

Questionnaires : Tests statistiques et visualisations des données

Minosoa Barphyl Andriamiharisoa

12 July, 2025

1. Test t de Student (PlantGrowth)

Objectif : Comparer deux groupes. *Fonction de base*: `t.test()`

1. Quelle est l'hypothèse nulle du test t de Student ?
2. Le test t est-il applicable ici ? Pourquoi ?
3. Que signifie une valeur de $p < 0.05$?
4. Que pouvez-vous déduire du boxplot concernant la distribution des poids dans les groupes ?
5. Le groupe "ctrl" a-t-il une moyenne significativement différente des autres ?

Remarque : Lors de votre analyse vous allez rencontrer des erreurs, le facteur `group` de `PlantGrowth` contient plus de deux niveaux, alors que le test t ne peut être utilisé qu'entre deux groupes, pensez à prendre en compte ce paramètre en ne prenant en compte que deux groupes.

2. ANOVA (Iris dataset)

Objectif : Comparer la longueur du sépale entre plusieurs espèces. *Fonction de base*: `aov()`

1. Quelle est l'hypothèse nulle de l'ANOVA ?
2. Pourquoi ne peut-on pas utiliser un test t pour ce cas ?
3. À quoi sert la fonction `aov()` dans ce contexte ?
4. Que révèle la boxplot sur la variabilité de la longueur de sépale entre les espèces ?
5. Quelle espèce semble avoir les sépales les plus longs ?

3. Régression linéaire (mtcars)

Objectif : Étudier l'effet de la puissance (`hp`) sur la consommation (`mpg`). *Fonction de base*: `lm()`

1. Quelle est la relation entre `hp` et `mpg` d'après la régression ?
2. Quelle est l'équation de la droite de régression ?

3. Comment interpréter le coefficient de la variable `hp` ?
4. Le modèle est-il significatif ?
5. Le nuage de points suit-il une tendance linéaire visible dans le graphique ?

4. Corrélation (Iris)

Objectif : Évaluer la corrélation entre deux variables continues. *Fonction de base*: `cor.test()`

1. Quelle est la différence entre corrélation et régression ?
2. La corrélation entre la longueur du sépale et celle du pétale est-elle forte ?
3. Quelle est la valeur du coefficient de corrélation de Pearson ?
4. La relation semble-t-elle linéaire selon le graphique ?
5. Peut-on prédire la longueur du pétale à partir de celle du sépale ? Pourquoi ?

5. Questions transversales (Synthèse)

1. Quel test utiliseriez-vous si vous aviez **plus de deux groupes** à comparer ?
2. Quels tests nécessitent des variables **catégorielles** comme facteur ?
3. Quel test choisiriez-vous si vos données ne suivent pas une distribution normale ?
4. Pourquoi est-il important de **visualiser les données** avant et après l'analyse statistique ?
5. Quel test vous paraît le plus utilisé dans vos travaux de recherche, et pourquoi ?

6. Pour obtenir des graphiques esthétiques, il est recommandé d'utiliser les fonctions disponibles sur le packages ggplot2.

7. Exporter vos resultats sous format Rmarkdown afin qu'on puisse les discuter pour la seance de la semaine prochaine.