

TD n°7 : Corrélation et régression

1 Corrélation

Dans ce TD, on cherche si connaître la taille d'un individu peut nous permettre de deviner ses performances en saut, en math et en gym.

1. Chargez le fichier "SautHauteur.csv" et stockez le dans la variable `donnees`
2. Représenter graphiquement les données grâce à la fonction `plot(donnees$Taille, donnees$Perf)`
3. Calculez la moyenne et l'écart type de [TAILLE] et de [PERF] (grâce à `mean()` et `sd()`).
4. Calculez la covariance de la Taille et de la Perf. Pour cela, utilisez la fonction `cov(donnees$Taille, donnees$Perf)`
5. En utilisant la formule du cours (avec la covariance et les écarts type), calculez le coefficient de corrélation.
6. **R** permet de calculer le coefficient de corrélation automatiquement. Pour cela, utilisez la fonction `cor(donnees$Taille, donnees$Perf)`

Pour voir si le coefficient de corrélation est significativement différent de zéro (et donc voir s'il existe un lien entre les variables), il faut calculer le petit p .

7. Pour calculer le petit p , utilisez la fonction `cor.test(donnees$Taille, donnees$Perf)`.

Cette fonction vous donne beaucoup d'informations.

8. Combien vaut le petit p ?
9. Combien vaut le coefficient de corrélation ?
10. Donnez l'intervalle de confiance de la corrélation au risque 95%

2 Régression linéaire

Faire une régression linéaire de [Y] en fonction de [X], c'est trouver la droite qui permet de prédire [Y] (la variable à expliquer) grâce à [X] (variable explicative). Cette droite est de la forme : $[Y] = b_0 + b_1 \times [X]$

La fonction qui permet de faire une régression linéaire est `lm()`. Pour l'utiliser, on doit lui donner la variable à expliquer (variable Y) puis le symbole tilde : `~` puis la variable explicative X. Dans notre cas, on cherche à expliquer la [PERF] et on veut l'expliquer grâce à la [TAILLE].

11. Calculez la droite de régression linéaire de [PERF] en fonction de [TAILLE]. Pour cela, utilisez l'instruction `lm(donnees$Perf~donnees$Taille)`.

Cette instruction vous donne deux choses : l'intercept (qui correspond à l'ordonnée à l'origine, que nous avons appelé b_0 dans le cours) et le coefficient de la variable `Taille` qui correspond à b_1 .

12. Donnez la droite de régression linéaire de [PERF] en fonction de [TAILLE].
13. Selon notre modèle, combien peut sauter quelqu'un qui mesure 2m ?

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

Corrélation et régression linéaire comportent donc quatre étapes :

- Représenter les variables graphiquement.
- Calculer le coefficient de corrélation
- Calculer le petit p associé.
- S'il existe un lien entre les variables, calculer la droite de régression linéaire.

Vous venez de faire tout cela pour les variables [PERF] en fonction de [TAILLE]. On cherche maintenant à savoir si la [TAILLE] peut prédire la [NOTEMATH], ou si [TAILLE] peut prédire [NOTEGYM].

14. Chargez le fichier "Math.csv" dans la variable donnees2 .
15. Faites une corrélation et s'il y a un lien une régression des [TAILLES] sur [NOTEMATH].
16. Chargez le fichier "Gym.csv" dans la variable donnees3
17. Faites une corrélation et s'il y a un lien une régression des [TAILLES] sur [NOTEGYM].