

STATISTIQUES

QUELQUES EXERCICE

Introduction : entête type d'un énoncé de partiel

Spécificités : L'énoncé de ce devoir est constitué des documents suivants :

- Ce fichier,
- Le fichier REPONSES.csv, devant contenir les réponses à vos questions et être renommé suivant la convention formulée par l'EPITA et déposé dans l'espace informatique tel que demandé par l'EPITA.

Il s'agit d'un devoir effectué à domicile. Nous vous demandons de répondre dans le fichier Excel, REPONSES.csv tel que précisé plus haut. Ce fichier doit impérativement être sauvegardé dans le format .csv couramment employé en France : séparateur point-virgule « ; », séparateur décimal virgule « , ».

Le fichier est constitué de deux colonnes.

- La première a été préalablement remplie et permet de retrouver l'endroit où vous devez apporter votre réponse. Dans cet énoncé, nous désignons systématiquement les éléments de cette première colonne entre guillemet. Par exemple, dans la ligne numéro 2 en colonne A est écrit NOM. Nous désignons cette première ligne par la ligne « NOM ». Cette première colonne n'emploie que les symboles courant du Français (pas de symbole mathématique, pas de lettre grecque).
- La deuxième (colonne B) contient votre réponse. Lorsque la réponse est numérique, vous ne donnez que le résultat numérique, sans aucun commentaire, en employant la virgule (,) comme séparateur décimal (exemple 3,14 mais pas 3.14). Lorsque la réponse n'est pas numérique, elle est alors constituée d'un ou plusieurs mots clés. Nous vous demandons de les écrire en MAJUSCULE sans aucun accent, séparés éventuellement par des caractères blancs.

Exercice 1.

L'objectif de cet exercice est de manipuler des commandes simples de R.

1.1 Création de vecteurs que nous décrirons par la suite.

Choisissez une constante K (par exemple $K = 1.25$ et un paramètre α , par exemple $\alpha = 0.957$, ne prenez pas une valeur trop grande.

Construisez les vecteurs suivants, ayant tous la même longueur :

Un intervalle : `INT <- 1:5000`

Une simulation de loi normale, prise en valeur absolue : `ABS.NORM <- abs(rnorm(5000))`

Une simulation de loi exponentielle : `EXP<-rexp(5000)`

Posez :

$$X <- -INT^\alpha \quad Y <- -ABS.NORM^\alpha \quad Z <- -EXP^\alpha$$

Pour chaque vecteur, interprété comme un échantillon de nombre, donnez :

La moyenne, la variance, l'écart-type, le premier quartile, la médiane, le troisième quartile.

Vous pouvez utiliser les commandes R : mean, var, sd, summary, boxplot.

Quels sont les coefficients de corrélation de chaque couple de vecteurs X, Y ; X, Z ; et enfin Y, Z ?

Exercice 2 :

Dans une usine agroalimentaire, les paquets de sucre emballés automatiquement doivent avoir un poids compatible avec celui affiché sur la boîte et lu par les consommateurs.

On suppose connu l'écart type du poids d'un paquet produit par les machines, et égal à 10 grammes. On a mesuré les poids d'un échantillon de 200 paquets et trouvé une moyenne de 495 grammes.

Donner un intervalle de confiance à 95% du poids d'un paquet de sucre sortant des usines ?
Même question, mais cette fois avec un intervalle de confiance à 99% ?

On suppose maintenant qu'on ne connaît pas l'écart type à priori des poids des paquets sortant des machines. Expérimentalement, sur l'échantillon précédent, on trouve un écart type de 17 grammes.

Donner un intervalle de confiance à 95% du poids d'un paquet de sucre sortant des usines, avec cette nouvelle hypothèse ? Même question, mais cette fois avec un intervalle de confiance à 99% ?

Exercice 3 :

On veut tester qu'un dé à 6 faces n'est pas truqué à l'aide d'un test de χ^2 .

On teste l'hypothèse suivante :

H_0 : la probabilité d'obtenir chacune des faces 1 à 6 lors du lancé d'un dé est de $1/6$.

On effectue 32 lancers de dés et on note les faces obtenues lors de ces lancers, ce qui donne le tableau suivant

Face du dé	1	2	3	4	5	6
Nombre de lancers	4	11	4	3	6	4

Peut-on accepter « à 99% » que le dé n'est pas truqué ?

Exercice 4 : une régression linéaire

Nous disposons du relevé présenté page suivante des valeurs des variables X et Y pour 24 individus.

Effectuez une régression linéaire de Y en fonction de X. reprenez les étapes des calculs qui ont été faits en cours.

Vous pouvez vous exercer avec Excel, ou R.

Le tableau est disponible dans un fichier séparé, au format csv, si nécessaire.

X	Y
62	280
262	1321
110	245
68	354
236	641
280	868
270	821
285	592
287	751
130	419
294	633
55	151
165	434
209	538
7	46
57	142
111	767
180	429
154	409
129	594
141	406
263	566
234	654
268	588