

Le but de ce TD est de vous familiariser avec le logiciel R.

Exercice 1 : 7.2 page 220. [8.2 page 208 dans la 2ème édition] (l'exercice est modifié)

À partir du fichier Coton.csv, effectuer une analyse statistique complète de la variable "pourcentage de coton" : histogramme, tableau d'effectifs (fréquences), diagramme boîte et moustaches (*box plot*), graphique quantile-quantile, graphique sur échelle de probabilité normale, calcul de moyenne, variance, écart type, coefficient de variation, médiane, Q1, Q3, écart interquartile.

Exercice 2 :

À partir du fichier Memoire.csv, on désire effectuer l'analyse statistique de la variable "stress".

- Calculer la moyenne l'écart type, le cinquième percentile, et le 95ième percentile.
- La variable est-elle normalement distribuée ? Répondre avec un graphique.
- Calculer la moyenne et la variance selon la variable "sexe".

Exercice 3 :

Le département administratif d'une compagnie qui développe des sites web est intéressé à déterminer quelles sont les variables qui ont un impact sur le nombre de sites web livrés aux clients ainsi que sur la taille du carnet de commandes. On a décidé d'analyser des données recueillies au sujet de 13 équipes qui ont travaillé au développement de 73 sites web, entre janvier 2001 et septembre 2002. Chacune des 13 équipes étudiées comportait trois membres. À la fin de la période de collecte des données, plus précisément aux deuxième et troisième trimestres de l'année 2002, le processus de développement des sites web au sein de la compagnie a subi des modifications.

On considère le fichier Sites_web.csv contenant les variables suivantes :

Numéro	Nom	Description
V1	no	numéro d'identification de l'observation (1 à 73)
V2	nbre de sites livrés	nombre de sites web complétés et livrés aux clients pendant le trimestre
V3	carnet de commandes	nombre de commandes en attente au terme du trimestre
V4	no. d'équipe	numéro d'identification de l'équipe (1 à 13)
V5	expérience de l'équipe	nombre de mois écoulés depuis la formation de l'équipe (comptabilisés au début du trimestre)
V6	processus modifié	variable indicatrice : 1 : pour les trimestres 2 et 3 de l'année 2002 0 : pour les autres trimestres
V7	année	2001 ou 2002
V8	trimestre	1, 2, 3, ou 4

- a) Calculer la moyenne, l'écart-type et la médiane pour le nombre de sites web livrés (V2) et pour la taille du carnet de commandes (V3). Laquelle de ces deux variables a le coefficient de variation le plus grand ?
- b) Tracer un nuage de points simultané (sur un même graphique) (*scatterplots*) du nombre de sites livrés (V2) sur l'axe vertical et du carnet de commandes (V3) sur l'axe horizontal pour chacune des équipes (V4). Refaire un graphique semblable (V2 versus V3) avec (V5) à la place de (V4) (expérience).
- c) Quelle est la relation entre l'expérience d'une équipe (V5) et le nombre de sites livrés (V2)? Et avec la taille de son carnet de commande (V3)?
- d) Est-il plausible de penser que les observations du nombre de sites livrés (V2) proviennent d'une population (distribution) normale ?
- e) Quel est l'effet associé à la modification du processus de développement des sites web effectuée aux deux derniers trimestres considérés :
 - sur le nombre de sites livrés (V2)?
 - sur le carnet de commandes (V3)?

Pour répondre, produire un tableau de statistiques descriptives (moyenne, écart type, médiane,...) des variables V2 et V3 avant et après la modification. Ajouter aussi un graphique boîte-moustaches illustrant la différence avant et après la modification (V6).