

TD N° 1

STATISTIQUE DESCRIPTIVE UNIVARIÉE ET BIVARIÉE

Questions. Parmi ces assertions, préciser celles qui sont vraies, celles qui sont fausses :

1. On appelle variable, une caractéristique que l'on étudie
2. La tâche de la statistique descriptive est de recueillir des données.
3. La tâche de la statistique descriptive est de présenter les données sous forme de tableaux, de graphiques et d'indicateurs statistiques.
4. Les valeurs des variables sont aussi appelées modalités.
5. Pour une variable qualitative, chaque individu statistique ne peut avoir qu'une seule modalité.
6. Pour faire des traitements statistiques, il arrive qu'on transforme une variable quantitative en variable qualitative.

Exercice 1. Donner la nature de chacune des variables suivantes :

- (i) Degré de satisfaction concernant le séjour estival dans un centre de vacances (Très élevé; Élevé; Modéré; Bas; Nul).
- (ii) La moyenne semestrielle d'un étudiant à la faculté de pharmacie.
- (iii) Le temps consacré par un étudiant de S2 pour suivre les informations sur un journal écrit ou télévisé.
- (iv) La section (A,B,C,...) du cours de biologie à laquelle est inscrit un étudiant.
- (v) Le nombre d'appels téléphoniques reçus en une journée par un étudiant.

Exercice 2. On a demandé à 300 jeunes collégiens quel était leur fruit préféré parmi les six fruits les plus consommés au Maroc : banane, nectarine, orange, pêche, poire, pomme. Voici les résultats obtenus :

Fruit	Banane	Nectarine	Orange	Pêche	Poire	Pomme
Effectif	72	33	30	36	45	84

- 1) Identifier la variable et préciser sa nature.
- 2) Dresser le tableau des fréquences.
- 3) Existe-t-il un mode ? si oui, donner le.
- 4) Faire deux représentations graphiques de cette variable.

Exercice 3. Une enquête en vue de la réduction du montant des allocations familiales, a été réalisée auprès d'un échantillon de 100 femmes de 40 ans. Cette enquête a donné les résultats suivants :

Nombre d'enfants (x_i)	Nombre de femmes (n_i)
0	10
1	20
2	20
3	30
4	20

- 1) Caractériser la distribution (la population et sa taille, l'individu, les modalités, le caractère (la variable) et son type).
- 2) Tracer le diagramme correspondant.
- 3) Définir et représenter la courbe cumulative croissante.
- 4) Donner la proportion (fréquence) des femmes ayant moins de 4 enfants.
- 5) Donner la fréquence des femmes ayant au moins 2 enfants.

Exercice 4. On a relevé la taille (en *cm*) de 50 étudiantes de la filière **SMI**, les résultats sont regroupés dans le tableaux suivant

Classe	[151.5, 155.5[[155.5, 159.5[[159, 5; 163, 5[[163, 5; 167, 5[[167, 5; 171, 5[
Effectif	10	12	11	7	10

1. Caractériser la distribution (*la population et sa taille, l'individu, la variable et son type*).
2. Dresser le tableau statistique complet (calculer les fréquences, les fréquences cumulées et les effectifs cumulés)
3. Tracer le diagramme correspondant.
4. Quelle est la classe modale ?
5. Définir et représenter la courbe cumulative croissante.
6. Calculer la moyenne et la variance.
7. Calculer le coefficient de variation. Interpréter le résultat.
8. Calculer la médiane ainsi que le premier et le troisième quantile.
9. Quelle est la fréquence des étudiantes ayant au moins 165cm ?

Exercice 5. Les données suivantes sont les frais d'électricité (en DH) durant le mois de mars pour un échantillon de 50 petits appartements dans une grande ville :

80	90	95	96	102	108	109	111	114	116
119	123	127	128	129	130	130	135	137	139
141	143	144	147	148	149	149	150	151	153
154	157	158	163	165	166	167	168	171	172
175	178	183	185	187	191	197	202	206	220

- 1) Calculer les quartiles Q_1 , Q_2 et Q_3 de cet échantillon.
- 2) Regrouper les données en classes d'amplitudes égales puis construire le tableau des fréquences et fréquences cumulées.
- 3) Tracer l'histogramme et le polygone des fréquences.

Exercice 6. Dans une enquête, menée auprès des étudiants de l'université Mohammed V, l'enquêteur relevait le temps (en minutes) mis par chaque répondant pour se rendre à l'université. Le tableau suivant résume les temps observés.

Classe	[21; 22[[22; 23[[23; 24[[24; 26[[26; 30[
Effectif	50	90	70	60	40

- 1) Dresser le tableau des fréquences cumulées.
- 2) Représenter ces données à l'aide d'un histogramme et tracer le polygone des fréquences.
- 3) Calculer la moyenne et la variance.
- 4) Calculer les quartiles.
- 5) Cette distribution est-elle symétrique ou asymétrique ?

Exercice 7. Une étude de budget a donné les résultats suivants :

Budget	[800, 1000[[1000, 1400[[1400, 1600[[1600, y[[y, 2400[[2400, x[
Fréq. cumulée	0.08	0.18	0.34	0.64	0,73	1

PARTIE 1 : Certaines données sont manquantes.

- 1) Calculer la borne manquante x sachant que l'étendue de la série est égale à 3200 euros.
- 2) Calculer la borne manquante y dans les deux cas suivants
 - a) le budget moyen est égal à 1995 euros,
 - b) le budget médian est égal à 1920 euros.

PARTIE 2 : Considérons maintenant que la borne manquante y est égale à 2000 euros.

- 3) Donner une représentation graphique de la distribution des budgets.
- 4) Calculer le budget moyen et médian.
- 5) Retrouver les effectifs n_i correspondant à chacune des tranches de budgets ainsi que l'effectif total n , sachant que :

$$\sum_{i=1}^n n_i x_i^2 = 4741200000 \quad \text{et} \quad \mathbb{V}(X) = 604044.$$

Exercice 8. On considère les statistiques (des "frais d'électricité (en DH)" durant le mois de mars pour un échantillon de 50 petits appartements dans une grande ville) données dans l'**Exercice 5**.

- 1) Construire le diagramme en boîte.
- 2) Interpréter les résultats.

Exercice 9. Une voiture roule pendant 200 kilomètres à 50km/h , puis pendant 100 kilomètres à 100km/h . Quelle est sa vitesse moyenne sur son trajet ?

Exercice 10. Soit le tableau suivant donnant la distribution du couple (X, Y) .

$X \backslash Y$	0	1
$[0.5, 1.5[$	21	8
$[1.5, 2.5[$	23	15
$[2.5, 3.5[$	10	23

- 1) Quelles sont les distributions marginales de X et de Y ?
- 2) Calculer les moyennes et les variances marginales de X et de Y .
- 3) Calculer le coefficient de variation marginale de Y . Interpréter.
- 4) Les variables X et Y sont-elles indépendantes ?
- 5) Calculer la moyenne et la variance de la variable $Z = 0.165X + 0.13Y$.

Exercice 11. Un responsable bancaire aimerait savoir s'il existe une relation entre le revenu annuel X et le montant d'argent Y consacré à l'épargne. Pour un échantillon de 10 familles, il a obtenu les résultats suivants (en 10^4DH)

X	12	15	13	10	10	14	16	18	16	14
Y	0,2	1,2	1	0,7	0,3	1	1,6	1,4	1,2	0,7

- 1) Calculer les moyennes \bar{x}, \bar{y} et les écart-types s_x, s_y :
- 2) Calculer le coefficient de corrélation et interpréter ce résultat.
- 3) Chercher l'expression de la droite de régression de Y en X .
- 4) Estimer le montant d'argent consacré à l'épargne par une famille ayant un revenu de 110000DH.

Exercices supplémentaires

Exercice 12. Les téléspectateurs sont invités à évaluer une émission en envoyant un message contenant l'une des lettres A, B, C ou D qui représentent respectivement "très bonne émission", "bonne émission", "mauvaise émission" et "très mauvaise émission" ; çà après les évaluations de 32 spectateurs :

B, B, A, C, A, D, A, A, B, C, D, D, C, A, B, B, C, A, D, C, A, A, B, A, C, D, B, B, C, D, B, A

- 1) Caractériser la variable.
- 2) Dresser le tableau de distribution des effectifs et des fréquences.
- 3) Tracer une représentation graphique associée.

Exercice 13. Les durées, en minutes et arrondies à l'entier le plus proche, enregistrées pour 22 communications, dans un centre d'appel, sont données dans le tableau suivant :

10	12	14	14	15	15	16	16	17	17	17
18	18	18	19	19	20	20	21	22	23	24

- 1) Établir le tableau des fréquences et fréquences cumulées, en utilisant cinq classes.
- 2) Construire l'histogramme de cette série statistique.
- 3) Représenter graphiquement la courbe cumulée croissante.
- 4) A quelle valeur sont inférieures 25% des durées observées ?
- 5) Calculer la moyenne et l'écart-type de cette distribution.
- 6) Calculer le coefficient de variation. Cette distribution statistique est-elle homogène ?

Exercice 14. Dans une usine spécialisée dans la conception d'appareils médicaux, la répartition des employés classés d'après leur âge, en années, est présentée dans le tableau suivant :

Age (en année)	[20, 25[[25, 30[[30, 35[[35, 40[[40, 45[[45, 50[[50, 55[
Nombre d'employés	18	54	72	84	36	22	14

- 1) Quelle est la nature de cette variable ?
- 2) Construire le tableau des fréquences.
- 3) Calculer l'âge médian des employés de l'usine.
- 4) Sachant que 70% de ces employés dépasse un âge x , en années, quelle est alors la valeur de x ?
- 5) Quelle est la moyenne et l'écart-type de cette distribution statistique.
- 6) Cette distribution est-elle asymétrique ? justifier votre réponse.

Exercice 15. L'analyse du taux de calcium X du sérum humain sur 100 personnes a donné les résultats suivants :

Taux de calcium en mg/l	[430, 440[[440, 450[[450, 460[[460, 470[[470, 480[[480, 490]
effectif	11	25	35	19	7	3

- 1) Dresser le tableau des fréquences et des fréquences cumulées.
- 2) Construire l'histogramme et le polygone des fréquences.
- 3) Déterminer la médiane M_e .
- 4) Calculer le taux de calcium moyen \bar{x} et l'écart type s .
- 5) En déduire le mode M_o à partir de la relation suivante : $\bar{x} - M_o = 3(\bar{x} - M_e)$.

Exercice 16. L'objectif de cet exercice est d'étudier le degré d'un certain type de polluants X , Sur 400 endroits différents, on a mesuré le degré (en $10ppm$) de ce polluant. Les résultats sont regroupés dans le tableau suivant :

$Classe_i$	c_i	n_i	N_i	f_i	F_i
[0, 10[55		
[10, 20[120		
[20, 30[220		
[30, 40[300		
[40, 50[360		
[50, 60[400		
Σ	—		—		—

— c_i : centre de la classe i — n_i : effectif — N_i : effectif cumulé — f_i : fréquence — F_i : fréquence cumulée.

- 1) Préciser la variable étudiée sa nature.
- 2) Recopier et compléter le tableau.
- 3) Construire l'histogramme de la distribution.
- 4) Quel est le degré modal de ce type de polluants ?
- 5) Calculer la moyenne et la variance. En déduire l'écart-type.
- 6) Tracer la courbe des fréquences cumulées et déterminer graphiquement les quantiles (Q_1, Me, Q_3)
- 7) Calculer la médiane et les quartiles Q_1 et Q_3
- 8) Calculer le coefficient de variation et conclure ?
- 9) Les responsable sanitaires ont décidé de soumettre indépendamment les 400 endroits à un type de désinfectants Y qui permettra de dégrader le degré du polluant. Sachant que la moyenne de Y vaut $\bar{y} = -5.25$ $10ppm$ et sa variance est égale à $S_Y^2 = 7.5$ ($10ppm$)², calculer la moyenne et l'écart-type de la nouvelle variable $Z = X + 2Y$.

Exercice 17. Dans le but d'évaluer la relation entre la *densité des grains semés* et le *rendement*, on a procédé à une série d'essais sur différentes parcelles d'une céréale. L'expérimentation a donné les résultats suivants :

x_i	150	250	350	450
z_i	57.06	60.73	62.73	63.48
y_i				

avec, x_i désigne le *nombre de grain semés* par m^2 et z_i désigne le *rendement* par hectare.

- 1) Calculer les nombres $y_i = \ln(64 - z_i)$, pour $i = 1, 2, 3, 4$.
- 2) Calculer le coefficient de corrélation linéaire entre x et y . Interpréter.
- 3) Déterminer l'équation de la droite de régression de y en x . En déduire une expression de z en fonction de x .