

# Correction Examen de Statistiques L2

Clément Narteau

29 janvier 2008

## Correction Exercice 1

1.

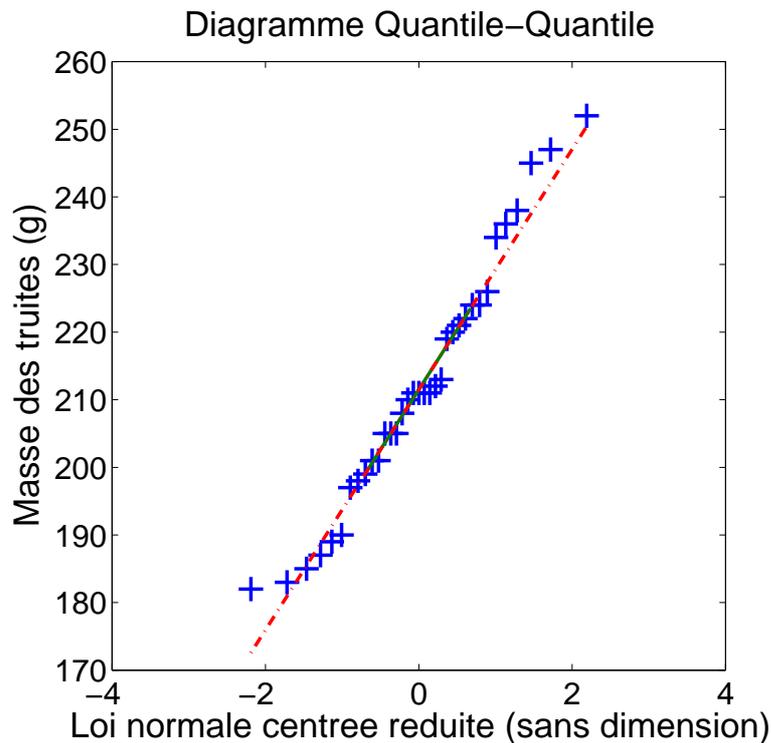
$$\bar{X} = 212.06 \text{ g}, \quad s^2 = 339.41 \text{ g}^2, \quad s = 18.42 \text{ g}.$$

La médiane est égale à 211 g

2.

$$Q_{20} = 197 \text{ g}, \quad Q_{40} = 205 \text{ g}, \quad Q_{60} = 212 \text{ g}, \quad Q_{80} = 224 \text{ g}.$$

3. Il suffit de classer dans l'ordre les données de masse et de les tracer en fonction des valeurs de  $F(X)$  que l'on peut lire dans le tableau. Cela donne la figure ci-dessous, à partir de laquelle on peut conclure positivement sur la normalité des données (i.e. la distribution des masses des truites suit une loi normale).



4. On calcule l'intervalle de confiance de  $\mu$

$$\bar{X} - z_c \frac{s}{\sqrt{N}} < \mu < \bar{X} + z_c \frac{s}{\sqrt{N}}$$

avec  $z_c = 1.96$  au risque de 5% et  $z_c = 2.33$  au risque de 1%. On obtient

$$205.95 < \mu_{95} < 218.16 \quad \text{et} \quad 204.04 < \mu_{99} < 220.08$$

## Correction Exercice 2

(a) Lorsque les nombres 1 et 2 sortent, le nouveau-né obtient le bac avec la mention passable. Donc

$$p_1 = \frac{2}{10} = \frac{1}{5}$$

(b) Lorsque les nombres 5 et 10 sortent, le nouveau-né n'obtient pas le bac. Donc la probabilité d'être recalé est

$$p = \frac{8}{10} = \frac{4}{5}$$

Les lois de probabilité de  $X$  et  $Y$  sont donc des lois binômiales de probabilité  $p$  avec des nombres de tirages qui sont respectivement  $n_X = 3$  et  $n_Y = 4$ .

$$f(X) = \mathcal{B}(n_X, p) = \mathcal{B}\left(3, \frac{4}{5}\right), \quad f(Y) = \mathcal{B}(n_Y, p) = \mathcal{B}\left(4, \frac{4}{5}\right).$$

(c) C'est la probabilité d'avoir 0 garçon et  $\{1, 2, 3\}$  filles, plus la probabilité d'avoir 1 garçon et  $\{2, 3\}$  filles, plus la probabilité d'avoir 2 garçons et  $\{3\}$  filles.

$$P = (C_4^0(1-p)^4) \times (1 - C_3^0(1-p)^3) + (C_4^1p(1-p)^3) \times (C_3^2p^2(1-p) + C_3^3p^3) + (C_4^2p^2(1-p)^2) \times (C_3^3p^3)$$

$$P \approx 0.106$$

(d) Pour dépasser l'objectif, il faut plus de  $0.8 \times 7$  bacheliers, c'est à dire au moins six bacheliers. Donc,

$$P = C_7^6p^6(1-p) + p^7 \approx 0.58$$

## Correction Exercice 3

On sait que le pourcentage de la population votant pour le candidat A peut être estimé à partir du pourcentage de l'échantillon grâce à la formule

$$\mu_p \in \left[ p - z_c \sqrt{\frac{p(1-p)}{N}}; p + z_c \sqrt{\frac{p(1-p)}{N}} \right]$$

Pour être élu, il faut que  $\mu_p > 0.5$ . On peut donc en déduire une valeur de  $N$  à partir de laquelle un candidat est sûr d'être élu dans 95% ( $z_c = 1.64$ ) ou 99% ( $z_c = 2.33$ ) des cas :

$$N > \frac{z_c^2 p(1-p)}{(p - \mu_p)^2}$$

(a) Pour  $p = 0.533$ , il faut au minimum 619 personnes (879 si on prend  $z_c = 1.96$ ).

(b) Pour  $p = 0.533$ , il faut au minimum 1237 personnes (1517 si on prend  $z_c = 2.57$ ).

(c) Il faut tout simplement trouver les racines de

$$f(p) = (p - \mu_p) - z_c \sqrt{\frac{p(1-p)}{N}}$$

avec  $\mu_p = 0.5$ ,  $N = 1350$  et  $z_c = 2.33$ . On a donc

$$\begin{aligned} f(p) &= 0 \\ (p - \mu_p)^2 &= z_c^2 \frac{p(1-p)}{N} \\ p^2 \left(1 + \frac{z_c^2}{N}\right) - p \left(2\mu_p + \frac{z_c^2}{N}\right) + \mu_p^2 &= 0 \end{aligned}$$

On calcule le déterminant

$$\Delta = \left(1 + \frac{z_c^2}{N}\right)^2 - 4\mu_p^2 \left(1 + \frac{z_c^2}{N}\right) \approx 0.004$$

et la racine

$$p_1 = \frac{(2\mu_p + \frac{z_c^2}{N}) + \sqrt{\Delta}}{2(1 + \frac{z_c^2}{N})} \approx 0.53159$$

Il faut donc au minimum

$$N_1 = p_1 \times N = 718 \text{ personnes}$$

exprimant une opinion favorable au candidat A.

Cela donne une proportion finale de

$$p = \frac{N_1}{N} = 0.53185$$

## Correction Exercice 4

1. Les hypothèses sont

$H_0$  : La femelle coucou adapte la taille de ses oeufs à la taille des oeufs présents dans les nids dans lesquels elle pond.

$H_1$  : La femelle coucou n'adapte pas la taille de ses oeufs à la taille des oeufs présents dans les nids dans lesquels elle pond.

Il faut utiliser le test de Student sur la différence de moyenne car les échantillons sont petits ( $N < 30$ ). On calcule pour cela

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad \text{avec} \quad \sigma = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}$$

Si  $t > t_c$  alors on refuse  $H_0$ .

2. On a  $N_1 = 12 \text{ mm}$ ,  $N_2 = 10 \text{ mm}$  oeufs,  $s_1^2 = 2.94 \text{ mm}$ ,  $s_2^2 = 0.93 \text{ mm}$ ,  $\bar{X}_1 = 28.49 \text{ mm}$ ,  $\bar{X}_2 = 29.99 \text{ mm}$ .  
Donc

$$\sigma = 1.49 \text{ mm},$$

et

$$t = 2.35.$$

Les valeurs seuils aux risques de 5% et de 1% sont de 2.09 et de 2.85. On rejette  $H_0$  au risque de 5%. On accepte  $H_0$  au risque de 1%. On peut donc conclure à un seuil de confiance significatif (i.e. 95%) que la femelle coucou adapte la taille de ses oeufs à la taille des oeufs présents dans les nids dans lesquels elle pond.

## Correction Exercice 5

Pour un échantillons de 64 exploitations, la distribution d'échantillonnage des moyennes suit une loi normale de moyenne  $\bar{X} = \mu = 1050 \text{ m}^3$  et d'écart type

$$s = \frac{\sigma}{\sqrt{N}} = \frac{180}{\sqrt{64}} = 22.5 \text{ m}^3.$$

Afin de trouver les proportions d'exploitations dépassant certains seuils  $X_c$ , il faut réduire les écarts

$$z(X_c) = \frac{\bar{X} - X_c}{s},$$

et lire dans la table à quelles proportions de la loi normale centrée réduite cela correspond.

1.

$$z(X_c = 1100) = 2.22, \quad P(\bar{X} > 1100) = 0.0131.$$

2.

$$z(X_c = 990) = -2.67, \quad P(\bar{X} < 990) = 0.0038.$$

3. Pour définir les classes de consommation, il faut déterminer les valeurs seuils pour les quantiles [0.2; 0.4; 0.6; 0.8] de la loi normale centrée réduite. On trouve des valeurs  $z = [-0.842; -0.253; 0.253; 0.842]$ . En utilisant la formule

$$X = z\sigma + \mu$$

on trouve les limites de consommation

$$[898.5; 1004.4; 1095.66; 1201.5]$$

4a. La taxe sera de  $\frac{53 \times 1000}{100} = 530$  euros

4b. Les 53 exploitants les plus économes en eau représentent  $p_1 = \frac{53}{677} = 7.9\%$  des exploitants. Les 100 exploitants les moins économes en eau représentent  $p_2 = \frac{100}{677} = 8.53\%$  des exploitants. On trouve dans les tables de la loi normale centrée réduite à quelles limites correspondent ces pourcentages. On trouve  $z_1 = -1.41$  et  $z_2 = 1.05$ . On trouve les limites de consommation grâce à la formule

$$X = z\sigma + \mu$$

et on trouve des consommations de  $795.9 \text{ m}^3$  et  $1238.9 \text{ m}^3$ .

## Correction Exercice 6

1. La moyenne théorique d'un dé à huit face est

$$\mu_8 = \sum_{i=1}^8 i \times p_i$$

avec  $p_i = \frac{1}{8}$ . On obtient  $\mu_8 = 4.5$ .

2. La moyenne observée est

$$\mu = \sum_{i=1}^8 i \times \frac{N_i}{N}$$

avec  $N = 96$ . On obtient  $\mu = 4.74$ .

3. Les hypothèses sont

$H_0$  : Le dé n'est pas pipé.

$H_1$  : Le dé est pipé.

On va faire un test du  $\chi^2$  sur la variabilité des résultats.

On calcule

$$\chi^2 = \sum_{j=1}^8 \frac{(o_j - e_j)^2}{e_j}$$

avec  $e_j = 96/8 = 12$ . On trouve

$$\chi^2 = 5.5.$$

Pour une loi du  $\chi^2$  avec 7 degrés de liberté, les valeurs seuil aux risques de 5% et 1% sont respectivement de 14.1 et 18.5. On peut donc conclure de manière très significative que le dé n'est pas pipé.

## Correction Exercice 7

Pour les population A et B individuellement, on applique directement la formule sur l'estimation de la moyenne d'une population à partir de celle de l'échantillon :

$$P\left(\bar{X} - z_c \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} + z_c \frac{\sigma}{\sqrt{N}}\right) = \Phi(z_c) - \Phi(-z_c),$$

avec  $z_c = 1.96$  et  $z_c = 2.57$  aux risques de 5% et 1%.

1.

$$P(163.3 \leq \mu_A \leq 172.7) = 0.05.$$

$$P(161.8 \leq \mu_A \leq 174.2) = 0.01.$$

2.

$$P(176.5 \leq \mu_B \leq 181.5) = 0.05.$$

$$P(175.7 \leq \mu_B \leq 182.3) = 0.01.$$

3. Pour les différences entre les population A et B, on applique la formule sur les intervalles de confiance des différences :

$$P(\bar{X}_1 - \bar{X}_2 - z_c \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} < \mu_{A-B} < \bar{X}_1 - \bar{X}_2 + z_c \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}),$$

avec  $z_c = 1.96$  et  $z_c = 2.57$  aux risques de 5% et 1%.

On trouve

$$P(-16.3 \leq \mu_{A-B} \leq 5.7) = 0.05.$$

$$P(-18 \leq \mu_{A-B} \leq -4) = 0.01.$$

On conclut que les éléments de B sont très significativement plus grands que les éléments de A.