

## TD 3&4. Analyse Bivariée

**Note IMPORTANTE :** dans tout ce qui suit, ce qui compte le plus est la rédaction des conclusions. NE LES OMETTEZ PAS !

Tous vos calculs doivent être faits dans un script R, tous vos résultats (et vos conclusions) doivent être notés dans un fichier Word. Le fichier Word ne doit pas contenir de code R. Il doit être clairement présenté et compréhensible pour quelqu'un ne connaissant rien à R.

### 1 Données

Nous allons travailler sur l'enquête **EPO2009 Fraude à l'université**. Reprenez votre script "analyse univarié". 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 Deux variables nominales

#### 2.1 $\chi^2$

Un  $\chi^2$  se calcule avec la fonction `chisq.test`. Cette fonction vous donne le  $\chi^2$ , le petit p associé et elle précise également si les conditions d'applications sont remplies : quand elles NE sont PAS remplies, un *Warning message* précisant que le  $\chi^2$  n'est pas valide apparait.

3. Dressez le tableau croisé des variables `Sexe` et `Filieres`.
4. Calculez leur  $\chi^2$  grâce à l'instruction `chisq.test(donnees$Sexe,donnees$Filiere)`.
5. Conclusion ?
6. Les conditions d'applications sont-elles remplies ?
7. Dans votre document word, notez le  $\chi^2$  et son petit p.

#### 2.2 Test exact de Fisher

Le test exact de Fisher se calcule avec la fonction `fisher.test`. Le test exact de Fisher ne vous donne que le petit p.

8. Calculez le test exact de Fisher pour les variables `Sexe` et `Filiere` grâce à l'instruction `fisher.test(donnees$Sexe,donnees$Filiere)`.

### 3 Deux variables numériques

#### 3.1 R de Pearson

Un r de Pearson se calcule avec la fonction `cor.test`. Cette fonction vous donne la corrélation (c'est à dire le r de Pearson) et le petit p associé. Pour vérifier les conditions d'application (normalité d'au moins une des deux variables), vous devez tracer les histogrammes des variables.

On cherche à établir s'il y a un lien entre les variables `Age` et une des variables mesurant la gravité de la triche (par exemple `GraviteEchangeBrouillon`)

9. Dressez l'histogramme de `Age`. La variable suit-elle une loi normale ?
10. Dressez l'histogramme de `GraviteEchangeBrouillon`. La variable suit-elle (à peu près) une loi normale ?

11. Tracer le nuage de points des deux variables à l'aide de `plot(donnees$Age~donnees$GraviteEchangeBrouillon)`.
12. Calculez la corrélation entre les deux variables grâce à la fonction `cor.test(donnees$Age, donnees$GraviteEchangeBrouillon)`.
13. Dans votre document word, notez la valeur de la corrélation, le petit p associé et votre conclusion.

### 3.2 R de Spearman

Le R de Spearman se calcule avec la même fonction de le R de Pearson, il suffit simplement de lui ajouter l'option `method=c("spearman")`.

14. Dressez les histogramme de `Age` et `GraviteCopier`. Conclusion ? Quel test faut-il utiliser ?
15. Tracez le nuage de points des deux variables.
16. Calculez leur corrélation de Pearson grâce à la fonction `cor.test(donnees$Age, donnees$GraviteCopier, method="spearman")`.
17. Conclusion ?

## 4 Une numérique et une nominale a deux classes

### 4.1 T de Student

Un t de Student se calcule avec la fonction `t.test`. Cette fonction vous donne le t et le petit p associé. Pour vérifier les conditions d'application :

- a. normalité de la variable continue OU effectifs supérieur à 30 : si les effectifs sont supérieurs a 30, il n'y a rien à faire. Sinon, il suffit de tracer l'histogramme de la variable pour vérifier qu'elle suit bien une loi normale.
- b. égalité des variances : il faut calculer la variance de la variable numérique pour une des modalités de la variable nominale, puis pour l'autre.

On cherche a établir s'il y a un lien entre les variables `Age` et `Redoublement`.

18. Quel est l'effectif de la population ?
19. Quelle est la variance des ages des hommes ? Celle des ages des femmes ? Pour cela, utilisez `var(donnee$Age[donnee$Redoublement==TRUE], na.rm=TRUE)` et `var(donnee$Age[donnee$Redoublement==FALSE], na.rm=TRUE)`.
20. Tracez les boites a moustache des deux modalités de `Redoublement` à l'aide de `boxplot(donnees$Age~donnees$Redoublement)`.
21. Calculez le t de Student grâce à la fonction `t.test(donnees$Age~donnees$Redoublement)`.
22. Dans votre document word, notez la valeur du t, le petit p associé et votre conclusion.

### 4.2 Test des rangs de Wilcoxon

Quand le test de Student n'est pas applicable, on doit calculer le test des rangs de Wilcoxon. Il se calcule avec la fonction `wilcox.test`.

Note : dans la présente étude, le t est toujours applicable (effectifs importants). Le test de Wilcoxon n'est donc pratiqué qu'à titre d'exercice.

23. Calculez le test des rangs de Wilcoxon grâce à la fonction `wilcox.test(donnees$Age~donnees$Redoublement)`.
24. Dans votre document word, notez la valeur du test, le petit p associé et votre conclusion.

## 5 Une numérique et une nominale a trois classes et plus

### 5.1 F de Fisher / ANOVA

Une ANOVA (analyse de variance) se calcule avec la fonction `summary(aov())`. Cette fonction vous donne le f et le petit p associé.

Pour vérifier les conditions d'application, la méthode est la même que pour le t de Student.

On cherche à établir s'il y a un lien entre les variables `Age` et `Filiere`.

25. Quel est l'effectif de la population ?
26. Quelle est la variance des ages des MS ? Des APA ? Des EM ?
27. Tracez les boîtes a moustaches.
28. Calculez le f de Fisher grace à `summary(aov(donnees$Age~donnees$Filiere))`.
29. Dans votre document word, notez la valeur du t, le petit p associé et votre conclusion.

### 5.2 Test de Kruskal-Wallis

Quand l'ANOVA et le f de Fisher ne sont pas calculable, on doit calculer le test de Kruskal et Wallis.

Note : dans la présente étude, l'ANOVA est toujours applicable (effectifs importants). Le test de Kruskal et Wallis n'est donc pratiqué qu'à titre d'exercice.

30. Calculez le test des rangs de Kruskal et Wallis grace à la fonction `kruskal.test(donnees$Age~donnees$Filiere)`.
31. Dans votre document word, notez la valeur du test, le petit p associé et votre conclusion.

## 6 Quel test ?

Pour chaque couple de variable :

32. Sexe,Redoublement
33. Age, GraviteVolerSujet
34. SurprendreTricher,Sexe
35. GraviteCopier,Sexe
36. GraviteVolerSujet,Sexe
37. Sexe,SanctionRecommencer
38. GraviteCopier,Filiere
39. GraviteEchangeBrouillon,GraviteCopier

vous devez :

- Si les variables sont nominales :
  - a. Dresser le tableau croisé des variables.
  - b. Calculer le  $\chi^2$ .
  - c. Si les conditions d'applications ne sont pas remplies, calculer le test exact de Fisher.
  - d. Dans votre document word, noter le test que vous avez utilisé, la valeur du petit p et votre conclusion. S'il s'agit du  $\chi^2$ , donnez aussi sa valeur.
- Si les variables sont numériques :
  - a. Déterminer s'il faut utiliser la corrélation de Pearson ou de Spearman.
  - b. Représenter graphiquement les variables.
  - c. Calculer la corrélation.
  - d. Dans votre document word, noter le test que vous avez utilisé, la valeur de la corrélation, le petit p et votre conclusion.
- Si une variable est numérique, l'autre nominale :
  - a. Déterminer s'il faut utiliser le T, le F, le Wilcoxon ou le Kuskal et Wallis.

- b. Représentez graphiquement les variables.
- c. Calculer la valeur du test correspondant.
- d. Dans votre document word, noter le test que vous avez utilisé, la valeur de la corrélation, le petit p et votre conclusion.

The end !