Université d'Ottawa · University of Ottawa
Faculté des sciences
Mathématiques et de statistique
Faculty of Science
Mathematics and Statistics

Examen final pour MAT 2779 **Date : 15 décembre 2014**

Durée : 3 heures **Professeur : Gilles Lamothe**

Nom : ____________________ Numéro d'étudiant :____________________
du siège :________

Une calculatrice non-graphique & non-programmable est permise. C'est un examen à livre fermé.
Il y a 25 questions à choix multiples. À la fin de l'examen, remettre seulement cette page. Vous pouvez garder le questionnaire.
Ecrire vos réponses aux questions dans le tableau suivant.

Question	Réponse	Question	Réponse
1	D	14	A
2	B	15	A
3	D	16	A ou B
4	B	17	D
5	C	18	B
6	A	19	C
7	B	20	B
8	D	21	E
9	C	22	E
10	A	23	B
11	C	24	D
12	E	25	C
13	E		

585, av. King-Edward
Ottawa (Ontario) K1N 6N5 Canada
585 King Edward Avenue
Ottawa, Ontario K1N 6N5 Canada
(613) 562-5864 • Téléc./Fax (613) 562-5776
Courriel/Email: uomaths@science.uottawa.ca

1. La durée moyenne de la gestation humaine est d'environ 40,5 semaines. On pense que le diabète de la mère peut influencer la durée de la gestation. Dans une étude constituée de 20 femmes diabétiques enceintes, il a été observé que la durée moyenne de gestation était de 38,8 semaines et que l'écart type de la durée était 5 semaines. Est-ce que les preuves que la durée moyenne de la gestation chez les femmes diabétiques n'est pas 40,5 semaines sont significatives ? Formuler un test d'hypothèses approprié à un niveau de signification de $\alpha = 0,05$. Déterminer la valeur P et donner vos conclusions.

   A) La valeur P est entre 0,01 et 0,025. Il y a des preuves significatives que la durée moyenne de la gestation chez les femmes diabétiques est différente de 40,5 semaines.
   B) La valeur P est entre 0,025 et 0,05. Il y a des preuves significatives que la durée moyenne de la gestation chez les femmes diabétiques est différente de 40,5 semaines.
   C) La valeur P est entre 0,05 et 0,10. Les preuves que la durée moyenne de la gestation chez les femmes diabétiques est différente de 40,5 semaines ne sont pas significatives.
   D) La valeur P est entre 0,10 et 0,20. Les preuves que la durée moyenne de la gestation chez les femmes diabétiques est différente de 40,5 semaines ne sont pas significatives.
   E) La valeur P est entre 0,20 et 0,40. Les preuves que la durée moyenne de la gestation chez les femmes diabétiques est différente de 40,5 semaines ne sont pas significatives.

   Solution : Soit μ la durée moyenne de la gestation chez les femmes diabétiques. On veut tester $H_0 : \mu = 40,5$ contre $H_1 : \mu \neq 40,5$. On sait que $n = 20$, $\bar{x} = 38,5$ et $s = 5$. La valeur observée de la statistique est
   $$t_0 = \frac{\bar{x} - 40,5}{s/\sqrt{n}} = \frac{38,8 - 40,5}{5/\sqrt{20}} = -1,52.$$
   Puisque nous avons une alternative bilatérale, la valeur P est $P = 2\,P(T > 1,52)$, où T suit une loi $t(19)$. Du tableau 17.4, on observe que 1,52 est une valeur entre 1,328 et 1,729, et leurs ordres correspondantes sont 0,10 et 0,05, respectivement. Alors, $P(T > 1,79)$ est entre 0,05 et 0,10. Donc, $0,1 < P < 0,20$. Puisque la valeur P est supérieure au niveau de signification, alors nos preuves contre H_0 ne sont pas significatives. La réponse est D.

2. Le vaccin Bacille Calmette-Guérin (BCG) pour la tuberculose est obliga-

toire pour les enfants d'âge scolaire dans de nombreux pays européens. Au Canada, avant la vaccination par le BCG, le patient est testé pour la tuberculose en utilisant un test cutané à la tuberculine, appelé le test de Mantoux. Les gens qui ont été vaccins par le BCG aura souvent un résultat positif au test de Mantoux, même s'ils n'ont pas la tuberculose. Par conséquent, le test de Mantoux n'est pas un outil très efficace pour détecter la tuberculose. Dans une étude récente, 12% des sujets avaient un résultat positif au test de Mantoux. Parmi ceux qui ont un résultat positif au test, seulement 10% avaient la tuberculose. D'autre part, 1% des patients avec un résultat négatif au test avaient la tuberculose. Dans cette étude, quel pourcentage des patients avaient la tuberculose ?

A) 1,10% B) 2,08% C) 0,88% D) 1,20% E) 13,03%

Solution : Soit TB l'événement qu'une personne de ce groupe ait la tuberculose. On veut

$$\begin{aligned} P(\text{TB}) &= P(\text{TB}|\text{Test+})P(\text{Test +}) + P(\text{TB}|\text{Test}-)P(\text{Test}-) \\ &= (0,10)(0,12) + (0,01)(0,88) = 0,0208. \end{aligned}$$

La réponse est B.

3. Une équipe de chercheurs ont étudié l'effet de la thérapie de la réminiscence pour les femmes âgées souffrant de dépression. Ils ont mesuré la dépression (sur une échelle de 1 à 10) chez 20 femmes de 60 ou plus résidant pendant 3 mois ou plus dans un établissement de soins de longue durée. Une plus grande cote de dépression est interprété comme une dépression plus sévère. La dépression fut mesuré aussi après l'intervention. Dans `R`, nous avons attribué les cotes pré-intervention et post-intervention aux variables `pre` et `post`, respectivement. Nous avons aussi calculé la différence entre la cote de dépression post-intervention et la cote de dépression pré-intervention, pour chaque sujet.

```
> d=post-pre
> summary(d)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -5.00   -3.00   -0.50   -1.35    0.00    1.00
> sd(d)
[1] 2.084403
```

Supposons que la différence entre les mesures post-intervention et pré-intervention est normalement distribuée. Utiliser les statistiques descriptives ci-haut pour pour calculer un intervalle de confiance à 95% pour la moyenne de la différence des cotes de dépression post-intervention et pré-intervention. Basé sur cet intervalle de confiance, est-ce que nous sommes confiant que la dépression est moins sévère en moyenne après l'intervention ?

A) I.C.=[0,37 ; 2,33]. Nous sommes confiant que la dépression est plus sévère après l'intervention en moyenne.
B) I.C.=[-2,33 ; 0,37]. Nous ne sommes pas confiant que la dépression est moins sévère après l'intervention en moyenne.
C) I.C.=[-2,53 ; -0,57]. Nous sommes confiant que la dépression est moins sévère après l'intervention en moyenne.
D) I.C.=[-2,33 ; -0,37]. Nous sommes confiant que la dépression est moins sévère après l'intervention en moyenne.
E) I.C.=[-2,33 ; -0,37]. Nous ne sommes pas confiant que la dépression est moins sévère après l'intervention en moyenne.

Solution : Nous avons des mesures appariées. On a défini la différence

$$D = \text{ dépression post-intervention } - \text{ dépression pré-intervention } .$$

On nous donne

$$n = 20, \quad \overline{d} = -1,35 \quad \text{et } s_d = 2,0844.$$

Un intervalle de confiance pour μ_D est

$$\overline{d} \pm t_{0,025;19} \frac{s_d}{\sqrt{n}} = [-2,33; -0,37],$$

où $t_{0,025;19} = 2,093$. Puisque toutes les valeurs dans l'intervalle sont négatives, alors nous sommes 95% confiant qu'en moyenne la dépression post-intervention est plus petite que la dépression pré-intervention. La réponse est D.

4. Une étude a été menée pour estimer la sensibilité et la spécificité d'une nouvelle procédure pour identifier une maladie rénale chez les patients souffrant d'hypertension. Parmi 54 sujets hypertendus atteints de la maladie, la nouvelle procédure a identifié la maladie chez 45 sujets. Parmi 83 sujets hypertendus sans la maladie, la nouvelle procédure a identifié la maladie chez 24 sujets. Supposons que nous avons un patient hypertendu d'une population avec une prévalence de 8% pour cette maladie rénale. La nouvelle procédure identifie la maladie rénale chez ce patient. Quelle

est la probabilité que le patient ait réellement cette maladie rénale ? Supposons que la sensibilité et la spécificité du test reste le même que dans l'étude mentionnée ci-haut.

A) 0,8820 B) 0,2004 C) 0,3737 D) 0,5732 E) 0,0545

Solution : Soit $T+$ l'événement que la nouvelle procédure identifie la maladie et M l'événement que la patient ait réellement cette maladie. On a

$$P(M) = 0,08; \; P(T+|M) = 45/54; \; P(T+|M') = 24/83.$$

On veut

$$\begin{aligned} P(M|T+) &= \frac{P(M \cap T+)}{P(T+)} = \frac{P(T+|M)P(M)}{P(T+)} \\ &= \frac{(45/54)(0,08)}{0,3327} = 0,2004 \end{aligned}$$

où

$$\begin{aligned} P(T+) &= P(T+|M)P(M) + P(T+|M')P(M') \\ &= (45/54)(0,08) + (24/83)(1-0,08) \\ &= 0,3327. \end{aligned}$$

La réponse est B.

5. La pression intra-oculaire est la pression du fluide à l'intérieur de l'oeil. Le glaucome est une maladie oculaire qui se manifeste par une pression intraoculaire élevée. La répartition de la pression intraoculaire dans la population générale suit une loi normale de moyenne 16 mm Hg et d'écart type de 3 mm Hg. Une pression intra-oculaire normale est considérée comme étant comprise entre 12 mm Hg et 20 mm Hg (inclusivement). Laquelle des commandes suivantes de `R` donne la probabilité qu'une personne choisie au hasard ait une pression intra-oculaire normale ? (Une seule réponse est correcte.)

   A) `qnorm(20,16,3)-qnorm(12,16,3)`
   B) `pnorm(20,3,16)-pnorm(12,3,16)`
   C) `pnorm(20,16,3)-pnorm(12,16,3)`
   D) `pnorm(20,16,3)-pnorm(11,16,3)`
   E) `pnorm(20,16,9)-pnorm(12,16,9)`

Solution : On veut calculer $P(12 \leq X \leq 20)$, où X suit une loi normale de moyenne $\mu = 16$ et d'écart type $\sigma = 3$. Cette probabilité est :

$$P(12 \leq X \leq 20) = P(X \leq 20) - P(X < 12) = P(X \leq 20) - P(X \leq 12)$$
$$= \text{pnorm}(20, 16, 3) - \text{pnorm}(12, 16, 3)$$

Puisque X est une variable aléatoire continue, on a $P(X < 12) = P(X \leq 12)$. La réponse est C.

6. Les autochtones du Canada ont un risque plus élevé de développer de nombreuses maladies chroniques par rapport au reste de la population. Dans une communauté autochtone particulière, 16% de la population ont la tuberculose, 20 % sont atteints de diabète et 8% ont les deux maladies. Quelle est la probabilité qu'une personne choisie au hasard dans cette communauté n'ait aucune de ces maladies ?

   A) 0,72 B) 0,28 C) 0,64 D) 0,85 E) 0,90

   Solution : Soit A l'événement que la personne ait la tuberculose et B l'événement que la personne soit atteint de diabète. On a $P(A) = 0,16$; $P(B) = 0,20$ et $P(A \cap B) = 0,08$. Par la règle de l'addition,

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,16 + 0,20 - 0,08 = 0,28.$$

   La probabilité qu'elle n'ait aucune de ces maladies est

   $$P(A' \cap B') = 1 - P(A \cup B) = 1 - 0,28 = 0,72$$

   La réponse est A.

7. La maladie à virus Ebola, anciennement connu sous le nom de fièvre hémorragique à virus Ebola, est une maladie grave, souvent mortelle chez les humains. On pense que les chauves-souris de la famille des Pteropodidae sont des hôtes naturels du virus Ebola. Le virus est introduit dans la population humaine par contact étroit avec les fluides corporels d'animaux infectés. La période d'incubation (l'intervalle de temps de l'infection par le virus avant l'apparition des symptômes) est compris entre 2 à 21 jours. Les données ci-après donne la durée d'incubation (en jours) de 16 patients infectés par le virus Ebola :

   4 5 6 6 7 8 9 9 11 12 13 15 15 17 20 21

   Calculer la médiane ($\tilde{x}$), le premier quartile (q_1) et le troisième quartile (q_3) pour cet échantillon. Donner les valeurs des valeurs aberrantes (s'il y en a).

A) $\tilde{x} = 10$; $q_1 = 5$; $q_3 = 15$; la valeur 21 est aberrante
B) $\tilde{x} = 10$; $q_1 = 6.25$; $q_3 = 15$; il n'y a pas de valeurs aberrantes
C) $\tilde{x} = 11$; $q_1 = 6$; $q_3 = 17$; la valeur 4 est aberrante
D) $\tilde{x} = 11$; $q_1 = 6,25$; $q_3 = 15$; les valeurs 4 et 5 sont aberrantes
E) $\tilde{x} = 11$; $q_1 = 5,25$; $q_3 = 15,5$; il n'y a pas de valeurs aberrantes

Solution : Ici $n = 16$. Voici les rangs de q_1, $\tilde{x}$ et q_3 :

$$(n+1)25\% = 4,25; \quad (n+1)50\% = 8,5; \quad (n+1)75\% = 12,75.$$

Alors,

$$\tilde{x} = (1-0,5)\,y_8 + 0,5\,y_9 = 0,5(9) + 0,5(11) = 10,$$
$$q_1 = 0,75\,y_4 + 0,25\,y_4 = 0,75(6) + 0,25(7) = 6,25$$

et

$$q_1 = 0,25\,y_{12} + 0,75\,y_{13} = 0,75(15) + 0,25(15) = 15$$

Pour trouver les valeur abberantes, on calcul les clôtures. La distance inter-quartile est $\text{DIQ} = q_3 - q_1 = 15 - 6.25 = 8,75$. Hence

clôture inf. $= q_1 - (1,5)\text{IQR} = 6,25 - (1,5)(8,75) = 6,25 - 13,125 = -6,875$

clôture sup. $= q_3 + (1,5)\text{IQR} = 15 + (1,5)(8,75) = 15 + 13,125 = 28,125$

Il y a aucune valeur à l'extérieure des clôtures. Alors, il n'y a pas de valeurs abberantes. La réponse est B.

8. En biochimie et la pharmacologie, un récepteur est une molécule de protéine habituellement trouvée dans la surface de la membrane plasmique d'une cellule qui reçoit des signaux chimiques à partir de l'extérieur de la cellule. D'un échantillon de 9 cellules, on a observé une moyenne de 1203 fmoles récepteurs par milligramme de protéine membranaire, avec une erreur type estimée de la moyenne de 64 fmoles. (Une fmole est égale à 10^{-15} mole). Basé sur ces données, donner un intervalle de confiance à 95% pour le montant moyen (en fmoles) de récepteurs par milligramme trouvés dans la protéine de la membrane de ces cellules. On suppose que la quantité de récepteurs par milligramme de protéine membranaire est normalement distribuée.

A) $[1077,3; 1329,7]$ B) $[1153,8; 1252,2]$ C) $[0; 1322,8]$
D) $[1055,4; 1350,6]$ E) $[1098,1; 1308,9]$

Solution : Soit μ le montant moyen (en fmoles) de récepteurs par milligramme trouvés dans la protéine de la membrane de ces cellules. Un intervalle de confiance pour μ est

$$\bar{x} \pm t_{0,025;8} \frac{s}{\sqrt{n}} = [1055,416; 1350,584],$$

où $t_{0,025;8} = 2,306$, $\overline{x} = 1203$ et l'erreur type estimée de la moyenne est $s/\sqrt{n} = 64$.

9. La sortie de `R` suivante est un sommaire du poids (en livres) pour 50 bébés qui sont nés prématurés dans un grand hôpital.

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.432   4.607   6.158   6.809   8.212  18.700
> sd(x)
[1] 3.420575
```

On applique une transformation logarithmique :

```
> y=log(x)
```

Voici des statistiques descriptive pour le poids transformé :

```
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3591  1.5276  1.8178  1.7990  2.1056  2.9285
> sd(y)
[1] 0.5064342
```

Calculer $a =$ la moyenne géométrique du poids (en livres) et $b =$ l'écart type géométrique du poids (en livres).

A) $a = 905,96$ et $b = 30,59$ B) $a = 6,809$ et $b = 3,42$
C) $a = 6,04$ et $b = 1,66$ D) $a = 1,79$ et $b = 0,50$
E) $a = 0,59$ et $b = 1,36$

Solution : La moyenne géométrique et l'écart type géométrique du poids en livres sont respectivement : $e^{\bar{y}} = e^{1,799} = 6,04$ et $e^{s_y} = e^{0,5064342} = 1,66$. La réponse est C.

10. Les données ci-dessous donne les poids à la naissance (en onces) pour six livraisons à l'Hôpital Civic. En supposant que le poids à la naissance suit une loi normale, trouver un intervalle de confiance à 90 % pour le poids moyen (en onces) à la naissance.

97   117   140   78   99   148

A) $[91,0;135,4]$ B) $[84,8;141,5]$ C) $[91,6;134,8]$
D) $[95,0;131,3]$ E) $[92,3;133,6]$

Solution : La moyenne de l'échantillon et l'écart type de l'échantillon sont respectivement :

$$\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 113,1667; \quad s = \sqrt{\frac{(\sum_{i=1}^{6} x_i^2) - (\sum_{i=1}^{6} x_i)^2/6}{6-1}} = 27,00679.$$

Un I.C. à 90% pour μ est

$$\overline{x} \pm t_{0,05;5}\,\frac{s}{\sqrt{n}} = [90,950;\ 135,383]$$

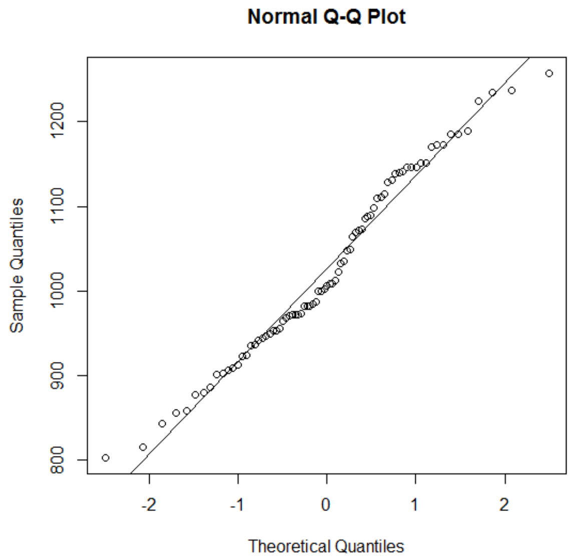
où $t_{0,05;5} = 2,015$. La réponse est A.

11. Le Dryas récent (ou le *grand gel*) a été un événement de refroidissement brusque de l'hémisphère Nord qui s'est produit il y a environ 12 000 ans et peut être attribuable à une modification des courants de l'océan Atlantique (COA) qui auraient cessé de convoyer de l'eau réchauffée de l'équateur vers l'Europe. Le moyen le plus commun de ralentir les COA implique la réduction de la densité de l'eau de surface océanique par une augmentation du débit d'eau douce dans l'Atlantique Nord. Pour prévoir si un tel événement peut se reproduise, la densité de l'eau de mer près de la surface est étroitement surveillée. La sortie de `R` suivante donne un sommaire des 79 mesures de la densité de l'eau de l'océan Atlantique près de la surface (en kg/m^3), à une latitude de 45 degrés nord :

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  802.5   948.1  1006.0  1026.0  1122.0  1258.0
> var(x)
[1] 12013.36
> sd(x)
[1] 109.61
```

Le diagramme ci-bas est le diagramme quantile-quantile pour ces données, qu'on ait produit avec les commandes suivantes :

```
> qqnorm(x)
> abline(mean(x),sd(x))
```



Si l'équation de la droite sur ce diagramme est $y = a + bz$, où y est le quantile de l'échantillon et z est le quantile théorique, quelle sont les valeurs de a et b ?

A) $a = 1006$ et $b = 109,61$ B) $a = 109,61$ et $b = 1026$
C) $a = 1026$ et $b = 109,61$ D) $a = 1026$ et $b = 173,90$
E) $a = 1006$ et $b = 12013,36$

Solution : L'équation est $y = \hat{\mu} + \hat{\sigma} z$ où $\hat{\mu} = \bar{x} = 1026$ et $\hat{\sigma} = s = 109,61$. Alors $a = 1026$ et $b = 109,61$. La réponse est C.

12. Voici les nombres d'attaques d'ours mortels en Amérique du Nord par décennie de 1900 à 1989.

$$2,\ 1,\ 4,\ 8,\ 6,\ 9,\ 9,\ 19,\ 20.$$

Calculer la moyenne et l'écart type du nombre d'attaques d'ours mortels en Amérique du Nord par décennie.

A) la moyenne est 8,667 et l'écart type est 5,6505
B) la moyenne est 8,0 et l'écart type est 19,0
C) la moyenne est 8,0 et l'écart type est 5,0
D) la moyenne est 8,667 et l'écart type est 46,0
E) La moyenne est 8,667 et l'écart type est 6,7823

Solution : La moyenne de cet échantillon est

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{78}{9} = 8.6667$$

et l'écart type est

$$s = \sqrt{\frac{(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2/n}{n-1}} = \sqrt{\frac{1044 - (78)^2/9}{9-1}} = 6,7823.$$

13. Aux Etats-Unis, les groupes sanguins ont la répartition suivante :

$$41\%\ O, \quad 31\%\ A, \quad 22\%\ B \quad \text{et} \quad 6\%\ AB.$$

Il est connu que O est un donneur universel, A peut seulement donner à A et AB, B peut seulement donner à B et AB, et AB peut seulement donner à AB. Si un patient qui a besoin d'une transfusion de sang reçoit le sang d'un donneur choisi au hasard, et que les deux personnes sont indépendants les uns des autres, quelle est la probabilité que la transfusion soit réussie ?

A) 0,6607 B) 0,3393 C) 0,4101 D) 0,7314 E) 0,5899

Solution : Soient A_1, A_2, A_3, A_4 les évenement que le donneur soit du groupe O, A, B, et AB, respectivement. Soient B_1, B_2, B_3, B_4 les événements que le patient qui reçoit la transfusion soit du groupe sanguin O, A, B, et AB, respectivement. L'événement A_i est indépendant de B_j, pour tout $i = 1, 2, 3, 4$ et $j = 1, 2, 3, 4$. L'événement que la transfusion est un réussie peut être écrit comme l'union d'événement mutuellement exclusifs :

$$C = (A_1 \cap B_1) \cup (A_1 \cap B_2) \cup (A_1 \cap B_3) \cup (A_1 \cap B_4) \cup (A_2 \cap B_2) \cup (A_2 \cap B_4) \cup (A_3 \cap B_3) \cup (A_3 \cap B_4) \cup (A_4 \cap B_4).$$

Alors,

$$\begin{aligned} P(C) &= P(A_1)P(B_1) + P(A_1)P(B_2) + P(A_1)P(B_3) + P(A_1)P(B_4) + \\ &\quad P(A_2)P(B_2) + P(A_2)P(B_4) + P(A_3)P(B_3) + P(A_3)P(B_4) + \\ &\quad P(A_4)P(B_4) \\ &= (0,41)(0,41) + (0,41)(0,31) + (0,41)(0,22) + (0,41)(0,06) + \\ &\quad (0,31)(0,31) + (0,31)(0,06) + (0,22)(0,22) + (0,22)(0,06) + \\ &\quad (0,06)(0,06) \\ &= 0,5899 \end{aligned}$$

La réponse est E.

14. 20 % des arbres dans une certaine forêt sont des érables. Dans cette forêt, 15% des érables sont des arbres matures, c'est-à-dire âgés entre 10 et 15 ans. Nous choisissons au hasard un arbre dans cette forêt. Quelle est la probabilité que c'est un érable âgé entre 10 et 15 ans ?

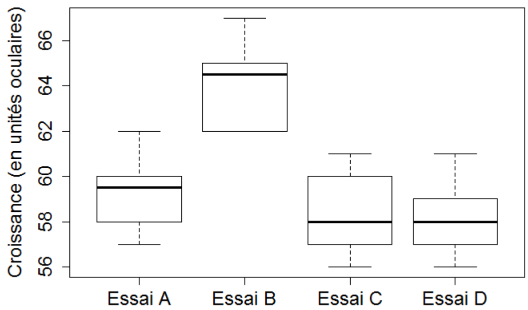
A) 0,03 B) 0,15 C) 0,20 D) 0,75 E) 0,175

Solution : Soit A l'événement que l'arbre soit un érable et B l'événement que l'arbre ait entre 10 et 15 ans. On sait que $P(A) = 0,2$ et $P(B|A) = 0,15$. Par la règle de la multiplication,

$$P(A \cap B) = P(B|A)P(A) = (0,15)(0,2) = 0,03$$

La réponse est A.

15. Les boîtes à moustaches comparatifs ci-bas montrent les effets de différents sucres sur la croissance des sections de pois cultivés en culture de tissu, mesurées en unités oculaires. (Une unité oculaire est 0,114 cm.) À l'essai A, 2% de glucose a été ajouté à la culture. À l'essai B, 2% de saccharose a été ajouté à la culture. À l'essai C, 1% de glucose et 2% de fructose a été ajouté à la culture. Enfin, à l'essai D, 1% de fructose a été ajouté à la culture.



Lequel des énoncés suivants est correct. (Seulement un de ces énoncés est correct.)

A) Les croissances médianes sont semblables pour les essais C et D.
B) Les distances interquartiles des croissances sont semblables pour les essais A, C et D, mais pas pour B.
C) Il y a des valeurs aberrantes aux essais A, C, et D, mais il n'y a pas

de valeur aberrante à l'essai B.
D) La distribution de la croissance pour l'essai B est approximativement symétrique.
E) La plus petite croissance fut à l'essai B.

Solution : La réponse est A.

16. Considérons un échantillon aléatoire de 50 femmes âgées de 20 à 29 ans. L'un des objectifs de l'étude est de décrire la répartition de l'indice de masse corporelle (IMC) pour ces femmes. Supposons que l'échantillon aléatoire est cueilli à partir d'une population de femmes ayant un IMC moyen de 26,8 et que l'écart type de la population est 7,42. Approximer la probabilité que l'IMC moyen de cet échantillon de 50 femmes sera plus grand que 29.

A) 0,0179 B) 0,0179 C) 0,6179 D) 0,3821 E) 0,0375

Solution : Soit $\overline{X}$ l'IMC moyen de cet échantillon. Par le théorème central limite,

$$Z = \frac{\overline{X} - 26,8}{7,42/\sqrt{50}}$$

suit approximativement une loi normale centrée et réduite. Alors,

$$\begin{aligned} P(\overline{X} > 29) &= 1 - P(\overline{X} \leq 29) \\ &= 1 - P\left(Z \leq \frac{29 - 26,8}{7,42/\sqrt{50}}\right) \\ &= 1 - P(Z \leq 2,10) \\ &\approx 1 - \Phi(2,10) \\ &= 1 - 0,9821 = 0,0179. \end{aligned}$$

La réponse est A ou B.

17. Une société pharmaceutique étudie un nouvel analgésique (médicament pour le soulagement de la douleur) sur un échantillon de six patients souffrant de migraine. Parmi ceux-ci, quatre patients ont rapporté que leurs migraines se sont atténuées après l'utilisation du médicament. Cependant, il est connu que 20 % des migraines vont s'atténuer de toute façon sans aucun traitement. Quelle est la probabilité que dans un échantillon de six patients souffrant de migraine, la migraine va s'atténuer sans aucun traitement pour exactement 4 des six patients ?

    A) 0,0016 B) 0,2534 C) 0,3523 D) 0,0154 E) 0,9992

    Solution : Soit X le nombre de patients qui auront une migraine qui va s'atténuer, parmi 6 patients sans traitement. Alors, X suit une loi binomiale avec $n = 6$ épreuves et p=0,2. On veut

$$P(X = 4) = \left(\begin{array}{c} 6 \\ 4 \end{array}\right) (0,2)^4\,(0,8)^2 = 0,01536$$

    La réponse est D.

18. La probabilité qu'un homme âgé de 40 à 55 va subir une crise cardiaque durant une période de 5 ans est 0,04. Un nouveau médicament est introduit dans l'espoir de réduire cette probabilité. Considérons une étude de 5 ans impliquant des hommes âgés de 40 à 55. Parmi les 2046 sujets donnés ce nouveau médicament, il y avait 56 sujets qui ont subi une crise cardiaque. Soit p soit la proportion d'hommes âgés de 40 à 55 utilisant ce nouveau médicament qui vont subir une crise cardiaque dans une période de 5 ans. Donner un intervalle de confiance à 95% pour p. Basé sur cet intervalle de confiance, sommes-nous confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque ?

    A) I.C.=[0,020 ; 0,034]. Nous ne sommes pas confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.
    B) I.C.=[0,020 ; 0,034]. Nous sommes confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.
    D) I.C.=[0,014 ; 0,041]. Nous sommes confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.
    D) I.C.=[0,041 ; 0,052]. Nous sommes confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.
    E) I.C.=[0,020 ; 0,056]. Nous ne sommes pas confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.

    Solution : Un intervalle de confiance à 95% pour p est

$$\hat{p} \pm 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0,020; 0,034]$$

    où $\hat{p} = 56/2046$ et $n = 2046$. Puisque toutes les valeurs dans l'intervalle sont inférieure à 0,04, nous confiant que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque.

19. Continuons avec la situation de la Question 18. Formuler l'hypothèse nulle et l'hypthèse alternative pour vérifier que le nouveau médicament est efficace pour réduire les risques d'une crise cardiaque. Ensuite, calculez la valeur P correspondante.

A) $H_0: \ p = 0,04$ contre $H_1: \ p > 0,04$. La valeur P est 0,9982.
B) $H_0: \ p = 0,04$ contre $H_1: \ p < 0,04$. La valeur P est 0,0036.
C) $H_0: \ p = 0,04$ contre $H_1: \ p < 0,04$. La valeur P est 0,0018.
D) $H_0: \ p = 0,04$ contre $H_1: \ p \neq 0,04$. La valeur P est 0,0036.
E) $H_0: \ p = 0,04$ contre $H_1: \ p < 0,04$. La valeur P est 0,0154.

Solution : On veut tester $H_0: \ p = 0,04$ contre $H_1: \ p < 0,04$. La valeur observée de la statistique du test est

$$z_0 = \frac{\hat{p} - 0,04}{\sqrt{(0,04)(0,96)/2046}} = -2,92,$$

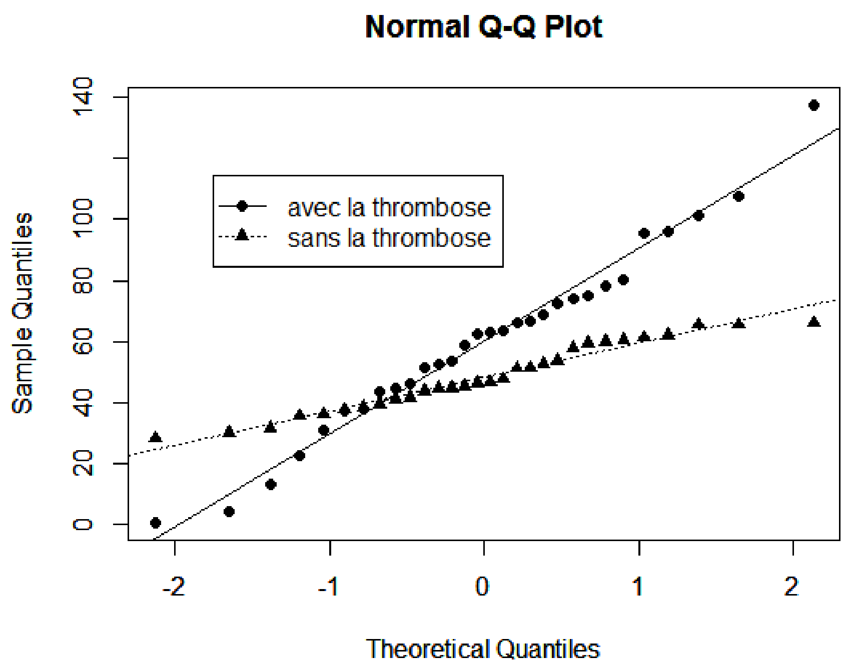
où $\hat{p} = 56/2046$. C'est une alternative unilatérale à la gauche, alors la valeur P est

$$P = P(Z < -2,92) = \Phi(-2,92) = 0,0018.$$

La réponse est C.

20. L'objectif d'une étude était de déterminer s'il y avait des concentrations différentes de l'anticardiolipine IgG (en mg/dl) chez les sujets avec et sans thrombose. Dans R, nous avons attribué la concentration d'IgG pour les 30 sujets avec la thrombose à la variable x. Alors que la concentration d'IgG pour les 30 sujets sans la thrombose a été attribuée à la variable y. Nous avons produit des statistiques descriptives pour les deux variables et des diagrammes quantile-quantile superimposés.

```
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.527  43.850  62.680  60.240  74.840 137.600
> summary(y)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
28.01   39.81   46.48   48.14   58.94   66.06
```



On veut utiliser la fonction `t.test()` pour vérifier que la concentration moyenne d'IgG pour les patients avec la thrombose est différente de la concentration moyenne d'IgG pour les patients sans la thrombose.
Lequel des énoncés suivant est correct ? (Seulement un des énoncés est correct.)

A) C'est évident que nous avons des échantillons de populations non-normales, alors on ne devrait pas utilisé la fonction `t.test`
B) C'est raisonnable de supposer que les populations sont normales et la bonne commande est `t.test(x,y)`
C) C'est raisonnable de supposer que les populations sont normales et la bonne commande est `t.test(x,y,var.equal=TRUE)`
D) C'est raisonnable de supposer que les populations sont normales et la bonne commande est `t.test(x,y,alternative="greater")`
E) C'est raisonnable de supposer que les populations sont normales et la bonne commande est `t.test(x,y,paired="TRUE")`

Solution : Il y a des tendances linéaires dans les deux diagrammes, alors c'est raisonnable de supposer que les populations sont normales. Alors, on peut utiliser la fonction `t.test()` pour comparer les moyennes. Alors A) est incorrect. Les pentes des droites sont très différentes, alors il n'est pas raisonnable de supposer l'égalité des variances. Alors, C) est incorrect. On n'a pas des mesures appariées. Alors, E) est incorrect. On veut tester une hypothèse nulle d'égalité des moyennes contre une alternative bilatérale. Alors, E est incorrect. La réponse est B.

21. Un vaccin expérimental contre le cancer a été mis au point pour réduire la taille d'une tumeur. Nous tenons à tester l'hypothèse nulle que le vaccin n'a pas d'effet sur la taille d'une tumeur, contre l'hypothèse alternative que le vaccin est efficace pour réduire la taille d'une tumeur. Soit μ_D la moyenne de la différence entre la taille du tumeur avant le vaccin et la taille du tumeur 3 mois après la vaccination. Formuler des hypothèses appropriées et interpréter une erreur de première espèce et une erreur de deuxième espèce dans ce contexte. Choisir l'énoncé qui est correct dans la liste ci-bas. (Un seul énoncé est correct.)

    A) $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. On commet une erreur de deuxième espèce, si on conclue que le vaccin est efficace à réduire la taille de la tumeur, mais dans la réalité le vaccin n'a pas d'effet.
    B) $H_0 : \mu_D = 0$ contre $H_1 : \mu_D < 0$. On commet une erreur de première espèce, si on conclue que le vaccin est efficace à réduire la taille de la tumeur, mais dans la réalité le vaccin n'a pas d'effet.
    C) $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. On commet une erreur de première espèce, si on conclue que le vaccin n'a pas d'effet sur la taille de la tumeur, mais dans la réalité le vaccin est efficace à réduire la taille du tumeur.
    D) $H_0 : \mu_D = 0$ contre $H_1 : \mu_D \neq 0$. On commet une erreur de première

espèce, si on conclue que le vaccin est efficace à réduire la taille du tumeur, mais dans la réalité le vaccin n'a pas d'effet.

E) $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. On commet une erreur de deuxième espèce, si on conclue que le vaccin n'a pas d'effet sur la taille de la tumeur, mais dans la réalité le vaccin est efficace à réduire la taille du tumeur.

Solution : On veut tester $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. Nous allons commettre une erreur de première espèce, si on conclue que le vaccin est efficace à réduire la taille du tumeur, mais dans la réalité le vaccin n'a pas d'effet. Nous allons commettre une erreur de deuxième espèce, si on conclue que le vaccin n'a pas d'effet sur la taille du tumeur, mais dans la réalité le vaccin est efficace à réduire la taille du tumeur. La réponse est E.

22. Une étude est menée pour étudier le lien entre le nombre X d'heures d'exercice par semaine et la pression artérielle systolique Y pour les hommes de 50 ans. Les données suivantes ont été obtenues sur les 10 sujets :

Nombre d'heures x_i	Pression artérielle systolique y_i	x_i^2	$x_i\, y_i$
4	120	16	480
10	110	100	1100
2	120	4	240
3	135	9	405
3	140	9	420
5	115	25	575
1	150	1	150
2	165	4	330
2	160	4	320
0	180	0	0

Pour ces données, nous avons :

$$\sum_{i=1}^{10} x_i = 32, \quad \sum_{i=1}^{10} y_i = 1395, \quad \sum_{i=1}^{10} x_i^2 = 172, \quad \sum_{i=1}^{10} x_i\, y_i = 4020$$

Donner la droite de régression estimée et estimer la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine.

A) La droite de régression estimée est $\hat{y} = 181,19 - (8,6)\,x$ et une estimation pour la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est 130.

B) La droite de régression estimée est $\hat{y} = 167,31 - (7,12)\,x$ et une estimation pour la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est 125.
C) La droite de régression estimée est $\hat{y} = 138,90 - (5,25)\,x$ et une estimation pour la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est 107.
D) La droite de régression estimée est $\hat{y} = 145,76 - (3,42)\,x$ et une estimation pour la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est 125.
E) La droite de régression estimée est $\hat{y} = 159,91 - (6,38)\,x$ et une estimation pour la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est 122.

Solution : Pour calculer $\hat{\beta}$ et $\hat{\alpha}$, on utilise les formules :

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i\, y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)/n}{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2/n} \quad \text{et} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

On obtient

$$\hat{\beta} = \frac{(4020) - (32)(1395)/10}{(172) - (32)^2/10} = -6,37931, \quad \hat{\alpha} = \frac{1395}{10} + (6,37931)\frac{32}{10} = 159,9138$$

Alors la droite de régression estimée est

$$\hat{y} = 159,91 - (6,38)x$$

Une estimation de la pression artérielle systolique moyenne pour un individu qui fait 6 heures d'exercices par semaine est

$$\hat{\mu}_{Y|x=6} = 159,91 - (6,38)(6) = 121,6.$$

La réponse est E.

23. Une hypothèse importante dans la recherche de l'hypertension est que la réduction du sodium peut abaisser la tension artérielle. Puisque la réduction de sodium est difficile de maintenir pendant une longue période de temps, des conseils diététiques sont parfois utilisés pour atteindre cet objectif. Des niveaux de sodium urinaires ont été obtenus de huit individus inscrits dans un programme de réduction de sodium. Ils ont mesurés la teneur de sodium avant l'intervention et une semaine après le début de l'intervention. Voici les données :

```
> Semaine0=c(7.85,12.03,21.84,13.94,16.68,41.78,14.97,12.07)
> Semaine1=c(5.59,8.5,4.55,12.78,10.69,23.51,5.46,11.95)
```

Le chercheur A utilise la commande suivante :

```
> t.test(Semaine0,Semaine1,alternative="greater")
```

Voici la résultat :

```
data:  Semaine0 and Semaine1
t = 1.6814, df = 11.267, p-value = 0.06008
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.4779779        Inf
sample estimates:
mean of x mean of y
 17.64500  10.37875
```

Le chercheur B utilise la commande suivante :

```
> t.test(Semaine0,Semaine1,paired=TRUE,alternative="greater")
```

Voici le résultat :

```
data:  Semaine0 and Semaine1
t = 2.8835, df = 7, p-value = 0.01177
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.492072      Inf
sample estimates:
mean of the differences
               7.26625
```

Supposons que chacun de ces chercheurs ont vérifié la condition d'application de normalité avant d'utiliser la fonction `t.test`.

Lequel de ces énoncés est correct ? (Seulement un énoncé est correct.)

A) Le chercheur A a utilisé la commande correctement. À $\alpha = 0,05$, il y a des preuves significatives que la teneur de sodium ait diminué en moyenne après l'intervention.
B) Le chercheur B a utilisé la commande correctement. À $\alpha = 0,05$, il y a des preuves significatives que la teneur de sodium ait diminué en

moyenne après l'intervention.
C) Le chercheur A a utilisé la commande correctement. À $\alpha = 0,05$, les preuves que la teneur de sodium ait diminué en moyenne après l'intervention ne sont pas significatives.
D) Le chercheur B a utilisé la commande correctement. À $\alpha = 0,05$, les preuves que la teneur de sodium ait diminué en moyenne après l'intervention ne sont pas significatives.
E) Ni l'un ni l'autre a utilisé la commande correctement. Une statistique t ne devrait pas être utilisé dans cette situation.

Solution : Nous avons des mesures appariées, alors le chercheur B a utilisé la commande correctement. Soit μ_D la différence entre la teneur de sodium avant et aprés l'intervention. Puisque dans la sortie du chercheur B, on observe une valeur P inférieure à 0,05, on a des preuves significatives contre $H_0 : \ \mu_D = 0$ en faveur de $H_1 : \ \mu_D > 0$. Alors, il y a des preuves significatives que la teneur de sodium ait diminué en moyenne après l'intervention. La réponse est B.

24. Le lien plante-eau joue un rôle important dans la physiologie des plantes. Nous considérons une expérience dans laquelle 16 jeunes plants de bouleau ont été inondées avec de l'eau pour une journée et 13 autres plants ont été utilisé comme un groupe témoin (c'est-à-dire des contrôles). À la fin de l'expérience, les racines de toutes les plantes ont été analysées pour le niveau de l'adénosine triphosphate (ATP), en tant que mesure pour le transfert d'énergie intracellulaire. Voici un sommaire des données :

	groupe inondé	groupe témoin
taille de l'échantillon	$n_1 = 16$	$n_2 = 13$
moyenne de l'échantillon	$\bar{x}_1 = 1,17$	$\bar{x}_2 = 1,91$
écart type de l'échantillon	$s_1 = 0,16$	$s_2 = 0,23$

Donner un intervalle de confiance à 90% pour $\mu_1 - \mu_2$, où μ_1 est l'ATP moyen pour les plants inondés et μ_2 est l'ATP moyen pour les plants non-inondés. Basé sur cet intervalle, peut-on conclure que les inondations causes une augmentation ou une diminution des niveaux moyens d'ATP ? (Supposons que le niveau d'ATP d'une plante inondée et le niveau d'ATP d'une plante non-inondée sont normalement distribués avec des variances égaux.)

A) [0,5673 ; 0,7614] ; l'inondation cause une augmentation du niveau moyen d'ATP.
B) [0,4532 ; 0,6719] ; l'inondation cause une augmentation du niveau

moyen d'ATP.
C) [-0,6182 ; -0,4820] ; l'inondation cause une diminution du niveau moyen d'ATP.
D) [-0,8635 ; -0,6165] ; l'inondation cause une diminution du niveau moyen d'ATP.
E) [-0,0346 ; 0,3471] ; on ne peut pas conclure que l'inondation cause une diminution ou une augmentation du niveau moyen d'ATP.
Solution : L'écart type pondéré est

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(15)(0.16)^2+(12)(0.23)^2}{16+13-2}} = 0.19425.$$

Un intervalle de confiance à 90% pour $\mu_1 - \mu_2$ est

$$\bar{x}_1 - \bar{x}_2 \pm t_{0,05;16+13-2} s_p \sqrt{1/n_1 + 1/n_2} = [-0,8635; -0,6165],$$

où $t_{0,05;16+13-2} = t_{0,05;27} = 1,703$. Puisque l'intervalle contient seulement des valeurs négatives, alors on est confiant que $\mu_1 < \mu_2$. On peut conclure que l'inondation cause une diminution du niveau moyen d'ATP. La réponse est D.

25. Le niveau moyen de pression artérielle systolique d'une certaine population est approximativement égale à 125 mm Hg. Un sujet d'intérêt récent est que l'utilisation extensive de contraceptif oral peut provoquer une réduction de la pression artérielle systolique moyenne. Une étude est planifiée pour vérifier cette hypothèse. Les n femmes qui ont participé à cette étude ont utilisé des contraceptifs orals pour une période de trois mois. À la fin de l'étude, leur pression artérielle systolique a été mesurée. Ces pressions artérielles systoliques furent attribuées à la variable `x` dans `R`. On affiche une sortie de `R` :

```
> mean(x)
[1] 120.4
> sd(x)
[1] 13.23
> t.test(x,mu=125,alternative="less")

        One Sample t-test

data:  x
t = -1.0998, df =?, p-value = 0.15
```

```
alternative hypothesis: true mean is less than 125
95 percent confidence interval:
    -Inf 128.067
sample estimates:
mean of x
    120.4
```

Déterminer le nombre n de participants pour cette étude.
(N.B. Dans la sortie ci-haut, on a remplacé le nombre de degrés de liberté par ?.)

A) 12 B) 40 C) 10 D) 32 E) 25

Solution : On veut tester $H_0 : \mu = 125$ contre $H_1 : \mu < 125$. La valeur observée de la statistique du test est

$$t_0 = \frac{\bar{x} - 125}{s/\sqrt{n}} = \frac{120,4 - 125}{13,23/\sqrt{n}}.$$

De la sortie de `R`, on a $t_0 = -1,0998$. Alors,

$$\frac{120,4 - 125}{13,23/\sqrt{n}} = -1,0998.$$

Donc,

$$n = \left(\frac{-1,0998 \times 13,23}{120,4 - 125} \right)^2 = 10,005$$

On peut conclure que $n = 10$. La réponse est C.