Enseignant : Sofiane MAAZI

Introduction aux logiciels pour les statistiques (IS2A3) - 2024/2025

TP2 : Manipulations avancées en R

Exercice 1 : Test statistique

Un fabriquant de gâteaux commercialise ses produits avec sur l'emballage la mention « Teneur moyenne en lipides inférieure ou égale à 20 grammes ». Le fabricant veut contrôler sa production car il pense que ses gâteaux respectent l'indication notée. Les résultats sont :

20 – 23 – 23 – 23 – 22 – 20 – 23

On suppose que la teneur en lipides en grammes d'un gâteau peut être modélisée par une var X suivant une loi normale. La problématique est la suivante : Peut-on affirmer, avec un faible risque de se tromper, que le fabriquant a raison ?

1. Explicitez les hypothèses à tester.
2. Déterminez la *p-valeur* associée..
3. Répondez à la problématique en précisant le degré de significativité en cas de rejet de H_0.

Exercice 2 : Regression lineaire simple

1. Importez le jeu de données *cars*.
2. Tracez le nuage de points constitué des points de coordonnées (speed, dist).
3. On considère le modèle de régression linéaire simple $dist = \alpha + \beta\ speed + \varepsilon$
   (a) Estimez $\alpha + \beta$ par la méthode des moindres carrés.
   (b) Ajoutez à votre nuage de points la droite d'équation $y = \alpha + \beta x$.

Exercice 3 : Simulations et lois de probabilite

Introduction

Pour chaque loi de probabilité, il existe 4 commandes essentielles. Chacune est obtenue en préfixant le nom de la loi (cf. annexe) par une lettre.

- ***d**nomloi :* fonction de densité (cas « continue ») ou fonction de probabilité (cas « discret »).
- ***p**nomloi* : fonction de répartition F où F(k)=P[X<k]
- ***q**nomloi :* fonction des quantiles. C'est la valeur pour laquelle F atteint certaine probabilité. Dans le cas discret, cette fonction renvoie le plus petit entier n tel que F(n) ≤ p.
- ***r**nomloi :* génère des valeurs aléatoires

Que font les 8 commandes suivantes ?

qnorm(0.975)	*pnorm(1.96)*	*runif(3)*	*rnorm(10,mean=5,sd=0.5)*
exp(232)	*rnorm(20)*	*rt(5,10)*	*x=seq(-3,3,0.1) ;pdf=dnorm(x) ;plot(x,pdf,type="l")*

Loi exponentielle

1. Écrivez une fonction *f* équivalente à *dexp* quand le paramètre $\lambda=0.1$.
2. Vérifiez que l'intégrale de *dexp* sur [0,100] est proche de 1.
3. Représentez le graphe de *dexp* et de la fonction de répartition associée.
4. Quelle est la probabilité que la durée de vie d'un noyau d'uranium dépasse 20 ans ?

Loi gamma

1. Écrire une fonction *g* équivalente à *dgamma* quand les paramètres $a=5$ et $b=1$.
2. Vérifiez que l'intégrale de *dgamma* sur [-100,100] est proche de 1.
3. Évaluez la fonction de répartition de X pour $x \in \{2,...,10\}$.
4. Déterminez le réel x vérifiant $P(X \leq x) = 0.92$

Loi normale

1. Écrivez une fonction *h* équivalente à *dnorm* quand les paramètres $\mu = 0$ et $\sigma = 1$.
2. Vérifiez que l'intégrale de *dnorm* sur [0,100] est proche de 1.
3. Représentez le graphe de *dnorm* et de la fonction de répartition associée.
4. Calculez $P(X \leq 2.2)$, $P(X \geq 1.7)$, $P(0.2 \leq X < 1.4)$ et $P(|X| \leq 1.96)$.
5. Déterminez le réel x vérifiant $P(X \leq x) = 0.98$
6. On suppose que poids d'une meule de fromage est une variable aléatoire Y suivant une loi normale de paramètres $\mu = 550$ et $\sigma=100$. L'unité est le gramme. Quelle est la probabilité que la meule pèse … moins de 650 grammes ? plus de 746 grammes ? entre 550 grammes et 600 grammes ?

Exercice 4 : Simulations de variables discretes

La fonction *sample()* prélève successivement un nombre donné d'éléments d'un ensemble de cardinal fini, avec ou sans remise : *sample(x, size, replace = FALSE, prob = NULL)*

1. Expliquez le rôle de chacun des paramètres de la fonction.
2. Écrivez une commande R qui renvoie le résultat
   (a) du lancer d'un dé.
   (b) du lancer de 2 dés.
   (c) de la somme des résultats du lancer de 2 dés.
   (d) du tirage du loto (5 numéros parmi 49).

3. Une urne contient B + N boules, dont B blanches et N noires. Écrivez une fonction *Urne(k; B; N)* qui modélise le résultat de k tirages sans remise d'une boule de l'urne.

4. On lance 5 dés cubiques équilibrés, puis on relance ceux qui n'ont pas fait 6, et ainsi de suite jusqu'à ce que les 5 dés affichent 6. Simulez une réalisation de cette expérience aléatoire en affichant les chiffres obtenus après chaque lancer et le nombre de lancers total.

Exercice 5 : *Simulations de variables continues*

Un échantillon de taille 50 d'une var X est présenté dans le tableau suivant : 0.498, 2.409, 0.577, 5.579, 2.507, 1.695, 2.396, 1.528, 0.365, 0.388 , 3.971, 0.984, 1.157, 12.283, 0.052, 3.939, 9.631, 3.727, 2.684, 0.513 , 2.540, 1.509, 1.542, 4.771, 1.478, 4.784, 3.266, 6.677, 3.034, 2.426 , 2.595, 0.472, 4.074, 4.530, 2.192, 0.418, 1.237, 1.929, 0.729, 0.127 , 6.429, 2.072, 7.844, 0.584, 1.769, 0.838, 1.590, 1.554, 5.853, 4.342

1. Représentez ces données par un histogramme des fréquences.
2. Tracez, sur le graphique précédent, les 3 densités de la loi exponentielle $E(\lambda)$ avec $\lambda \in \{0.33, 0.5, 3\}$. Quelle est la densité la plus adaptée ?

3. Simulez un échantillon de taille 50 d'une variable aléatoire Y suivant une loi exponentielle de paramètre $\lambda=0.33$ et représentez le avec un histogramme des fréquences.

Exercice 5 : ***Diagramme quantile-quantile***

On condidère 2 variables aléatoires indépendantes X_1 et X_2 . Illustrez les situations suivantes :

(a) Si X_1 et X_2 suivent la même loi exponentielle $\varepsilon(\lambda)$ alors X_1+X_2 suit une loi gamma $\Gamma(\lambda,2)$.

(b) Si X_1 et X_2 suivent respectivement une loi gamma $\Gamma(m_1, \lambda)$ et $\Gamma(m_2, \lambda)$ *alors* X_1+X_2 *suit une loi gamma* $\Gamma(m_1+m_2, \lambda)$.