

STT1000 - Exercices # 1

Automne 2021

Tout comme en cours, on utilise les conventions : Y_1, \dots, Y_n sont des variables aléatoires, y_1, \dots, y_n sont les observations (données). Les échantillons aléatoire et observé associés seront notés $\mathbf{Y} = (Y_1, \dots, Y_n)$ et $\mathbf{y} = (y_1, \dots, y_n)$. 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 données parfois de grande taille. Ces jeux de données sont tous stockés au sein du chier STT1000.RData (disponible sous la plateforme moodle) qu'il convient de charger de la façon suivante :

```
1 > url = "http://freakonometrics.free.fr/STT1000.R"
2 > download.file(url, "STT1000.R")
3 > source("STT1000.R")
```

Exercice 1 – (*Iris*)

```
1 > install.packages(MASS)
2 > library(MASS)
3 > data("iris")
```

L'ensemble de données sur l'iris donne les mesures en centimètres des variables longueur et largeur des sépales et longueur et largeur des pétales, respectivement, pour 150 fleurs de chacune des 3 espèces d'iris. Les espèces sont *Iris setosa*, *versicolor* et *virginica*. Le tableau ci-dessus résume les length du sepal 150 iris.

Sepal.Length	effectif
(4, 4.5]	5
(4.5, 5]	27
(5, 5.5]	27
(5.5, 6]	30
(6, 6.5]	31
(6.5, 7]	18
(7, 7.5]	6
(7.5, 8]	6

150

1. Quels sont les types des variables [Species] et [Sepal.Length] ?
2. Dessiner l'histogramme de ces données.
3. Quelle distribution peut on suggerer
4. Trouver une valeur approximative de la moyenne et la variance des poids des Sepal.Length.
5. Utiliser la distribution Normal pour modeliser le Sepal.Length. Realisez le tableau ou chaque ligne du tableau contient les intervalles de length . Calculez a partir des donnees la distribution observee. Calculez ensuite le Sepal.Length., et calculez les probabilites theoriques selon une loi normal. Comparez les probabilites empiriques et theoriques. Est-ce que la loi normal vous parait un modele adequat ?

temps entre arrivées	fréq.	probabilité empirique	probabilité théorique
(4, 4.5]	5		
(4.5, 5]	27		
(5, 5.5]	27		
(5.5, 6]	30		
(6, 6.5]	31		
(6.5, 7]	18		
(7, 7.5]	6		
(7.5, 8]	6		
	150		

note that $F(\frac{-13\sqrt{2}}{12}) = 0.06275324$, $F(\frac{-2\sqrt{2}}{3}) = 0.1728893$, $F(\frac{-\sqrt{2}}{4}) = 0.3618368$, $F(\frac{\sqrt{2}}{6}) = 0.5931681$, $F(\frac{7\sqrt{2}}{12}) = 0.795302$. $F(\sqrt{2}) = 0.9213504$, $F(\frac{17\sqrt{2}}{12}) = 0.9774365$, $F(\frac{11\sqrt{2}}{6}) = 0.9952391$

Exercice 2 – Une expérience a été menée en Suède en 1961-1962 pour évaluer l’effet d’une limitation de vitesse sur le taux d’accidents sur autoroute. L’expérience a été menée sur 92 jours chaque année, appariés de sorte que le jour j en 1962 était comparable au jour j en 1961. Certains jours, la limite de vitesse était en vigueur et appliquée, tandis que d’autres jours il n’y avait pas de limite de vitesse et les voitures avaient tendance à être conduit plus vite. Les jours de limitation de vitesse avaient tendance à être en blocs contigus.

y : nombre d’accidents de la circulation pour ce jour. limit:y avait-il une limitation de vitesse ?(0 is non 1 is oui) Pour la variable y:

1. Tracer le histogram
2. Quelle distribution peut on suggerer
3. Si Y suit une loi de poisson Estimer son parametre (utiliser maximum de vraisemblance)
4. Si Y suit une loi normale Estimer ses parametre (utiliser maximum de vraisemblance)
5. Si Limite suit une loi bernouilli Estimer ses parametre (utiliser maximum de vraisemblance)
6. Utiliser la distribution Poisson et normal pour modeliser le y. Realisez le tableau ou chaque ligne du tableau contient les intervalles de y . Calculez a partir des donnees la distribution observee.Calculez les probabilites theoriques. Comparez les probabilites empiriques et theoriques.

Limits	effectif
0	115
1	69
	184

y	effectif
7	1
8	2
9	8
10	2
11	4
12	7
13	4
14	8
15	13
16	10
17	10
18	10
19	9
20	8
21	11
22	8
23	4
24	8
25	9
26	5
27	5
28	2
29	6
30	2
31	4
32	4
34	2
35	1
36	2
37	2
38	1
39	2
40	2
41	2
42	2
44	1
47	1
48	1
49	1

184

y	fréq.	probabilité empirique	probabilité théorique Normal	probabilité théorique Poisson
(5, 10]	13			
(10, 15]	36			
(15, 20]	47			
(20, 25]	40			
(25, 30]	20			
(30, 35]	11			
(35, 40]	9			
(40, 45]	5			
(45, 50]	3			
	185	100%	100%	100%