STATISTIQUE DESCREPTIVE A UNE DIMENSION

Dr. BEGHRICHE A/FATEH

2019 /2020

Plan

Plan

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une série statistique : variable qualitative

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)
5. Etude d'une sérié statistique : variable quantitative

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)
5. Etude d'une sérié statistique : variable quantitative
   - Données condensées (k est petit par rapport à N) ou Varible discrète

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une série statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)
5. Etude d'une série statistique : variable quantitative
   - Données condensées (k est petit par rapport à N) ou Varible discrète
   - Distribution des fréquences

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une série statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)
5. Etude d'une série statistique : variable quantitative
   - Données condensées (k est petit par rapport à N) ou Varible discrète
   - Distribution des fréquences
   - Représentation graphique

Plan

1. Introduction
2. Préparation d'une étude statistique
   - Population/Echontillon/Individu
   - Variable /Modalité/ Nature
   - Collecte des données
3. Etude d'une sérié statistique : variable qualitative
   - Distribution des fréquences
4. Représentation graphique
   - Diagramme en colonnes(tuyaux d'orgue)
   - Diagrammes à secteurs (secteurs angulaires)
5. Etude d'une sérié statistique : variable quantitative
   - Données condensées (k est petit par rapport à N) ou Varible discrète
   - Distribution des fréquences
   - Représentation graphique
   - Fréquences cummulées

Introduction

Introduction

La statistique descriptive est un ensemble de méthodes et de technique qui consistent à classer, synthéser et présenter des données par des tableaux, des graphes et des paramètres afin d'avoir une idée plus précise, de les rendre conséquemment exploitables et en tirer le maximum d'informations.

Préparation d'une étude statistique

La population statistique est l'ensemble sur lequel porte l'étude ou la prévision,et **l'échantillon** représente la fraction de cette population qui est réellement observée ou étudiée.

La population statistique est l'ensemble sur lequel porte l'étude ou la prévision,et **l'échantillon** représente la fraction de cette population qui est réellement observée ou étudiée.

- La notion **d'individu** est très large : les éléments d'un échantillon ou d'une population sont appelés généralement des individus, cependant cette notion peut être remplacé par plusieurs dénominations : unité statistique, sujet, objet, élément, observation, mesure, doses,. . ...ext.

Toute étude statistique s'intéresse à un ou plusieurs caractères d'une population. Ce sont les **variables statistiques**.Toute valeur pouvant être prise par une variable s'appelle **modalité** de cette variable.

Toute étude statistique s'intéresse à un ou plusieurs caractères d'une population. Ce sont les **variables statistiques**.Toute valeur pouvant être prise par une variable s'appelle **modalité** de cette variable.
Reconnaître la nature des variables est important puisque leur traitement en dépend.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.Dans ces deux cas la variable mesure une certaine quantité, on dit qu'elle est quantitative. Mais, il existe des variables qui ne sont pas quantitatives, on dit qu'elles sont qualitatives.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.Dans ces deux cas la variable mesure une certaine quantité, on dit qu'elle est quantitative. Mais, il existe des variables qui ne sont pas quantitatives, on dit qu'elles sont qualitatives.

Example

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.Dans ces deux cas la variable mesure une certaine quantité, on dit qu'elle est quantitative. Mais, il existe des variables qui ne sont pas quantitatives, on dit qu'elles sont qualitatives.

Example

- Le poids des vaches d'un troupeau donné est une variable (quantitative) continue.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.Dans ces deux cas la variable mesure une certaine quantité, on dit qu'elle est quantitative. Mais, il existe des variables qui ne sont pas quantitatives, on dit qu'elles sont qualitatives.

Example

- Le poids des vaches d'un troupeau donné est une variable (quantitative) continue.
- Le nombre de vache, du troupeau précédent, atteints d'une certaine maladie est une variable (quantitative) discrète.

Une variable pouvant prendre n'importe quelle valeur d'un intervelle est dite continue.
Par contre, une variable ne pouvant prendre que des valeurs isolées est dite discrètes.Dans ces deux cas la variable mesure une certaine quantité, on dit qu'elle est quantitative. Mais, il existe des variables qui ne sont pas quantitatives, on dit qu'elles sont qualitatives.

Example

- Le poids des vaches d'un troupeau donné est une variable (quantitative) continue.
- Le nombre de vache, du troupeau précédent, atteints d'une certaine maladie est une variable (quantitative) discrète.
- La couleur des yeux des étudiants de 1ére année sciences vétérinaires est variable qualitative.

Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.

Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.Mais la plus souvent ceci n'est pas possible;

Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.Mais la plus souvent ceci n'est pas possible; soit le temps d'étude ne le permet pas, Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.Mais la plus souvent ceci n'est pas possible; soit le temps d'étude ne le permet pas, soit le coût de l'opération est trop élevé, Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.Mais la plus souvent ceci n'est pas possible; soit le temps d'étude ne le permet pas, soit le coût de l'opération est trop élevé,soit c'est carrément impossible.

Les données peuvent être, parfois, obtenues par recensement. C'est à dire que toutes les unités de la population sont observées.Mais la plus souvent ceci n'est pas possible; soit le temps d'étude ne le permet pas, soit le coût de l'opération est trop élevé,soit c'est carrément impossible.Ainsi, par exemple, l'orsqu'on s'interesse à la durée de vie de lampes produites par une usine, il n'est pas raisonnable de tester toutes les lampes produites. Autrement dit, l'échantillonage s'impose trés souvent. Il faut donc extraire de la population un sous ensemble qui sera observé et dont les résultats constituent l'échantillon.

Etude d'une sérié statistique : variable qualitative

Etude d'une sérié statistique : variable qualitative

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.Nous nous contentrons, seulement, de voir comment:

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.Nous nous contentrons, seulement, de voir comment:
- Calculer les fréquances relatives.

Etude d'une sérié statistique : variable qualitative

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.Nous nous contentrons, seulement, de voir comment:

- Calculer les fréquances relatives.

- Visualiser par des représentations graphiques

Etude d'une série statistique : variable qualitative

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.Nous nous contentrons, seulement, de voir comment:

- Calculer les fréquances relatives.

- Visualiser par des représentations graphiques

Sur un échantillon de taille n, supposons que l'on a obtenu $k$modalités différentes $(k \leq n)$ puisque certaines observation peuvent se répéter.

Nous n'allons pas nous étendre beaucoup dans d'étude du variable qualitative.Nous nous contentrons, seulement, de voir comment:

- Calculer les fréquances relatives.
- Visualiser par des représentations graphiques

Sur un échantillon de taille n, supposons que l'on a obtenu $k$modalités différentes $(k \leq n)$ puisque certaines observation peuvent se répéter.Nous notons par $(x_i)_{1 \leq i \leq k}$ l'ensemble des observation deux à deux diférentes ainsi obtenu.

Definition

- Nous appelons fréquence (absolue) de la donnée x_i,et on note n_i, le nombre de fois que cette donnée a été observée.

Definition

- Nous appelons fréquence (absolue) de la donnée x_i,et on note n_i, le nombre de fois que cette donnée a été observée.
-Nous appelons fréquence relative de la donnée x_i,et on note f_i le nombre donnée par:

Definition

- Nous appelons fréquence (absolue) de la donnée x_i,et on note n_i, le nombre de fois que cette donnée a été observée.

-Nous appelons fréquence relative de la donnée x_i,et on note f_i le nombre donnée par:

$f_i = \frac{n_i}{N}$ où N est le nombre total des données.

Definition

- Nous appelons fréquence (absolue) de la donnée x_i,et on note n_i, le nombre de fois que cette donnée a été observée.

-Nous appelons fréquence relative de la donnée x_i,et on note f_i le nombre donnée par:

$f_i = \frac{n_i}{N}$ où N est le nombre total des données.

Il est clair que $\sum_{i=1}^{k} n_i = N$ et $\sum_{i=1}^{k} f_i = 1.$

Etude d'une série statistique : variable qualitative

Distribution des fréquences

Example

L'étude de l'état civil de 40 employés d'une entreprise a conduit aux résultats suivants:

Numéro	*Etat civil*	*Numéro*	*Etat civil*	*Numéro*	*Etat civil*	*Num*
1	*Marié*	11	*veuf*	21	*Marié*	31
2	*Marié*	12	*Marié*	22	*Célibataire*	32
3	*Célibataire*	13	*Célibataire*	23	*Marié*	33
4	*Divorcé*	14	*Marié*	24	*veuve*	34
5	*Marié*	15	*Marié*	25	*Marié*	35
6	*Célibataire*	16	*Marié*	26	*Divorcé*	36
7	*Célibataire*	17	*Marié*	27	*Célibataire*	37
8	*Marié*	18	*Célibataire*	28	*Marié*	38
9	*Marié*	19	*Célibataire*	29	*Marié*	39
10	*Divorcé*	20	*Célibataire*	30	*Marié*	40

La distribution des fréquences s'écrit:

La distribution des fréquences s'écrit:

Modalités	*Fréquences*	*Fréquences relatives*
x_i	n_i	f_i
Marié(e)	20	0.50
Célibataire	12	0.30
Divorcé(e)	6	0.15
veuf(ve)	2	0.05
	$\sum n_i = 40$	$\sum f_i = 1$

Représentation graphique

A chaque modalité correspond un rectangle dont la hauteur est égale à la fréquence (absolue ou relative) associée à cette modalité. Les réctangles ont des largeurs égales et sont séparés les uns des autres par des distances égales.
pour l'exemple précédent, il vient

On utilise un disque que l'on partage en autant de secteurs que de modalités prises par notre variable. Les surfaces de ces secteurs sont proportionnelles aux fréquences des modalités. L'angle au centre d'un sécteur est égal à $f_i * 360$ où f_i représente la fréquence relative de la modalité par ce secteur.Pour notre exemple

Etude d'une sérié statistique : variable quantitative

Comme il a été déja expliqué, nous voulons rendre un ensemble de données expoitable.

Comme il a été déja expliqué, nous voulons rendre un ensemble de données expoitable.
Nous commencerons par les ranger par ordre non décroissant.

Comme il a été déja expliqué, nous voulons rendre un ensemble de données expoitable.
Nous commencerons par les ranger par ordre non décroissant.
Si la taille de l’échantillon est petite (infèrieure à 20 à titre indicatif), il est unutile de dresser le tableau de fréquences et de tracer un diagramme puisque cela n’ajoutera pas grand chose à la clarté de la présentation des données.

Comme il a été déja expliqué, nous voulons rendre un ensemble de données expoitable.
Nous commencerons par les ranger par ordre non décroissant.
Si la taille de l'échantillon est petite (infèrieure à 20 à titre indicatif), il est unutile de dresser le tableau de fréquences et de tracer un diagramme puisque cela n'ajoutera pas grand chose à la clarté de la présentation des données.Dans le cas où la taille de l'échantillon est grande, nous distingnons deux cas.

Dans ce cas, il devient utile de faire une étude du même genre que dans le cas qualitatif.

Dans ce cas, il devient utile de faire une étude du même genre que dans le cas qualitatif.Autrement dit, nous dressons le tableau de fréquences et nous traçons le diagramme en bâtons.

Chaque modalité est affecté de sa fréquence (absolue ou relative) qui est définie et notée exactement comme dans le cas qualitatif.

Chaque modalité est affecté de sa fréquence (absolue ou relative) qui est définie et notée exactement comme dans le cas qualitatif.

Example

Une clinique a enregistré le nombre de frères et soeurs de chacun de ses patients atteints de la maladie contagieuse qu'est la varicelle durant l'année 2000. Les données brutes obtenues sont:

Chaque modalité est affecté de sa fréquence (absolue ou relative) qui est définie et notée exactement comme dans le cas qualitatif.

Example

Une clinique a enregistré le nombre de frères et soeurs de chacun de ses patients atteints de la maladie contagieuse qu'est la varicelle durant l'année 2000. Les données brutes obtenues sont:

2	1	3	0	6	0	1	2	3	1
3	0	2	0	4	1	0	4	0	2
1	1	3	2	3	3	2	1	1	1
0	1	2	4	1	2	2	7	3	2
0	1	1	2	5	5	3	4	3	0
1	2	2	3	0	1	2	0	2	2

Ici $N = 60$ et $k = 8$ (petit par rapport à N). La distribution des fréquences est données dans le tableau suivant:

Ici $N = 60$ et $k = 8$ (petit par rapport à N). La distribution des fréquences est données dans le tableau suivant:

x_i	n_i	f_i
0	11	0.183
1	15	0.250
2	16	0.267
3	10	0.167
4	4	0.067
5	2	0.033
6	1	0.017
7	1	0.016

La repréentation graphique est donnée par le diagramme en bâtons. Sur l'axe des abscisses sont représentées les modalités de la variable et sur l'axe des ordonnées sont représentées les fréquences.

La repréentation graphique est donnée par le diagramme en bâtons. Sur l'axe des abscisses sont représentées les modalités de la variable et sur l'axe des ordonnées sont représentées les fréquences. Partant de la valeur x_i, nous traçons un segment parallèlle à l'axe des ordonnées (bâton), de longeur égale à la valeur de la fréquence x_i.

La repréentation graphique est donnée par le diagramme en bâtons. Sur l’axe des abscisses sont représentées les modalités de la variable et sur l’axe des ordonnées sont représentées les fréquences. Partant de la valeur x_i, nous traçons un segment parallèlle à l’axe des ordonnées (bâton), de longeur égale à la valeur de la fréquence x_i.
Pour l’exemple précédent, nous obtenons le graphe suivant:

Dans cet exemple, nous aurions pu nous questionner sur le pourcentage de patients dont le nombre de frères et soeurs est infèrieure à 3, par exemple.

Dans cet exemple, nous aurions pu nous questionner sur le pourcentage de patients dont le nombre de frères et soeurs est infèrieure à 3, par exemple. Ceci nous amène à cumulet les fréquences et à introduire la notion suivante.

- Nous appellons fréquence (absolue) (resp relative) cumulée da la donnée x_i notée N_i (resp F_i),la somme de la fréquence (absolue)(resp relativee) de cette donnée et des fréquences (absolue) (resp relatives) des données qui lui sont infèrieures.

- Nous appellons fréquence (absolue) (resp relative) cumulée da la donnée x_i notée N_i (resp F_i),la somme de la fréquence (absolue)(resp relativee) de cette donnée et des fréquences (absolue) (resp relatives) des données qui lui sont infèrieures.

-L'effectif cumulé associé à chaque valeur x_i ainsi que sa fréquence relative cumulée s'écrivent sur la ligne qui se trouve en dessous de la case où figure x_i.

- Nous appellons fréquence (absolue) (resp relative) cumulée da la donnée x_i notée N_i (resp F_i),la somme de la fréquence (absolue)(resp relativee) de cette donnée et des fréquences (absolue) (resp relatives) des données qui lui sont infèrieures.

-L'effectif cumulé associé à chaque valeur x_i ainsi que sa fréquence relative cumulée s'écrivent sur la ligne qui se trouve en dessous de la case où figure x_i.

- La distribution des fréquences (absolue) (resp relative) cumulées est la donnée des différentes modalités affectées chacune de sa fréquences (absolue) (resp relative) cumulée.

- Pour l'exemple précedent, la distribution des fréquences cumulées est donnée par le tableau suivant:

- Pour l'exemple précedent, la distribution des fréquences cumulées est donnée par le tableau suivant:

x_i	N_i	F_i
0	11	0.183
1	26	0.433
2	42	0.700
3	52	0.867
4	56	0.934
5	58	0.967
6	59	0.984
7	60	1

- Pour l'exemple précedent, la distribution des fréquences cumulées est donnée par le tableau suivant:

x_i	N_i	F_i
0	11	0.183
1	26	0.433
2	42	0.700
3	52	0.867
4	56	0.934
5	58	0.967
6	59	0.984
7	60	1

Nous déduisons, par exemple, que 86.7% des patients ont un nombre de frères et soeurs infèrieurs ou égal à 3.

-La représentation graphique concernant les fréquences cumulées est donnée par le polygône des fréquences cumulées qui est construit en escalier.

-La représentation graphique concernant les fréquences cumulées est donnée par le polygône des fréquences cumulées qui est construit en escalier. Nous plaçons les points dont les abcisses sont les modalités et les ordonnées sont leurs fréquences cumulées.

-La représentation graphique concernant les fréquences cumulées est donnée par le polygône des fréquences cumulées qui est construit en escalier. Nous plaçons les points dont les abcisses sont les modalités et les ordonnées sont leurs fréquences cumulées.Partant de chacun de ces points, nous traçons des segments horizontaux s'arrêtant en regard de la valeur de la modalité suivante (pour montrer que la variable ne peut prendre aucune autre valeur entre ces deux modalités).

-La représentation graphique concernant les fréquences cumulées est donnée par le polygône des fréquences cumulées qui est construit en escalier. Nous plaçons les points dont les abcisses sont les modalités et les ordonnées sont leurs fréquences cumulées.Partant de chacun de ces points, nous traçons des segments horizontaux s'arrêtant en regard de la valeur de la modalité suivante (pour montrer que la variable ne peut prendre aucune autre valeur entre ces deux modalités).puis nous joignons ces segments par des segments verticaux.

Pour l'exemple précédent, nous obtenons le polygône des fréquences suivant:

Pour l'exemple précédent, nous obtenons le polygône des fréquences suivant:
L'utilisation des fréquences relatives, pour le polygône des fréquences cummulées, à l'avantage de n'avoir, en ordonnées, que des valeurs comprises entre 0 et 1 et permettre la comparaison entre des série différentes.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.Ce cas se produit aussi lorsque la variable est discrète mais que l'ensemble de ses modalités possède un grand effectif.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.Ce cas se produit aussi lorsque la variable est discrète mais que l'ensemble de ses modalités possède un grand effectif. Il est clair que dans ce cas une étude semblable à celle effectuée juste au dessus s'avère inutile.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.Ce cas se produit aussi lorsque la variable est discrète mais que l'ensemble de ses modalités possède un grand effectif. Il est clair que dans ce cas une étude semblable à celle effectuée juste au dessus s'avère inutile.
Il est alors indiqué de regrouper les données en classes en prenant le soin que la précision que l'on perd ne nuise pas trop à l'étude.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.Ce cas se produit aussi lorsque la variable est discrète mais que l'ensemble de ses modalités possède un grand effectif. Il est clair que dans ce cas une étude semblable à celle effectuée juste au dessus s'avère inutile.
Il est alors indiqué de regrouper les données en classes en prenant le soin que la précision que l'on perd ne nuise pas trop à l'étude. C'est sans doute qui se pose alors le choix de ces classes.

Notamment lorsque la variable observée est continue les observations sont presque toutes différentes.Ce cas se produit aussi lorsque la variable est discrète mais que l'ensemble de ses modalités possède un grand effectif. Il est clair que dans ce cas une étude semblable à celle effectuée juste au dessus s'avère inutile.

Il est alors indiqué de regrouper les données en classes en prenant le soin que la précision que l'on perd ne nuise pas trop à l'étude. C'est sans doute qui se pose alors le choix de ces classes. Il n'y a pas de règle absolument stricte d'un tel choix mais certaines indications de nous guider.

- Soient $([b_i, b_{i+1}[)_{0 \leq i \leq m-1}$ les m classes à définir.

- Soient $([b_i, b_{i+1}[)_{0\leq i\leq m-1}$ les m classes à définir.
- m varie en général entre 5 et 20 mais de préférence il se situe entre 6 et 12.

- Soient $([b_i, b_{i+1}[)_{0\leq i\leq m-1}$ les m classes à définir.
- m varie en général entre 5 et 20 mais de préférence il se situe entre 6 et 12.
- m doit être choisi de sorte que les fréquences des classes ne soient pas toutes trés petites.

- Soient $([b_i, b_{i+1}[)_{0\leq i\leq m-1}$ les m classes à définir.
- m varie en général entre 5 et 20 mais de préférence il se situe entre 6 et 12.
- m doit être choisi de sorte que les fréquences des classes ne soient pas toutes trés petites.
- Ces intervalles doivent contenir toutes les données.

- Soient $([b_i, b_{i+1}[)_{0\leq i\leq m-1}$ les m classes à définir.
- m varie en général entre 5 et 20 mais de préférence il se situe entre 6 et 12.
- m doit être choisi de sorte que les fréquences des classes ne soient pas toutes trés petites.
- Ces intervalles doivent contenir toutes les données.Pour cela la plus petite donnée doit être plus grande que b_0 et la plus grande plus petite que b_m.

- Soient $([b_i, b_{i+1}[)_{0\leq i\leq m-1}$ les m classes à définir.
- m varie en général entre 5 et 20 mais de préférence il se situe entre 6 et 12.
- m doit être choisi de sorte que les fréquences des classes ne soient pas toutes trés petites.
- Ces intervalles doivent contenir toutes les données.Pour cela la plus petite donnée doit être plus grande que b_0 et la plus grande plus petite que b_m.
- Si possible prendre la même longueur pour toutes les classes, qu'on choisit de préférence un amplitude de $5, 10, 100, 1000$ (pour faciliter les calculs que nous aurons à mener plus loin).

- Eviter la concentration des données aux bornes des intervalles car cela pourrait fausser certains calculs ultérieurs.

- Eviter la concentration des données aux bornes des intervalles car cela pourrait fausser certains calculs ultérieurs.

La formule de Sturges nous donne une indication sur le choix de m en fonction de n, elle est donnée par:

- Eviter la concentration des données aux bornes des intervalles car cela pourrait fausser certains calculs ultérieurs.

La formule de Sturges nous donne une indication sur le choix de m en fonction de n, elle est donnée par:

$$m = 1 + 3.322 \log_{10} n.$$

Le tableau suivant donne m (d'aprés la formule de sturges) pour les valeurs de n les plus utilisées dans la pratique.

Le tableau suivant donne m (d'aprés la formule de sturges) pour les valeurs de n les plus utilisées dans la pratique.

n	m
$10 \prec n \leq 22$	5
$22 \prec n \leq 44$	6
$44 \prec n \leq 90$	7
$90 \prec n \leq 180$	8
$180 \prec n \leq 360$	9
$360 \prec n \leq 720$	10
$720 \prec n \leq 1000$	11

b_0 sera choisi le plus proche possible (en étant inférieure) de la plus petite des données.

b_0 sera choisi le plus proche possible (en étant inférieure) de la plus petite des données.

- Il reste alors à définir la longeur des intervalles.

b_0 sera choisi le plus proche possible (en étant inférieure) de la plus petite des données.

- Il reste alors à définir la longeur des intervalles.Une bonne idée sur ce nombre peut être donnée par le calcul de $\frac{E}{m}$ où E est l'étendue, qui est égale à la différence entre la plus grande et la plus petite des observations.

Example

Un club d'athlétisme a noté la taille en centimètres de tous ses membres. Les données, rangées en ordre croissant, sont:

Example

Un club d'athlétisme a noté la taille en centimètres de tous ses membres. Les données, rangées en ordre croissant, sont:

142.4	162.1	172.1	178.3	181.2	188.5
148.7	163.4	172.3	178.5	181.4	189.1
151.5	165.1	172.7	179.0	182.1	190.1
153.6	165.2	173.4	179.2	183.5	191.4
156.1	166.3	175.2	179.6	184.2	192.5
158.2	167.1	175.3	179.8	186.1	193.3
159.9	170.3	176.1	180.2	187.2	196.2
160.5	171.1	177.2	180.4	188.3	197.1
161.3	171.2	177.4	180.9	188.4	205.2

Example

Ici $n = 54$ et la règle de Sturges donne $m = 7$.

Example

Ici $n = 54$ et la règle de Sturges donne $m = 7$.

L'étendue de la variable est $205.2 - 142.4 = 62.8$ et $\frac{62.8}{7} \backsimeq 9$.

Example

Ici $n = 54$ et la règle de Sturges donne $m = 7$.
L'étendue de la variable est $205.2 - 142.4 = 62.8$ et $\frac{62.8}{7} \backsimeq 9$. On peut prendre la longueur des classes égale à 10 et la borne inférieure de la première classes égale à 140 .

Example

Ici $n = 54$ et la règle de Sturges donne $m = 7$.
L'étendue de la variable est $205.2 - 142.4 = 62.8$ et $\frac{62.8}{7} \backsimeq 9$. On peut prendre la longueur des classes égale à 10 et la borne inférieure de la première classes égale à 140 . Ceci nous conduit à la distribution des fréquences suivante (la fréquence d'une classe étant le nombre d'individus pour lesquels la variable prend ses valeurs dans cette classe).

Classes	*Fréquences* n_i	*Fréquences relatives* f_i
$[140, 150[$	2	0.037
$[150, 160[$	5	0.093
$[160, 170[$	8	0.148
$[170, 180[$	18	0.333
$[180, 190[$	14	0.259
$[190, 200[$	6	0.111
$[200, 210[$	1	0.019
Totaux	54	1.000

Les définitions de ces notions restent les même pour les variables statistiques continues comme pour les variables discrètes.

Les définitions de ces notions restent les même pour les variables statistiques continues comme pour les variables discrètes.
Nous avons, de ce fait:

Les définitions de ces notions restent les même pour les variables statistiques continues comme pour les variables discrètes.
Nous avons, de ce fait:
L'effectif cumulé correspondant à la valeur X_i *est le nombre des individus ayant une valeur inférieure ou égal* X_i.

Les définitions de ces notions restent les même pour les variables statistiques continues comme pour les variables discrètes.
Nous avons, de ce fait:
L'effectif cumulé correspondant à la valeur X_i *est le nombre des individus ayant une valeur inférieure ou égal* X_i.
Autrement dit, c'est la somme des effectifs qui se sont accumulés en atteignant cette valeur.

Ce qui s'écrit:

Ce qui s'écrit:

$$N_i = \sum_{p=1}^{i} n_p$$

Ce qui s'écrit:

$$N_i = \sum_{p=1}^{i} n_p$$

De même,

Ce qui s'écrit:

$$N_i = \sum_{p=1}^{i} n_p$$

De même,
La fréquence cumulée correspondant à la valeur X_i *est la fréquence des individus ayant une valeur inférieure ou égal à* X_i.

Ce qui s'écrit:

$$N_i = \sum_{p=1}^{i} n_p$$

De même,
La fréquence cumulée correspondant à la valeur X_i *est la fréquence des individus ayant une valeur inférieure ou égal à* X_i.
C'est à dire:

Ce qui s'écrit:

$$N_i = \sum_{p=1}^{i} n_p$$

De même,
La fréquence cumulée correspondant à la valeur X_i *est la fréquence des individus ayant une valeur inférieure ou égal à* X_i.
C'est à dire:

$$F_i = \sum_{p=1}^{i} f_p$$

Example

Un club d'athlétisme , temps de réaction au son

Dans le cas d'une variable statistique continue le diagramme différentiel s'appelle **histogramme**.

Dans le cas d'une variable statistique continue le diagramme différentiel s'appelle **histogramme**.Achaque classe est associé un rectangle dont la largeur est l'amplitude de la classse et dont la hauteur est le rapport de la fréquence sur l'amplitude $(\frac{n_i}{a_i} ou \frac{f_i}{a_i})$.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l’étude de la variable statistique discrète, que le diagramme intégral n’est autre que la courbe représentative de la fonction cumulative.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l'étude de la variable statistique discrète, que le diagramme intégral n'est autre que la courbe représentative de la fonction cumulative. Nous avons constaté que c'est une courbe en escalier.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l'étude de la variable statistique discrète, que le diagramme intégral n'est autre que la courbe représentative de la fonction cumulative. Nous avons constaté que c'est une courbe en escalier. Ce genre de croissance par sauts est du au fait qu'entre deux valeurs de la variable il n'y a aucune accumulation. Toute accumulation supplémentaire se fait lorsqu'on dépasse une valeur de la variable.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l'étude de la variable statistique discrète, que le diagramme intégral n'est autre que la courbe représentative de la fonction cumulative. Nous avons constaté que c'est une courbe en escalier. Ce genre de croissance par sauts est du au fait qu'entre deux valeurs de la variable il n'y a aucune accumulation. Toute accumulation supplémentaire se fait lorsqu'on dépasse une valeur de la variable.

Dans le cas de la variable continue, cela ne se passe pas de la même manière.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l'étude de la variable statistique discrète, que le diagramme intégral n'est autre que la courbe représentative de la fonction cumulative. Nous avons constaté que c'est une courbe en escalier. Ce genre de croissance par sauts est du au fait qu'entre deux valeurs de la variable il n'y a aucune accumulation. Toute accumulation supplémentaire se fait lorsqu'on dépasse une valeur de la variable.

Dans le cas de la variable continue, cela ne se passe pas de la même manière. En effet, en allant d'une extrémité de la classe à l'autre on accumule les individus de cette classe. Il y a donc croissance de la fonction cumulative à l'intérieur même de la classe et non seulement aux extrémités.

Avant de définir le diagramme intégral nous avons à apporter quelques précisions importantes. Nous avons vu, lors de l'étude de la variable statistique discrète, que le diagramme intégral n'est autre que la courbe représentative de la fonction cumulative. Nous avons constaté que c'est une courbe en escalier. Ce genre de croissance par sauts est du au fait qu'entre deux valeurs de la variable il n'y a aucune accumulation. Toute accumulation supplémentaire se fait lorsqu'on dépasse une valeur de la variable.

Dans le cas de la variable continue, cela ne se passe pas de la même manière. En effet, en allant d'une extrémité de la classe à l'autre on accumule les individus de cette classe. Il y a donc croissance de la fonction cumulative à l'intérieur même de la classe et non seulement aux extrémités.Comment se fait cette croissance?

Nous avous supposé que la distribution est uniforme.

Nous avous supposé que la distribution est uniforme. Ce qui siginfie que la croissance est constante d'une extrémité de la classe à l'autre.

Nous avous supposé que la distribution est uniforme. Ce qui siginfie que la croissance est constante d'une extrémité de la classe à l'autre. Une croissance constante est synonyme de croissance linéaire; c'est à dire suivant une droite.

Nous avous supposé que la distribution est uniforme. Ce qui siginfie que la croissance est constante d'une extrémité de la classe à l'autre. Une croissance constante est synonyme de croissance linéaire; c'est à dire suivant une droite.
La courbe cumulative est une ligne brisée faite de segment de droite joignants les points dont les cordonnées sont : en abscisse, les extrèmités des classes et en ordonnées les fréquences cumilées correspondantes à ces extrèmités.

Nous avous supposé que la distribution est uniforme. Ce qui siginfie que la croissance est constante d'une extrémité de la classe à l'autre. Une croissance constante est synonyme de croissance linéaire; c'est à dire suivant une droite.
La courbe cumulative est une ligne brisée faite de segment de droite joignants les points dont les cordonnées sont : en abscisse, les extrèmités des classes et en ordonnées les fréquences cumilées correspondantes à ces extrèmités.

Example

Un club d'athlétisme , temps de réaction au son

Caractéristiques de position centrale

Nous avons déja donné des moyens de présenter des données quantitatives.

Caractéristiques de position centrale

Nous avons déja donné des moyens de présenter des données quantitatives. Mais cela ne suffit pas, nous avons aussi besoin d'avoir une idée sur l'ordre de grandeur des données et sur la position où elles semblent se rassembler.

Caractéristiques de position centrale

Nous avons déja donné des moyens de présenter des données quantitatives. Mais cela ne suffit pas, nous avons aussi besoin d'avoir une idée sur l'ordre de grandeur des données et sur la position où elles semblent se rassembler.Cela est donné par les paramètres de tendence centrale.

C'est un paramètre trés utilisé par tous.

C'est un paramètre trés utilisé par tous.Pensez à vos moyennes durant vos années d'études et qui ont servi à juger vos aptitudes!.

C'est un paramètre trés utilisé par tous.Pensez à vos moyennes durant vos années d'études et qui ont servi à juger vos aptitudes!. La moyenne est la valeur centrale autour de laquelle se répartissent les données.

C'est un paramètre trés utilisé par tous.Pensez à vos moyennes durant vos années d'études et qui ont servi à juger vos aptitudes!.
La moyenne est la valeur centrale autour de laquelle se répartissent les données.La moyenne d'une série $(x_i)_{1 \leq i \leq n}$,noté par $\overline{X}$, se calcule par:

C'est un paramètre trés utilisé par tous.Pensez à vos moyennes durant vos années d'études et qui ont servi à juger vos aptitudes!. La moyenne est la valeur centrale autour de laquelle se répartissent les données.La moyenne d'une série $(x_i)_{1\leq i\leq n}$,noté par $\overline{X}$, se calcule par:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

pour les données condensées par:

pour les données condensées par:

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} f_i x_i \text{ où } (x_1, x_2, ...x_k) \text{ sont les différentes modalités.}$$

pour les données condensées par:

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} f_i x_i \text{ où } (x_1, x_2, ...x_k) \text{ sont les différentes modalités.}$$

Et pour les données groupées par:

$$\frac{\sum_{i=1}^{n} n_i c_i}{N} = \sum_{i=1}^{n} f_i c_i \text{ où } (c_1, c_2, ...c_k) \text{ sont les différents milieux des classes.}$$

Calculer la moyenne de la variable statistique dont le tableau de distribtion de fréquences est donné ci dessous:

Caractéristiques de position centrale

Exemple

Calculer la moyenne de la variable statistique dont le tableau de distribtion de fréquences est donné ci dessous:

Modalités x_i	*Fréquences* n_i	$n_i x_i$
12	3	36
14	6	84
16	10	160
18	16	288
20	11	220
25	6	150
29	3	87
Totaux	55	1023

Calculer la moyenne de la variable statistique dont le tableau de distribtion de fréquences est donné ci dessous:

Modalités x_i	*Fréquences* n_i	$n_i x_i$
12	3	36
14	6	84
16	10	160
18	16	288
20	11	220
25	6	150
29	3	87
Totaux	55	1023

$$\overline{X} = \frac{1023}{55} = 18.64$$

C'est la valeur qui se répète le plus dans l'échantillon, C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0,

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

$$M_0 = b_{m_0} + \left(\frac{\triangle_1}{\triangle_1 + \triangle_2}\right) L_{m_0}$$

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

$$M_0 = b_{m_0} + \left(\frac{\triangle_1}{\triangle_1 + \triangle_2}\right) L_{m_0}$$

où b_{m_0} :est la borne inférieure de la classe modale;

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

$$M_0 = b_{m_0} + \left(\frac{\triangle_1}{\triangle_1 + \triangle_2}\right) L_{m_0}$$

où b_{m_0} :est la borne inférieure de la classe modale;
$\triangle_1$: est la différence entre la fréquence de la classe modale et la fréquence de la classe précédente;

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

$$M_0 = b_{m_0} + \left(\frac{\triangle_1}{\triangle_1 + \triangle_2}\right) L_{m_0}$$

où b_{m_0} :est la borne inférieure de la classe modale;
$\triangle_1$: est la différence entre la fréquence de la classe modale et la fréquence de la classe précédente;$\triangle_2$:est la différence entre la fréquence de la classe modale et la fréquence de la classe qui suit;

C'est la valeur qui se répète le plus dans l'échantillon,Autrement dit pour les données condensées, c'est la modalité qui a la plus grande fréquence.Quant au cas des données groupées en classes, on détermine d'abord la classe modale qui est la classe de plus grande fréquence puis le mode, noté M_0, Il se calcul sur l'histogramme par:

$$M_0 = b_{m_0} + \left(\frac{\triangle_1}{\triangle_1 + \triangle_2}\right) L_{m_0}$$

où b_{m_0} :est la borne inférieure de la classe modale;
$\triangle_1$: est la différence entre la fréquence de la classe modale et la fréquence de la classe précédente;$\triangle_2$:est la différence entre la fréquence de la classe modale et la fréquence de la classe qui suit;L_{m_0} :est la longeur de la classe modale.

Example

-Données condensées:

Modaliés	*Fréquences*
12	3
14	6
16	10
18	16
20	11
25	6
29	3

Example

-Données condensées:

Modaliés	*Fréquences*
12	3
14	6
16	10
18	16
20	11
25	6
29	3

$M_0 = 18$car c'est la modalité qui a la plus grande fréquence.

Example

- Données groupées:

Classes	*fréquences*
$[30, 40[$	4
$[40, 50[$	7
$[50, 60[$	11
$[60, 70[$	12
$[70, 80[$	8
$[80, 90[$	5

La classe modale est $[60, 70[$ car elle a la plus grande fréquence. $b_{m_0} = 60$, $\triangle_1 = 1, \triangle_2 = 4, L_{m_0} = 10$, d'où $M_0 = 60 + \frac{1}{5} * 10 = 62$

Remarque

Le mode peut ne pas être unique. Certaines distribution peuvent présenter plusieurs valeurs dont les effectifs sont égaux et qui sont les plus grands de la distribution.

Example

Remarque

Le mode peut ne pas être unique. Certaines distribution peuvent présenter plusieurs valeurs dont les effectifs sont égaux et qui sont les plus grands de la distribution.

Example

250 étudiants ont été soumis à un test faisant état du réflexe musculaire. Chaque étudiant se voit attribuer une note en fonction du temps mis pour réagir. Les résultats ont été classés comme suit:

Remarque

Le mode peut ne pas être unique. Certaines distribution peuvent présenter plusieurs valeurs dont les effectifs sont égaux et qui sont les plus grands de la distribution.

Example

250 étudiants ont été soumis à un test faisant état du réflexe musculaire. Chaque étudiant se voit attribuer une note en fonction du temps mis pour réagir. Les résultats ont été classés comme suit:

Nombre de points X_i	1	2	3	4	5	6	7	8
Nombre d'étudiants n_i	17	30	45	38	45	45	28	2

Remarque

Le mode peut ne pas être unique. Certaines distribution peuvent présenter plusieurs valeurs dont les effectifs sont égaux et qui sont les plus grands de la distribution.

Example

250 étudiants ont été soumis à un test faisant état du réflexe musculaire. Chaque étudiant se voit attribuer une note en fonction du temps mis pour réagir. Les résultats ont été classés comme suit:

Nombre de points X_i	1	2	3	4	5	6	7	8
Nombre d'étudiants n_i	17	30	45	38	45	45	28	2

Nous observons ici que nous avons trois modes; les valeurs 3; 5 et 6. Ils correspondent tous les trois au plus grand effectif.

C'est la valeur de la variable statistique qui partage la population en deux populations d'éffectifs égaux.

C'est la valeur de la variable statistique qui partage la population en deux populations d'éffectifs égaux.Autrement dit, les individus appartenant à la première moitié de la population ont des valeurs infèrieures à la médiane. les individus appartenant à la dexième moitié lui ont des valeurs supèrieures.

Pour déterminer la médiane nous avons deux moyens:

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\frac{N}{2} = \frac{150}{2} = 75$

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\frac{N}{2} = \frac{150}{2} = 75$
Nous le trouvons encadré par deux valeurs de l'effectif cumulé : N_{i-1} et N_i

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\quad \frac{N}{2} = \frac{150}{2} = 75$
Nous le trouvons encadré par deux valeurs de l'effectif cumulé : N_{i-1} et N_i
Dans notre exemple $N_{i-1} = 35$ et $N_i = 76$

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\quad \frac{N}{2} = \frac{150}{2} = 75$
Nous le trouvons encadré par deux valeurs de l'effectif cumulé : N_{i-1} et N_i
Dans notre exemple $N_{i-1} = 35$ et $N_i = 76$
Ces effectifs cumulés sont écrits sur les lignes.

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\frac{N}{2} = \frac{150}{2} = 75$
Nous le trouvons encadré par deux valeurs de l'effectif cumulé : N_{i-1} et N_i
Dans notre exemple $N_{i-1} = 35$ et $N_i = 76$
Ces effectifs cumulés sont écrits sur les lignes. La valeur de la variable inscrite dans la case comprise entre ces deux lignes est la médiane.

Pour déterminer la médiane nous avons deux moyens:
Utiliser la colonne des effectifs cumulés (fréquences absolues cumulées):
Il suffit de situer où se trouve $\frac{N}{2}$.
Reprenons l'exemple 1: où se trouve $\frac{N}{2}$? $\frac{N}{2} = \frac{150}{2} = 75$
Nous le trouvons encadré par deux valeurs de l'effectif cumulé : N_{i-1} et N_i
Dans notre exemple $N_{i-1} = 35$ et $N_i = 76$
Ces effectifs cumulés sont écrits sur les lignes. La valeur de la variable inscrite dans la case comprise entre ces deux lignes est la médiane.
Pour notre exemple $M = 16$.

Utiliser la colonne des fréquenes relatives cumulées :

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

$$F(M) = 0.5$$

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

$$F(M) = 0.5$$

Ainsi, nous devons chercher les valeurs F_{i-1} et F_i qui encadrent 0.5

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

$$F(M) = 0.5$$

Ainsi, nous devons chercher les valeurs F_{i-1} et F_i qui encadrent 0.5
Dans notre exemple $F_{i-1} = 0.233$ et $F_i = 0.506$

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

$$F(M) = 0.5$$

Ainsi, nous devons chercher les valeurs F_{i-1} et F_i qui encadrent 0.5
Dans notre exemple $F_{i-1} = 0.233$ et $F_i = 0.506$
Donc on trouve $M = 16$
Nous pouvons aussi nous servir du diagramme intégral pour situer la médiane.

Utiliser la colonne des fréquenes relatives cumulées :
Quand on se trouve réfère à la définition de la médiane M alors la fonction cumulative donne:

$$F(M) = 0.5$$

Ainsi, nous devons chercher les valeurs F_{i-1} et F_i qui encadrent 0.5
Dans notre exemple $F_{i-1} = 0.233$ et $F_i = 0.506$
Donc on trouve $M = 16$
Nous pouvons aussi nous servir du diagramme intégral pour situer la médiane. Il suffit de déterminer l'abscisse du point qui a 0.5 comme ordonnée.

Dans le cas des données groupées il faut d'abord situer la classe de la médiane.

Dans le cas des données groupées il faut d'abord situer la classe de la médiane.
nous savons que la médiane est la valeur de la variable statistique qui partage la population en deux effectifs égaux.

Dans le cas des données groupées il faut d'abord situer la classe de la médiane.
nous savons que la médiane est la valeur de la variable statistique qui partage la population en deux effectifs égaux.C'est la valeur M telle que $F(M) = 0.5$ où F est la fonction cumulative.

Dans le cas des données groupées il faut d'abord situer la classe de la médiane.
nous savons que la médiane est la valeur de la variable statistique qui partage la population en deux effectifs égaux.C'est la valeur M telle que $F(M) = 0.5$ où F est la fonction cumulative.Ainsi, à la premiére extrémité de la classe où se trouve la médiane nous n'avons pas encore atteint 0.5 comme fréquence cumulée et à la deuxiéme extrémité nous auront dépassé 0.5.

Dans le cas des données groupées il faut d'abord situer la classe de la médiane.
nous savons que la médiane est la valeur de la variable statistique qui partage la population en deux effectifs égaux.C'est la valeur M telle que $F(M) = 0.5$ où F est la fonction cumulative.Ainsi, à la premiére extrémité de la classe où se trouve la médiane nous n'avons pas encore atteint 0.5 comme fréquence cumulée et à la deuxiéme extrémité nous auront dépassé 0.5.La définition de la classe de la médiane est donc:
C'est la classe $[\mathbf{e}_{i-1}, e_i[$ telleque

$$F(\mathbf{e}_{i-1}) \prec 0.5 \; et \; F(\mathbf{e}_i) \succ 0.5$$

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + (\frac{0.5 - F_{m_e-1})}{f_{m_e}})L_{m_e}$$

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + \left(\frac{0.5 - F_{m_e - 1})}{f_{m_e}}\right) L_{m_e}$$

où

b_{m_0} : est la borne inférieure de la classe médiane;

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + (\frac{0.5 - F_{m_e - 1})}{f_{m_e}})L_{m_e}$$

où

b_{m_0} : est la borne inférieure de la classe médiane;
$F_{m_e - 1}$: est la fréquence relative cumulée de la classe qui précède la classe médiane.

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + (\frac{0.5 - F_{_{m_e}-1})}{f_{m_e}})L_{m_e}$$

où

b_{m_0} : est la borne inférieure de la classe médiane;

$F_{_{m_e}-1}$: est la fréquence relative cumulée de la classe qui précède la classe médiane.

f_{m_e} : est la fréquence relative de la classe médiane;

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + (\frac{0.5 - F_{m_e - 1})}{f_{m_e}})L_{m_e}$$

où
b_{m_0} : est la borne inférieure de la classe médiane;
$F_{m_e - 1}$: est la fréquence relative cumulée de la classe qui précède la classe médiane.
f_{m_e} : est la fréquence relative de la classe médiane;
L_{m_0} : est la longeur de la classe médiane.

Une fois que la classe de la médiane est déterminée, nous pouvons calculer la médiane par interpolaion linéaire alors:

$$M = b_{m_0} + \left(\frac{0.5 - F_{m_e - 1})}{f_{m_e}}\right) L_{m_e}$$

où

b_{m_0} : est la borne inférieure de la classe médiane;
$F_{m_e - 1}$: est la fréquence relative cumulée de la classe qui précède la classe médiane.
f_{m_e} : est la fréquence relative de la classe médiane;
L_{m_0} : est la longeur de la classe médiane.

Example

1) Données condensées utilisées en exemple pour le calcul du mode:

x_i	n_i	f_i	F_i
12	3	0.055	0.055
14	6	0.109	0.164
16	10	0.192	0.346
18	16	0.291	$0.637 \succ 0.5$
20	11	0.200	
25	6	0.109	
29	3	0.054	
Totaux	55	1.00	

$M = 18$ puisque c'est la première valeur dont la fréquence relative cumulée dépasse 0.5.

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$$

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$

Comme on a définit la médiane pour répartir la population en moitiés on

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$

Comme on a définit la médiane pour répartir la population en moitiés on

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$

Comme on a définit la médiane pour répartir la population en moitiés on

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$

Comme on a définit la médiane pour répartir la population en moitiés on

Example

2) Données groupées suivantes:

Classes	*fréquences*	*fréquences relatives*	*fréquences relatives cumulées*
$[30, 40[$	4	0.085	0.085
$[40, 50[$	7	0.149	0.234
$[50, 60[$	11	0.234	0.468
$[60, 70[$	12	0.255	$0.723 \succ 0.5$
$[70, 80[$	8	0.170	
$[80, 90[$	5	0.107	

La classe médiane est $[60, 70[$ car c'est dans cette classe que F_i dépasse, pour la première fois la valeur 0.5, la médiane est :

$M = 60 + \frac{0.5 - 0.468}{0.255} * 10 = 61.3.$

Comme on a définit la médiane pour répartir la population en moitiés on

Pour déterminer les quartiles on procède de la même manière que pour la médiane.
Il suffit, pour le premier quartile, de situer où se trouve $\frac{N}{4}$ sur la colonne des effectifs cumulés (ou 0.25 sur la colonne des fréquences relatives cumulées).

Pour déterminer les quartiles on procède de la même manière que pour la médiane.
Il suffit, pour le premier quartile, de situer où se trouve $\frac{N}{4}$ sur la colonne des effectifs cumulés (ou 0.25 sur la colonne des fréquences relatives cumulées).
Pour le troisième quartile, nous avons à situer $\frac{3N}{4}$ comme effectif cumulé (où 0.75 comme fréquence relative cumulée).

Dans le cas des données groupées la procédure est la même que pour la médiane.
il s'agit d'abord d'identifier les classes où se trouve les quartiles et de les calculer ensuite par interpolation linéaire.
Comme le premier quartile Q_1cumule le quart de la population alors il se trouve dans la classe $[\mathbf{e}_{i-1}, e_i[$ qui vérifie les conditions:

$$F(\mathbf{e}_{i-1}) \prec 0.25 \ et \ F(\mathbf{e}_i) \succ 0.25$$

Pour calculer Q_1 nous allons utiliser la formule de l'interpolaion linéaire

$$Q_1 = b_{m_0} + (\frac{0.25 - F_{m_e - 1})}{f_{m_e}})L_{m_e}$$

Le troisiéme quartile Q_3 cumule les trois quarts de la population et donc sa classe est celle qui vérifie

$$F(\mathbf{e}_{i-1}) \prec 0.75 \ et \ F(\mathbf{e}_i) \succ 0.75$$

La valeur de Q_3 et donc

$$Q_3 = b_{m_0} + (\frac{0.75 - F_{m_e -1})}{f_{m_e}})L_{m_e}$$

Caractéristiques de dispersion

Si les paramètres de tendances centrale donnent une idée sur les valeurs centrales autour desquelles se répartissent les données, il nous faut plus que cela puisque des sériés trés différentes peuvent avoir la même moyenne.

Caractéristiques de dispersion

Si les paramètres de tendances centrale donnent une idée sur les valeurs centrales autour desquelles se répartissent les données, il nous faut plus que cela puisque des sériés trés différentes peuvent avoir la même moyenne. Ainsi pour avoir une idée plus précise sur les données il est intéressant de connaitre la dispersion de ces données autour de la moyenne.

Caractéristiques de dispersion

Si les paramètres de tendances centrale donnent une idée sur les valeurs centrales autour desquelles se répartissent les données, il nous faut plus que cela puisque des sériés trés différentes peuvent avoir la même moyenne. Ainsi pour avoir une idée plus précise sur les données il est intéressant de connaitre la dispersion de ces données autour de la moyenne. Par exemple les deux sériés -1, 0, 1 et -100, 0, 100 ont toutes deux la même moyenne mais dans le premier exemple les données se reppochent beaucoup plus de la moyenne. Si par exemple, en plus de la connaissance de la moyenne on sait que les écarts des données à la moyenne sont petits on peut déduire que les données se concenant autour de la moyenne.

La variance de l'échantillon, notée s^2, est égale à :

La variance de l'échantillon, notée s^2, est égale à :

$$var(x) = s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{N}.$$

La variance de l'échantillon, notée s^2, est égale à :

$$var(x) = s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{N}.$$

- En données condensées:

$$s^2 = \sum_{i=1}^{n} \frac{n_i(x_i - \overline{x})^2}{N}.$$

La variance de l'échantillon, notée s^2, est égale à :

$$var(x) = s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{N}.$$

- En données condensées:

$$s^2 = \sum_{i=1}^{n} \frac{n_i (x_i - \overline{x})^2}{N}.$$

- En données groupées:

$$s^2 = \sum_{i=1}^{n} \frac{n_i (c_i - \overline{x})^2}{N}$$

où c_i est le milieu de la $i^{éme}$ classe et n_i sa fréquence.

Remarque

Pour le calcul, il est plus souvent pratique d'utiliser l'expression équivalente de la variance et qui est donnée par:

$$s^2 = \sum_{i=1}^{n} \frac{n_i x_i^2}{N} - \overline{x}^2$$

La variance est un paramètre de dispersion trés utilisé mais son unité est le carré des unités des meusures, c'est pourquoi sa racine carré est introduite et s'appelle l'écart type, on note s.

C'est la différence entre la plus grande donnée et la plus petite est aussi un paramétre de dispersion .Seulement elle donne une idée, certes rapide, mais grossiére de la dispersion des données puis qu'elle ne tient compte que des données extrémes.
pour l'exemple pécédent: E=29-12=17.

Il est donné par la différence entre les quartiles d'ordre 3 et d'ordre 1. Cet intervelle contient la moitié des observation qui se situent au centre de la série.
pour l'exemple precident
$E = Q_3 - Q_1 = 20 - 16$

Ajustement Linéaire-Regression

Ajustement Linéaire-Regression

Les coefficients de la droite de regression se calculent toujours par:

$$\begin{aligned} a &= \frac{Cov(X,Y)}{Var(X)} \\ b &= \overline{Y} - a\overline{X} \end{aligned}$$

Ajustement à l'aide l'une fonction exponentielle

Example

une société a mis en vente un produit qui a connu un grand succés. les ventes enregistrées les 7 premies mois été rapportées dans le tableau suivant:

Example

une société a mis en vente un produit qui a connu un grand succés. les ventes enregistrées les 7 premies mois été rapportées dans le tableau suivant:

Mois	1	2	3	4	5	6	7
Nombre d'article vendus	10	35	88	280	802	2350	7300

Ajustement à l'aide l'une fonction exponentielle

Example

Ajustement à l'aide l'une fonction exponentielle

Example

Nous constatons sur cet exemple que la progression de la variable Y est trés rapide au point qu'il semblent évident qu'elle ne peut etre ajustée par une droite.

Ajustement à l'aide l'une fonction exponentielle

Example

Nous constatons sur cet exemple que la progression de la variable Y est trés rapide au point qu'il semblent évident qu'elle ne peut etre ajustée par une droite. Les nombres triplent presque à chaque mois. L'ajustement le plus approprié sera à l'aide d'une fonction exponentielle de la forme

Example

Nous constatons sur cet exemple que la progression de la variable Y est trés rapide au point qu'il semblent évident qu'elle ne peut etre ajustée par une droite. Les nombres triplent presque à chaque mois. L'ajustement le plus approprié sera à l'aide d'une fonction exponentielle de la forme

$$Y = BA^{X}$$

Ajustement à l'aide l'une fonction exponentielle

Example

Ajustement à l'aide l'une fonction exponentielle

Example

Nous allons voir que ce cas se ramène au précédent par une transformation logarithmique

Ajustement à l'aide l'une fonction exponentielle

Example

Nous allons voir que ce cas se ramène au précédent par une transformation logarithmique

$$\ln Y = \ln(BA^X)$$

Ajustement à l'aide l'une fonction exponentielle

Example

Nous allons voir que ce cas se ramène au précédent par une transformation logarithmique

$$\ln Y = \ln(BA^X)$$

Ce qui revient à

Ajustement à l'aide l'une fonction exponentielle

Example

Nous allons voir que ce cas se ramène au précédent par une transformation logarithmique

$$\ln Y = \ln(BA^X)$$

Ce qui revient à

$$\ln Y = X \ln A + \ln B$$

Ajustement à l'aide l'une fonction exponentielle

Ajustement à l'aide l'une fonction exponentielle

Reamaquez que ceci est l'équation d'une droite et nous savons faire un ajustement à l'aide d'une droite.

Ajustement à l'aide l'une fonction exponentielle

Reamaquez que ceci est l'équation d'une droite et nous savons faire un ajustement à l'aide d'une droite. Notre variable explicative est x et notre variable expliquée est $\ln Y$.

Ajustement à l'aide l'une fonction exponentielle

Reamaquez que ceci est l'équation d'une droite et nous savons faire un ajustement à l'aide d'une droite. Notre variable explicative est x et notre variable expliquée est $\ln Y$.

$$y = ax + b$$

Ajustement à l'aide l'une fonction exponentielle

Reamaquez que ceci est l'équation d'une droite et nous savons faire un ajustement à l'aide d'une droite. Notre variable explicative est x et notre variable expliquée est $\ln Y$.

$$y = ax + b$$

ou

$$\begin{aligned} a &= \ln A \\ b &= \ln B \end{aligned}$$

Ajustement à l'aide l'une fonction exponentielle

Lorsque ces coefficients seront déterminés. nous revenons à nos coefficients de départ par:

Ajustement à l'aide l'une fonction exponentielle

Lorsque ces coefficients seront déterminés. nous revenons à nos coefficients de départ par:

$$\begin{aligned} A &= e^{a} \\ B &= e^{b} \end{aligned}$$

Ajustement à l'aide l'une fonction puissance

Ajustement à l'aide l'une fonction puissance

Ajustement à l'aide l'une fonction puissance

Lorsque nous devons ajuster deux variables X et Y par une fontion puissance de la forme:

Ajustement à l'aide l'une fonction puissance

Lorsque nous devons ajuster deux variables X et Y par une fontion puissance de la forme:

$$Y = BX^a$$

Ajustement à l'aide l'une fonction puissance

Lorsque nous devons ajuster deux variables X et Y par une fontion puissance de la forme:

$$Y = BX^a$$

nous pouvons revenir au cas de l'ajustement à l'aide d'une droite par la tronsformation suivante:

Ajustement à l'aide l'une fonction puissance

Lorsque nous devons ajuster deux variables X et Y par une fontion puissance de la forme:

$$Y = BX^a$$

nous pouvons revenir au cas de l'ajustement à l'aide d'une droite par la tronsformation suivante:

$$\ln Y = \ln(BX^a)$$

Ajustement à l'aide l'une fonction puissance

Ajustement à l'aide l'une fonction puissance

C'est-à-dire

$$\ln Y = a \ln X + \ln B$$

C'est-à-dire

$$\ln Y = a \ln X + \ln B$$

Nous avons, de cette maniére, l'équation d'une droite

Ajustement à l'aide l'une fonction puissance

C'est-à-dire

$$\ln Y = a \ln X + \ln B$$

Nous avons, de cette maniére, l'équation d'une droite

$$y = ax + b$$

ou

$$\begin{aligned} x &= \ln X \\ y &= \ln Y \end{aligned}$$

Ajustement à l'aide l'une fonction puissance

Lorsque les coefficients a et b seront obtenus, le retour aux coefficients initiaux se fait ainsi par:

Ajustement à l'aide l'une fonction puissance

Lorsque les coefficients a et b seront obtenus, le retour aux coefficients initiaux se fait ainsi par:

$$\begin{aligned} a &= a \\ B &= e^{b} \end{aligned}$$

Merci pour votre attention.