Série 11

17 mai 2017

Exercice 1. Rappelez-vous de l'exercice 2 de la série précédente. 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, 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 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 des parties (b) et (c) de l'exercice 2 de la série précédente, est-ce que vous voyez une correspondance entre tests et intervalles de confiance ?

(g) Calculez la valeur p_{obs}.

(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 beaucoup plus souvent que les valeurs critiques.

(i) Répondez à toutes les questions précédentes pour une alternative unilatérale (vous pouvez choisir la direction).

(j) La consommation de l'ancien modèle de voiture était de 14.5 litres. Est-ce qu'on peut dire que le nouveau modèle est plus efficace ? Ecrivez le modèle, les hypothèses nulle et alternative, et testez à un seuil de signification de 5 %.

En ce moment, les tests statistiques et les intervalles de confiance peuvent vous paraître comme étant des méthodes assez imparfaites. Et justement, car elles le sont ! Les statistiques ne peuvent pas donner des réponses 100 % *correctes, parce qu'on l'utilise quand on n'a pas* 100 % *des informations dont on a besoin. Chaque méthode statistique prend en considération la possibilité d'erreurs. C'est à vous de décider quelles erreurs sont plus graves que les autres. Avec des informations incomplètes vous ne pouvez pas complètement éliminer la possibilité d'erreurs . Mais c'est le rôle du statisticien de faire de son mieux en choisissant le modèle et les méthodes pour utiliser les informations disponibles d'une manière la plus efficace possible (par exemple donner un intervalle de confiance qui est court et, en même temps, dont la probabilité d'erreur grave est petite).*

Exercice 2. Pour estimer le résultat d'un possible référendum sur l'indépendance du Québec, un sondage d'opinion a été effectué. Parmi les 800 Québécois interrogés, 55 % se sont prononcés en faveur de l'indépendance du Québec. Ce sondage montre-t-il que la majorité de la population est favorable à l'indépendance ?

Exercice 3. 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 par x_i les résultats obtenus pour le groupe 1, et par 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 $X_1, \dots, X_n$ et $Y_1, \dots, Y_m$ deux échantillons tels que $X_i \sim \mathrm{N}(\mu_X, \sigma^2)$ et $Y_j \sim \mathrm{N}(\mu_Y, \sigma^2)$ pour $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, m\}$, où les variables X_i sont indépendantes entre elles, les variables Y_j sont indépendantes entre elles, et les variables X_i sont indépendantes avec les variables Y_j. On a

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

Exercice 4. L'effet de deux soporifiques A et B est mesuré par le nombre d'heures de sommeil additionnels par rapport à la moyenne habituelle. Les soporifiques ont été administrés aux mêmes 10 personnes, à des temps suffisamment espacés afin 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

Est-ce qu'on peut dire que l'effet du soporifique B a une moyenne de 3 heures de plus que celui du soporifique A ?
Si on note par x_i les résultats obtenus pour le soporifique A, et par 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 : supposez que $X_i \sim \mathrm{N}(\mu_X, \sigma_X^2)$ et $Y_i \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ et définissez $Z_i = X_i - Y_i \sim \mathrm{N}(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}[X_i, Y_i])$. En travaillant directement avec les variables $Z_1, \dots, Z_{10}$, vous obtenez un problème similaire à celui de la question 1.