

# STATISTIQUES POUR MEDECINS

## Statistiques descriptives



Christophe Combescure  
*Unité d'Appui Méthodologique du CRC*

# *Objectifs de la séance*

- ◆ Connaître les différents types de données
- ◆ Connaître les méthodes d'analyses descriptives

Chapitres Petrie/Sabin  
1 – 4 – 5 – 6 – 7

# Etudes cliniques

- ◆ Objectif:
  - répondre à une question de recherche de manière **objective** et **scientifique**
  - affirmations basées sur des **données mesurables**
- ◆ Comment parvenir à cet objectif ?
  - utilisation de **méthodes** adéquates et reproductibles
    - dessin d'étude (essai randomisé, étude cas-témoins,...)
    - choix des variables à mesurer
    - méthodes de mesure des variables
  - déroulement de l'étude
    - sélection des participants
    - suivi
    - recueil des données
  - **analyse des données**

# Analyse statistique

## ◆ Buts:

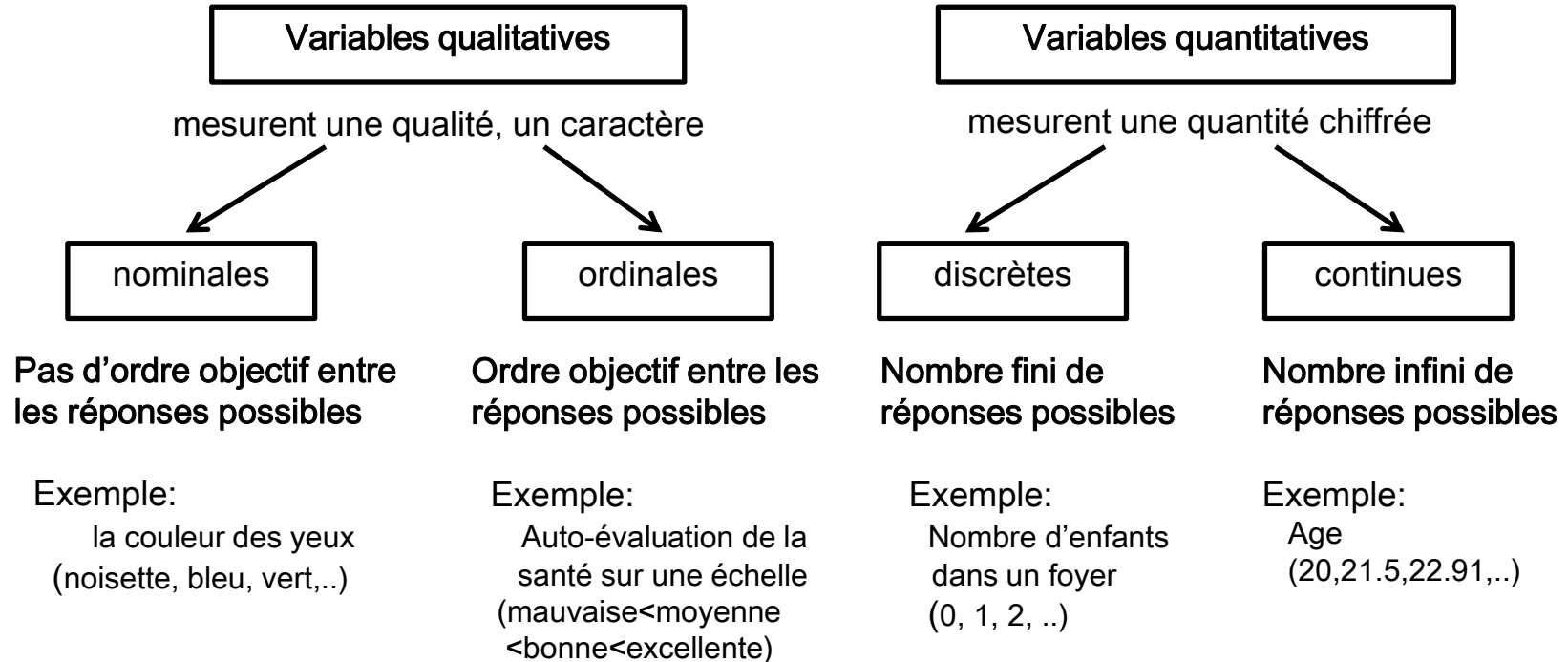
- **décrire** les données: participant.es, résultats principaux
- répondre à la question de recherche (à partir du cours 3)

## ◆ Analyse descriptive d'une variable:

- **représentation graphique** des données
  - diagramme en bâton, histogramme, boxplot, nuage de point, ...
- **statistiques** résumant les données
  - fréquences, mode, moyenne, médiane, écart type, quartiles, ...
- les méthodes appropriées dépendent de la nature de la variable

## ◆ Savoir interpréter la description d'une variable

# Types de variables (1)



# Types de variables (2)

- ◆ Pour les variables quantitatives (discrètes et continues):
  - la différence entre 2 valeurs de la variable a toujours du sens
  - le ratio entre 2 valeurs de la variable n'a pas toujours un sens
  - exemple: variable «année civile»
    - différence entre 2 années civiles = durée
    - le ratio entre 2 années civiles ne s'utilise pas
- ◆ On distingue
  - les variables quantitatives de type **ratio**
    - le ratio a une signification
    - exemples: âge, taille, poids corporel
  - les variables quantitatives de type **intervalle**
    - seule la différence a une signification
    - exemples: année civile, température corporelle en degrés Celsius

# Questionnaire

Quel est votre sexe ? (cochez une case)

Femme ..... 1

Homme ..... 2

← **qualitative, nominale**

Quelle est votre année de naissance ?

20 \_\_\_\_

← **quantitative, discrète, intervalle**

Quelle est votre taille ? (en centimètres, pieds nus)

\_\_\_\_ cm

← **quantitative, continue, ratio**

Dans l'ensemble, pensez-vous que votre santé est...

Excellente ..... 1

Très bonne ..... 2

Bonne ..... 3

Médiocre ..... 4

Mauvaise ..... 5

← **qualitative, ordinale**

# Description d'une variable qualitative ou quantitative discrète (1)

## ◆ Tableau de fréquences

Laquelle de ces 6 spécialités vous attire le plus en ce moment ? (vous pourrez encore changer d'avis !)

- Médecine adulte..... 1 →
- Pédiatrie..... 2 →
- Chirurgie..... 3 →**
- Gynécologie-obstétrique..... 4 →
- Psychiatrie..... 5 →
- Dentaire..... 6 →

Distribution

N (326)	%
52	16.0
51	15.6
122	<b>37.4</b>
24	7.4
34	10.4
43	13.2

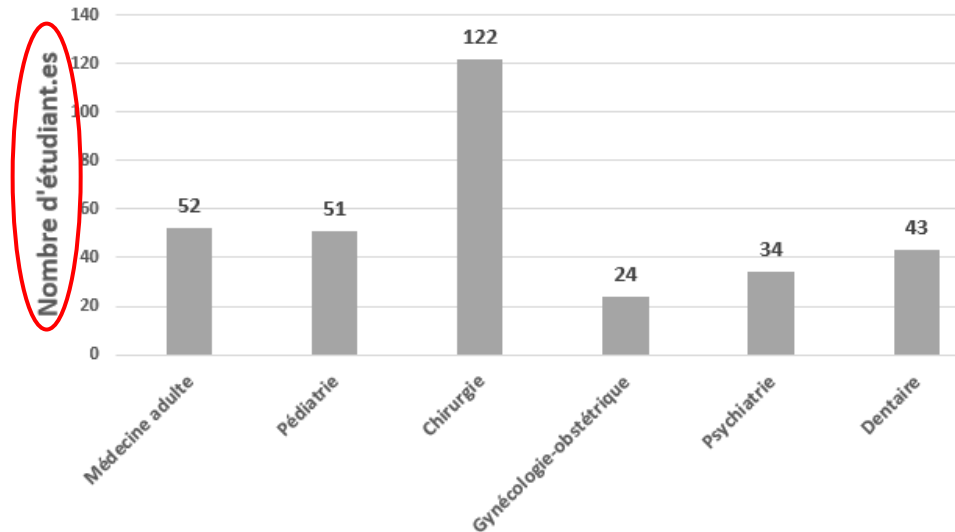
données  
d'une  
volée  
précédente

Mode (réponse la plus fréquente)

# Description d'une variable qualitative ou quantitative discrète (2)

- ◆ Représentation graphique de la distribution : diagramme en bâton (bar chart)

Nombre d'étudiants ou pourcentage

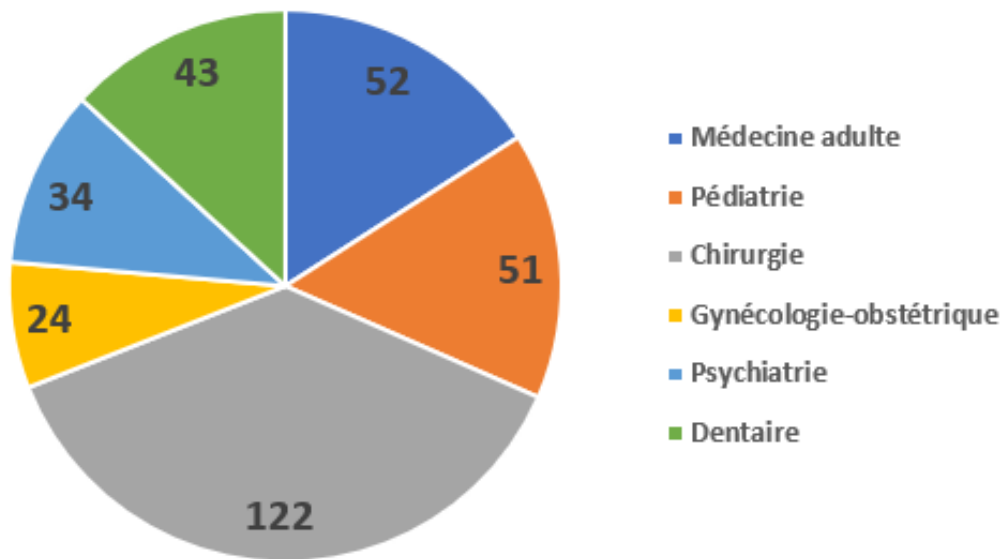


	<b>N (326)</b>	<b>%</b>
Médecine adulte	52	16.0
Pédiatrie	51	15.6
Chirurgie	122	37.4
Gyn.-obst.	24	7.4
Psychiatrie	34	10.4
Dentaire	43	13.2

Si la variable est ordinaire ou quantitative discrète, l'ordre des modalités doit être respecté

# Description d'une variable qualitative ou quantitative discrète (3)

◆ Représentation graphique de la distribution : diagramme circulaire (pie-chart)



	N (326)	%
Médecine adulte	52	16.0
Pédiatrie	51	15.6
Chirurgie	122	37.4
Gyn.-obst.	24	7.4
Psychiatrie	34	10.4
Dentaire	43	13.2

# Tableau croisé de fréquence (1)

- ◆ Distribution de la variable «Spécialité» en fonction de la variable «Genre»

Laquelle de ces 6 spécialités vous attire le plus en ce moment ? (vous pourrez encore changer d'avis !)

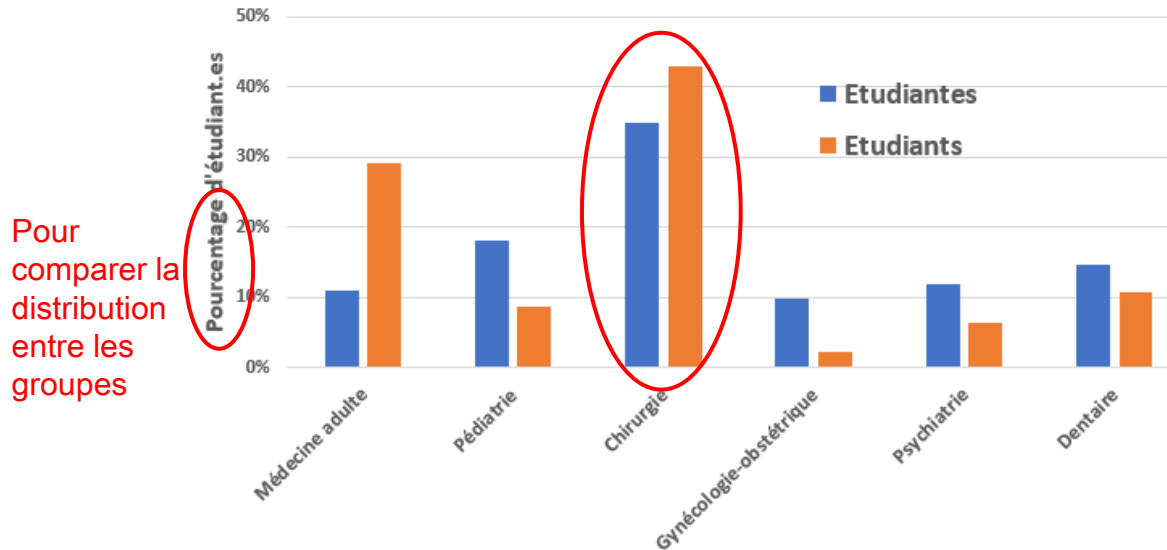
- Médecine adulte..... 1
- Pédiatrie ..... 2
- Chirurgie..... 3
- Gynécologie-obstétrique..... 4
- Psychiatrie..... 5
- Dentaire..... 6

% des spécialités parmi les étudiantes et parmi les étudiants

	Etudiantes	Etudiants
	25 (11.0%)	27 (29.0%)
	41 (18.1%)	8 (8.6%)
	79 (34.8%)	40 (43.0%)
	22 (9.7%)	2 (2.1%)
	27 (11.9%)	6 (6.5%)
	33 (14.5%)	10 (10.8%)
	<b>227</b>	<b>93</b>

# Tableau croisé de fréquence (2)

- ◆ Représentation graphique: diagramme en bâton



35% des étudiantes préfère la spécialité «chirurgie» et 43% des étudiants préfère la spécialité «chirurgie»

## Tableau croisé de fréquence (3)

- ◆ Distribution de la variable «Genre» en fonction de la variable «Spécialité»

Laquelle de ces 6 spécialités vous attire le plus en ce moment ? (vous pourrez encore changer d'avis !)

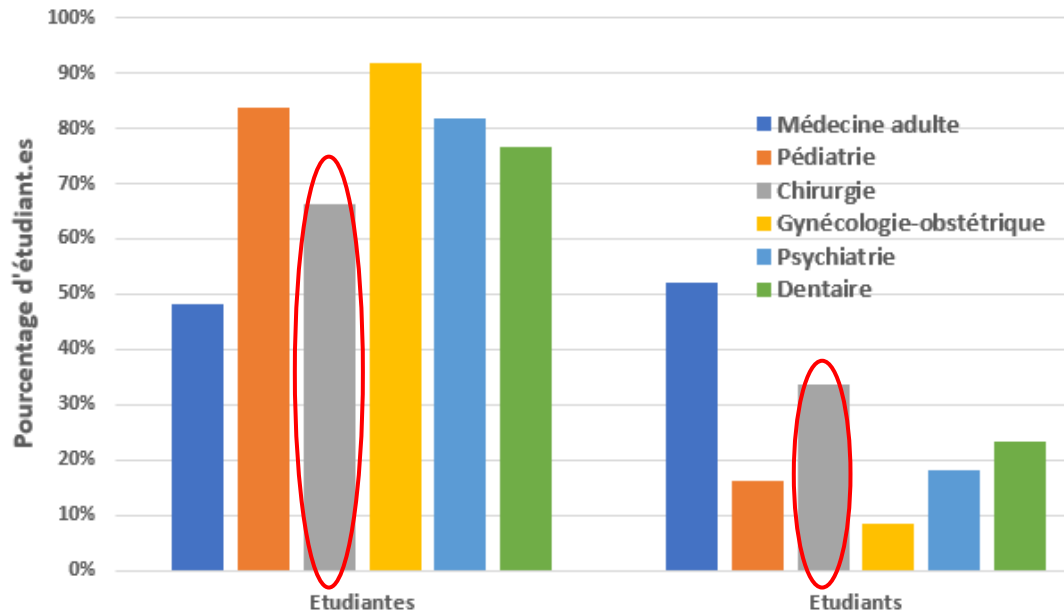
- Médecine adulte.....  1
- Pédiatrie.....  2
- Chirurgie.....  3
- Gynécologie-obstétrique.....  4
- Psychiatrie.....  5
- Dentaire.....  6

% des spécialités parmi les étudiantes et parmi les étudiants

	Etudiantes	Etudiants	
	25 (48.1%)	27 (51.9%)	52
	41 (83.7%)	8 (16.3%)	49
	79 (66.4%)	40 (33.6%)	119
	22 (91.7%)	2 (8.3%)	24
	27 (81.8%)	6 (18.2%)	33
	33 (76.7%)	10 (23.3%)	43

# Tableau croisé de fréquence (4)

- ◆ Représentation graphique: diagramme en bâton

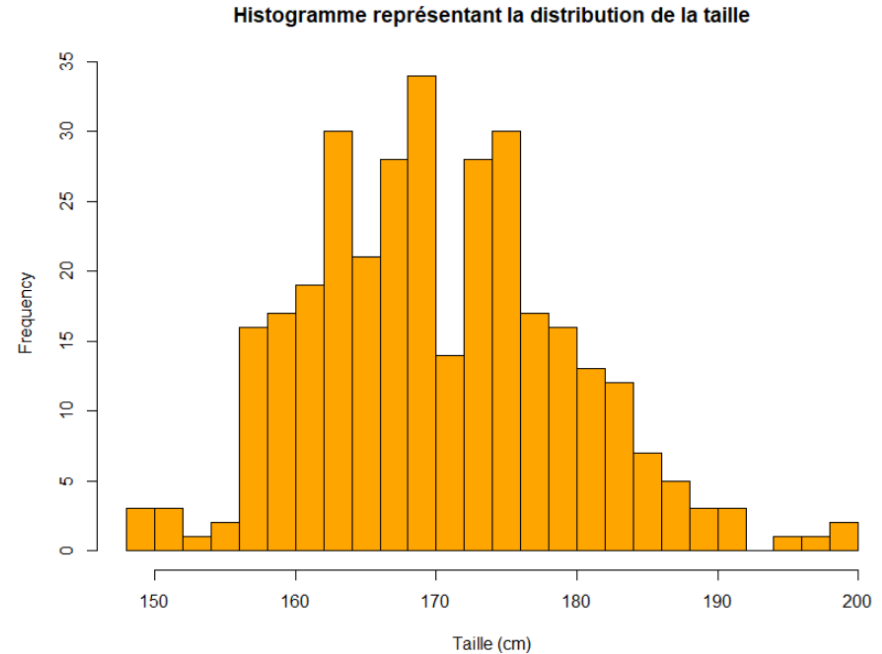


Parmi les personnes préférant la spécialité «chirurgie», 66% sont des étudiantes et 34% sont des étudiants

# Variable quantitative continue

## Histogramme

- ◆ Représentation graphique de la distribution: histogramme
  - l'étendue des valeurs possibles est découpée en intervalles
  - dans chaque intervalle, la surface de la barre est proportionnelle au nombre de sujets dans cet intervalle
  - souvent, les intervalles sont de même longueur et la hauteur de la barre est proportionnelle au nombre de sujets dans l'intervalle correspondant



# Variable quantitative continue

## Moyenne

- ◆ Statistiques: **Tendance centrale** de la distribution
  - **moyenne**
  - valeur « typique » de l'échantillon
  - sensible aux valeurs extrêmes
  - applicable aussi aux variables quantitatives discrètes

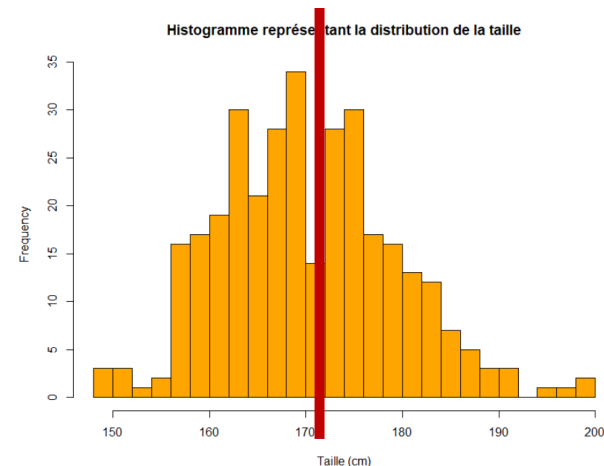
Moyenne de la taille = 170.7 cm

La distribution de la taille est centrée sur la valeur de 170.7 centimètres

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

$x_i$ : donnée mesurée de la variable X chez le sujet i

n: nombre de sujets



# Variable quantitative continue

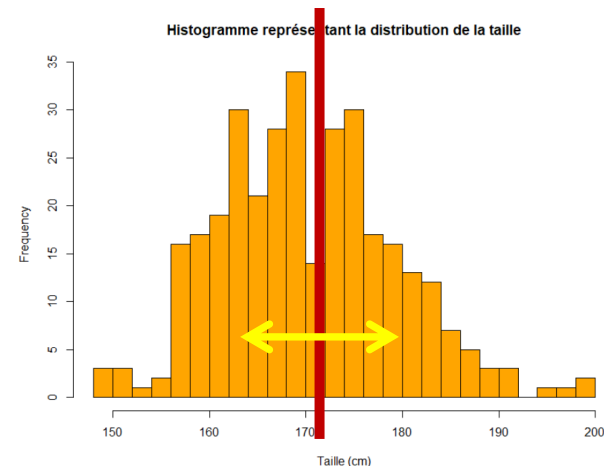
## Ecart type

- ◆ Statistiques: **dispersion** de la distribution
  - variance (notée:  $S^2$ ) et **écart type** (noté:  $S$ )
  - mesure de la dispersion des valeurs autour de la moyenne
  - l'écart type s'exprime dans l'unité de la variable

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$s = \sqrt{s^2}$$

Ecart type de la taille = 9.1 cm



# Variable quantitative continue

## Médiane (1)

### ◆ Médiane:

- valeur qui sépare en 2 parties égales la distribution:
  - autant d'observations au-dessus de la médiane qu'en dessous
- mesure la **tendance centrale** de la distribution
- comment faire:
  - arranger les observations de la plus petite à la plus grande
  - si  $n$  est impair, médiane = valeur qui se situe au milieu
  - si  $n$  est pair, médiane = moyenne des 2 valeurs centrales

# Variable quantitative continue

## Médiane (2)

◆ Exemple de calcul d'une médiane:

– Valeurs mesurées (nombre impair):

- 8, 17, 10, 1, 12, 11, 5, 14, 6

– Valeurs classées par ordre croissant:

- 1, 5, 6, 8, 10, 11, 12, 14, 17

**Valeur centrale=10**

– Valeurs mesurées (nombre pair):

- 8, 17, 10, 1, 12, 11, 5, 14, 6, 9

– Valeurs classées par ordre croissant:

- 1, 5, 6, 8, 9, 10, 11, 12, 14, 17

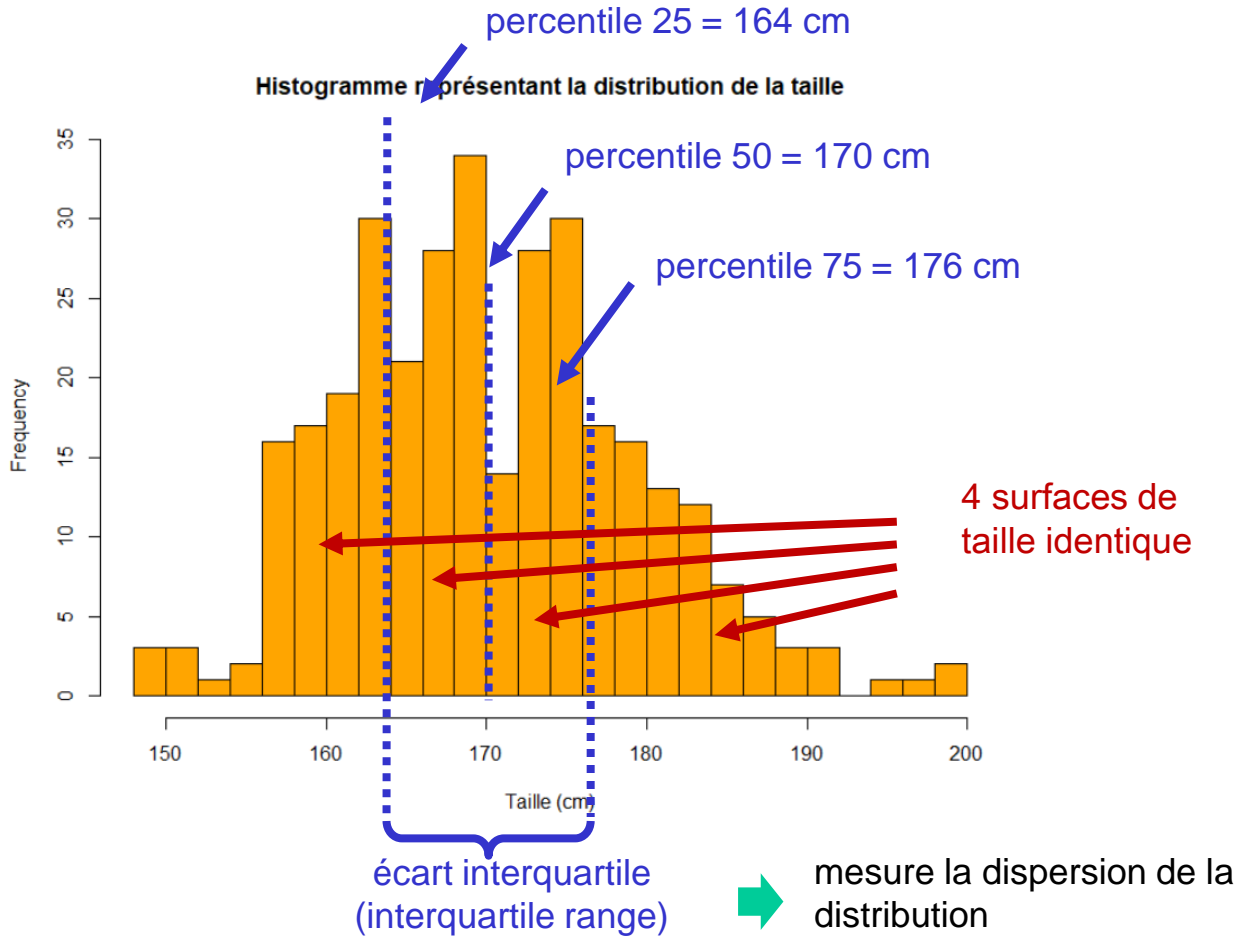
**Moyenne des valeurs  
centrales=9.5**

# *Variable quantitative continue*

## *Percentile (1)*

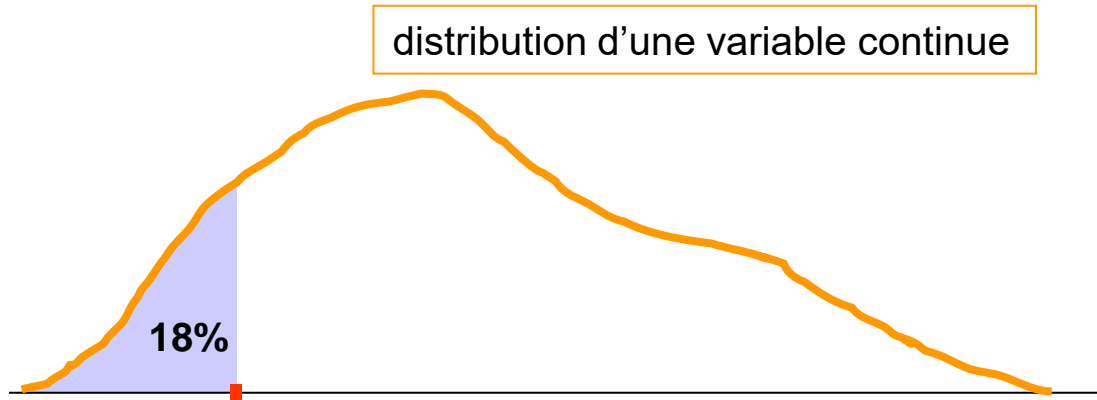
### ◆ Percentiles:

- $k^{\text{ème}}$  percentile: valeur correspondant à  $k/100$  de la distribution (arrangée de la plus petite valeur à la plus grande)
- Médiane =  $50^{\text{ème}}$  percentile
- Quartiles
  - premier quartile =  $25^{\text{ème}}$  percentile
  - deuxième quartile =  $50^{\text{ème}}$  percentile (= médiane)
  - troisième quartile =  $75^{\text{ème}}$  percentile



# Variable quantitative continue

## Percentile (2)

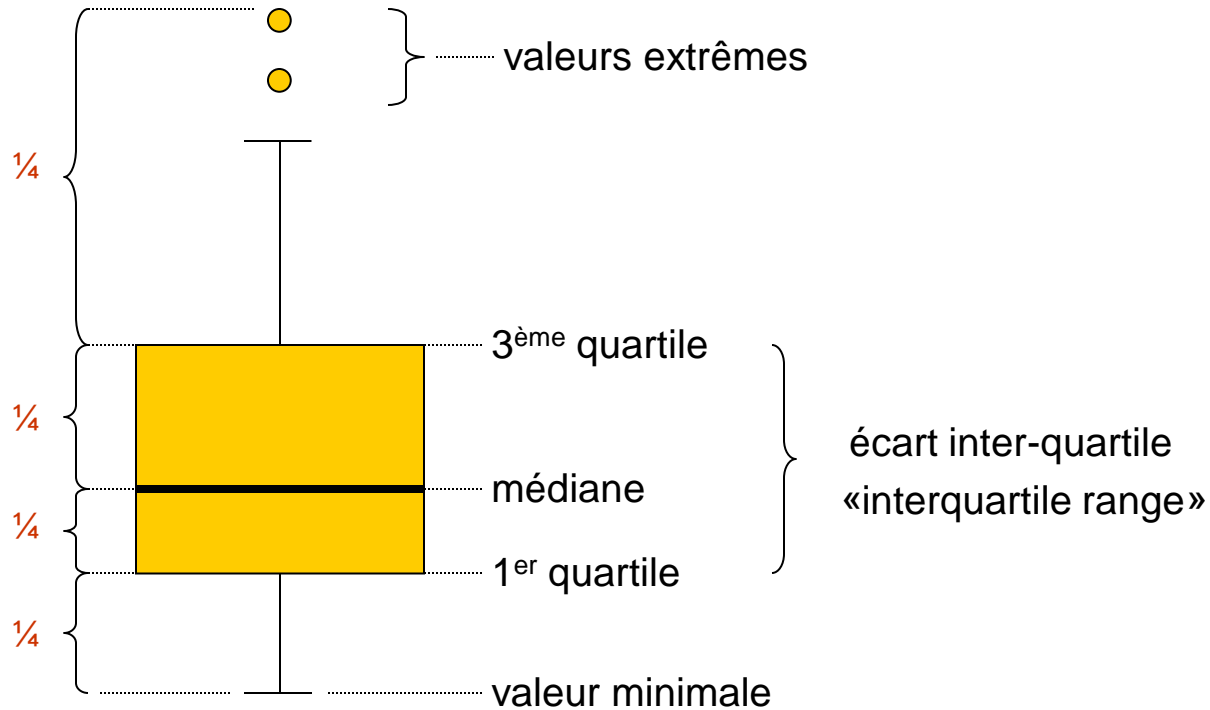


percentile 18

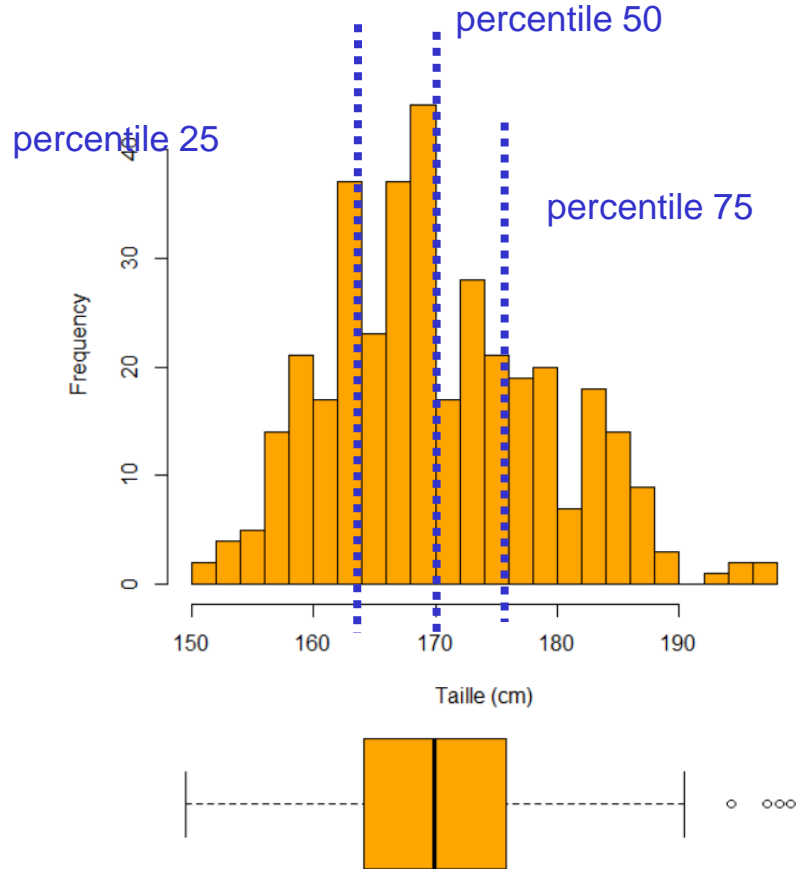
valeur en dessous de laquelle se trouvent 18% des observations

# Variable quantitative continue

## Box-plot



# Percentiles et box-plot



# Comparaison visuelle entre 2 groupes

## Histogramme

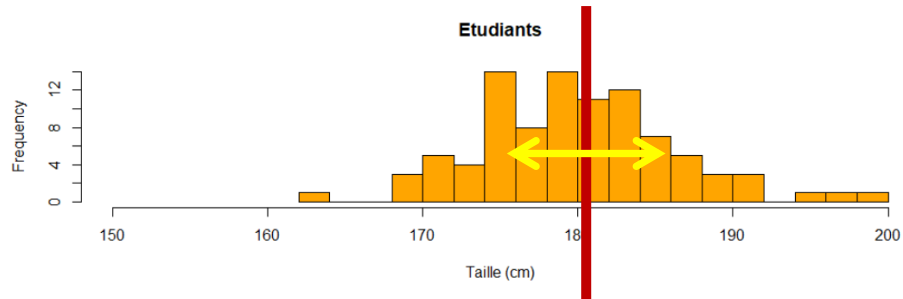
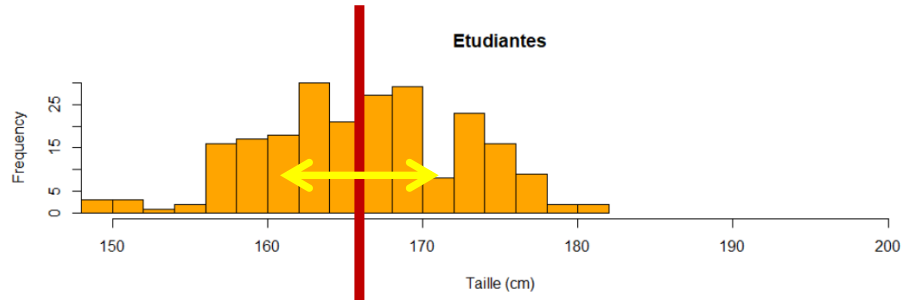
- ◆ Relation entre une variable qualitative et une variable quantitative continue

Relation entre genre et taille

Taille moyenne (écart type):

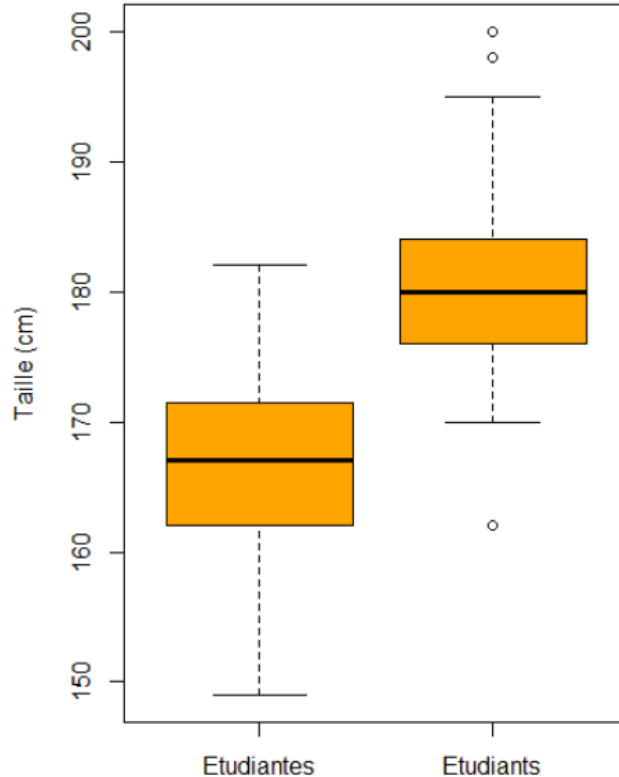
- étudiantes = 166.6 cm (6.5 cm)
- étudiants = 180.5 cm (6.4 cm)

Chez les participant-es à l'enquête, la taille moyenne est plus grande de 13.9 cm chez les étudiants que chez les étudiantes



# Comparaison visuelle entre 2 groupes

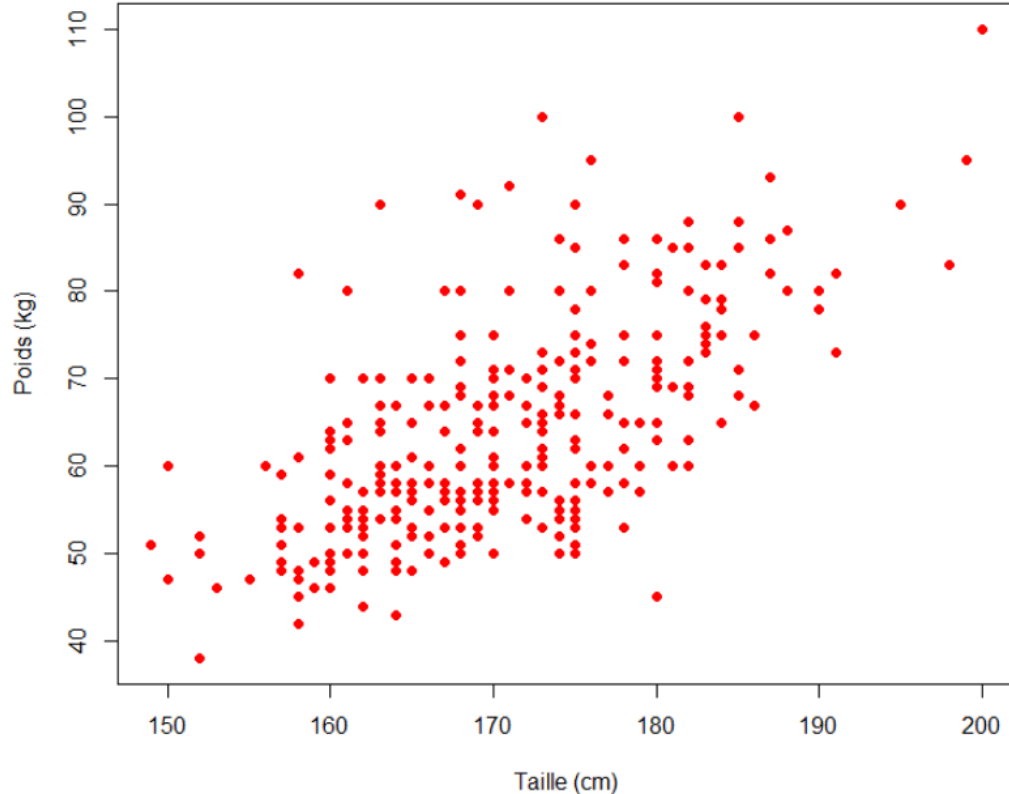
## Box-plot



Utile pour comparer graphiquement la distribution d'une variable entre deux groupes

# Relation entre 2 variables quantitatives continues

## Nuage de points (scatter plot)



Utile pour représenter graphiquement la relation entre deux variables quantitatives mesurées chez les mêmes sujets

# *Descriptions en recherche clinique*

- ◆ Les participant.es et les résultats principaux sont souvent décrits par:
  - ◆ **intervention** (intervention expérimentale versus placebo ou intervention de référence) dans un essai clinique randomisé (cours 7)
  - ◆ **niveau d'exposition** (par exemple non exposé versus exposé) dans les études prospectives cherchant à déterminer des facteurs de risque d'un problème de santé (cours 10)
  - ◆ **issue** (par exemple caractéristiques des participant-es qui décèdent versus participant-es qui survivent) notamment dans les études cas-témoins (cours 12)

# Association entre 2 variables

Deux variables sont associées

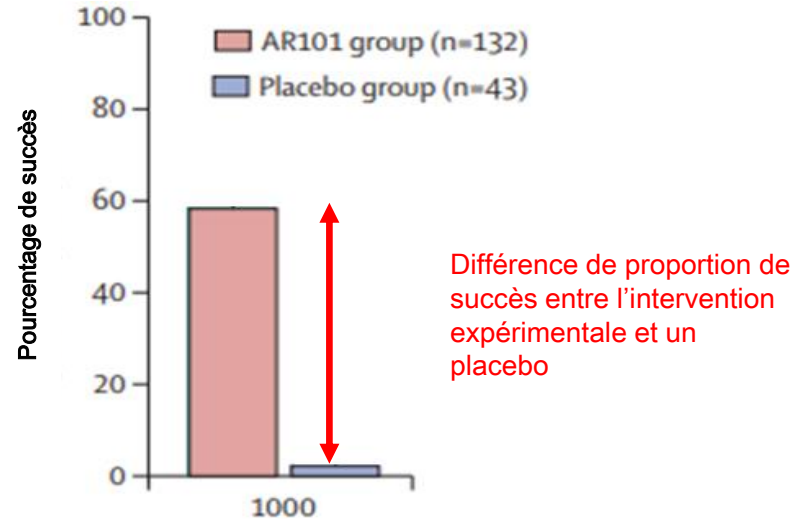
=

la distribution de l'une des variables dépend de la valeur de l'autre variable

- ◆ L'effet d'une intervention expérimentale par rapport à une intervention contrôle sur un critère de jugement est l'association entre la variable «Intervention» et le critère de jugement (outcome)
- ◆ Exemples:
  - proportion de décès plus basse avec le traitement A qu'avec le traitement B
  - niveau moyen de cholestérol plus avec le traitement A qu'avec le traitement B
- ◆ Une exposition est un facteur de risque (ou facteur protecteur) si elle est associée au problème de santé étudié
- ◆ Exemple:
  - proportion de fumeur-ses plus élevée chez les personnes ayant un cancer du poumon que chez celles n'ayant pas de cancer du poumon ou proportion de personnes avec un cancer de poumon plus élevée chez les fumeur-ses que chez les non fumeur-ses

# Mesures d'association

- ◆ L'ampleur de l'association est une information importante
- ◆ Mesure d'association: mesure quantitative
  - ◆ différence ou ratio de proportion
  - ◆ différence de moyenne
  - ◆ etc.
- ◆ L'ampleur de l'association peut être représentée graphiquement



# *Exemples d'analyses descriptives*

- ◆ 2 essais cliniques randomisés
  - ◆ un essai avec un critère de jugement qualitatif
  - ◆ un essai avec un critère de jugement quantitatif continu
- ◆ 1 étude transversale
- ◆ 1 étude cas-témoins
- ◆ 1 étude de cohorte
- ◆ Les exemples du cours sont donnés pour illustrer les méthodes / propos. Ils **ne** sont **pas** à connaître pour l'examen.

# 1<sup>er</sup> essai clinique randomisé (1)

Efficacy and safety of oral immunotherapy with AR101 in European children with a peanut allergy (ARTEMIS): a multicentre, double-blind, randomised, placebo-controlled phase 3 trial

*Lancet Child Adolesc Health* 2020; 4: 728-39

- ◆ Est-ce qu'une immunothérapie (AR101) a un effet sur la tolérance aux arachides chez des enfants souffrant d'allergie aux arachides par rapport à un placebo ?
- ◆ Formulation de la question de recherche avec les critères PICO:
  - ◆ Population: enfants souffrant d'allergies aux arachides
  - ◆ Intervention expérimentale: immunothérapie orale, 300 mg/jour d'AR101 pendant 12 semaines
  - ◆ Comparateur: placebo
  - ◆ Outcome (critère de jugement): tolérance à une dose de 1 gramme de protéine d'arachide
- ◆ Randomisation 3:1 : attribution aléatoire de l'intervention avec 3x plus d'enfants recevant l'immunothérapie que d'enfants recevant le placebo

# 1<sup>er</sup> essai clinique randomisé (2)

Description des participant.es  
par groupe d'intervention

	AR101 group (n=132)	Placebo group (n=43)
Sex		
Male	68 (52%)	27 (63%)
Female	64 (48%)	16 (37%)
Mean age, years	9.0 (3.7)	9.5 (3.9)
Age group		
4-11 years	97 (73%)	30 (70%)
12-17 years	35 (27%)	13 (30%)
History of pre-trial peanut anaphylaxis	58 (44%)	22 (51%)
History of asthma	56 (42%)	14 (33%)
Atopic dermatitis	78 (59%)	22 (51%)
Allergic rhinitis	63 (48%)	16 (37%)
Food allergies other than peanut	81 (61%)	21 (49%)
Data are n (%),		

**Table 1: Baseline demographics and clinical characteristics of participants**

Globalement, autant de filles que de garçons

Environ 30% sont des adolescent.es

Presque la moitié ont eu une réaction allergique sévère  
aux arachides avant l'essai

Plus d'un tiers ont des antécédents d'asthme

Etc.

Avec la randomisation, il est attendu que la distribution  
d'une caractéristique soit similaire dans les 2 groupes

Les différences observées sont dues au hasard

Plus le nombre de participant.es est grand, plus le  
risque d'observer des déséquilibres importants est  
faible

# 1<sup>er</sup> essai clinique randomisé (3)

Critère de jugement (outcome) = tolérance à une dose de 1 g de protéine d'arachide

Test à la fin de l'essai pour déterminer si l'enfant tolère une dose de 1 g de protéine d'arachide

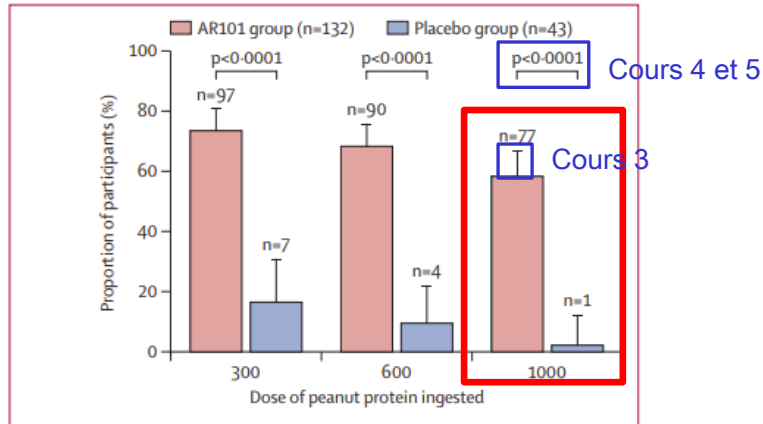


Figure 2: Proportion of participants who tolerated at least 300 mg, 600 mg, or 1000 mg peanut protein as a single dose in the exit double-blind placebo-controlled food challenge

		Groupe	
		AR101	Placebo
Tolérance à une dose de 1 g	Non	55 (41.7%)	42 (97.7%)
	Oui	77 (58.3%)	1 (2.3%)
Total		132	43

Dans cet essai, on observe une plus grande fréquence de participants tolérant une dose de 1 g de protéine d'arachide sous AR101 (58.3%) que sous placebo (2.3%).

Ces résultats suggèrent un effet bénéfique massif de l'AR101 par rapport à un placebo.

# 1<sup>er</sup> essai clinique randomisé (4)

		Groupe	
		AR101	Placebo
Tolérance à une dose de 1 g	Non	55 (41.7%)	42 (97.7%)
	Oui	77 (58.3%)	1 (2.3%)
Total		132	43

L'ampleur de l'effet de l'AR101 par rapport à un placebo peut être décrite par

- la différence absolue du pourcentage de tolérants entre les 2 groupes:
  - $58.3 - 2.3 = 56.0\%$
  - la fréquence de participant.es tolérant est plus grande (en absolue) de 56.0% dans le groupe AR101 que dans le groupe placebo
- le ratio du pourcentage de tolérants entre les 2 groupes :
  - $58.3 / 2.3 = 25.3$
  - la fréquence de participant.es tolérant est 25.3 fois plus grande dans le groupe AR101 que dans le groupe placebo

Mesures  
d'association

# 2<sup>ème</sup> essai clinique randomisé (1)

Physical Fitness Training in Patients with Subacute Stroke  
(PHYS-STROKE): multicentre, randomised controlled, endpoint  
blinded trial

BMJ 2019;366:l5101

- ◆ Est-ce qu'une série d'exercices physiques a un effet sur la mobilité chez des adultes après un accident vasculaire cérébral (AVC) non aigu par rapport à des séances de relaxation ?
  - ◆ Formulation de la question de recherche avec les critères PICO:
    - ◆ **P**opulation: adultes venant d'avoir un AVC non aigu
    - ◆ **I**ntervention expérimentale: réhabilitation standard + fitness (5 sessions de 25 mn par semaine pendant 4 semaines)
    - ◆ **C**omparateur: réhabilitation standard + relaxation (5 sessions de 25 mn par semaine pendant 4 semaines)
    - ◆ **O**utcome (critère de jugement): vitesse maximale de marche et index de Barthel
- Variables quantitatives continues**
- ◆ Index de Barthel: mesure de la performance dans les activités quotidiennes
  - ◆ Randomisation 1:1 : attribution aléatoire de l'intervention avec un nombre de participant.es équilibré dans les 2 groupes

## 2<sup>ème</sup> essai clinique randomisé (2)

Description des participant.es par groupe + Total d'intervention

**Table 1 | Baseline characteristics of participants stratified by aerobic physical fitness training or relaxation sessions (control group). Values are numbers (percentages) unless stated otherwise**

Characteristic	Aerobic physical fitness training (n=105)	Relaxation sessions (n=95)	Total cohort (n=200)
Mean (SD) age (years)	69 (12)	70 (11)	69 (12)
Women	45 (43)	36 (38)	81 (41)
Median (interquartile range) time from stroke to intervention (days)*	30 (17-39)	27 (17-41)	28 (17-40)
Anterior circulation stroke	84 (80)	72 (76)	156 (78)
Hemiparesis on admission	98 (93)	89 (94)	188 (94)
Median (interquartile range) NIHSS score†	9 (5-12)	7 (5-11)	8 (5-12)
Ischaemic stroke	91 (87)	90 (95)	181 (91)
Treatment with alteplase‡	34 (37)	27 (30)	61 (34)
Cause of ischaemic stroke‡:			
Large artery atherosclerosis	17 (19)	19 (21)	36 (20)
Cardioembolism	18 (20)	18 (20)	36 (20)
Small vessel occlusion	16 (18)	15 (17)	31 (17)
Other causes	3 (3)	4 (4)	7 (4)
Undetermined causes	34 (37)	28 (31)	62 (34)
Competing causes	3 (3)	6 (7)	9 (5)
Previous stroke	27 (27)	27 (28)	54 (27)

Pour les variables quantitatives continues, il est fréquent de préférer la médiane à la moyenne lorsque la distribution de la variable est très asymétrique

Caractéristiques ne concernant que les patient-es ayant eu un AVC ischémique

Pourcentages calculés sur le nombre de patient-es avec un AVC ischémique

## 2<sup>ème</sup> essai clinique randomisé (3)

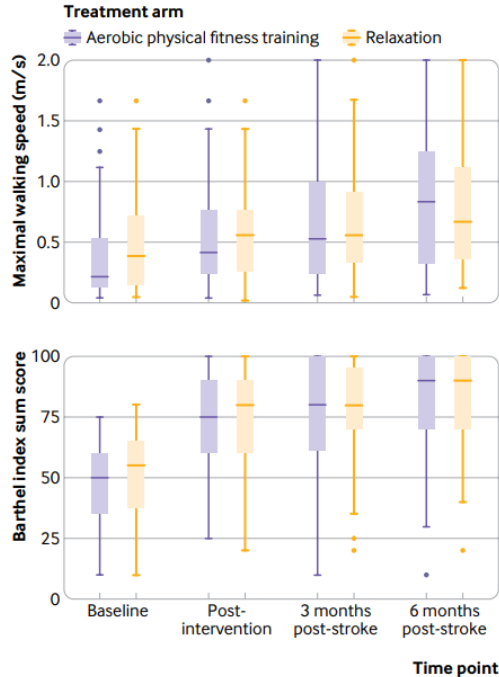


Fig 3 | Boxplot showing medians (interquartile ranges) for maximal walking speed assessed by 10 m walk test (top panel), and Barthel index score (bottom panel) for each study visit and intervention group. Data are based on measurements without multiple imputation. Number of participants at each scheduled study visit was: baseline (n=105 in aerobic physical fitness training group, n=95 in relaxation group-), post-intervention (n=87, n=85), three months post-stroke (n=89, n=77), and six months post-stroke (n=80, n=65). Dots represent outliers

### Critères de jugement:

- vitesse maximale de marche
- index de Barthel: mesure de la performance dans les activités quotidiennes

### Les box plots montrent que globalement dans l'étude:

- la vitesse maximale de marche s'améliore progressivement avec le temps dans les 2 groupes
- l'index de Barthel s'améliore aussi dans les 2 groupes, surtout au début du suivi
- l'amélioration semble similaire dans les 2 groupes

Les boxplots ne montrent pas l'évolution des valeurs de chaque participant.e individuellement

## 2<sup>ème</sup> essai clinique randomisé (4)

Table 2 | Results for primary efficacy outcome of **change in** maximal walking speed and Barthel index score from baseline to three months after stroke by aerobic physical fitness training or relaxation sessions (control group)

Primary outcomes	Aerobic physical fitness training (n=105)	Relaxation sessions (n=95)
Mean (95% CI) maximal walking speed (m/s)	0.4 (0.3 to 0.4)	0.3 (0.2 to 0.4)
Mean (95% CI) Barthel index score	30 (24 to 36)	30 (23 to 36)

En moyenne, la vitesse maximale de marche s'améliore entre le début l'étude et 3 mois après l'intervention de :

- 0,4 m/s dans le groupe «Fitness»
- 0,3 m/s dans le groupe «Relaxation»



Mesure d'association: + 0,1 m/s dans le groupe «Fitness» par rapport au groupe «Relaxation»

L'amélioration de la vitesse maximale de marche est plus grande de 0,1 m/s dans le groupe «Fitness» que dans le groupe «Relaxation»

L'effet de «Fitness» (par rapport à «Relaxation») observé dans cet essai est faible

# Etude transversale (1)

- ◆ Swissnoso – Centre national de prévention des infections
- ◆ Enquête sur les infections nosocomiales 2022
- ◆ Objectif: évaluer la prévalence d'infections nosocomiales en Suisse

## Participating hospitals' characteristics

	Hospitals, N		Patients, N	
All	108	100%	13916	100%
Small size hospitals (<200 beds)	76	70.4%	4416	31.7%
Medium-size hospitals (200-650 beds)	24	22.2%	4857	34.9%
Large-size hospitals (>650 beds)	8	7.4%	4643	33.4%
University hospitals	5	4.6%	3404	24.5%
Primary hospitals	53	49.1%	3633	26.1%
Secondary hospitals	36	33.3%	4794	34.4%
Tertiary hospitals	10	9.3%	4934	35.5%
Specialized hospitals	9	8.3%	555	4.0%
Public hospitals	63	58.3%	10834	77.9%
Private non-for-profit hospitals	21	19.4%	1793	12.9%
Private for-profit hospitals	24	22.2%	1289	9.3%

Attention:

ce n'est **pas** un tableau croisé de fréquence

ce tableau décrit

- 1) les hôpitaux
- 2) les patient.es

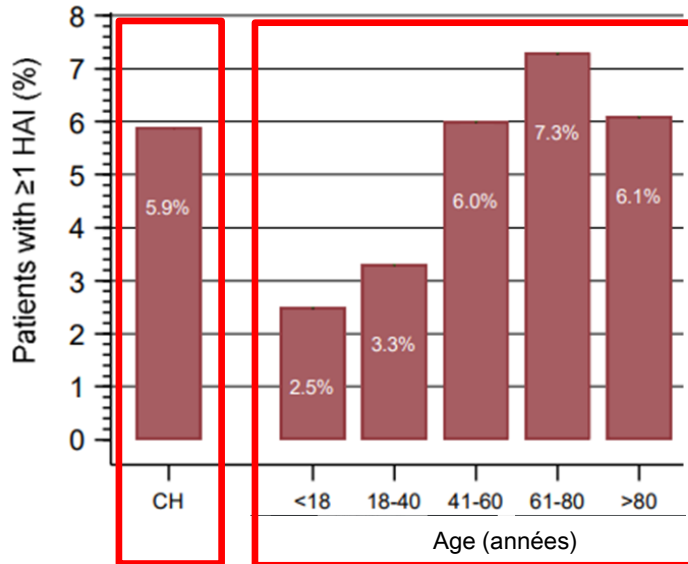
Beaucoup de «petits» hôpitaux participent à l'enquête (70%) mais contribuent à 32% des participant-es

Cela reflète la réalité des hospitalisations en Suisse

# Etude transversale (2)

	Age (années)				
	< 18	18-40	41-60	61-80	>80
≥ 1 infection noso. Non	97.5%	96.7%	94.0%	92.7%	95.9%
Oui	2.5%	3.3%	6.0%	7.3%	6.1%

Dans cette étude, on observe une prévalence globale d'infections nosocomiales de 5.9%



Représentation graphique du pourcentage de patients avec ≥ 1 infection nosocomiale par tranche d'âge

Permet de visualiser l'association entre l'âge des patient.es et la survenue d'au moins une infection nosocomiale

On observe que, globalement, cette proportion augmente avec l'âge

Attention: ce diagramme en bâton **ne** représente **pas** la distribution de l'âge des participant.es (somme des % ≠ 100)

# *Etude cas-témoins (1)*

Risk Factors and Outcomes of Invasive Aspergillosis in  
Kidney Transplant Recipients: A Case-Control Study of  
United States Renal Data System Data

Clinical Infectious Diseases®

2023;76(8):1431-9

- ◆ Population: receveur.ses d'une greffe de rein
- ◆ Problème de santé: infection invasive à aspergillus
- ◆ Objectif: identifier des facteurs de risque d'infections invasives à aspergillus
  
- ◆ Etude cas-témoins
  - les cas sont les patient.es transplanté.es qui ont développé une infection invasive à aspergillus
  - les témoins sont les patient.es transplanté.es qui **n'ont pas** développé une infection invasive à aspergillus
  - pour chaque cas inclus dans l'étude, 4 témoins sont recrutés dans le même centre à peu près au moment

## Description des participant.es par issue

Characteristic	Cases of Aspergillosis (n = 359)	Matched Controls (n = 1436)	
<b>Recipient characteristics</b>			
Age at transplant, y, mean (SD)	56.1 (12.6)	53.9 (15.2)	+2.2 ans chez les cas en moyenne
<b>Sex</b>			
Male	228 (63.5)	846 (58.9)	
Female	131 (36.5)	590 (41.1)	
<b>Blood type</b>			
A	133 (37.1)	573 (39.9)	distribution similaire
B	44 (12.3)	159 (11.1)	
AB	17 (4.7)	88 (6.1)	
O	162 (45.1)	603 (42.0)	
<b>Indication for transplant</b>			
Diabetic nephropathy	133 (37.1)	405 (28.2)	
Hypertensive nephropathy	25 (7.0)	70 (4.9)	
Glomerular disease	72 (20.1)	283 (19.7)	
Polycystic kidney disease	20 (5.6)	171 (11.9)	
Other	109 (30.4)	507 (35.3)	hémodialyse 3x plus longue chez les cas
Duration of pretransplant hemodialysis, y, mean (SD)	4.2 (4.5)	1.3 (2.6)	
ECl score at diagnosis, mean (SD)	4.6 (1.9)	2.8 (2.3)	
<b>Comorbid conditions at diagnosis</b>			
Congestive heart failure	84 (23.4)	109 (7.6)	comorbidité 3x plus fréquente chez les cas
Peripheral vascular disease	29 (8.1)	67 (4.7)	
Cancer	14 (3.9)	37 (2.6)	
Anemia	21 (5.9)	35 (2.4)	
Coagulopathy	51 (14.2)	80 (5.6)	

## *Etude cas-témoins (3)*

- ◆ les expositions peuvent être des variables quantitatives ou qualitatives
- ◆ le problème de santé étudié est une variable binaire (cas/témoin)
- ◆ l'association entre une exposition et le problème de santé peut être mesurée par la différence de proportion ou de moyenne de l'exposition entre les cas et les témoins
- ◆ une exposition plus fréquente ou plus élevée chez les cas que chez les témoins suggère que l'exposition est un facteur de risque
- ◆ plus la différence entre les cas et les témoins est grande, plus l'association est forte

# *Etude de cohorte (1)*

Risk of preterm birth, small for gestational age at birth, and stillbirth after covid-19 vaccination during pregnancy:  
population based retrospective cohort study [BMJ 2022;378:e071416](#)

- ◆ Objectif: évaluer le risque de naissance prématurée chez des femmes ayant reçu au moins une dose de vaccin contre le Covid-19 pendant la grossesse
- ◆ Etude non randomisée: les données ont été collectées sans que la prise en charge soit impactée par la participation à l'étude

# Etude de cohorte (2)

## Exposition

Characteristics	≥1 vaccine dose during pregnancy (n=43 099)*	No vaccine during pregnancy (n=42 063)*
Maternal age (years)		
<25	1836 (4.3)	5041 (12.0)
25-29	8652 (20.1)	11 686 (27.8)
30-34	19 127 (44.4)	15 223 (36.2)
35-39	11 287 (26.2)	8 118 (19.3)
≥40	2197 (5.1)	1995 (4.7)
Mean (SD) age	32.9 (4.4)	31.3 (5.3)
Pre-existing medical condition§		
Yes	5253 (12.2)	4221 (10.0)
Thyroid disease	2957 (6.9)	2061 (4.9)
Asthma	1683 (3.9)	1596 (3.8)
Diabetes	483 (1.1)	427 (1.0)
Chronic hypertension	414 (1.0)	347 (0.8)
Heart disease	56 (0.1)	44 (0.1)
Smoked during pregnancy		
No	40 566 (94.1)	37 537 (89.2)
Yes	1461 (3.4)	3707 (8.8)
Missing‡	1072 (2.5)	819 (1.9)

Absence de randomisation

=>

La distribution des caractéristiques peut différer entre les exposés et les non exposés

Il y a 2.5x plus de fumeuses chez les femmes non vaccinées pendant la grossesse  
Or le tabagisme pendant la grossesse est un facteur de risque connu de prématurité (risque x 2 ou 3)

## Etude de cohorte (3)

Outcomes	No vaccine during pregnancy (n=42 063)	≥1 vaccine dose during pregnancy* (n=43 099)
Preterm birth <37 weekst		
No of participants	41 879	42 992
No (%) with outcome (rate/100 live births)	2907 (6.9)	2812 (6.5)

Le pourcentage de naissance prématuré est:

6.9% dans le groupe des femmes non vaccinées

6.5% dans le groupe des femmes vaccinées

Mesure de l'association entre la vaccination et le risque de naissance prématuré:

différence de proportions: -0.4% chez les vaccinées

ratio de proportions:  $6.5/6.9 = 0.94$  (chez les vaccinées, le % de prématurité est 0.94 fois le % chez les non vaccinées)

L'association observée est faible

Attention: il se peut que cette faible association apparente soit en partie expliquée par la plus grande fréquence de fumeuses chez les non vaccinées (effet de confusion : cours 13)

# Bien d'autres représentations graphiques...

## Enquête sur les infections nosocomiales

Figure 18 : HAI types (proportions) in all participating hospitals in the two national PPS (SSI: surgical-site infection, LRTI: lower respiratory tract infection, BSI: bloodstream infection, UTI: urinary tract infection, GI: gastro-intestinal infection, EENT: eye, ear, nose, throat or mouth infection, SYS: systemic infection, NEO: neonatal infection, other: other infection)

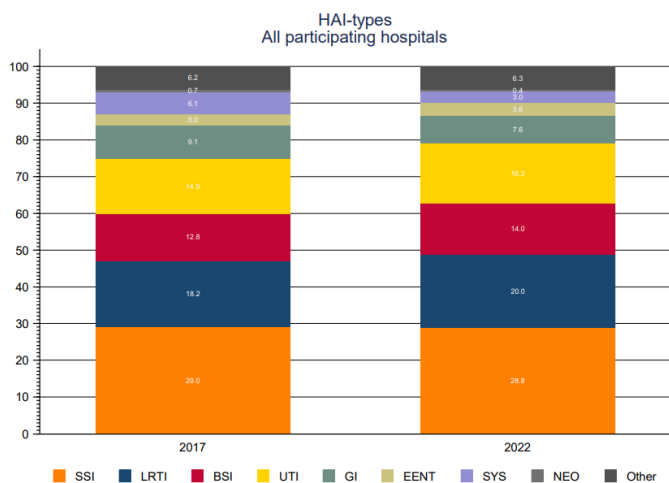
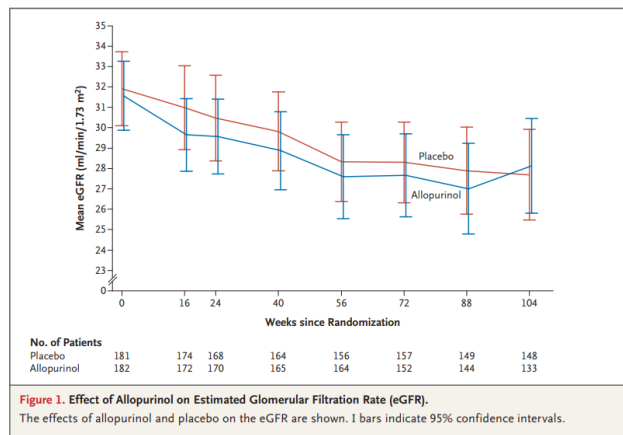


Diagramme en bâton empilé



Représentation de la moyenne d'une variable au cours du suivi



## *et d'autres mesures d'association*

- ◆ Odds ratio (cours 12), hazard ratio (cours 11), coefficient de corrélation (cours 6)....
- ◆ Le choix de la mesure d'association dépend:
  - du type d'études cliniques
  - de la nature des variables
- ◆ Vous verrez les mesures d'association les plus couramment utilisées dans les prochains cours mais il en existe d'autres
- ◆ L'objectif est toujours de mesurer quantitativement l'ampleur de la relation entre 2 variables

# *Description ne vaut pas preuve*

- ◆ Les participant.es à l'étude ne sont qu'une partie de la population étudiée (= échantillon)
- ◆ Les résultats observés dans l'étude sont sujets à une imprécision
- ◆ Une association (un effet) observée dans l'étude ne signifie pas forcément que cette association existe dans la population
- ◆ Une association observée dans l'étude peut être partiellement ou totalement expliquée par des différences entre les exposé.es et les non exposé.es (par ex. fréquence de tabagisme différente)
- ◆ La description de l'association est utile mais **pas suffisante** pour répondre à la question de recherche
- ◆ Inférence statistique: cours 3 à 5

# *Pour résumer (1)*

- ◆ Description des caractéristiques participant.es pour:
  - comprendre la constitution de l'échantillon sur lequel se base l'étude
  - détecter des différences entre les groupes comparés
- ◆ Description des variables importantes pour les comprendre:
  - expositions, critères de jugement
- ◆ Description des associations:
  - pour évaluer l'effet d'une intervention, une relation entre une exposition et un problème de santé,...
  - pour décrire **objectivement** les résultats
  - avec des mesures d'association
  - insuffisante pour répondre à la question de recherche

## *Pour résumer (2)*

- ◆ Description avec:
  - des statistiques (tableau de fréquence, moyenne, écart type,...)
  - des représentations graphiques (diagramme en bâton, box-plot,...)
- ◆ Les méthodes descriptives dépendent de la nature des variables décrites
- ◆ Dans les articles scientifiques médicaux, les représentations graphiques sont utilisées pour appuyer/faciliter les messages les plus importants
- ◆ Ce cours présente les méthodes descriptives les plus courantes et vous trouverez dans des articles d'autres méthodes (en particulier d'autres représentations graphiques)

# Conclusions

- ◆ Description des données est une phase initiale importante de l'analyse des données
- ◆ Statistiques descriptives
  - résumé simple et informatif des données
- ◆ Représentations graphiques:
  - message visuel clair
  - montrer toutes les données
  - rapport information/encre élevé
- ◆ **Un médecin doit savoir interpréter les analyses descriptives**

Message principal  
souvent apparent avec  
des moyens simples

# *Objectifs et lectures pour la prochaine fois*

## ◆ Comprendre les notions suivantes

- Echantillonnage
- Types d'erreurs (aléatoire et systématique)
- Estimation de paramètres
- Intervalle de confiance

Chapitres Petrie/Sabin  
7 – 10 – 11