

### TD de Statistique Descriptive SMI(3)

#### Exercices sur la “Statistique Descriptive Univariée”

**Exercice 1 :** Classer les variables ci-dessous selon leur type (nature) : Poids, Couleur des yeux, Nombre de frères et soeurs, Marque de PC portable, Profession, Nationalité, Taille, Mention, Revenu mensuel, Langue maternelle.

**Exercice 2 :** On s'intéresse à la quantité de livres lus dans l'année, par un groupe d'étudiants ( $A = \text{peu}$ ,  $B = \text{moyen}$ ,  $C = \text{beaucoup}$ ,  $D = \text{exceptionnel}$ ). Soit la liste des prénoms des étudiants de ce groupe, suivis de l'indication de leur appétit de lecture :

Zineb ( $C$ ), Youssef ( $C$ ), Ali ( $A$ ), Yassir ( $B$ ), Amine ( $A$ ), Catherine ( $B$ ), Jean ( $C$ ), Loubna ( $B$ ), Leila ( $B$ ), Fatima ( $C$ ), Taha ( $D$ ), Farida ( $B$ ), Bouchra ( $A$ ), Malak ( $C$ ), Naima ( $C$ ), Abdellah ( $C$ ), Akram ( $C$ ), Souad ( $D$ ), Taoufiq ( $C$ ), Yahya ( $C$ ).

1. Quelle est la nature de la variable “appétit de lecture” ?
2. Construire le tableau représentatif de cette distribution.
3. Représenter cette distribution à l'aide des diagrammes adéquats.

**Exercice 3 :** Le gérant d'un magasin vendant des articles de consommation courante a relevé pour un article particulier qui semble connaître une forte popularité, le nombre  $X$  d'articles vendus par jour. Son relevé des observations sur 52 jours se présente comme suit :

7 ; 13 ; 8 ; 10 ; 9 ; 12 ; 10 ; 8 ; 9 ; 10 ; 6 ; 14 ; 7 ; 15 ; 9 ; 11 ; 12 ; 11 ; 12 ; 5 ; 14 ;  
11 ; 8 ; 10 ; 14 ; 12 ; 8 ; 5 ; 7 ; 13 ; 12 ; 16 ; 11 ; 9 ; 11 ; 11 ; 12 ; 12 ; 15 ; 14 ; 5 ;  
14 ; 9 ; 9 ; 14 ; 13 ; 11 ; 10 ; 11 ; 12 ; 9 ; 15.

- a. Quel est le type de cette variable statistique ?
- b. Représenter le tableau d'effectifs et d'effectifs cumulés, associé à cette série.  
Que représente l'effectif dans ce cas ?  
Que représente l'effectif cumulé ?

$X$	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Effectifs															
Effectifs cumulés															

- c. Calculer la médiane et la moyenne de cette série.
- d. Quels sont l'étendue et l'écart-type de cette distribution.
- e. Donner les trois quartiles  $Q_1$ ,  $Q_2$  et  $Q_3$  puis tracer le diagramme en boîte correspondant à cette série statistique.
- f. Existe-t-il des valeurs aberrantes ?

**Exercice 4 :** La taille en centimètres (cm) de 60 étudiants d'une école d'ingénieurs a été relevée. La collecte des données a fourni les résultats suivants :

167   169   171   172   167   171   169   171   172   165   173   169  
 174   168   164   170   175   173   169   173   170   171   176   170  
 172   163   171   171   174   168   163   165   168   168   173   169  
 170   168   171   170   164   170   172   172   167   170   169   164  
 167   171   173   168   167   177   171   172   159   169   165   176

1. Quelle est la population concernée par l'étude ? Quel est l'échantillon ? Quelle est la variable étudiée ? Donner sa nature.
2. Remplissez le tableau suivant en utilisant 5 classes de même amplitude (la borne inférieure de la 1<sup>ère</sup> classe sera prise égale 155 et la borne supérieure de la dernière classe égale à 180).

$X$	$n_i$	$N_i$	$c_i$	$a_i$	$h_i$
[.....; .....[					
[.....; .....[					
[.....; .....[					
[.....; .....[					
[.....; .....[					
[.....; .....[					
Total					

3. Quelle est l'étendue de cette distribution ?
4. Tracer l'histogramme correspondant à ces données. Y a-t-il une classe modale ?
5. Quelle est la valeur approchée de la moyenne des tailles ?
6. Quel est l'écart-type des tailles ?
7. Quelle est la médiane des tailles ?
8. Donner des valeurs approchées des deux autres quartiles. En déduire l'écart interquartile ?
9. Dessiner le diagramme en boîte correspondant à cette distribution.
10. Y a-t-il des valeurs aberrantes ?

**Exercice 5 :** Dans le cadre d'une étude sur le temps  $X$  de mémorisation (en minutes), on a choisi au hasard, un échantillon de 28 élèves qu'on a soumis à un test. Les résultats obtenus sont donnés dans le tableau suivant :

$X$	$n_i$	$N_i$	$c_i$	$a_i$	$h_i$
[3; 4[	3				
[4; 5[	7				
[5; 6[	5				
[6; 11[	5				
[11; 13[	6				
[13; 17[	2				
Total					

1. Quelle est la population étudiée ? Quelle est la taille de l'échantillon ? Préciser la variable étudiée et son type .
  2. Compléter le tableau ci-dessus puis dessiner le diagramme correspondant à cette distribution. Que constate-t-on ?
  3. Calculer la moyenne, la variance et l'écart-type de  $X$ .
  4. Déterminer la médiane. Que représente cet indice ? Expliquer brièvement pourquoi il y a une différence non négligeable entre la moyenne et la médiane.
- .....

### Exercices sur la “Statistique Descriptive Bivariée”

**Exercice 1 :** On considère 10 élèves qui jouent à quatre jeux.

Soient  $Y$  la variable représentant “le nombre de jeux auxquels un élève joue” et  $X$  la variable représentant “le gain ou la perte ” (+1 s’il gagne un bon point,  $-1$  s’il perd un bon point et 0 s’il n’est ni gagnant ni perdant).

Les résultats obtenus sont résumés dans le tableau suivant :

$X/Y$	1	2	3	4	$n_{i.}$
$-1$	0	1	2	2	
$0$	1	1	0	1	
$1$	0	1	1	0	
$n_{.j}$					

1. Compléter le tableau ci-dessus.
2. Calculer  $Cov(X, Y)$  puis le coefficient de corrélation  $r_{XY}$ . Que dire du lien entre ces deux variables ?

### Exercice 2 :

Une étude faite au sein d’une entreprise a relevé  $X$  le salaire mensuel (en milliers de dirhams) et  $Y$  l’ancienneté de ses ouvriers.

Les résultats obtenus sont résumés dans le tableau ci-dessous :

$X(\times 10^3) / Y$	$[0; 8[$	$[8; 16[$	$[16; 24[$	$[24; 32[$	$n_{i.}$	$f_{i.}$
$[2; 3[$	5	6	1	0		
$[3; 4[$	2	4	3	3		
$[4; 5[$	0	2	4	10		
$n_{.j}$						
$f_{.j}$						

1. Etudier les séries marginales (leurs moyennes et variances). Commenter.
2. Déterminer si les variables  $X$  et  $Y$  sont indépendantes.
3. Etudier les séries conditionnelles  $X/y_3$  et  $Y/x_2$  et représenter les résultats pour chaque groupe de séries conditionnelles.

**Exercice 3 :** Une enquête menée, pour le compte d’une entreprise, a permis d’établir le nombre  $Y$  d’acheteurs (en milliers), d’un produit selon le montant de son prix de vente  $X$  (en dirhams). Les résultats de l’enquête sont donnés dans le tableau suivant :

$x_i$	10	15	20	30	40
$y_i$	3.75	2.8	2	1	0.5

1. Représenter le nuage de points  $(x_i, y_i)$ ;  $i = 1, \dots, 4$ .
2. On cherche un ajustement affine.
  - a. Donner l'équation de la droite de régression.
  - b. Tracer cette droite dans le même plan que précédemment.
  - c. Utiliser cet ajustement pour prédire le nombre d'acheteurs pour un produit vendu à 25 dirhams.

**Exercice 4 :** On considère le tableau de données suivant :

$X$	1	2	7	4	6
$Y$	5	4	1	3	2

1. Tracer le nuage de points.
2. Deviner le signe et la valeur du coefficient de corrélation.
3. Calculer le coefficient de corrélation, la pente et l'ordonnée à l'origine de la droite de régression.

**Exercice 5 :** Une usine produit des pièces d'une machine. Pour chaque pièce (individu), on dispose du coût  $Y$  de sa production (en dirhams) et du temps de sa réalisation  $X$  (en heures). On considère 5 pièces pour lesquelles on mesure ces deux caractères. Les résultats sont résumés dans le tableau suivant :

Individu	1	2	3	4	5
Temps $X$ (en heures)	2	3	5.2	2	4
Coût $Y$ (en dhs)	10	16	23	12	18

1. Calculer les moyennes respectives des variables statistiques  $X$  et  $Y$ .
2. Calculer les écart-types de  $X$  et de  $Y$ .
3. Calculer la covariance des variables  $X$  et  $Y$ .
4. En supposant qu'il existe une corrélation linéaire entre  $X$  et  $Y$ , déterminer cette droite de régression.
5. Une nouvelle pièce est réalisée en 6 heures. Estimer le coût de production de cette pièce en utilisant la droite de régression.