

Tous droits réservés Tutorat Santé Brestois ©  
Toute diffusion et reproduction, totale ou  
partielle, de ce document est interdite

# Statistiques descriptives

Stage de Pré-Rentrée 2022  
Pôle Biostatistiques

D'après le cours du Professeur Morin



## Petit message d'avertissement avant de commencer :

Nous vous rappelons que ce diaporama, réalisé par des étudiants, est une aide et **non un support de cours officiel** et ne peut donc pas être considéré comme un ouvrage de référence lors de l'examen de PASS ou de L.AS.

Il se base sur le **cours de l'année précédente** qui peut être **amené à être modifié** dans sa forme et son contenu au bon vouloir du professeur.

Have fun ;)

# Introduction

Population : collectif étudié

Individu : éléments qui composent la population

Échantillon : partie de la population

Variable aléatoire : propriétés observées sur les individus

Modalités : différentes valeurs que peut prendre une variable

Variables aléatoires qualitatives	Variables aléatoires quantitatives
Caractère nominal : ne se chiffre pas Nombre fini de modalités  <i>Ex : couleur des yeux, des cheveux...</i>	Caractère numérique : dénombrable <ul style="list-style-type: none"><li>• Discrète : nombre fini de modalités <i>Ex : nbr d'enfants par famille</i></li><li>• Continue : nombre infini de modalités <i>Ex : le poids, taille, âge...</i></li></ul>

Les variables aléatoires **continues** sont discrétisées en classes.

# Indicateurs numériques

## A. Indicateur de tendance centrale

### Moyenne, me ou $\bar{x}$

→ sensible aux valeurs extrêmes

$$\bar{x} = \frac{1}{N} \times \sum_{i=1}^{i=k} n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n_1 + n_2 + \dots + n_k}$$

Quotient de la somme de toute les valeurs de la série par l'effectif total  $N$

### Médiane (ou centile 50)

→ attirée par les grands effectifs  
→ pas toujours une valeur de la série

→ moins sensible aux valeurs extrêmes que la moyenne

Sépare la série statistique en 2 → valeur de la variable au dessus ou au dessous de laquelle il y a 50% des individus

N pair :

$$K = \frac{N}{2} \text{ et } C_{50} = \frac{(V_k + V_{k+1})}{2}$$

N impair :

$$K = \frac{(N - 1)}{2} \text{ et } C_{50} = V_{k+1}$$

### Mode

Valeur de la variable avec le plus grand effectif

# Indicateurs numériques

## B. Mesure de tendance dispersion

### Variance = $Se^2$

+ la dispersion est grande + la variable est grande = prépondérance au grand écart

→ unité de la variable de départ au **carré**

$$1/N \sum (v_i - me)^2 = (1/N \sum v_i^2) - me^2$$

Moyenne de la variable au carrée - moyenne au carré

### Ecart type = $Se$

→ dispersion des valeurs autour de la moyenne

→ même unité que la variable de départ

$$\sqrt{Se^2}$$

Il y a aussi l'étendue, et l'écart moyen en VA

# Indicateurs numériques

## C. Exemple

Calcul de variance :

A Brest, une équipe de chercheurs enquête sur le nombre de dauphin aperçu dans la mer d'Iroise par les touristes au cours de l'été. A la fin de l'étude, on obtient le tableau suivant :

Nb de dauphins aperçus	0	1	2	3
Effectifs	<b>125</b>	<b>85</b>	<b>102</b>	<b>12</b>

1) On calcule la moyenne,

$$me = \frac{x_1n_1 + x_2n_2 + \dots}{N} = \frac{0 \times 125 + 1 \times 85 + 2 \times 102 + 3 \times 12}{325} = \frac{325}{325} = 1$$

# Indicateurs numériques

## C. Exemple

Calcul de variance :

A Brest, une équipe de chercheurs enquête sur le nombre de dauphin aperçu dans la mer d'Iroise par les touristes au cours de l'été. A la fin de l'étude, on obtient le tableau suivant :

Nb de dauphins aperçus	0	1	2	3
Effectifs	125	85	102	12

- 1) On calcule la moyenne = 1
- 2) On calcule la moyenne de la **variable au carré**,

$$\frac{125 \times 0 + 85 \times 1^2 + 102 \times 2^2 + 12 \times 3^2}{325} = \frac{125 \times 0 + 85 \times 1 + 102 \times 4 + 12 \times 9}{325} = \frac{6013}{25} \approx 1,85$$

# Indicateurs numériques

## C. Exemple

Calcul de variance :

A Brest, une équipe de chercheurs enquête sur le nombre de dauphin aperçu dans la mer d'Iroise par les touristes au cours de l'été. A la fin de l'étude, on obtient le tableau suivant :

Nb de dauphins aperçus	0	1	2	3
Effectifs	<b>125</b>	<b>85</b>	<b>102</b>	<b>12</b>

- 1) On calcule la moyenne = 1
- 2) On calcule la moyenne de la **variable au carré**
- 3) On calcule la variance,  $Se^2 = \text{moyenne de la variable au carré} - \text{carré de la moyenne}$

$$Se^2 = 1,85 - 1^2 = 0,85$$



# Tables

Information morphologique élevée sur la répartition des valeurs. On peut utiliser des variables qualitatives et quantitatives.

## A. Tables de fréquences simples

Montre l'effectif ou la fréquence pour les modalités d'une variable. Donne la répartition des valeurs dans chaque classe. Pour des variables qualitatives ou **quantitatives mises en classe**.

Hospitalisation (j)	≤1	2 à 3	> 3
<b>Effectifs</b>	72	29	14
<b>Fréquences</b>	0,626	0,252	0,122
<b>Fréquences (%)</b>	62,6	25,2	1,22

N = 115

# Tables

## B. Tables de fréquences cumulées

Chaque valeur de fréquence est remplacée par la somme de toutes les **précédentes** pour les tables cumulées **croissantes**, ou la somme des **suivantes** pour les tables cumulées **décroissantes**.

Hospitalisation (j)	≤1	2 à 3	> 3
Effectifs	72	29	14
Cumul <b>croissant</b>	72	$72 + 29 =$ 101	$72 + 29 + 14 =$ 115
Cumul <b>décroissant</b>	$14 + 29 + 72 =$ 115	$14 + 29 =$ 43	14

# Tables

## C. Tables croisées (ou de contingence)

Pour visualiser la relation entre :

- 2 variables quantitatives ou
- 2 variables qualitatives ou
- 1 variable quantitative et 1 variable qualitative

Montre un possible lien entre les variables = relation de **dépendance**

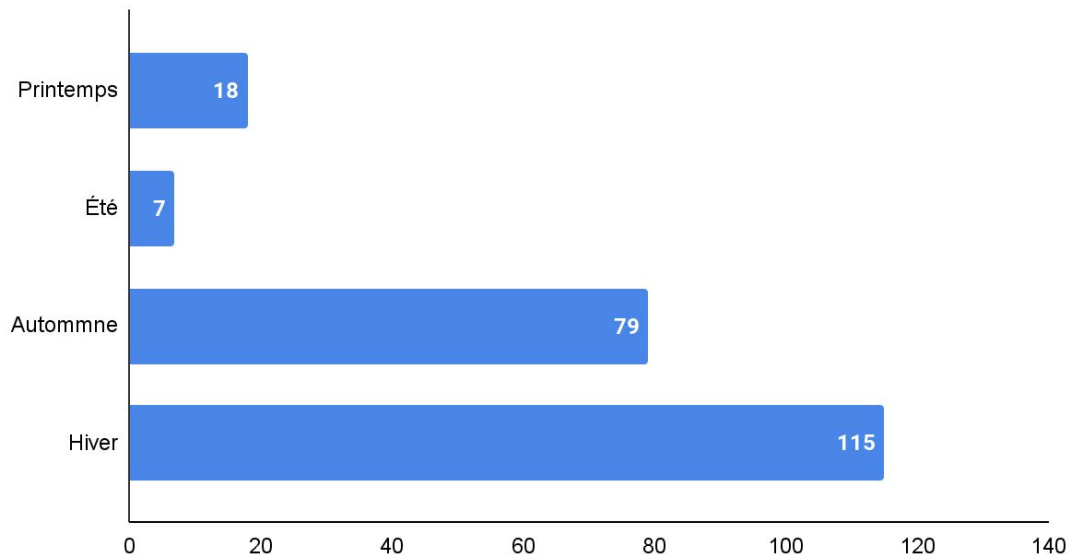
Diag \ Taux psa	$\leq$ seuil	$\geq$ seuil	
	HBP	253	38
Cp	3	24	27
Sain	328	54	382
	584	116	700

**Effectifs marginaux**

# Graphiques

## A. Graphique en barre

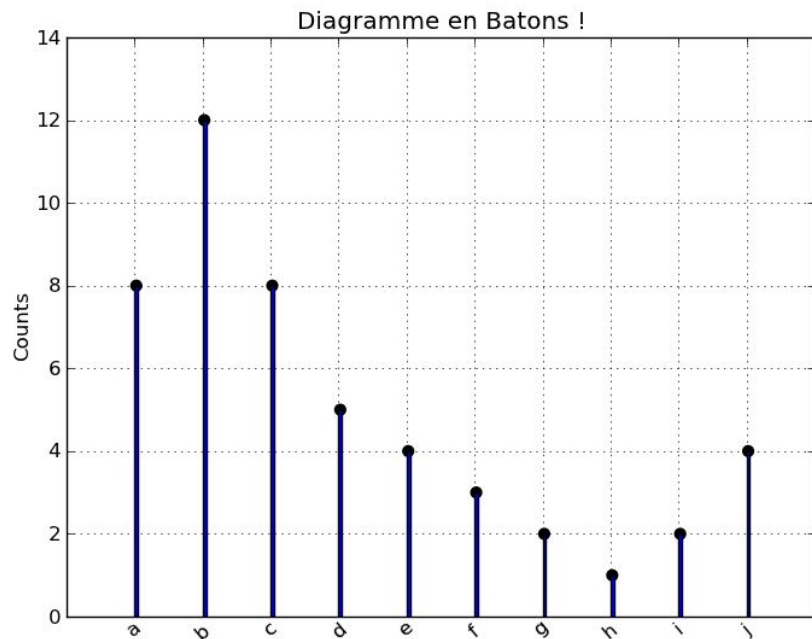
Consultations pour une angine



- Variable **qualitative** ou **quantitative mise en classe**
- **Hauteur** des barres **proportionnelle** à l'effectif mais pas la surface

# Graphiques

## B. Graphique en bâton

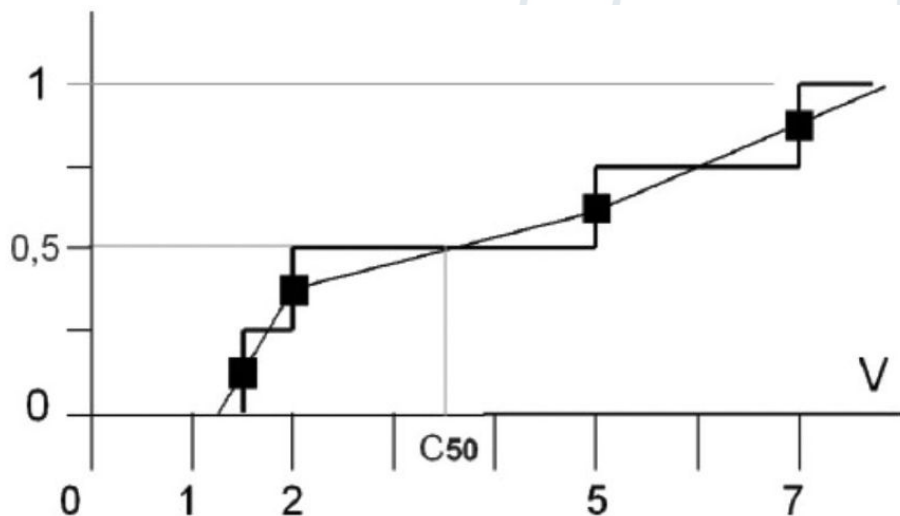


- Variables **quantitatives discrètes**
- Comme le graphique en barre mais avec plus de modalités
- **Hauteur proportionnelle** à l'effectif

# Graphiques

## C. Graphique des centiles

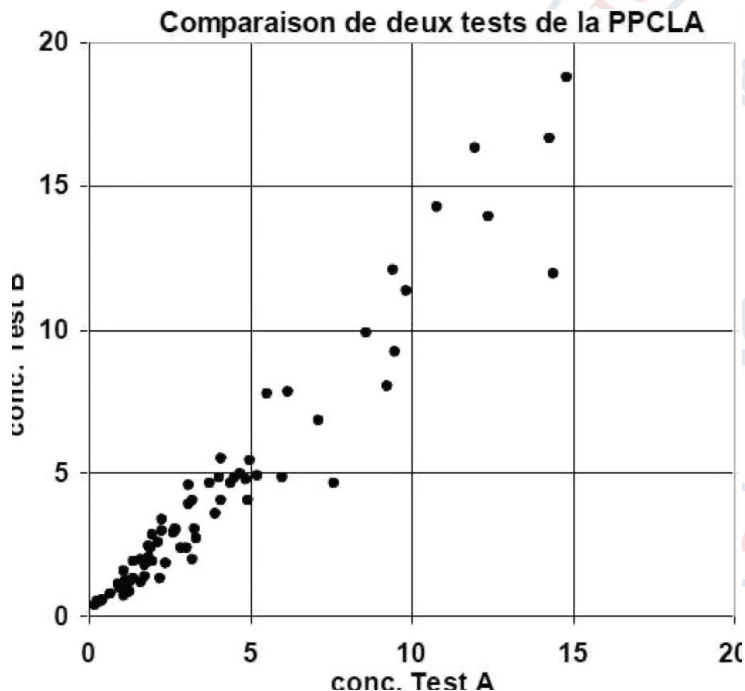
- Sur l'axe des abscisses :  $n$  valeurs mises en ordre croissant
- Sur l'axe des ordonnées : graduation de 0 à 1 correspondant aux centiles



- Variable **quantitative**
- Indique la médiane (C50), le quartile 1 (C25), le quartile 3 (C75)...
- Image expérimentale de la **fonction de répartition**

# Graphiques

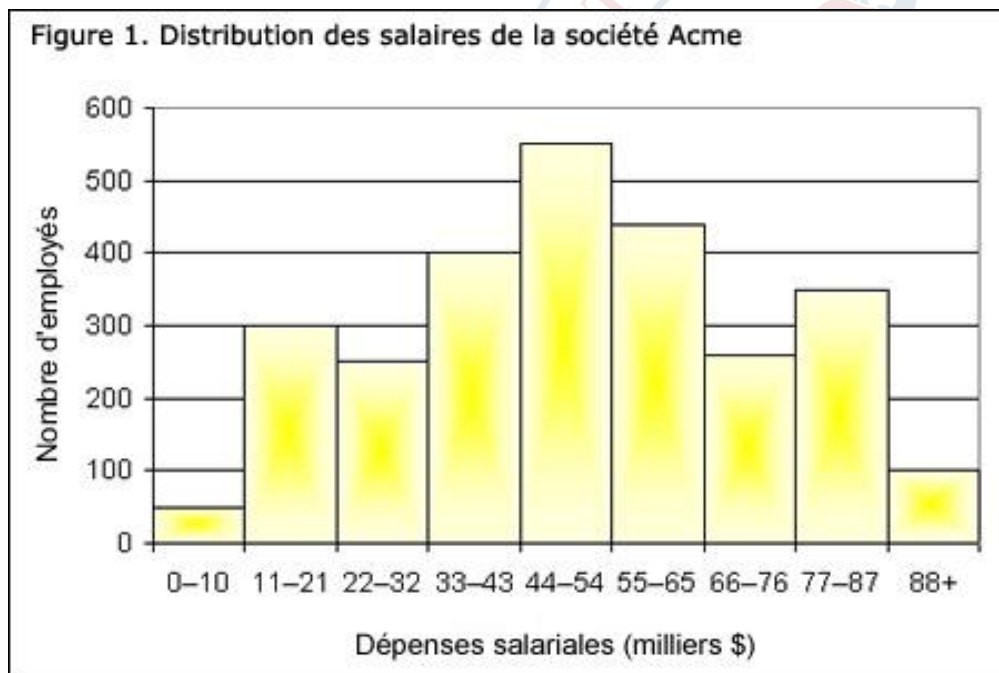
## D. Graphique cartésien



- Variables quantitatives **continues** (forme de courbe)
- Variables quantitatives **discrètes** (nuage de points)
- Relation entre 2 variables X et Y quantitatives = étude de **corrélacion**
- Si les points sont **ordonnés** = **dépendance**

# Graphiques

## E. Histogramme

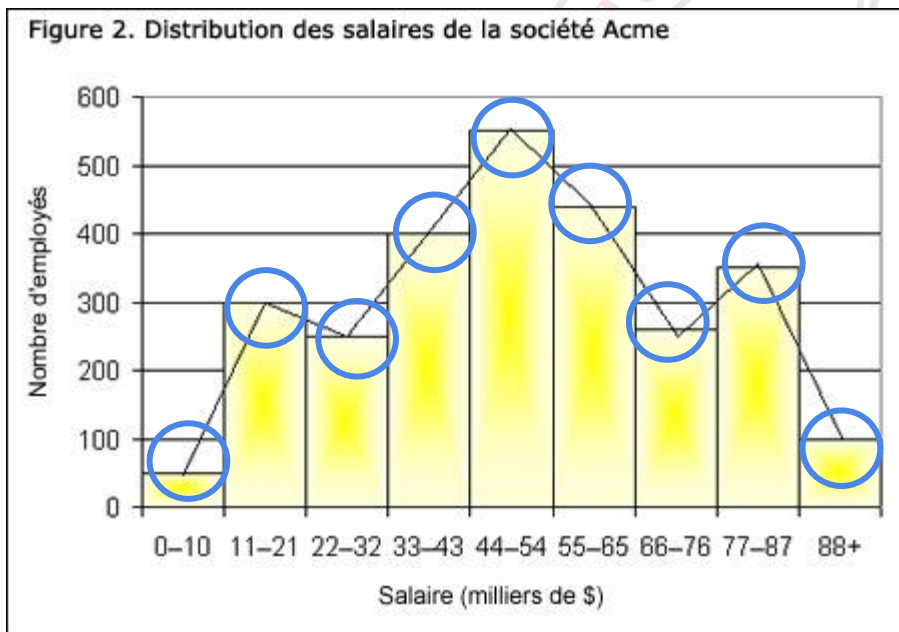


- Variable **quantitative** discrétisée en **classes**
- **!** Les classes ne sont pas toujours de la même largeur **!**
- **Surface** des rectangles **proportionnelle** à l'effectif de la classe
- Représentation d'une **fonction de densité**



# Graphiques

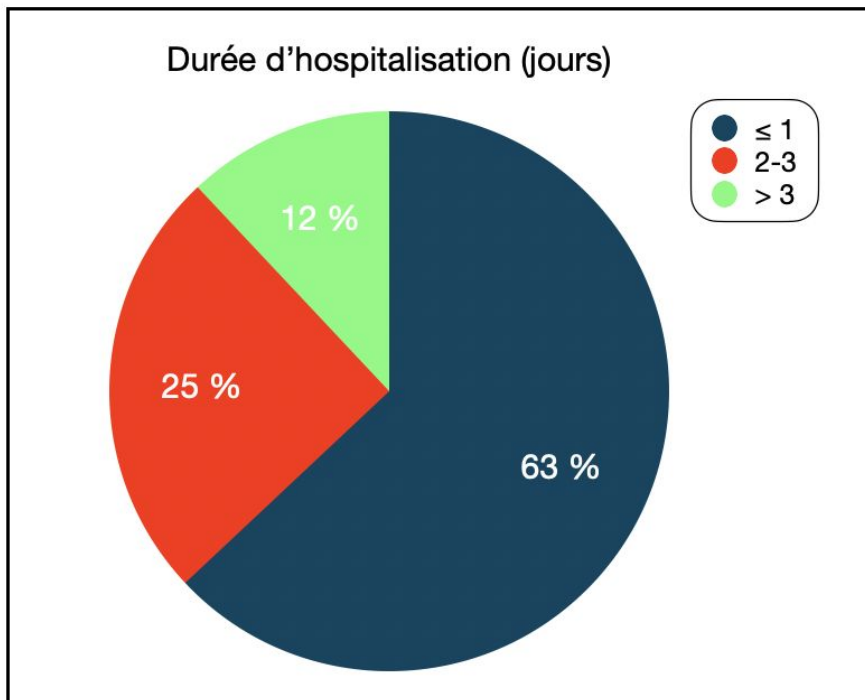
## F. Polygone des fréquences



- Variable **quantitative** mise en **classes**
- Obtenu en traçant des segments joignant les **milieux des sommets** des rectangles
- **Surface sous le polygone = Surface des rectangles**
- Fonction de **densité**

# Graphiques

## G. Graphe circulaire



- Variable **quantitatives et qualitatives**
- Disque dont la surface correspond à la taille de l'échantillon
- **Secteur et angle** du disque **proportionnelle** à la **fréquence** de la modalité
- Variable à faible nombre de modalités

# Graphiques

## H. Récapitulatif

	Qualitative	Quantitative			Entre 2 variables
		Discrète	Continue	Continue en classe	
Tables fréquences	✓	✓		✓	
Tables croisées	✓	✓		✓	✓
Graphe en barre	✓				
Graphe bâton		✓			
Graphe cartésien		✓	✓	✓	✓
Graphe des centiles		✓	✓	✓	
Graphe circulaire	✓	✓		✓	
Histogramme				✓	



# Cours imprimés

