Université de Lorraine - UFR MIM - 2015/2016
Cours MATLAB

MATLAB 7

Statistiques élémentaires

J-P. Croisille

1- Espérance de vie
On considère le tableau des données de l'espérance de vie dans 12 pays d'Amérique du Sud.

	Pays	Esp. vie
1	Argentine	71.5
2	Bolivie	63.5
3	Brésil	62
4	Chili	75
5	Colombie	72.5
6	Equateur	70.5
7	Guyane	65
8	Paraguay	73.5
9	Pérou	66
10	Surinam	69.5
11	Uruguay	74.5
12	Venezuela	73

1) Entrer les données dans le vecteur `X` par `X=[71.5 63.5 62 75 72.5 70.5 65 73.5 66 69.5 74.5 73];` Vérifier le contenu de `X`.
2) Calculer la taille `n` de l'échantillon (utiliser `size`).
3) Calculer la moyenne empirique `m`, la variance empirique `s` ainsi que la médiane `md` de l'échantillon, en utilisant les fonctions `mean,std,median`. Quelle est la différence entre moyenne et médiane ? Quelle est la différence entre variance et écart-type ?
4) Représenter ces données à l'aide d'un histogramme par les commandes
`figure(1);hist(X)`. Cette commande répartit les données en 10 classes.
5) Représenter dans une autre figure un second histogramme avec seulement 4 classes par `hist(X,4)`. On conservera le premier histogramme avec 10 classes à l'écran. Rajouter le titre suivant aux deux histogrammes. `title('Esperance de vie en Amerique du Sud');`

[0]Les données sont extraites des références suivantes:

- Samuels & Witmer: "Statistics in life sciences".
- N. Weiss: "Introductory Statistics".

2- Concentration en créatine phosphokinase
On considère le tableau de données suivant (concentration en créatine phosphokinase) chez 36 volontaires masculins.

121	82	100	151	68	58
95	145	64	201	101	163
84	57	139	60	78	94
119	104	110	113	118	203
62	83	67	93	92	110
25	123	70	48	95	42

1) Entrer le tableau des données dans le tableau `data`.
2) Le classement de ces données en 10 classes fournit le tableau suivant

Serum CK	Fréquence
20-39	1
40-59	4
60-79	7
80-99	8
100-119	8
120-139	3
140-159	2
160-179	1
180-199	0
200-219	2

2) Entrer le tableau suivant:
`frequence=[1 4 7 8 8 3 2 1 0 2];`
3) Représenter ce classement des données à l'aide d'un premier histogramme en rouge par la commande `bar(frequence,'r')`.
4) On souhaite à présent que les extrêmités des classes soient représentées en abscisse. Entrer `edges=[20 40 60 80 100 120 140 160 180 200 220];` puis `[frequence1]=histc(data,edges);` et `bar(edges,n,'histc');`. 5) Représenter dans une autre figure la répartition avec un diagramme de type camembert ("pie-chart" en anglais) en procédant ainsi:
a) Entrer le tableau `data=[1 4 7 8 8 3 2 1 0 2];`
b) Entrer le tableau correpondants des chaînes de caractères: `titre={'20-39','40-59','60-79','80-99','100-119','120-`
`'140-159','160-179','180-199','200-219'};`
c) Entrer les commandes:
`pie(data,titre)`;

3- Contrôle de la normalité d'un échantillon
En statistique, il est important de procéder à une analyse visuelle des données d'un échantillon. Par exemple, on peut vérifier qu'un échantillon est distribuée de façon approximativement normale en utilisant un "normal plot".
1) Générer la donnée de $n = 100$ nombres aléatoires avec `x=normrnd(0,1,100,1);`.
2) Représenter un histogramme de x avec 10 classes. Conclusion ?
3) Vérifier la normalité approximative de `x` à l'aide de la commande `normplot(x)`.

4- Représentation graphique des données multivariées
Un jeu de données multivariées consiste en une matrice $X \in \mathbb{M}_{n,p}(\mathbb{R})$. Cette matrice représente n individus (nombre de lignes) ayant chacun p caractères.
1) L'importation de la matrice X donnée en format texte se fait par l'une des commandes `tblread,tdfread`.

Par exemple, `X=tblread('pullover.dat');`. On utilisera la syntaxe convenable permettant de lire un fichier comportant comme séparateur `;`. (Voir `help tblread` et `help tdfread`).
2) En utilisant les commandes `mean`, `cov` et `corrcoef`, calculer le vecteur moyen $m \in \mathbb{R}^p$, la matrice des covariances `C` ainsi que la matrice des corrélations `R` de chacune des données `pullover.dat`, `mandible.dat`, `frenchfood.dat`.
3) Il y a plusieurs façons courantes pour représenter graphiquement un échantillon de n individus possédant p caractères. Par exemple
a) tracer les "courbes d'Andrew", en utilisant `andrewsplot`.
b) tracer les "boxplots" en utilisant `boxplot`.
c) tracer une matrice histogramme/ scatterplot en utilisant `gplotmatrix`.
d) tracer un diagramme de type "visages de Chernoff" en utilisant `glyphplot`.