STT1000 - Exercices # 2

Automne 2021

Tout comme en cours, on utilise les conventions : $Y_1, \cdots, Y_n$ sont des variables aléatoires, $y_1, \cdots, y_n$ sont les observations (données).

Les échantillons aléatoire et observé associés seront notés $\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ et $\boldsymbol{y} = (y_1, \cdots, y_n)$. La notation Z sera (principalement) réservée pour une variable aléatoire suivant une loi normale centrée réduite, $Z \sim \mathcal{N}(0, 1)$.

Certains exercices impliquent des jeux de donnees parfois de grande taille. Ces jeux de donnees sont tous stockes au sein du chier STT1000.RData (disponible sous la plateforme moodle) qu'il convient de charger de la facon suivante :

Exercice 1 – ***(Poids joueurs)***
Le tableau ci-dessus resume les poids exprimes en livres des 45 joueurs qui se sont presentes au camp d'entrainement de l'equipe de football d'une universite Canadienne en juin dernier. Le jeu de donnees brut est stocke dans le vecteur poids et celui range en classe dans le vecteur `cut.poids`

poids	effectif
$[190, 200)$	15
$[200, 210)$	12
$[210, 220)$	9
$[220, 235)$	6
$[235, 255)$	3

1. Dessiner l'histogramme de ces donnees.
2. La variable poids semble-t-elle suivre une loi normale ?
3. Trouver une valeur approximative de la proportion des joueurs dont le poids etait superieur a 230 livres.
4. Trouver une valeur approximative de la moyenne des poids des joueurs.

Exercice 2 – ***(Publicité)***
Lors d'une enquête sur la publicite televisee on a demande a 16 sujets d'evaluer, en minutes, le temps des coupures publicitaires pendant le lm du dimanche soir sur une chaîne. Voici les resultats (stockes dans la variable `pub` en R :

```
[1] 28 29 30 30 31 32 32 32 33 33 33 33 34 34 35 39
```

1. Trouver la mediane, le premier quartile, et le troisieme quartile de ces donnees.
2. Tracer le diagramme en boîte de ces donnees.
3. La loi normale est-elle un modele approprie ?

Exercice 3 – ***(Naissances)***
Le jeu de donnees "presente en annexe (et qui fait partie du fichier `STT1000.RData` sous le nom `naissances`), provient d'un hopital de Brisbane (cote est de l'Australie). Des naissances (44) ont etes enregistrees sur une periode de 24 heures ; pour chacune des naissances, nous observons : Heure de naissance (format hh :mm) ([Heure]) ; Sexe du nouveau ne ([Sexe]) ; Poids a la naissance en grammes ([Poids]) ; Nombre de minutes apres minuit pour chaque naissance ([Temps]).

1. Quels sont les types des variables [Sexe] et [Poids] ?
2. Representez les donnees concernant le sexe a l'aide d'un diagramme en batons et d'un tableau

3. Evaluez la mediane.
4. Realisez un histogramme de la distribution du poids ; choisissez le nombre de classes qui vous semble adequat pour la distribution.
5. Decrivez la distribution de poids par sexe. Estimez l'etendue, l'intervalle interquartile, la mediane et la moyenne de poids par sexe.

Exercice 4 – ***(Naissances - suite)***
Dans cette partie vous aller resoudre les questions a la main, toute fois, vous pouvez utiliser R. Pour minimiser un certain nombre de calculs, des tableaux sont fournis apres chaque item. Vous pouvez vous en servir

1. Vous remarquerez que la proportion des nouveaux nes garçons est plus elevee que la proportion attendue par la loi de la genetique (la probabilite que le sexe d'un nouveau ne soit un garçon ou une fille est environ de 1/2). Est-ce que le resultat observe est si inhabituel que cela ? An repondant a cette question, calculez la probabilite d'observer au moins 26 enfants de meme sexe. On suppose que la naissance d'une fille ou d'un garçon est equiprobable.
2. On peut utiliser la distribution geometrique pour modeliser le nombre de naissances jusqu'a l'obtention d'un garçon (incluant celui-ci). En ne considerant les donnees que jusqu'a la derniere naissance d'un garcon (donc en ne considerant que les 41 premieres observations), construisez un tableau ou vous compterez le nombre d'essais avant d'obtenir un garcon (le tableau en question a la meme forme que le tableau de l'exemple 7 du chapitre 1 ; les lignes du tableau correspondent au nombre d'essais possibles : 1,2,3,4 et 5+. Utilisez ensuite la probabilite observee que le sexe d'un nouveau ne soit un garcon (sur les 41 naissances) durant cette experience pour calculer la probabilite theorique donne selon une loi geometrique. Comparez les probabilites empiriques et theoriques. Est-ce que la loi geometrique vous parait un modele adequat ?

nb naissance jusqu'à la naissance d'un garçon	fréq.	probabilité empirique	probabilité théorique
1	18		
2	3		
3	4		
4	0		
5+	1		
total	26	100%	100&

3. La distribution de Poisson peut etre utilisee pour modeliser le nombre de naissance par heures. Realisez un tableau de facon similaire a la question precedente ou chaque ligne du tableau contient ici le nombre de naissances par heure (0,1,2,3,4,5+). Calculez a partir des donnees la distribution observee. Calculez ensuite le nombre de naissance moyen par heure , et calculez les probabilites theoriques selon une loi de Poisson de parametre . Comparez les probabilites empiriques et theoriques. Est-ce que la loi geometrique vous parait un modele adequat ?

nb naissance par heure	fréq.	probabilité empirique	probabilité théorique
0	3		
1	8		
2	6		
3	4		
4	3		
5+	0		
total	24	100%	100&

4. Apres avoir ajuste la distribution de Poisson aux temps d'arrivee, il est naturel d'essayer d'ajuster la distribution exponentielle aux temps d'attente entre les arrivees. Vous allez encore construire un tableau similaire aux questions precedentes. Ici, chaque ligne du tableau correspondra a un intervalle de temps : [0,19.5 min), [19.5,39 min), [39.58.5 min), [58.5, 78 min), 78+ min. Analysez les donnees pour compter les frequences observees et deduisez-en les probabilites empiriques. Sachant que la premiere naissance arrive a 00h05 et la derniere a 23h55, calculez le temps moyen entre arrivee 1=, et utilisez cet estime pour calculez les probabilites theoriques pour chaque classe. Comparez les probabilites empiriques et theoriques. Est-ce que la loi geometrique vous parait un modele adequat ?

temps entre arrivées	fréq.	probabilité empirique	probabilité théorique
$[0, 19.5)$	18		
$[19.5, 39)$	12		
$[39, 58.5)$	5		
$[58.5, 78)$	6		
78+	2		
total	43	100%	100%