MAT 2779
Examen Partiel (version B)

2 novembre 2016 **Professeur Gilles Lamothe**
Durée : 80 minutes

Numéro d'étudiant :________________

Nom de famille : ________________ **Prénom :** ________________

C'est un examen à livre fermé. Une feuille de formules et le tableau pour la loi normale centrée et réduite sont incluses avec l'examen. Seule une calculatrice non-programmable et non-graphique est permise. Notez votre choix de réponse pour chaque question dans le tableau ci-dessous.

Question	Réponse
1	E
2	A
3	B
4	E
5	C
6	B
7	C
8	D
9	B
10	A
11	A
12	D

N.B. : À la fin de l'examen, remettre seulement cette page. Vous pouvez garder le questionnaire.

****

1. Un test de dépistage est appliqué à 200 patients souffrant d'une certaine maladie. Ce test est également appliqué à 200 personnes choisies au hasard dans la communauté qui ne possèdent pas la maladie. Le tableau suivant est un sommaire des résultats du test :

	Avec la maladie	Sans la maladie	Total
Test positif	166	18	184
Test négatif	34	182	216
Total	200	200	400

   Quelle est la spécificité de ce test?

   A) 0,82 B) 0,10 C) 0,73 D) 0,89 E) 0,91

   $$\text{spécificité} = P(\text{Test} - |\text{sans la maladie}) = \frac{182}{200} = 0{,}91$$

   La réponse est E.

2. Supposons que la police ait mis en place une surveillance radar sur une rue particulière. Ils ont distribué un grand nombre de contraventions sur cette rue. Soit x la différence entre la vitesse mesurée du véhicule et la limite de vitesse. D'un très grand échantillon d'infractions sur cette rue, nous avons calculé des statistiques descriptives pour x :

moyenne	35 km/h
écart type	3,9 km/h
DIQ	5,2 km/h
max	53 km/h

   Supposons que dans cette juridiction, la loi prévoit une amende de 100$ plus 10$ par km/h au dessus de la limite de vitesse. Donc, la valeur de la contravention y (en dollars) est calculée comme $y = 100 + 10\,x$. Donner les statistiques descriptives correspondantes pour la valeur de la contravention y.

   Laquelle des colonnes suivantes A, B, C, D, ou E est correcte?

	A	B	C	D	E
moyenne	450\$	450\$	45\$	450\$	450\$
écart type	39\$	139\$	39\$	39\$	139\$
DIQ	52\$	152\$	52\$	152\$	52\$
max	630\$	630\$	63\$	63\$	630\$

Solution: Nous avons une fonction linéaire $y = a\,x + b$, avec un facteur de redimension $a = 10$ et un terme de décalage $b = 100$. Nous devons prendre à la fois la redimension et le décalage, quand on calcule la moyenne et le maximum :

$$\text{moyenne} = 100 + 10\,(35) = 450\$ \quad \text{et} \quad \text{max} = 100 + 10\,(53) = 630\$.$$

Avec une redimension d'un facteur de 10, la distance entre les valeurs est 10 fois plus grande. Mais nous ajoutons 100 à toutes les valeurs, ce décalage n'a pas d'effet sur la distance entre les valeurs. Ainsi, les mesures de dispersion ne sont pas affectés par le décalage. Cela signifie que lorsque nous considérons les mesures de la dispersion tels que l'écart type et la DIQ, nous avons seulement besoin de considérer la redimension par un facteur de 10. Le décalage n'a pas d'effet sur les mesures de dispersion. Donc, nous obtenons

$$\text{écart type} = 10\,(3,9) = 39\$ \quad \text{et} \quad \text{DIQ} = 10\,(5,2) = 52\$.$$

La réponse est A.

3. Un certain type de cancer se trouve chez les femmes âgées de plus de 60 avec une probabilité de 6,5 %. Il existe un test sanguin pour la détection de la maladie, mais le test n'est pas parfait. Ce test a un taux de faux négatif de 10 % et un taux de faux positifs de 4,5 %. Une femme âgée de plus de 60 a pris le test et a reçu un résultat de test favorable, c'est-à-dire un résultat négatif. Quelle est la probabilité qu'elle a ce type de cancer?

A) 0,1000 B) 0,0072 C) 0,0025 D) 0,0450 E) 0,0145

Solution: Soit M l'événement d'avoir la maladie et soit $-$ et $+$ les événements d'avoir un résultat négatif et positif au test, respectivement.

On a $P(M) = 0{,}065$, le taux des faux négatif= $P(-|D) = 0{,}1$ et le taux des faux positifs est $= P(+|D') = 0{,}045$. On calcul en premier la probabilité d'avoir un résultat négatif :

$$P(-) = P(-|M)P(M)+P(-|M')P(M') = (0.1)(0.065)+(1-0.045)(1-0.065) = 0{,}8994.$$

On veut la probabilité suivante :

$$P(M|-) = \frac{P(M \cap -)}{P(-)} = \frac{P(-|M)P(M)}{P(-)} = \frac{(0{,}1)(0{,}065)}{0{,}8994} = 0{,}0072.$$

La réponse est B.

4. Voici la taille (en cm) pour 9 étudiantes sélectionnées au hasard.

160	159	164	169	176	163	161	160	178

Lequel des énoncés suivants est correct?

A) 159 est la seule valeur aberrante.
B) 178 est la seule valeur aberrante.
C) 178 et 176 sont des valeurs aberrantes.
D) 159 et 178 sont des valeurs aberrantes.
E) Il n'y a pas des valeurs aberrantes.

solution: Nous allons arranger les valeurs en ordre croissant :

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
159	160	160	161	163	164	169	176	178

On calcul le rang pour q_1 et q_3, respectivement :

$$(n+1)25\% = 2{,}5 \quad \text{et} \quad (n+1)25\% = 7{,}5.$$

Alors,

$$q_1 = 0{,}5\,y_2 + 0{,}5\,y_3 = 0{,}5\,(160) + 0{,}5\,(160) = 160$$

et

$$q_3 = 0{,}5\,y_7 + 0{,}5\,y_8 = 0{,}5\,(169) + 0{,}5\,(176) = 172{,}5.$$

Alors, la distance interquartile est DIQ $= q_3 - q_1 = 172{,}5 - 160 = 12{,}5$. Alors,

$$\begin{aligned} \text{clôture sup.} &= q_3 + 1{,}5\text{DIQ} = 172{,}5 + 1{,}5\,(12{,}5) = 191{,}25 \\ \text{clôture inf.} &= q_1 - 1{,}5\text{DIQ} = 160 - 1{,}5\,(12{,}5) = 141{,}25 \end{aligned}$$

Puisque toutes les valeurs sont entre les clôtures, alors il n'y a pas de valeurs aberrante. La réponse est E.

5. Le test d'acide vanillylmandélique est un test d'urine qui a été développé pour identifier la présence du neuroblastome, une maladie rare et grave. Ce test donne un résultat **positif** dans 70 % des cas avec une présence du neuroblastome. 8 enfants atteints du neuroblastome sont administrés ce test. Nous voulons calculer la probabilité que le test donne un résultat **négatif** pour au moins 2 enfants parmi ces 8 enfants. Laquelle des commandes suivantes avec `R` donne cette probabilité?

   A) pbinom(2,8,0.3)
   B) 1-pbinom(1,8,0.7)
   C) 1-pbinom(1,8,0.3)
   D) pbinom(2,8,0.7)
   E) pbinom(1,8,0.7)

   Solution: Soit X le nombre de résultats négatifs parmi les $n = 8$ résultats. X suit une loi binomiale avec $n = 8$ et $p = 0{,}3$. On veut $P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1) = 1 - pbinom(1, 8, 0.3)$. Alors, C est la réponse.

6. Un chercheur rapporte que les souris vont vivre une moyenne de 40 mois lorsque leurs régimes sont fortement limités et enrichis avec des vitamines et des protéines. En supposant que les durées de vie de ces souris sont normalement distribuées avec un écart type de 6,3 mois, trouver la probabilité qu'une souris va vivre entre 37 et 49 mois.

   A) 0,3888 B) 0,6080 C) 0,7205 D) 0,5845 E) 0,3920

   Solution: Soit X la durée de vie d'une sourie. X suit une loi normale de moyenne $\mu = 40$ et d'écart type $\sigma = 6{,}3$. On veut

$$\begin{aligned} P(37 < X < 49) &= P\left(\frac{37-40}{6{,}3} < \frac{X-\mu}{\sigma} < \frac{49-40}{6{,}3}\right) \\ &= \Phi(1{,}43) - \Phi(-0{,}48) \\ &= 0{,}9236 - 0{,}3156 = 0{,}6080 \end{aligned}$$

   La réponse est B.

7. Les personnes à risque pour le cancer du côlon sont conseillé d'avoir une coloscopie proche des 50 ans. Considérons 250 patients à risque pour le cancer du côlon. Pour chaque patient, on détermine si le patient a eu une coloscopie et si le patient a le cancer du côlon. Voici un sommaire des données.

	Cancer du côlon : oui	Cancer du côlon : oui	Total
Coloscopie : oui	x	50-x	50
Coloscopie : non	5-x	195+x	200
Total	5	245	250

   Supposons que nous sélectionnons un des 250 patients au hasard. Soit A l'événement que le patient sélectionné ait le cancer du côlon et B l'événement que le patient sélectionné ait le cancer du côlon. Dans le tableau, x désigne le nombre de patients atteints d'un cancer du côlon qui a eu une coloscopie. Trouver la valeur de x tel que A et B sont des événements indépendants. Pour cette valeur de x, peut-on dire que A' et B' sont également des événements indépendants?

A) $x = 5$. Non, for $x = 5$, A' et B' ne sont pas indépendants.
B) $x = 5$. Oui, pour $x = 5$, A' et B' sont aussi indépendants.
C) $x = 1$. Oui, pour $x = 1$, A' et B' sont aussi indépendants.
D) $x = 1$. Non, pour $x = 1$, A' et B' ne sont pas indépendants.
E) Il n'y a pas de valeur de x afin que A et B soient indépendants.

Solution: Par définition, A et B sont indépendants si $P(A \cap B) = P(A)P(B)$. Mais, $P(A \cap B) = x/250$ et $P(A)P(B) = (5/250)(50/250)$. Alors, l'indépendance de A et B est équivalent à $x = 250\,(5/250)(50/250) = 1$. Vérifions que A' et B' sont inépendants, si $x = 1$. On a $P(A' \cap B') = 196/250 = 0{,}784$ et $P(A')P(B') = (245/250)(200/250) = 0{,}784$. Puisque $P(A' \cap B') = P(A')P(B')$ alors A' et B' sont indépendants. La réponse est C.

8. Dans une ville, 6 % de la population fume des cigares, 14 % de la population fume la cigarette, et 84 % de la population ne fume pas la cigarette et ne fume pas des cigares. Soit A l'événement qu'un résident choisi au hasard fume des cigares et soit B l'événement qu'un résident choisi au hasard fume la cigarette. Lequel des énoncés suivants est **incorrect**?

A) $P(A \cup B') = 0{,}90$
B) $P(A \cap B') < P(A' \cap B)$
C) $P(A' \cup B') = 0{,}96$
D) $P(A \cap B) = P(A)P(B)$
E) $P(A \cup B) = 0{,}16$

Solution: On sait que $P(A) = 0{,}06$, $P(B) = 0{,}14$, et $P(A' \cap B') = 0{,}84$. Alors, $P(A \cup B) = 1 - P(A' \cap B') = 1 - 0{,}84 = 0{,}16$, alors le choix (E) est correct. Par la règle de l'addition, on a

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0{,}06 + 0{,}14 - 0{,}16 = 0{,}04.$$

Ensuite, on calcul

$$
\begin{aligned}
P(A' \cup B') &= 1 - P(A \cap B) = 1 - 0{,}04 = 0{,}96, \\
P(A \cap B') &= P(A) - P(A \cap B) = 0{,}06 - 0{,}04 = 0{,}02, \\
P(A' \cap B) &= P(B) - P(A \cap B) = 0{,}14 - 0{,}04 = 0{,}10, \\
P(A \cup B') &= 1 - P(A' \cap B) = 1 - 0{,}10 = 0{,}90.
\end{aligned}
$$

Alors, les choix (A), (B) et (C) sont corrects. Notons que

$$P(A \cap B) = 0{,}04 \neq P(A)P(B) = (0{,}06)(0{,}14) = 0{,}0084,$$

alors D est incorrect. La réponse est D.

9. Le bureau de Santé publique d'Ottawa cueille des échantillons d'eau à tous les jours pendant l'été. Il y a un avis qui interdit la nage, si la concentration de bactéries est plus que 200 E. coli par 100 ml. Supposons que la concentration de bactéries est une variable aléatoire normale de moyenne $\mu = 120$ et d'écart type $\sigma = 50$. Quelle est la probabilité qu'il y ait au moins deux jours avec un avis qui interdit la nage à la plage Mooney's Bay au cours des 5 prochaines années durant la fête du Canada?

   A) 0,246 B) 0,027 C) 0,378 D) 0,055 E) 0,163

   Solution: Soit X la concentration de bactéries E. coli dans l'eau (en nombre d'E. coli par 100 ml). On suppose que X suit une loi normale de moyenne $\mu = 120$ et d'écart type $\sigma = 50$. On calcul

$$
\begin{aligned}
P(X > 200) &= P\left(\frac{X - 120}{50} > \frac{200 - 120}{50}\right) \\
&= P(Z > 1{,}6) = 1 - 0{,}9452 = 0{,}0548.
\end{aligned}
$$

   Soit Y le nombre de jours avec un avis qui interdit la nage parmi $n = 5$ jours. Y suit une loi binomiale avec $n = 5$ et $p = P(X > 200) = 0{,}0548$. On veut

$$
\begin{aligned}
P(Y \geq 2) &= 1 - [P(Y = 0) + P(Y = 1)] \\
&= 1 - (0{,}0548)^0(1 - 0{,}0548)^5 - 5(0{,}0548)^1(1 - 0{,}0548)^4 \\
&= 0{,}027
\end{aligned}
$$

La réponse est B.

10. Il y a deux rats mâles et deux rats femelles dans une cage. Nous choisissons au hasard deux rats (sans remise) de cette cage. Soit X soit le nombre de mâles parmi les deux rats sélectionnés. Soit F_X la fonction de répartition de X et f_X la loi de probabilité de X. Lequel des énoncés suivants est **incorrect**? (Un seul énoncé est incorrect.)

    A) $F_X(3) = 0$
    B) $F_X(0) > 0$
    C) $f_X(x) = F_X(x)$ pour tout $x < 0$ D) $f_X(3) = 0$
    E) $f_X(0) > 0$

    Solution: X est une variable discrète avec l'image $\{0, 1, 2\}$, où $f_X(0) = P(X = 0) > 0$, $f_X(1) = P(X = 1) > 0$ et $f_X(2) = P(X = 2) > 0$. On a

    $$F_X(0) = P(X \leq 0) = P(X = 0) = f_X(0) > 0,$$
    $$F_X(3) = P(X \leq 3) = 1 \quad \text{et} \quad f_X(3) = P(X = 3) = 0.$$

    Puisque $P(X \geq 0) = 1$, alors $F(x) = P(X \leq x) = 0$ et $f(x) = P(X = x) = 0$ pour tout $x < 0$. La réponse est A.

11. Chez les humains, la couleur des cheveux est déterminée par un gène dont l'allèle pour les cheveux foncés (D) est dominant sur l'allèle pour les cheveux roux (d), tandis que la couleur des yeux est déterminée par un gène dont l'allèle pour les yeux bruns (B) est dominant sur l'allèle pour les yeux bleus (b). Une femme a les yeux bleus et elle a les cheveux foncés, étant hétérozygote pour le gène de la couleur des cheveux. Un homme a les cheveux roux et il a les yeux bruns étant hétérozygote pour le gène de la couleur des yeux. La femme a le type sanguin AB et l'homme a le type sanguin O. Calculer la probabilité que l'enfant du croisement de cet homme et cette femme ait des yeux bleus, des cheveux roux et le type sanguin A.

(Supposer l'assortiment indépendant de ces gènes.)

N.B. : Le type sanguin est déterminé par les allèles I^A, I^B, i d'un gène noté I : I^A et I^B sont dominants sur i, mais I^A et I^B sont codominant uns avec les autres. Un personne de type A peut avoir le génotype I^AI^A ou I^Ai, une personne de type B peut avoir le génotype I^BI^B ou I^Bi, une personne de type O a le génotype ii, et une personne de type AB a le génotype I^AI^B.

A) 1/8 B) 1 C) 1/4 D) 1/2 E) 0

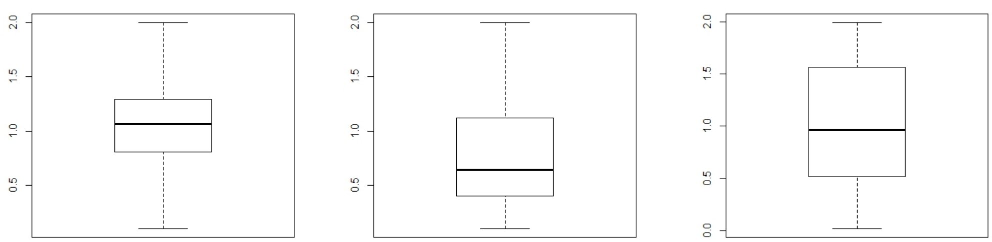
solution: Le génotype pour le type sanguin est $I^A\,I^B$ pour la femme et ii pour l'homme. Alors, le croisement peut donner les génotypes $I^A\,i$ et $I^B\,i$. Les phénotypes correspondants sont type sanguin A et type sanguin B (chacun avec une probabilité de 1/2).

Le génotype de l'homme pour les cheveux est dd et c'est Dd pour la femme. Le croisement peut donner les génotypes Dd et dd. Les phénotypes correspondants sont cheveux foncés et cheveux roux. La probabilité pour les cheveux foncés est 1/2 et c'est 1/2 pour les cheveux roux.

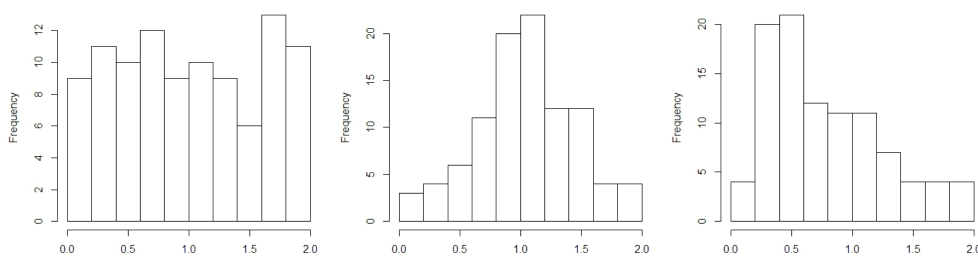
Le génotype de l'homme pour les yeux est Bb et c'est bb pour la femme. Le croisement peut donner les génotypes Bb et bb. Les phénotypes correspondants sont yeux bruns et yeux bleus (chacun avec une probabilité de 1/2).

Alors, (par l'assortiment indépendant de ces gènes), la probabilité que leur enfant ait les yeux bleus, les cheveux roux et le type sanguin A est $(1/2)(1/2)(1/2) = 1/8$. La réponse est A.

12. Les espèces les plus communes d'abeilles sudoripares sont vertes, rouges et jaunes. Pour chaque couleur, on sélectionne un échantillon de 100 abeilles de cette couleur et on mesure leur longueur (en pouces). Nous créons ensuite 3 ensembles de données de 100 observations chacune, appelé *vert*, *rouge* et *jaune*. Voici les diagrammes à boîte et moustaches et histogrammes pour ces ensembles de données. Les étiquettes des variables sont manquantes pour les histogrammes, mais sont incluses pour les diagrammes à boîte et moustaches. Notre tâche est d'identifier les étiquettes manquantes. Lequel des énoncés suivants est correct?



(a) vert (b) rouge (c) jaune



(d) (e) (f)

A) hist. (f) est pour vert, (d) est pour rouge, et (e) est pour jaune.
B) hist. (d) est pour vert, (e) est pour rouge, et (f) est pour jaune.
C) hist. (f) est pour vert, (e) est pour rouge, et (d) est pour jaune.
D) hist. (e) est pour vert, (f) est pour rouge, et (d) est pour jaune.
E) hist. (d) est pour vert, (f) est pour rouge, et (e) est pour jaune.

Solution: Tous les échantillons ont la même étendue, alors les moustaches ne peuvent pas être utilisés pour la correspondance. Notez que l'un de l'histogramme a une asymétrie négative, tandis que les deux autres histogrammes sont approximativement symétrique. On identifie une médiane en bas du centre pour l'asymétrie négative. Par conséquent, rouge correspond à (f). Pour les distributions symétriques, nous pouvons comparer la dispersion pour la correspondance. Notez que les valeurs dans l'histogramme (e) sont moins dispersées (c.-à-d. plus concentrées au centre) en comparaison aux valeurs de l'histogramme (d). Lorsque l'on compare les diagrammes (a) et (b), nous devrions remarquer que la boîte dans le diagramme (a) est plus petite (c.-à-d. des valeurs moins dispersées). Par conséquent, vert correspond à (e) et jaune correspond à (d).

La réponse est D.