

## Statistique descriptive\*

### 1. Statistique uni-dimensionnelle : variable quantitative discrète

On note l'âge (arrondi à l'année près) des 48 salariés d'une entreprise, dont la statistique brute est :  
43 29 57 45 50 29 37 59 46 31 46 24 33 38 49 31  
62 60 52 38 38 26 41 52 60 49 52 41 38 26 37 59  
57 41 29 33 33 43 46 57 46 33 46 49 57 57 46 43

- Représenter dans un tableau statistique : les *effectifs*, les *effectifs cumulés*, les *fréquences* et les *fréquences cumulées*, calculés sur la série des données donnée ci-dessus.
- Représenter le *diagramme en bâtons*, puis le *diagramme cumulatif*, associés à la série de données.
- Résumer la série de données à l'aide d'une *boîte-à-moustache*.
- Calculer quelques caractéristiques de la série de données : la *moyenne*, la *variance* et l'*écart-type empiriques*.

### 2. Statistique uni-dimensionnelle : variable quantitative continue

Le tableau ci-dessous donne, pour l'année 1987, la répartition des exploitations agricoles françaises selon la SAU (surface agricole utilisée) exprimée en hectares.

SAU (en ha)	fréquences (%)
moins de 5	24
de 5 à 10	10,9
de 10 à 20	17,8
de 20 à 35	20,3
de 35 à 50	10,2
plus de 50	16,8

- Représenter la *fonction de répartition empirique*, puis l'*histogramme* (ou la fonction de densité empirique) associées à la variable continue SAU.
- Déterminer le *mode*, puis calculer la valeur *moyenne* et enfin calculer par interpolation linéaire la valeur de la *médiane*.

\* D'après le cours de statistique descriptive réalisé par A. Baccini de l'Univ. P. Sabatier, Toulouse

### 3. Statistique multi-dimensionnelle : variables quantitatives

Les Tableaux de l'Économie Française, INSEE, 1989, pag. 109, montrent pour les 51 secteurs d'activité, le nombre total d'entreprises (NB), l'effectif salarié (EF) et le chiffre d'affaires hors -taxes en millions de francs (CA). Le tableau ci-dessous reproduit une partie de l'ensemble de données.

<i>code</i>	<i>secteur</i>	<i>NB</i>	<i>EF</i>	<i>CA</i>
4	production de combustibles minéraux solides et cokéfaction	19	49251	14111
5	production de pétrole et gaz naturel	120	46594	306293
6	production et distribution d'électricité	731	129723	138389
...	...	...	...	...
52	industrie du caoutchouc	746	87121	37502
53	transformation de matières plastiques	3232	102437	58122
54	industrie diverses	10171	84012	38071

a) Calculer sur l'échantillon présenté ci-dessus : l'étendue, la moyenne, la variance, ainsi que l'écart-type empiriques des variables NB, EF et CA.

b) Représenter le nuage de points en utilisant les variables CA (en ordonnées) et EF (en abscisses).

c) Calculer la covariance des variables  $X$  et  $Y$  :  $cov(X,Y)$ , puis le coefficient de corrélation linéaire :  $r_{XY}$  pour  $X=EF$  et  $Y=CA$ .

d) Calculer la matrice des variances-covariances  $S$  et la matrice des corrélations  $R$ , entre les 3 variables de l'étude.

e) Représenter sous Matlab la partie inférieure du tableau de nuages correspondant aux variables NB, EF et CA.

### 4. Statistique bi-dimensionnelle : 1 variable quantitative et 1 variable qualitative

Les individus étant 19 chiens prémédiqués au pentobarbital, on veut étudier l'effet sur leur rythme cardiaque (variable  $Y$ ) de 2 facteurs croisés (variable  $X$ ). L'effet est mesuré par le temps entre 2 battements de coeur successifs. Les facteurs sont :

- la pression d'administration de  $CO_2$  qui peut être élevée (E) ou faible (F) ;
- la présence (0) ou l'absence (1) d'halothane.

Le tableau ci-dessous contient le rythme cardiaque, exprimé en ms, à l'issue des expériences

réalisées pour chacune des conditions ainsi définies.

N° chien	E et 0	F et 0	E et 1	F et 1
1	426	609	556	600
2	253	236	392	395
3	359	433	349	357
4	432	431	522	600
5	405	426	513	513
6	324	438	507	539
7	310	312	410	456
8	326	326	350	504
9	375	447	547	548
10	286	286	403	422
11	349	382	473	497
12	429	410	488	547
13	348	377	447	514
14	412	473	472	446
15	347	326	455	468
16	434	458	637	524
17	364	367	432	469
18	420	395	508	531
19	397	556	645	625

- Préciser la nature des variables X et Y, puis spécifier pour la variable qualitative ses *modalités*.
- Calculer la *moyenne* et l'*écart-type* de la variable Y pour la population globale. Calculer les *moyennes* et les *écarts-types partiels* pour chacune des sous-populations.
- Résumer les données du tableau ci-dessus au moyen des *boîtes parallèles*. Comparer les boîtes et déduire l'influence de X sur les valeurs de Y.
- Calculer la *variance totale*, la *variance expliquée* et la *variance résiduelle*. Déduire le *rapport de corrélation*.