

MATH0487-1 - Éléments de statistique

Partie 2 du projet personnel

Généralités

Le projet portera sur l'étude des résultats de première session du cours d'*éléments de probabilité* lors de l'année académique 2015-2016^{1 2}. Lors de la deuxième partie de ce projet consacrée aux statistiques inférentielles, vous devrez d'une part étudier différents estimateurs et d'autre part réaliser des tests d'hypothèses.

Ce travail devra être réalisé individuellement. Chaque étudiant devra rendre une archive `.zip` contenant d'une part un rapport au format pdf et d'autre part ses codes sources MATLAB. Les rapports inutilement longs sont à proscrire. Il n'est ni nécessaire d'écrire une introduction, ni de faire des rappels des questions posées, ni de prévoir une table des matières. Toute sous-question posée dans l'énoncé devra comporter un élément de réponse dans le rapport, en justifiant votre raisonnement. Vous devez rendre un code source MATLAB pour toutes les sous-questions et le mettre en annexe du rapport. Toute forme de plagiat sera sanctionnée.

Cette partie du projet doit être rendue pour le jeudi 08/12/2016 23 :59 via la plateforme

`submit.run.montefiore.ulg.ac.be`

Au delà de la deadline il ne sera plus possible de soumettre les projets.

Présentation du problème

Vous disposez du même fichier que pour la partie 1 du projet. Nous nous intéressons en particulier à la note finale des étudiants, qui pour simplifier les choses se calcule par la moyenne des neuf colonnes du fichier de données. Dans ce projet, nous considérons à nouveau que les résultats des données complètes représentent la population. Nous tirons plusieurs échantillons i.i.d. pour estimer différents paramètres et réaliser des tests d'hypothèses. L'objectif premier de ce projet est d'étudier plus précisément les estimateurs et de s'interroger sur les hypothèses sous-jacentes. Le second objectif est d'apprendre à réaliser des tests d'hypothèses et à comprendre les limitations de cet outil.

-
1. on ne considère que les étudiants ayant obtenu une cote non nulle lors de l'examen écrit
 2. les points ont été arrondis de manière à n'avoir que des nombres entiers

Questions

1. Estimation

Tirer 100 échantillons i.i.d. de 20 notes finales d'étudiants.

- (a) Calculez pour chaque échantillon la moyenne m_X et sauvegardez les 100 valeurs dans une nouvelle variable. Utilisez cette nouvelle variable pour estimer le biais et la variance de l'estimateur m_X de la note finale moyenne de la population.
- (b) Calculez pour chaque échantillon la médiane $median_X$ et sauvegardez les 100 valeurs dans une nouvelle variable. Utilisez cette nouvelle variable pour estimer le biais et la variance de l'estimateur $median_X$ de la note finale moyenne de la population.
- (c) Répétez les deux points précédents avec des échantillons i.i.d. de taille 50. Que constatez-vous ? Interprétez.
- (d) Construisez pour chaque échantillon de taille 20 un intervalle de confiance à 95% de la note finale moyenne de la population à partir de m_X en faisant l'hypothèse que la variable parente est Gaussienne et
 - i. en utilisant la loi de student pour construire l'intervalle.
 - ii. en utilisant la loi de Gauss pour construire l'intervalle.

Vérifiez dans les deux cas combien des 100 intervalles de confiance contiennent la valeur de la population. Interprétez. Etait-il raisonnable de supposer que la variable parente soit Gaussienne ?

2. Tests d'hypothèse

Dans cette partie du travail, nous imaginons que les autorités de l'ULg ainsi que 6 instituts de sondages externes s'intéressent au cours de probabilité du professeur Wehenkel et veulent tester l'hypothèse H_0 = "un quart des étudiants ont obtenu une note finale strictement inférieure à 10/20" versus l'hypothèse alternative H_1 = "plus d'un quart des étudiants ont obtenu une note finale strictement inférieure à 10/20". Tous les organismes tirent un échantillon i.i.d. de 20 notes finales d'étudiants³ et utilisent le même seuil de signification $\alpha = 5\%$. Un article dans la gazette locale critiquant le cours de probabilité du professeur Wehenkel sera publié si et seulement si au moins un institut de sondage externe rejete l'hypothèse H_0 .

Tirer 100 fois 7 échantillons i.i.d. de 20 notes finales d'étudiants⁴.

- (a) Effectuez dans chaque cas le test d'hypothèse demandé. Dans combien de cas les autorités de l'ULg ont-elles rejeté l'hypothèse ? Comparez cette valeur à α .
- (b) Dans combien de cas y a-t-il eu un article dans la gazette locale ? Comparez cette valeur à celle de la question précédente. Interprétez.
- (c) Quelle(s) méthode(s) aurait-on pu utiliser pour éviter que les instituts de sondage externes soient avantagés par rapport aux autorités de l'ULg⁵ ?

3. dont ils déduisent la proportion d'étudiants ayant obtenu moins de 10/20

4. le premier des sept correspondant à chaque fois aux autorités de l'ULg

5. en ce sens qu'à eux six ils avaient plus de chance de tomber sur un échantillon rejetant H_0