République algérienne démocratique et populaire

Université Mohamed BOUDIAF M’sila

Département génie mécanique

**Conférence sur :**

Statistiques descriptives.

M. Houari ZEGGANE

Février 2020

1. Statistique descriptive uni variée

La statistique est une méthode scientifique qui consiste à réunir des données chiffrées sur des ensembles nombreux, puis à analyser, à commenter et à critiquer ces données. En d’autres termes, c’est une science qui a pour objectif : la planification du projet ; la collecte, la codification, la saisie, le traitement et l’analyse des données ; la publication des résultats.

* 1. Série statistique qualitative

Lorsque le caractère étudié est qualitatif, chaque caractère sera indexé et pour chaque variété du caractère, on indiquera les nombres de la population ayant cette variété : c’est une série statistique qualitative.

**Exemple :**

On considère comme population 100 nouveau-nés et le caractère est le sexe.

On indexe les garçons par G et les filles par F.

La série sera par exemple :

G:63, F:37

* 1. Série statistique quantitative

Lorsque le caractère étudié est exprimable directement par un nombre, l’énumération des nombres exprimant la valeur de ce caractère pour chaque membre de la population étudiée est une série statistique quantitative.

Exemple :

On considère comme population 11 adolescents et le caractère est la taille exprimée en centimètres. La série est obtenue par simple énumération :

155,147,153,154,155,148,151,162,144,159,156.

|  |  |
| --- | --- |
| Adolescents | La taille |
| 1 | 155 |
| 2 | 147 |
| 3 | 153 |
| 4 | 154 |
| 5 | 155 |
| 6 | 147 |
| 7 | 151 |
| 8 | 162 |
| 9 | 147 |
| 10 | 159 |
| 11 | 156 |

Ou

|  |  |
| --- | --- |
| taille | effectifs |
| 155 | 2 |
| 147 | 3 |
| 153 | 1 |
| 154 | 1 |
| 151 | 1 |
| 162 | 1 |
| 159 | 1 |
| 156 | 1 |

* 1. Population-variable
     1. Population : ensemble des observations.

C’est l’ensemble des individus (ou unités statistiques) présentant un caractère commun. Pour une thématique donnée, la population regroupe toujours la totalité des individus relatifs à cette thématique (notion d'exhaustivité).

* + 1. Échantillon :

C’est un sous-ensemble construit et représentatif d'une population donnée.

* + 1. Variable :

C’est la (les) caractéristique(s) de l'individu intégrant la population étudiée.

* 1. Les fréquences relative et cumulée

On appelle effectif de la modalité 𝐱𝐢 , le nombre 𝐧𝐢 . Il est aussi **appelé fréquence absolue**.

**La fréquence relative** est le nombre *fi*  tel que : La fréquence d’une modalité est l’effectif divisé par le nombre d’unités d’observation

.

*fi* peut être présenter en pourcentage

**Fréquence cumulée** :

La fréquence cumulée croissante est cependant le nombre 𝑭𝒊 tel que :

***Exemple***

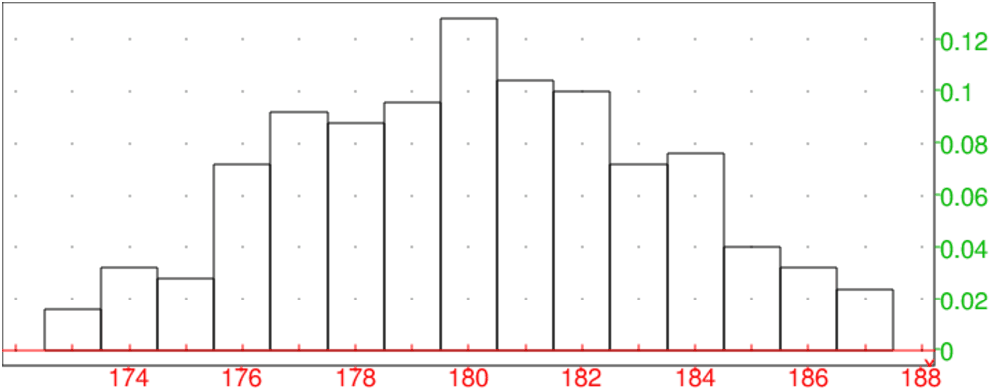
|  |  |
| --- | --- |
| *xi* | *ni* |
| 1 | 8 |
| 2 | 18 |
| 3 | 14 |
| 4 | 10 |
| Total | **50** |

***Solution***

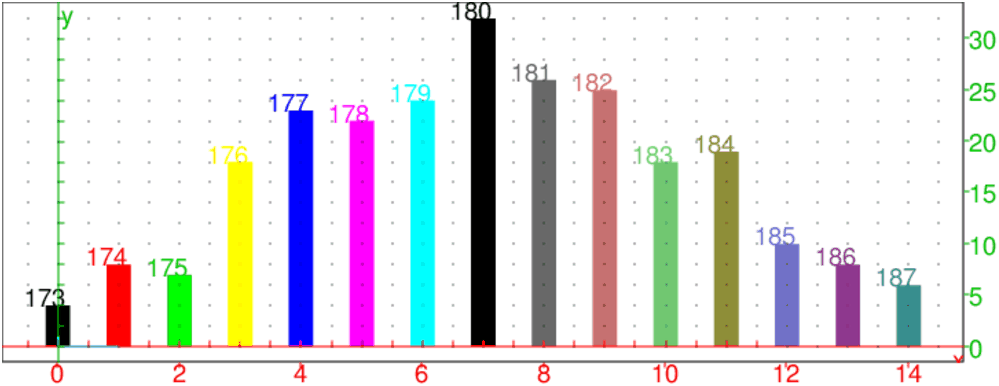
|  |  |  |  |
| --- | --- | --- | --- |
| *xi* | *ni* | *fi* | *Fréquence cumulée Fc* |
| 1 | 8 | 8/50=0.16 | 0.16 |
| 2 | 18 | 18/50=0.36 | 0.16+0.36=0.52 |
| 3 | 14 | 14/50=0.28 | 0.28+0.52=0.8 |
| 4 | 10 | 10/50=0.2 | 0.2+0.8=1 |
| Total | **50** | **50/50=1** |  |

* + 1. Présentation graphique

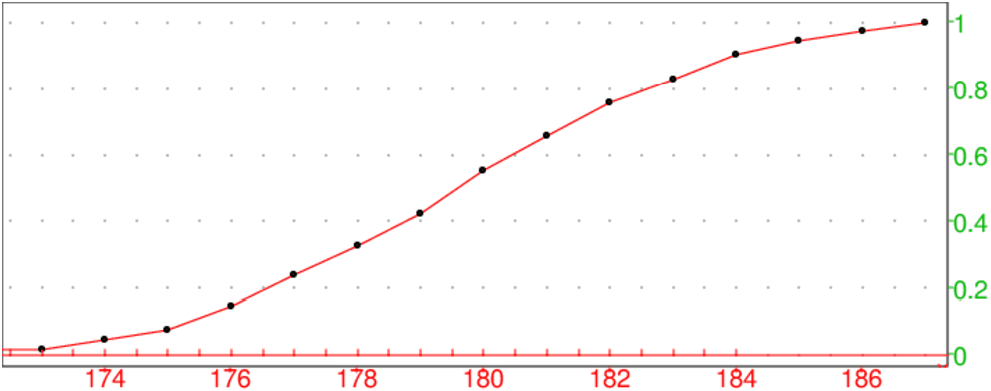
1. Histogramme



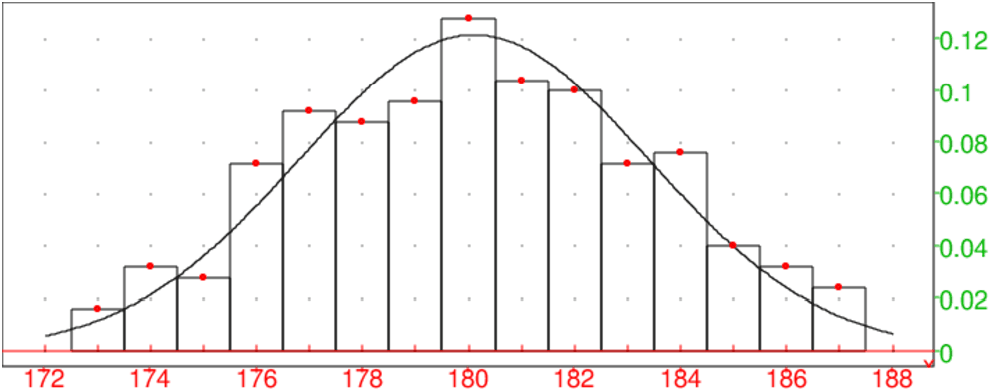
1. Diagramme en bâtons



1. Diagramme pour la fréquence cumulée



1. Diagramme pour la distribution



* + 1. Les quartiles

Les quartiles sont trois valeurs du caractère qui partage la série statistique en quatre groupes de même effectif :

* Le 1-ier quartile est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.25.
* Le 2-ième quartile est confondu avec la médiane.0.5
* Le 3-ième quartile est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.75

**Exercice**

Soient les données suivantes :

|  |  |
| --- | --- |
| Taille | Effective |
| 173 | 4 |
| 174 | 8 |
| 175 | 7 |
| 176 | 18 |
| 177 | 23 |
| 178 | 22 |
| 179 | 24 |
| 180 | 32 |
| 181 | 26 |
| 182 | 25 |
| 183 | 18 |
| 184 | 19 |
| 185 | 10 |
| 186 | 8 |
| 187 | 6 |

* Calculer les fréquences relative et cumulée.
* Présenter résultats sous forme d’histogramme.
* Déterminer les quartiles (1,2,3)

**Solution**

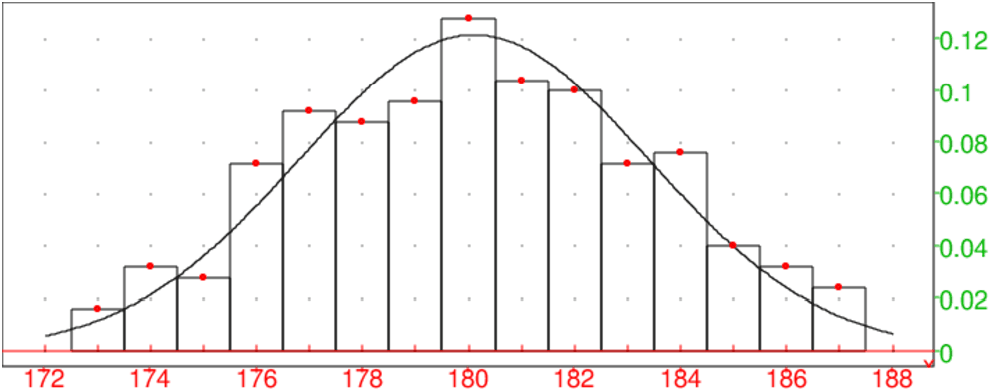
|  |  |  |  |
| --- | --- | --- | --- |
| Taille | Effective | fi | Fc |
| 173 | 4 | 0,016 | 0,016 |
| 174 | 8 | 0,032 | 0,048 |
| 175 | 7 | 0,028 | 0,076 |
| 176 | 18 | 0,072 | 0,148 |
| 177 | 23 | 0,092 | 0,24 |
| 178 | **22** | **0,088** | **0,328** |
| 179 | 24 | 0,096 | 0,424 |
| 180 | **32** | **0,128** | **0,552** |
| 181 | 26 | 0,104 | 0,656 |
| 182 | **25** | **0,1** | **0,756** |
| 183 | 18 | 0,072 | 0,828 |
| 184 | 19 | 0,076 | 0,904 |
| 185 | 10 | 0,04 | 0,944 |
| 186 | 8 | 0,032 | 0,976 |
| 187 | 6 | 0,024 | 1 |

Les digrammes

Histogramme de fréquence : taille-fréquence.

Histogramme taille-fréquence cumulée.

Courbe taille-fréquence cumulée.



Distribution de fréquence.

* 1. Caractéristiques de tendance centrale des données

Les paramètres de tendance centrale ou « mesures de tendance centrale » sont des grandeurs susceptibles de représenter au mieux un ensemble de données. L'appellation «tendance centrale » vient du fait que ces paramètres donnent une idée de ce qui se passe au centre d'une distribution, d'un ensemble de données. On distingue trois mesures de tendance centrale :

* La moyenne ;
* Le mode ;
* La médiane.
  + 1. La moyenne

|  |  |
| --- | --- |
| Désignations | Notation courante |
| Moyenne arithmétique |  |
| Moyenne géométrique | , |
| Moyenne harmonique | , |
| Moyenne quadratique | , |

***Moyenne arithmétique***

***Simple***

 : moyenne arithmétique.

n : nombre d’observation.

 : valeur observée pour l’individu (i).

***Pondérée***

Pi : le poids.

**Exemple**

* Calcul des notes des étudiants avec des coefficients des matières.
* Coefficient de ruissèlement moyen pour une surface hétérogène.

***Moyenne géométrique***

La moyenne géométrique est un instrument permettant de calculer des taux moyens notamment des taux moyens annuels. Son utilisation n’a un sens que si les valeurs ont un caractère multiplicatif.

**Exemple**

Les prix de l’immobilier ancien ont augmenté ces trois dernières années de la façon suivante : 2, 4, 8.

*Moyenne harmonique*

On utilise la moyenne harmonique lorsqu’on veut déterminer un rapport moyen dans des domaines où il existe des liens de proportionnalité inverse.

• Pour une distance donnée, le temps de trajet est d’autant plus court que la vitesse est élevée. • Un loyer dans le parc privé est d'autant plus élevé que la taille ou la surface du logement est petite.

*Moyenne quadratique*

Une moyenne qui trouve des applications lorsque l'on a affaire à des phénomènes présentant un caractère sinusoïdal avec alternance de valeurs positives et de valeurs négatives.

* + 1. La mode

Le mode, 𝑴𝒐 d’une série statistique est la valeur du caractère la plus fréquente ou dominante dans l'échantillon.

**Exemple :**

1. Donner le mode des séries des données suivantes 𝑆1 = {9, 10, 9, 9, 11, 10, 11, 11, 11, 9}

|  |  |
| --- | --- |
| Individus | Effectifs |
| 9 | 4 |
| 10 | 2 |
| 11 | 4 |

Le mode est de **Mo = 4**

Par contre, si l'on est en présence de données groupées en classes, le mode se rapportera à la classe comportant le plus grand nombre d'individus : on parlera alors de classe modale.

Xm : limite inférieure de la classe modale ;

i : amplitude de la classe modale ;

Δi : écart d’effectif entre la classe modale et la classe inférieure la plus proche

Δs : écart d’effectif entre la classe modale et la classe supérieure la plus proche

**Exemple :**

|  |  |
| --- | --- |
| Individus | Effectifs |
| [0-5[ | 10 |
| [5-10[ | 15 |
| [10-15[ | 12 |
| *[15-25[* | *25* |
| [25-30[ | 10 |

La classe modale : [15-25[

Xm : 15 ;

i : 25-15=10;

Δi : 25-12= 10 : écart d’effectif entre la classe modale et la classe inférieure la plus proche

Δs : 25-10= 15 écarts d’effectif entre la classe modale et la classe supérieure la plus proche

* + 1. La médiane

La **médiane** partage la série statistique en deux groupes de même effectif. C’est une valeur du caractère à partir de laquelle l’effectif des valeurs qui lui sont inférieures est supérieur ou ègal à l’effectif des valeurs qui lui sont supérieures

Exemple :

[140,145,146,147] : ma médiane est 146

[140,145,146] est 145). La médiane est 145.

* 1. Caractéristiques de dispersion
     1. Étendue

 L’étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

Étendue de X = Xmax - Xmin

* + 1. La variance et l’écart type (pour un échantillon)

La variance, notée **( x) ²** est la moyenne du carré des écarts à la moyenne.

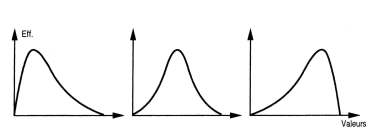
Pour un échantillon.

Pour une population.

L’écart type S

* + 1. Coefficient de variation (CV)
    2. Coefficient de forme (coefficient de d’asymétrie de Fisher(skewness)).

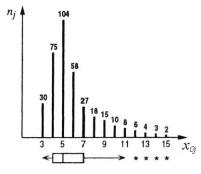
On distingue trois types de distributions selon qu'elles sont dissymétriques (asymétriques) à gauche (graphique de gauche), symétriques (graphique du milieu) ou dissymétriques (asymétriques) à droite (graphique de droite).



m3<0

m3=0

m3>0



pour une population.

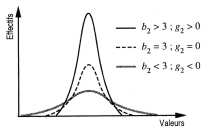
pour une échantillon.

m3<dissymétrie à gauche.

m3> dissymétrie à droite.

m3 =0 distribution symétrique.

* + 1. Coefficient de forme (coefficient d’aplatissement) (kurtosis).



Pour une population.

Pour un échantillon.

1. Statistique descriptive bivariée
   1. Série statistique bivariée

On s’intéresse `a deux variables *x* et *y*. Ces deux variables sont mesurées sur les *n* unités d’observation. Pour chaque unité, on obtient donc deux mesures. La série statistique est alors une suite de *n* couples des valeurs prises par les deux variables sur chaque individu :

* 1. Deux variables quantitatives
     1. Représentation graphique de deux variables

Dans ce cas, chaque couple est composé de deux valeurs numériques. Un couple de nombres (entiers ou réels) peut toujours être représenté comme un point dans un plan.

**Exemple 1 :** On mesure le poids y et de taille xde 20 individus.

|  |  |  |  |
| --- | --- | --- | --- |
| Yi (kg) | Xi (cm) | Yi (kg) | Xi (cm) |
| 60 | 155 | 75 | 180 |
| 61 | 162 | 76 | 175 |
| 64 | 157 | 78 | 173 |
| 67 | 170 | 80 | 175 |
| 68 | 164 | 85 | 179 |
| 69 | 162 | 90 | 175 |
| 70 | 169 | 96 | 180 |
| 71 | 170 | 96 | 185 |
| 72 | 178 | 98 | 189 |
| 73 | 173 | 101 | 187 |

**Figure.** Le nuage de points.

* + 1. Analyse des variables

Les variables *x* et  *y* peuvent être analysées s´séparément. On peut calculer tous les paramètres dont les moyennes et les variances :

; ; ;

* + 1. Covariance

La covariance est définie

* + 1. Corrélation

Le coefficient de corrélation rxy est la covariance divisée par deux écarts types marginaux

Le coefficient de détermination r2 est donné par la relation suivante :

Le coefficient de corrélation mesure la d´dépendance linéaire entre deux variables :

* −1 ≤ rxy ≤ 1
* 0 ≤ ≤ 1

– Si le coefficient de corrélation est positif, les points sont alignés le long d’une droite croissante.

– Si le coefficient de corrélation est négatif, les points sont alignés le long d’une droite d´décroissante.

– Si le coefficient de corrélation est nul ou proche de zéro, il n’y a pas de d´dépendance linéaire.

On peut cependant avoir une dépendance non-linéaire avec un coefficient de corrélation nul.