Statistiques 3 Corrigé des exercices

Exercice 2.1

Soit une variable Z distribuée selon une loi normale centrée-réduite. Calculer la probabilité pour la variable Z de se trouver dans chacun des cas suivants :

1. Entre z = 0 et z = 1.2.

$$P(0 < Z < 1.2) = P(Z < 1.2) - P(Z < 0)$$

= 0.8849 - 0.5
= 0.3849

2. Entre z = -0.68 et z = 0.

$$P(-0.68 < Z < 0) = P(Z < 0) - P(Z < -0.68)$$
$$= P(Z < 0) - (1 - P(Z < 0.68))$$
$$= 0.5 - (1 - 0.7517) = 0.2517$$

3. Entre z = -0.46 et z = 2.21.

$$P(-0.46 < Z < 2.21) = P(Z < 2.21) - P(Z < -0.46)$$
$$= P(Z < 2.21) - (1 - P(Z < 0.46))$$
$$= 0.9864 - (1 - 0.6772) = 0.6636$$

4. Entre z = 0.81 et z = 1.94.

$$P(0.81 < Z < 1.94) = P(Z < 1.94) - P(Z < 0.81)$$

= 0.9738 - 0.7910 = 0.1828

5. A gauche de z = -0.6.

$$P(Z < -0.6) = 1 - P(Z < 0.6)$$

= 1 - 0.7257 = 0.2743

6. A droite de z = -1.28.

$$P(Z > -1.28) = P(Z < 1.28)$$

= 0.8997

7. A droite de z=2.05 et à gauche de z=-1.44.

$$P(2.05 < Z < -1.44) = 0$$

C'est impossible car 2.05 > -1.44.

8. A droite de z = 2.05 ou à gauche de z = -1.44.

$$P(Z > 2.05) + P(Z < -1.44) = (1 - P(Z < 2.05) + (1 - P(Z < 1.44)))$$

= $(1 - 0.9798) + (1 - 0.9251) = 0.0951$

Exercice 2.2

Soit Z, une variable aléatoire suivant une loi normale centrée et réduite N(0,1).

- 1. Déterminer la probabilité des événements suivants :
 - (a) Z < 2.14

$$P(Z < 2.14) = 0.9838$$

(b) Z < -2.14

$$P(Z < -2.14) = 1 - P(Z < 2.14)) = 1 - 0.9838 = 0.0162$$

(c) $Z \in [-2.14 ; 2.14]$

$$P(-2.14 < Z < 2.14) = 2(P(Z < 2.14) - 0.5)$$

= $2P(Z < 2.14) - 1$
= $2 \cdot 0.9838 - 1 = 0.9676$

2. Pour quel seuil a a-t-on P(Z < a) = p lorsque p prend les valeurs suivantes :

(a)
$$p = 0.8686$$

$$a = 1.12$$

(b)
$$p = 0.9719$$

$$a = 1.91$$

(c)
$$p = 0.2912$$

$$P(Z < a) = 0.2912$$

 $P(Z < -a) = 1 - 0.2912 = 0.7088$
 $-a = 0.55$
 $a = -0.55$

Soit X, une variable aléatoire suivant une loi normale N(2,1).

- 3. Déterminer les probabilités suivantes :
 - (a) P(X < 3.98)

$$P(X < 3.98) = P(Z < \frac{3.98 - 2}{1})$$

= $P(Z < 1.98) = 0.9761$

(b)
$$P(X < 1.5)$$

$$P(X < 1.5) = P(Z < \frac{1.5 - 2}{1})$$

$$= P(Z < -0.5)$$

$$= 1 - P(Z < 0.5)$$

$$= 1 - 0.6915 = 0.3085$$

(c)
$$P(X \in [2; 3])$$

$$\begin{split} P(2 < X < 3) &= P(X < 3) - P(X < 2) \\ &= P(Z < \frac{3-2}{1}) - P(Z < \frac{2-2}{1}) \\ &= P(Z < 1) - P(Z < 0) \\ &= 0.8413 - 0.5 = 0.3413 \end{split}$$

4. Pour quelle valeur de a a-t-on :

(a)
$$P(X < a) = 0.7454$$

$$P(X < a) = 0.7454$$

$$P(Z < \frac{a-2}{1}) = 0.7454$$

$$a-2 = 0.66$$

$$a = 2.66$$

(b)
$$P(-a+2 < X < a+2) = 0.7458$$

$$P(-a+2 < X < a+2) = 0.7458$$

$$P(X < a+2) - P(X < -a+2) = 0.7458$$

$$P(Z < \frac{(a+2)-2}{1}) - P(Z < \frac{(-a+2)-2}{1}) = 0.7458$$

$$P(Z < a) - P(Z < -a) = 0.7458$$

$$P(Z < a) - (1 - P(Z < a)) = 0.7458$$

$$2P(Z < a) - 1 = 0.7458$$

$$P(Z < a) = 0.8729$$

$$a = 1.14$$

Soit Y, une variable aléatoire suivant une loi normale N(2,9).

- 5. Déterminer les probabilités suivantes :
 - (a) P(Y < 4.25)

$$P(Y < 4.25) = P(Z < \frac{4.25 - 2}{3})$$

= $P(Z < 0.75) = 0.7734$

(b) P(Y < 1.5)

$$P(Y < 1.5) = P(Z < \frac{1.5 - 2}{3})$$

$$\approx P(Z < -0.17)$$

$$= 1 - P(Z < 0.17)$$

$$= 1 - 0.5675 = 0.4325$$

(c)
$$P(|Y| < 1)$$

$$P(Y < 1) - P(Y < -1) = P(Z < \frac{1-2}{3}) - \frac{-1-2}{3})$$

$$\approx P(Z < -0.33) - P(Z < -1)$$

$$= (1 - P(Z < 0.33) - (1 - P(Z < 1)))$$

$$= P(Z < 1) - P(Z < 0.33)$$

$$= 0.8413 - 0.6293 = 0.212$$

- 6. Pour quelle valeur de a a-t-on :
 - (a) P(Y < a) = 0.64

$$P(Y < a) = 0.64$$

$$P(Z < \frac{a-2}{3}) = 0.64$$

$$\frac{a-2}{3} \approx 0.36$$

$$a = 3 \cdot 0.36 + 2 = 3.08$$

(b)
$$P(a < Y < 2.5) = 0.3788$$

$$P(Y < 2.5) - P(Y < a) = 0.3788$$

$$P(Z < \frac{2.5 - 2}{3}) - P(Z < \frac{a - 2}{3}) = 0.3788$$

$$P(Z < 0.17) - P(Z < \frac{a - 2}{3}) \approx 0.3788$$

$$0.5675 - P(Z < \frac{a - 2}{3}) \approx 0.3788$$

$$P(Z < \frac{a - 2}{3}) \approx 0.5675 - 0.3788$$

$$P(Z < \frac{a - 2}{3}) \approx 0.1887$$

$$1 - P(Z < \frac{a - 2}{3}) \approx 1 - 0.1887$$

$$P(Z < -\frac{a - 2}{3}) \approx 0.8113$$

$$-\frac{a - 2}{3} \approx 0.88$$

$$a \approx 2 - 3 \cdot 0.88 = -0.64$$

(c)
$$P(-a < Y < a) = 0.9$$

$$P(-a < Y < a) = 0.9$$

$$P(Y < a) - P(Y < -a) = 0.9$$

$$P(Z < \frac{a-2}{3}) - P(Z < -\frac{-a-2}{3}) = 0.9$$

Pas de solution direct car il n'est pas possible d'utiliser la symétrie. En faisant varier la valeur de a il est possible de trouver la solution : $a\approx 5.917$

Exercice 2.3

1.

2.

$$P(X > 36) = P(Z > \frac{36 - 30}{6})$$

$$= P(Z > 1)$$

$$= 1 - P(Z < 1)$$

$$= 1 - 0.8413 = 0.1587$$

3.

$$P(Y < 30) = P(Z < \frac{30 - 36}{12})$$

$$= P(Z < -0.5)$$

$$= 1 - P(Z < 0.5)$$

$$= 1 - 0.6915 = 0.3085$$

Exercice 2.4

1. Soit la variable X suivant une loi normale N(10,4).

$$P(8 < X < 13) = P(\frac{8-10}{2} < Z < \frac{13-10}{2})$$

$$= P(Z < 1.5) - P(Z < -1)$$

$$= P(Z < 1.5) - (1 - P(Z < 1))$$

$$= 0.9332 - (1 - 0.8413) = 0.7745$$

2. Soit la variable Q_{10} suivant une loi du chi-2 à 10 degrés de liberté.

$$P(Q_{10} < 12.55) = 0.75$$

 $P(Q_{10} > 18.31) = 1 - 0.95 = 0.05$
 $P(Q_{10} < 18.31) = 0.95$

$$P(12.55 < Q_{10} < 18.31) = P(Q_{10} < 18.31) - P(Q_{10} < 12.55)0.95 - 0.75 = 0.2$$

Trouver la valeur a telle que $P(Q_{10} < a) = 0.975$.

$$a = 20.48$$

3. Soit la variable T_8 suivant une loi de Student à 8 degrés de liberté.

$$P(T_8 < 0.546) = 0.7$$

$$P(T_8 < -0.546) = 1 - P(T_8 < 0.546) = 1 - 0.7 = 0.3$$

$$P(-0.546 < T_8 < 1.86) = P(T_8 < 1.86) - P(T_8 < -0.546)$$

$$= P(T_8 < 1.86) - (1 - P(T_8 < -0.546))$$

$$= 0.95 - (1 - 0.7) = 0.65$$

$$P(T_8 > b) = 0.6$$

 $P(T_8 < -b) = 0.6$
 $-b = 0.262$
 $b = -0.262$

4.

$$P(F_{8,5} < c) = 0.95$$

 $c = 4.82$

5.

$$P(F_{5,8} < d) = 0.95$$

 $d = 3.69$

6.

$$P(F_{12,3} < e) = 0.99$$

 $e = 27.05$

7.

$$P(F_{3,12} < f) = 0.99$$

 $f = 5.95$

Exercice 2.5

- 1. Loi du Chi-2
- 2. Loi de Poisson
- 3. Loi Binomiale
- 4. Loi Normale

Exercice 3.1

1.

$$\dot{Y} = 295$$

$$\overline{Y} = \frac{1475}{5} = 295$$

2.

$$\dot{Y} = 295$$

$$\overline{Y} = \frac{2668}{9} \approx 296.4$$

3.

$$\hat{\sigma}^2 = \frac{n}{n-1}s^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right)$$

a) cinq observations

$$s^{2} = \frac{435625}{5} - 295^{2} = 100$$
$$\hat{\sigma}^{2} = \frac{5}{5 - 1}100 = 125$$
$$\hat{\sigma} = \sqrt{125} \approx 11.18$$

Pour la moyenne de l'échantillon, nous avons

$$\hat{\sigma}_{\bar{y}_1} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{11.18}{\sqrt{5}} = 5$$

b) neuf observations

$$s^{2} = \frac{792274}{9} - 296.4^{2} \approx 151.14$$
$$\hat{\sigma}^{2} = \frac{9}{9-1}151.14 \approx 170.03$$
$$\hat{\sigma} \approx \sqrt{170.03} \approx 13.04$$

Pour la moyenne de l'échantillon, nous avons

$$\hat{\sigma}_{\bar{y}_2} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{13.04}{\sqrt{9}} \approx 4.35$$

4.

$$\dot{Y} = 295$$

$$\overline{Y} = \frac{2416}{9} \approx 268.4$$

La médiane est ici plus pertinente que la moyenne comme estimateur de μ car elle est plus robuste.

Exercice 3.4

a) Intervalle de confiance pour μ avec σ^2 inconnu :

$$\mu = \bar{x} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \hat{\sigma}_{\bar{X}}$$

La moyenne se calcule ainsi

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

Dans le cas où les données sont groupées par classes, cette formule s'écrit

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{c} n_i \cdot x_i$$

où n_i représente la fréquence de la classe i et c le nombre de classes.

Nous avons donc

$$\overline{x} = \frac{230 \cdot 1 + 248 \cdot 2 + \dots + 3 \cdot 6}{688} = \frac{1469}{688} \approx 2.1352$$

La variance de l'estimateur \bar{X} se calcule de la manière suivante

$$\hat{\sigma}_{\overline{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

et l'estimation de la variance de la population se calcule ainsi

$$\hat{\sigma} = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{\sqrt{n-1}}$$

Dans le cas où les données sont groupées par classes, cette formule s'écrit

$$\hat{\sigma} = \frac{\sqrt{\sum_{i=1}^{c} n_i \cdot (x_i - \overline{x})^2}}{\sqrt{n-1}}$$

Nous avons donc

$$\hat{\sigma} = \frac{\sqrt{230 \cdot (1 - 2.1352)^2 + 248 \cdot (2 - 2.1352)^2 + \dots + 3 \cdot (6 - 2.1352)^2}}{\sqrt{687}} \approx 1.8075$$

et par conséquent

$$\hat{\sigma}_{\overline{x}} = \frac{1.8075}{\sqrt{688}} \approx 0.0415$$

Pour $\alpha=5\%$ et n=688, nous avons

$$t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(687)} \approx z_{0.975} \approx 1.96$$

Nous obtenons alors

$$\mu = \bar{x} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \hat{\sigma}_{\bar{X}} = 2.1352 \pm 0.0415 \cdot 1.96 = 2.1352 \pm 0.0813$$

En d'autres termes, l'intervalle de confiance pour $\alpha=5\%$ est donné par [2.0539;2.2165].

b) Intervalle de confiance pour σ^2 avec μ inconnu :

$$\left[\frac{\sum (x_i - \overline{x})^2}{q_{1-\frac{\alpha}{2}}^{(n-1)}}; \frac{\sum (x_i - \overline{x})^2}{q_{\frac{\alpha}{2}}^{(n-1)}}\right]$$

n = 688,

n - 1 = 687,

 $1 - \frac{\alpha}{2} = 0.975$

 $\frac{\alpha}{2} = 0.025.$

Avec la fonction EXCEL KHIDEUX.INVERSE nous pouvons trouver que

$$q_{0.975}^{(687)} = 761.53$$
 $q_{0.025}^{(687)} = 616.26$

Notez que cette fonction d'EXCEL définit la probabilité comme le risque de première espèce.

Pour des données groupées par classes, la formule

$$\sum_{i=1}^{c} (x_i - \overline{x})^2$$

s'écrit

$$\sum_{i=1}^{c} n_i \cdot (x_i - \overline{x})^2$$

Nous avons donc

$$\sum_{i=1}^{c} n_i \cdot (x_i - \overline{x})^2 = 230 \cdot (1 - 2.1352)^2 + 248 \cdot (2 - 2.1352)^2 + \dots + 3 \cdot (6 - 2.1352)^2 \approx 812.43$$

L'intervalle de confiance pour la variance est donné par

$$\left[\frac{812.43}{761.53} \; ; \; \frac{812.43}{616.26}\right] = [1.07 \; ; \; 1.32]$$

Exercice 4.6

C'est un cas particulier du test de la moyenne avec variance connue et un échantillon de taille 1.

$$\sigma_0 = \sigma_1 = 20$$

$$\mu_0 = 100$$

$$\mu_1 = 80$$

Nous choisissons la statistique

$$z_0 = \frac{x - \mu_0}{\sigma_0} = \sim N(0, 1)$$

1. $\alpha = 0.05$

$$x_c = \mu_0 + z_\alpha \cdot \sigma_0$$

$$= \mu_0 + z_{0.05} \cdot \sigma_0$$

$$= 100 + (-1.645) \cdot 20$$

$$= 67.1$$

2. $\alpha = 0.01$

$$x_c = \mu_0 + z_\alpha \cdot \sigma_0$$

$$= \mu_0 + z_{0.01} \cdot \sigma_0$$

$$= 100 + (-2.33) \cdot 20$$

$$= 53.4$$

3. $\alpha = 0.05$

Le risque de seconde espèce est donné par

$$\beta = P(x_1 > x_c \mid x_1 \sim N(\mu_1 = 80, \sigma_1^2 = 20))$$

$$= P(z_1 > \frac{x_c - 80}{20} \mid z_1 \sim N(1, 1))$$

$$= P(z_1 > \frac{67.1 - 80}{20})$$

$$= P(z_1 > -0.645)$$

$$\approx 0.74$$

4. $\alpha = 0.01$

$$\beta = P(x_1 > x_c \mid x_1 \sim N(\mu_1 = 80, \sigma_1^2 = 20))$$

$$= P(z_1 > \frac{x_c - 80}{20} \mid z_1 \sim N(1, 1))$$

$$= P(z_1 > \frac{53.4 - 80}{20})$$

$$= P(z_1 > -1.33)$$

$$= 0.9082$$

Exercice 6.3

Le test de la somme des rangs de Wilcoxon vérifie l'égalité des médianes de deux populations dans le cas de données indépendantes :

$$H_0$$
: $med(X) = med(Y)$
 H_1 : $med(X) \neq med(Y)$

1. Classer les observation par ordre croissant.

	Bâtiment		Automobile	
	rang	salaire	rang	salaire
1	1.5	12		
2			1.5	12
3	3 5.5	15		
4	5.5	16		
5			5.5	16
6			5.5	16
7			5.5	16

2. Calculer w, la somme des rangs du plus petit échantillon ordonné par ordre croissant.

$$w = 1.5 + 3 + 5.5 = 10$$

3. Calculer w_i , la somme des rangs du plus petit échantillon ordonné par ordre décroissant.

$$w_i = n_X (n_X + n_Y + 1) - w = 3(3 + 4 + 1) - 10 = 14$$

4. Calculer w_{\min} le minimum entre w et w_i .

$$w_{\min} = 10$$

5. Trouver le seuil de rejet dans la table.

Pour $n_x=3$ et \mathbf{n}_y =5, le seuil de rejet pour $\alpha=5\%$ est de 6. Par conséquent, pour $n_x=3$ et $n_y=4$, le seuil de rejet pour $\alpha=5\%$ est inférieur à 6.

6. Déterminer le résultat du test :

Accepter H_0 car w_{\min} est strictement supérieur au seuil de rejet. Statistiquement, il n'y a pas de différence significative à 5% entre les salaires horaires des deux branches.

Exercice 4.3

Question 1.

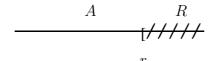
1.

$$H_0: \mu = \mu_0 = 50$$
 contre $H_1: \mu = \mu_1 > 50$

- 2. $\alpha = 0.05$
- 3. Test de la moyenne avec variance connue :

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma_{\bar{X}}} \sim N(0, 1)$$

- 4. Déterminer la région critique :
 - Test unilatéral à droite :



- Seuil de rejet :

$$r = z_{1-\alpha} = z_{0.95} = 1.645$$

5. Valeur observée de la statistique :

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.87 - 50}{\frac{5}{\sqrt{12}}} = 2.68$$

6. 2.68 > 1.645

L'hypothèse H_0 est rejetée.

Question 2.

4. Dans ce cas particulier, le seuil de rejet ne dépend pas de la taille de l'échantillon :

$$r = z_{1-\alpha} = z_{0.95} = 1.645$$

5. La valeur observée de la statistique dépend de la taille de l'échantillon :

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.87 - 50}{\frac{5}{\sqrt{100}}} = 7.74$$

6. 7.74 > 1.645

L'hypothèse H_0 est rejetée.

Exercice 4.4

Question 1.

1.

$$H_0: \mu=\mu_0=27$$
 contre
$$H_1: \mu=\mu_1>27$$

- 2. $\alpha = 0.05$
- 3. Test de la moyenne avec variance inconnue :

$$T_0 = \frac{\overline{X} - \mu_0}{\hat{\sigma}_{\bar{X}}} \sim St_{n-1}$$

4. Déterminer la région critique :

- Test unilatéral à droite.
- Seuil de rejet :

$$r = t_{0.95}^{(50)} = 1.676$$

5. Valeur observée de la statistique :

$$\hat{\sigma}_{\overline{x}} = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{\sqrt{n} \cdot \sqrt{n-1}} = \frac{\sqrt{816}}{\sqrt{51} \cdot \sqrt{50}} = 0.5657$$

$$t_0 = \frac{\overline{x} - \mu_0}{\hat{\sigma}_{\overline{x}}} \frac{28 - 27}{0.5657} = 1.768$$

6. 1.768 > 1.676

L'hypothèse \mathcal{H}_0 est rejetée. En d'autres termes, l'âge moyen a augmenté.

Question 2.

3. Test de la moyenne avec variance connue :

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma_{\bar{X}}} \sim N(0, 1)$$

4. Seuil de rejet :

$$r = z_{0.95} = 1.645$$

5. Valeur observée de la statistique :

$$\sigma_{\overline{x}} = \frac{\sigma_0}{\sqrt{n}} = \frac{5}{\sqrt{51}} = 0.7$$

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma_{\bar{x}}} \frac{28 - 27}{0.7} = 1.43$$

6. 1.43 < 1.645

L'hypothèse H_0 est acceptée. En d'autres termes, l'âge moyen n'a pas augmenté.

Question 3.

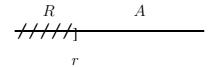
1.

$$H_0: \sigma = \sigma_0 = 5$$
 contre
$$H_1: \sigma = \sigma_1 < 5$$

- 2. $\alpha = 0.05$
- 3. Test de la variance avec moyenne inconnue :

$$Q_0 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma_0^2} \sim \chi_{n-1}^2$$

- 4. Déterminer la région critique :
 - Test unilatéral à gauche :



- Seuil de rejet :

$$r = q_{\alpha}^{(n-1)} = q_{0.05}^{(50)} = 34.76$$

5. Valeur observée de la statistique :

$$q_0 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} = \frac{816}{5^2} = 32.64$$

6. 32.64 < 34.76

L'hypothèse H_0 est rejetée. En d'autres termes l'écart-type est plus petit que 5.

Par conséquent, le résultat de la Question 1 est plus correct que celui de la Question 2.