LSTAT2100 - Exercices - Série 1

Énoncés

Exercice 1

Nous nous intéressons à la couleur des yeux d'une certaine population. Pour $n = 300$ individus, pris au hasard, nous avons observé les chiffres suivants.

bleu	marron	noir	vert
48	122	95	35

(a) Ces couleurs sont-elles toutes réparties de manière uniforme (équiprobables) ? Répondez à cette question eu utilisant le test de LR. Calculez la $p-$valeur de ce dernier en utilisant (i) la théorie asymptotique, et (ii) des simulations.

(b) Peut-on dire que les yeux foncés (marron et noir) sont deux fois plus probables que les yeux clairs (bleu et vert) ? Utilisez le test de Pearson.

(c) Les proportions des yeux bleus et des yeux verts sont-elles les mêmes ? Utilisez le test LR. Détaillez votre approche et vos calculs.

Exercice 2

Voici les fréquences ($Freq$) du nombre de passages de bus ($nBus$) par heure à un point d'arrêt. Ces données concernent une période de 30 heures réparties sur 5 jours de la semaine.

nBus	0	1	2	3	4	5
Freq	1	5	6	10	4	4

(a) Supposons que $nBus \sim Pois(\mu)$, $\mu > 0$. Testez l'hypothèse $H_0 : \mu = 3$ vs H_1: $\mu \neq 3$. Pour ce faire, utilisez les trois tests classiques vus dans le cours. Calculez un intervalle de confiance pour μ.

(b) En supposant que la fréquence moyenne de passage est de $\mu = 3$ par heure, testez l'ajustement d'une loi de Poisson ($H_0 :$ $nBus \sim Pois(3)$) aux données ? *Proposez un test pertinent.*

(c) Refaites le même test que celui de la question précédente mais, cette fois, sans supposer que $\mu = 3$; càd testez $H_0 :$ $nBus \sim Pois(\mu)$, pour une μ fixe mais inconnue.

Exercice 3

Dans cet exercice, nous allons utiliser le jeu de données mdata.csv. Ce jeu de données comprend des observations liées à des clients qui ont contracté un crédit.

`mdata.csv` contient de nombreuses variables, mais nous n'en utiliserons que quelques-unes dans cette série et dans la série qui suit. Parmi ces variables, il y a

- `repay`: variable binaire prenant les valeurs “Default” ou “NotInDefault” pour un crédit remboursé ou non à temps.
- `account`: flux mensuels moyens sur le compte courant du client; “<0”, “[0-200)”, “>=200”, ou “No acc”, si le client n’a pas de compte courant auprès de la banque.
- `tel`: le client dispose-t-il d’un numéro de téléphone fixe ? “Yes” ou “No”.
- `employm`: la situation professionnelle du client en termes de durée de l’emploi (actuel) en années; “No or <1”, pour un client sans-emploi ou qui travaille mais depuis moins d’un an, “[1-4)”, “[4-7)” , ou “>=7”, pour un client qui exerce son activité depuis au moins 7 ans.

Charger et examiner les données. Pour ce faire, vous pouvez utiliser la commande suivante (on présume que votre répertoire de travail comprend `Data/mdata.csv`).

```r
mdata <- read.csv(file = "Data/mdata.csv")
str(mdata)
```

(a) Donnez le tableau des fréquences pour `repay`. Représentez ce tableau à l’aide d’un graphique approprié.

(b) Construisez la table de contingence croisant les variables `repay` et `account`, et estimez les proportions $P(repay|account)$, pour les différentes valeurs de `(repay, account)`.

(c) Faites un graphique pour représenter la distribution (marginale) de `account` et la distribution (conditionnelle) de `repay|account`. Que suggère ce graphique quant à l’association entre ces deux variables ?

(d) Testez l’indépendance entre `account` et `repay`.

(e) Calculez le rapport de cotes entre `tel` et `repay`. Quelle information peut-on en tirer sur l’association entre ces deux variables ? Compléter par un test statistique (d’indépendance) fondé sur le ratio calculé.

Exercice 4

Soit le tableau de contingence suivant

X/Y	1	2
1	5	48
2	34	251

La signification des variables X et Y n’est pas importante pour la suite.

(a) Testez l’indépendance entre X et Y.

(b) Soit $p_{ij} = P(X = i, Y = j)$, $i, j = 1, 2$. Considérons l’hypothèse suivante

$$H_0 : \; p_{11} = \theta^2, p_{12} = p_{21} = \theta(1 - \theta), \text{ et } p_{22} = (1 - \theta)^2$$

Montrez que sous H_0, X et Y sont indépendantes et identiquement distribuées.

(c) En supposant un échantillonnage multinomial simple, donnez l’estimateur du maximum de vraisemblance de θ et calculez-le.

(d) Proposez une statistique de test pour tester H_0. Effectuez le test et concluez.