



**UMF**  
UNIVERSITATEA DE  
MEDICINĂ ȘI FARMACIE  
IULIU HAȚIEGANU  
CLUJ-NAPOCA

# Statistiques (Mesures) Descriptives (1)

# Plan du cours

## STATISTIQUES DESCRIPTIVES

Mesures de tendance centrale

Mesures de position

Mesures de dispersion (cours 05)

Mesures de symétrie et aplatissement (cours 05)

Rapport, proportion & Taux

Représentations graphiques des données quantitatives

Représentations graphiques des qualitatives

# Généralités: Qu'est-ce que les statistiques ?

- « There are three kinds of lies: lies, damned lies, and statistics. »  
(Mark Twain qui l'a attribuée au Benjamin Disraeli)
- **La statistique** : domaine de la mathématique qui a pour but la collecte, la description et l'analyse des données afin d'en extraire des conclusions (inférences).
- **Une statistique** : un nombre/valeur calculé à partir d'un ensemble des données
- **Les statistiques** : collection de nombres/valeurs relatif à un ensemble des données

## Notions de base de statistique: *population cible, accessible*

- **La Population** = ensemble d'objets ou de personnes étudiées,

### Exemples:

- ⊗ ensemble des patients atteints de COVID-19 hospitalisés
- ⊗ ensemble des implants dentaires

Population *cible*: la population on désire a étudier, la population auquel on veut extrapoler le résultats a la fin de l'étude sur un échantillon

Population *accessible*: la population auquel on a accès: cabinet médical/dentaire, hôpital, école, ...

**Population cible**



**Population accessible**

Patients atteints de COVID-19 hospitalisés



Patients atteints de COVID-19 suivi dans un hôpital

# Notions de base de statistique: *échantillon, individu*

- **L'échantillon** = partie de la population sur laquelle on va porter l'enquête

## Exemples:

- ◎ 200 patients hospitalisés pour COVID-19 dans le service de médecine interne du CHU Pitié-Salpêtrière
- ◎ 100 étudiants de la MGFR inscrits à l'UMF Cluj-Napoca
- Choisie par des différentes méthodes échantillonnage (voir les cours suivants)
- **L'individu** = unité statistique = l'élément (objet, personne/famille ou un événement) sur lequel on va porter l'étude

Population cible



Population accessible



Echantillon



# Notions de base de statistique:

## Variables statistiques versus Données

**Variable:** la caractéristique (démographique, clinique, ...) étudié sur les différents individus (appartenant a un échantillon)

- ✓ Variables qualitatives (dichotomiques, ordinales, nominales) et quantitatives (discrètes, continues)
- ✓ Variables indépendantes et variables dépendantes

**Donnée:** la « valeur »(peut-être un nombre ou une modalité/catégorie de type « oui »/ « non ») de la variable

- une ou plusieurs variables peuvent être mesuré sur un individu ->
- **Série statistique de données** = valeurs des variables mesurées lors d'une étude

# Notions de base de statistique:

## Type d'une série statistique

Nombre de variables	Série statistique
1	<b>uni</b> variée
2	<b>bi</b> variée
3	<b>tri</b> variée
>3	<b>multi</b> variée

Nombre de variables <b>quantitatives</b>	Série statistique
1	uni <b>dimensionnelle</b>
2	bi <b>dimensionnelle</b>
3	tri <b>dimensionnelle</b>
>3	multi <b>dimensionnelle</b>

# Type d'une série statistique: exemples

- unidimensionnelle: âges (ans)

	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>
Âges (ans)	20	45	70	67	60	35	55

- bidimensionnelle: âges (ans) et Délai entre le début de la COVID-19 et l'hospitalisation (jours)

	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>
Âges (ans)	20	45	70	67	60	35	55
Délai entre le debut de la COVID-19 et l'hospitalisation	10	6	4	12	2	10	5

- univarié: Embolie pulmonaire (oui=1;non=0)

	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>
Embolie pulmonaire	0	1	0	1	1	1	0



# Mesures (statistiques) descriptives de tendance centrale

Moyenne arithmétique

Mediane

Mode

Valeur centrale

Moyenne pondérée

Moyenne géométrique

!!!! Statistiques observées sur l'échantillon  $\approx$  paramètre de population

# Échantillon versus Population: **Statistiques** vs. **Paramètres**

Population → **paramètre**

Moyenne arithmétique calculée sur une population

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

$x_i$  = la valeur de la variable quantitative pour le  $i$ -ème patient  
 $N$  = taille de la population

Échantillon → **statistique**

Moyenne arithmétique calculée sur un échantillon

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

$x_i$  = la valeur de la variable quantitative pour le  $i$ -ème patient  
 $n$  = taille de l'échantillon

# Moyenne arithmétique (Ma)

## Avantages

- ✓ Toute valeur de la série est prise en compte dans le calcul de la moyenne
- ✓ la somme des écart à la moyenne est nulle:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

## Désavantages

- ✗ peut être influencée par les valeurs extrêmes
- ✗ modification d'une seule valeur de la série de données va influencer (changer) la moyenne
- ✗ n'a de sens que pour une variable quantitative

# Médiane (Me): définition et calcul

- **Médiane:** la valeur qui partage la série des individus en 2 groupes d'effectifs égaux
- **Comment trouver/determiner le médiane?**
  - triez / ordonnez les données en ascendant
  - regardez la taille de l'échantillon («  $n$  »)

$$Me = \begin{cases} x_{\frac{n+1}{2}} & \text{si } n \text{ est impaire,} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{si } n \text{ est paire.} \end{cases}$$

# Médiane (Me):

## Avantages

- ✓ variables quantitatives et qualitatives ordinales
- ✓ peu sensible aux valeurs extrêmes contrairement à la moyenne

## Désavantages

- ✗ elle ne se prête pas à des opérations algébriques
- ✗ ne s'applique pas aux variables qualitatives nominales

# Exemple de calcul: moyenne et médiane

	continue	continue ~ discrète	dichotomique	ordinaire
Patient	Taille (cm)	Age (années)	Sexe	Catégories d'IMC <sup>(a)</sup>
1	170	50	F	Corpulence normale
2	160	45	F	Pré-obésité
3	187	38	H	Obésité de classe I
4	172	25	H	Corpulence normale
5	157	65	F	Pré-obésité
6	175	56	H	Corpulence normale

Variable: Âge

**Moyenne arithmétique (Ma) =?**

$Ma = (50 + 45 + 38 + 25 + 65 + 56) / 6 = 46,5$  ans

**Médiane (Me)=?**

Série statistique ordonnée: 25, 38, **45, 50**, 56, 65

Taille de l'échantillon: nombre paire = 6 =>

$Me = (45+50) / 2 = 47,5$

**Paire (2, 4, ...)**

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

# MODE: définition et calcul

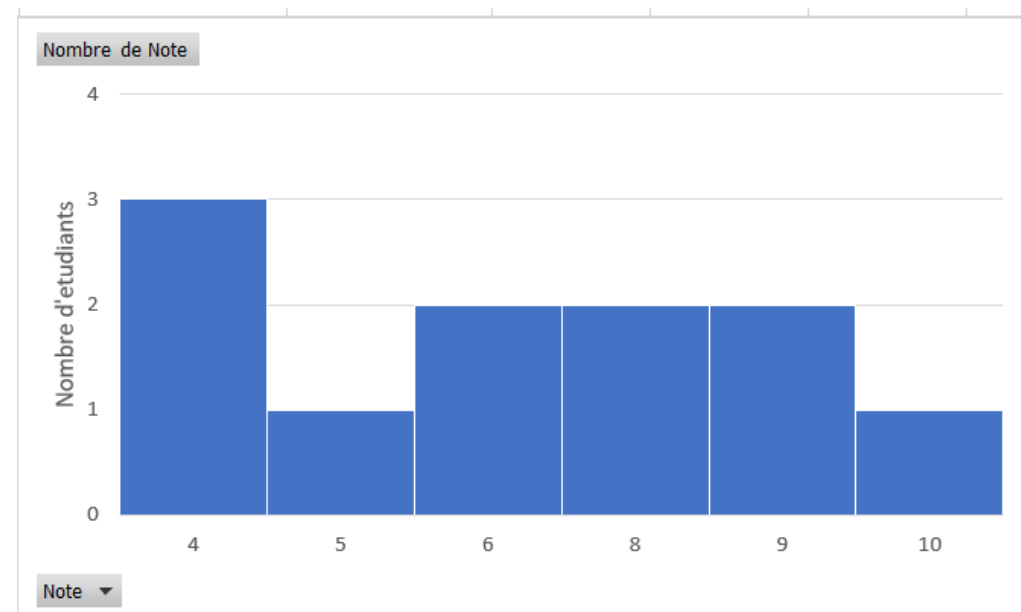
- la valeur (donnée) de la série statistique ayant la fréquence la plus élevée (la valeur de la série qui revient le plus souvent)
- il n'y a pas de formule mathématique pour le calculer
- une série statistique de données peut posséder plusieurs valeurs modales.

# MODE: définition et calcul

Les notes obtenues à l'examen pratique de Biostat par un groupe de 11 étudiants:

④, 9, 5, 8, 6, ④, 9, 10, 8, 6, 5, ④

- Mode: 4 → série unimodale



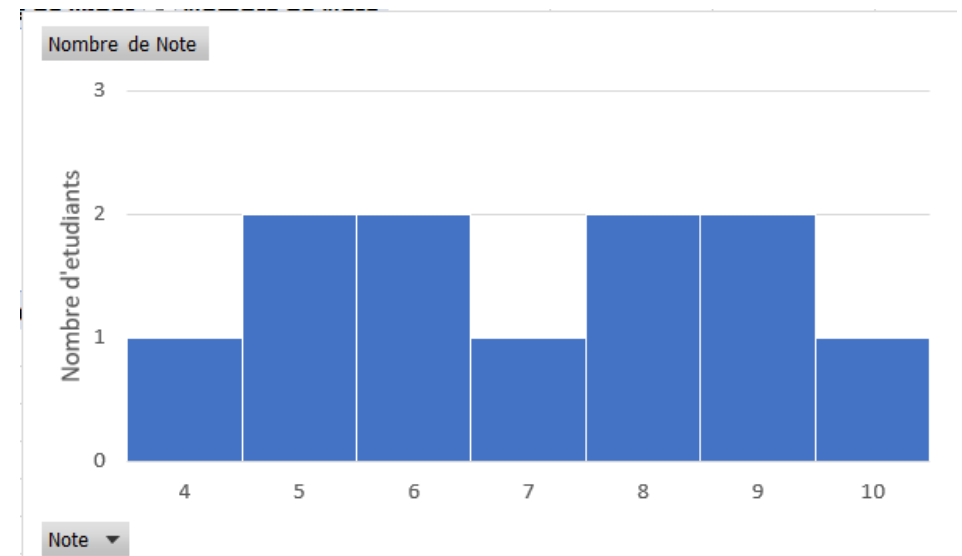


# MODE: exemple

Les notes obtenues à l'examen pratique de Biostat par un groupe de 11 étudiants:

4, 9, 5, 8, 6, 7, 9, 10, 8, 6, 5

- Mode: 5, 6, 8, 9
- → série multimodale (plurimodale)



# MODE (Mo): propriétés

## Avantages

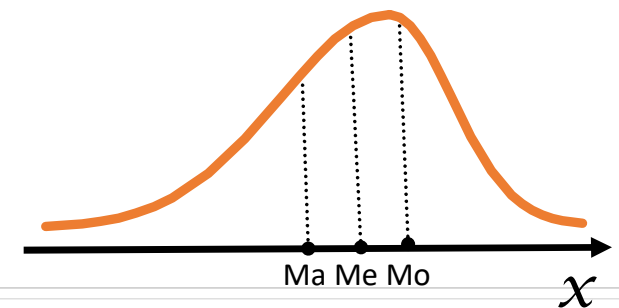
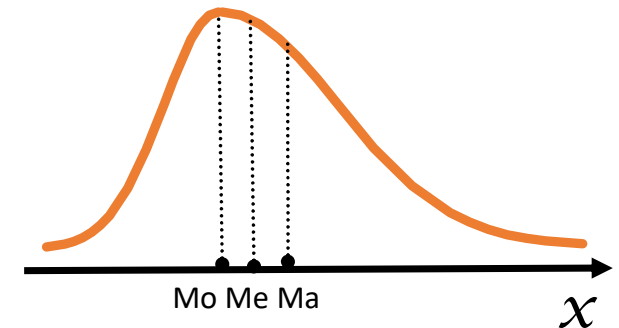
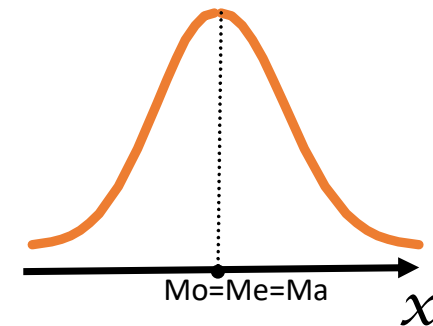
- ✓ faible sensibilité aux valeurs extrêmes de la série.
- ✓ peut être un indicateur d'une série de données hétérogène
- ✓ si les données sont hétérogènes (série bimodale), il vaut mieux deux valeurs modales qu'une médiane

## Désavantages

- ✗ ne se prête pas aux calculs (Transformer l'échelle de mesure de la série de données statistiques:  $X'' = C \cdot X$ ,  $C = \text{constante}$ )

# Positions relatives de la moyenne, médiane et mode

- Distribution **symétrique**:  
Mode  $\approx$  Moyenne  $\approx$  Médiane
- Distribution **asymétrique (étalée) à droite**:  
Mode  $<$  Médiane  $<$  Moyenne
- Distribution **asymétrique à gauche**:  
Mode  $>$  Médiane  $>$  Moyenne



# Autres mesures de la tendance centrale

- **Autres mesures de la tendance centrale** (on doit pas les apprendre)

- **La moyenne pondérée:**  $m_x = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$

- ex. Le calcul de la note finale dans la faculté en fonction des nombre de crédits

- **La Moyenne Geometrique**

$$\sqrt[n]{x_1 x_2 \dots x_n}$$

- **La moyenne armonique**

$$\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

- **Valeur Centrale**

$$\frac{x_{\max} + x_{\min}}{2}$$

# Autres mesures de la tendance centrale

## Moyenne pondérée

### • Moyenne pondérée

$$m_x = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

	Crédits	10	7	6	4	3		
N.	Nom, surnom (initiales)	Anatomie	Biologie cellulaire	Biochimie	Biophysique	Informatique médicale	Moyenne arithmétique	Moyenne pondérée
1	B R	7	7	7	8	8	7.40	7.23
2	B A	8	8	8	9	9	8.40	8.23
3	B S	9	9	7	7	10	8.40	8.43
4	B R	10	6	8	8	8	8.00	8.20
5	B A	8	7	9	9	9	8.40	8.20
6	C D	9	8	6	9	6	7.60	7.87
7	C E	5	5	7	10	7	6.80	6.27
8	C R	6	10	5	10	5	7.20	7.17
9	C S	5	5	6	5	7	5.60	5.40
10	C T	6	10	10	6	9	8.20	8.03

### Moyenne arithmétique

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

- ✓ La moyenne arithmétique est un cas particulier de la moyenne pondérée ( $w_i$ , poids égaux)

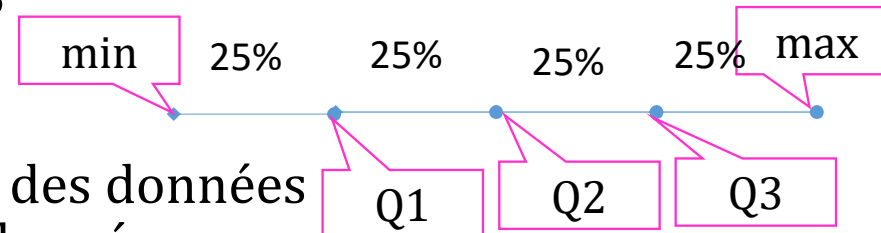
# MESURES DE POSITION: quantiles et percentiles

- **Les quantiles** ( $Q_1, Q_2, \dots, Q_{q-1}$ ): valeurs remarquables qui partagent la série de données ordonnées en  $q$  sous-ensembles (groupes) consécutifs égaux.
- **le quantile d'ordre  $\alpha$  ( $0 < \alpha < 1$ )** est la valeur  $x_\alpha$  telle qu'une proportion  $\alpha$  (%) des données soit plus petite que  $x_\alpha$
- **les plus utilisées quantiles: les quartiles, les quintiles, les deciles, les centiles**
- **le  $p^{\text{ième}}$  percentile:**
  - divise la série en deux sous-ensembles (tel qu'au plus  $p\%$  des valeurs sont en-dessous d'elle et au plus  $(100-p)\%$  sont au-dessus)
  - est le quantile d'ordre  $\alpha = p/100$

# MESURES DE POSITION: quantiles & percentiles

• **Les quartiles ( $Q_1, Q_2, Q_3$ ):** divisent la série de données en quatre groupes ayant la même proportion des données .

- le 1<sup>er</sup> quartile ( $Q_1$ ) sépare les 25% inférieurs des données
  - 25% des valeurs sont  $\leq Q_1$ , 75% sont  $\geq Q_1$
- le 2<sup>e</sup> quartile est la médiane de la série (50%)
  - 50% des valeurs sont  $\leq Q_2$  / médiane, 50% sont  $\geq Q_2$  / médiane
- le 3<sup>e</sup> quartile sépare les 75% inférieurs des données
  - 75% des valeurs sont  $\leq Q_3$ , 25% sont  $\geq Q_3$



• **Les quintiles ( $V_1, V_2, V_3, V_4$ ):**

- le 1<sup>er</sup> quintile ( $V_1$ ) sépare les 20% inférieurs des données
- le 2<sup>e</sup> quintile sépare les 40% inférieurs des données
- le 3<sup>e</sup> quintile sépare les 60% inférieurs des données
- le 4<sup>e</sup> quintile sépare les 80% inférieurs des données

# MESURES DE POSITION: déciles & centiles

- **Les déciles ( $D_1, \dots, D_9$ ):**

- le 1<sup>er</sup> décile sépare les 10% inférieurs des données
- le 2<sup>e</sup> décile sépare les 20% inférieurs des données
- ...
- le 9<sup>e</sup> décile sépare les 90% inférieurs des données

- **Les centiles ( $C_1, \dots, C_{99}$ ): permettent de scinder la série de données en 100 sous-ensembles égaux)**

- le 1<sup>er</sup> centile ( $C_1$ ) sépare le 1% inférieurs des données
- ....
- le 99<sup>e</sup> centile ( $C_{99}$ ) sépare le 99% inférieurs des données



# MESURES DE DISPERSION: l'écart/l'intervalle interquartile

- **Intervalle interquartile IQR: [Q1, Q3]**

- C'est l'intervalle entre le première quartile (Q1) et le 3<sup>ème</sup> quartile (Q3)
- D'habitude il est mis après la médiane dans les résultats d'articles scientifiques
  - Ex: Cholestérol mg/dl (médiane, IQR): 189 [162; 211]

- **Ecart interquartile EQR: Q3-Q1**

- Parfois certain chercheurs montre seulement la différence entre la quartile 3 et la quartile 1 (étendu ou écart inter quartile  $EQR=Q3-Q1$ ) au lieu des deux valeurs
  - Ex: Cholestérol mg/dl (médiane, EQR): 189 (49)
- La meilleure représentation est si on montre les quartiles (Q1, Q3), pas seulement le EQR

# MESURES DE POSITION: exemple

Variable:  
Note de l'  
ex.pratique

X <sub>1</sub>	9
X <sub>2</sub>	6
X <sub>3</sub>	4
X <sub>4</sub>	9
X <sub>5</sub>	4
X <sub>6</sub>	8
X <sub>7</sub>	8
X <sub>8</sub>	9
X <sub>9</sub>	7
X <sub>10</sub>	4
X <sub>11</sub>	10
X <sub>12</sub>	10

Trier

	A	B
1	X <sub>10</sub>	4
2	X <sub>3</sub>	4
3	X <sub>5</sub>	4
4	X <sub>2</sub>	6
5	X <sub>9</sub>	7
6	X <sub>6</sub>	8
7	X <sub>7</sub>	8
8	X <sub>1</sub>	9
9	X <sub>4</sub>	9
10	X <sub>8</sub>	9
11	X <sub>11</sub>	10
12	X <sub>12</sub>	10

Me=8

IQR=[Q<sub>1</sub>, Q<sub>3</sub>] = [5.5, 9]

$$Me = [X_{12/2} + X_{(12/2+1)}] / 2 = (X_6 + X_7) / 2 = (8 + 8) / 2 = 8$$

Sur Excel:

(Mediane) Me:

=MEDIAN(B1:B12)

• Intervalle interquartile IQR: [Q<sub>1</sub>, Q<sub>3</sub>]

Q<sub>3</sub>=QUARTILE(B1:B12,3)

Q<sub>1</sub>=QUARTILE(B1:B12,1)

# Mesures descriptives utilisés dans un article SCIENTIFIQUE

**Objectifs de l'étude** (lien vers l'étude: <https://pubmed.ncbi.nlm.nih.gov/33271421/>)

- 1) décrire les caractéristiques de base de patients COVID-19 atteints d'embolie pulmonaire et
- 2) identifier les facteurs de risque associés à l'embolie pulmonaire dans le contexte de la COVID-19

Serum biomarkers			
Troponin hs, n-99th percentile; peak	2.2 (0.2–17.5)	28 (8–180)	2 (0.2–9)
Median (IQR)			
Acute cardiac injury	69 (31)	15 (47)	54 (28)
D-dimer, peak; ng/ml	625	1819	555
Median (IQR)	(90–2050)	(568–5017)	(13–1530)
Pro-BNP, peak; pg/ml	4616 ± 7800	4076 ± 3402	4765 ± 3654

Baseline characteristics of the study population.

Variables	Total (n = 224)	Pulmonary embolism (n = 32)	No pulmonary embolism (n = 192)
<b>Demographics</b>			
Age, years	69 ± 14	67 ± 15	69 ± 13
Male	127 (62)	20 (62)	120 (62)
<b>Symptoms</b>			
Fever	153 (68)	25 (78)	128 (67)
Dyspnoea	158 (70)	24 (75)	134 (70)
Chest discomfort	69 (31)	10 (31)	59 (31)
Cough	85 (38)	10 (31)	75 (39)
Syncope	21 (9)	4 (12)	17 (9)
Symptoms onset to hospitalization, days	6 (2–10)	7 (3–11)	3 (1–6)
Median (IQR)			
Late presentation	99 (44)	22 (69)	77 (40)

Note. Categorical data are presented as numbers (%). Continuous data are presented as *mean ± SD or median (IQR), as appropriate*.

Source: Scudiero F, Silverio A, Di Maio M, Russo V, Citro R, Personeni D, Cafro A, D'Andrea A, Attena E, Pezzullo S, Canonico ME, Galasso G, Pitì A, Parodi G; Cov-IT Network. Pulmonary embolism in COVID-19 patients: prevalence, predictors and clinical outcome. Thromb Res. 2021 Feb;198:34-39. doi: 10.1016/j.thromres.2020.11.017.

19/10/2022

# MESURES DESCRIPTIVES: CAS D'UNE VARIABLE QUALITATIVE

- Le rapport
- La proportion
- Le taux

## Le rapport

- Est une expression de la forme:  $x : y$
- $x$  et  $y$  représentent un nombre d'événements
  - à un moment donné
  - ou dans un intervalle de temps donné
- Exemple: Deux personnes sur 10 consultées par jour par un médecin de famille ont une tension artérielle systolique (TAS) supérieure à la normale.
  - Quel est le rapport TAS normale / TAS pathologique?
    - TAS normale / TAS pathologique =  $8/2 = 4$  (un sujet avec des valeurs pathologiques de TAS → 4 individus avec des valeurs normales de TAS)

# MESURES DESCRIPTIVES: CAS D'UNE VARIABLE QUALITATIVE

- **La proportion:** Est une expression de la forme:  $(x / (x + y)) * 100$
- Est un rapport dans lequel le numérateur est contenu dans le dénominateur
- Il s'exprime d'habitude en pourcentage
- Ex.

- dans une étude sur 200 sujets 50 ont développé une maladie

La proportion des sujets malades est:

$$(50 / 200) * 100 = 25\%$$

20% des cancers sont dus au tabagisme

70% des cancers du poumon sont dus au tabagisme

Le nombre des nouveaux cancers va s'agrandir avec 70% dans les prochaines 20 années

World Health Organisation - <http://www.who.int/>

# MESURES DESCRIPTIVES: CAS D'UNE VARIABLE QUALITATIVE

## Le TAUX

- est une expression de la forme:  $x / y * k$
- montre le risque d'apparition d'un événement dans un temps donné
- $$\text{Taux} = \frac{\text{nombre d'événements dans une période spécifique}}{\text{la population au risque pour ces événements dans la même période}} \times \text{base}$$
- **Le taux** doit
- s'exprimer en fonction d'une unité du temps
- être calculé
- pour une aire géographique (place)
- pour un groupe de sujets bien défini

## MESURES DESCRIPTIVES: CAS D'UNE VARIABLE QUALITATIVE

- **Le TAUX:**
- Le numérateur  $x$  représente le numéro d'événements dans une période du temps bien défini (décès, maladies, ...)
- Le dénominateur  $y$  représente la taille de la population au risque dans une période du temps, dans un lieu
- La constante  $k$  (la base / le multiplicateur) est une puissance de 10, qui dépend de la taille de  $x$  et  $y$
- Ex. pour un taux de 0,00024 apparitions d'une maladie par an dans une population définie, on peut choisir  $k = 10\,000$ , donc le taux est 2,4 malades pour 10 000 sujets par année.

# Le TAUX:

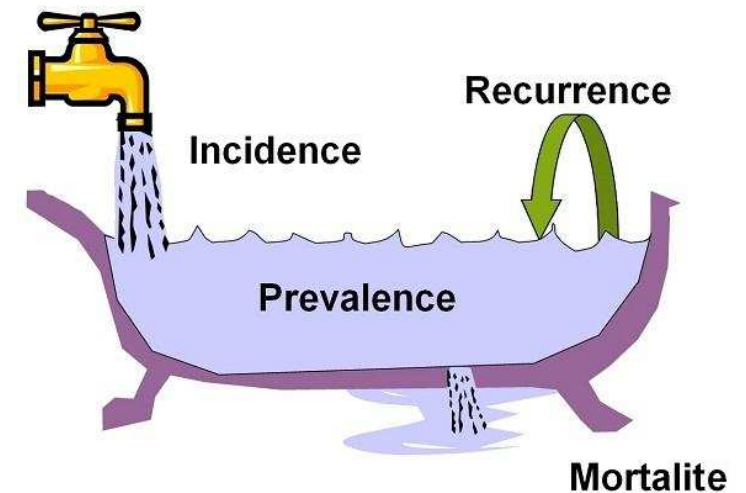
- Les plus utilisées taux sont:
  - de morbidité
  - de mortalité
  - de natalité
- Les taux de **morbidité**
  - L'évènement ici est l'apparition/la présence d'un **maladie**
  - Ils mesurent la fréquence des maladies dans une population dans une période du temps, dans un lieu
  - Les plus utilisées sont:
    - prévalence instantanée (de moment)
    - prévalence de période
    - le taux de l'incidence
    - le taux de l'attaque



# Le taux de morbidité: la prevalence

## La **prévalence**

- mesure la fréquence d'une maladie **toutes** les cas (les nouveaux et anciennes cas),
  - a un moment donné - instantanée (de moment)
  - dans une intervalle du temps – de période
- dans un population, dans un place
- (en réalité elle est un proportion)
- Ex:
  - la prévalence de l'asthme chez les enfants variait de moins de 5% à plus de 20%\* (diffèrent pays)



<https://www.studyblue.com/notes/note/n/epidemiology-lecture-1/deck/7404763>

\* WHO [https://www.euro.who.int/data/assets/pdf\\_file/0012/96996/3.1.pdf](https://www.euro.who.int/data/assets/pdf_file/0012/96996/3.1.pdf)

# Le taux de morbidité

## L'incidence

- mesure la fréquence des **nouveaux cas** d'une maladie, dans une intervalle du temps, dans un population, dans un place
- Les nouveaux cas d'une maladie sont:
  - des nouveaux cas réels dans la population initiale
  - la migration des cas déjà malades
- ex. diabète 4,3 pour 1000 sujets de 18-44 ans dans 2017-2018 dans USA  
<https://www.cdc.gov/diabetes/data/statistics-report/newly-diagnosed-diabetes.html>

# MÉTHODES DE PRÉSENTATION DES DONNÉES

- Tableaux:
  - une variable
  - deux ou plus des variables
- Diagrammes
  - une variable
  - deux/trois variables

# TABLEAUX

- Tableau associé a **une seule variable**: **TABLEAU DE FREQUENCES**
  - qualitative / quantitative discrète;
  - quantitative continue (regroupements en classes ou intervalles des valeurs);
- Tableau associé aux **deux variables** qualitatives : **TABLEAU DE CONTINGENCE**  
÷ permet de voir un relation entre les deux variables
- Tableau des **indicateurs/mesures (statistiques) descriptives**
  - moyenne, déviation standard
  - médiane, quartile 1 (Q1), quartile 3 (Q3), intervalle interquartile (IQR)
  - permet de décrire une seule variable ou de comparer des groups

# COMMENT PEUT EN OBTENIR LE TABLEAU DE FRÉQUENCES ?

- On regroupe les données de la série statistique
  - par classes dans un tableau (voir la table pour signification de la classe)
    - indiquant la répartition des individus selon le caractère (la variable) étudié ;

<b>Variable qualitative ou quantitative discrète</b>	<b>Variable quantitative</b>
Classe est une catégorie/ valeur de la variable	Classe devient intervalles de classe ou classe de valeurs

## TABLEAU DE FRÉQUENCES : cas d'une variable qualitative nominale

On montre les fréquences absolues et relatives.

Fréquences  
absolues

Fréquences  
relatives

Causes de décès	Nombre de patients	Pourcentage (%)
L'asphyxie à la naissance	527	26.30
Blessures obstétricales	92	4.59
Pneumonie	181	9.03
Malformations congénitales	598	29.84
Autres causes	606	30.24
<b>Total</b>	<b>2004</b>	<b>100.00</b>

Table 1. La distribution des causes de décès chez les nouveaux nés

## TABLEAU DE FRÉQUENCES : cas d'une variable qualitative ordinale

Que faut-il indiquer pour chaque classe ?	Définitions
fréquence <b>absolue</b> ( $n_i$ )	<b>nombre</b> d'individus de la classe
fréquence <b>relative</b> ( $f_i$ )	<b>proportion</b> de sujets de la population ou de l'échantillon appartenant a une certaine classe ;
fréquence <b>absolue cumulée croissante</b> de la classe ( $x_i$ )	<b>effectif</b> d'individus dans la population (ou l'échantillon) pour lesquelles le caractère étudié peut prendre une valeur inférieur ou égale a $x_i$
fréquence <b>absolue cumulée décroissante</b> de la classe ( $x_i$ )	<b>effectif</b> d'individus dans la population (ou l'échantillon) pour lesquelles le caractère étudié peut prendre une valeur supérieur ou égale a $x_i$
fréquence <b>relative cumulée croissante</b> de la classe ( $x_i$ )	<b>proportion</b> d'individus dans la population (ou l'échantillon) pour lesquelles le caractère étudié peut prendre une valeur inférieur ou égale a $x_i$
fréquence <b>relative cumulée décroissante</b> de la classe ( $x_i$ )	<b>proportion</b> d'individus dans la population (ou l'échantillon) pour lesquelles le caractère étudié peut prendre une valeur supérieur ou égale a $x_i$

Table 2. Définition de la fréquence absolue, fréquence relative, fréquence absolue cumulée croissante, fréquence relative cumulée croissante

## TABLEAU DE FRÉQUENCES : cas d'une variable qualitative **ordinaire**

- On montre les fréquences absolues et relatives et aussi les fréquences cumulées:

X (Variable )	Fréquence absolue ( $n_i$ )	Fréquence relative (%) ( $f_i$ )	Fréquence absolue cumulée croissante ( $N_i$ )	Fréquence relative cumulée croissante ( $F_i$ ) (%)
$x_1$	$n_1$	$f_1$	$N_1 = n_1$	$F_1 = f_1$
$x_2$	$n_2$	$f_2$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
...	...	...	...	...
$x_k$	$n_k$	$f_k$	$N_k = n$	$F_k = 100$
Total	$n$	100		

Excel:

✓ **COUNTIF**  
✓ **DATA-PIVOT TABLE**

Table 3. Définition de la fréquence absolue, fréquence relative, fréquence absolue cumulée croissante, fréquence relative cumulée croissante



## TABLEAU DE FRÉQUENCES : **cas d'une variable qualitative ordinale**

Consommation d'alcool	Fréquences absolues	Fréquences relatives (%)	Fréquences absolues cumulées croissante	Fréquences relatives cumulées croissante (%)
nulle	53	6,74	53	6,74
faible	178	22,65	231	29,39
modérée	450	57,25	681	86,64
importante	84	10,69	765	97,33
excessive	21	2,67	786	100,00
<b>Total</b>	786	100,00		

Table 4. La distribution des types de consommation d'alcool dans un échantillon des sujets hypertendus

## TABLEAU DE CONTINGENCE : CAS DE DEUX VARIABLES QUALITATIVES

		Diabète		Total
		Oui	Non	
Obésité	Oui	24	54	<b>78</b>
	Non	2	10	<b>12</b>
	<b>Total</b>	<b>26</b>	<b>64</b>	<b>90</b>

- Montre la relation entre deux variables (ici Obésité et Diabète)
- Montre si une caractéristique a une distribution égale ou non entre les catégories d'un autre variable
- Plus des diabétiques chez les obèses que chez les non obèses

		Diabète		Total %
		Oui %	Non %	
Obésité	Oui	30,77	69,23	<b>100,0</b>
	Non	16,67	83,33	<b>100,0</b>
	<b>Total</b>	<b>28,89</b>	<b>71,11</b>	<b>100,0</b>

Excel:

✓ **DATA-PIVOT TABLE**

R (R Commander):

✓ **Statistics/Contingency  
Tables/Two-way table...**

EpiInfo:

✓ **STATISTICS/TABLES**

## TABLEAU DE FRÉQUENCES: **cas d'une SEULE variable quantitative** (regroupements en classes des valeurs)

Classes du poids des nouveau nées (g)	Fréquences absolues	Fréquences relatives (%)	Fréquences relatives cumulées croissante (%)
<= 2500	14	2.31	2.31
(2500, 2700]	31	5.12	7.43
(2700, 2900]	76	12.54	19.97
(2900, 3200]	170	28.05	48.02
(3200, 3400]	97	16.01	64.03
(3400, 3600]	94	15.51	79.54
(3600, 3800]	62	10.23	89.77
(3800, 4000]	31	5.12	94.88
> 4000	31	5.12	100.00
Total	606	100.00	

Table 6. La distribution du poids chez les nouveaux nées

# TABLEAU DE FRÉQUENCES: **cas d'une SEULE variable quantitative** (regroupements en classes des valeurs)

- ✓ il n'existe pas de règle commune acceptée pour choisir le nombre de classes
- ✓ il existe quelques formules empiriques pour déterminer le nombre de classes ( $n_c$ ) :

$$n_c = \lfloor \sqrt{n} \rfloor$$

$$n_c = \lfloor 1 + 3,3 \lg n \rfloor$$

- ✓ en général, on va travailler avec des classes de même amplitude
- ✓ on utilise ce que les logiciels nous offrent d'habitude

X (Variable )	Fréquence absolue ( $n_i$ )	Fréquence relative (%) ( $f_i$ )	Fréquence absolue cumulée croissante ( $N_i$ )	Fréquence relative cumulée croissante ( $F_i$ ) (%)
$[a_1, a_2]$	$n_1$	$F_1$	$N_1 = n_1$	$F_1 = f_1$
$(a_2, a_3]$	$n_2$	$f_2$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
...	...	...	...	...
$(a_k, a_{k+1}]$	$n_k$	$f_k$	$N_k = n$	$F_k = 100$
Total	$n$	100		

EXCEL:

✓ **TOOLS-DATA ANALYSIS-HISTOGRAMME**

# Description des données **avec des graphiques**

Variable(s) qualitative(s)	Variable (s) quantitative (s)
<b>Pour décrire une SEULE variable qualitative</b> <ul style="list-style-type: none"> <li>- La diagramme à secteurs (camembert/pie)</li> <li>- La diagramme en barres colonnes)</li> </ul>	<b>Pour décrire une SEULE variable quantitative</b> <ul style="list-style-type: none"> <li>- L'histogramme</li> <li>- Les graphiques des indicateurs: <ul style="list-style-type: none"> <li>- La graphique boîte à moustaches (le box-plot)</li> <li>- Le graphique du moyenne et déviation standard</li> <li>- Le graphique des quantiles</li> <li>- Le graphique linéaire</li> </ul> </li> </ul>
<b>Relation entre DEUX variables qualitatives:</b> <ul style="list-style-type: none"> <li>- La diagramme en barres/colonnes</li> </ul>	<b>Relation entre DEUX variables quantitatives:</b> <ul style="list-style-type: none"> <li>- Le graphique nouage des points</li> </ul>

# Principes pour la bonne réalisation des tableaux

- le **titre** du tableau doit être concis, précis et clair;
- chaque ligne et chaque colonne du tableau doit être
  - concis et clairement **étiquetée**;
- le **total** doit être mentionné dans chaque tableau;
- les **codes** et les symboles doivent être expliqués dans une note;

# Principes pour la bonne réalisation des graphiques

Toute représentation graphique doit avoir:

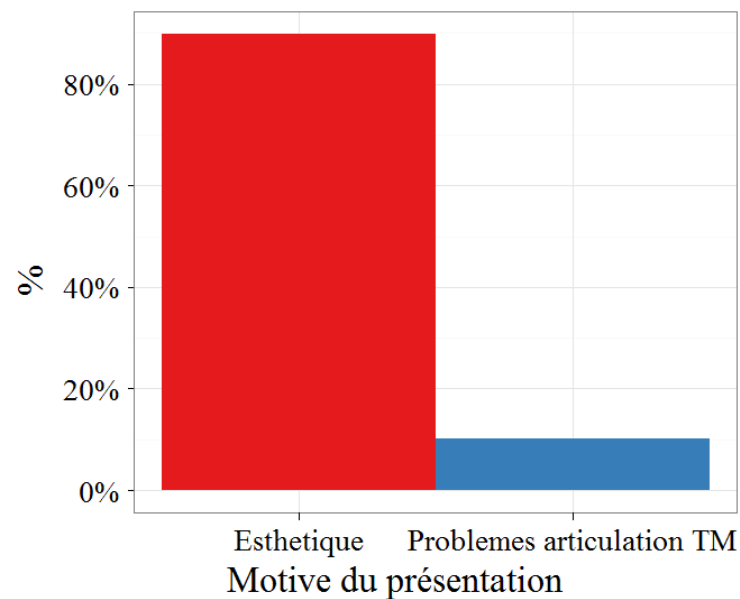
- ✓ titre clair, concis et précis
- ✓ définitions des axes, sans abréviations (à l'exception des unités de mesure)
- ✓ unités de mesure
- ✓ légende (s'il faut)
- ✓ tous les symboles, des abréviations ou lettres utilisées dans la figure doivent être expliqués clairement dans la légende

# GRAPHIQUES: cas d'une variable QUALITATIVE

- Représentent des aires proportionnelles aux fréquences (absolues/relatives) des catégories (classes, modalités) de la variable statistique
- Types
  - Diagrammes camembert - à secteurs circulaires
    - (difficile pour le cerveau)
  - Diagrammes en colonnes/barres – la meilleure
- Conseils:
  - 2D vaut mieux que 3D,
  - Pourcentages sont meilleures que les valeurs absolues pour comprendre les relations
    - mais les valeurs absolues sont très importantes aussi



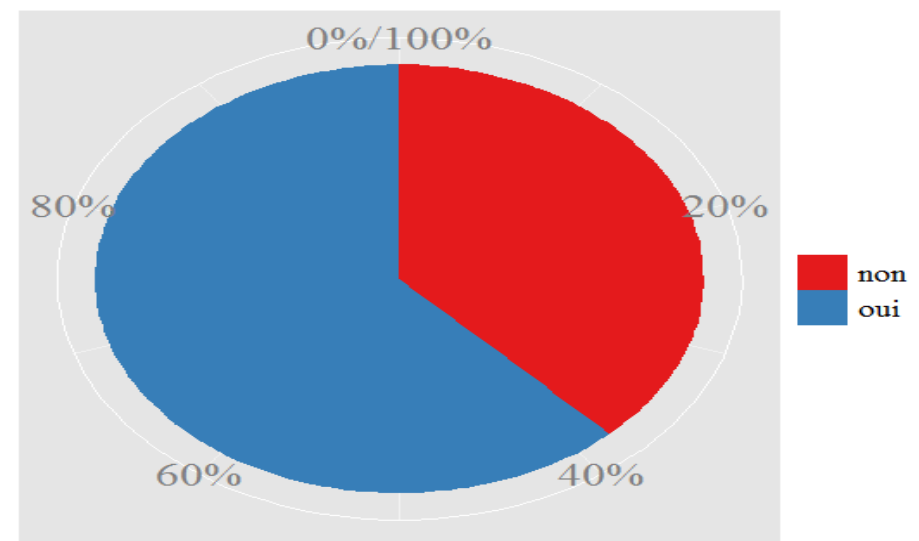
## Cas d'une variable qualitative : Diagramme en barres et camembert/sectorielle



Motive du présentation pour  
l'intervention chirurgicale maxillo-faciale

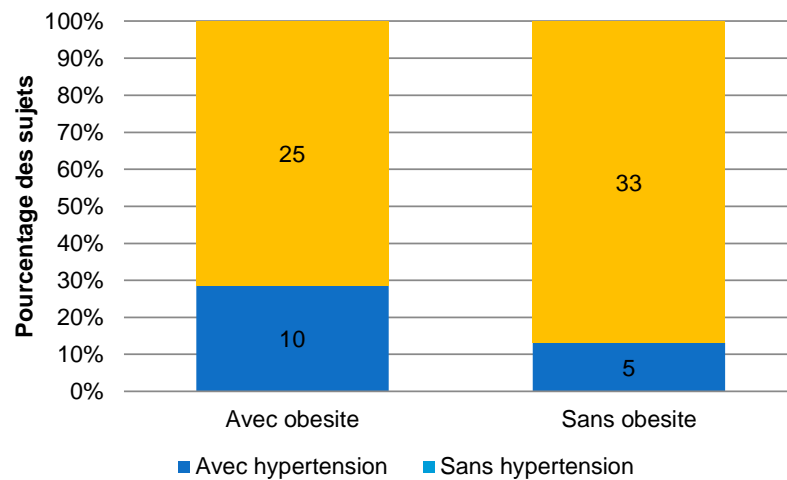
Excel: **INSERT/GRAPH/BAR ou PIE**

Traitement orthodontique fixe, avant l'intervention  
chirurgicale

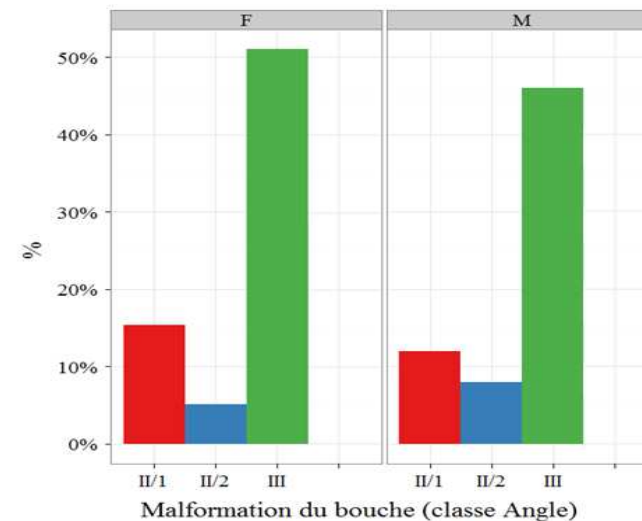


# GRAPHIQUES: La relation entre deux variables qualitatives

L'hypertension en fonction de l'obésité



La malformation du buche en fonction du sexe



Excel:

✓ **DATA-PIVOT TABLE + INSERT CHART: type: COLUMN - 100% STACKED COLUMN**

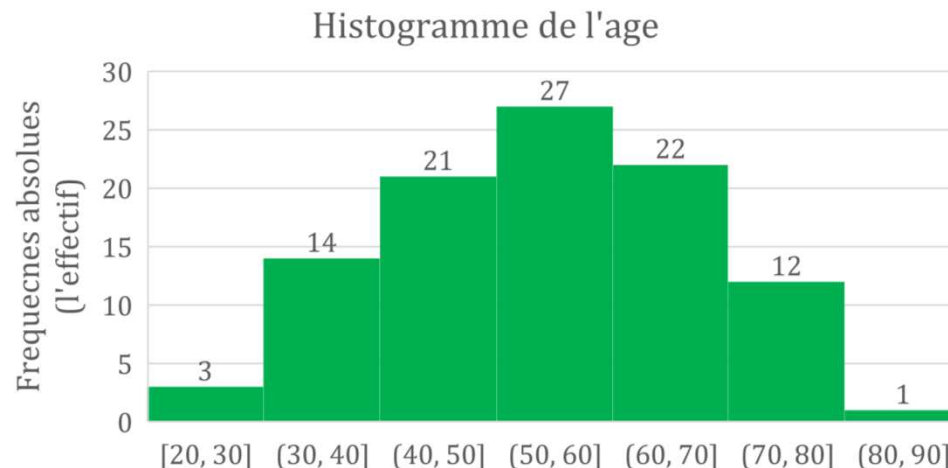
R (R Commander):

✓ **KMggplot2/Bar chart ... avec stratum variable**

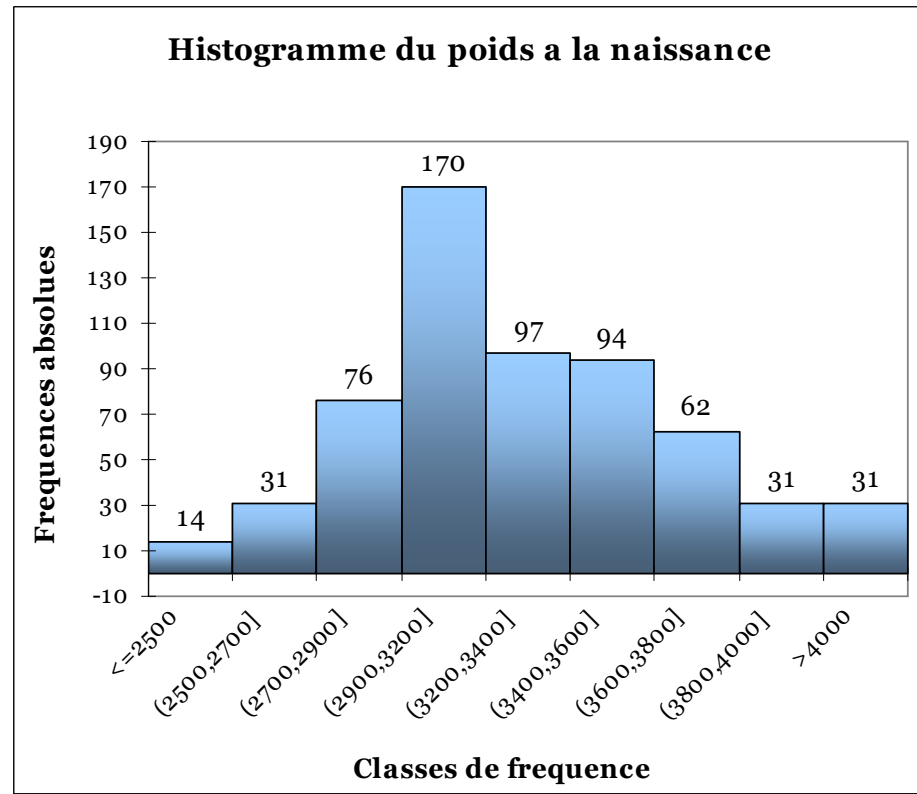
# Cas d'une variable quantitative continue: **HISTOGRAMME**

- **Histogramme** = rectangles juxtaposés
  - chacune des bases est égale à l'intervalle de chaque classe
  - la hauteur est la fréquence absolue/relative
  - l'aire de chaque rectangle est proportionnelle
    - aux effectifs ou aux fréquences de la classe correspondante;

	A	B
1	Id_patient	Age (ans)
2	1	59
3	2	57
4	3	63
5	4	48
6	5	70
7	6	72
8	7	39
9	8	45
10	9	64
11	10	34
12	11	55
13	12	53
14	13	63
15	14	55
16	15	71
17	16	68
18	17	48
19	18	37
20	19	54
21	20	56
22	21	52
23	22	68
24	23	68
25	24	87
26	25	34
27	26	48



# Cas d'une variable quantitative continue: l'histogramme



Excel:

- ✓ **INSERT/CHART/HISTOGRAM**
- ✓ **TOOLS/DATA ANALYSIS/HISTOGRAM**

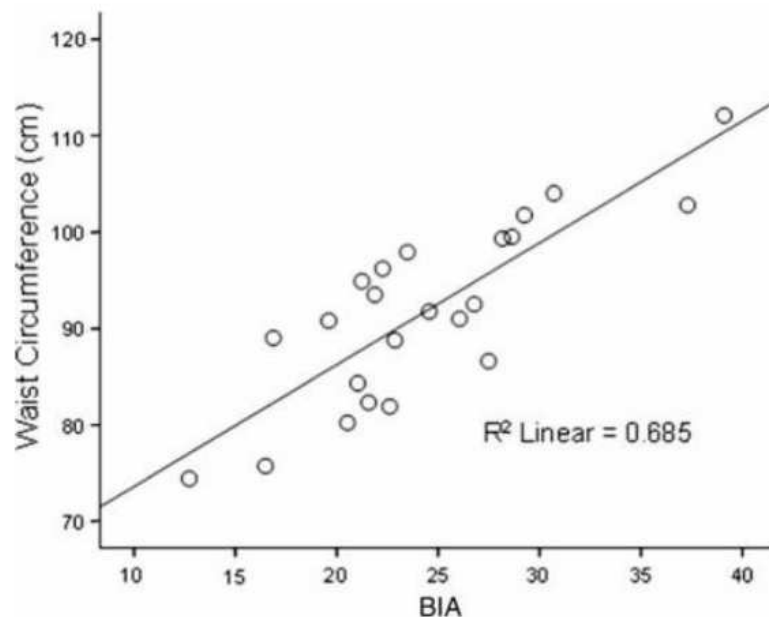
EpiInfo:

- ✓ **STATISTICS/GRAPH/HISTOGRAM**

R (R Commander):

- ✓ **KMggplot2/Histogram...**
- ✓ **ou Graphs/Histogram...**

# Graphique par nouage de points (engl. Scatter plot): relation entre deux **variables quantitatives**



**Figure 2** Scatter plot of BIA versus waist circumference. A regression line is drawn. The coefficient of determination ( $r^2$ ) was 0.685.

Source:

Eriks-Hoogland I, Hilfiker R, Baumberger M, Balk S, Stucki G, Perret C. Clinical assessment of obesity in persons with spinal cord injury: validity of waist circumference, body mass index, and anthropometric index. J Spinal Cord Med.

2011;34(4):416-22. doi: 10.1179/2045772311Y.0000000014.

Lien vers l'article:

<https://pubmed.ncbi.nlm.nih.gov/21903016/>

BIA = mesure la composition corporelle en fonction de la vitesse à laquelle un courant électrique se déplace à travers le corps

Waist circumference = Circonférence abdominale (cm)

# Exemple de questions

*Nombre de réponses correctes par question: 1-4.*

## **E1. Pour une série statistique de données quantitatives:**

- A. la moyenne est une mesure de centralité
- B. la somme des écarts à la moyenne est nulle
- C. la médiane est une mesure de l'asymétrie
- D. les trois quartiles sont des indicateurs de tendance centrale
- E. les trois quartiles sont des indicateurs de localisation/position

**R1: A, B, E**

### *Solution de l'exercice:*

- > la Moyenne est une mesure de tendance centrale qui a la propriété que la somme des écarts entre les valeurs de la série de données et la Moyenne est égale à 0 (voir diapo 8).
- > les trois quartiles (Q1, Q2, Q3) sont des mesures descriptives de localisation (de position) parce que ils nous donnent une idée de la localisation des données brutes

# Exemple de questions

**E2. Les valeurs de pression artérielle systolique (exprimées en mm Hg) dans un échantillon de 10 patients sont:**

**120, 100, 110, 120, 130, 160, 130, 120, 140, 160. La moyenne arithmétique, la médiane, le mode et la valeur centrale sont:**

- A. 129 – 125 – 120 – 130
- B. 130 – 125 – 130 – 125
- C. 120 – 130 – 120 – 125
- D. 129 – 130 – 120 – 130
- E. 125 – 125 – 120 – 130

**R2: A**

*Solution de l'exercice -> Moyenne arithmétique (Ma) =?*

$$Ma = (120 + 100 + 110 + 120 + 130 + 160 + 130 + 120 + 140 + 160) / 10 = 129 \text{ mmHg}$$

*-> Médiane (Me) =? Série statistique ordonnée: 100, 110, 120, 120, 120, 130, 130, 140, 160, 160*

$$\text{Taille de l'échantillon: nombre paire} = 10 \Rightarrow \frac{\frac{X_{10} + X_{10}}{2} + 1}{2} = \frac{X_5 + X_6}{2} = 125$$

*-> Mode (Mo) =? la valeur (donnée) de la série statistique ayant la fréquence la plus élevée (la valeur de la série qui revient le plus souvent) = 120*

$$\text{-> Valeur centrale} = ? \frac{x_{\max} + x_{\min}}{2} = 130$$

# *Exemple de questions*

**E3. Les valeurs de pression artérielle systolique (exprimées en mm Hg) dans un échantillon de 10 patients sont: 120, 100, 110, 120, 130, 160, 130, 120, 140, 160.**

**Lesquels des énoncés suivants sont corrects:**

- A. Moyenne arithmétique = médiane = valeur modale
- B. valeur modale < médiane < moyenne arithmétique
- C. valeur modale > médiane > moyenne arithmétique
- D. La série est asymétrique à droite (négative)
- E. La série est asymétrique à gauche (positive)

**R3: B,D**



# Exemple de questions

**E4.** Les valeurs de pression artérielle systolique (exprimées en mm Hg) dans un échantillon de 10 patients sont: 120, 100, 110, 120, 130, 160, 130, 120, 140, 160. Il est connu que la valeur du premier quartile (Q1) est égale a 120, la deuxième quartile (Q2) est égal à 125 et le troisième quartile (Q3) est égal à 137.5.

**Lesquels des affirmations suivants sont corrects:**

- A.  $Q2 - Q1 = 5$
- B.  $Q3 - Q2 = 12$
- C.  $Q3 - Q2 = 12,5$
- D. la distribution est approximativement symétrique
- E. la distribution est asymétrique

**R4: A, C, E**

*Solution de l'exercice: -> Dans une distribution symétrique, les quartiles sont situés à égale distance de chaque côté de la médiane =>  $Q2 - Q1 = Q3 - Q2$  ; dans notre cas, les deux termes ne sont pas égaux alors on dit que la série de données est asymétrique.*

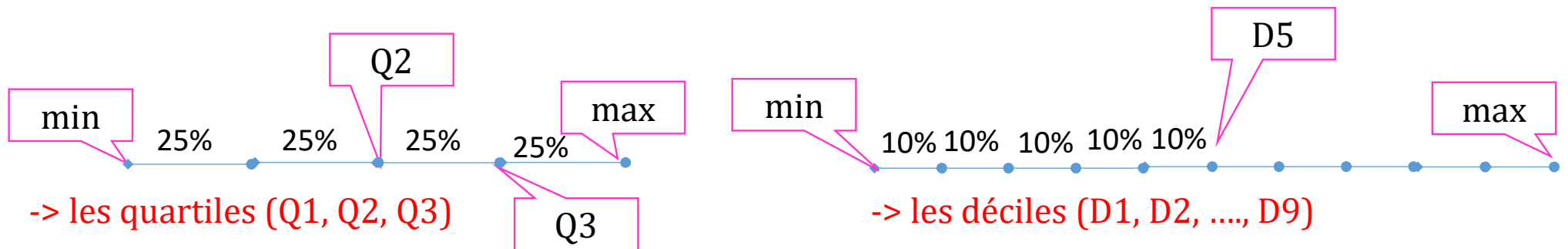
# Exemple de questions

**E5. Pour une série statistique de données d'une variable continue ou ordinale:**

- A. on peut calculer quatre quartiles
- B. l'intervalle interquartile (IQR) contient 50% des observations
- C. le cinquième décile est égal à la médiane
- D. 45% des observations sont supérieures au premier quartile
- E. 75% des observations sont supérieures au premier quartile

**R5: B, C, E**

*Solution de l'exercice:*



# *Exemple de questions*

**E6. Pour une variable statistique ayant la distribution symétrique:**

- A. la moyenne est égale à la première quartile
- B. 50% des observations sont supérieures à la moyenne
- C.  $Q3 - Q2 = Q2 - Q1$
- D.  $Q3 - Q2 = 3*(Me - Q1)$
- E. Mode  $\approx$  Moyenne  $\approx$  Médiane

**R6: B, C, E**

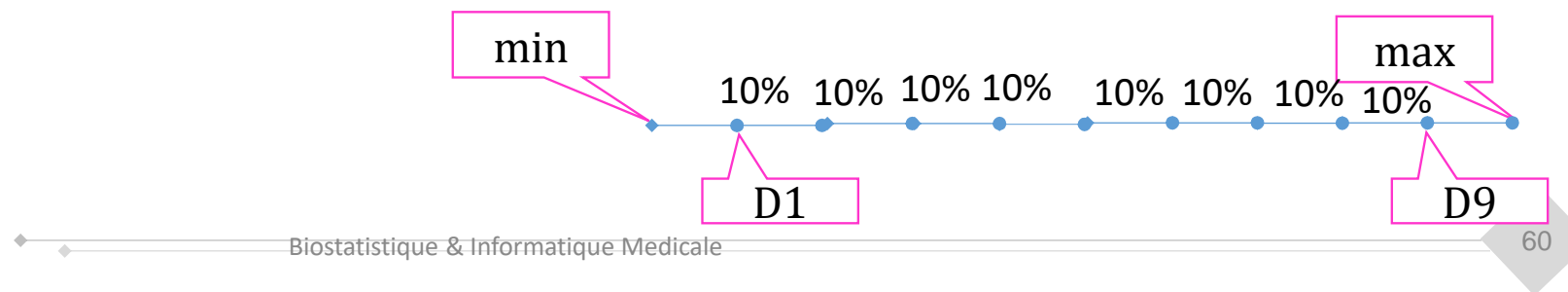
# Exemple de questions

**E7. Pour la distribution d'une variable statistique continue:**

- A. l'histogramme est la représentation graphique des fréquences (relatives ou absolues) cumulées
- B. 15% des observations sont comprises entre le troisième quartile et le neuvième décile
- C. la médiane est égale à la deuxième quartile
- D. l'intervalle inter décile  $[D1, D9]$  contient 90% des observations
- E. l'intervalle inter décile  $[D1, D9]$  contient 80% des observations

**R7: B, C, E**

*Solution de l'exercice: ->*



# Exemple de questions

**E8. Pour minimiser l'influence des valeurs extrêmes :**

- A. on choisit la médiane à la moyenne
- B. on choisit la médiane à l'amplitude(=max-min)
- C. on choisit l'intervalle [min, max] à l'intervalle interquartile [Q1, Q3]
- D. on choisit l'intervalle inter-décile [D1, D9] à l'intervalle [min, max]
- E. on choisit l'intervalle inter-centile [C1, C99] à l'intervalle [min, max]

**R8: A, B, D, E** *Solution de l'exercice: ->*

-> les Deciles



# *Exemple de questions*

**E9. Lesquelles des réponses suivantes sont correctes :**

- A. une bonne graphique pour le niveau de anxiété (mesurée sur une échelle avec des valeurs de 0 à 100) est le graphique histogramme
- B. une bonne graphique pour montrer le lien entre la sténose carotidienne et l'hypertension artérielle est la diagramme en colonnes (barres)
- C. la présence d'une allergie aux antibiotiques peut être bien représentée avec un graphique camembert
- D. la présence d'un infarctus du myocarde (oui/non) peut être bien représentée avec un graphique en barres
- E. une bonne graphique pour le niveau de la intelligence (réduite, normale, supérieure) est l'histogramme

**R9: A, B, C, D** (voir les explications sur le diapo suivant)

# Exemple de questions

*Solution de l'exercice: ->*

-> Variable Niveau de anxiété (mesurée sur une échelle avec des valeurs de 0 à 100) est une variable quantitative continue, donc les valeurs de la variable peuvent être représentées à l'aide d'un histogramme (rectangles juxtaposés/adjacents avec des bases égales et l'hauteur proportionnelle à la fréquence relative de chaque intervalle de classe de sorte que l'aire de l'histogramme est égale à 1).

-> le lien entre les variables sténose carotidienne (variable qualitative dichotomique défini par présence/absence) et l'hypertension artérielle (variable qualitative dichotomique défini par oui/non) sera bien représenté graphiquement par le diagramme en colonnes (barres)

-> Variable Allergie aux antibiotiques (oui/non) peut être bien représentée avec un graphique camembert

-> Variable Infarctus du myocarde (oui/non) peut être bien représentée avec un graphique en barres

-> Variable Niveau de la intelligence (défini par réduite, normale, supérieure) peut être bien représenté par un graphique en bar/colonnes/camembert mais pas l'histogramme (utile dans le cas d'une variable quantitative continue)

## *Exemple de questions*

**E10.** Les valeurs suivantes représentent les jours d'incubation d'un éventuel contact avec l'agent étiologique jusqu'à l'apparition d'une maladie infectieuse: 7, 3, 5, 9, 10, 6, 8, 4, 5, 3, 7, 6, 8, 7, 10, 10, 3, 3, 5, 6, 7, 8. Laquelle des valeurs suivantes correspond à la fréquence relative cumulée de 68.2%:

- A. 7
- B. 5
- C. 10
- D. 8
- E. Ne peut être déterminé sur la base des informations disponibles

**R10. A** (voir les explications sur le diapo suivant)



# Exemple de questions

## *Solution de l'exercice:*

Conformément aux formules de la fréquence relative et freq. relative croissante (voir diapo 38), on obtient les resultants suivants:

jours d'incubation	Fréquence absolue	Fréquence relative	Fréquence relative cumulée croissante
3	4	0.182	0.182
4	1	0.045	0.227
5	3	0.136	0.364
6	3	0.136	0.500
7	4	0.182	0.682
8	3	0.136	0.818
9	1	0.045	0.864
10	3	0.136	1.000

# *Exemple de questions*

**E11. Un médecin de famille a 1 800 personnes (918 femmes) enregistrées a son cabinet médical. En 2015 au cabinet sont enregistrées 506 femmes âgées de 15 à 45 ans et 11 nombre de naissances. Il y a eu 52 crises cardiaques (infarctus du myocarde) et 19 décès enregistrées dans le bureau de ce médecin de famille en 2015.**

- A. le sex-ratio (femmes : hommes) était égal a 1.04
- B. taux de natalité (pour 1000 personnes) est égale a 6.11
- C. taux de fertilité (pour 1000 personnes) est égale a 21.74
- D. l'incidence de l'infarctus du myocarde est égale a 28.89 per mille
- E. le taux de mortalité enregistré dans le cabinet de ce médecin de famille est égal a 10.56

**R11. A, B, C, E (voir les explications sur le diapo suivant)**

# Exemple de questions

## *Solution de l'exercice:*

-> le sex-ratio est une expression du forme: x: y , ou x=effectif de femmes et y=effectif d'hommes;

->Taux de natalité (pour 1000 personnes)=  $\frac{\text{nombre d'evenements dans une periode}}{\text{population au risque de ces evenements dans la meme periode}} \cdot k = \frac{11}{1800} \cdot 1000$

->Taux de fertilité (pour 1000 personnes)=  $\frac{\text{nombre d'evenements dans une periode}}{\text{population au risque de ces evenements dans la meme periode}} \cdot k = \frac{11}{506} \cdot 1000$

->Taux de fertilité (pour 1000 personnes)=  $\frac{\text{nombre d'evenements dans une periode}}{\text{population au risque de ces evenements dans la meme periode}} \cdot k = \frac{11}{506} \cdot 1000$

->l'incidence de l'infarctus du myocarde (pour 1000 personnes)=

$$\frac{\text{nombre d'evenements NOUVEAUX dans une periode}}{\text{population au risque de ces evenements dans la meme periode}} \cdot k$$

-> on ne sait pas (il n'est pas précisé) si les 52 crise cardiaques sont des nouveaux cas ou parmi elles il y a aussi des anciennes patients cardiaques.

->Taux de mortalité (pour 1000 personnes)=  $\frac{\text{nombre d'evenements (deces) dans une periode}}{\text{population au risque de ces evenements dans la meme periode}} \cdot k = \frac{19}{1800} \cdot 1000$

# Exemple de questions

**E12. Regardez la table suivant depuis une article scientifique médical. Il compare différents caractéristiques (âge, sexe, symptômes de covid-19, etc.) chez les sujets diabétiques et ceux non-diabétiques. Répondez à la question de la diapositive suivante:**

	Total (n=193)	Diabetes (n=48)	Non-diabetes (n=145)	P value*
	Number (%)	Number (%)	Number (%)	
Age, median (IQR), years	64 (49 to 73)	70 (62 to 77)	60 (43 to 71)	<0.001
Sex				
Male	114 (59.1)	33 (68.8)	81 (55.9)	0.115
Female	79 (40.9)	15 (31.3)	64 (44.1)	
Symptoms				
Fever	173 (89.6)	43 (89.6)	130 (89.7)	0.989
Cough	135 (69.9)	37 (77.1)	98 (67.6)	0.214
Dyspnea	115 (59.6)	33 (68.8)	82 (56.6)	0.136
Pectoralgia	10 (5.2)	1 (2.1)	9 (6.2)	0.458
Diarrhea	51 (26.4)	10 (20.8)	41 (28.3)	0.311
Nausea	14 (7.3)	2 (4.2)	12 (8.3)	0.528
Vomiting	5 (2.6)	2 (4.2)	3 (2.1)	0.788
Anorexia	68 (35.2)	21 (43.8)	47 (32.4)	0.154
Headache	21 (10.9)	5 (10.4)	16 (11.0)	0.905
Fatigue	101 (52.3)	28 (58.3)	73 (50.3)	0.337
Comorbidities				
Hypertension	94 (48.7)	29 (60.4)	65 (44.8)	0.061
Cardiovascular disease	73 (37.8)	24 (50.0)	49 (33.8)	0.045
Cerebrovascular disease	31 (16.1)	13 (27.1)	18 (12.4)	0.016
Chronic kidney disease	8 (4.1)	5 (10.4)	3 (2.1)	0.036
Chronic pulmonary disease	4 (2.1)	0 (0.0)	4 (2.8)	0.574
Chronic liver disease	14 (7.3)	4 (8.3)	10 (6.9)	0.739
Chronic liver disease	1 (0.5)	0 (0.0)	1 (0.7)	1.000
Exposure to disease	76 (39.4)	17 (35.4)	59 (40.7)	0.517
ICU patients	92 (47.7)	32 (66.7)	60 (41.4)	0.002
Mechanical ventilation treatment†	110 (57.0)	39 (81.3)	71 (49.0)	<0.001
Length of hospital stay, median (IQR), days	13 (7 to 16)	10 (6 to 13)	13 (9 to 18)	0.001
Mortality	108 (56.0)	39 (81.3)	69 (47.6)	<0.001

\*P values indicate differences between diabetes and non-diabetes. A value of  $p < 0.05$  was considered statistically significant.  
 †Non-invasive mechanical ventilation and invasive mechanical ventilation were included. Non-invasive mechanical ventilation included bilevel positive airway pressure ventilation or high-flow nasal cannula oxygen therapy.  
 ICU, intensive care unit.

Source: Alkundi A, Mahmoud I, Musa A, Naveed S, Alshawwaf M. Clinical characteristics and outcomes of COVID-19 hospitalized patients with diabetes in the United Kingdom: A retrospective single centre study. Diabetes Res Clin Pract. 2020 Jul;165:108263. doi: 10.1016/j.diabres.2020.108263. Epub 2020 Jun 10.

Lien vers l'article:

<https://pubmed.ncbi.nlm.nih.gov/32531325/>

# *Exemple de questions*

**Lesquelles des réponses suivantes sont correctes :**

- A. La variable Age (year – années) est décrit dans la table de résultats a l'aide des fréquences absolues
- B. La variable Age (year – années) est décrit dans la table de résultats a l'aide des mesures de tendance centrale et position
- C. L'âge moyen est égale a 70 ans chez les patients diabétiques atteints d'une forme sévère de COVID-19
- D. L'âge médian est égale a 70 ans chez les patients diabétiques atteints d'une forme sévère de COVID-19
- E. 50 % des patients atteints d'une forme sévère de COVID-19 avaient entre 49 et 73 ans.

**R12: B, D, E (voir les explications sur le diapo suivant)**

# *Exemple de questions*

## *Solution de l'exercice:*

-> parce que l'âge est une variable quantitative continue, il sera décrite par les mesures descriptives de tendance centrale, dispersion et de position; dans cet article scientifique, l'âge est décrit à l'aide de la médiane et l'intervalle interquartile (IQR).

# *Exemple de questions*

**E13. Regardez aux même tableau donné à l'exercice E12. Lesquelles des réponses suivantes sont correctes :**

- A. La variable Sexe (défini par homme et femme) est décrit dans la table de résultats a l'aide des fréquences absolues et fréquences relatives
- B. La variable Mortalité était une variable qualitative dichotomiques décrite par des fréquences absolues et fréquences relatives
- C. La variable Symptômes était une variable qualitative nominales décrite par des fréquences absolues et fréquences relatives
- D. 58.3% des patients diabétiques atteints d'une forme sévère de COVID-19 présentaient des symptômes de fatigue
- E. La fréquence relative de la Dyspnée était plus élevée chez les patients diabétiques comparés à ceux non diabétiques.

**R13: A, B, D,E (voir les explications sur le diapo suivant)**

# Exemple de questions

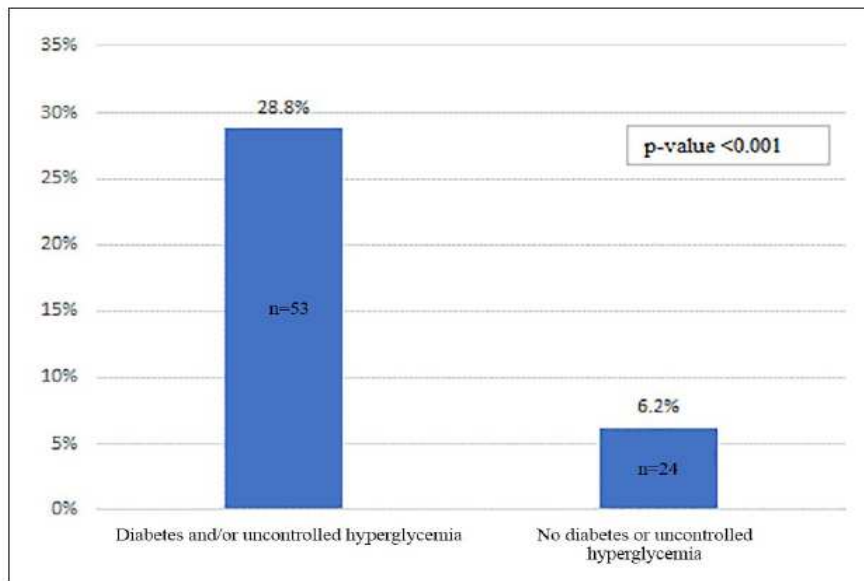
## *Solution de l'exercice:*

- > parce que la variable Sexe est une variable qualitative, il sera décrite par les fréquences absolues (effectif de chaque modalité de la variable) et les fréquences relatives (%).
- > On ne peut pas parler de la variable Symptômes comme une variable qualitative nominale (défini par les modalités: fièvre, toux, dyspnée, nausée, etc.) parce que les modalités ne sont pas mutuellement exclusives (un patient peut présenter simultanément, par exemple fièvre et la toux; voir les pourcentages).
- > La variable Dyspnée est une variable qualitative dichotomique (défini par oui/non) et la fréquence relative de la Dyspnée était plus élevée chez les patients diabétiques comparés à ceux non diabétiques (68.8 % versus 56.6%).



# Exemple de questions

**E14. Regardez au graphique suivant depuis un article scientifique médical. Il compare la fréquence de décès chez les sujets diabétiques et / ou l'hyperglycémie incontrôlée et patients sans diabète ou hyperglycémie. Répondez à la question de la diapositive suivante:**



Bode B, Garrett V, Messler J, McFarland R, Crowe J, Booth R, Klonoff DC. Glycemic Characteristics and Clinical Outcomes of COVID-19 Patients Hospitalized in the United States. J Diabetes Sci Technol. 2020 Jul;14(4):813-821. doi: 10.1177/1932296820924469.

Lien vers l'article:

<https://pubmed.ncbi.nlm.nih.gov/32389027/>

**Figure 3.** Mortality rates among patients who were discharged or died comparing diabetes and/or uncontrolled hyperglycemia ( $n = 184$ ) with patients without diabetes or hyperglycemia ( $n = 386$ ).

# *Exemple de questions*

**Lesquelles des réponses suivantes sont correctes :**

- A. Le graphique présenté est un diagramme en secteurs
- B. Le graphique présenté est un diagramme en colonnes/barres
- C. Le graphique présenté est une histogramme
- D. Le graphique nous montre la proportions de diabétiques et non diabétiques
- E. 6.2% représente la proportion de patients sans diabète ou hyperglycémie qui ont décédé.

**R14: B, E**

MERCI POUR VOTRE  
ATTENTION!



**Contact:** Conf.univ.dr. Mihaela Iancu  
Email: [miancu@umfcluj.ro](mailto:miancu@umfcluj.ro)