

2 exercices de statistique à résoudre avec Rstudio

A enregistrer la réponse sous la forme de deux fichiers .Rmd et .pdf.

Données

Le fichier `employee.csv` est une base de données d'employés administratifs, retraitée d'après la base fournie par Andrew F. Siegel.¹ Ainsi le genre est codé 0 pour le sexe féminin et 1 pour le masculin. Deux formations de niveau progressif sont proposées de temps en temps sur la base du volontariat. La variable `training` compte le nombre de formations effectuées par l'employé considéré. La variable `experience` donne le nombre de ses années d'expérience au poste occupé. Le salaire annuel (\$) est enregistré dans la variable `salary`.

Exercice 1

Nous allons nous intéresser à la loi t de Student dans cet exercice.

- Vérifier que la taille de l'échantillon est de $n = 71$ observations.
- En notant $T_{d.d.l.}$ une variable aléatoire suivant une loi t de Student à d.d.l. degrés de liberté, calculer la probabilité $P(T_{69} > 0)$ ². Commenter.
- Trouver le point $q_{0.8}$ tel que $P(T_{69} < q_{0.8}) = 0.8$ (80% des observations sont en dessous).
- Trouver le point t_{α} tel que $P(|T_{69}| \geq t_{\alpha}) = \alpha$ où $\alpha = 5\%$.

¹*Practical Business Statistics*, 6th ed. (Elsevier, 2012).

²Indice : utiliser l'argument `lower.tail`.

Exercice 2

Considérez l'expérience comme variable explicative et le salaire comme variable expliquée.

- a. Calculer la moyenne de chacune de ces variables ainsi que leur écart-type respectif.
- b. Représenter par un graphique en nuage de points la relation entre ces deux variables et décrire la relation qui les relie.
- c. Calculer et interpréter leur coefficient de corrélation. Commenter par rapport au graphique en nuage de points.
- d. Donner l'équation de la droite de régression qui relie les deux variables et la tracer sur le graphique en nuage de points.
- e. Calculer l'erreur type S_{b_1} de la pente b_1 de la droite de régression.
- f. En déduire l'intervalle de confiance à 95% de la pente b_1 .
- g. Tester au seuil de 5% si la pente est significativement différente de 0. Interpréter le résultat.
- h. Quelle part de la variabilité des salaires peut-elle être expliquée par le fait que certains employés aient plus d'expérience que les autres ?
- i. À quel salaire annuel pourrait-on s'attendre d'un employé avec 8 années d'expérience ? Et d'un employé avec 3 années d'expérience ?
- j. Examiner les conditions de la régression en vous appuyant sur les résidus.
- k. Créer `employee_f` une base qui regroupe toutes les employées et `employee_m` le reste de l'échantillon. Calculer la pente de la régression dans chaque sous-groupe et commenter.