ Toutes les moyennes sont significativement égales.

Exercice

★★ Soit la table d'analyse de la variance suivante obtenue à l'aide de Rcmdr :

```
>summary(ANOVAModel)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Raclette	4	926.12	231.53	238.72	0.0001
Residuals	48	46.55	0.97		

À partir de la table d'ANOVA, quelle affirmation parmi les suivantes est vraie ? Une seule réponse est bonne.

☐ La proportion de la variance totale expliquée par le facteur Raclette est égale à 0,0001.☐ Il y a 4 niveaux différents pour le facteur Raclette.☐ La variance du facteur Raclette est environ 20 fois plus grande que la variance du facteur Residuals.☐ L'hypothèse nulle supposait que les moyennes des différents niveaux du facteur Raclette étaient toutes égales.☐ L'hypothèse nulle supposait que les moyennes des différents niveaux du facteur Residuals étaient toutes égales.☐ Il s'agit d'une ANOVA à blocs aléatoires.☐ Il manque la ligne d'interaction entre Raclette et Residuals.

Exercice

★★ Un ingénieur veut vérifier s'il existe une différence de rigidité d'un certain type de cylindre, selon le pourcentage de *Vachopouliotum* qui le compose.

Il compare donc la rigidité de 6 pourcentages différents (10%, 20%, 25%, 30%, 35%, et 40%) en produisant 5 cylindres pour chaque pourcentage. La table d'ANOVA suivante est produite par Rcmdr. Dans celle-ci, certaines quantités ont été remplacées par AAA, BBB, CCC et ???.

Nous considérons que les postulats du modèle sont respectés.

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pourcentage	???	278.85	???	AAA	0.01
Residuals	BBB	343.20	???		
Total	???	CCC			

(a) Que vaut AAA ?

☐ 55,77

- ☐ 278,85
☐ 3,90
☐ 0,8125
☐ Il n'est pas possible de trouver la valeur de AAA avec les informations données.

(b) Que vaut BBB ?

- ☐ 5
☐ 6
☐ 24
☐ 25
☐ 29
☐ 30
☐ Il n'est pas possible de trouver la valeur de BBB avec les informations données.

(c) Que vaut CCC ?

- ☐ 1,231
☐ 64,35
☐ 622,05
☐ 0,8125
☐ Il n'est pas possible de trouver la valeur de CCC avec les informations données.

(d) Au seuil de 5%, quelle conclusion pouvons-nous tirer de la table d'ANOVA ? Une seule réponse est bonne.

- ☐ Il n'est pas possible de tirer de conclusion car nous ne connaissons pas les valeurs remplacées par ???.
☐ Il faudrait répéter l'expérience plus de fois pour pouvoir conclure.
☐ Toutes les moyennes sont significativement différentes.
☐ La rigidité moyenne des cylindres d'au moins un pourcentage est significativement différente de celle d'au moins un autre pourcentage.
☐ Toutes les moyennes sont significativement égales.

Exercice

★★ Soit la table d'ANOVA suivante obtenue à l'aide de Rcmdr :

```
>summary(ANOVAModel)
      Df  Sum Sq Mean Sq  F value    Pr(>F)
Fromages  4   926.12   231.53   238.72  0.0001
Residuals 48    46.55     0.97
```

À partir de la table d'ANOVA, quelle affirmation parmi les suivantes est vraie ?

- ☐ La proportion de la variance totale expliquée par le facteur Fromages est égale à 0,0001.
☐ Nous avons 4 niveaux différents pour le facteur Fromages.
☐ La variance du facteur Fromages est environ 20 fois plus grande que la variance du facteur Residuals.
☐ L'hypothèse nulle supposait que les moyennes des différents niveaux du facteur Fromages étaient toutes égales.

Exercice

★★★ Soit la table d'ANOVA suivante obtenue à l'aide de Rcmdr :

```
>summary(ANOVAModel)
      Df  Sum Sq Mean Sq  F value    Pr(>F)
Fromages  4   926.12   231.53   238.72 0.000001
Residuals 48    46.55     0.97
```

À partir de la table d'ANOVA, quelle affirmation parmi les suivantes est vraie ?

- ☐ L'estimation de la moyenne du facteur Fromages est égale à 231,53.

- ☐ Nous avons 49 observations dans le jeu de données.
- ☐ La variance du facteur Fromages est environ 20 fois plus grande que la variance du facteur Residuals.
- ☐ L'hypothèse nulle supposait que les moyennes des différents niveaux du facteur Fromages étaient toutes égales.

Exercice

★★★ Soit la table d'analyse de la variance suivante obtenue à l'aide de Rcmdr :

```
>summary(ANOVAModel)
      Sum Sq   Df  Mean Sq   F value    Pr(>F)
Cantons   926.12    4    231.53    238.72  0.000001
Residuals  446.55   48      0.97
```

À partir de la table d'analyse de la variance, quelle affirmation parmi les suivantes est **vraie** ?

- ☐ L'estimation de la moyenne du facteur Cantons est égale à 231,53.
- ☐ Nous avons 49 observations dans le jeu de données.
- ☐ La variance du facteur Cantons est environ 2 fois plus grande que la variance du facteur Residuals.
- ☐ Au seuil $\alpha = 0,05$, nous pouvons affirmer que les moyennes de la variable réponse à chaque niveaux du facteur Cantons ne sont pas toutes égales.
- ☐ Au seuil $\alpha = 0,05$, nous pouvons affirmer que les variances de la variable réponse à chaque niveaux du facteur Cantons ne sont pas toutes égales.

Exercice

★★★ Le directeur d'une entreprise veut évaluer si le temps (en minute) pour compléter un assemblage est le même pour ses quatre employés. Chacun des employés a effectué cinq assemblages dans les mêmes conditions. Les résultats (temps d'assemblage) obtenus sont enregistrés. On vous donne le modèle suivant d'ANOVA à un facteur

$$Y_{ij} = \mu + \tau_i + E_{ij},$$

où Y_{ij} est le temps d'assemblage de l'employé i pour l'essai j et les E_{ij} sont i.i.d. de loi $N(0, \sigma^2)$, $i=1, 2, 3, 4, j=1, \dots, 5$. Le directeur vous demande de tester si le temps moyen est le même pour les quatre employés. (Supposez que tous les postulats de l'ANOVA à un facteur sont respectés).

Voici la sortie de Rcmdr de quelques statistiques descriptives et la table ANOVA où des valeurs ont été remplacées par des lettres.

```
> with(donnees, numSummary(Temps, groups=employé, statistics=c("mean", "sd")))
      mean      sd
Employé 1    41    4.00000
Employé 2    46    2.12132
Employé 3    40    2.54951
Employé 4    38    2.54951

> summary(AnovaModel.1)
      Df    Sum Sq   Mean Sq   F value    Pr(>F)
Employé  A1    173.8    57.92     A4    0.00337
Residuals A2      A3      8.38
```

- (a) (2 points) Soient μ_1, μ_2, μ_3 et μ_4 les moyennes théoriques du temps d'assemblage de chaque employé et μ la moyenne globale, énoncez l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1) du test.
- (b) (5 points) Calculez les valeurs A1, A2, A3 et A4 de la table ANOVA. Justifiez vos réponses.
- (c) (3 points) Y a-t-il une différence significative, au seuil $\alpha = 5\%$, entre les temps moyens d'assemblage par les quatre employés ? (Choisissez la bonne réponse et complétez la phrase.)

- ☐ Oui car
- ☐ Non car
- ☐ On ne peut pas répondre car

(d) (2 points) Supposons que l'on ait répondu "Oui" en (c) et que l'on vous donne

$PPDS = t_{0.025;16} \sqrt{\frac{2MSE}{5}} = 3.88$. Au seuil $\alpha = 5\%$, choisissez tous les énoncés qui sont vrais. (Deux points seront accordés uniquement si les choix correspondent à la solution exacte.)

- ☐ Le temps moyen d'assemblage de l'employé 1 est significativement différent de celui des autres employés.
- ☐ Le temps moyen d'assemblage de l'employé 2 est significativement différent de celui des autres employés.
- ☐ Le temps moyen d'assemblage de l'employé 3 est significativement différent de celui des autres employés.
- ☐ Le temps moyen d'assemblage de l'employé 4 est significativement différent de celui des autres employés.
- ☐ Le temps moyen d'assemblage de l'employé 2 est significativement différent de celui de l'employé 4 uniquement.
- ☐ Tous les employés ont des temps moyens d'assemblage différents les uns des autres.

Exercice

★★★ Le directeur d'une entreprise veut évaluer si le temps (en minute) pour compléter un assemblage est le même pour ses quatre employés. Chacun des employés a effectué cinq assemblages dans les mêmes conditions. Les résultats (temps d'assemblage) obtenus sont enregistrés. On vous donne le modèle suivant d'ANOVA à un facteur

$$Y_{ij} = \mu + \tau_i + E_{ij},$$

où Y_{ij} est le temps d'assemblage de l'employé i pour l'essai j et les E_{ij} sont i.i.d. de loi $N(0, \sigma^2)$, $i = 1, \dots, 4$, $j = 1, \dots, 5$. Le directeur vous demande de tester si le temps moyen est le même pour les quatre employés. (Supposez que tous les postulats de l'ANOVA à un facteur sont respectés).

Voici la sortie de Rcmdr de quelques statistiques descriptives et la table ANOVA où des valeurs ont été remplacées par des lettres.

```
> with(donnees, numSummary(Temps, groups=employé, statistics=c("mean", "sd")))
```

	mean	sd
Employé 1	41	4.00000
Employé 2	46	2.12132
Employé 3	40	2.54951
Employé 4	38	2.54951

```
> summary(AnovaModel.1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Employé	AA	173.8	57.92	DD	0.00337
Residuals	BB	CC	8.38		

- (a) (2 points) Soient μ_1, μ_2, μ_3 et μ_4 les moyennes théoriques du temps d'assemblage de chaque employé et μ la moyenne globale, énoncéz l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1) du test.
- (b) (5 points) Calculez les valeurs AA, BB, CC et DD de la table ANOVA. Justifiez vos réponses.
AA=BB=CC=DD=
- (c) (3 points) Y a-t-il une différence significative, au seuil $\alpha = 5\%$, entre les temps moyens d'assemblage par les quatre employés ? (Choisissez la bonne réponse et complétez la phrase.)

- ☐ Oui car
- ☐ Non car
- ☐ On ne peut pas répondre car

Exercice

★★★ Soit le tableau suivant dans lequel certaines informations ont été remplacées par AAA et BBB :

i	1	2	3	4	5	Somme
x_i	-3	-2	0	1,5	3,5	0
y_i	-1,1	AAA	1,1	2	2,5	???
$x_i y_i$	3,3	BBB	0	3	8,75	???
x_i^2	9	4	0	2,25	12,25	27,5

On suppose que le modèle de régression linéaire simple peut être utilisé et on veut estimer les paramètres de régression.

(a) (8 points) Sachant que $\hat{\beta}_0 = 1$, que vaut $\hat{\beta}_1$?

$\hat{\beta}_1 =$

(b) (8 points) En supposant que le modèle de régression estimé est $\hat{y} = 1 + 0,511x$ et en utilisant les informations supplémentaires données, donnez un intervalle de confiance à 95% pour la valeur moyenne de Y en $x = 0$.

i	1	2	3	4	5
x_i	-3	-2	0	1,5	3,5
$\hat{y}_i$	-0,533	-0,022	1	1,767	2,789
$y_i - \hat{y}_i$	-0,567	0,522	0,1	0,233	-0,289
$(y_i - \hat{y}_i)^2$	0,321	0,272	0,01	0,054	0,084

(c) (4 points) Sachant que $\sum_{i=1}^5 (y_i - \bar{y})^2 = 7,92$ et en supposant le même modèle qu'en (b), quel pourcentage de la variabilité des Y est expliqué par la régression ?

- ☐ Environ 9,4%
- ☐ Environ 62,6%
- ☐ Environ 87,5%
- ☐ Environ 90,6%
- ☐ Environ 96,9%

Exercice

★★★ Soit la table d'analyse de la variance à un facteur suivante obtenue à l'aide de Rcmdr :

>summary(ANOVAModel)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fromage	2	170.45	85.225	5.704	0.0251
Residuals	9	134.467	14.941		

À partir de la table d'ANOVA, quelle affirmation parmi les suivantes est vraie ? Une seule réponse est bonne.

- ☐ La proportion de la variance totale expliquée par le facteur Fromage est égale à 0,0251.
- ☐ La valeur-p du test d'égalité des moyennes est égale à 5,704%.
- ☐ La variance du facteur Fromage vaut 170,45.
- ☐ La valeur critique du test d'hypothèses au seuil 5% que l'on fait avec cette table d'ANOVA est égale à 4,26.
- ☐ L'hypothèse nulle spécifie que les variances des différents niveaux du facteur Fromage sont toutes égales.
- ☐ Il s'agit d'une ANOVA à blocs aléatoires.
- ☐ Il manque la ligne d'interaction entre Fromage et Residuals pour pouvoir rejeter ou non H_0 .

Module 10 - Régression linéaire simple

Exercice

★★ Afin d'estimer le volume de bois disponible dans la forêt entre Burtigny et Saint-Oyens, nous aimerions modéliser le volume de bois d'un arbre à l'aide de sa circonférence. Nous disposons du volume (Y) et de la circonférence (X) de 31 arbres. Afin d'étudier le modèle $Y_i = \beta_0 + \beta_1 X_i + E_i$, en supposant que les postulats de régression linéaire sont respectés, nous avons les résultats suivants :

```
lm(formula = volume ~ circonfer, data = foret)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-36.943	6.546373	-5.643	0.000001
circonfer	5.066	2.456783	2.062	0.024135

Residual standard error: 8.53 on 30 degrees of freedom

Multiple R-squared: 0.81

- (a) Quelle est l'équation de la droite de régression permettant de prédire le volume de bois d'un arbre de circonférence précise ?
- (b) Calculez la prévision ponctuelle du volume de bois d'un arbre ayant une circonférence de 11.
- (c) Gérard aimerait estimer le volume de bois de l'arbre devant sa maison. Cet arbre a une circonférence de 14. La prévision ponctuelle du volume de bois de cet arbre est environ 33,98. Choisissez l'intervalle de valeurs ayant 95% de chances de contenir le volume de bois de cet arbre de circonférence 14.

- ☐ $33,98 \pm t_{0,05;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33,98 \pm t_{0,025;1} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33,98 \pm t_{0,025;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33,98 \pm t_{0,025;29} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33,98 \pm t_{0,025;1} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$

- (d) Quelle affirmation parmi les suivantes est vraie ?

- ☐ La circonférence d'un arbre et le volume de celui-ci ne sont pas unis par une relation linéaire significative au seuil 5%.
- ☐ La corrélation entre le volume de bois et la circonférence des arbres est 0,81.
- ☐ La circonférence d'un arbre et le volume de celui-ci sont indépendants.
- ☐ La relation linéaire entre le volume de bois d'un arbre et sa circonférence est significative au seuil 5%.

Exercice

★★ Dans la forêt amazonienne, on veut prédire la taille des arbres (Y) en mètres à partir de leur diamètre (X) en centimètres. Pour ce faire, un statisticien dispose de 11 mesures (X_i, Y_i) et à l'aide de Rcmdr, il a ajusté le modèle de régression suivant

$$Y_i = \beta_0 + \beta_1 X_i + E_i, \quad i = 1, \dots, 11.$$

Le statisticien vous montre les figures suivantes :

Figure A

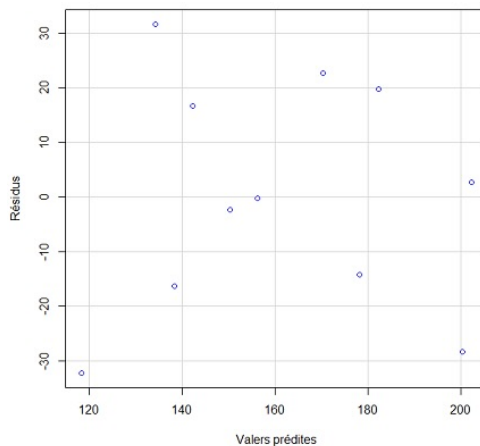


Figure B

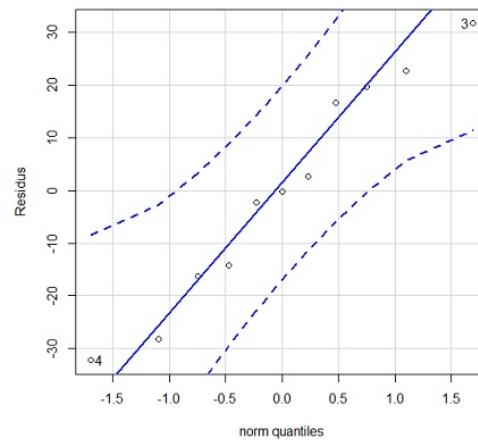
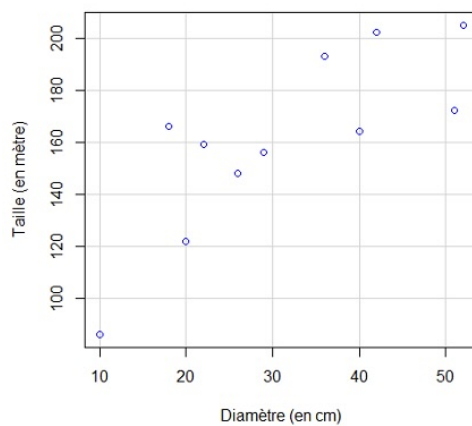


Figure C



(a) (1 point) Choisissez une seule figure permettant de valider le postulat de la linéarité ?

☐ Figure A ☐ Figure B ☐ Figure C

(b) (1 point) Choisissez une seule figure permettant de valider le postulat de l'homoscédasticité ?

☐ Figure A ☐ Figure B ☐ Figure C

(c) (1 point) Choisissez une seule figure permettant de valider le postulat de la normalité ?

☐ Figure A ☐ Figure B ☐ Figure C

Exercice

★★

Soit Y , la quantité d'or (en onces) extraite d'une quantité x de minerai (en tonnes). On vous dit que Y suit une loi normale de moyenne $10 + 10x$ onces et de variance 100 onces².

(a) (6 points) Un camion rapporte 5 tonnes de minerai. Quelle est la probabilité qu'on y extraira plus de 70 onces d'or ?

(b) (3 points) Dix camions rapportent chacun 5 tonnes de minerai. Soit Q_i , la quantité d'or extraite du chargement du camion i . On suppose que $Q_1, \dots, Q_{10}$ sont indépendantes et toutes de même loi que Y .

Quelle est la loi de $Q_{Tot} = \sum_{i=1}^{10} Q_i$?

- ☐ $Q_{Tot} \sim N\left(6; \frac{100}{10}\right)$
- ☐ $Q_{Tot} \sim N\left(60; \frac{100}{10}\right)$
- ☐ $Q_{Tot} \sim N\left(600; \frac{100}{10}\right)$
- ☐ $Q_{Tot} \sim N(6; 1000)$
- ☐ $Q_{Tot} \sim N(60; 1000)$
- ☐ $Q_{Tot} \sim N(600; 1000)$
- ☐ $Q_{Tot} \sim N(6; 10000)$
- ☐ $Q_{Tot} \sim N(60; 10000)$
- ☐ $Q_{Tot} \sim N(600; 10000)$
- ☐ On ne peut pas déterminer cette loi car n est trop petit.

Exercice

★★★ Afin d'estimer le volume de bois disponible dans la forêt entre Burtigny et Saint-Oyens, nous aimerions modéliser le volume de bois d'un arbre à l'aide de sa circonférence. Nous disposons du volume (Y) et de la circonférence (X) de 31 arbres. Afin d'étudier le modèle $Y_i = \beta_0 + \beta_1 X_i + E_i$, en supposant que les postulats de régression linéaire sont respectés, nous avons les résultats suivants :

```
lm(formula = volume ~ circon, data = foret)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-36.943	6.546373	-5.643	0.000001
circon	5.066	2.456783	2.062	0.024135

- (a) Quelle est l'équation de la droite de régression permettant de prédire le volume de bois d'un arbre de circonférence précise ?
- (b) Calculez la prévision ponctuelle du volume de bois d'un arbre ayant une circonférence de 11.
- (c) Gérard aimerait estimer le volume de bois de l'arbre devant sa maison. Cet arbre a une circonférence de 14. La prévision ponctuelle du volume de bois de cet arbre est environ 33.98. Choisissez l'intervalle de valeurs ayant 95% de chances de contenir le volume de bois de cet arbre de circonférence 14.

- ☐ $33.98 \pm t_{0,05;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;1} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;29} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;1} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$

- (d) Quelle affirmation parmi les suivantes est vraie au seuil 1% ?

- ☐ La circonférence d'un arbre et le volume de celui-ci ne sont pas unis par une relation linéaire significative.
- ☐ Nous ne pouvons pas prédire la valeur du volume de bois d'un arbre à partir de sa circonférence.
- ☐ La circonférence d'un arbre et le volume de celui-ci sont indépendants.
- ☐ La relation linéaire entre le volume de bois d'un arbre et sa circonférence est significative.

Exercice

★★★ Un limnologue a étudié certaines propriétés physico-chimiques de l'eau du lac Mistassini en 1973. En particulier, il a étudié la quantité d'oxygène dissous Y (en ppm) à diverses profondeurs X (en pieds), à l'aide de la méthode de Winkler. Les résultats obtenus sont les suivants :

X	10	20	30	50	110
Y	10,4	10,4	10,5	10,9	11,4

Il a aussi la sortie Rcmdr suivante, dans laquelle certaines valeurs ont été remplacées par ????

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.249367	0.064875	157.987	5.59e-07
X	0.010696	????	????	????

Residual standard error: 0.09117 on 3 degrees of freedom

Multiple R-squared: ????, Adjusted R-squared: ????

F-statistic: 86.99 on ??? and ??? DF, p-value: ????

- (a) Quelle est l'équation de la droite de régression estimée?
- (b) (2 points) Donnez une estimation pour la quantité moyenne d'oxygène dissous lorsque la profondeur du lac est de 10 pieds. Justifiez votre réponse.
- (c) La prévision ponctuelle à une profondeur de 25 pieds est de 10,5168. Choisissez l'intervalle de valeurs ayant 95% de chances de contenir la quantité moyenne d'oxygène dissous à cette profondeur.

☐ $10,5168 \pm t_{0,05;3} \sqrt{MS_E \left[\frac{1}{n} + \frac{(25 - \bar{x})^2}{S_{XX}} \right]}$

☐ $10,5168 \pm t_{0,025;1} \sqrt{MS_E \left[\frac{1}{n} + \frac{(25 - \bar{x})^2}{S_{XX}} \right]}$

☐ $10,5168 \pm t_{0,025;3} \sqrt{MS_E \left[\frac{1}{n} + \frac{(25 - \bar{x})^2}{S_{XX}} \right]}$

☐ $10,5168 \pm t_{0,025;3} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(25 - \bar{x})^2}{S_{XX}} \right]}$

☐ $10,5168 \pm t_{0,025;1} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(25 - \bar{x})^2}{S_{XX}} \right]}$

- (d) Les données présentent-elles de l'évidence, au seuil de 5%, qu'il y a une relation linéaire significative entre la quantité d'oxygène dissous et la profondeur du lac? Justifiez votre réponse.

Exercice

★★★ Dans la forêt amazonienne, on veut prédire la taille des arbres (Y) en mètres à partir de leur diamètre (X) en centimètres. Pour ce faire, un statisticien dispose de 11 mesures (X_i, Y_i) et à l'aide de Rcmdr, il a ajusté le modèle de régression suivant

$$Y_i = \beta_0 + \beta_1 X_i + E_i, \quad i = 1, \dots, 11.$$

Il a obtenu les sorties ci-dessous (où certains éléments ont été remplacés par un point d'interrogation). On suppose que tous les postulats du modèle sont respectés.

```
> RegModel.1 <- lm(Y~X, data=Foret)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	98.2787	17.4889	5.619	0.000326
X	1.9998	0.5129	3.899	?

Residual standard error: 22.4 on 9 degrees of freedom

Multiple R-squared: ?, Adjusted R-squared: 0.5868

F-statistic: 15.2 on 1 and 9 DF, p-value: ?

```
> Anova(RegModel.1, type="II")
```

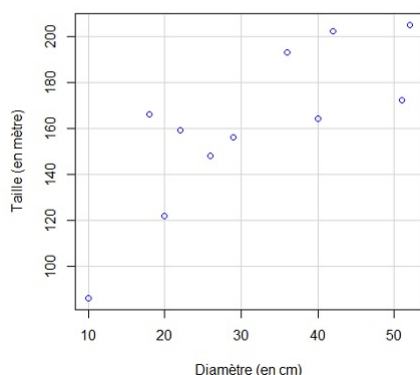
Anova Table (Type II tests)

Response: Y

	Sum Sq	Df	F value	Pr(>F)
X	7625.5	1	15.203	?
Residuals	4514.2	9		

- (a) Avant d'interpréter ces résultats, le statisticien vous montre les figures suivantes :

Figure C



i. (1 point) Choisissez une figure permettant de valider le postulat de la normalité ?

☐ Figure A ☐ Figure B ☐ Figure C

ii. (1 point) Choisissez une figure permettant de valider le postulat de l'homoscédasticité ?

☐ Figure A ☐ Figure B ☐ Figure C

iii. (1 point) Choisissez une figure permettant de valider le postulat de la linéarité ?

☐ Figure A ☐ Figure B ☐ Figure C

(b) (3 points) Calculez un intervalle de confiance à 90% pour l'effet d'une augmentation d'un centimètre de diamètre sur la taille moyenne des arbres. Justifiez votre réponse.

(c) (2 points) Existe-t-il une relation linéaire significative entre le diamètre et la taille des arbres au seuil 10% ? Justifiez votre réponse.

☐ Oui. ☐ Non.

Justification :

(d) (2 points) Calculez le coefficient de corrélation échantillonnal entre le diamètre et la taille des arbres. Justifiez votre réponse.

(e) (2 points) L'une des observations de ce jeu de données est $(x, y) = (25, 135)$. Calculez le résidu associé à cette observation. Justifiez votre réponse.

(f) Déterminez si les indications suivantes sont vraies ou fausses.

Dans le modèle de régression linéaire simple $Y_i = \beta_0 + \beta_1 X_i + E_i, \dots$

i. (1 point) les estimations de β_0 et β_1 sont obtenues en minimisant la valeur de $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$.

☐ Vrai ☐ Faux

ii. (1 point) l'estimation de la variance σ^2 représente la variabilité des observations autour de $\bar{y}$ et se calcule avec $\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 2)$.

☐ Vrai ☐ Faux

iii. (1 point) si tous les points sont sur la droite de régression de Y par rapport à X, le coefficient de corrélation échantillonnal entre Y et X est égal à 1 ou -1.

☐ Vrai ☐ Faux

Exercice

★★★

Afin d'estimer le volume de bois disponible dans la forêt entre Burtigny et Saint-Oyens, on aimerait modéliser le volume de bois d'un arbre à l'aide de sa circonférence puisqu'il est beaucoup plus facile de mesurer la circonférence

d'un arbre que sa hauteur. On dispose du volume (Y) et de la circonférence (X) de 31 arbres. Afin d'étudier le modèle $Y_i = \beta_0 + \beta_1 X_i + E_i$, en supposant que les postulats de régression linéaire sont respectés, vous disposez des résultats suivants :

Residual standard error: 4.252 on 29 degrees of freedom
Multiple R-squared: 0.9353, Adjusted R-squared: 0.9331
F-statistic: 419.4 on 1 and 29 DF, p-value: 0.00014

$$\sum_{i=1}^{31} x_i = 410,70; \quad \sum_{i=1}^{31} x_i^2 = 5\,736,55; \quad \sum_{i=1}^{31} x_i y_i = 13\,887,86; \quad \sum_{i=1}^{31} y_i = 935,30; \quad \sum_{i=1}^{31} y_i^2 = 36\,324,99.$$

- (a) (7 points) Donnez l'équation de la droite de régression permettant de prédire le volume de bois d'un arbre de circonférence précise. **Donnez les détails de votre démarche.**
- (b) (2 points) Donnez le pourcentage de la variabilité totale du volume qui est expliquée par la circonférence.
- (c) (3 points) Calculez la prévision ponctuelle du volume de bois d'un arbre ayant une circonférence de 11. **Donnez les détails de votre démarche.**
- (d) (2 points) Gérard aimerait estimer le volume de bois de l'arbre devant sa maison. Cet arbre a une circonférence de 14. La prévision ponctuelle du volume de bois de cet arbre est environ 33.98. Choisissez l'intervalle de valeurs ayant 95% de chances de contenir le volume de bois de cet arbre de circonférence 14.

- ☐ $33.98 \pm t_{0,05;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;1} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;29} \sqrt{MS_E \left[\frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;29} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$
- ☐ $33.98 \pm t_{0,025;1} \sqrt{MS_E \left[1 + \frac{1}{n} + \frac{(14 - \bar{x})^2}{S_{XX}} \right]}$

- (e) (1 point) Si Gérard avait un arbre ayant une circonférence de 18, l'intervalle de valeurs calculé à la question précédente serait...

- ☐ plus court
- ☐ de même longueur
- ☐ plus long
- ☐ on ne peut pas savoir la différence de longueur

- (f) (2 points) Quelle affirmation parmi les suivantes est vraie au seuil 1% ?

- ☐ La circonférence d'un arbre et le volume de celui-ci ne sont pas unis par une relation linéaire significative.
- ☐ On ne peut pas prédire la valeur du volume de bois d'un arbre à partir de sa circonférence.
- ☐ La circonférence d'un arbre et le volume de celui-ci sont indépendants.
- ☐ La relation linéaire entre le volume de bois d'un arbre et sa circonférence est significative.

Exercice

★★★★ Dans la forêt amazonienne, on veut prédire la taille des arbres (Y) en mètres à partir de leur diamètre (X) en centimètres. Pour ce faire, un statisticien dispose de 31 mesures (X_i, Y_i) et à l'aide de Rcmdr, il a ajusté le modèle de régression suivant

$$Y_i = \beta_0 + \beta_1 X_i + E_i, \quad i = 1, \dots, 31.$$

Il a obtenu les sorties ci-dessous (où certains éléments ont été remplacés par un point d'interrogation). On suppose que tous les postulats du modèle sont respectés.

```
> RegModel.1 <- lm(Y~X, data=Foret)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	102.2655	14.6302	6.99	1.1e-07
x	1.6514	0.4652	3.55	?

 Residual standard error: 22.15 on 29 degrees of freedom
 Multiple R-squared: ?, Adjusted R-squared: 0.2789
 F-statistic: 12.6 on 1 and 29 DF, p-value: ?

> Anova(RegModel.1, type="II")

Anova Table (Type II tests)

Response: y

	Sum Sq	Df	F value	Pr(>F)
x	6181.8	1	12.603	?
Residuals	14225.1	29		

(a) Calculez un intervalle de confiance à 90% pour l'effet d'une augmentation d'un centimètre de diamètre sur la taille moyenne des arbres. Justifiez votre réponse.

(b) Existe-t-il une relation linéaire significative entre le diamètre et la taille des arbres au seuil 10%? Justifiez votre réponse.

☐ Oui. ☐ Non.

Justification :

(c) Calculez le coefficient de corrélation échantillonnal entre le diamètre et la taille des arbres. Justifiez votre réponse.

(d) L'une des observations de ce jeu de données est $(x, y) = (25, 135)$. Calculez le résidu associé à cette observation. Justifiez votre réponse.

(e) Déterminez si les indications suivantes sont vraies ou fausses.

Dans le modèle de régression linéaire simple $Y_i = \beta_0 + \beta_1 X_i + E_i$, ...

i. les estimations de β_0 et β_1 sont obtenues en minimisant la valeur de $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$.

☐ Vrai ☐ Faux

ii. l'estimation de la variance σ^2 représente la variabilité des observations autour de $\bar{y}$ et se calcule avec

$$\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 2).$$

☐ Vrai ☐ Faux

iii. si tous les points sont sur la droite de régression de Y par rapport à X, le coefficient de corrélation échantillonnal entre Y et X est égal à 1 ou -1 (en supposant que tous les points n'ont pas la même valeur de Y).

☐ Vrai ☐ Faux

Module 11 - Régression linéaire multiple

Exercice

★★ Lors d'une expérience, on cherche à construire un modèle pour expliquer et prédire une variable réponse Y par des variables explicatives X_1, X_2, X_3, X_4 et X_5 . Pour cela, on a recueilli les données et on a considéré le modèle de régression linéaire multiple suivant :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + E_i,$$

où les E_i sont i.i.d et de loi $N(0, \sigma^2)$. Vous ajustez ce modèle aux données recueillies et vous obtenez les sorties de Rcmdr suivantes, avec quelques éléments remplacés par des points d'interrogation.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.5405	0.1927	7.992	0.0001
X1	0.1532	0.0752	?????	0.0423
X2	0.0666	0.0489	1.361	0.1743
X3	0.3769	0.0832	4.533	0.0001
X4	0.3681	0.0589	6.246	0.0001
X5	0.1462	0.0839	1.743	0.0821

Residual standard error: 1.019 on 402 degrees of freedom
 Multiple R-squared: 0.5339, Adjusted R-squared: ????
 F-statistic: 92.11 on 5 and 402 DF, p-value: 0.00001

- (a) Quelle est la valeur de la statistique du test pour tester la nullité de β_1 , lorsque les autres variables sont dans le modèle ?
☐ 2.037 ☐ -1.251 ☐ 3.798 ☐ -3.029
- (b) Au seuil de 5%, l'effet de la variable indépendante X_2 sur la variable réponse Y est-il significatif, lorsque les autres variables sont dans le modèle ?
☐ Oui. ☐ Non. ☐ Nous n'avons pas assez d'information pour le savoir.
- (c) Vous voulez maintenant tester, au seuil 1%,

H_0 : aucune des variables X_1, X_2, X_3, X_4, X_5 n'a un effet sur la variable réponse Y

H_1 : au moins une variable indépendante a un effet sur Y .

i. Quelle est la valeur-p associée à ce test ?

ii. Quelle est votre conclusion ?

- ☐ On rejette l'hypothèse nulle au seuil 1%.
- ☐ On ne rejette pas l'hypothèse nulle au seuil 1%.
- ☐ Aucune des 5 variables explicatives n'a un effet sur Y au seuil 1%.
- ☐ Nous ne pouvons pas conclure sur ce test d'hypothèses car nous n'avons pas assez d'information.

Exercice

★ Lors d'une expérience, on cherche à construire un modèle pour expliquer et prédire une variable réponse Y par des variables explicatives X_1, X_2, X_3, X_4 et X_5 . Pour cela, on a recueilli les données et on a considéré le modèle de régression linéaire multiple suivant :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + E_i,$$

où les E_i sont i.i.d et de loi $N(0, \sigma^2)$. Vous ajustez ce modèle aux données recueillies et vous obtenez les sorties de Rcmdr suivantes, avec quelques éléments remplacés par des points d'interrogation.

```
> summary(RegModel.2)
```

Call:

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = Donnees)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.54046	0.19274	7.992	0.000001
X1	0.15321	0.07522	?????	0.0423
X2	0.06662	0.04895	1.361	0.1743
X3	0.37692	0.08315	4.533	0.000001
X4	0.36810	0.05893	6.246	0.000001
X5	0.14618	0.08388	1.743	0.0821

Residual standard error: 1.019 on 402 degrees of freedom

Multiple R-squared: 0.5339, Adjusted R-squared: ????

F-statistic: 92.11 on 5 and 402 DF, p-value: 0.00001

- (a) (2 points) Quelle est la valeur de la statistique du test pour tester la nullité de β_1 , lorsque les autres variables sont dans le modèle ?
☐ 2.037 ☐ -1.251 ☐ 3.798 ☐ -3.029
- (b) (2 points) Au seuil de 5%, l'effet de la variable indépendante X_2 sur la variable réponse Y est-il significatif, lorsque les autres variables sont dans le modèle ?
☐ Oui. ☐ Non. ☐ Nous n'avons pas assez d'information pour le savoir.
- (c) (2 points) Quelle est la valeur de R^2 ajusté du modèle ci-dessus ?
☐ 0.7815 ☐ 0.5281 ☐ 0.2037 ☐ 0.6963
- (d) Vous voulez maintenant tester, au seuil 1%,

H_0 : aucune des variables X_1, X_2, X_3, X_4, X_5 n'a un effet sur la variable réponse Y

H_1 : au moins une variable indépendante a un effet sur Y .

i. (1 point) Quelle est la valeur p (p-value) associée à ce test ?

ii. (1 point) Quelle est votre conclusion ?

- ☐ On rejette l'hypothèse nulle au seuil 1%.
- ☐ On ne rejette pas l'hypothèse nulle au seuil 1%.
- ☐ Aucune des 5 variables explicatives n'a un effet sur Y au seuil 1%.
- ☐ Nous ne pouvons pas conclure sur ce test d'hypothèses car nous n'avons pas assez d'information.

Exercice

★★★ Soit le modèle de régression linéaire multiple suivant :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + E_i,$$

où les E_i sont indépendants et identiquement distribués de loi $\mathcal{N}(0, \sigma^2)$. Nous ajustons ce modèle aux données recueillies et nous obtenons les sorties suivantes, avec deux éléments remplacés par AAA et BBB.

```
> summary(RegModel.2)
```

Call:

```
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = Donnees)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.54046	0.19274	7.992	0.000001
X1	0.15321	0.07522	AAA	0.0423
X2	0.06662	0.04895	1.361	0.1743
X3	0.37692	0.08315	4.533	0.000001
X4	0.36810	0.05893	6.246	0.000001
X5	0.14618	0.08388	1.743	0.0821

Residual standard error: 1.019 on 402 degrees of freedom

Multiple R-squared: 0.5339, Adjusted R-squared: BBB

F-statistic: 92.11 on 5 and 402 DF, p-value: 0.000012

(a) Quelle est la valeur de AAA ?

- ☐ 2.037 ☐ -1.251 ☐ 3.798 ☐ -3.029

(b) Au seuil de 5%, l'effet de la variable indépendante X_2 sur la variable réponse Y est-il significatif, lorsque les autres variables sont dans le modèle ?

- ☐ Oui. ☐ Non. ☐ Nous n'avons pas assez d'information pour le savoir.

(c) Quelle est la valeur de BBB ?

- ☐ 0.7815 ☐ 0.5281 ☐ 0.2037 ☐ 0.6963

(d) Nous voulons maintenant tester, au seuil 1%,

H_0 : aucune des variables X_1, X_2, X_3, X_4, X_5 n'a un effet sur la variable réponse Y

H_1 : au moins une variable indépendante a un effet sur Y.

i. Quelle est la valeur-p associée à ce test ?

- ☐ 0,000001.
- ☐ 0,0423.
- ☐ 0,1743.
- ☐ 0,0821
- ☐ 0,000012.
- ☐ 0,01
- ☐ 0,05
- ☐ 0,10
- ☐ Nous n'avons pas assez d'information pour le savoir.

ii. Quelle est la conclusion ?

- ☐ Nous rejetons l'hypothèse nulle au seuil 1%.
- ☐ Nous ne rejetons pas l'hypothèse nulle au seuil 1%.

- ☐ Aucune des 5 variables explicatives n'a un effet sur Y au seuil 1%.
- ☐ Nous ne pouvons pas conclure sur ce test d'hypothèses car nous n'avons pas assez d'information.