Énoncés des Travaux dirigés de Statistiques

Julian Tugaut[1]

1. Si vous trouvez des erreurs de Français ou de mathématiques ou bien si vous avez des questions et/ou des suggestions, envoyez-moi un mail à julian.tugaut@univ-st-etienne.fr

Exercice 1

À la suite d'une interrogation portant sur 40 étudiants, on a relevé les notes suivantes :

$$14, 14, 6, 14, 11, 10, 15, 7, 14, 9, 9, 13, 12, 9, 10, 15, 8, 10, 16, 8,$$
$$12, 9, 7, 11, 10, 12, 11, 16, 11, 10, 10, 9, 7, 10, 13, 10, 9, 9, 16, 12\,.$$

1. Grouper les valeurs de cette série pour la présenter sous la forme d'une série à valeurs isolées. Déterminer également les fréquences, les effectifs cumulés et les fréquences cumulées.
2. Construire le diagramme en bâtons des effectifs ainsi que celui des effectifs cumulés.
3. Calculer la moyenne arithmétique.
4. Déterminer graphiquement la médiane, le premier quartile et le troisième quartile.
5. Déterminer le(s) mode(s).
6. Déterminer la variance et l'écart-type.

Exercice 2

Tracer l'histogramme de la série suivante de mille salaires :

Classe de salaires	Nombre de salaires
$[800, 1400[$	100
$[1400, 1800[$	200
$[1800, 2200[$	350
$[2200, 3000[$	250
$[3000, 5000[$	100

Exercice 3

On a noté sur 6 800 individus les couleurs des cheveux et des yeux. On a construit le tableau de contingence :

Table 1 – Tableau de contingence

couleur des cheveux / couleur des yeux	blond	châtain	noir	roux
bleu	1 768	807	189	47
gris ou vert	946	1 387	746	53
marron	115	438	288	16

Calculer les fréquences conditionnelles des couleurs des yeux par rapport aux couleurs des cheveux dans cet échantillon. La couleur des yeux est-elle indépendante de celle des cheveux dans cet échantillon ?

Exercice 4

On suppose qu'un estimateur T_n d'un paramètre $\theta \in \mathbb{R}$ a pour loi : $\left(1 - \frac{1}{n}\right)\delta_\theta + \frac{1}{n}\delta_n$.

1. Cet estimateur est-il convergent ?

2. Calculer l'erreur quadratique moyenne de T_n.

Exercice 5

On considère une variable aléatoire discrète X dont la loi de probabilité est définie par

$$\mathbb{P}(X = k) = \frac{\theta^{k-1}}{(1+\theta)^k}, \quad \forall k \in \mathbb{N}^*.$$

Ici, θ est un paramètre strictement positif que l'on ne connait pas.

1) Calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.

2) On se propose d'estimer le paramètre θ à partir d'une réalisation numérique $(x[1], \cdots, x[n])$ d'un n-échantillon $(X_1, \cdots, X_n)$ de la loi de X.

(a) Trouver l'estimateur du maximum de vraisemblance $\widehat{\Theta}_n$ du paramètre θ.

(b) L'estimateur $\widehat{\Theta}_n$ est-il sans biais ?

Exercice 6

On considère un échantillon issu d'une variable aléatoire de densité

$$f(x) := \frac{1}{\theta}\exp\left(-\frac{x-\gamma}{\theta}\right)\mathbb{1}_{[\gamma;+\infty[}(x),$$

où $\theta > 0$ et $\gamma \in \mathbb{R}$ sont deux paramètres inconnus.

1) Déterminer l'estimateur du maximum de vraisemblance des paramètres θ et γ.

2) Déterminer un estimateur des paramètres θ et γ par la méthode des moments.

Exercice 7

On considère un n-échantillon $(X_1, \cdots, X_n)$ issu de la loi $p\mathcal{U}_{[0;a]} + (1-p)\mathcal{U}_{[0;b]}$, c'est-à-dire que X_i suit la loi uniforme sur $[0;a]$ avec probabilité p et elle suit la loi uniforme sur $[0;b]$ avec probabilité $1-p$. On suppose $0 < a < b$ avec a et b connus et fixés.

1) Déterminer la fonction de répartition de X_1.

2) Déterminer sa densité.

3) Soit N_a la variable aléatoire égale au nombre d'individus compris entre 0 et a. Quelle est la loi de N_a ? En déduire son espérance et sa variance.

4) Le paramètre p est inconnu. Déterminer l'estimateur du maximum de vraisemblance : $\widehat{p}$.

Exercice 8

Soient $(X_i)_i$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{U}_{[0;\theta]}$. Construire un intervalle de confiance asymptotique sur le paramètre θ centré sur le paramètre $\frac{\theta}{2}$.

Exercice 9

Soient $(X_i)_i$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{U}_{[-\theta;\theta]}$. Construire un intervalle de confiance asymptotique sur le paramètre θ, de la forme $[c;+\infty[$.

Exercice 10

On effectue une enquête, durant une épidémie de grippe, dans le but de connaître la proportion p de personnes présentant ensuite des complications graves. On observe un échantillon représentatif de 400 personnes et pour un tel échantillon, 40 individus ont présenté des complications.

1. Donner un intervalle de confiance pour p au risque 5%.

2. On désire que la valeur estimée $\hat{p}$ diffère de la proportion inconnus exacte p de moins de 0.005 avec une probabilité égale à 95%. Quel sera l'effectif d'un tel échantillon ?

Exercice 11

On observe le nombre de filles dans 320 fratries de cinq enfants :

Nombre de filles	Nombre de fratries
0	18
1	56
2	110
3	88
4	40
5	8

1) *En supposant que le genre d'un enfant suit une loi de Bernoulli de paramètre* $\frac{1}{2}$, *quelle est la loi du nombre de filles dans une fratrie de cinq enfants ?*

2) *La distribution du nombre de filles suit-elle une loi binomiale de paramètres* $m := 5$ *et* $p := \frac{1}{2}$ *au seuil de risque de* 5% *? Conclure.*

Exercice 12

Au départ d'une course de chevaux, il y a habituellement huit positions de départ et la position numéro un est la plus proche de la palissade. On soupçonne qu'un cheval a plus de chances de gagner quand il porte un numéro faible, c'est-à-dire quand il est plus proche de la palissade intérieure. Voici les données de 144 courses :

Numéro de départ	1	2	3	4	5	6	7	8
Nombre de victoires d'un cheval ayant ce numéro	31	17	16	27	15	12	13	13

L'hypothèse nulle (H_0) est l'hypothèse d'équiprobabilité des numéros de départ dans les places de premier à l'arrivée.

Nous prendrons comme hypothèse alternative (H_1) la non équiprobabilité des numéros de départ.

Nous supposons que l'échantillon de 144 courses est un échantillon aléatoire.

Tester l'hypothèse (H_0) contre l'hypothèse (H_1) au risque d'erreur 5%.

Exercice 13

On a enregistré, pendant $n = 200$ jours, le nombre d'appels téléphoniques qui ont eu lieu entre 18h00 et 18h15 dans un standard téléphonique. On a obtenu les résultats suivants :

Nb d'appels	0	1	2	3	4	5	6	7	8	9	10	11	12
Nb de jours	1	16	31	37	41	30	23	13	6	1	0	1	0

1) Calculer la moyenne et la variance de cette distribution.
2) Par quelle loi de probabilité peut-on ajuster la distribution observée ?
3) Quel test statistique faut-il utiliser pour vérifier le bien-fondé de ce choix ? Quelle est la conclusion, au risque de 5% ?

Exercice 14

Dans une région tropicale, sur une population P d'individus porteurs de parasitose, on a recensé : 52% porteurs de paludisme ; 16.2% porteurs de bilharziose ; 3.8% porteurs de filariose ; 4.2% porteurs de toxoplasmose et 23.8% porteurs d'amibiase.

À la suite d'une action sanitaire entreprise dans cette région, on examine un échantillon de 200 malades porteurs de parasitoses et on observe : 88 porteurs de paludisme ; 38% porteurs de bilharziose ; 12% porteurs de filariose ; 6% porteurs de toxoplasmose et 56% porteurs d'amibiase.

Peut-on considérer, au coefficient de sécurité de 99%, que l'action sanitaire entreprise a modifié la répartition des parasitoses ?

Exercice 15

On a mesuré la taille (x) et le poids (y) de 300 adultes extraits par un tirage au hasard non exhaustif d'une population. Les données sont présentées dans le tableau de contingence suivant :

Table 2 – Tableau de contingence (poids/taille)

Poids (kg) \ Taille (cm)	[150; 160[	[160; 170[	[170; 180[	[180; 190[	[190; 200[
[45; 55[	2	1	0	0	0
[55; 65[	7	8	4	2	0
[65; 75[	5	15	22	7	1
[75; 85[	2	12	63	19	5
[85; 95[	0	7	28	32	12
[95; 105[	0	2	10	20	7
[105; 115[	0	0	1	4	2

Déterminer l'équation de la droite de régression au sens des moindres carrés du caractère y par rapport au caractère x dans cet échantillon.