

Module : Analyse des données

BOUYACOUB Brahim

Table des matières



I - Chapitre 2 : Analyse descriptive	3
1. Les mesures de tendance	3
1.1. La moyenne	4
1.2. La médiane	4
1.3. Le mode	4
2. Les mesures de dispersion	4
2.1. L'écart type	5
2.2. La variance	5
2.3. L'étendue	6
2.4. Le coefficient de variation	6
3. La représentation graphique des données	6
3.1. Diagramme en bâton	7
3.2. Diagramme circulaire	7
3.3. Histogramme	8
3.4. Nuage de point	8
4. Activités d'apprentissages	9
4.1. Exercice 01	9
4.2. Exercice 02	10
4.3. Exercice 03	10

Chapitre 2 : Analyse descriptive



Les mesures de tendance	3
Les mesures de dispersion	4
La représentation graphique des données	6
Activités d'apprentissages	9

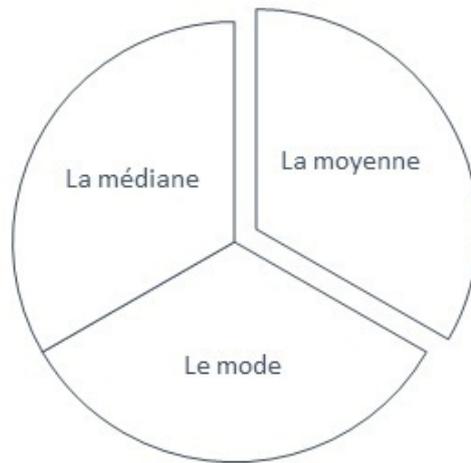
L'analyse descriptive comme le nom l'indique sert à analyser et décrire des données pour obtenir un rendu final. Ce sont de simples calculs mathématiques qui permettent de dégager des données une réelle tendance positive ou négative des résultats.

1. Les mesures de tendance

La moyenne	4
La médiane	4
Le mode	4

Les mesures de tendance centrale (la moyenne, la médiane et le mode) vous aident à saisir, à l'aide d'un nombre unique, ce qui est typique d'un ensemble de données.





Les mesures de tendances

1.1. La moyenne

En mathématiques, la moyenne est un outil de calcul permettant de résumer une liste de valeurs numériques en un seul nombre réel.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{\sum_{n=1}^N X_n}{N}.$$

La moyenne

1.2. La médiane

La médiane d'un ensemble de valeurs (échantillon, population, distribution de probabilités) est une valeur x qui permet de couper l'ensemble des valeurs en deux parties égales

$$M = \begin{cases} X_{N/2} & \text{si } N \text{ est pair} \\ X_{\lfloor N/2 \rfloor + 1} & \text{si } N \text{ est impair} \end{cases}$$

La médiane

1.3. Le mode

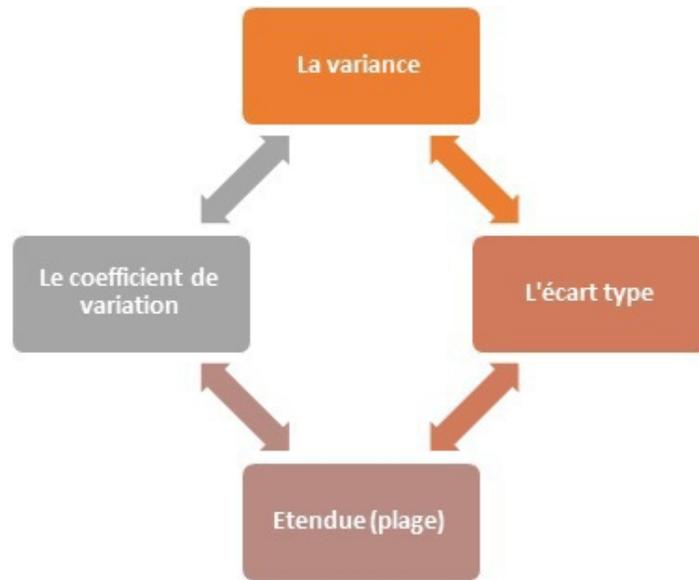
Il est défini comme la valeur la plus fréquente dans la série d'observation.

(Cette valeur n'est pas nécessairement unique).

2. Les mesures de dispersion

L'écart type	5
La variance	5
L'étendue	6
Le coefficient de variation	6

Comme le nom l'indique, les indicateurs de dispersions permettent de mesurer comment les données se «répartissent».



Les mesures de dispersion

2.1. L'écart type

Il est défini comme la racine carrée de la variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \sum_{n=1}^N (X_n - \bar{X})^2}$$

L'écart type

2.2. La variance

La variance est une mesure de la dispersion des valeurs d'un échantillon ou d'une distribution de probabilité.

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N (X_n - \bar{X})^2$$

La variance

2.3. L'étendue

L'étendu d'une série statistique est l'écart entre sa plus grande valeur et sa plus petite. En d'autre terme, l'étendu c'est la différence entre la valeur la plus élevé et la valeur la plus bas.

$$e = \max X - \min X$$

L'étendue

2.4. Le coefficient de variation

Le coefficient de variation (CV) est le rapport de l'écart-type à la moyenne. Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande. Il est généralement exprimé en pourcentage.

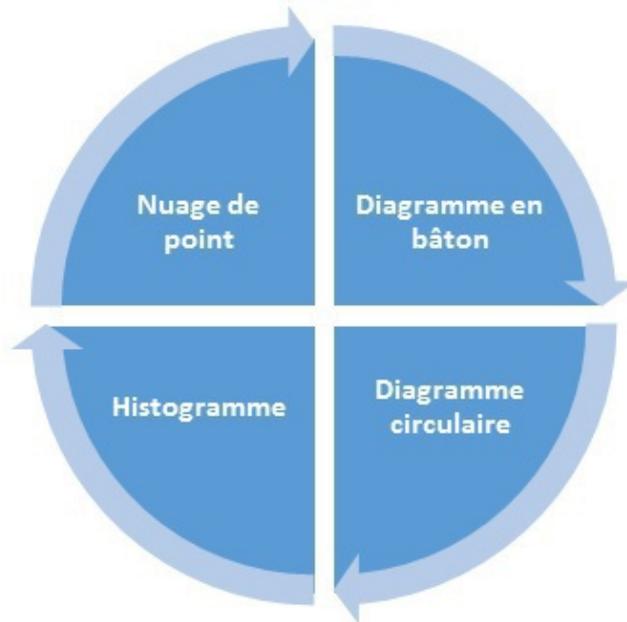
$$CV = \frac{\text{l'écart type}}{\text{la moyenne}} * 100$$

Le coefficient de variation

3. La représentation graphique des données

Diagramme en bâton	7
Diagramme circulaire	7
Histogramme	8
Nuage de point	8

Une représentation graphique de données ou visualisation de données est un résumé visuel des données.



Les types graphiques

3.1. Diagramme en bâton

Définition

Les diagrammes en bâtons sont utilisés pour représenter des séries à variable discrète.

Exemple

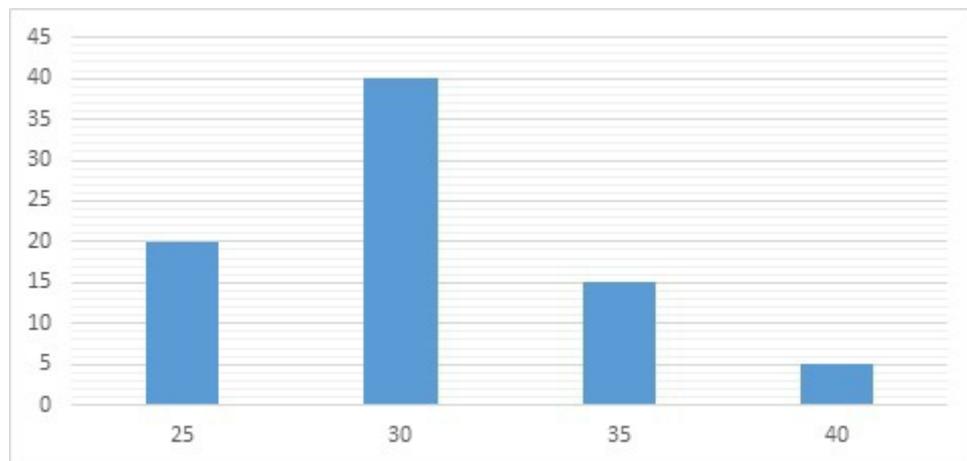


Diagramme en bâton

3.2. Diagramme circulaire

Définition

Un diagramme circulaire admet pour support un disque découpé en secteurs dont les aires sont proportionnelles aux pourcentages des différents constituants de la population statistique.

Exemple

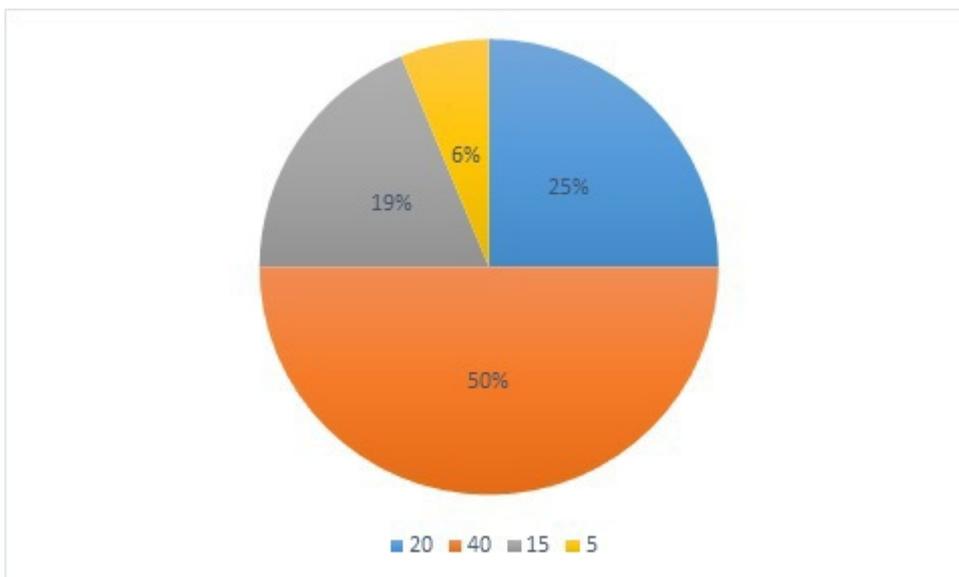


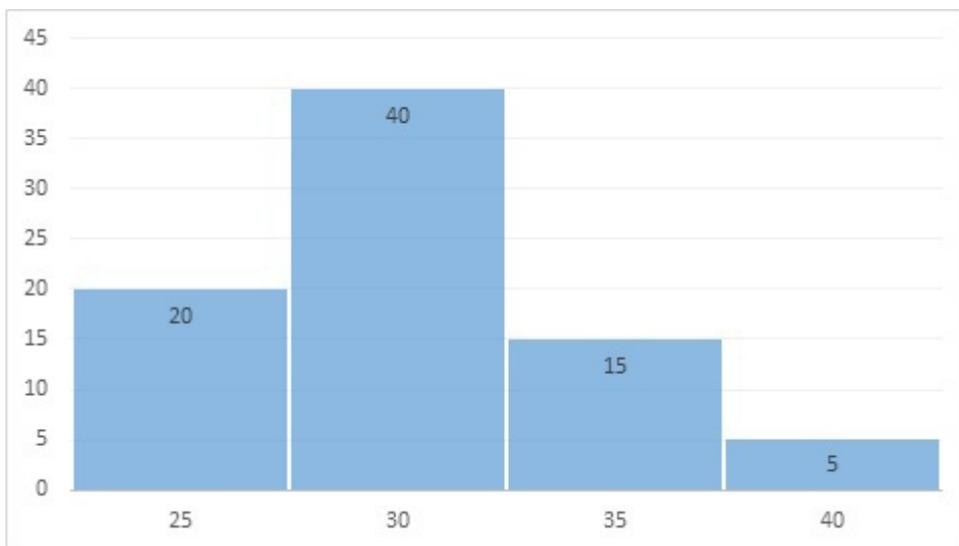
Diagramme circulaire

3.3. Histogramme

Définition

L'histogramme est un moyen rapide pour étudier la répartition d'une variable.

Exemple



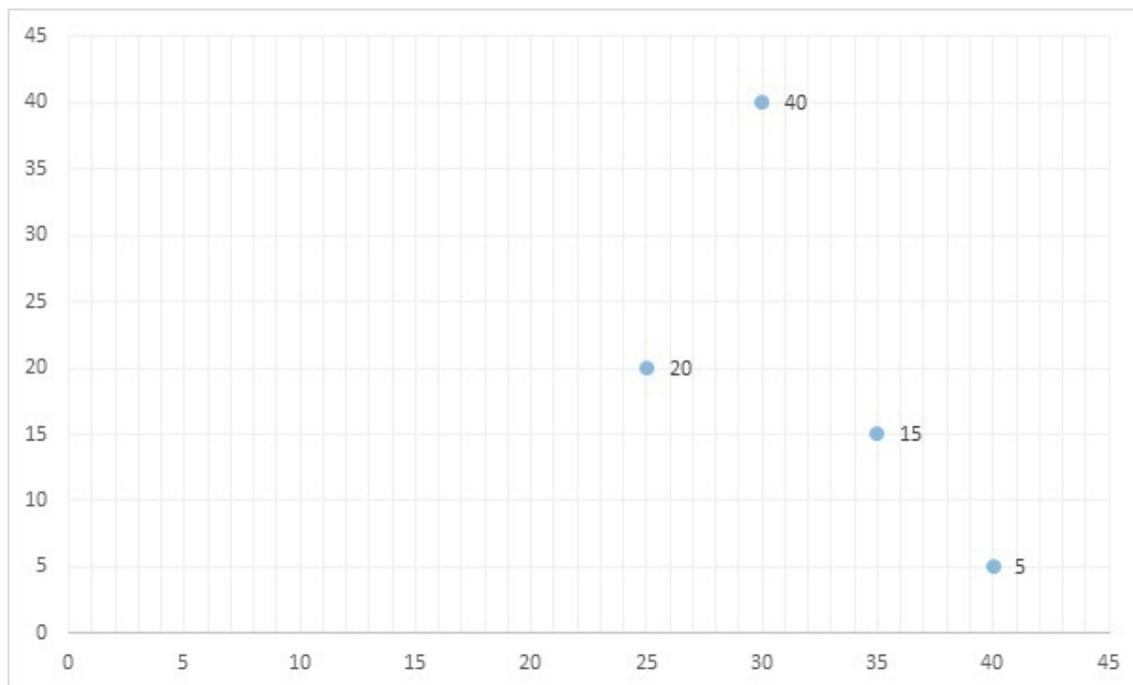
Histogramme

3.4. Nuage de point

Définition

Un nuage de points est une représentation de données dépendant de plusieurs variables.

 **Exemple**



Nuage de point

4. Activités d'apprentissages

Exercice 01	9
Exercice 02	10
Exercice 03	10

4.1. Exercice 01

Les données suivants représente le taux d'inflation durant la période 2000 jusqu'à 2005 :

Années	2000	2001	2002	2003	2004	2005
Taux d'inflation	4,5	5	8	11	5	7

Question

- 1) Calculer l'étendu du taux d'inflation durant la période 2000 - 2005.
- 2) Calculer le coefficient de variation du taux d'inflation.
- 3) Calculer les paramètres classiques de tendance centrale et de dispersion des observations (Faire une analyse descriptive).
- 4) Réaliser un diagramme en bâton du taux d'inflation.
- 5) Réaliser le graphe du nuage de point du taux d'inflation durant cette période.

4.2. Exercice 02

Les chiffres suivants représentent les taux de la variable de la croissance économique :

1	2	1	2	5	0	1	3	2	2	5	0	2	1
2	4	3	2	4	0	2	1	1	2	3	5	2	0
0	3	2	1	0	6	2	1	3	2	0	3	4	1

Question

- 1) Calculer les paramètres classiques de tendance centrale et de dispersion des observations (Faire une analyse descriptive).
- 2) Déterminer la distribution de fréquences sous forme d'un tableau et d'un graphique.

4.3. Exercice 03

Les données suivantes correspondent aux indicateurs économiques tels que la croissance économique et l'inflation en Algérie durant la période 2010 et 2018 :

Année	PIB	inflation
2010	3,63	3,91
2011	2,89	4,52
2012	3,37	8,89
2013	2,77	3,25
2014	3,79	2,92
2015	3,76	4,78
2016	3,3	6,4
2017	1,6	5,59
2018	2,1	4,27

Question

- 1) Déterminer les caractéristiques de la tendance centrale et de dispersion des variables PIB et Inflation.
- 2) Représenter graphiquement ces données pour montrer l'évolution du PIB avec l'inflation.
- 3) Quel est le pourcentage cumulé de la fréquence du PIB en 2016.
- 4) Quel est le pourcentage cumulé de la fréquence de l'inflation en 2014.