

STT-1920 MÉTHODES STATISTIQUES

SOLUTIONS DES EXERCICES DU CHAPITRE 3

NUMÉRO 1. Supposons que la loi normale avec moyenne μ et variance σ^2 soit un bon modèle pour décrire la distribution d'une certaine variable d'intérêt dans une certaine population. Supposons qu'on obtienne un échantillon aléatoire de taille 7 à partir de cette population. On pose

$$A = \frac{\bar{X} - \mu}{\sigma/\sqrt{7}} \quad B = \frac{\bar{X} - \mu}{S/\sqrt{7}} \quad C = \frac{6S^2}{\sigma^2}$$

- (a) Quelle est la distribution de la variable aléatoire A ?
- (b) Quelle est la distribution de la variable aléatoire B ?
- (c) Quelle est la distribution de la variable aléatoire C ?
- (d) Que vaut $\mathbb{E}[A]$? $\mathbb{E}[B]$? $\mathbb{E}[C]$? $\mathbb{V}\text{ar}[A]$? $\mathbb{V}\text{ar}[B]$? $\mathbb{V}\text{ar}[C]$?
- (e) Obtenez $\mathbb{P}[1.04 < A \leq 1.65]$, $\mathbb{P}[-1.34 < A \leq 0.35]$ et $\mathbb{P}[A > 1.25]$.
- (f) Obtenez $\mathbb{P}[0.718 < B \leq 1.44]$, $\mathbb{P}[B \leq -0.906]$ et $\mathbb{P}[B > 3.143]$.
- (g) Obtenez $\mathbb{P}[5.35 < C \leq 10.65]$, $\mathbb{P}[C \leq 1.64]$ et $\mathbb{P}[C > 2.20]$.
- (h) Dessinez le graphe de la densité de la variable A et indiquez sur ce graphe la probabilité $\mathbb{P}[1.04 < A \leq 1.65]$.
- (i) Dessinez le graphe de la densité de la variable B et indiquez sur ce graphe la probabilité $\mathbb{P}[0.718 < B \leq 1.44]$.
- (j) Dessinez le graphe de la densité de la variable C et indiquez sur ce graphe la probabilité $\mathbb{P}[C > 2.20]$.

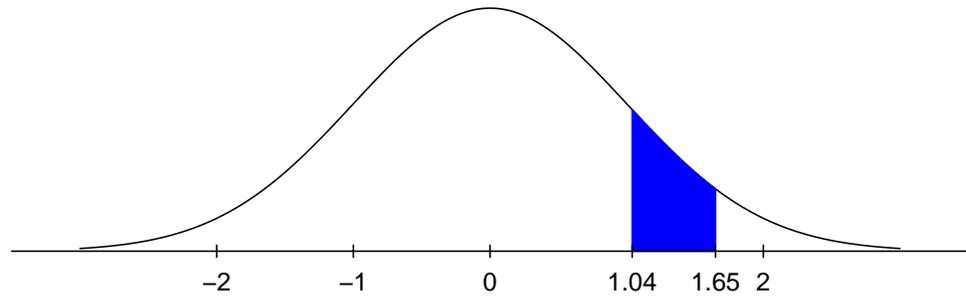
SOLUTION.

- (a) $A \sim N(0, 1)$.
- (b) $B \sim t_6$.
- (c) $C \sim \chi_6^2$.
- (d) $\mathbb{E}[A] = 0$, $\mathbb{E}[B] = 0$, $\mathbb{E}[C] = 6$, $\mathbb{V}\text{ar}[A] = 1$, $\mathbb{V}\text{ar}[B] = 3/2$, $\mathbb{V}\text{ar}[C] = 12$.
- (e) Avec la table de la loi $N(0, 1)$ ou avec R-Commander, j'obtiens
 - $\mathbb{P}[1.04 < A \leq 1.65] = \Phi(1.65) - \Phi(1.04) = 0.0997$,
 - $\mathbb{P}[-1.34 < A \leq 0.35] = \Phi(0.35) - \Phi(-1.34) = 0.5467$,
 - $\mathbb{P}[A > 1.25] = 1 - \Phi(1.25) = 0.1056$.
- (f) Avec la table de la loi de Student ou avec R-Commander, j'obtiens
 - $\mathbb{P}[0.718 < B \leq 1.44] = 0.15$,
 - $\mathbb{P}[B \leq -0.906] = 0.20$,
 - $\mathbb{P}[B > 3.143] = 0.01$.

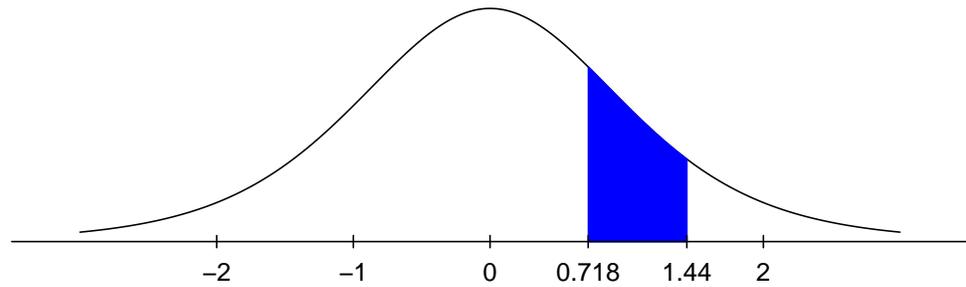
(g) Avec la table de la loi du khi-deux ou avec R-Commander, j'obtiens

- $\mathbb{P}[5.35 < C \leq 10.65] = 0.40$,
- $\mathbb{P}[C \leq 1.64] = 0.05$,
- $\mathbb{P}[C > 2.20] = 0.90$.

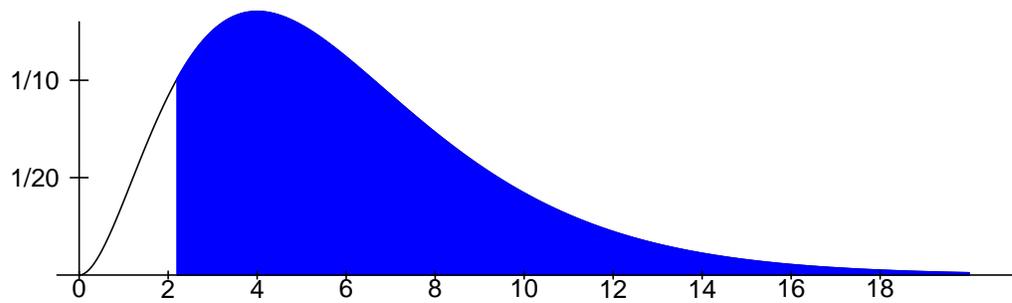
(h) $\mathbb{P}[1.04 < A \leq 1.65] = 0.0997$



(i) $\mathbb{P}[0.718 < B \leq 1.44] = 0.15$



(j) $\mathbb{P}[C > 2.20] = 0.90$



NUMÉRO 2. On a mesuré les poids de 16 kiwis provenant d'une ferme de Bay of Plenty en Nouvelle-Zélande. Voici ces 16 poids, en grammes :

65.06	71.44	67.93	69.02	67.28	62.34	66.23	64.16
68.56	70.45	64.91	69.90	65.52	66.75	68.54	67.90

On suppose que la loi normale avec moyenne μ et variance σ^2 est un bon modèle pour décrire la distribution des poids des kiwis de Bay of Plenty.

- (a) Estimez la moyenne théorique μ à partir de ces données.
- (b) Quelle est l'erreur type associée à l'estimation précédente ?
- (c) Estimez la variance théorique σ^2 à partir de ces données.
- (d) Quelle est l'erreur type associée à l'estimation précédente ?
- (e) Obtenez un intervalle de confiance de niveau 95% pour la moyenne μ .
- (f) Obtenez un intervalle de confiance de niveau 90% pour la variance σ^2 .
- (g) Obtenez un intervalle de confiance de niveau 90% pour l'écart-type σ .
- (h) Tracez le diagramme en boîte, l'histogramme et le *normal Q-Q plot* de ces 20 observations. Est-ce qu'il vous semble raisonnable de supposer que la loi normale est un bon modèle pour la population des poids des kiwis de Bay of Plenty ?
- (i) Avec le logiciel R, faites le test de normalité de Shapiro et Wilk. Quelles sont vos conclusions ?

SOLUTION.

- (a) L'estimation de la moyenne théorique μ est $\bar{x} = 67.249$ grammes.
- (b) L'erreur type associée à l'estimation précédente est $s/\sqrt{n} = 2.455/\sqrt{16} = 0.614$ grammes.
- (c) L'estimation de la variance théorique σ^2 est $s^2 = 6.029$ grammes².
- (d) L'erreur type associée à l'estimation précédente est $\sqrt{2} s^2/\sqrt{n-1} = \sqrt{2} \times 6.029/\sqrt{15} = 2.201$ grammes².
- (e) J'utilise l'intervalle

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

avec $n = 16$, $\alpha = 0.05$, $\bar{x} = 67.249$ et $s = 2.455$. La table me donne $t_{n-1, \alpha/2} = t_{15, 0.025} = 2.131$. J'obtiens l'intervalle (65.94, 68.56).

- (f) J'utilise l'intervalle

$$\left(\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2} s^2, \frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} s^2 \right)$$

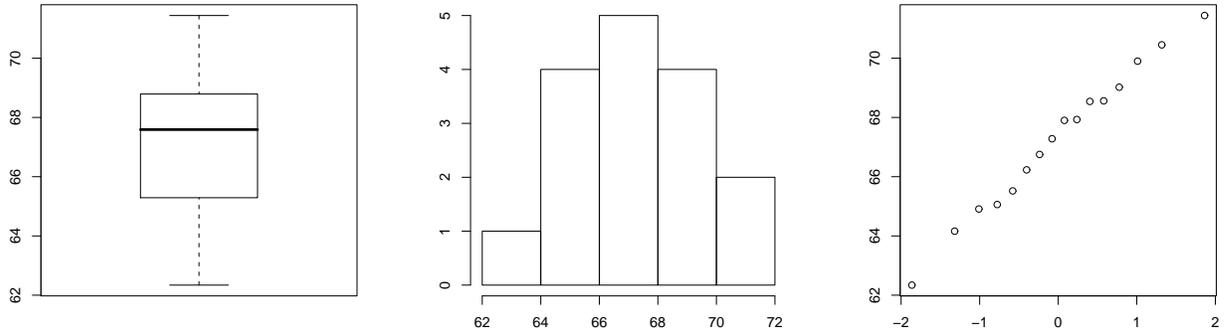
avec $n = 16$, $\alpha = 0.10$, et $s^2 = 6.029$. La table me donne $\chi_{n-1, \frac{\alpha}{2}}^2 = \chi_{15, 0.05}^2 = 25.00$ et $\chi_{n-1, 1-\frac{\alpha}{2}}^2 = \chi_{15, 0.95}^2 = 7.261$. J'obtiens l'intervalle (3.62, 12.45).

(g) J'utilise l'intervalle

$$\left(\sqrt{\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}} s \right)$$

et j'obtiens l'intervalle (1.90, 3.53).

(h) Voici les graphes. Rien ne suggère que la loi normale n'est pas appropriée.



(i) Je fais le test de Shapiro et Wilk avec R-Commander et j'obtiens
 $W = 0.9877$, $p\text{-value} = 0.9971$

Conclusion : Le $p\text{-value}$ est très grand ! Il n'y a pas lieu de rejeter l'hypothèse de normalité.

NUMÉRO 3. On désire estimer la proportion de chevreuil d'Anticosti qui sont atteints du syndrome MO-22. Nous allons obtenir un échantillon aléatoire de chevreuils d'Anticosti et nous allons déterminer combien parmi eux sont atteints de ce syndrome. Nous allons ensuite calculer un intervalle de confiance de niveau 90% pour la proportion théorique p des chevreuils d'Anticosti atteints de ce syndrome.

- Quelle doit être la taille de l'échantillon si on veut être assuré que l'erreur de notre intervalle de confiance (c'est-à-dire la demi-longueur) soit au plus 0.03 ?
- On obtient un échantillon de 57 chevreuils. Parmi ces 57 chevreuils, il y en a 23 qui sont atteints du syndrome MO-22. Estimez la proportion p .
- Quelle est l'erreur type associée à l'estimation précédente ?
- Obtenez un intervalle de confiance de niveau 90% pour la proportion p .

SOLUTION.

- On veut la taille d'échantillon qui nous assure que l'erreur de notre intervalle de confiance (c'est-à-dire sa demi-longueur) sera au plus 0.03. On résout l'équation

$$\frac{z_{\alpha/2}}{2\sqrt{n}} = 0.03$$

avec $z_{\alpha/2} = z_{0.05} = 1.645$. On obtient $n = 752$.

(b) On a $n = 57$ et $\hat{p} = 23/57$. L'estimation de la proportion p est $\hat{p} = 23/57 = 0.4035$.

(c) L'erreur type associée à l'estimation précédente est

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.4035(1-0.4035)}{57}} = 0.065.$$

(d) J'utilise l'intervalle

$$\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

et j'obtiens l'intervalle (0.297, 0.510).

NUMÉRO 4. Le fichier `Serpents.xls` (voir section *Notes de cours et exercices* du site web du cours) contient les poids en grammes (colonne 1) et les longueurs en centimètres (colonne 2) de 147 serpents nouveaux-nés. On suppose que ces 147 serpents constituent un échantillon aléatoire de taille $n = 147$ issu de la population de tous les serpents nouveaux-nés de cette espèce. On suppose que la loi normale avec moyenne μ_1 et variance σ_1^2 est un bon modèle pour décrire la distribution des poids dans cette population et que loi normale avec moyenne μ_2 et variance σ_2^2 est un bon modèle pour décrire la distribution des longueurs dans cette population.

- Obtenez un intervalle de confiance de niveau 95% pour μ_1 .
- Obtenez un intervalle de confiance de niveau 95% pour σ_1 .
- L'hypothèse de normalité de la distribution des poids est-elle raisonnable? Justifiez votre réponse.
- Obtenez un intervalle de confiance de niveau 95% pour μ_2 .
- Obtenez un intervalle de confiance de niveau 95% pour σ_2 .
- L'hypothèse de normalité de la distribution des longueurs est-elle raisonnable? Justifiez votre réponse.

SOLUTION.

(a) J'utilise la formule

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

avec $n = 147$, $\alpha = 0.05$, $\bar{x} = 34.882$ et $s = 0.956$. J'obtiens l'intervalle (34.726, 35.038).

(b) J'utilise l'intervalle

$$\left(\sqrt{\frac{n-1}{\chi_{n-1, \alpha/2}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, 1-\alpha/2}^2}} s \right)$$

avec $n = 147$, $\alpha = 0.05$ et $s = 0.956$. J'obtiens l'intervalle (0.858, 1.080).

- (c) Avec le logiciel R-Commander, je fais le test de Shapiro et Wilk. J'obtiens $W = 0.9943$, $p\text{-value} = 0.8315$

Conclusion : Le $p\text{-value}$ est très grand ! Il n'y a pas lieu de rejeter l'hypothèse de normalité.

- (d) J'utilise la formule

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

avec $n = 147$, $\alpha = 0.05$, $\bar{x} = 8.740$ et $s = 0.376$. J'obtiens l'intervalle $(8.679, 8.801)$.

- (e) J'utilise l'intervalle

$$\left(\sqrt{\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}} s \right)$$

avec $n = 147$, $\alpha = 0.05$ et $s = 0.376$. J'obtiens l'intervalle $(0.337, 0.425)$.

- (f) Avec le logiciel R-Commander, je fais le test de Shapiro et Wilk. J'obtiens $W = 0.9917$, $p\text{-value} = 0.5508$

Conclusion : Le $p\text{-value}$ est très grand ! Il n'y a pas lieu de rejeter l'hypothèse de normalité.

NUMÉRO 5. Voici 16 observations obtenues à partir d'une certaine population :

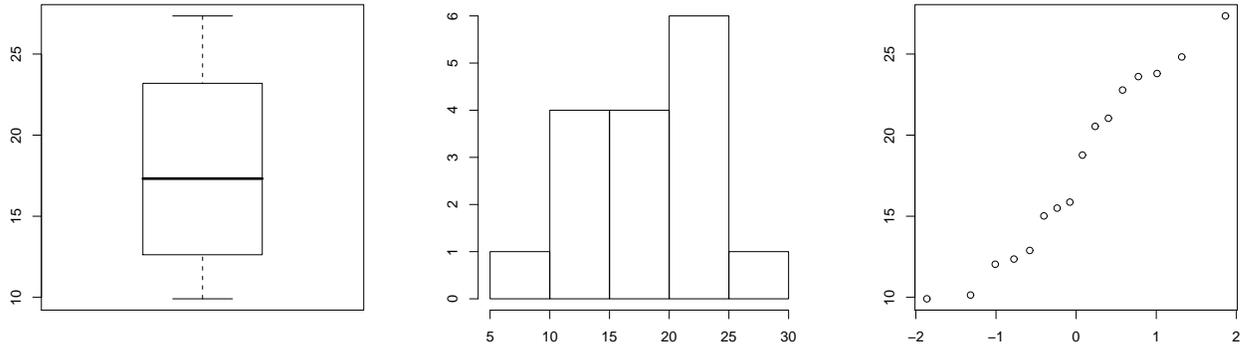
24.82	21.04	10.14	15.87	12.36	27.35	15.02	23.80
9.91	20.54	12.04	15.50	12.89	22.78	23.61	18.77

Que peut-on dire au sujet de la population ? Plus précisément, répondez aux questions suivantes :

- La loi normale est-elle un modèle approprié ?
- Que peut-on dire au sujet de la moyenne de cette population ?
- Que peut-on dire au sujet de l'écart-type de cette population ?

SOLUTION.

- La loi normale est-elle un modèle approprié ? Voici le boxplot, l'histogramme et le *normal Q-Q plot*. Il n'y a pas lieu de mettre en doute l'hypothèse de normalité.



Je fais le test de Shapiro et Wilk avec R-Commander et j'obtiens
 $W = 0.9384$, $p\text{-value} = 0.3295$

Conclusion : Le $p\text{-value}$ est très grand ! Il n'y a pas lieu rejeter de l'hypothèse de normalité !

- (b) L'estimation de la moyenne de la population est $\bar{x} = 17.90$. L'erreur type associée à cette estimation est $s/\sqrt{n} = 5.647/\sqrt{16} = 1.412$. On peut calculer un intervalle de confiance pour la moyenne de la population à l'aide de la formule

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

Par exemple, avec $\alpha = 0.05$ on obtient l'intervalle de confiance à 95% : (14.89, 20.91).

- (c) L'estimation de l'écart-type de la population est $s = 5.647$. On peut calculer un intervalle de confiance pour l'écart-type de la population à l'aide de la formule

$$\left(\sqrt{\frac{n-1}{\chi_{n-1, \alpha/2}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, 1-\alpha/2}^2}} s \right)$$

Par exemple, avec $\alpha = 0.05$ on obtient l'intervalle de confiance à 95% : (4.172, 8.740).

NUMÉRO 6. On obtient un échantillon aléatoire de taille n . On suppose que la loi $N(\mu, \sigma^2)$ est un bon modèle pour la population à partir de laquelle on échantillonne.

Pour chaque item de la liste de gauche, associez un item de la liste de droite.

- | | |
|---|------------------------------------|
| 1. l'échantillon aléatoire (pas encore observé) | (a) μ |
| 2. l'échantillon aléatoire (observé) | (b) σ^2 |
| 3. la moyenne échantillonnale (pas encore observée) | (c) $N(0, 1)$ |
| 4. la distribution de la moyenne échantillonnale | (d) χ_{n-1}^2 |
| 5. la moyenne échantillonnale (observée) | (e) σ^2/n |
| 6. la variance échantillonnale (pas encore observée) | (f) σ/\sqrt{n} |
| 7. la variance échantillonnale (observée) | (g) \bar{X} |
| 8. l'écart-type échantillonnal (pas encore observé) | (h) $2\sigma^4/(n-1)$ |
| 9. l'écart-type échantillonnal (observé) | (i) s |
| 10. l'espérance de la moyenne échantillonnale | (j) x_1, x_2, \dots, x_n |
| 11. la variance de la moyenne échantillonnale | (k) $\sqrt{2} \sigma^2/\sqrt{n-1}$ |
| 12. l'écart-type de la moyenne échantillonnale | (l) $N(\mu, \sigma^2)$ |
| 13. l'espérance de la variance échantillonnale | (m) X_1, X_2, \dots, X_n |
| 14. la variance de la variance échantillonnale | (n) S^2 |
| 15. l'écart-type de la variance échantillonnale | (o) S |
| 16. (la valeur théorique de) l'erreur type associée à \bar{x} | (p) p |
| 17. (l'estimation de) l'erreur type associée à \bar{x} | (q) \bar{x} |
| 18. (la valeur théorique de) l'erreur type associée à s^2 | (r) r |
| 19. (l'estimation de) l'erreur type associée à s^2 | (s) $Q2$ |
| 20. la distribution de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ | (t) $\sqrt{2} s^2/\sqrt{n-1}$ |
| 21. la distribution de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ | (u) s/\sqrt{n} |
| 22. la distribution de $\frac{(n-1)S^2}{\sigma^2}$ | (v) $N(\mu, \sigma^2/n)$ |
| 23. la moyenne de la population | (w) t_{n-1} |
| 24. la variance de la population | (x) S |
| 25. l'écart-type de la population | (y) s^2 |
| 26. la distribution de la population | (z) σ |
| 27. la médiane de la population | |
| 28. la médiane échantillonnale | |

RÉPONSES. Voici les bonnes associations :

l'échantillon aléatoire (pas encore observé)	\longleftrightarrow	X_1, X_2, \dots, X_n
l'échantillon aléatoire (observé)	\longleftrightarrow	x_1, x_2, \dots, x_n
la moyenne échantillonnale (pas encore observé)	\longleftrightarrow	\bar{X}
la distribution de la moyenne échantillonnale	\longleftrightarrow	$N(\mu, \sigma^2/n)$
la moyenne échantillonnale (observée)	\longleftrightarrow	\bar{x}
la variance échantillonnale (pas encore observé)	\longleftrightarrow	S^2
la variance échantillonnale (observée)	\longleftrightarrow	s^2
l'écart-type échantillonnal (pas encore observé)	\longleftrightarrow	S
l'écart-type échantillonnal (observé)	\longleftrightarrow	s
l'espérance de la moyenne échantillonnale	\longleftrightarrow	μ
la variance de la moyenne échantillonnale	\longleftrightarrow	σ^2/n
l'écart-type de la moyenne échantillonnale	\longleftrightarrow	σ/\sqrt{n}
l'espérance de la variance échantillonnale	\longleftrightarrow	σ^2
la variance de la variance échantillonnale	\longleftrightarrow	$2\sigma^4/(n-1)$
l'écart-type de la variance échantillonnale	\longleftrightarrow	$\sqrt{2} \sigma^2/\sqrt{n-1}$
(la valeur théorique de) l'erreur type associée à \bar{x}	\longleftrightarrow	σ/\sqrt{n}
(l'estimation de) l'erreur type associée à \bar{x}	\longleftrightarrow	s/\sqrt{n}
(la valeur théorique de) l'erreur type associée à s^2	\longleftrightarrow	$\sqrt{2} \sigma^2/\sqrt{n-1}$
(l'estimation de) l'erreur type associée à s^2	\longleftrightarrow	$\sqrt{2} s^2/\sqrt{n-1}$
la distribution de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	\longleftrightarrow	$N(0, 1)$
la distribution de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$	\longleftrightarrow	t_{n-1}
la distribution de $\frac{(n-1)S^2}{\sigma^2}$	\longleftrightarrow	χ_{n-1}^2
la moyenne de la population	\longleftrightarrow	μ
la variance de la population	\longleftrightarrow	σ^2
l'écart-type de la population	\longleftrightarrow	σ
la distribution de la population	\longleftrightarrow	$N(\mu, \sigma^2)$
la médiane de la population	\longleftrightarrow	μ
la médiane échantillonnale	\longleftrightarrow	$Q2$

NUMÉRO 7. Le fichier `0oiseaux.xls` (voir section *Notes de cours et exercices* du site web du cours) contient les informations suivantes sur 49 moineaux :

Colonne 1 : un code d'identification.

Colonne 2 : la variable dichotomique **survie** (oui ou non).

Colonne 3 : la variable **longueur** (du bout de la queue jusqu'à la tête en mm).

Colonne 4 : la variable **etendue** (du bout d'une aile à l'autre en mm).

Colonne 5 : la variable **tete** (longueur de la tête en mm).

Colonne 6 : la variable **humerus** (longueur de l'humérus en mm).

Colonne 7 : la variable **sternum** (longueur du sternum en mm).

Nous allons supposer que ces 49 moineaux constituent un échantillon aléatoire issu d'une certaine population de moineaux. La variable dichotomique (colonne 2) est

la réponse à un certain traitement : ou bien le moineau survit au traitement, ou bien il meurt. Les cinq autres variables (colonnes 3 à 7) sont des variables quantitatives de type continu. Dans le présent exercice, nous examinons les six variables individuellement. Les liens entre ces différentes variables seront examinés plus tard.

- (a) Dénotons par p la proportion théorique de survie. Voici deux façons équivalentes d'interpréter cette proportion p . Si toute la population de moineaux était soumise au traitement, la proportion de moineaux qui survivraient au traitement serait p . Si on choisit un moineau au hasard à partir de cette population et si on lui administre le traitement, la probabilité qu'il survive est p . On veut tester $H_0 : p = 1/2$ vs $H_1 : p < 1/2$.
- Énoncez clairement la règle de décision au seuil 5%.
 - Avec les données du fichier `oiseaux.xls`, est-ce qu'on accepte ou est-ce qu'on rejette H_0 au seuil 5% ?
 - D'après l'approximation par la loi normale, quel est le p -value ?
 - Obtenez la valeur exacte du p -value à l'aide de la distribution binomiale.
- (b) On se demande si la loi normale est un bon modèle pour décrire les distributions des cinq variables quantitatives. Pour chacune de ces cinq variables,
- dessinez le diagramme en boîte (boxplot) ;
 - dessinez l'histogramme ;
 - dessinez le graphe quantile-quantile gaussien (*normal Q-Q plot*) ;
 - performez le test de normalité de Shapiro et Wilk.
- Quelles sont vos conclusions ?
- (c) Obtenez un intervalle de confiance de niveau 95% pour la moyenne théorique de la variable `etendue`.
- (d) Obtenez un intervalle de confiance de niveau 95% pour l'écart-type théorique de la variable `etendue`.
- (e) Pour la variable `tete`, testez $H_0 : \mu = 31$ mm vs $H_1 : \mu \neq 31$ mm. Quel est votre p -value ?
- (f) Pour la variable `humerus`, testez $H_0 : \sigma = 1/2$ mm vs $H_1 : \sigma > 1/2$ mm. Quel est votre p -value ?

Vous pouvez consulter la solution clip sur le site web du cours. Voici une solution traditionnelle.

SOLUTION DE LA PARTIE (a)

- (i) On rejette H_0 si $Z \leq -z_\alpha$. Ici la statistique Z est donnée par

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\hat{p} - (1/2)}{\sqrt{\frac{(1/2)(1/2)}{49}}}$$

Avec $\alpha = 0.05$, on obtient $z_\alpha = z_{0.05} = 1.645$.

(ii) Avec $\hat{p} = 21/49 = 3/7 = 0.4286$, on obtiens

$$Z_{obs} = \frac{(3/7) - (1/2)}{\sqrt{\frac{(1/2)(1/2)}{49}}} = -1.00.$$

Il n'y a pas lieu de rejeter H_0 .

(iii)

$$p\text{-value} = \mathbb{P}[Z \leq Z_{obs}] = \mathbb{P}[Z \leq -1.00] = 0.1587.$$

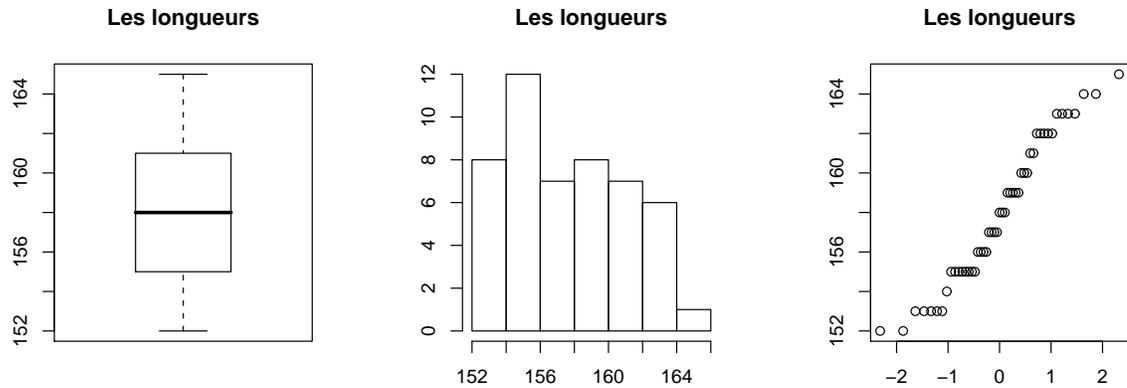
(iv) Avec $N \sim \text{binomiale}(49, 1/2)$, on obtient

$$p\text{-value} = \mathbb{P}[Z \leq 21] = \sum_{k=0}^{21} \binom{49}{k} (1/2)^k (1 - (1/2))^{49-k} = 0.1958.$$

SOLUTION DE LA PARTIE (b)

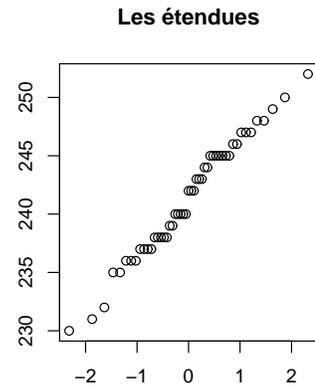
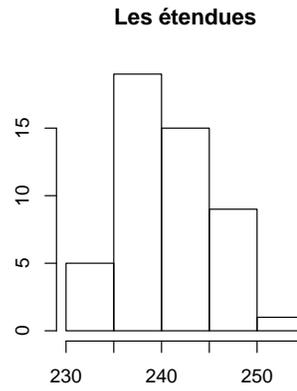
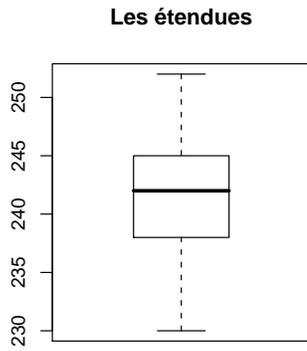
Pour chacune des variables, je fais le diagramme en boîte, l'histogramme, le graphe quantiles-quantiles et le test de Shapiro et Wilk. Voici les résultats.

Pour la variable longueur :



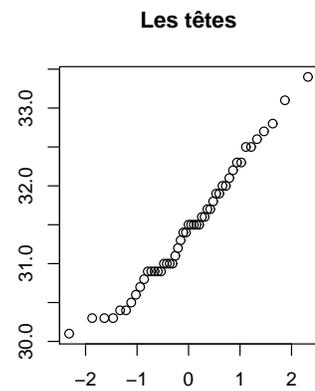
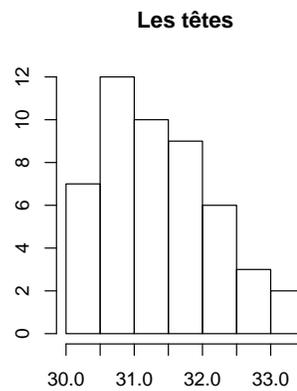
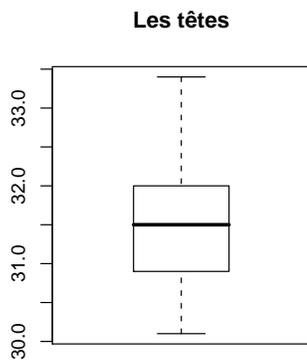
```
Shapiro-Wilk normality test
data : longueur
W = 0.9509, p-value = 0.04014
```

Pour la variable etendue :



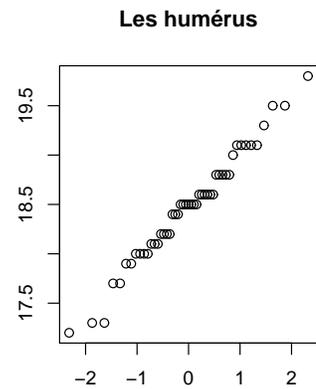
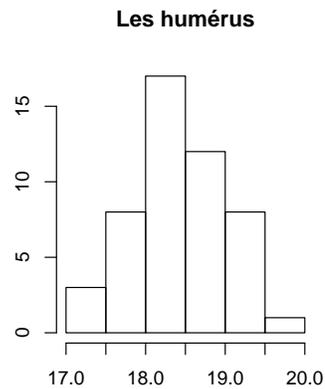
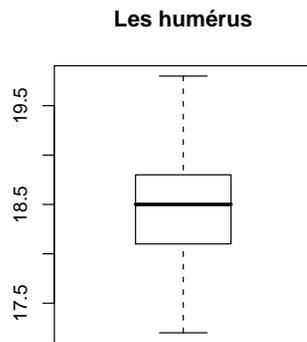
Shapiro-Wilk normality test
 data : etendue
 $W = 0.9789$, p-value = 0.5192

Pour la variable tete :



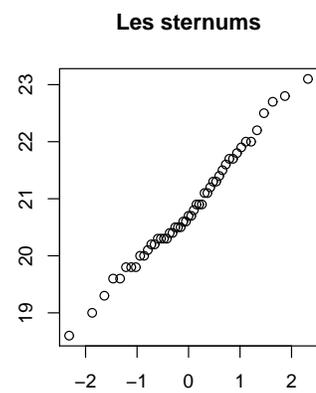
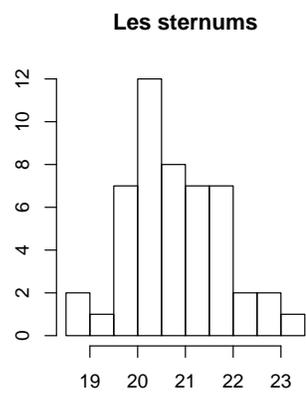
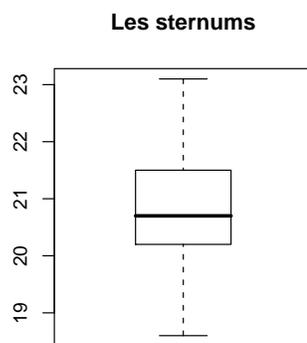
Shapiro-Wilk normality test
 data : tete
 $W = 0.9738$, p-value = 0.3391

Pour la variable humerus :



Shapiro-Wilk normality test
 data : humerus
 W = 0.9810, p-value = 0.6067

Pour la variable sternum :



Shapiro-Wilk normality test
 data : sternum
 W = 0.9868, p-value = 0.8533

CONCLUSIONS : La loi normale est un bon modèle pour les variables *etendue*, *tete*, *humerus* et *sternum*. La loi normale n'est pas un bon modèle pour la variable *longueur*.

SOLUTION DE LA PARTIE (c)

J'utilise la formule

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

avec $n = 49$, $\alpha = 0.05$, $\bar{x} = 241.33$ et $s = 5.068$. J'obtiens l'intervalle (239.87, 242.78).

SOLUTION DE LA PARTIE (d)

J'utilise l'intervalle

$$\left(\sqrt{\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}} s \right)$$

avec $n = 49$, $\alpha = 0.05$ et $s = 5.068$. J'obtiens l'intervalle (4.23, 6.33).

SOLUTION DE LA PARTIE (e)

C'est un test bilatéral sur la moyenne d'une loi normale. Le seuil α n'est pas spécifié. Nous examinerons le cas $\alpha = 0.01$ et le cas $\alpha = 0.05$. On utilise la règle de décision suivante :

$$\text{On rejette } H_0 \text{ si } |T| \geq t_{n-1, \alpha/2}$$

avec

$$T = \frac{\bar{X} - \mu_o}{S/\sqrt{n}}.$$

Ici on a $\mu_o = 31$ et $n = 49$. On obtient $\bar{x} = 31.459$ et $s = 0.795$. La valeur observée de notre statistique de test est donc

$$T_{obs} = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} = \frac{31.459 - 31}{0.795/\sqrt{49}} = 4.044.$$

Avec seuil $\alpha = 0.05$, on a $t_{n-1, \alpha/2} = t_{48, 0.025} = 2.01$.

Avec seuil $\alpha = 0.01$, on a $t_{n-1, \alpha/2} = t_{48, 0.005} = 2.68$.

Dans les deux cas on rejette H_0 . En fait, le *p-value* est très petit :

$$\begin{aligned} p\text{-value} &= \mathbb{P}_{H_0}[|T| \geq 4.044] \\ &= 2 \times \mathbb{P}_{H_0}[T \geq 4.044] \\ &= 2 \times \text{la surface à droite de 4.044 sous la } t_{48} \\ &= 2 \times (0.000095) \\ &= 0.00019. \end{aligned}$$

SOLUTION DE LA PARTIE (f)

On peut réécrire les hypothèses sous la forme suivante : $H_0 : \sigma^2 = 1/4$ vs $H_1 : \sigma^2 > 1/4$. Il s'agit donc d'un test unilatéral (à droite) sur la variance d'une loi normale. Le seuil α n'est pas spécifié. Nous examinerons le cas $\alpha = 0.01$ et le cas $\alpha = 0.05$. On utilise la règle de décision suivante :

$$\text{On rejette } H_0 \text{ si } U \geq \chi_{n-1, \alpha}^2$$

avec

$$U = \frac{(n-1)S^2}{\sigma_o^2}.$$

Ici on a $\sigma_o^2 = 1/4$ et $n = 49$. On obtient $s^2 = 0.3184$. La valeur observée de notre statistique de test est

$$U_{obs} = \frac{(49 - 1)s^2}{\sigma_o^2} = \frac{48 \times (0.3184)}{0.25} = 61.136.$$

Avec seuil $\alpha = 0.05$, on a $\chi_{n-1,\alpha}^2 = \chi_{48,0.05}^2 = 65.17$.

Avec seuil $\alpha = 0.01$, on a $\chi_{n-1,\alpha}^2 = \chi_{48,0.025}^2 = 73.68$.

Dans les deux cas, il n'y a pas lieu de rejeter H_0 . En fait, le p -value est grand :

$$\begin{aligned} p\text{-value} &= \mathbb{P}_{H_0}[U \geq 61.136] \\ &= \text{la surface à droite de } 61.136 \text{ sous la } \chi_{48}^2 \\ &= 0.0965. \end{aligned}$$

NUMÉRO 8. On considère un champ récemment ensemencé. On suppose que la loi normale est un bon modèle pour décrire la hauteur des plants dans ce champ. Certains plants sont malades. On obtient un échantillon aléatoire de 36 plants. Parmi ces 36 plants, 5 sont malades et 31 sont en santé. Voici les 36 hauteurs, en cm :

68.6	66.8	66.7	65.8	68.4	55.7	59.0	62.3	60.1
75.6	67.0	74.1	61.3	66.6	57.5	64.4	72.1	65.0
53.1	58.7	63.2	66.9	64.8	57.6	68.9	60.7	64.7
63.5	62.4	59.6	64.3	67.9	68.9	72.1	62.4	65.5

Ces données sont disponibles sur le site web du cours. Il s'agit du fichier `Plants.xls`.

- (a) Obtenez une estimation de la hauteur moyenne des plants de ce champ.
- (b) Calculez l'erreur type associée à l'estimation obtenue à la partie (a).
- (c) Obtenez un intervalle de confiance de niveau 95% pour la hauteur moyenne des plants de ce champ.
- (d) Obtenez une estimation pour l'écart-type des hauteurs des plants dans ce champ.
- (e) Obtenez un intervalle de confiance de niveau 95% pour l'écart-type des hauteurs des plants dans ce champ.
- (f) Obtenez une estimation pour la proportion des plants malades dans ce champ.
- (g) Calculez l'erreur type associée à l'estimation obtenue à la partie (f).
- (h) Obtenez un intervalle de confiance de niveau 95% pour la proportion de plants malades dans ce champ.
- (i) À la partie (h), quelle taille d'échantillon aurait-il fallu utiliser pour être sûr d'obtenir un intervalle de longueur au plus 0.04 ?
- (j) Dessinez l'histogramme des 36 hauteurs. Si vous le faites à la main, choisissez des classes raisonnables.
- (k) Dessinez le diagramme en boîte des 36 hauteurs.

SOLUTION. Voir la solution clip sur le site web du cours.

NUMÉRO 9.

- (a) Obtenez le quantile d'ordre 0.90 de la loi normale standard.
- (b) Obtenez le quantile d'ordre 0.90 de la loi normale de moyenne 50 et d'écart-type 6.
- (c) Obtenez le quantile d'ordre 0.90 de la loi de Student avec 20 degrés de liberté.
- (d) Obtenez le quantile d'ordre 0.90 de la loi du khi-deux avec 20 degrés de liberté.
- (e) Pour chacune des 4 questions ci-dessus, dessinez avec soin le graphe de la distribution et indiquez sur ce graphe la quantile que vous avez obtenu.

Remarque : « Quantile d'ordre 0.90 » et « 90^e centile » sont deux expressions qui veulent dire la même chose.

SOLUTION.

- (a) 1.2816
- (b) 57.6893
- (c) 1.3253
- (d) 28.4120
- (e) Semblable aux parties (h), (i) et (j) du numéro 1.

NUMÉRO 10. Un certain génotype est supposé être présent chez 20% des plants d'une certaine population. On soupçonne que la proportion réelle, disons p , est supérieure à cette valeur théorique de 20%. On veut tester $H_0 : p = 0.20$ contre $H_1 : p > 0.20$. On obtient un échantillon aléatoire de 25 plants et on détermine le nombre N de plants qui ont le génotype en question. On utilise la règle de décision suivante :

On rejette H_0 si $N \geq 10$.

- (a) Quel est le seuil exact de cette règle de décision ?
- (b) On réalise l'expérience et on obtient $N_{\text{obs}} = 13$. Quel est le p -value ?
- (c) Au seuil 1%, est-ce qu'on accepte ou est-ce qu'on rejette H_0 ?
- (d) Complétez la phrase suivante : Si l'hypothèse nulle avait été vraie, on se serait attendu à ce que N soit environ -----, plus ou moins environ -----.

SOLUTION.

- (a) Seuil = $\mathbb{P}_{H_0}[N \geq 10] = \sum_{k=10}^{25} \binom{25}{k} (0.20)^k (0.80)^{25-k} = 0.01733$.
- (b) p -value = $\mathbb{P}_{H_0}[N \geq 13] = \sum_{k=13}^{25} \binom{25}{k} (0.20)^k (0.80)^{25-k} = 0.00037$.
- (c) Au seuil 1%, on rejette H_0 .
- (d) Si l'hypothèse nulle avait été vraie, on se serait attendu à ce que N soit environ $np_o = 25 \times 0.2 = 5$, plus ou moins environ $\sqrt{np_o(1-p_o)} = \sqrt{25(0.2)(0.8)} = 2$.

NUMÉRO 11. On suppose que la loi normale $N(\mu, \sigma^2)$ est un bon modèle pour décrire la distribution de la quantité d'eau absorbée par un plant de tomate durant la première semaine du mois d'août (sous des conditions expérimentales contrôlées). On veut tester $H_0 : \mu = 11$ contre $H_1 : \mu \neq 11$, au seuil 5%. On réalise l'expérience suivante : à l'aide d'un dispositif complexe, on mesure la quantité d'eau absorbée par n plants choisis au hasard. À partir des résultats de cette expérience, on doit prendre une décision : accepter ou rejeter H_0 .

(a) Énoncez clairement la règle de décision au seuil 5%.

On mesure les quantités d'eau absorbée par 15 plants. Voici les résultats :

10.98 10.86 8.93 8.48 9.65 9.84 11.19 12.31 12.63 9.09 8.24 9.99 7.35 10.80 9.35

(b) Calculez la valeur observée de votre statistique de test.

(c) Au seuil 5%, est-ce qu'on accepte ou est-ce qu'on rejette H_0 ?

(d) À l'aide de la table appropriée, que peut-on dire du p -value ?

(e) Refaites les parties (b), (c) et (d) avec R-Commander.

(f) La loi normale est-elle un bon modèle ? Avec R-Commander, examinez l'histogramme, le boxplot et le graphe quantile-quantile. Puis faites le test de Shapiro et Wilk.

SOLUTION.

(a) On rejette H_0 si $\left| \frac{\bar{X}-11}{S/\sqrt{n}} \right| \geq t_{n-1,0.025}$

(b) Ici on a $n = 15$ et on obtient $\bar{x} = 9.9793$ et $s = 1.4875$. La valeur observée de notre statistique de test est donc $T_{obs} = \frac{9.9793-11}{1.4875/\sqrt{15}} = -2.6575$.

(c) La valeur critique est $t_{14,0.025} = 2.145$. On rejette H_0 .

(d) Le p -value est $\mathbb{P}_{H_0}[|T| \geq 2.6575] = 2 \times \mathbb{P}_{H_0}[T \geq 2.6575] = 0.01875$. (J'ai utilisé R).

(f) Les graphiques ne mettent aucunement en doute l'hypothèse de normalité. Le test de Shapiro et Wilk donne un très grand p -value (0.9655). Il n'y a pas lieu de rejeter l'hypothèse de normalité.

NUMÉRO 12. Voici un échantillon de taille 17 obtenu à partir d'une certaine population :

18.22 20.83 26.88 21.42 19.36 35.67 20.40 20.39 23.89
23.21 18.43 21.57 22.91 24.54 21.62 31.33 17.70

Est-ce que la loi normale est un bon modèle pour cette population ? Avec R-Commander, on fait le test de Shapiro et Wilk. On obtient les résultats suivants :

$$W_{obs} = 0.8465 \quad \text{et} \quad p\text{-value} = 0.009487.$$

Le p -value étant plus petit que 1%, on rejette l'hypothèse de normalité.

- (a) On examine l'histogramme, le boxplot et le graphe quantile-quantile. On constate la présence de deux observations aberrantes. Après un examen attentif, on constate que les observations 31.33 et 35.67 ont été mal enregistrées! Les bonnes valeurs sont 21.33 et 25.67. Si on refait notre test de Shapiro et Wilk avec 31.33 et 35.67 remplacées par 21.33 et 25.67, la valeur de notre statistique de Shapiro et Wilk sera-t-elle plus petite ou plus grande que 0.8465? Notre *p-value* sera-t-il plus petit ou plus grand que 0.009487? Répondez à ces deux questions sans utiliser R-Commander.
- (b) Avec R-Commander, analysez ces données, d'abord avec 31.33 et 35.67, puis avec 21.33 et 25.67. Examinez l'histogramme, le boxplot, le graphe quantile-quantile. Faites le test de Shapiro et Wilk.

SOLUTION. Voir la solution clip sur le site web du cours.

NUMÉRO 13. Un biologiste doit réaliser une expérience de grande envergure dans le but de comparer plusieurs variétés de blé. Pour son expérience, il utilise des milliers de pots de terres noires riches en divers nutriments. Pour les fins de cette étude, il est important que la concentration de potassium ne varie peu d'un pot à l'autre. Le biologiste a déterminé que l'écart-type ne doit pas dépasser la valeur 3.0. La compagnie qui lui vend ces pots de terre l'assure que l'écart-type des teneurs en potassium ne dépasse pas 3.0. On obtient un échantillon aléatoire de 21 pots et on mesure la teneur en potassium pour chacun de ces pots. Voici les résultats :

51.6	48.8	56.9	57.2	53.3	55.8	54.7
51.7	60.2	51.2	56.9	54.9	56.3	55.9
50.3	52.1	56.7	62.1	62.0	52.8	54.7

Un simple calcul nous donne l'écart-type échantillonnal $s = 3.59$. C'est plus grand que la valeur 3.0. Le biologiste devrait-il s'inquiéter?

- (a) Faites un test approprié. Quelles sont les hypothèses H_0 et H_1 ?
- (b) Quelle est la règle de décision au seuil 5%?
- (c) Avec les données ci-dessus, quel est le *p-value*?
- (d) La procédure utilisée ci-dessus est valide à condition que la loi normale soit un bon modèle pour décrire la distribution des teneurs en potassium. L'hypothèse de normalité vous semble-t-elle raisonnable? Expliquez.

SOLUTION.

- (a) $H_0 : \sigma^2 = 9$ et $H_1 : \sigma^2 > 9$.
- (b) On rejette H_0 si $(n - 1)S^2/9 \geq \chi_{n-1,0.05}^2$.
- (c) La valeur observée de notre statistique de test est $20 \times (3.59)^2/9 = 28.64$. La valeur critique est $\chi_{n-1,0.05}^2 = \chi_{20,0.05}^2 = 31.41$. Au seuil 5%, il n'y a pas lieu de rejeter H_0 . Le *p-value* est $\mathbb{P}_{H_0}[20S^2/9 \geq 28.64] = 0.095$.
- (d) Les graphiques (histogramme, diagramme en boîte, graphe quantile-quantile) sont cohérents avec l'hypothèse de normalité. Le test de Shapiro et Wilk donne un *p-value* de 0.5063. Il n'y a donc pas lieu de rejeter l'hypothèse de normalité.