

TD 5 Mesurer la subjectivité

1 Données

Nous allons travailler sur l'enquête **EPO2009 Fraude à l'université**. Reprenez votre script "analyse univariée". Pour cela :

1. Dans R, allez dans *Fichier* → *Ouvrir un script* puis ouvrez votre script de la semaine dernière.
2. Sélectionnez tout votre code et envoyez-le vers la console (grâce aux touches Ctrl+R).

2 Construction d'un score

Nous allons transformer les items `Copier`, `Communiquer`, `EchangeBrouillon`, `QuestionProf`, `AntisecheSMS`, `Internet`, `Cours`, `GarderCopie`, `RetardExam`, `PreparerSalle` et `VolerSujet` en un score. Pour cela, il faut indiquer à R que ces 12 variables sont ordonnées.

3. Ordonnez les 12 variables (comme cela avait été fait dans le TD 2 pour la variable `Copier`).
4. Vérifiez que les variables sont maintenant ordonnées (instruction `str`).

Maintenant que les variables sont ordonnées, R peut les transformer en variables discrètes en numérotant les modalités de 1 à k grâce à l'instruction `as.integer`.

5. Ecrivez `as.integer(donnees$Copier)`. Qu'obtenez-vous ?
6. Ecrivez `as.integer(donnees$Copier)-1`. Qu'obtenez-vous ?
7. Quelle est la différence ? Laquelle des deux formules conviendra le mieux à vos besoins ?

Nous allons créer de nouvelles variables (nouvelles colonnes) qui contiendront les variables recodées en numérique. Pour cela, il suffit de choisir un nom pour la nouvelle colonne et de dire à R quoi mettre dedans.

8. Première étape, il faut choisir le nom de la colonne qui doit servir à contenir la variable `Copier` recodé en numérique. On peut par exemple prendre `CopierNUM`.

La création de cette nouvelle colonne se fait grâce à l'instruction `donnees$CopierNUM <- as.integer(donnees$Copier)-1`.

9. Faites de même pour les 11 autres variables mesurant la triche : créez 11 autres colonnes.

Tout est maintenant prêt pour le calcul final du score. Dans une nouvelle colonne, nous allons additionner les 12 colonnes que nous venons de créer :

10. Le nom de cette nouvelle colonne est `ScoreTriche`. Pour la créer, utilisez l'instruction :
`donnees$ScoreTriche <- donnees$CopierNUM + donnees$CommuniquerNUM + ...`
`... + donnees$VolerSujetNUM`

Le score de triche est maintenant calculé. Comment se comporte-t-il ? Pour le savoir...

11. Faites l'analyse univariée de la variable `ScoreTriche`.

3 Validation

3.1 Vérification que la valeur est conforme

On veut vérifier si ce que mesure notre score est bien relié à la fraude. Pour cela, on va le comparer à d'autres mesures de fraude.

12. Dans toutes les variables autres que celles qui composent l'échelle, trouvez-en une qui est a priori indicatrice d'une tendance à la fraude.
13. Cette variable est-elle significativement liée à notre score de fraude ?
14. Conclusion ?

3.2 Qualité des items

Pour tous les items, il faut vérifier s'ils présentent un effet plafond ou plancher.

15. Présentez graphiquement tous les items. Pour cela, on pourra diviser la fenêtre graphique en trois lignes et 4 colonnes grâce à l'instruction `par(mfrow=c(3,4))`.
16. Puis pour chaque item, tracez sa représentation graphique.
17. Quels items sont bons ? Quels items sont mauvais ?

3.3 Sensibilité au changement

Il faudrait faire une enquête non anonyme et refaire passer le questionnaire une semaine plus tard...

3.4 Uni-dimensionnalité

Pour savoir si notre échelle est unidimensionnelle, on trace le diagramme des valeurs propres. Cela se fait avec la fonction `scree.plot`. `scree.plot` travaille sur l'ensemble des données qui ont servi à la construction de l'échelle. Il va donc falloir préciser les numéros des colonnes de `donnees` contenant les variables de triche recodées en numérique.

18. Grâce à l'instruction `name(donnees)`, repérez les numéros des colonnes CopierNUM (que l'on notera `d`) et VolerSujetNUM (que l'on notera `f`).
19. Écrivez `summary(donnees[,d :f])`. Qu'obtenez-vous ?
20. Vous pouvez maintenant utiliser `scree.plot(donnees[,d :f])`. Combien de dimension(s) a notre échelle ?
21. Pour vérifier le nombre de dimension, on peut demander à R de faire des simulations. Pour demander 20 simulations, ajoutez l'option `simu=20` à l'instruction `scree.plot` précédente. Avez-vous confirmation du nombre de dimension ?

3.5 Cohérence interne

La cohérence interne (mesurée par le α de Cronbach) est un concept compliqué mais est très simple à calculer. Cela se fait avec l'instruction `cronbach`, la encore appliquée sur les données qui ont servi à la construction de l'échelle.

22. Calculez l' α de Cronbach de notre échelle.
23. Un α de 0.8 est bon, 0.7 est moyen, 0.6 est mauvais. Comment est notre échelle ?

3.6 Redondance des items

Les items ne doivent pas être trop corrélés entre eux. Pour vérifier, on utilise la fonction `var` appliquée à l'ensemble des colonnes composant l'échelle.

24. Calculez les corrélations entre les colonnes (rappel : pour éliminer les manquantes, il existe une option...)
25. Quelle est la plus grande corrélation (en dehors de la diagonale) ?