

Bilan_Jour3

Cédric HASSEN-KHODJA

Analyse de variance à un facteur

Exercice: Un chercheur s'intéresse à l'effet de différents programmes d'exercice sur la fréquence cardiaque maximale atteinte par des individus. Il a recruté 60 participants qu'il a réparti en 3 groupes de 20 personnes chacun. Chaque groupe a suivi un programme d'exercice différent pendant 8 semaines. À la fin de la période d'entraînement, le chercheur a mesuré la fréquence cardiaque maximale atteinte par chaque participant lors d'un test d'effort. Les données sont stockées dans un fichier CSV nommé "programme.csv" avec les colonnes "programme" et "fréquence_cardiaque_max".

But: Déterminer s'il existe des différences significatives entre les trois programmes d'entraînement (A, B et C) en termes de la fréquence cardiaque maximale atteinte par les participants.

Regression multiple

Exercice: Un chercheur s'intéresse à l'impact de l'âge, du sexe et du niveau d'éducation sur le salaire annuel d'une personne. Il a recueilli des données sur 12 individus et a enregistré leur âge (en années), leur sexe (0 pour les femmes et 1 pour les hommes), leur niveau d'éducation (en années d'études) et leur salaire annuel (en milliers de dollars). Les données sont stockées dans un fichier CSV nommé "social.csv".

1. Importez les données dans R et effectuez une analyse exploratoire pour identifier les variables importantes et visualiser les relations entre les variables.
2. Effectuez une régression multiple pour prédire le salaire annuel en fonction de l'âge, du sexe et du niveau d'éducation.
3. Interprétez les coefficients de régression pour chaque variable et déterminez si les variables sont significativement liées au salaire annuel.
4. Évaluez la qualité de l'ajustement du modèle et faites des prévisions sur de nouvelles données.

Regression logistique

Exercice: Une entreprise souhaite étudier les facteurs qui influencent la décision des clients d'acheter ou non un produit. Le dataset contient les données de 1000 clients, y compris leur âge, leur sexe, leur revenu annuel et s'ils ont effectué un achat ou non. La variable cible est une variable binaire : 1 si le client a effectué un achat et 0 sinon. Les données sont présentées dans le fichier "produit.csv".

1. Importez les données dans R et effectuez une analyse exploratoire pour identifier les variables importantes et visualiser les relations entre les variables.
2. Effectuez une régression logistique pour prédire la probabilité d'un achat en fonction de l'âge, du sexe et du revenu annuel.
3. Interprétez les coefficients de régression pour chaque variable et déterminez si les variables sont significativement liées à la probabilité d'un achat.
4. Évaluez la qualité de l'ajustement du modèle et faites des prévisions sur de nouvelles données.

Test t pour comparer les moyennes de deux groupes indépendants

Exercice: Un chercheur souhaite étudier l'effet d'un nouveau médicament sur la tension artérielle. Il a recruté 50 participants atteints d'hypertension artérielle, divisés en deux groupes : un groupe a reçu le nouveau médicament et l'autre groupe a reçu un placebo. Après une semaine, le chercheur a mesuré la tension artérielle de chaque participant. Les données sont présentées dans le fichier "tension_arterielle.csv".

1. Importez les données dans R et effectuez une analyse exploratoire pour visualiser les différences potentielles entre les deux groupes.
2. Effectuez un test t pour comparer les moyennes de tension artérielle entre le groupe qui a reçu le médicament et le groupe qui a reçu le placebo.
3. Interprétez les résultats du test t et déterminez s'il y a une différence significative entre les moyennes des deux groupes.
4. Évaluez l'effet potentiel du médicament sur la tension artérielle en calculant la taille de l'effet.

La taille de l'effet correspond au rapport entre la différence de moyenne entre deux groupes et la variance de l'expérience.

Test t apparié

Exercice: Un chercheur a voulu évaluer si un nouveau traitement augmentait la concentration sanguine d'une certaine protéine. Il a mesuré les concentrations de protéine chez 10 participants avant et après le traitement. Le chercheur veut savoir si le traitement a un effet significatif sur la concentration de protéine. Les données sont présentées dans le fichier "conc_prot.csv".

Analyse de variance à deux facteurs

Exercice: On s'intéresse à l'effet de deux types de traitement (A et B) sur la croissance de plantes dans différents types de sol (Sol 1, Sol 2, Sol 3). Pour cela, un essai expérimental a été réalisé avec un échantillon de 60 plantes. Les plantes ont été réparties au hasard, chacun correspondant à un type de sol différent. Dans chaque groupe, les plantes ont été réparties au hasard en deux sous-groupes, l'un recevant le traitement A et l'autre le traitement B. Après une période de croissance de 2 mois, la taille des plantes a été mesurée en cm.

Les données sont stockées dans le fichier "plantes.csv" avec les variables suivantes :

- taille : taille des plantes en cm
- traitement : type de traitement (A ou B)
- sol : type de sol (Sol 1, Sol 2 ou Sol 3)

Effectuez une analyse de variance à deux facteurs pour comparer les effets du traitement et du sol sur la croissance des plantes et interprétez les résultats.

GLMM

Question biologique: Vous avez étudié l'impact de la température (variable continue) et de l'altitude (variable catégorielle) sur la présence d'une espèce de papillon dans différentes zones d'étude. Vous souhaitez utiliser un modèle GLMM pour prendre en compte la corrélation entre les observations dans une même zone d'étude.

Exercice: Utilisez R pour construire un modèle GLMM qui prédit la présence ou l'absence d'une espèce de papillon en fonction de la température et de l'altitude, en tenant compte de la corrélation entre les observations dans une même zone d'étude. Interprétez les coefficients de régression.

χ^2

Exercice: Une étude a été menée pour déterminer s'il existe une association entre le tabagisme et la maladie pulmonaire chronique (MPC). Les données ont été collectées auprès de deux groupes : un groupe de fumeurs et un groupe de non-fumeurs.