Examen de biostatistique pour les étudiantEs en pharmacie

Janvier 2010

Figure 1

1. La processionnaire du pin (*Thaumetopea pityocampa*) est un insecte de l'ordre des lépidoptères. Lorsqu'elles cherchent de la nourriture, les processionnaires se déplacent en file indienne (v. Figure 1), la cohésion de la file en déplacement étant assurée par le contact tactile de soie à soie.

Si l'on dispose une file de processionnaires sur le rebord d'un vase, on peut appondre la première chenille à la dernière. De cette façon, les chenilles tournent en rond jusqu'à ce que l'une d'elles décide de quitter la file, s'estimant mal guidée vu l'absence manifeste de nourriture.

Utilisons le modèle suivant: le temps T pendant lequel une chenille marche sans quitter la file suit une loi normale de moyenne égale à 5 heures et de variance égale à 2 heures^2. Avec une file de 20 chenilles, quelle est la probabilité que les insectes marchent en cercle pendant au moins 3 heures? (On supposera l'indépendance des temps de désertion des 20 chenilles.)

2. Retrouvons nos 20 processionnaires quelque temps plus tard, sous forme de papillons. Toutes sont des femelles. Leur vie sous cette forme est brève et intense: en l'espace de quelques heures, elles vont naître, connaître l'amour, pondre et mourir.

Chaque femelle pond une quantité d'oeufs dont la distribution a une moyenne de 200 oeufs et un écart-type de 50 oeufs.

En supposant que chaque papillon est indépendant des autres, quelle est la probabilité que le nombre total d'oeufs pondus par les 20 femelles soit compris entre 3800 et 4000? (Utiliser l'approximation normale.)

3. Les processionnaires ont des poils très urticants.

Rodolphe élève des processionnaires. Par ailleurs, c'est un grand amateur de panse de brebis farcie au anchois (PBFA) (v. Figure 2). A tel point qu'il se demande si en nourrissant des processionnaires avec de la PBFA, elles atténueront leurs propriétés urticantes. Il divise donc ses insectes en deux groupes et nourrit le premier avec de la PBFA et l'autre avec des aiguilles de pin pendant trois mois. Ensuite, il prend au hasard 10 insectes dans chaque groupe et compte les poils urticants sur chacun d'eux. Il obtient les résultats suivants (en milliers de poils):

Figure 2

PBFA (x_i):	123	156	171	148	160	155	139	164	140	153
Aiguilles de pin (y_i):	177	169	140	177	168	156	157	161	188	166

En supposant que les nombres de poils urticants dans chacun des groupes sont des variables normales, indépendantes et de même variance, tester au niveau 5% s'il y a une différence entre le nombre moyen de poils urticants dans le premier et le deuxième groupe.

Indications:
- $\hat{\sigma}_x^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 194.77$ ($\bar{x}$ est la moyenne des x_i).
- $\hat{\sigma}_y^2 = \frac{1}{9}\sum_{i=1}^{10}(y_i - \bar{y})^2 = 180.10$ ($\bar{y}$ est la moyenne des y_i).

Solution de l'examen de biostatistique pour les étudiantEs en pharmacie de janvier 2010

1. Soit T_i le temps pendant lequel la chenille i continue à tourner. On a

$$T_i \sim \mathcal{N}(5, 2).$$

Pour que le cercle se maintienne, il faut que toutes les chenilles restent. Soit T_c le temps pendant lequel le cercle se maintient. T_c est le temps de la chenille qui quitte le cercle en premier, donc le plus petit des T_i :

$$T_c = \min_i (T_i).$$

On a donc :

$$\begin{aligned} P(T_c \geq 3) &= P(\min_i (T_i) \geq 3) \\ &= P\big((T_1 \geq 3) \cap (T_2 \geq 3) \cap \ldots \cap (T_{20} \geq 3)\big), \end{aligned}$$

et grâce à l'indépendance des T_i on obtient

$$P(T_c \geq 3) = P(T_1 \geq 3) \times P(T_2 \geq 3) \times \ldots \times P(T_{20} \geq 3).$$

Pour chaque chenille i, on a

$$\begin{aligned} P(T_i \geq 3) &= P\left(\frac{T_i - 5}{\sqrt{2}} \geq \frac{3-5}{\sqrt{2}}\right) \\ &= P\left(Z \geq \sqrt{2}\right), \end{aligned}$$

où $Z = \frac{T_i - 5}{\sqrt{2}} \sim \mathcal{N}(0, 1)$. A l'aide de la table, on trouve

$$P\left(Z \geq \sqrt{2}\right) \approx 0.9207.$$

On trouve donc $P(T_c \geq 3) \approx (0.9207)^{20} \approx 0.1916$.

2. Soit N_i le nombre d'oeufs pondus par la femelle i. Pour chaque i, on a

$$\mu(N_i) = 200 \text{ et } \sigma(N_i) = 50$$

et donc

$$\sigma^2(N_i) = 2500.$$

Soit $T = \sum_{i=1}^{20} N_i$ le nombre total d'oeufs. Les N_i étant indépendants, les espérances et les variances s'additionnent et donc

$$\begin{array}{lll} \mu(T) & = 20 \times 200 & = 4000 \quad \text{et} \\ \sigma^2(T) & = 20 \times 2500 & = 50000. \end{array}$$

On a alors

$$P(3800 \le T \le 4000) = P\left(\frac{3800 - 4000}{\sqrt{50000}} \le \frac{T - 4000}{\sqrt{50000}} \le \frac{4000 - 4000}{\sqrt{50000}}\right).$$

Approximation normale :

$$Z = \frac{T - 4000}{\sqrt{50000}} \sim \mathcal{N}(0, 1).$$

Avec la table, on obtient alors

$$P\left(\frac{-200}{\sqrt{50000}} \le Z \le 0\right) \approx P(-0.89 \le Z \le 0) \approx 0.3133.$$

(Détail :

$$\begin{aligned} P(-0.89 \le Z \le 0) &= P(Z \le 0) - P(Z \le -0.89) \\ &= P(Z \le 0) - (1 - P(Z \le 0.89)) \\ &\approx 0.5 - 1 + 0.8133 \\ &= 0.3133) \end{aligned}$$

3. On fait un t-test non apparié car on compare deux populations, avec deux échantillons et une seule mesure par individu.
Soient

- $\mu(X)$ = moyenne du nombre de poils dans la population nourrie à la PBFA
- $\mu(Y)$ = moyenne du nombre de poils dans la population nourrie aux aiguilles de pin

Hypothèses :

- $H_0 : \mu(X) = \mu(Y)$
- $H_1 : \mu(X) \neq \mu(Y)$

On calcule les quantités nécessaires à l'évaluation de la statistique de test :

- $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 150.9$
- $\bar{y} = \frac{1}{10} \sum_{i=1}^{10} y_i = 165.9$
- $\hat{\sigma}_d = \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \sqrt{\frac{(n_x-1)\hat{\sigma}_x^2 + (n_y-1)\hat{\sigma}_y^2}{n_x+n_y-2}} \approx 6.1227$
  ($n_x = n_y = 10$, $\hat{\sigma}_x^2 = 194.77$ et $\hat{\sigma}_y^2 = 180.10$; voir donnée)

On calcule alors la statistique

$$t = \frac{\bar{x} - \bar{y}}{\hat{\sigma}_d} \approx -2.4499.$$

On la compare à $t_{\alpha/2, n_x+n_y-2} = t_{0.025,18} = -2,101$ (table) ($\alpha = 5\%$ est le seuil du test).
Comme $t < t_{\alpha/2, n_x+n_y-2}$, l'hypothèse d'égalité des moyennes dans la population est rejetée.