

Corrigé 11

Problème 1

Une émission de radio permet à ses auditeurs de s'exprimer. Afin de gérer les appels, l'animateur désire établir un modèle sur le nombre d'appels par heure à son émission. Il veut vérifier si le nombre d'appels est distribué selon une distribution de Poisson de moyenne 5 appels/heure. Pour cela, il a collecté les données suivantes lors des 100 dernières heures de ses émissions :

Nombre d'appels /heure	0	1	2	3	4	5	6	7	8	≥ 9
Fréquence	0	2	4	10	7	14	11	13	5	34

```
observe <- c(6, 10, 7, 14, 11, 13, 5, 34)
x <- (2:9)
theorique <- dpois(x, lambda=5)
theorique[1] <- ppois(2, lambda=5)
theorique[length(x)] <- 1-ppois(8, lambda=5)
chisq.test(observe,p=theorique)
# Valeurs théoriques
chisq.test(observe,p=theorique)$expected
```

- En tant que consultant, aidez-le à effectuer le test avec un niveau de signification de 0.05.

Comme la fréquence par cellule doit être au moins 5, il est nécessaire de grouper les cellules 0,1 et 2.

Nombre d'appels /heure	≤ 2	3	4	5	6	7	8	≥ 9
Fréquence observée	6	10	7	14	11	13	5	34
Fréquence théorique	12.46	14.04	17.55	17.55	14.62	10.44	6.53	6.81

Les fréquences théoriques sont trouvées par $100 * Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$, sachant que $X \sim \mathcal{P}(5)$. Pour la dernière cellule, il suffit de calculer $100 * (1 - F_X(8))$

La règle de décision est la suivante : si la statistique χ^2 est plus grande que la valeur critique, alors H_0 est rejeté.

Valeur critique : $\chi_{0.05}^{(7)} = 14.0671$

La statistique vaut $\chi^2 = 122.02$. Comme $\chi^2 = 122.02 > 14.0671 = \chi_{0.05}^{(7)}$, l'hypothèse nulle est rejetée. Ainsi, le nombre d'appels ne suit pas une loi de Poisson de moyenne 5 appels par heure.

- Quel type d'erreur pourriez-vous commettre ?

L'hypothèse nulle étant rejetée, une erreur de première espèce pourrait avoir été commise

Problème 2

Les amendes d'ordre des automobilistes à Genève sont représentées ci-dessous en fonction du sexe du conducteur.

Infraction	Homme	Femme
Vitesse	240	160
Parcage	80	40
Feux grillés	32	18
Service anti-pollution	11	9
Autres	5	4

Utilisez un niveau de signification de 0.05 pour déterminer si les deux variables sont indépendantes.

La valeur critique vaut $\chi_{0.05}^{(4)} = 9.4877$

La statistique $\chi^2 = 2.356$

Comme la valeur de la statistique est inférieure au seuil critique, l'hypothèse nulle n'est pas rejetée.

Ainsi, on peut supposer que les deux variables sont indépendantes.

Note : comme l'hypothèse nulle n'est pas rejetée, aucune erreur de type I n'est commise. Par conséquent, il est inutile de s'inquiéter au sujet de la fréquence inférieure à 5 de la dernière cellule du tableau.

Valeurs théoriques :

```
amendes <- matrix(c(240,160,80,40,32,18,11,9,5,4),ncol=2,byrow=TRUE)
colnames(amendes)<-c("Homme","Femme")
rownames(amendes)<-c("Vitesse","Parcage","Feux grillé","Service AP", "Autre")
# Valeurs théoriques
res <- chisq.test(amendes)
rowSums(amendes); colSums(amendes); sum(amendes)
res$expected; res$residuals^2;
res
```

	Vitesse	Parcage	Feux grillé	Service AP	Autre
	400	120	50	20	9
Homme	Femme				
368	231				
[1]	599				

	Homme	Femme
Vitesse	245.742905	154.257095
Parcage	73.722871	46.277129
Feux grillé	30.717863	19.282137
Service AP	12.287145	7.712855

Autre	5.529215	3.470785
	Homme	Femme
Vitesse	0.13420919	0.21380512
Parcage	0.53446566	0.85144312
Feux grillé	0.05351528	0.08525378
Service AP	0.13483546	0.21480281
Autre	0.05065256	0.08069325

Pearson's Chi-squared test

data: amendes

X-squared = 2.3537, df = 4, p-value = 0.671