TP4 MST

Nicolas Brunel, Anastase Charantonis, Christian Kahindo

April 2019

Test de Student

Simuler un échantillon i.i.d $\mathcal{S}_n = (x_1, \ldots, x_n)$ de taille $n = 20$, et dont la loi commune est une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = 1$ et $\sigma^2 = 2$ (on rappelle que la fonction R pour simuler *rnorm*).

1. Nous voulons tester si la moyenne de l'échantillon μ est égale à $\mu_0 = 1$, ou plutôt égale à $\mu_1 = 1.5$. On suppose que la variance σ^2 est inconnue. Pour répondre à cette question, on va faire un test statistique avec un niveau de significativité $\alpha = 5\%$ (appelé encore risque de 1ère espèce).
   (a) Les hypothèses du test sont $H_0 : \mu = \mu_0$ et $H_1 : \mu = \mu_1$. Rappeler la définition de α et à quoi il correspond.
   (b) Donner la forme de la zone de rejet W, pour $\alpha = 5\%$ (on pourra utiliser le lemme de Neyman-Pearson, vu en cours).
   (c) Programmer la règle de décision associée $\delta(\mathcal{S}_n, \alpha, \mu_0, \mu_1)$ (écrire une fonction R paramétrée par les moyennes, α, et $\mathcal{S}_n$).
2. Simuler $N = 100$ échantillons $\mathcal{S}_n^1, \ldots, \mathcal{S}_n^N$ (toujours tel que $\mathcal{N}(\mu, \sigma^2)$, avec $\mu = 1$ et $\sigma^2 = 2$).
   (a) On rappellera la loi de la variable aléatoire $\delta(\mathcal{S}_n, \alpha, \mu_0, \mu_1)$. Appliquer la règle de décision du test de Student sur $S_n^i, i = 1, \ldots, 100$. Qu'observez vous? .
   (b) Faire varier $\alpha = 0.2, 0.1, 0.05, 0.01$: comment la zone de rejet est-elle modifiée ?
   (c) Pour $\alpha = 0.2, 0.1, 0.05, 0.01$, appliquer la règle de décision $\delta(\mathcal{S}_n^i, \alpha, \mu_0, \mu_1)$, $i = 1, \ldots, N$.
3. On va simuler $N = 100$ échantillons $\mathcal{S}_n^{'1}, \ldots, \mathcal{S}_n^{'N}$, mais qui suivent maintenant une loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu = 1.5$ et $\sigma^2 = 2$.
   (a) On rappellera la loi de la variable aléatoire $\delta(\mathcal{S}_n^{'i}, \alpha, \mu_0, \mu_1)$. Appliquer la règle de décision du test de Student sur $\mathcal{S}_n^{'i}, i = 1, \ldots, 100$. Qu'observez vous?

(b) Rappeler la définition et calculer théoriquement la puissance du test β, en fonction de α, μ_0, μ_1.

(c) On fixe $\alpha = 0.05$, et on fait varier l'hypothèse alternative $H_1 : \mu = \mu_1$. Simuler $N = 100$ échantillons $\mathcal{S}_n^{'1}, \ldots, \mathcal{S}_n^{'N}$ en faisant varier la moyenne $\mu = \mu_1 \in \{1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$ et appliquer la règle de décision $\delta(\mathcal{S}_n^{'i}, \alpha, \mu_0, \mu_1)$, $i = 1, \ldots, N$. Tracer en fonction de μ_1 le pourcentage de bonne décision et comparer avec les résultats de la question précédente.

4. On va utiliser la fonction R *t.test* qui permet de faire le test d'une hypothèse simple $H_0 : \mu = \mu_0$, contre une hypothèse multiple (ou composite) $H_1 : \mu > \mu_0$ (ou $\mu \neq \mu_0$).

(a) Pour un échantillon $\mathcal{S}_n = (x_1, \ldots, x_n)$ de la question 2, utiliser la fonction *t.test* pour faire le test vu ci-dessus. On lira attentivement l'aide de la fonction pour comprendre les inputs et outputs : à quoi correspond la valeur "t". A quoi correspond *df* ?

(b) Si on note $x \mapsto F_{n-1}^T(x)$, la fonction de répartition d'une loi de Student à $n-1$ degrés de libertés, alors la p-value donnée par la fonction *t.test* est égale à $p - value = 1 - F_{n-1}^T(t)$, où t est la valeur donnée précédemment. En se rappelant la forme de la zone de rejet W, et le caractère monotone d'une fonction de répartition, expliquer comment la p-value permet de prendre une décision au niveau $\alpha = 0.05$.

(c) Reprendre les N échantillons de la question 2, et utiliser la p-value pour étudier l'impact de α variant dans $\alpha = 0.2, 0.1, 0.05, 0.01$.

(d) La fonction *t.test* permet de calculer l'intervalle de confiance au niveau $1 - \alpha$. Rappeler comment l'intervalle de confiance. Sur les N échantillons de la question 2, dans combien de cas 1 est dans l'intervalle de confiance. Est ce normal ?