

# Une librairie pour la statistique descriptive avec Asymptote

Dariussh GHORNBANZADEH

<http://www.cnam.fr/math/Membres/ghorbanzadeh/>

4 mai 2010 (version 1.1)

## 1 Introduction

Dans ce document nous allons présenter une librairie pour le traitement des données (🌐 [statistique descriptive](#)) réalisée avec 🌐 [Asymptote](#). Nous distinguons trois types de données : Données Qualitatives, Données Discrètes et Données Continues. Afin d'inclure les résultats dans un document (via  $\LaTeX$ ), nous proposons une *IHM* s'appuyant sur le module `xkeyval`.

Pour utiliser la librairie, le document  $\LaTeX$  doit être compilé avec la commande :

```
latex -enable-write18
```

Pour les Données Qualitatives la librairie contient : `histogramme des données` et `représentation graphique circulaire des données (camembert)`.

Pour les Données Discrètes la librairie contient : `analyse des données`, `histogramme des données`, `représentation graphique de la fonction de répartition des données`, et `représentation graphique circulaire des données (camembert)`.

Pour les Données Continues la librairie contient : `analyse des données`, `histogramme des données` et `représentation graphique de la fonction de répartition empirique des données`.

## 2 Architecture de la librairie

Les données à analyser peuvent être stockées dans un fichier externe, sinon, à l'aide du package `filecontents` l'utilisateur peut inclure les données dans le corps du document. Par exemple (dans le préambule) :

```

\usepackage{filecontents} % dans le préambule
\begin{filecontents*}{MoisSyracuse.dat}
janvier 2010
f\`evrier 2010
mars 2010
avril 2010
\end{filecontents*}

```

```

\usepackage{filecontents} % dans le préambule
\begin{filecontents*}{VisitSyracuse.dat}
25421
45847
53895
43719
\end{filecontents*}

```

### 3 Données Qualitatives

L'IHM pour les Données Qualitatives contient deux commandes :

#### ① Histogramme des données :

```
\HistDQ[donQual=..., donnees=..., captionfig=...]
```

pour les données enregistrée ci-dessus, on utilisera :

```
\HistDQ[donQual=jourAsymp.dat, donnees=downAsymp.dat, captionfig=caption pour la figure]
```

par défaut le paramètre `captionfig` est fixé à `captionfig=Histogramme des données`

#### ② Graphique circulaire des données :

```
\CirculaireDQ[donQual=..., donnees=..., captionfig=...]
```

pour les données enregistrée ci-dessus, on utilisera :

```
\CirculaireDQ[donQual=jourAsymp.dat, donnees=downAsymp.dat, captionfig=caption pour la figure]
```

par défaut le paramètre `captionfig` est fixé à `captionfig=Graphique circulaire des données`

# Exemples pour les Données Qualitatives

## Histogramme des Données Qualitatives

Exemple

1

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
\usepackage[colorlinks=true,urlcolor=blue]{hyperref}
\newcommand{\SSyracuse}{%
\href{http://melusine.eu.org/syracuse/}{Syracuse} }
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}

\HistDQ[donQual=MoisSyracuse.dat, donnees=VisitSyracuse.dat,
captionfig=histogramme du nombre de visiteurs du site
\SSyracuse entre janvier 2010 et avril 2010]

\end{document}
```

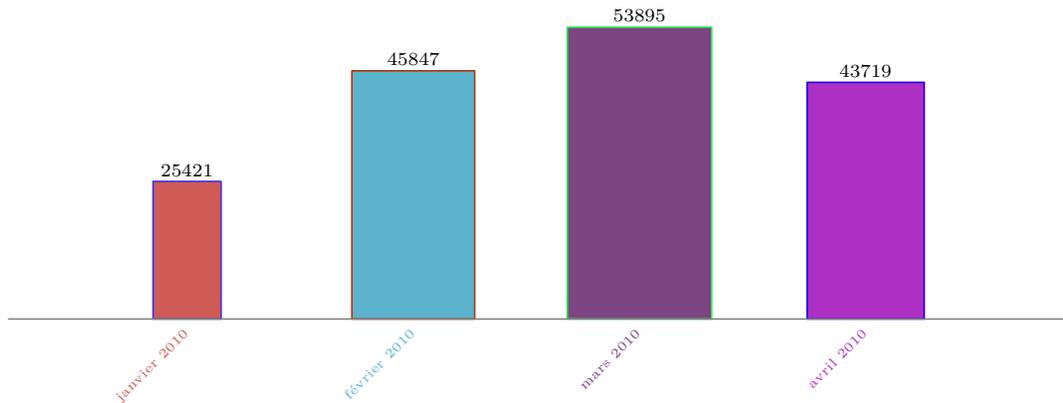


Figure 1: histogramme du nombre de visiteurs du site [Syracuse](http://melusine.eu.org/syracuse/) entre janvier 2010 et avril 2010

## Graphique circulaire des données Qualitatives

Exemple

2

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%\usepackage{hyperref}
%\usepackage{url}
\usepackage[colorlinks=true,urlcolor=blue]{hyperref}
\newcommand{\SSyracuse}{%
\href{http://melusine.eu.org/syracuse/}{Syracuse} }
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}

\CirculaireDQ[donQual=MoisSyracuse.dat, donnees=VisitSyracuse.dat,
captionfig=Graphique corculaire du nombre de visiteurs du site
\SSyracuse entre janvier 2010 et avril 2010]

\end{document}
```

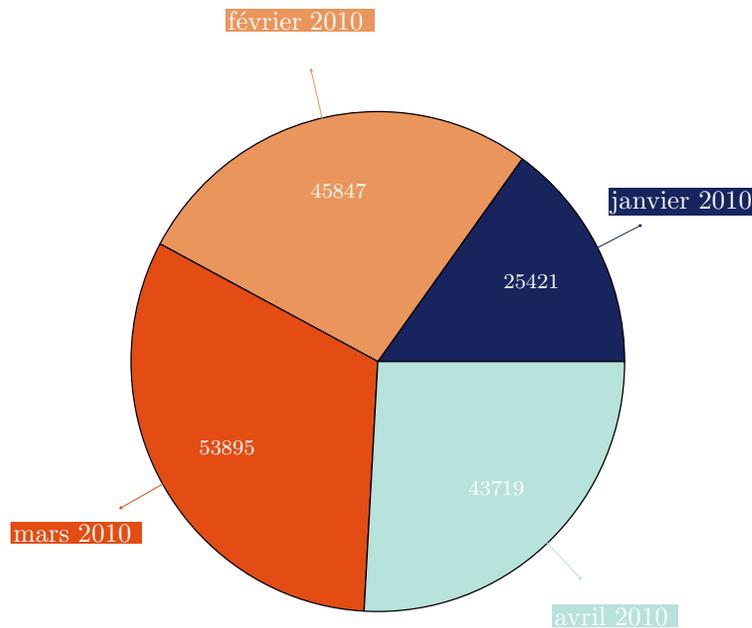


Figure 2: Graphique circulaire du nombre de visiteurs du site [Syracuse](http://melusine.eu.org/syracuse/) entre janvier 2010 et avril 2010

## 4 Données Discrètes

L'IHM pour les Données Discrètes contient cinq commandes :

### 1 Analyse des données [\AnalyseD\[donnees=fichier des données\]](#)

Pour la série statistique  $X = (x_1, \dots, x_n)$ , la commande [\AnalyseD](#) fournit le tableau suivant :

Nombre de données	$n$
Min	$\text{Min}\{x_1, \dots, x_n\}$
Max	$\text{Max}\{x_1, \dots, x_n\}$
Moyenne Arithmétique	$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n x_i$
Variance	$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X}_n)^2$
Médiane	 <a href="#">voir Median</a>
Mode	 <a href="#">voir Mode</a>
Moyenne Géométrique	 <a href="#">voir geometric mean</a>
Moyenne Quadratique	 <a href="#">voir Moyenne quadratique</a>
Moyenne Harmonique	 <a href="#">voir harmonic mean</a>
Skewness	 <a href="#">voir Skewness</a>
Kurtosis	 <a href="#">voir Kurtosis</a>

### 2 Histogramme des données [\HistD\[donnees=fichier des données\]](#)

### 3 Fonction de répartition des données [\FDRD\[donnees=fichier des données\]](#)

### 4 Graphique circulaire des données [\Circulaire\[donnees=fichier des données\]](#)

### 5 Histogramme et tableau contenant l'analyse des données

[\HTabD\[donnees=fichier des données\]](#)

## Exemples pour les Données Discrètes

### Analyse des données Discrètes

Exemple

3

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[francais]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\AnalyseD[donnees=discrets.dat]
\end{document}
```

Le résultat est le tableau suivant :

Résultats de l'analyse des données	
Nombre de données	58
Min	0
Max	20
Moyenne Arithmétique	9.362069
Variance	22.196492
Médiane	10
Mode	13
Moyenne Géométrique	7.638286
Moyenne Quadratique	10.480688
Moyenne Harmonique	6.136901
Skewness	-0.212706
Kurtosis	-0.740672

## Histogramme des données Discrètes

Exemple

4

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[francais]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\HistD[donnees=discrets.dat]
\end{document}
```

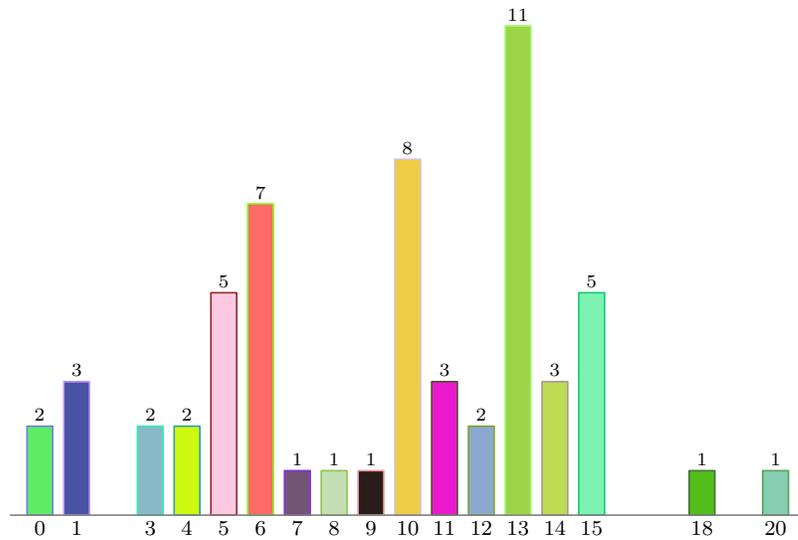


Figure 3: Histogramme des données

## Fonction de répartition des données discrètes

Exemple

5

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\FDRD[donnees-discrets.dat]
\end{document}
```

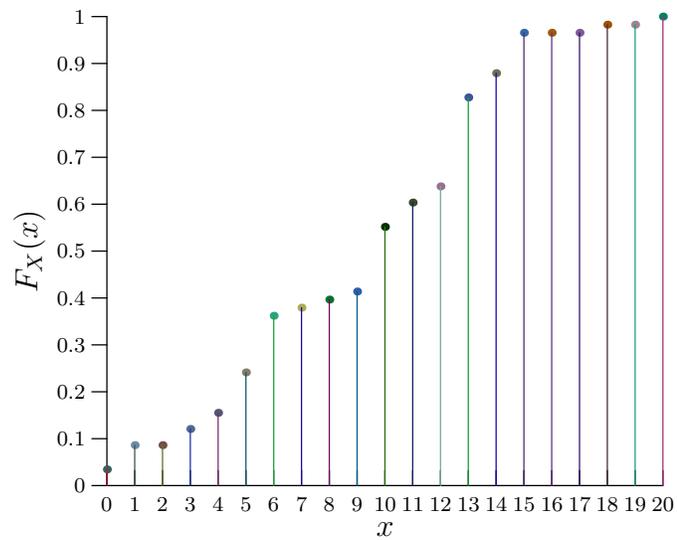


Figure 4: Fonction de répartition des données

## Représentation graphique circulaire des données discrètes

Exemple

6

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\Circulaire[donnees=discrets.dat]
\end{document}
```

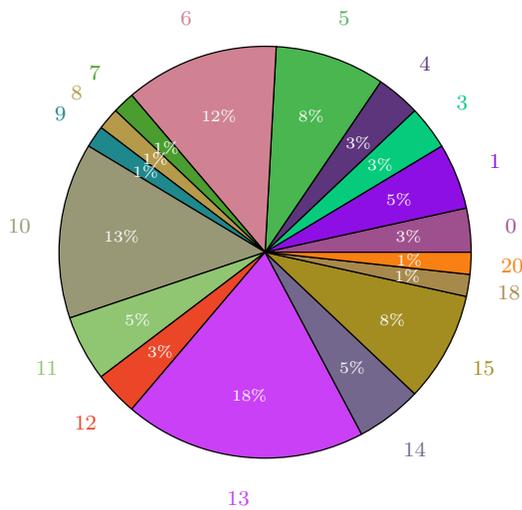


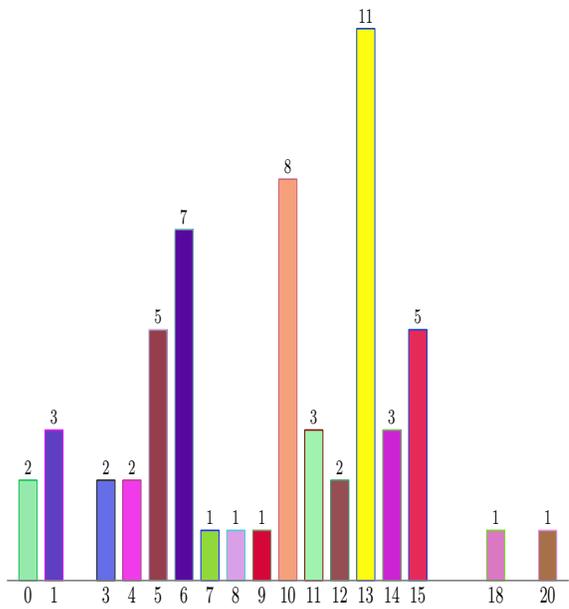
Figure 5: Graphique circulaire des données

## Histogramme et tableau contenant l'analyse des données Discrètes

Exemple  
**7**

```

\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\HTabD[donnees=discrets.dat]
\end{document}
    
```



Résultats de l'analyse des données	
Nombre de données	58
Min	0
Max	20
Moyenne Arithmétique	9.362069
Variance	22.196492
Médiane	10
Mode	13
Moyenne Géométrique	7.638286
Moyenne Quadratique	10.480688
Moyenne Harmonique	6.136901
Skewness	-0.212706
Kurtosis	-0.740672

Figure 6: Histogramme des données.

## 5 Données Continues

L'IHM pour les Données Continues contient quatre commandes :

### 1 Analyse des données

`\AnalyseC[donnees=fichier des données, nbClasses=nombre de classes]`

Pour la série statistique  $X = (x_1, \dots, x_n)$ , la commande `\AnalyseC` fournit le tableau suivant :

Nombre de données	$n$
Min	$\text{Min}\{x_1, \dots, x_n\}$
Max	$\text{Max}\{x_1, \dots, x_n\}$
Moyenne Arithmétique	$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n x_i$
Variance	$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X}_n)^2$
Moyenne Géométrique	 voir <a href="#">geometric mean</a>
Moyenne Quadratique	 voir <a href="#">Moyenne quadratique</a>
Moyenne Harmonique	 voir <a href="#">harmonic mean</a>
Médiane	 voir <a href="#">Median</a>
Skewness	 voir <a href="#">Skewness</a>
Kurtosis	 voir <a href="#">Kurtosis</a>
Classe Modale	 voir <a href="#">classe modale</a>

### 2 Histogramme des données

`\HistC[donnees=fichier des données, nbClasses=nombre de classes]`

### 3 Fonction de répartition des données

`\FDRC[donnees=fichier des données]`

### 4 Histogramme et tableau contenant l'analyse des données

`\HTabC[donnees=fichier des données, nbClasses=nombre de classes]`

# Exemples pour les Données Continues

## Analyse des données continues

Exemple

8

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% avril 2010 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% http://www.cnam.fr/depts/maths/Membres/ghorbanzadeh/
% Compiler avec l'option --shell-escape
% pour MikTeX : latex -enable-write18
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\AnalyseC[donnees=continues.dat , nbClasses=21]
\end{document}
```

Le résultat est le tableau suivant :

Résultats de l'analyse des données	
Nombre de données	200
Min	3.121417
Max	13.428487
Moyenne Arithmétique	7.926819
Variance	8.281799
Moyenne Géométrique	7.396028
Moyenne Quadratique	8.451652
Moyenne Harmonique	7.055507
Médiane	6.356319
Skewness	0.556529
Kurtosis	-1.228384
Classe Modale	[6.066294; 6.557107]

## Histogramme des données continues

Exemple

9

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\HistC[donnees=continues.dat,nbClasses=21]
\end{document}
```

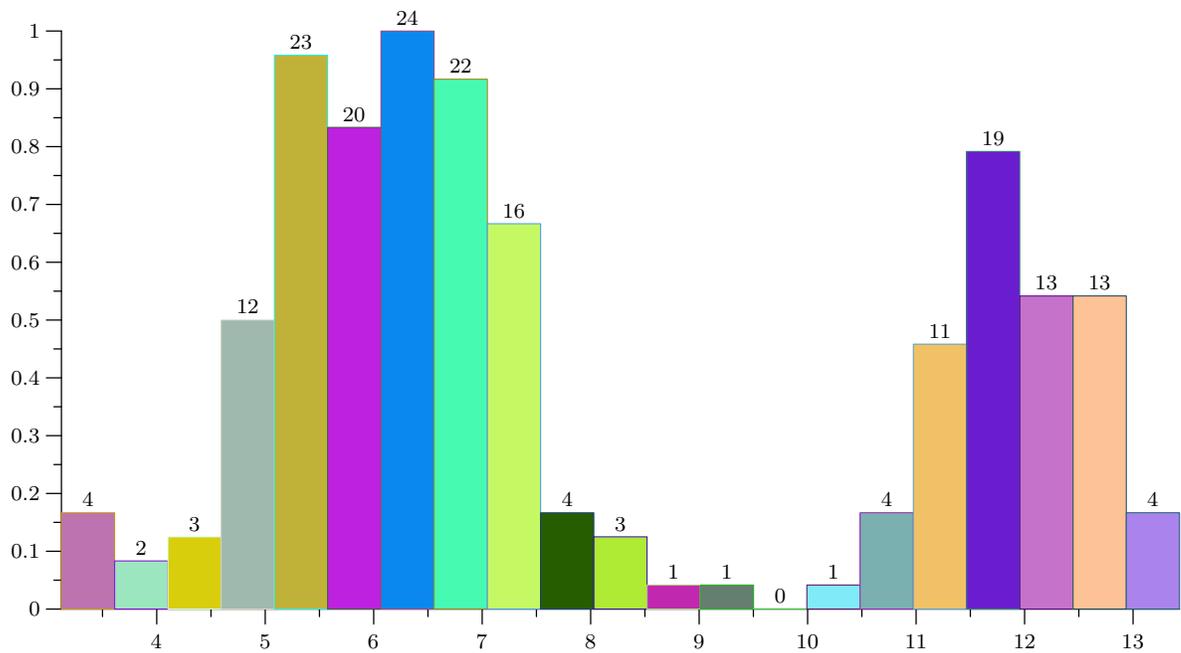


Figure 7: Histogramme des données réparties en 21 classes.

Exemple

10

### Fonction de répartition empirique des données continues

Pour la série statistique des données continues  $X = (x_1, \dots, x_n)$  la fonction de répartition empirique associée est définie par :

$$F_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{]-\infty, x]}(x_i)$$

```
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\FDRC[donnees=continues.dat]
\end{document}
```

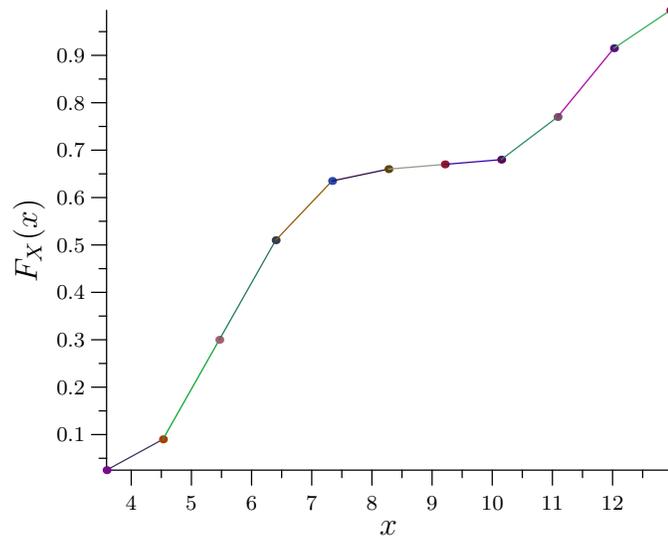


Figure 8: Fonction de répartition empirique des données

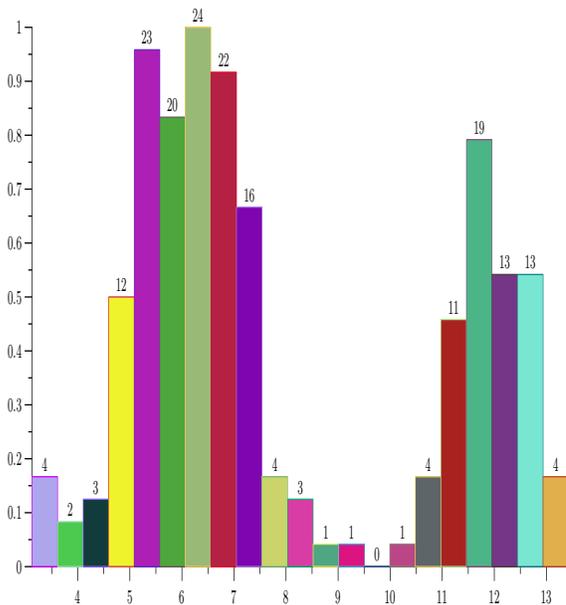
## Histogramme et tableau contenant l'analyse des données continues

Exemple

# 11

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Compiler avec l'option --shell-escape
% pour MikTeX : latex -enable-write18
\documentclass[12pt,a4paper]{article}
\usepackage[latin1]{inputenc}
\usepackage[français]{babel}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Appel au fichier contenant les macros
\input{statistique-Descriptive.tex}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
\HTabC[donnees=continues.dat, nbClasses=21]
\end{document}
    
```



Résultats de l'analyse des données	
Nombre de données	200
Min	3.121417
Max	13.428487
Moyenne Arithmétique	7.926819
Variance	8.281799
Moyenne Géométrique	7.396028
Moyenne Quadratique	8.451652
Moyenne Harmonique	7.055507
Médiane	6.356319
Skewness	0.556529
Kurtosis	-1.228384
Classe Modale	[6.066294; 6.557107]

Figure 9: Histogramme des données réparties en 21 classes.