

Probabilités et statistiques-séance 2

Licence 1-semester 1

Allegret Audrey

Maître de Conférences - Université de Toulon, LEAD

15 septembre 2021

Introduction

La statistique descriptive a deux approches pour décrire un jeu de données observées :

Introduction

La statistique descriptive a deux approches pour décrire un jeu de données observées :

- une **approche graphique** qui a pour objectif de fournir des représentations graphiques permettant de visualiser la distribution des données.

Introduction

La statistique descriptive a deux approches pour décrire un jeu de données observées :

- une **approche graphique** qui a pour objectif de fournir des représentations graphiques permettant de visualiser la distribution des données.
- une **approche quantitative** qui a pour but de calculer des indices numériques caractérisant la répartition des données, les tendances, la dispersion, la concentration, etc.

Introduction

La statistique descriptive a deux approches pour décrire un jeu de données observées :

- une **approche graphique** qui a pour objectif de fournir des représentations graphiques permettant de visualiser la distribution des données.
- une **approche quantitative** qui a pour but de calculer des indices numériques caractérisant la répartition des données, les tendances, la dispersion, la concentration, etc.

Selon le type de variable statistique étudié, on a recours à des graphiques différents.

Diagrammes à secteurs circulaires

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Diagrammes à secteurs circulaires

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Diagrammes à secteurs circulaires

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Ce sont des disques découpés en secteurs dont l'angle est proportionnel aux proportions (ou fréquences) de chaque modalité.

Diagrammes à secteurs circulaires

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Ce sont des disques découpés en secteurs dont l'angle est proportionnel aux proportions (ou fréquences) de chaque modalité.

Le secteur total étant de 360° , si f_i est la fréquence de la i -ième modalité, on la représente par un secteur d'angle α_i défini comme ceci :

Diagrammes à secteurs circulaires

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Ce sont des disques découpés en secteurs dont l'angle est proportionnel aux proportions (ou fréquences) de chaque modalité.

Le secteur total étant de 360° , si f_i est la fréquence de la i -ième modalité, on la représente par un secteur d'angle α_i défini comme ceci :

$$\alpha_i = f_i * 360 = \frac{n_i}{N} * 360$$

Exemple

On utilise les données suivantes : **Taux de réussite au baccalauréat en 2013 dans l'académie de Lille**. Si on isole les trois grands types de baccalauréats, on obtient les résultats suivants (onglet Feuil1) :

Exemple

On utilise les données suivantes : **Taux de réussite au baccalauréat en 2013 dans l'académie de Lille**. Si on isole les trois grands types de baccalauréats, on obtient les résultats suivants (onglet Feuil1) :

Type	Total	Proportions	Angles
Bac général	19772	46.79%	168°
Bac techno	9043	21.40%	77°
Bac pro	13439	31.81%	115°

Exemple

On utilise les données suivantes : **Taux de réussite au baccalauréat en 2013 dans l'académie de Lille**. Si on isole les trois grands types de baccalauréats, on obtient les résultats suivants (onglet Feuil1) :

Type	Total	Proportions	Angles
Bac général	19772	46.79%	168°
Bac techno	9043	21.40%	77°
Bac pro	13439	31.81%	115°

A noter ici que $N=42254$.

Exemple

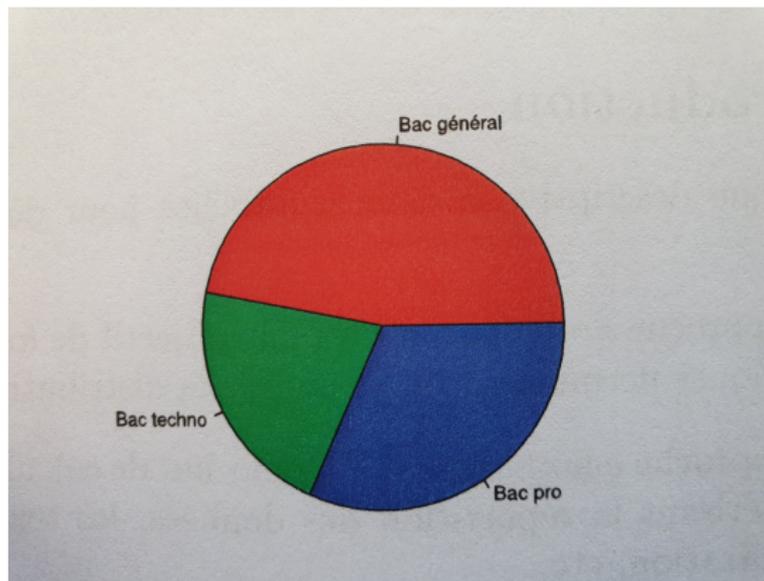
On utilise les données suivantes : **Taux de réussite au baccalauréat en 2013 dans l'académie de Lille**. Si on isole les trois grands types de baccalauréats, on obtient les résultats suivants (onglet Feuil1) :

Type	Total	Proportions	Angles
Bac général	19772	46.79%	168°
Bac techno	9043	21.40%	77°
Bac pro	13439	31.81%	115°

A noter ici que $N=42254$.

La représentation graphique est donnée ci-dessous.

Diagrammes à secteurs circulaires



Exemple sous excel

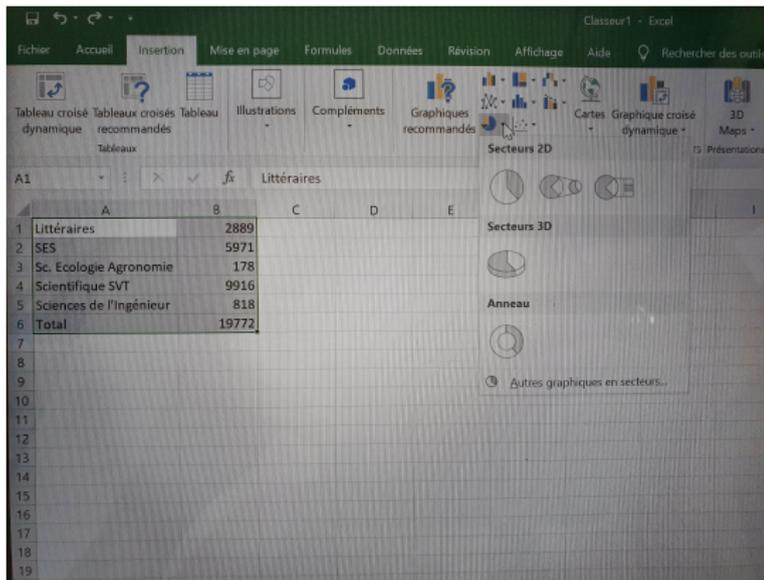
Réalisons un diagramme à secteurs circulaires pour les sous-catégories du baccalauréat général avec les données du tableau ci-dessous (onglet Feuil4) :

Exemple sous excel

Réalisons un diagramme à secteurs circulaires pour les sous-catégories du baccalauréat général avec les données du tableau ci-dessous (onglet Feuil4) :

Type	Effectifs	Proportions	Angles
Littéraires	2889	14.61%	54°
SES	5971	30.20%	108°
Sc. écologie et agronomie	178	0.90%	4°
Scft. SVT	9916	50.15%	180°
Sciences de l'ingénieur	818	4.14%	14°
Ensemble	19772	100%	360°

Diagrammes à secteurs circulaires

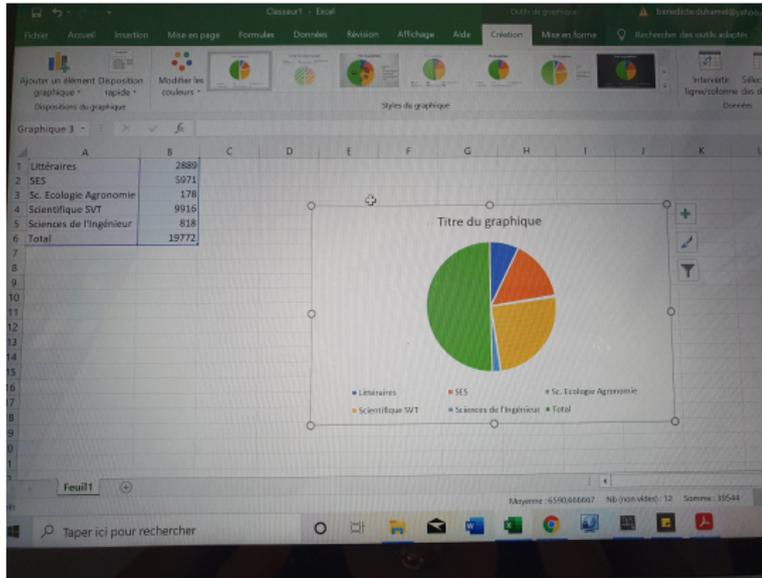


The screenshot shows the Microsoft Excel interface. The 'Insertion' ribbon is active, displaying various chart options. The data table is as follows:

	A	B	C	D	E
1	Littéraires	2889			
2	SES	5971			
3	Sc. Ecologie Agronomie	178			
4	Scientifique SVT	9916			
5	Sciences de l'Ingénieur	818			
6	Total	19772			
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

The 'Insertion' ribbon shows the 'Graphiques recommandés' group, which includes options for 'Secteurs 2D', 'Secteurs 3D', and 'Anneau'. The 'Secteurs 2D' group is currently selected, showing three different 2D pie chart styles. Below the chart options, there is a link for 'Autres graphiques en secteurs...'. The 'Tableaux' group is also visible, showing 'Tableau croisé dynamique', 'Tableaux croisés recommandés', and 'Tableau dynamique'.

Diagrammes à secteurs circulaires



Diagrammes en bâtons

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Diagrammes en bâtons

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Diagrammes en bâtons

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Les modalités sont représentées en abscisse et les effectifs correspondants sont représentés par des lignes ou des bandes verticales dont la hauteur est proportionnelle à la valeur.

Diagrammes en bâtons

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Les modalités sont représentées en abscisse et les effectifs correspondants sont représentés par des lignes ou des bandes verticales dont la hauteur est proportionnelle à la valeur.

C'est donc la hauteur des lignes ou des bandes qui permet d'apprécier les tailles relatives des différentes modalités.

Diagrammes en bâtons

Ils conviennent pour représenter des **variables qualitatives** ou des **variables quantitatives discrètes**.

Il est préférable qu'il y ait un nombre restreint de modalités pour que le graphique reste lisible.

Les modalités sont représentées en abscisse et les effectifs correspondants sont représentés par des lignes ou des bandes verticales dont la hauteur est proportionnelle à la valeur.

C'est donc la hauteur des lignes ou des bandes qui permet d'apprécier les tailles relatives des différentes modalités.

Diagrammes en bâtons

Dans le cas d'une variable qualitative, la position des modalités en abscisse n'a pas de signification particulière.

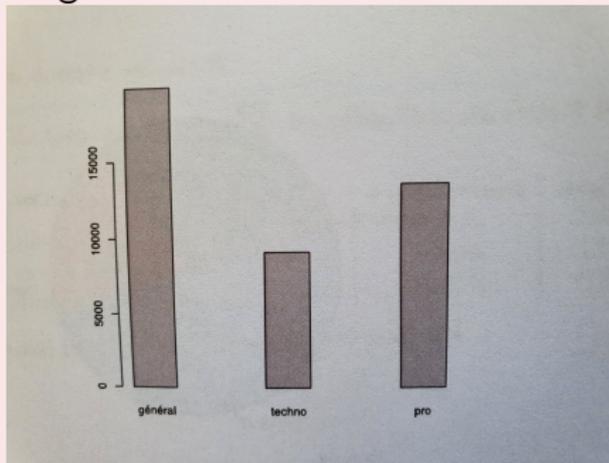
Diagrammes en bâtons

Dans le cas d'une variable qualitative, la position des modalités en abscisse n'a pas de signification particulière.

Si la variable est ordinale, on placera les modalités dans leur ordre naturel.

Diagrammes en bâtons

Reprenons l'exemple des catégories du baccalauréat. On obtient le diagramme suivant :



Diagrammes en bâtons

A noter que l'on peut aussi bien faire les diagrammes en proportions plutôt qu'en effectifs.

Diagrammes en bâtons

Fréquences et effectifs étant proportionnels (dans le rapport N), l'aspect visuel est rigoureusement identique. Seules changent les valeurs sur l'axe vertical.

Diagrammes en bâtons

Fréquences et effectifs étant proportionnels (dans le rapport N), l'aspect visuel est rigoureusement identique. Seules changent les valeurs sur l'axe vertical.

Un avantage des diagrammes en bâtons par rapport aux diagrammes circulaires est qu'ils permettent de représenter plusieurs distributions en parallèle.

Diagrammes en bâtons

Fréquences et effectifs étant proportionnels (dans le rapport N), l'aspect visuel est rigoureusement identique. Seules changent les valeurs sur l'axe vertical.

Un avantage des diagrammes en bâtons par rapport aux diagrammes circulaires est qu'ils permettent de représenter plusieurs distributions en parallèle.

Pour une même modalité, on peut placer côte à côte plusieurs lignes ou bandes verticales, correspondant à des sous-ensembles différents.

Diagrammes en bâtons

Fréquences et effectifs étant proportionnels (dans le rapport N), l'aspect visuel est rigoureusement identique. Seules changent les valeurs sur l'axe vertical.

Un avantage des diagrammes en bâtons par rapport aux diagrammes circulaires est qu'ils permettent de représenter plusieurs distributions en parallèle.

Pour une même modalité, on peut placer côte à côte plusieurs lignes ou bandes verticales, correspondant à des sous-ensembles différents.

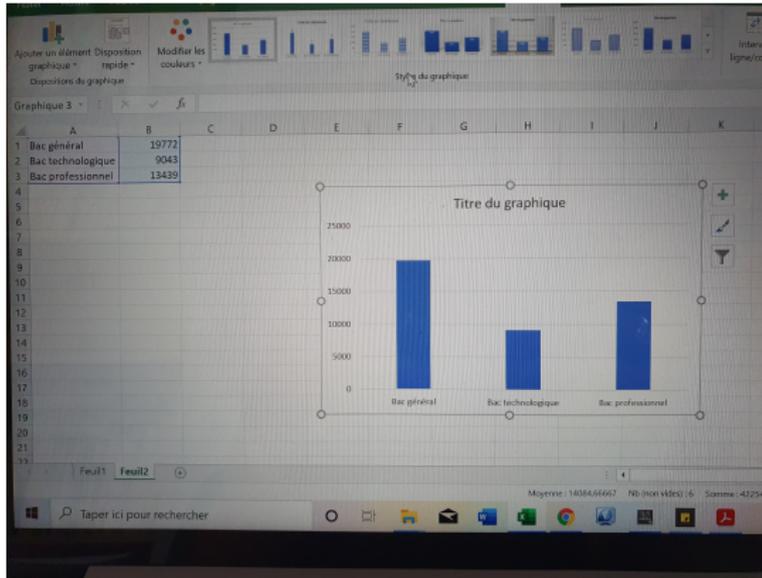
Un autre mode de représentation consiste à empiler les valeurs verticalement en faisant plusieurs segments.

Diagrammes en bâtons sur excel

The screenshot shows the Microsoft Excel interface with the 'Insertion' tab selected. The 'Graphiques recommandés' group is active, displaying a dropdown menu for 'Histogrammes'. The menu options are: Histogramme 2D, Histogramme 3D, Barres 2D, Barres 3D, and Autres histogrammes... The spreadsheet data is as follows:

	A	B	C	D	E
1	Bac général	19772			
2	Bac technologique	9043			
3	Bac professionnel	13439			
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					

Diagrammes en bâtons sur excel



Exemple sous excel

Le tableau suivant donne les proportions de réussite au baccalauréat dans l'académie de La Réunion pour les filles et les garçons de 2005 à 2011. Les valeurs sont exprimées en pourcentage.

Exemple sous excel

Le tableau suivant donne les proportions de réussite au baccalauréat dans l'académie de La Réunion pour les filles et les garçons de 2005 à 2011. Les valeurs sont exprimées en pourcentage.

Année	2005	2006	2007	2008	2009	2010
Filles	60.1	59.7	63.8	63.3	65.5	65.9
Garçons	42.8	44.2	43.5	47.1	48.4	49.4

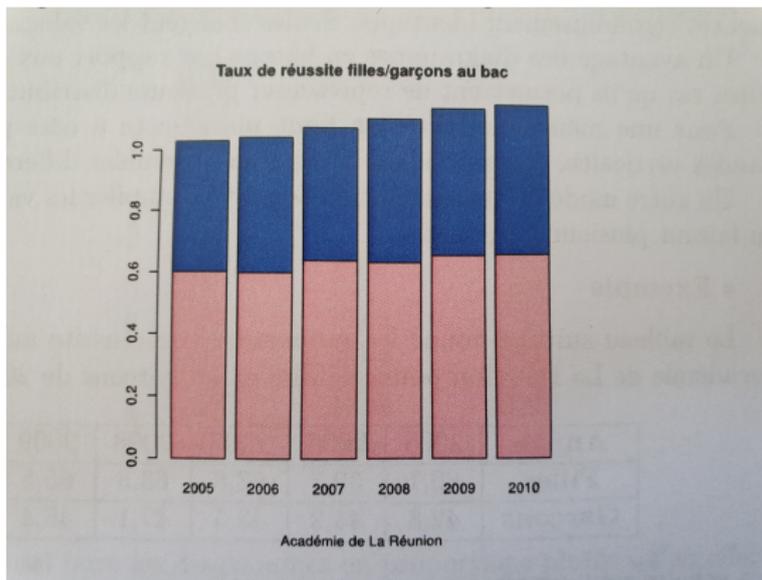
Exemple sous excel

Le tableau suivant donne les proportions de réussite au baccalauréat dans l'académie de La Réunion pour les filles et les garçons de 2005 à 2011. Les valeurs sont exprimées en pourcentage.

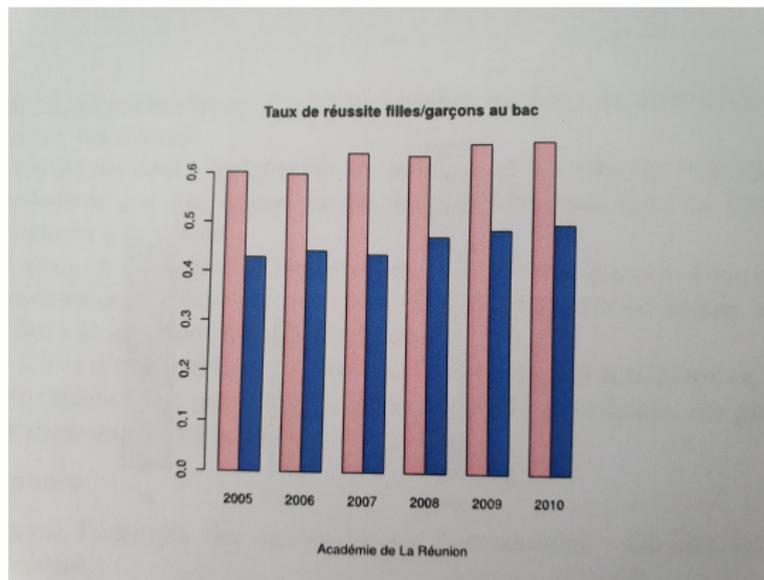
Année	2005	2006	2007	2008	2009	2010
Filles	60.1	59.7	63.8	63.3	65.5	65.9
Garçons	42.8	44.2	43.5	47.1	48.4	49.4

Dans les diagrammes ci-dessous les valeurs sont placées côte à côte et les valeurs sont empilées verticalement.

Diagrammes en bâtons sur excel



Diagrammes en bâtons sur excel



Diagrammes d'effectifs cumulés

Les diagrammes d'effectifs cumulés représentent la répartition de la distribution des effectifs.

Histogrammes

Les histogrammes sont des graphiques qui permettent de visualiser les proportions au moyen de rectangles verticaux.

Histogrammes

Les histogrammes sont des graphiques qui permettent de visualiser les proportions au moyen de rectangles verticaux.

Ils concernent les **variables quantitatives discrètes** ou les **variables quantitatives continues** qu'on regroupe en classes contiguës.

Histogrammes

Les histogrammes sont des graphiques qui permettent de visualiser les proportions au moyen de rectangles verticaux.

Ils concernent les **variables quantitatives discrètes** ou les **variables quantitatives continues** qu'on regroupe en classes contiguës.

Un histogramme peut être dessiné en effectifs ou en fréquences : comme ce sont des grandeurs proportionnelles, cela ne change pas l'allure du graphique mais seulement les valeurs portées sur l'axe vertical.

Histogrammes

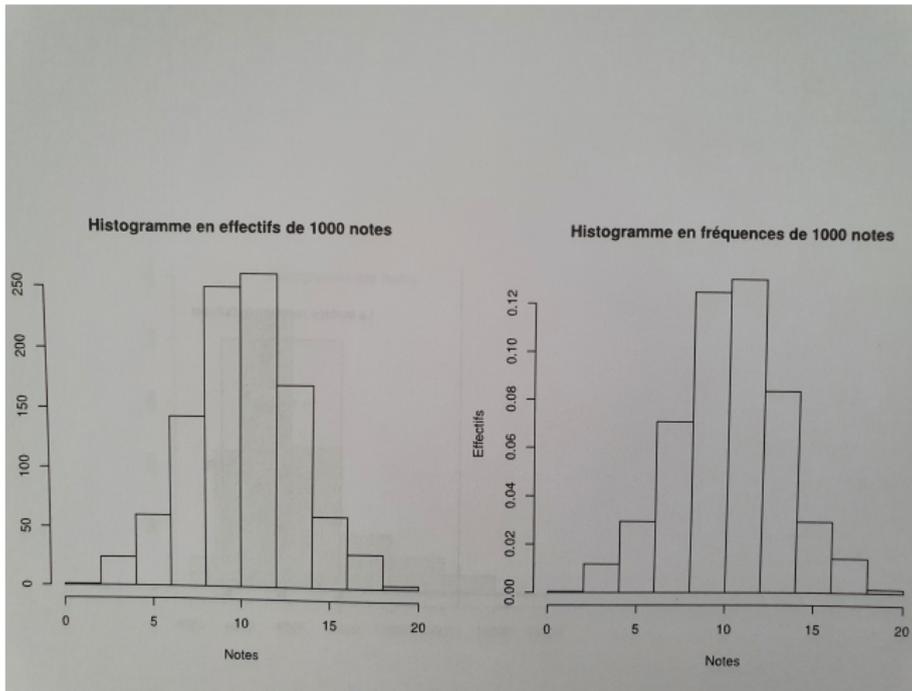
Les histogrammes sont des graphiques qui permettent de visualiser les proportions au moyen de rectangles verticaux.

Ils concernent les **variables quantitatives discrètes** ou les **variables quantitatives continues** qu'on regroupe en classes contiguës.

Un histogramme peut être dessiné en effectifs ou en fréquences : comme ce sont des grandeurs proportionnelles, cela ne change pas l'allure du graphique mais seulement les valeurs portées sur l'axe vertical.

Voici un histogramme correspondant à des notes obtenues à un examen par 1000 étudiants.

Histogrammes



Histogrammes

Le principe de construction d'un histogramme consiste à découper les données en classes et à dessiner des rectangles dont la surface est proportionnelle aux effectifs (ou aux fréquences).

Histogrammes

Le principe de construction d'un histogramme consiste à découper les données en classes et à dessiner des rectangles dont la surface est proportionnelle aux effectifs (ou aux fréquences).

La base des rectangles correspond à chaque intervalle $[e_i; e_{i+1}[$. La largeur de ces intervalles est l'amplitude $a_i = e_{i+1} - e_i$

Histogrammes

Le principe de construction d'un histogramme consiste à découper les données en classes et à dessiner des rectangles dont la surface est proportionnelle aux effectifs (ou aux fréquences).

La base des rectangles correspond à chaque intervalle $[e_i; e_{i+1}[$. La largeur de ces intervalles est l'amplitude $a_i = e_{i+1} - e_i$

Si on désigne la hauteur par h_i , la surface du rectangle est alors :

Histogrammes

Le principe de construction d'un histogramme consiste à découper les données en classes et à dessiner des rectangles dont la surface est proportionnelle aux effectifs (ou aux fréquences).

La base des rectangles correspond à chaque intervalle $[e_i; e_{i+1}[$. La largeur de ces intervalles est l'amplitude $a_i = e_{i+1} - e_i$

Si on désigne la hauteur par h_i , la surface du rectangle est alors :

$$S_i = a_i * h_i$$

Histogrammes

Le principe de construction d'un histogramme consiste à découper les données en classes et à dessiner des rectangles dont la surface est proportionnelle aux effectifs (ou aux fréquences).

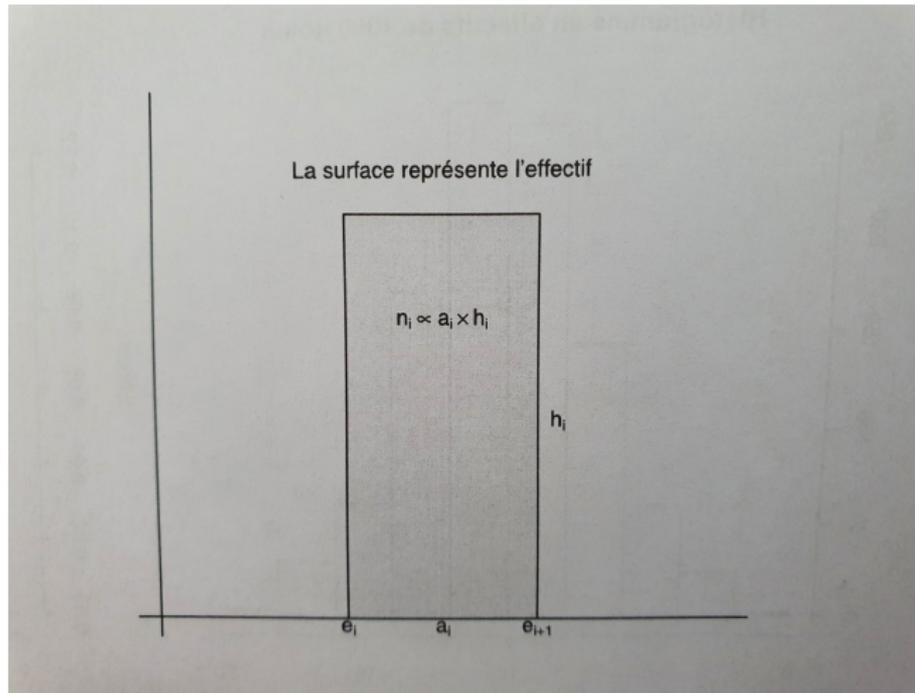
La base des rectangles correspond à chaque intervalle $[e_i; e_{i+1}[$. La largeur de ces intervalles est l'amplitude $a_i = e_{i+1} - e_i$

Si on désigne la hauteur par h_i , la surface du rectangle est alors :

$$S_i = a_i * h_i$$

Cette valeur doit correspondre à l'effectif n_i (pour un histogramme en effectifs) ou à la fréquence f_i (pour un histogramme en fréquences).

Surface histogramme



Histogrammes

Dans le cas d'un histogramme en fréquences (ou proportions), la surface S_i s'interprète comme la fréquence f_i c'est-à-dire la proportion des observations qui se trouvent dans l'intervalle $[e_i; e_{i+1}[$.

Histogrammes

Dans le cas d'un histogramme en fréquences (ou proportions), la surface S_i s'interprète comme la fréquence f_i c'est-à-dire la proportion des observations qui se trouvent dans l'intervalle $[e_i; e_{i+1}[$.

L'intérêt de cette représentation est qu'on peut représenter la proportion d'observations qui sont dans plusieurs intervalles contigus en additionnant les surfaces des rectangles correspondants.

Histogrammes

Dans le cas d'un histogramme en fréquences (ou proportions), la surface S_i s'interprète comme la fréquence f_i c'est-à-dire la proportion des observations qui se trouvent dans l'intervalle $[e_i; e_{i+1}[$.

Histogrammes

Dans le cas d'un histogramme en fréquences (ou proportions), la surface S_i s'interprète comme la fréquence f_i c'est-à-dire la proportion des observations qui se trouvent dans l'intervalle $[e_i; e_{i+1}[$.

L'intérêt de cette représentation est qu'on peut représenter la proportion d'observations qui sont dans plusieurs intervalles contigus en additionnant les surfaces des rectangles correspondants.

Exemple

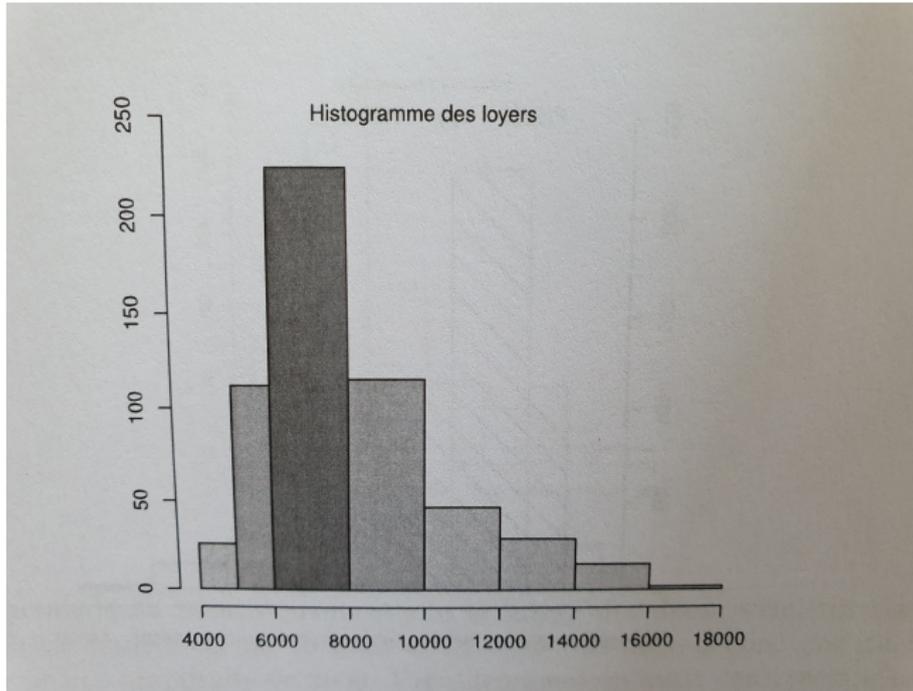
On a relevé le loyer annuel (en milliers) de 500 domiciles d'une agglomération et obtenu le tableau d'effectifs suivant :

Exemple

On a relevé le loyer annuel (en milliers) de 500 domiciles d'une agglomération et obtenu le tableau d'effectifs suivant :

Classes	Effectifs
[4 ;5[13
[5 ;6[56
[6 ;8[224
[8 ;10[115
[10 ;12[46
[12 ;14[29
[14 ;16[15
[16 ;18[2

Surface histogramme



Exemple

On remarque que les deux premières classes ont une amplitude de 1 (c'est-à-dire 1000 euros) tandis que les suivantes ont une amplitude de 2 (c'est-à-dire 2000 euros).

Exemple

On remarque que les deux premières classes ont une amplitude de 1 (c'est-à-dire 1000 euros) tandis que les suivantes ont une amplitude de 2 (c'est-à-dire 2000 euros).

Cela a pour conséquence que les deux premiers rectangles sont deux fois plus hauts et en particulier que le deuxième et le quatrième ont approximativement la même hauteur.

Exemple

On remarque que les deux premières classes ont une amplitude de 1 (c'est-à-dire 1000 euros) tandis que les suivantes ont une amplitude de 2 (c'est-à-dire 2000 euros).

Cela a pour conséquence que les deux premiers rectangles sont deux fois plus hauts et en particulier que le deuxième et le quatrième ont approximativement la même hauteur.

En effet le deuxième correspond à la valeur $n_2 = 56$ qui a été multipliée par 2, à savoir 112, tandis que le quatrième correspond à la valeur $n_4 = 115$.

Exemple

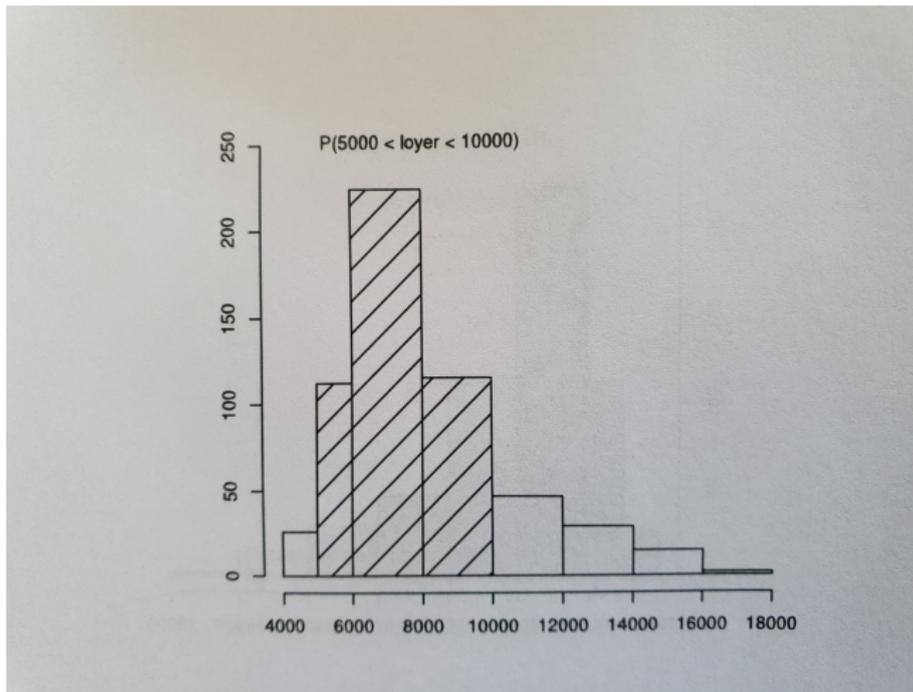
L'intérêt de cette représentation est qu'on peut représenter la proportion d'observations qui sont dans plusieurs intervalles contigus en additionnant les surfaces des rectangles correspondants.

Exemple

L'intérêt de cette représentation est qu'on peut représenter la proportion d'observations qui sont dans plusieurs intervalles contigus en additionnant les surfaces des rectangles correspondants.

Par exemple, dans le diagramme ci-dessous, la zone hachurée correspond à la proportion $P(5000 \leq X \leq 10000)$ où X est le loyer.

histogramme



Interprétation des hauteurs

La grande différence entre les diagrammes en bâtons et les histogrammes est que dans les premiers ni est représenté en hauteur tandis que, dans les seconds, il est représenté en surface.

Interprétation des hauteurs

La grande différence entre les diagrammes en bâtons et les histogrammes est que dans les premiers ni est représenté en hauteur tandis que, dans les seconds, il est représenté en surface. Quelle est alors la signification de la hauteur dans un histogramme ?

Interprétation des hauteurs

La grande différence entre les diagrammes en bâtons et les histogrammes est que dans les premiers n_i est représenté en hauteur tandis que, dans les seconds, il est représenté en surface. Quelle est alors la signification de la hauteur dans un histogramme ?

On a :

$$n_i = a_i * h_i$$

$$\Rightarrow h_i = \frac{n_i}{a_i}$$

Interprétation des hauteurs

La grande différence entre les diagrammes en bâtons et les histogrammes est que dans les premiers n_i est représenté en hauteur tandis que, dans les seconds, il est représenté en surface. Quelle est alors la signification de la hauteur dans un histogramme ?

On a :

$$n_i = a_i * h_i$$

$$\Rightarrow h_i = \frac{n_i}{a_i}$$

Le rapport $d_i = \frac{n_i}{a_i}$ est la densité de classe c_i .

Interprétation des hauteurs

Donc lorsqu'un rectangle est plus haut qu'un autre, c'est que la densité de son intervalle est plus grande, autrement dit qu'il comporte plus de données à amplitude égale.

Interprétation des hauteurs

Les histogrammes appréciés en hauteur donnent un aperçu de la densité de répartition des données.