



## Analyse bivariée entre variables qualitatives

*Durée : 3h*

*L'objectif de ce TP est d'étudier le lien éventuel entre deux séries de données qualitatives.*

### Exercice 1

#### *Niveau de diplôme en fonction du sexe*

Le tableau suivant donne la répartition des salariés d'une entreprise selon le niveau de formation X et le sexe Y.

X \ Y	F	H
Bac+3	45	49
Bac+5	16	11
Bac+8	4	6

- 1) Quels calculs est-il possible de faire avec ce type de tableau ? Etablissez le tableau complet des effectifs observés en ajoutant les effectifs marginaux. Que représentent  $n_{3\bullet}$  et  $n_{\bullet 1}$  ?
- 2) Calculez le tableau des fréquences observées.
- 3) Calculez le tableau des profils lignes et celui des profils colonnes.
- 4) Faites une phrase interprétant les fréquences suivantes :  $f_{31}$ ,  $f_{3|1}$ ,  $f_{1|3}$ ,  $f_{3\bullet}$  et  $f_{\bullet 1}$
- 5) Etude du tableau des profils lignes.
  - a) Avec quelles fréquences peut-on comparer les profils lignes ?
  - b) Y-a-t-il des modalités de la variable niveau d'étude qui s'écartent du profil moyen ?
- 6) Faites la même étude avec les profils colonnes. Que pouvez-vous en conclure sur le lien entre ces variables ?
- 7) Calculez la distance du chi-deux pour confirmer votre conclusion.

### Exercice 2

#### *Durée du Chômage en fonction de l'âge et le sexe*

Dans un premier temps, nous allons étudier le lien entre l'âge et la durée du chômage tous sexes confondus.

- 1) Etablir que l'âge et la durée de chômage sont deux variables fortement liées.

- 2) Etudier cette liaison au travers les profils lignes. Quelles modalités de la durée du chômage s'écartent le plus du profil moyen ?
- 3) Etudier cette liaison au travers des profils colonnes. Quelles modalités de la variable âge s'écartent le plus du profil moyen ?
- 4) Faites la même étude par sexe

### Exercice 3

### Utilisation de R

Nous allons utiliser le jeu de données [Lenses](#) constitué de 24 observations et 5 variables qualitatives :

- Recommendation : (1) the patient should be fitted with hard contact lenses, (2) the patient should be fitted with soft contact lenses, (3) the patient should not be fitted with contact lenses.
- Age of the patient: (1) young, (2) pre-presbyopic.
- Spectacle prescription: (1) myope, (2) hypermetrope.
- Astigmatic: (1) no, (2) yes.
- Tear production rate: (1) reduced, (2) normal

Nous allons étudier le lien entre les variables Recommendation et Age.

- 1) Déterminer le tableau de contingence des effectifs et des fréquences. Ajouter la somme de lignes et des colonnes (cf. TD1).
- 2) Profils lignes
  - a) Calculer les profils lignes à l'aide de la fonction `prop.table`.
  - b) Vérifier que la somme des lignes est égale à 1.
  - c) Faire une représentation graphique (avec et sans le profil moyen)
  - d) Que pouvez-vous en conclure ?
- 3) Profils colonnes  
Calculer les profils colonnes et faites une représentation graphique. Que pouvez-vous en conclure ?
- 4) Indépendance des variables ?
  - a) Calculer le tableau des fréquences théoriques puis des effectifs théoriques.
  - b) En déduire la distance du chi-deux.
  - c) Que pouvez-vous conclure sur le lien entre les deux variables ?

Autres jeux de données : [Nursery](#) / [Shuttle Landing Control](#)