Série 11

Exercice 1. Un contrôle de qualité est effectué sur la production de pailles d'une usine. Les pailles produites dans cette usine sont vendues dans des paquets de 100 pièces, et on voudrait estimer le pourcentage de pièces défectueuses dans un paquet.

On prélève 18 paquets aléatoirement et on compte le nombre de pièces défectueuses dans chaque paquet. Soit Y_i le nombre de pièces défectueuses dans le $i^{\text{ième}}$ paquet. On a observé les résultats suivants :

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}
0	1	1	1	1	1	0	1	2	1	0	1	1	2	0	0	0	2

(a) Proposez un modèle statistique pour résoudre ce problème d'estimation.

(b) Donnez l'estimateur du maximum de vraisemblance pour le pourcentage cherché.

(c) Donnez un intervalle de confiance approximatif à 95 % pour le pourcentage cherché, basé sur l'approximation de la distribution de l'estimateur du maximum de vraisemblance pour un échantillon de grande taille.

Exercice 2. Soient $X_1, \dots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$ pour $\lambda > 0$, et $X_{(1)} = \min(X_1, \dots, X_n)$. Montrer que $\lambda n X_{(1)} \sim \text{Exp}(1)$. Expliquer pourquoi il s'agit d'un pivot et donner un intervalle de confiance de niveau $1-\alpha$ pour λ basé sur ce pivot.

Exercice 3. Soient $Y_1, \dots, Y_n$ des observations indépendantes et identiquement distribuées selon une distribution dont la densité est

$$f_\theta(y) = \frac{\theta^3}{2} y^2 \exp\{-\theta y\}, \theta > 0, y \geq 0.$$

(i). Trouver l'estimateur du maximum de vraisemblance $\widehat{\theta}_{\text{ML}}$ de θ.

(ii). Trouver l'information observée et l'utiliser afin de construire un intervalle de confiance approximatif au niveau de 95% pour θ.

Exercice 4. Rappelez-vous l'exercice 5 de la série 10. Une entreprise d'automobiles s'intéresse à la consommation d'un nouveau modèle de voiture. Pour 12 voitures de ce modèle, on mesure la consommation d'essence, en litres, pour parcourir 100 km, et on obtient les résultats suivants :

$$14.60, 11.21, 11.56, 11.37, 13.68, 15.07, 11.06, 16.58, 13.37, 15.98, 12.07, 13.22.$$

La moyenne et la variance empirique de l'échantillon sont $\bar{x}_{12} = 13.31$ et $s_{12}^2 = \frac{1}{11}\sum_{i=1}^{12}(x_i - \bar{x})^2 = 3.69$, respectivement.

En supposant un modèle normal pour les données, on veut tester l'hypothèse nulle que la consommation moyenne est égale à 12.2 litres contre l'hypothèse alternative que la consommation n'est pas égale à 12.2 litres.

(a) Ecrivez le modèle ainsi que les hypothèses nulle et alternative d'une manière formelle.

(b) Quelle statistique de test pouvez-vous utiliser ?

(c) Quelles valeurs de la statistique de test considérez-vous comme étant "extrêmes" ?

(d) Testez à un seuil (ou niveau) de signification de 5 %.

(e) Testez à un seuil de signification de 10 %. Commentez sur une éventuelle différence.

(f) Au vu des résultats des parties (d) et (e) de cet exercice et de l'exercice 1 de la série précédente, voyez-vous une correspondance entre tests et intervalles de confiance ?

(g) Calculez la valeur p_{obs}. *Si on n'a pas couvert la définition de p_{obs} en classe, on peut reporter cette partie et la suivante à la semaine prochaine.*

(h) Refaites les parties (d) et (e) en utilisant l'approche basée sur p_{obs}. Tenant compte des résultats des parties (d) et (e), expliquez à quoi la valeur p_{obs} correspond. *En anglais on l'appelle "p−value", et on l'utilise plus souvent que les valeurs critiques.*

(i) La consommation de l'ancien modèle de voiture était de 15 litres. Peut-on dire que le nouveau modèle de voiture a une consomation différente de l'ancien ? Si oui, recommanderiez vous le remplacement de l'ancien modèle de voiture ? Ecrivez le modèle statistique, les hypothèses nulle et alternative, et testez à un seuil de signification de 5 %.

Exercice 5. Nous considérons deux groupes de personnes âgées. Le premier groupe est constitué de personnes ayant un faible risque de chute (groupe 1) et le second est constitué de personnes ayant un risque élevé de chute (groupe 2). Une analyse de mobilité a été effectuée en mesurant le temps, en secondes, nécessaire pour passer d'une position assise à une position debout. On a obtenu les résultats suivants :

Groupe 1	2.94	3.34	2.96	3.14	2.42	2.92
Groupe 2	3.11	3.27	4.16	3.85	4.99	3.51

Cette analyse indique-t-elle une différence significative entre les deux groupes ?
Si on note x_i les résultats obtenus pour le groupe 1 et y_i les résultats obtenus pour le groupe 2, on a $\bar{x}_6 = 2.95$, $\bar{y}_6 = 3.82$, $s_x^2 = 0.094$, $s_y^2 = 0.478$, et $s_{xy}^2 = \frac{1}{5}\sum_{i=1}^{6}(x_i - \bar{x})(y_i - \bar{y}) = -0.162$.
Indication : supposez que $X_i \sim \mathrm{N}(\mu_X, \sigma^2)$ et $Y_i \sim \mathrm{N}(\mu_Y, \sigma^2)$, et testez $H_0 : \mu_X - \mu_Y = 0$ contre $H_1 : \mu_X - \mu_Y \neq 0$. Afin de déduire une statistique de test, utilisez le fait suivant.

Soient $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_m$ deux échantillons tels que $X_i \sim \mathcal{N}(\mu_X, \sigma^2)$ et $Y_j \sim \mathcal{N}(\mu_Y, \sigma^2)$ pour $i \in \{1, \ldots, n\}$ et $j \in \{1, \ldots, m\}$, où les variables X_i sont indépendantes entre elles, les variables Y_j sont indépendantes entre elles, et les variables X_i sont indépendantes des variables Y_j. On a

$$\sqrt{n+m-2}\sqrt{\frac{nm}{n+m}}\,\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sim t_{n+m-2}.$$

Expliquer pourquoi cette indépendance est raisonnable dans ce cas.

Exercice 6. L'effet de deux soporifiques A et B est mesuré par le nombre d'heures de sommeil additionnel par rapport à la moyenne habituelle. Les soporifiques ont été administrés aux mêmes 10 personnes, à des temps suffisamment espacés pour que les effets du premier soporifique aient disparu une fois que le deuxième a été administré. On note x_i et y_i l'effet du soporifique A et B, respectivement, sur la i-ème personne.

Personne	1	2	3	4	5	6	7	8	9	10
Soporifique A	0.7	−1.6	−0.2	−1.2	−0.1	3.4	3.7	0.8	0	2
Soporifique B	1.9	−0.8	1.1	0.8	−0.1	4.4	5.5	1.1	4.6	3.4

Peut-on dire que l'effet du soporifique B a une moyenne de 3 heures de plus que celui du soporifique A ?
Si on note x_i les résultats obtenus pour le soporifique A, et y_i les résultats obtenus pour le soporifique B, on a $\bar{x}_{10} = 0.75$, $\bar{y}_{10} = 2.19$, $s_x^2 = 3.20$, $s_y^2 = 4.63$, et $s_{xy}^2 = 3.11$.
Indication : supposer que $X_i \sim \mathrm{N}(\mu_X, \sigma_X^2)$ et $Y_i \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Définir $Z_i = X_i - Y_i$ et supposer que $Z_i \sim \mathrm{N}(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}[X_i, Y_i])$ sont indépendantes. En travaillant directement avec les variables $Z_1, \ldots, Z_{10}$, on obtient un problème similaire à celui de l'exercice 1.