

BE1 : Introduction aux statistiques descriptives

Cette 1^{ère} séance sur machine avec Matlab a pour but de passer en revue plusieurs notions nécessaires aux cours de Traitement de données et de Statistiques.

Les exercices peuvent se faire en binôme ainsi que le compte rendu qui est demandé à l'issue de la séance.

I. Questions de cours

Pour répondre aux questions de cours vous pouvez utiliser internet et/ou des documents/livres

1.1. Donner les définitions des grandeurs suivantes : moyenne, écart-type et variance

1.2. Donner la définition du coefficient de corrélation entre 2 variables ? Que permet cette grandeur ?

II. Exercices

Les exercices sont à réaliser en utilisant Matlab. Il est important de retenir le nom des fonctions, ou du moins qu'elles existent, car elles vous seront utiles pour l'ensemble des BEs. Ci-dessous, vous trouverez une liste de quelques fonctions. Pour rappel, si vous souhaitez comprendre une fonction, ses variables d'entrée et de sortie, vous devez utiliser la fonction *help* de Matlab.

Exemple : >> help mean

Liste de fonctions :

mean, std, var,
mode, median,
min, max,
hist,
boxplot,
quartile, prctile,

corrcoef, cov

find,
isnan
xlsread

Exercice 1:

Les données se trouvent dans le fichier : Exercice1.xls. Vous trouverez 2 jeux de données (un par colonne) ; on parle également de 2 échantillons.

Etude de l'échantillon 1:

1/ Donner la valeur minimale, la valeur maximale, la moyenne et médiane.

A partir de cet échantillon, créer un nouvel échantillon identique pour lequel vous modifierez les valeurs min et max. Donner à nouveau la moyenne et la médiane. Quel commentaire pouvez-vous formuler ?

2/ Tracer la boîte à moustaches (ou diagramme en boîte). Il est ici important de comprendre les différents éléments qui apparaissent sur ce diagramme

3/ Calculer les quartiles. Que vaut l'intervalle interquartile ?

Etude des 2 échantillons séparément :

3/ Pour chaque échantillon (reprenez les échantillons originaux et pas celui avec les valeurs extrêmes modifiées) donner la moyenne et l'écart-type.

4/ Pour chaque échantillon tracer les histogrammes. A quelle loi de probabilité vous font penser ces histogrammes ?

Exercice 2:

Les données se trouvent dans le fichier : Exercice2.xls. Vous trouverez la taille et le poids de 11 individus.

1/ Calculer le coefficient de corrélation ? (à l'aide de la fonction Excel mais également à la main à partir de la formule de la corrélation)

2/ Que peut-on ici conclure ?

Exercice 3:

Récupérer un jeu de données de Matlab : *patients*

>> load patients

On s'intéressera principalement aux variables suivantes :

Gender, Age, Height et Weight

Comment décrire ces données ? Proposer des résumés et des graphiques