

TD n°8 : T de Student

1 Comparaison d'un groupe à une population

1.1 Á la main...

Statistique et informatique est une matière initialement difficile mais ou il est au final possible de ramasser pas mal de points, tout le monde sait ça. Aussi, l'an dernier, certains étudiants ont adopté une stratégie pour augmenter leurs chances d'avoir des bonnes notes :

- Groupe A (groupe de soutien) : un groupe de 9 étudiants a décidé de faire du soutien mutuel : une fois par semaine, ils se sont réunis dans une salle et se sont mutuellement posés des questions sur les points qu'ils n'avaient pas compris.

À la fin de l'année, les résultats ont été les suivants :

- **Groupe A** (groupe de soutien) : notes 12, 16, 13, 14, 15, 14, 13, 14, 14, 15
- Amphi : moyenne générale de l'amphi = 9.17.

La question est de savoir si la stratégie du groupe A a été efficace, c'est à dire de savoir si la moyenne du groupe A diffère significativement de la moyenne de la population.

1. Chargez en mémoire le fichier `GroupeA.csv` et stockez le dans la variable `Donnees`.
2. Quelle est la population considérée ? Donnez sa moyenne.
3. Représentez graphiquement l'histogramme de la variable note. Quelle loi suit-elle (à peu près) ? Donnez le nom de la loi, sa moyenne et son écart type.
4. Quelle loi suit la moyenne des groupes de taille 9 ? Donnez le nom de la loi, sa moyenne et son écart type.
5. Calculez le T de Student pour le groupe A.
6. Quel est le DDL du groupe A ?
7. En consultant la table sur internet, trouvez le petit p .
8. Conclusion ?

1.2 Sous R

Sous **R**, la fonction qui calcule le T de Student est `t.test()`

9. Stockez la moyenne générale de l'amphi dans `moy`
10. Représentez graphiquement la boîte à moustache de la variable `Note`
(rappel : `boxplot(donnee$Note, ylim=c(0,20))`).
11. **R** ne sait pas qu'une note varie en 0 et 20 mais on peut le lui préciser. Pour cela, il faut fixer les limites de l'axe des y . On le fait grâce à l'instruction `boxplot(donnee$Note, ylim=c(0,20))`.
12. On aimerait, sur le même graphique, ajouter une ligne rouge indiquant la position de la moyenne générale. Pour cela, on utilise l'instruction `lines()`. Cette instruction ajoute une ligne sur un graphe existant. Pour tracer une ligne entre les points $(0, moy)$ et $(2, moy)$, utilisez l'instruction `lines(c(0,2), c(moy,moy), col="red")`

Le graphique qu'on obtient permet de situer les notes du groupe A par rapport à la moyenne.

13. Pour comparer un groupe à une moyenne générale, on utilise le T de Student. Pour cela, on doit lui préciser le groupe et la moyenne (en lui donnant `mu=moy`). Au final, on doit utiliser l'instruction `t.test(donnee$Note, mu=moy)`
14. Combien vaut le T ?
15. Combien vaut le petit p associé ?
16. La moyenne du groupe A est-elle significativement différente de celle du reste de l'amphi ?

1.3 Y a que la première fois qui coûte...

L'année précédent, un groupe (appelé **Groupe B**) de 31 étudiants a remarqué que la compréhension des CM n'était pas toujours évidente. Ils avaient donc mutualisé les notes de cours, chacun photocopiant ses notes personnelles pour les distribuer aux autres. Cette année là, la moyenne générale de l'amphi avait été de 8.93.

17. Chargez en mémoire le fichier `GroupeB.csv` et stockez le dans la variable `Donnees2`.
18. Représentez graphiquement l'histogramme de la variable note. Suit-elle une loi normale ?
19. Quelle loi suit la moyenne des groupes de taille 31 ? Donnez le nom de la loi, sa moyenne et son écart type.
20. Représentez graphiquement la boîte à moustache de la variable note, en ajoutant la moyenne générale en rouge.
21. La stratégie du groupe B est-elle efficace ?

2 Comparaison de deux groupes entre eux

Les APA sont-ils meilleurs ou moins bon que les managements sportifs ? Pour le savoir, on veut comparer la moyenne des APA a la moyenne des MS.

2.1 Á la main

22. Chargez le fichier `NoteAPA-MS.csv` et stockez le en mémoire.
23. On veut calculez la moyenne des notes des APA. Pour cela, il faut préciser à **R** qu'on veut la moyenne des notes (en utilisant `mean(donnees$Note)`) MAIS uniquement des notes des APA, c'est à dire uniquement les notes de ceux dont l'option est égale a `APA`. Pour cela, on va ajouter entre crochet une condition qui sera `donnees$option=="APA"`. Au final, la syntaxe est `mean(donnees$Note[donnees$option=="APA"])`.
24. Selon le même principe, calculez la moyenne des MS.
25. Calculez l'écart type des APA.
26. Calculez l'écart type des MS.
27. Calculez la variance commune entre les APA et les MS.
28. Calculez le T de Student entre les APA et les MS.
29. Quel est le DDL associé ?
30. Combien vaut le petit p ?
31. Conclusion ?

2.2 Sous R

R permet de calculer le T de Student en utilisant la fonction `t.test()`. Cette fois-ci, plus besoin de lui préciser la moyenne générale (que de tout façon nous ne connaissons pas). Par contre, on doit lui dire ce qui permet de définir les groupes. Dans notre cas, c'est l'option.

32. Représentez graphiquement les notes des APA et des MS côte à côte. Pour cela, utilisez `boxplot(donnees$Note~donnees$option)`.
33. Calculez le T de Student entre les APA et le MS. Pour cela, utilisez `t.test(donnees$Note~donnees$option)`.
34. Par défaut, **R** suppose que les variances des groupes sont différentes et calcule donc un T de Student légèrement modifié. Si on veut que **R** calcule le vrai T de Student, il faut ajouter l'option `var.equal=TRUE`. Calculez le vrai T de Student.
35. Combien vaut le T de Student
36. Combien vaut le petit p associé.
37. Conclusion ? Une option est-elle meilleure que l'autre ?

2.3 Il n'y a que la première fois qui coûte...

Les APA 2006 sont-ils meilleur ou moins bon que les APA 2005 ?

38. Pour le savoir, chargez le fichier `NoteAPA2005-2006.csv`
39. Représentez graphiquement les deux années côte à côte.
40. Une année est-elle significativement meilleur que l'autre ?