Statistiques décisionnelles : TD 7

Le but de cette feuille de TD est de vous faire découvrir quelques tests de comparaison des moyennes pour des échantillons gaussiens. Dans beaucoup de situations, vous avez déjà les connaissances requises pour construire les tests adéquats. Ces tests font partie du cours et vous devez savoir les construire et vous en servir.

1 Rappels

Pour construire un test, une démarche classique consiste à trouver une statistique de test T qui dépend des données, dont on connait la loi sous H_0. On déterminera la forme de la zone de rejet grâce au comportement de cette statistique T sous H_1 (test bilatéral ou unilatéral ?), et le seuil de rejet grâce aux quantiles de T sous H_0.

Théorème 1 (Théorème de Cochran)**.** *Soit $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi $\mathcal{N}(m, \sigma^2)$, alors*

1. $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *et* $V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$ *sont indépendantes.*
2. $\frac{n-1}{\sigma^2}V_n$ *suit une loi* $\chi^2(n-1)$.

Théorème 2 (Théorème de Student)**.** *Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi $\mathcal{N}(m, \sigma^2)$, soit $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$, alors*

$$\frac{\sqrt{n}}{\sqrt{V_n}}\left(\overline{X_n} - m\right) \sim t(n-1).$$

2 Exercices

Exercice 1 (1 échantillon gaussien)**.** *On dispose d'un échantillon des n variables aléatoires i.i.d. $(X_1, \ldots, X_n)$, loi $\mathcal{N}(m, \sigma^2)$ où m et σ^2 sont inconnus. Attention, dans les questions suivantes, les zones de rejet doivent être adaptées à l'hypothèse alternative.*

1. *Construire un test de niveau α pour $H_0 : m = m_0$ contre $H_1 : m \neq m_0$.*
2. *Comment modifiez-vous la zone de rejet précédente pour faire un test de niveau α pour $H_0 : m = m_0$ contre $H_1 : m < m_0$?*
3. *Comment modifiez-vous la zone de rejet précédente pour faire un test de niveau α pour $H_0 : m = m_0$ contre $H_1 : m > m_0$?*
4. *Construire un test de niveau α pour $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 \neq \sigma_0^2$.*

Application : *On observe un échantillon de* 11 *variables gaussienne i.i.d. dont les valeurs sont rapportées ci-dessous*

$-4.3 \quad -1.17 \quad 0.39 \quad 0.23 \quad -2.13 \quad -3.13 \quad 3.74 \quad 3.17 \quad -0.28 \quad -2.58 \quad -0.79$

Sur cet échantillon, la moyenne empirique vaut -0.62 et la variance empirique vaut 6.2 *. Peut-on dire que cet échantillon est de moyenne nulle au niveau* 5% *?*

Exercice 2 (Égalité des moyennes pour deux échantillons gaussiens appariés)**.** *On dispose de deux échantillons appariés $X_1, \dots, X_n$ et $Y_1, \dots, Y_n$ où les $X_i \sim \mathcal{N}(m_1, \sigma_1^2)$ et les $Y_i \sim \mathcal{N}(m_2, \sigma_2^2)$. On cherche à construire un test de niveau α pour $m_1 = m_2$ contre $m_1 \neq m_2$ (ou bien $m_1 < m_2$ ou $m_1 > m_2$).*

1. *On pose pour $i \in \{1, \dots, n\}$, $Z_i = X_i - Y_i$. Montrer que $Z_i \sim \mathcal{N}(m_1 - m_2, \sigma^2)$? Pourquoi ne peut-on pas en déterminer la variance en fonction de σ_1 et σ_2 ?*
2. *Construire un test de niveau α pour $H_0 : m_1 - m_2 = 0$ contre $H_1 : m_1 - m_2 \neq 0$.*
3. *En déduire un test de niveau α pour l'égalité des espérances contre $m_1 \neq m_2$.*

Application : *On observe les notes de 46 élèves en L3MIASHS en statistiques mathématique et en simulation, et on suppose que la répartition des notes est gaussienne[1]. On appelle $Z_1, \dots, Z_{46}$ les écarts entre les moyennes dans ces deux matières. La moyenne observée est de 1.5 et la variance empirique de cet échantillon est 8.*

Peut-on dire que l'espérance de la note des élèves en statistiques mathématiques est supérieure à celle correspondant au cours de simulation ?

Exercice 3 (Égalité des moyennes pour deux échantillons gaussiens indépendants, de variances connues)**.** *On dispose de deux échantillons i.i.d. $X_1, \dots, X_{n_1}$ et $Y_1, \dots, Y_{n_2}$, et les deux échantillons sont indépendants. De plus, les variables $X_i \sim \mathcal{N}(m_1, \sigma_1^2)$ et les variables $Y_j \sim \mathcal{N}(m_2, \sigma_2^2)$ où m_1 et m_2 sont inconnus* ***mais*** *σ_1 et σ_2 sont connus. On cherche à construire un test de niveau α pour $m_1 = m_2$ contre $m_1 \neq m_2$ (ou bien $<$ ou $>$).*

1. *Quelles sont les espérance, variance et la loi de $\bar{X}_{n_1}$? Même question pour $\bar{Y}_{n_2}$.*
2. *Quelle est la loi de $\bar{X}_{n_1} - \bar{Y}_{n_2}$?*
3. *Construire un test de niveau α pour $m_1 = m_2$ contre $m_1 \neq m_2$.*

Application : *Il est rare que la variance des mesures soit connues à l'avance. C'est le cas lorsqu'on fait des mesures avec un appareil étalonné. On dispose de deux objets A et B très légers (dont le poids est de quelques milligrammes) et on souhaite déterminer si A est plus lourd que B[2]. Malheureusement, les balances dont on dispose sont précises à l'ordre du milligramme et les mesures sont donc entachées d'une erreur. On dispose de 20 balances du même fabriquant, qui garanti que le poids affiché en milligramme est le vrai poids plus une erreur gaussienne de variance 4. On utilise alors 10 balances pour effectuer 10 mesures de A et les 10 autres pour effectuer 10 autres mesures de B. La moyenne des masses mesurées pour A est de $10mg$ et celle pour B est de $9mg$.*

1. *Expliquez pourquoi cette situation correspond à la situation théorique ci-dessus.*
2. *Déterminez la masse de A est supérieur à celle de B, au niveau 2%.*

3 Conclusion

Ces tests sont à connaître, mais ils ne demandent pas de connaissance nouvelle par rapport au premier semestre. Ce qui est important pour moi, c'est que vous soyez autonomes dans le choix et la construction de vos tests. Vous devez savoir identifier quel test est pertinent à partir d'une situation pratique.

1. Les notes sont fictives, et l'hypothèse de répartition gaussienne est un peu farfelue ici.
2. Ici je ne fais pas de distinction poids/masse, mais si vous travaillez un jour avec des physiciens ou des chimistes, vous verrez qu'ils n'apprécient pas tous cette confusion.