

# Semaine 10

## Rappel de la statistique descriptive

# La statistique

## La statistique

est la science de la **variabilité**.



21	21	23	22	23	24
21	18	23	19	18	20
24	20	20	19	19	22
21	18	20	23	17	20
25	23	21	14	18	29

Hauteur (en cm)  
des *Onobrychis*  
*viciifolia*

Variabilité

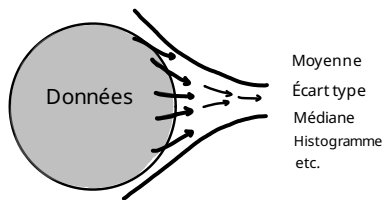


- 1 de la **confusion** : Comment résumer les données ? Comment décrire la variabilité ?
- 2 de l'**incertitude** : Jusqu'où peut-on généraliser les résultats d'une expérience ?
- 3 des **questions scientifiques** : Quelles sont les sources de variabilité ?

# Statistique descriptive et inférentielle

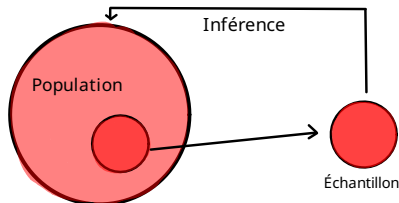
## La statistique descriptive

est un ensemble de méthodes qui traitent les données de sorte à les résumer et à extraire de l'information utile. Elle adresse le problème de **confusion**.



## La statistique inférentielle

est un ensemble de méthodes qui tentent de généraliser les résultats d'une expérience au reste de la population. Elle adresse le problème de **l'incertitude**.



# Vocabulaire

## Une population

est l'ensemble des individus d'intérêt d'une étude.

## Une variable

est une caractéristique d'intérêt *mesurable* d'une population.

Âge, poids, salaire, appréciation (d'un cours, d'un service, etc.),  
nombre de points marqués

## Les modalités

sont les **valeurs** *possibles* d'une variable.

## Un échantillon

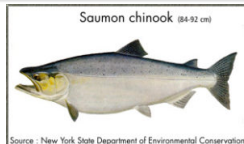
est un ensemble de quelques individus *représentatifs* de la population.

# Exemple

## Mise en situation

Des chercheur·e·s s'intéressent aux espèces, à la couleur, à la taille et au poids des poissons du lac Ontario.

- Individus : poissons
- Population : poissons du lac Ontario
- Variables : Espèces, couleur, taille, poids
- Modalités
  - ▶ espèce : Truite arc-en-ciel, saumon chinook, etc.
  - ▶ couleur : Gris, picoté, etc.
  - ▶ taille :  $]0, \infty[$  (en cm)
  - ▶ poids :  $]0, \infty[$  (en kg)



# Types de variables

## Deux types principaux

**Qualitative** et **quantitative**

Une variable qualitative mesure une **qualité**.

Appréciation d'un service

très insatisfait, insatisfait, satisfait, très satisfait

Couleur préférée : rouge, bleu, etc.

Une variable quantitative mesure une **quantité**.

Âge, poids, taille, température, salaire

# Quelle est le type de chaque variable ?

- Temps pour courir 100m
- Nombre d'enfants d'une famille
- Quantité de pluie tombée
- Le niveau de satisfaction des clients
- Le nombre de jours ensoleillés dans le mois d'août
- La couleur préférée

# Types de variables qualitatives

## Nominale

Les valeurs possibles sont des noms, des étiquettes ou des catégories sans ordre.

P.ex. : sexe, mode de transport, espèce.

## Ordinale

Valeurs possibles **ordonnées**.

P.ex. : Échelle de 1 à 10, nombre d'étoiles, échelle de satisfaction.

## Exemple

Une compagnie aérienne souhaite connaître le niveau de satisfaction de ses clients sur leur service. Elle leur demande

Êtes-vous satisfait de votre vol aujourd'hui ?

**Très insatisfait    Insatisfait    Satisfait    Très satisfait**

### Résultats

Très insatisfait	Insatisfait	Satisfait	Très satisfait
7	13	34	19

$X$  = Niveau de satisfaction du vol de cette journée

$X$  est une variable qualitative.

# Exemple (provenant de statcan)

## Mode de transport habituel utilisé par les Canadiens pour se rendre au travail

Mode de transport pour se rendre au travail	Nombre de personnes
Automobile, camion ou fourgonnette (conducteur)	9 929 470
Automobile, camion ou fourgonnette (passager)	923 975
Transport en commun	1 406 585
À pied	881 085
Bicyclette	162 910
Autres moyens	146 835

# Types de variables quantitative

## Variables quantitatives discètes

prennent un nombre **limité** de valeurs *possibles* (un ensemble au plus dénombrable).

P.ex. : 0,1,2,3,4,5,6,...

$$\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}, \frac{6}{10}, \frac{7}{10}, \frac{8}{10}, \frac{9}{10}, \frac{10}{10}$$

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}$

## Variables continues

prennent des valeurs possibles dans un *intervalle*.

P.ex. :  $[a,b]$

$[a, \infty), (-\infty, b)$

$\mathbb{R}$

# Tableaux de fréquence

Exemple : Pour une longue distance (plus de 500km), vous préférez vous déplacer en

autobus    auto    avion    autre ?

Observations : B B A B V T T V B A A A

# Tableaux de fréquence

Exemple : Pour une longue distance (plus de 500km), vous préférez vous déplacer en

autobus    auto    avion    autre ?

Observations : B B A B V T T V B A A A

Auto (A)	4
Bus (B)	4
Avion (V)	2
Autre (T)	2

# Tableaux de fréquence

Exemple : Pour une longue distance (plus de 500km), vous préférez vous déplacer en

autobus   auto   avion   autre ?

Observations : B B A B V T T V B A A A

Mode de transport	Fréquence absolue	Fréquence relative
Auto (A)	4	33,3%
Bus (B)	4	33,3%
Avion (V)	2	16,6%
Autre (T)	2	16,6%
Total	12	100%

# Tableaux de fréquence

Exemple : Pour une longue distance (plus de 500km), vous préférez vous déplacer en

autobus auto avion autre ?

Observations : B B A B V T T V B A A A

Mode de transport	Fréquence absolue	Fréquence relative
Auto (A)	4	33,3%
Bus (B)	4	33,3%
Avion (V)	2	16,6%
Autre (T)	2	16,6%
Total	12	100%

Un tableau de fréquence

contient le nombre de fois que chaque modalité a été observée.

# Tableau de fréquence

## Le tableau de fréquence

- sert à présenter l'information de façon **simple et concise**
- permet de faire d'autres analyses statistiques (plus tard dans la session)

## La fréquence relative

- peut s'interpréter comme une *probabilité*.  
(Attention ! Il faut un grand nombre de données pour que cela fonctionne.)
- On l'appelle aussi une *proportion*.

## Exemple : fréquence relative et probabilité

Groupe sanguin	Fréquence relative
O	46%
A	42%
B	9%
AB	3%

Répartition des groupes sanguins au Canada

### Interprétation

Si on choisit une personne au hasard, il y a une probabilité de 46% que son groupe sanguin soit  $O+$  ou  $O-$ .

### Attention

Cette interprétation est seulement valable si le nombre de données est assez grand !

# Données en vrac vs données regroupées

## Taille d'un échantillon de 100 personnes (en cm)

166.8, 174.5, 160.3, 152.3, 176.6, 174.7, 166.9, 166.0, 159.8, 170.0, 155.4,  
161.9, 152.0, 164.8, 163.9, 159.0, 158.5, 151.7, 157.0, 175.4, 164.6, 169.5,  
160.9, 156.2, 167.5, 172.7, 158.9, 175.5, 167.1, 170.0, 152.4, 156.3, 175.0,  
157.5, 161.8, 162.9, 166.5, 168.3, 156.0, 158.7, 164.5, 170.9, 156.2, 177.4,  
172.0, 159.7, 166.8, 156.5, 161.5, 166.2, 173.6, 168.3, 165.1, 157.7, 162.3,  
160.3, 165.5, 176.3, 162.0, 149.5, 163.2, 158.8, 158.0, 162.3, 153.6, 160.5,  
174.2, 156.4, 163.9, 163.1, 155.3, 166.8, 156.0, 160.9, 154.5, 159.2, 157.7,  
156.7, 160.8, 174.3, 165.5, 167.7, 177.9, 167.3, 160.6, 161.2, 154.9, 166.7,  
159.3, 179.4, 167.4, 172.4, 154.4, 162.2, 157.4, 156.6, 156.7, 151.9, 160.2, 155.2

Taille (en cm)	nombre
[149, 155[	10
[155, 161[	36
[161, 167[	27
[167, 173[	14
[173, 180[	13

# Données en vrac vs données regroupées

## Données en vrac

L'information de chacune des données est connue.

- Plus précis
- Plus difficile à lire
- Temps de calcul plus long

## Données regroupées

Les données sont regroupées en *classes*. Dans ce cas, seul le nombre de données par classe est connue.

- Moins précis
- Plus facile à lire
- Plus facile pour les calculs

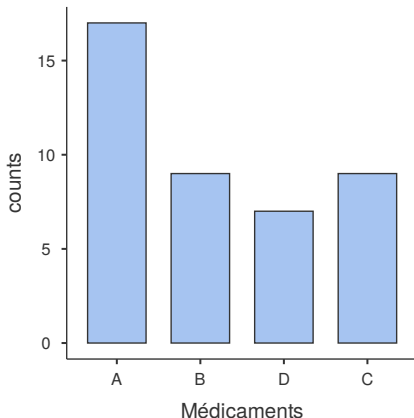
Souvent, on ne choisit pas sous quelle forme on reçoit les données.

# Diagramme à bandes

## Un diagramme à bande

est une représentation graphique des données d'une variable **qualitative** ou d'une variable **quantitative discrète**.

Médicaments	Nb. pers.
A	17
B	9
C	9
D	7

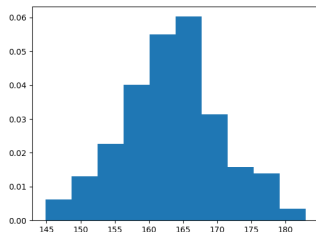


# Histogramme de densité

## Un histogramme densité

est une représentation graphique des données regroupées d'une variable continue telle que

- 1 à chaque classe est associée une bande ;
- 2 une bande est au-dessus de l'intervalle de sa classe ;
- 3 la largeur d'une bande est la largeur de sa classe ;
- 4 l'**aire** d'une bande est la fréquence relative de sa classe.

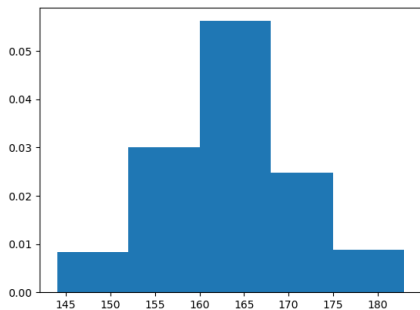


## Histogramme : exemple

Taille	Freq. abs.	Freq. rel.	Densité
[144, 152[	20		
[152, 160[	72		
[160, 168[	135		
[168, 175[	52		
[175, 183]	21		

# Histogramme

Taille	Freq. abs.	Freq. rel.	Densité
[144, 152[	20	6,6%	0,8%
[152, 160[	72	24%	3%
[160, 168[	135	45%	5,6%
[168, 175[	52	17,3%	2,5%
[175, 183]	21	7%	0,9%



Histogramme de densité  
de la taille (en cm)  
d'un échantillon de Canadiens·ne·s

# Différences entre un diagramme à bande et un histogramme

## Diagramme à bande

- la largeur des bandes est esthétique seulement

Utiliser pour les variables

- qualitatives
- quantitatives discrètes

## Histogramme de densité

- la largeur des bandes est la longueur de la classe
- l'aire de la bande est importante

Utiliser pour les variables

- quantitatives continues

# Représentation graphique

## La distribution des données

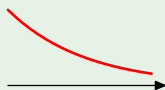
réfère à la **répartition** des différentes valeurs possibles.

- Les données suivent-elles un motif connu ?

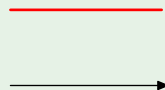
## Voici quelques distributions vues dans le cours



Loi normale



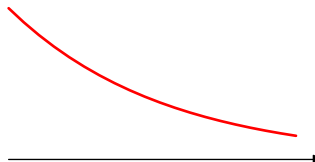
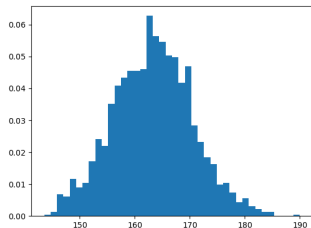
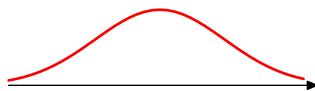
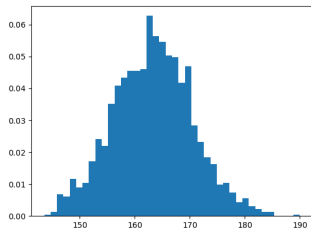
Loi exponentielle



Loi uniforme

# Exemple 1

Comparaison d'un échantillon de la taille de 2000 personnes (fictif) contre des distributions connues



## Exemple 2

On mesure la quantité de technetium-99 dans un patient à intervalle régulier jusqu'à obtenir 100 observations.

