Test de chi-deux

Visseho Adjiwanou, PhD.

20 March 2023

## 1. Tableau croisé et Chi-deux

A partir des données du tableau suivant:

| id | Sexe | Emploi | Age |
| --- | --- | --- | --- |
| 1 | Masculin | Temps plein | 27 |
| 2 | Féminin | Non | 34 |
| 3 | F | Partiel | 22 |
| 4 | F | Non | 38 |
| 5 | M | Partiel | 30 |
| 6 | F | Temps plein | 27 |
| 7 | M | Temps plein | 33 |
| 8 | M | Temps plein | 39 |
| 9 | F | Temps plein | 36 |
| 10 | F | Non | 25 |
| 11 | M | Partiel | 31 |
| 12 | M | Partiel | 28 |
| 13 | M | Non | 26 |
| 14 | F | Partiel | 21 |
| 15 | F | Temps plein | 29 |
| 16 | F | Temps plein | 34 |
| 17 | F | Non | 20 |
| 18 | F | Non | 32 |
| 19 | M | - | 27 |
| 20 | F | Partiel | 33 |
| 21 | M | Temps plein | 22 |
| 22 | M | Non | 38 |
| 23 | F | Non | 35 |
| 24 | M | Temps plein | 35 |
| 25 | F | Temps plein | 27 |
| 26 | M | Temps plein | 24 |
| 27 | F | Temps plein | 37 |
| 28 | F | Non | 38 |
| 29 | F | Temps plein | 29 |
| 30 | F | Temps plein | 32 |
| 31 | F | Non | 20 |
| 32 | M | Partiel | 24 |
| 33 | M | Non | 26 |
| 34 | M | Partiel | 31 |
| 35 | M | Temps plein | 32 |
| 36 | F | Partiel | 24 |

1. Dire quel est le type de chaque variable
2. Conjecturer sur la question de recherche qui a poussé à la collecte de ces données
3. Dresser un tableau croisé liant les deux variablesde sexe et d’emploi, préciser les informations importantes du tableau (nom des colonnes et des lignes, effectifs, et effectifs marginaux)
4. Quelle conclusion tirez-vous sur ces deux variables? Quel est le problème avec cette conclusion?
5. Présenter un tableau de pourcentage marginal et commentez à nouveau.
6. Quel est le type de relation que vous observez entre ces deux variables?
7. Diriez-vous que cette relation est forte, modérée, faible ou nulle?
8. Calculer le Chi-deux
9. Déterminer le degré de liberté
10. Quelle conclusion tirez-vous sur votre confiance au résultat?
11. Calculer si possible les mesures d’intensité suivante:

* le coefficient de contingence de Pearson
* le V de Cramer
* le phi
* le coefficient de prédiction de Guttman
* Gamma de Goodman et de Kruskal
* le Dyx de Somers
* le tau-b et
* le tau-c

1. Si vous devez étudier la relation entre l’âge et l’emploi, comment procéderiez-vous?

## 2. Test de chi-deux avec R

1. Entrer les données précédentes dans R. Je vous ai fait cela. Utiliser le fichier qui s’appelle **Emploi\_femme.csv**
2. Présenter le tableau de fréquence de la variable sexe, et commentez
3. Présenter le tableau de fréquence de la variable emploi, et commentez
4. Quelle est la relation entre le sexe et l’emploi?
5. Cette relation est-elle statistiquement significative?
6. Créer une nouvelle variable d’emploi que vous notez **emploi2** qui prend les valeurs suivantes:

* Si emploi prend les valeurs “non”” ou “temps partiel”“, emploi2 prend la valeur”ne travaille pas”
* Si emploi prend la valeur “Temps plein”, emploi2 prend la valeur “Travaille”

1. Créer aussi la nouvelle variable **groupe\_age** qui prends deux valeurs:

* Moins de 30 ans
* 30 ans et plus

1. Étudiez à nouveau la relation entre le sexe et l’emploi, cette fois-ci séparément pour les moins de 30 ans et les 30 ans et plus

## 3. Mesure d’association

La plupart des mesures d’association ne sont pas dans BlueSky. Je n’ai vu que le chi-carré. Vous pouvez utiliser d’autres packages pour vous aider.

Installer les packages suivant dans BlueSky: - **summarytools** 9https://cran.r-project.org/web/packages/summarytools/vignettes/introduction.html) et - **DescTools** (<https://www.rdocumentation.org/packages/DescTools/versions/0.99.44>)

* En utilisant DescTools, calculer les statistiques suivantes:

1. A partir de la réponse 1.8 ou 2.5, calculer les mesures d’intensité suivantes:

* le coefficient de contingence de Pearson
* le V de Cramer
* le phi
* le coefficient de prédiction de Guttman
* Gamma de Goodman et de Kruskal
* le Dyx de Somers
* le tau-b et
* le tau-c

1. Maintenant, analyser l’association entre la variable groupe\_age créée au 2.7 et l’employabilité.

* Présenter le tableau qui montre cette association
* Quel est le degré de liberté de ce tableau?
* Est-ce que cette association est significative au seuil de 1%?
* Calculer les mesures d’intensité suivantes:
  + le coefficient de contingence de Pearson
  + le V de Cramer
  + le phi
  + le coefficient de prédiction de Guttman
  + Gamma de Goodman et de Kruskal
  + le Dyx de Somers
  + le tau-b et
  + le tau-c