MAT 2779, Introduction à la biostatistique

Devoir 3

Écheance : le vendredi 18 novembre avant 15h

Total=100 points

Partie I) *Total*=80 points [20 points par question]

Exercice 10.4 : On calcul la moyenne de l'échantillon: $\bar{x} = \sum x_i/n = 216{,}625$. (Puisqu'on suppose une population normale avec un écart type connu $\sigma = 28$, alors notre IC sera basé sur z). Un intervalle de confiance à 96% pour μ est

$$\bar{x} \pm z_{0{,}02} \frac{\sigma}{\sqrt{n}} = [196{,}29; 236{,}97].$$

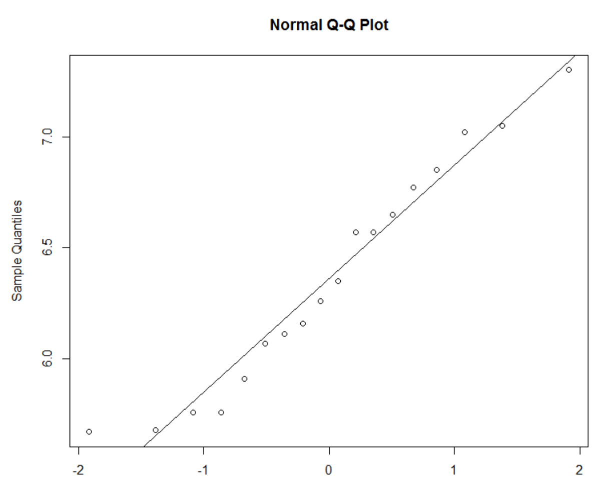
où $z_{0{,}02} = 2{,}055$. (N.B. $z_{0{,}02}$ est une valeur tel que $P(Z > z_{0{,}02}) = 0{,}02$. Alors, on veut une valeur $z_{0{,}02}$ tel que $\Phi(z_{0{,}02}) = 0{,}98$.)

Exercice 10.7 :

(a) On a utilisé les commandes suivantes pour construire le diagramme quantile-quantile pour vérifier la normalité du pH.

```
x=c(6.11,6.26,5.67,5.76,7.3,5.68,6.57,6.57,6.07,
5.76,5.91,6.16,7.02,6.35,6.77,6.65,7.05,6.85)
qqnorm(x)
abline(mean(x),sd(x))
```

Le diagramme quantile-quantile est ci-bas. La tendance est linéaire avec un petit écart à l'extrémité gauche. C'est raisonable de supposer que le pH est normalement distribué.



(b) Avec les $n = 18$ observations, on calcul la moyenne $\overline{x} = 6,3617$ et l'écart type $s = 0{,}5107$.

Un intervalle de confiance à 95% pour le pH moyen est

$$\bar{x} \pm t_{0{,}025;17} \frac{s}{\sqrt{n}} = 6{,}3617 \pm 2{,}110 \, \frac{0{,}5107}{\sqrt{18}} = [6{,}11; 6.62],$$

où $t_{0{,}025;17} = 2,110$. (N.B. Nous supposons que la population est normale et σ est inconnu, alors notre IC est basé sur t.)

Exercice 10.8 : On peut calculé une estimation pour le taux d'efficacité p médicament : $\hat{p} = 119/170 = 0{,}7$.

(a) Un intervalle de confiance à 95% pour le taux d'efficacité est

$$\hat{p} \pm z_{0{,}025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.631; 0.769],$$

où $z_{0{,}025} = 1{,}96$, $n = 170$, et $\hat{p} = 0{,}7$.

(b) Pour avoir un plus grand pourcentage des échantillons qui auront un IC qui contient le taux d'efficacité de la population, on doit avoir des intervalles qui sont plus longs. Alors, un IC à 98% sera plus long.

(c) Un intervalle de confiance à 98% pour le taux d'efficacité est

$$\hat{p} \pm 2.326 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.618; 0.782].$$

où $z_{0{,}01} = 2{,}326$, $n = 170$, et $\hat{p} = 0{,}7$.

Exercice 16.16 :

(a) La moyenne de léchantillon est $\bar{x} = 0{,}844$ est une estimation de la moyenne de la population. La moyenne géométrique de léchantillon $g = e^{\bar{y}} = e^{-1{,}3193} = 0{,}2673$ est une estimation de la moyenne géométrique de la population.

(b) Un intervalle de confiance à 95% pour la concentration moyenne est

$$\bar{x} \pm z_{0{,}025} \frac{s}{\sqrt{n}} = 0.8441 \pm 1.96 \frac{2.1014}{\sqrt{524}} = [0{,}664; 1{,}024],$$

où $z_{0{,}025} = 1{,}96$. (N.B. Ici ce n'est pas raisonnable de supposer que la population est normale. Mais, nous avons un échantillon de grande taille, alors nous pouvons utiliser un IC basé sur z.)

(c) Un intervalle de confiance à 95% pour la moyenne du logarithme de la concentration d'arsenic est

$$\bar{y} \pm z_{0,025}\,\frac{s_y}{\sqrt{n}} = -1{,}3193 \pm 1{,}96\frac{1{,}52}{\sqrt{524}} = [-1{,}44945; -1{,}18915].$$

Alors, un intervalle de confiance à 95% pour la moyenne géométrique de l'arsenic est

$$[e^{-1.44945}; e^{-1.18915}] = [0{,}235; 0{,}304].$$

(d) L'histogramme de la concentration a une forte asymétrie positive. C'est difficile de le voir dans l'histogramme, mais il est fort probable que la moyenne est plus grande qu'une grande proportion des concentrations. Alors, la moyenne n'est probablement pas une bonne description d'une concentration typique. Cependant, l'histogramme du logarithme de la concentration est approximativement symétrique, alors la moyenne des logs devrait être proche de la médiane des logs. Alors, en prenant l'exponentielle, la moyenne géométrique de la concentration sera proche de la concentration médiane. Alors, la moyenne géométrique sera une meilleure description d'une concentration typique en comparaison à la moyenne.

Partie II) *Total*=20 points

Résoudre la question suivante en utilisant une calculatrice TI-30, TI-34, Casio FX-260 ou Casio FX-300.

Question 1

(a) On veut

$$\begin{aligned} P(X > 9) &= P\left(\frac{X-\mu}{\sigma} > \frac{9-8{,}3}{1{,}7}\right) \\ &= P\left(Z > 0{,}41\right) \\ &= 1 - \Phi(0{,}41) \\ &= 1 - 0{,}6591 = 0{,}3409 = 34{,}09\%. \end{aligned}$$

(b) Soit $\overline{X}$ la masse moyenne (en grammes) d'un échantillon de $n = 10$ souris. Puisque la population est normale, alors $\overline{X}$ suit une loi normale de moyenne $\mu_{\overline{X}} = 8{,}3$ et d'écart type $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1,7/\sqrt{10}$. Alors,

$$Z = \frac{\overline{X} - 8{,}3}{1,7/\sqrt{10}} \text{ suit une loi N(0,1).}$$

On veut

$$\begin{aligned} P(7{,}5 < \overline{X} < 8{,}3) &= P\left(\frac{7{,}5-8{,}3}{1{,}7/\sqrt{10}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{10}} < \frac{8{,}9-8{,}3}{1{,}7/\sqrt{10}}\right) \\ &= P\left(-1{,}49 < Z < 1{,}12\right) \\ &= \Phi(1{,}12) - \Phi(-1{,}49) \\ &= 0.8686 - 0.0681 = 0.8005 = 80.05\%. \end{aligned}$$

(c) Soit $\overline{X}$ la masse moyenne (en grammes) d'un échantillon de $n = 50$ souris. Puisque la population est normale, alors $\overline{X}$ suit une loi normale de moyenne $\mu_{\overline{X}} = 8{,}3$ et d'écart type $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1,7/\sqrt{50}$. Alors,

$$Z = \frac{\overline{X} - 8{,}3}{1,7/\sqrt{50}} \text{ suit une loi N(0,1).}$$

On veut

$$P(\overline{X} < 7{,}5) = \Phi\left(\frac{7{,}5 - 8{,}3}{1{,}7/\sqrt{50}}\right) = \Phi(-3{,}33) = 0{,}0004.$$

(d) Puisque la population n'est pas normale, alors il est possible que probabilité de la partie (b) ne soit pas une bonne approximation puisque $n = 10$ est une petite taille d'échantillon. Mais, en (c), la taille de l'échantillon est grande (c.-à-d. $n > 30$). Alors, il est fort probable que la probabilité calculée en (c) soit une bonne approximation.