
Devoir Maison : Utilisation d'une calculatrice programmable et comparaison avec les fonctions R

Le but est ici d'apprendre à se servir de la calculatrice programmable du bac pour faire tourner les procédures disponibles dans les calculatrices TI ou Casio, permettant d'obtenir rapidement des calculs de probabilités ou quantiles (menu proba des calculatrices) et de faire de la description statistique d'échantillons (à 1 ou 2 échantillons), des calculs d'intervalles de confiance, ou des tests paramétriques sur 1 ou 2 échantillons (comparaisons de moyennes ou de variances dans le cas de deux échantillons).

Les données sont présentées dans le fichier `notes.csv` envoyé par mail et déposé sur Chamilo dans le répertoire TP/Donnees-TP. Ce document sera en ligne dans le répertoire TP-Enoncés. Des modes d'emplois des calculatrices TI et Casio sont en ligne dans le répertoire Doc-TD sous le nom Aides-calculatrices (les 2 premières pages concernent les TI et les deux dernières les Casio). Les données proposées sont fictives et ont été simulées pour l'exercice. La colonne "Id" indique le numéro du candidat pour un certain bac en 2019, la colonne "math" son résultat (sur 20) à l'épreuve de mathématiques et la colonne "franc" sa note de français à l'écrit l'année précédente, la dernière colonne "lycée" désigne le lycée où il a suivi sa terminale. Ces échantillons ont été tirés dans la population totale des élèves de chacun des deux lycées et lorsque l'on parle par exemple de la moyenne des élèves du lycée A en mathématiques, il s'agit de la moyenne inconnue sur toute la population des candidats du lycée A sur laquelle on a prélevé un petit échantillon de taille 12. *Pour chacune des questions posées il faudra savoir répondre avec la procédure de la calculatrice et avec R et exprimer clairement à l'oral le résultat obtenu.*

Exercice 1 : Calculs de probabilités et Saisie de données

Indiquer pour chaque question le nom de votre calculatrice le nom du programme utilisé ainsi que la procédure R permettant de produire vos résultats.

1. Calculer la probabilité qu'une variable normale de moyenne 2 et d'écart-type 3 appartienne à l'intervalle $[-1, 5]$ avec le programme de la calculatrice et vérifier le résultat avec R (fonction `pnorm`).
2. Calculer la probabilité qu'une variable normale de moyenne 2 et d'écart-type 3 appartienne à l'intervalle $[-4, 8]$ avec le programme de la calculatrice et vérifier le résultat avec R (fonction `pnorm`).
3. Calculer la probabilité qu'une variable de Student à 13 degrés de liberté appartienne à l'intervalle $[-\infty, 3]$ avec le programme de la calculatrice et vérifier le résultat avec R (fonction `pt`).
4. Calculer le quantile d'ordre 2,5% d'une normale centrée réduite et vérifier le résultat à l'aide de la table statistique (préciser la page de lecture et la position dans l'abaque de la valeur lue).

5. Calculer le quantile d'ordre 2,5% d'une Student à 11 degrés de liberté et vérifier le résultat à l'aide de la table statistique (préciser la page de lecture et la position dans l'abaque de la valeur lue).
6. Calculer la probabilité qu'une variable du chi-deux à 11 degrés de liberté dépasse la valeur 12.
7. Calculer le quantile d'ordre 2,5% de la variable précédente et vérifier sa valeur sur les tables statistiques.
8. Apprendre à se servir du menu Edit et saisir dans la première liste les notes de mathématiques obtenues pour les candidats du lycée A et dans la colonne suivante les notes de français des mêmes candidats (lycée A). Saisir ensuite dans les deux colonnes suivantes les notes obtenues en mathématiques (liste 3) puis en français (liste 4) des candidats du second lycée B.
9. Dans la colonne suivante (liste 5), calculer la différence entre note de mathématiques et note de français pour les candidats du lycée A puis faire de même dans la liste 6 pour les candidats du lycée B.
10. A l'aide du menu statistique à une variable, retrouver la taille, la moyenne, l'écart-type empirique, l'écart-type empirique corrigé (l'estimation sans biais de l'écart-type noté s' dans le cours) pour chacun des 4 échantillons de départs (listes 1 à 4) et pour les échantillons des différences (listes 5 et 6).
11. Quelle fonction de la calculatrice utilisera-t-on pour calculer un intervalle de confiance sur la moyenne d'un échantillon gaussien avec variance connue ou avec variance inconnue ?
12. Quelle fonction de la calculatrice utilisera-t-on pour calculer un intervalle de confiance sur la moyenne d'un échantillon de Bernoulli (cas d'une proportion inconnue) ?
13. Quels sont les échantillons appariés ou indépendants parmi les quatre saisis dans les quatre premières listes (il y a 6 couples d'échantillons possibles et pour chaque dire s'ils sont appariés ou indépendants) ?
14. Indiquer pour chacun des couples précédent si leur comparaison présente un intérêt ou non en précisant lorsque la comparaison est intéressante quel effet elle permettra d'établir (celui de la matière ou du lycée).

Exercice 2 : Estimation et Intervalles de confiance

Pour chacune des questions il faut savoir décrire à l'oral les conditions requises (s'il y en a) pour appliquer la routine de R ou le programme de la calculatrice et exprimer clairement et oralement le résultat obtenu.

1. Calculer l'estimation sans biais de la note moyenne en mathématiques d'un candidat venant du lycée A. Calculer également l'estimation sans biais de l'écart-type.
2. Même question pour un candidat venant du lycée B.

3. Calculer l'intervalle de confiance de niveau 90% pour la note moyenne en mathématiques des candidats du lycée A et sachant que la variance est connue et vaut 4 (rappel : R et la calculette demandent la valeur de l'écart-type).
4. Calculer l'intervalle de confiance de niveau 90% pour la note moyenne en mathématiques des candidats du lycée B et sachant que la variance est connue et vaut 4 (rappel : R et la calculette demandent la valeur de l'écart-type).
5. Calculer l'intervalle de confiance de niveau 90% pour la note moyenne en mathématiques des candidats du lycée A dans le cas où la variance est supposée inconnue.
6. Calculer l'intervalle de confiance de niveau 90% pour la note moyenne en mathématiques des candidats du lycée B dans le cas où la variance est supposée inconnue.
7. Calculer l'intervalle de confiance de niveau 90% pour la probabilité qu'un candidat vienne du lycée A.
8. Calculer l'intervalle de confiance de niveau 90% pour la probabilité qu'un candidat vienne du lycée B.

Exercice 3 : Tests sur un échantillon

Pour chacune des questions suivantes il faut savoir décrire à l'oral et en langage clair :

- les conditions requises pour appliquer le test (les hypothèses de modélisation)
 - les hypothèses du test: l'hypothèse nulle \mathcal{H}_0 et l'hypothèse alternative \mathcal{H}_1 .
 - la procédure R et le programme de la calculatrice utilisé pour répondre
 - la p-valeur du test
 - la conclusion littérale à la question posée
1. On se demande si la note moyenne en mathématiques des élèves du lycée A est la même que celles des élèves du bac 2019 sur tous les lycées et que l'on sait valoir $\mu_0 = 13$. Faire le test statistique permettant de répondre à cette question.
 2. Si la réponse à la question précédente est non, faire le test unilatéral permettant de montrer que les élèves du lycée A sont moins bons ou meilleurs en moyenne que ceux de n'importe quel lycée français.
 3. même question que la 1 pour le lycée B.
 4. même question que la 2 pour le lycée B.
 5. On se demande si en moyenne dans le lycée A les résultats en mathématiques diffèrent des résultats en français ? Si oui quel est le sens de l'effet de la matière ?
 6. Même question que dans la précédente mais pour le lycée B ?
 7. La proportion d'élèves du lycée A qui obtiennent une note en mathématiques supérieure à 12 est-elle de plus de 50% ?

8. La proportion d'élèves du lycée B qui obtiennent une note en mathématiques supérieure à 12 est-elle de plus de 50% ?

Exercice 4 : Tests de comparaison de deux échantillons

Comme dans l'exercice précédent et pour chacune des questions suivantes il faut savoir décrire à l'oral et en langage clair :

- échantillons appariés ou indépendants
 - comparaison des moyennes ou des variances
 - les conditions requises pour appliquer le test (les hypothèses de modélisation)
 - les hypothèses du test: l'hypothèse nulle \mathcal{H}_0 et l'hypothèse alternative \mathcal{H}_1 .
 - la procédure R et le programme de la calculatrice utilisé pour répondre
 - la p-valeur du test
 - la conclusion littérale à la question posée
1. Dans le lycée A les candidats obtiennent-ils en moyenne une meilleure note en mathématiques qu'en français ?
 2. Même question pour le lycée B.
 3. Les résultats sont-ils en moyenne meilleurs dans le lycée A que dans le lycée B sur l'épreuve de mathématiques ?
 4. Les résultats sont-ils en moyenne meilleurs dans le lycée A que dans le lycée B sur l'épreuve de français ?
 5. Proposer un test permettant de savoir s'il était raisonnable de poser l'hypothèse d'égalité des variances pour pouvoir faire le test de comparaison des moyennes entre les deux lycées pour l'épreuve de mathématiques. Peut-on du coup utiliser le résultat du test de comparaison des moyennes de la questions 3 ?
 6. Même question que la précédente pour l'utilisation du test de comparaison des moyennes de la questions 4.
 7. Quelle routine R utilise-t-on pour diagnostiquer l'hypothèse de modélisation requise pour pouvoir faire le test de la question 1 ?
 8. Quelle routine R utilise-t-on pour diagnostiquer l'hypothèse de modélisation requise pour pouvoir faire le test de la question 2 ?