

## EXERCICES SUR LE CHAPITRE 4 : données catégorielles

### Comparaison de 2 proportions (données pairées)

### Association entre 2 variables catégorielles

#### FORMULES

#### Comparaison de 2 proportions (données pairées) :

- $(\pi_2 - \pi_1 | \text{données}) \approx N\left(\frac{n_{21} - n_{12}}{n}, \frac{s_D^2}{n}\right)$  avec  $s_D^2 = \frac{n_{21} + n_{12}}{n} - \frac{(n_{21} - n_{12})^2}{n^2}$
- $\frac{n_{21} - n_{12}}{n} \pm 1.96 \sqrt{\frac{s_D^2}{n}}$
- Test de McNemar :  $\frac{(n_{21} - n_{12})^2}{n_{21} + n_{12}} < X_1^2(0.95)$

#### Association entre 2 variables catégorielles :

- **Test d'indépendance**  $G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J O_{ij} \log\left(\frac{O_{ij}}{A_{ij}}\right) \leq X_{(I-1)(J-1)}^2(0.95)$   
avec  $O_{ij}$  et  $A_{ij}$  valeurs observées et attendues respvt
- **Quantification de la force d'association** :  $\log(RR) = \log\left(\frac{p_1}{p_2}\right) \pm 1.96 \sqrt{\frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}}}$

#### EXERCICES À FAIRE À LA SÉANCE DE TP

##### Comparaison de 2 proportions (données pairées)

1. Un sondage auprès de 50 élèves d'un lycée (comportant plusieurs milliers d'élèves) a été effectué avant et après la remise des résultats de juin. Deux questions leur ont été posées. La première était posée avant la remise des résultats et leur demandait s'ils pensaient avoir réussi leur année. 40 élèves ont répondu « oui ». La seconde question était posée après la remise des résultats et leur demandait s'ils avaient réussi leur année : 18 élèves ont répondu non. Seuls 28 élèves avaient estimé correctement avoir réussi leur année.
  - a. Ces données suggèrent-elles que, dans le lycée, la proportion de personnes pensant réussir leur année est égale à la proportion réelle de réussites?
  - b. Donnez un ensemble de valeurs plausibles pour cette différence de pourcentages.
  - c. Répondez à la question a) en utilisant un test.
2. Une étude cross-over a pour objectif de comparer l'efficacité de deux médicaments contre les migraines. Codons le succès du médicament par 1 et l'échec par 0. Pour un premier groupe de 50 sujets, le traitement A est administré à la première migraine et le traitement B à la seconde. Les résultats pour cette séquence A-B sont les suivants: 6 ont (1-1), 25 ont (1-0), 10 ont (0-1) et 9 ont (0-0). Pour le deuxième groupe de 50 sujets les traitements A et B sont administrés dans l'ordre inverse. Parmi ces 50 nouveaux sujets, les résultats pour la séquence B-A sont les suivants: 10 ont (1-1), 20 ont (0-1), 12 ont (1-0) et 8 ont (0,0). En oubliant l'ordre d'administration des traitements c'est-à-dire en supposant que le temps entre les deux migraines est suffisamment long pour que le premier traitement n'ait plus d'effet,

- a. Déterminez s'il y a lieu penser que les deux traitements ont la même efficacité à partir d'un intervalle de crédibilité? Explicitez votre conclusion.
- b. Déterminez s'il y a lieu penser que les deux traitements ont la même efficacité en utilisant un test adéquat.

### Association entre 2 variables catégorielles

3. Dans l'étude de Framingham (USA), on a suivi une cohorte d'hommes âgés de 35 à 44 ans, classés en deux groupes à risque en fonction de leur taux de cholestérol de départ (soit  $>2,5$  g/l, soit  $\leq 2,5$  g/l). L'étude s'est intéressée à la survenue de l'infarctus du myocarde. 603 hommes sont considérés et les résultats sont les suivants : parmi 135 hommes ayant un taux de cholestérol de départ  $> 2,5$  g/l, 10 ont eu un infarctus du myocarde. De plus, 447 hommes ont un taux de cholestérol de départ  $\leq 2,5$  g/l et n'ont pas eu un infarctus du myocarde.
  - a. Effectuez une analyse descriptive des données. (suggestion: estimez les pourcentages conditionnels en ligne). Représentez vos résultats graphiquement.
  - b. À partir de cette étude, peut-on déterminer si le cholestérol influence l'occurrence de l'infarctus du myocarde?
  - c. Estimez et donnez un intervalle de crédibilité pour le rapport de risques. Que peut-on en conclure ?
  - d. Estimez et donnez un intervalle de crédibilité pour la différence de risques. Que peut-on en conclure?
  
4. Lors d'une étude sur les carences en iode ( $\mu\text{g/l}$ ) d'enfants mis en crèches, on a voulu déterminer également s'il existait un lien entre les habitudes alimentaires et le niveau de carence en iode des mères. Pour cela, chaque mère impliquée dans l'étude a rempli un questionnaire portant sur ses habitudes alimentaires. Il y avait notamment la question suivante: « A quelle fréquence mangez-vous du poisson? »(Jamais, au plus 1 fois/mois,  $>1$  fois/mois). Voici les résultats obtenus:

Consommation de Poisson	Carence		
	Absence	Modérée	Sévère
Jamais	3	12	15
Au plus 1 fois/mois	12	60	15
Plus d'1 fois/mois	60	96	27

- a. Effectuez une analyse descriptive des données. (suggestion: estimez les pourcentages conditionnels en ligne). Représentez vos résultats graphiquement.
- b. Peut-on, à partir de ces données, suggérer que la consommation de poisson influence le niveau de carence?
- c. Un médecin s'intéresse uniquement au risque d'une carence élevée lorsqu'on ne mange pas de poisson (respectivement: lorsqu'on en mange au plus 1 fois par mois). A l'aide de rapports, comparez ces deux risques avec le risque de carence lors d'une consommation  $>1$  fois/mois. Que peut-on en conclure?

## EXERCICES COMPLÉMENTAIRES

### Comparaison de 2 proportions (données pairées)

1. On a demandé à 200 personnes, choisies au hasard parmi de nombreux chefs d'entreprise, si elles approuvaient la politique de leur gouvernement: 35% des personnes questionnées la désapprouvent. Quelques mois plus tard, la monnaie du pays est violemment attaquée sur les marchés des changes et subit une dévaluation de 5%. On repose alors la même question aux mêmes chefs d'entreprise: désormais, 60% approuvent la politique du gouvernement. De plus, 50 des 200 personnes questionnées désapprouvent l'action du gouvernement avant et après la dévaluation de la monnaie.
  - a. Quelle est la probabilité que la crise monétaire ait augmenté le pourcentage des chefs d'entreprises approuvant la politique du gouvernement ?
  - b. Donnez un ensemble de valeurs plausibles pour la différence de pourcentages concernée.
  - c. Peut-on dire que l'action du gouvernement durant la crise a significativement modifié la satisfaction des chefs d'entreprises ? Répondez à cette question à l'aide d'un test adéquat.

### Association entre 2 variables catégorielles

2. Un grand fabricant de cigarettes désire savoir si le fait de fumer est lié à la consommation de chocolat. À cette fin, une enquête a été réalisée auprès de 1000 personnes. Parmi 600 mangeurs de chocolat, 200 personnes ne fument pas. De plus, 100 personnes ne consomment ni chocolat, ni cigarette.
  - a. Effectuez une analyse descriptive des données. Représentez vos résultats graphiquement.
  - b. À partir de ce sondage, peut-on déterminer si le fait de fumer est lié à la consommation de chocolat?
  - c. Dans la population concernée, comparez les proportions de consommateurs de chocolat chez les fumeurs et chez les non-fumeurs de deux manières différentes.