

Questions

1. Interprétez le tableau de contingence en termes de sur- ou sous-représentations.
2. Calculez à la main une statistique du chi2 de cellule dont on peut penser *a priori* qu'elle est élevée (justifiez votre choix).
3. Interprétez la valeur de la statistique du chi2 globale en comparant aux quantiles de la loi du chi2 avec le bon nombre de degrés de liberté.
4. Interprétez la p-valeur du test.
5. Rejetez-vous l'hypothèse d'indépendance entre les variables au seuil de 5 %, de 1 % ?

Exercice 1

1. Il y a une surreprésentation des enquêtrices parmi les femmes, dans la mesure où elles représentent 22,29 % des femmes contre 16,24 % de l'ensemble des salariés de l'Insee.
2. Les cellules dont la valeur de chi2 est la plus élevée sont *a priori* celles présentant un écart important à la situation d'indépendance et un effectif élevé.

sexe	categorie	n ^{obs}	n ^{exp}	cell_chi2
Femme	Enquêtrice	704	512,91	71,19
Femme	A	579	753,86	40,56

3. La valeur de la statistique du chi2 est de 368,2542, à comparer aux quantiles d'une loi du chi2 à $(2 - 1) * (5 - 1) = 4$ degrés de liberté. Le quantile à 99,9 % d'une loi du chi2 à 4 degrés de liberté vaut 18,47 : $368,2542 > 18,47$ aussi on peut rejeter l'hypothèse nulle d'indépendance entre les variables au seuil de 0,1 %.
4. La p-valeur du test du chi2 est inférieure à 0,01 %, aussi on peut rejeter l'hypothèse nulle d'indépendance entre les variables au seuil de 0,01 %.
5. Le test du chi2 conduit à rejeter l'hypothèse nulle d'indépendance au seuil de 0,01 % : on la rejette donc *a fortiori* aux seuils de 1 % et 5 %. On peut donc considérer qu'il existe une relation statistiquement significative entre sexe et catégorie d'emploi à l'Insee.

Exercice 2

1. Il y a une surreprésentation des actifs occupés parmi les diplômés du supérieur, dans la mesure où ils représentent 69,29 % des diplômés du supérieur contre 47,20 % de l'ensemble de la population.
2. Les cellules dont la valeur de chi2 est la plus élevée sont *a priori* celles présentant un écart important à la situation d'indépendance et un effectif élevé.

ACTEU	sup	n ^{obs}	n ^{exp}	cell_chi2
Actif occupé	1	282	192,12	42,05
Actif occupé	0	545	634,88	12,73

3. La valeur de la statistique du chi2 est de 108,0529, à comparer aux quantiles d'une loi du chi2 à $(3 - 1) * (2 - 1) = 2$ degrés de liberté. Le quantile à 99,9 % d'une loi du chi2 à 2 degrés de liberté vaut 13,82 : $108,0529 > 13,82$ aussi on peut rejeter l'hypothèse nulle d'indépendance entre les variables au seuil de 0,1 %.

4. La p-valeur du test du chi2 est inférieure à 0,01 %, aussi on peut rejeter l'hypothèse nulle d'indépendance entre les variables au seuil de 0,01 %.

5. Le test du chi2 conduit à rejeter l'hypothèse nulle d'indépendance au seuil de 0,01 % : on la rejette donc *a fortiori* aux seuils de 1 % et 5 %. On peut donc considérer qu'il existe une relation statistiquement significative entre diplôme et position sur le marché du travail.

Remarque : on est ici au stade de la corrélation, il est vraisemblable qu'un effet de génération affectant conjointement la part de diplômés et le taux d'activité perturbe l'analyse. Une analyse trivariée avec l'âge comme troisième variable serait ici judicieuse.