

Épreuve partielle #1

RÉPONDRE À TOUTES LES QUESTIONS.

(20 points)

1) Les cinq énoncés suivants s'appliquent à un test d'adéquation basé sur le khi-deux de Pearson X^2 . Qualifier chaque énoncé de vrai ou de faux en justifiant brièvement votre réponse.

- a) Lorsque le seuil observé est 0.03, l'hypothèse nulle est rejetée au seuil 0.02.
- b) Pour un test de seuil 0.06, le seuil observé est ≤ 0.06 .
- c) Le seuil observé est la probabilité que l'hypothèse nulle soit correcte.
- d) Supposons que $X^2 = 8.5$ avec 4 degrés de liberté. Alors le seuil observé est ≤ 0.05 .
- e) Le seuil observé est une variable aléatoire indépendante du seuil α .

(20 points)

2) Soit X le nombre de garçons dans une famille de deux enfants. On désire vérifier si ce nombre suit une loi binomiale $\mathcal{B}(2, p)$, où $p = P(\text{garçon})$ est une probabilité à déterminer. À partir d'un échantillon aléatoire de 120 familles de deux enfants, on a observé les fréquences suivantes.

X	0	1	2
Fréquence	24	64	32

- a) La fonction de vraisemblance peut s'écrire

$$L(p) = \prod_{j=1}^{120} \binom{2}{x_j} p^{x_j} (1-p)^{2-x_j}, \quad 0 < p < 1.$$

Posons $N_i = \#\{x_j = i : j = 1, \dots, 120\}$, $i = 0, 1, 2$. Réexprimer $L(p)$ en termes des N_i .

- b) Montrer que l'estimateur de vraisemblance maximale \hat{p} est égal à $8/15$.
- c) On souhaite tester l'adéquation du modèle binomial pour X avec le khi-deux de Pearson. Comment calcule-t-on les fréquences espérées?
Rappel. Les N_i sont binomiales.
- d) On fait le test d'adéquation au seuil $\alpha = 0.05$.

- 1. Donner une formule exprimant le seuil observé.
- 2. Quelle est la décision ?

(20 points)

3) Un total de 960 matches ont été disputés dans la Ligue nationale de hockey pendant la saison 92-93. Pour chaque match, on dispose du nombre total X de buts comptés. Il est raisonnable de penser qu'un modèle approprié pour X est celui de Poisson.

- Expliquer de manière suffisamment détaillée comment tester l'adéquation du modèle de Poisson. Bien préciser l'hypothèse H_0 , les distributions en jeu et la manière de faire les calculs.
- On dispose des résultats des 840 matches de la saison 84-85. En supposant que le modèle de Poisson soit valide pour X , comment pourrait-on tester l'hypothèse qu'un modèle de Poisson de même paramètre s'applique aux deux saisons ? Encore une fois, préciser l'hypothèse testée, les distributions en jeu et la manière de faire les calculs.

N.B. On rappelle que la fonction de masse de la loi de Poisson de paramètre $\lambda > 0$ s'écrit

$$f(x|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{sinon.} \end{cases}$$

(20 points)

4) Y a-t-il un lien entre le niveau de revenu des parents et le niveau de scolarité de leurs enfants ? Pour répondre à cette question, on dispose des données suivantes recueillies chez 185 jeunes adultes ayant terminé leurs études.

	Niveau de scolarité			
Revenu des parents en \$	secondaire	cégep	université	Total
< 15 000	28	8	2	38
15 000 – 35 000	27	25	10	62
35 000 – 60 000	12	20	15	47
> 60 000	3	15	20	38
Total	70	68	47	185

- Quel type d'échantillonnage a produit ce tableau ?
- Répondre à la question posée au début. On pourra faire l'inférence au seuil $\alpha = 0.05$.
- Supposons que les 185 sujets proviennent de 4 échantillons indépendants, soit un pour chaque catégorie de revenu. Brièvement, quelle analyse statistique serait appropriée ?

(20 points)

5) Le tableau de fréquences suivant rend compte d'un sondage effectué auprès de quelques étudiants d'université (120 hommes, 120 femmes) afin de savoir s'ils sont favorables à une grève (Oui) ou opposés à celle-ci (Non).

Sexe	Faculté	Grève	
		Oui	Non
masculin	Sciences et génie	25	75
	Sciences humaines	15	5
féminin	Sciences et génie	5	15
	Sciences humaines	80	20

- Utiliser le rapport de cotes pour faire une analyse descriptive de ces données. Il s'agit ici de décrire ce que suggèrent les rapports de cote, sans parler de test ni d'intervalle de confiance. On considérera le tableau marginal aussi bien que les tableaux conditionnels.
- Reprendre maintenant l'analyse à la lumière du programme et de la sortie SAS ci-après.

Jean-Claude Massé
Professeur

```
data exemple;
input sexe faculte greve frequency @@;
cards;
1 1 1 25 1 1 2 75 1 2 1 15 1 2 2 5
2 1 1 5 2 1 2 15 2 2 1 80 2 2 2 20
;
proc freq order=data;weight frequency;
tables sexe*faculte*greve/ chisq measures cmh;
run;
proc freq order=data;weight frequency;
tables faculte*greve/ chisq measures cmh;
run;
*****
```

Statistics for Table 1 of faculte by greve

Controlling for sexe=1

Statistic	DF	Value	Prob
Chi-Square	1	18.7500	<.0001
Likelihood Ratio Chi-Square	1	17.8030	<.0001
Mantel-Haenszel Chi-Square	1	18.5938	<.0001

Fisher's Exact Test

Cell (1,1) Frequency (F)	25
Left-sided Pr <= F	3.646E-05
Right-sided Pr >= F	1.0000
Two-sided Pr <= P	3.646E-05

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.1111	0.0367	0.3367
Cohort (Col1 Risk)	0.3333	0.2183	0.5090
Cohort (Col2 Risk)	3.0000	1.3925	6.4630

Statistics for Table 2 of faculte by greve

Controlling for sexe=2

Statistic	DF	Value	Prob
Chi-Square	1	24.4034	<.0001
Likelihood Ratio Chi-Square	1	22.2990	<.0001
Mantel-Haenszel Chi-Square	1	24.2000	<.0001

Fisher's Exact Test

Cell (1,1) Frequency (F)	5
Left-sided Pr <= F	3.911E-06
Right-sided Pr >= F	1.0000
Two-sided Pr <= P	3.911E-06

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.0833	0.0271	0.2566
Cohort (Col1 Risk)	0.3125	0.1454	0.6718
Cohort (Col2 Risk)	3.7500	2.3518	5.9794

Summary Statistics for faculte by greve

Controlling for sexe

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	42.4931	<.0001
2	Row Mean Scores Differ	1	42.4931	<.0001
3	General Association	1	42.4931	<.0001

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	0.0968	0.0439	0.2132
(Odds Ratio)	Logit	0.0964	0.0438	0.2123

Breslow-Day Test for

Homogeneity of the Odds Ratios

Chi-Square	0.1277
DF	1
Pr > ChiSq	0.7208

Statistics for Table of faculte by greve

Statistic	DF	Value	Prob
Chi-Square	1	70.5391	<.0001
Likelihood Ratio Chi-Square	1	74.5158	<.0001
Mantel-Haenszel Chi-Square	1	70.2452	<.0001

Fisher's Exact Test

Cell (1,1) Frequency (F)	30
Left-sided Pr <= F	1.049E-17
Right-sided Pr >= F	1.0000
Two-sided Pr <= P	2.097E-17

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.0877	0.0479	0.1605
Cohort (Col1 Risk)	0.3158	0.2286	0.4363
Cohort (Col2 Risk)	3.6000		

The FREQ Procedure

Summary Statistics for faculte by greve

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	70.2452	<.0001
2	Row Mean Scores Differ	1	70.2452	<.0001
3	General Association	1	70.2452	<.0001

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	0.0877	0.0479	0.1605
	Logit	0.0877	0.0479	0.1605