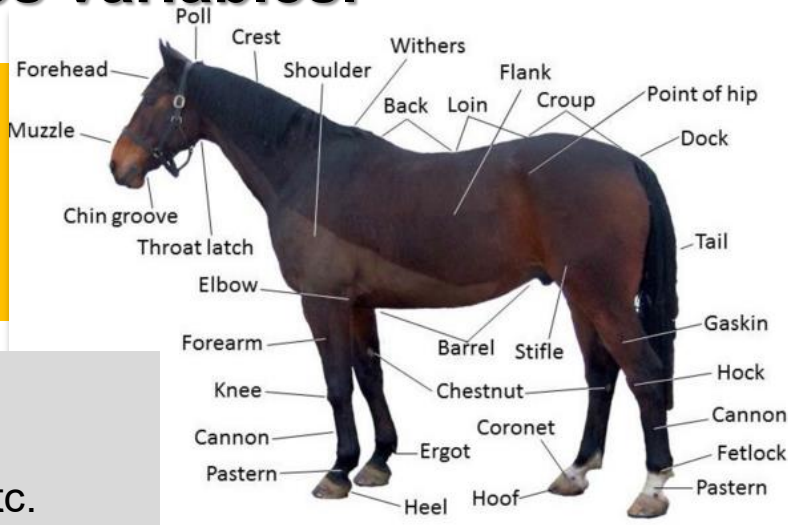
Concepts statistiques essentiels

NOTIONS DE BASE

- (Caractéristiques) variables
- Séries (rangs) statistiques
- Population statistique
- Echantillon statistique
- Randomisation
- Inférences statistiques

Les statistiques étudient des ensembles d'observations effectuées sur des objets appelés unités statistiques présentant certaines caractéristiques variables.

Les traits ou propriétés de l'organisme vivant sont nommés caractéristiques en statistiques et sont inclus dans la catégorie générale des données.



- L'unité statistique: LE CHEVAL
- Caractéristiques variables: âge, couleur, poids, race, etc.

- L'organisme vivant, le système le plus complexe de la nature, ne peut pas être décrit exactement par des mots ou des nombres, autant que nous pouvons en utiliser.
- L'homme n'est capable de décrire que certains traits ou propriétés de l'organisme vivant.
- Nous devons être conscients que même si le nombre de traits du corps qui étudie l'anatomie, la biochimie, la biophysique, la physiologie et toutes les autres disciplines est énorme, nous ne pourrons jamais le décrire exactement.

1 Details

Représentations (codages) d'informations au moyen de symboles ou de signaux

Définition:
La caractéristique est une propriété commune à tous les individus d'une population statistique donnée.

Les caractéristiques sont ce que nous entendons dans la langue commune par attribut, qualité.

Les unités statistiques peuvent être classées, ordonnées ou mesurées par rapport à ces caractéristiques.

La statistique étudie des multitudes d'observations effectuées sur des objets nommés unités statistiques qui présentent certaines caractéristiques variables.

Les traits ou les propriétés de l'organisme vivant sont nommées en statistique caractéristiques et entrent dans la catégorie générale de données.

La caractéristique est une propriété commune à tous les individus d'une population statistique donnée.

Exemple:

Dans un hôpital, les patients (unités statistiques) présentent les caractéristiques suivantes:

- **le groupe sanguin. - peut uniquement être classé (affecté aux classes, aux variantes respectives: A, B, 0, AB).**
- **la note qu'il attribue à la qualité des soins reçus. - peut être classée et ordonnée.**
- **poids à l'admission. - peut être classé, ordonné et mesuré**

Du point de vue de la statistique médicale, les caractéristiques sont de deux types fondamentaux: quantitatives et qualitatives.

Caractéristiques:

Quantitatives (peuvent être mésurées) VALEURS

Qualitatives (peuvent seulement être classées): VARIANTES

Ordinales (peuvent seulement être ordonnées): RANGS

Une caractéristique qui change d'un individu à l'autre ou chez le même individu dans le temps ou comme réponse aux conditions environnementales, de maladie, de traitement, etc, s'appelle variable.

Une variable est un attribut qui décrit une personne ou un objet.
Les variables ont des valeurs qui peuvent varier de la personne ou de l'objet.
La valeur prise par la variable s'appelle **DONNEE.**

TYPES DE CARACTERISTIQUES:

Quantitatif (peut être mesuré)

- **poids**
- **hauteur**
- **nombre de patients,**
- **temps d'hospitalisation,**
- **la durée de fonctionnement d'un appareil,**
- **le coût d'un traitement,**
- **le nombre de jours d'hospitalisation.**

VALEURS

Ordinales (peut seulement être ordonnées):

- note ou qualificatif pour les soins de santé,
- la hiérarchie des médecins dans un hôpital,
- la gravité d'une pathologie, l'état de santé, la performance d'un cadre médical;

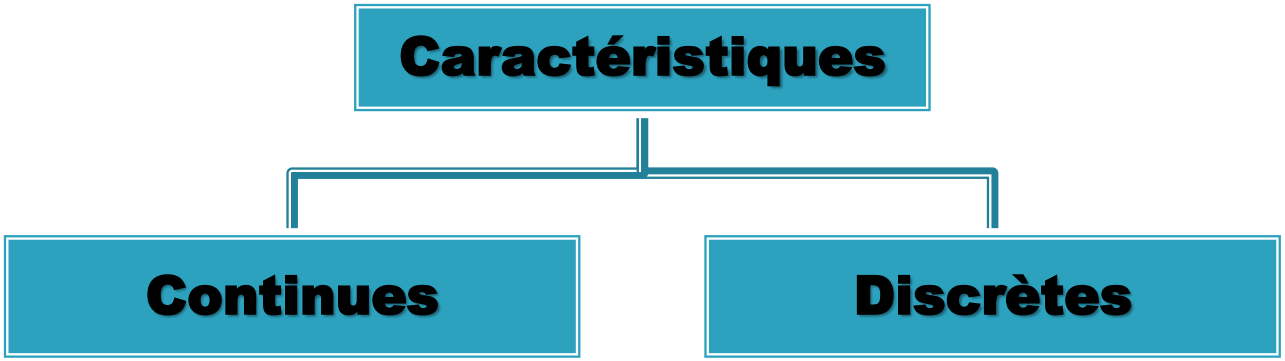
RANGS

Qualitative (peut seulement être classé):

- **groupe sanguin,**
- **RH,**
- **le sexe,**
- **diagnostic médical,**
- **spécialité médicale,**
- **pathologie,**
- **traitement**
- **type de service médical,**
- **hôpital,**
- **médecin**

VARIANTES

Caractéristiques quantitatives



CONTINUES

- Elles ne sont pas limitées à un nombre précis de valeurs (nombres entiers / catégories)
- Théoriquement, elles peuvent prendre un nombre infini de niveaux mesurés
- Elles autorisent l'utilisation de nombres décimaux
- Exemples: poids, taille.

DISCRETES

- Elles sont présentées dans des catégories distinctes
- Elles sont exprimées en nombres entiers (ne prennent pas en charge les décimales)
- Exemples: non. des naissances (veaux) TAS ou valeurs TAD.

Séries statistiques

Les ensembles d'observations s'appellent des **séries statistiques** ou des **chaînes statistiques**.

En fonction du nombre de variables étudiées, une série peut être:

- **uni-variée**, lorsqu'une seule variable est observée
- **bi-variée** lorsque deux variables sont observées;
- **multi-variée** lorsque plusieurs variables sont observées.

Exemples de séries statistiques:

F	M	M	F	F	F	M	M	F	M

Série statistique **UNI-VARIEE**

TAS mmHg	140	110	120	160	130	150	140
TAD (mm Hg	60	80	70	80	90	75	60

Serie statistique **BI-VARIEE** de volume 2 de la pression artérielle

Sexe	f	f	m	f	m	f	f
poids	58	62	85	92	102	75	62
hauteur	1,52	1,65	1,87	1,70	1,92	1,48	1,60
TAS mmHg	140	110	120	160	130	150	140
TAD (mm Hg	60	80	70	80	90	75	60

Serie statistique **MULTI-VARIEE** de volume 5

VARIABLES ET CONSTANTES

La constante

Une quantité qui suppose uniquement une valeur est appelée constante.

Les constantes sont généralement marquées avec les premières lettres de l'alphabet.

Par exemple: valeur $\pi = 3,1415$ ou $e = 2,7182$

Les constants ne font pas l'objet d'étude de la statistique.

La notion de variable statistique

Définition:

Une caractéristique qui change d'un individu à l'autre ou d'un même individu au fil du temps ou en réponse à des conditions environnementales, à une maladie liée à la médication, etc. est appelée variable.

Definition

Une valeur caractéristique de la réalité qui peut prendre différentes valeurs

Âge, poids, état de santé, etc.

Qu'est-ce qu'une variable?

Un attribut décrivant une personne ou un objet.
Des valeurs qui peuvent varier d'une personne à l'autre.

La valeur que prend la variable s'appelle **DONNÉES**.

Exemples de variables (caractéristiques)

Etat de santé

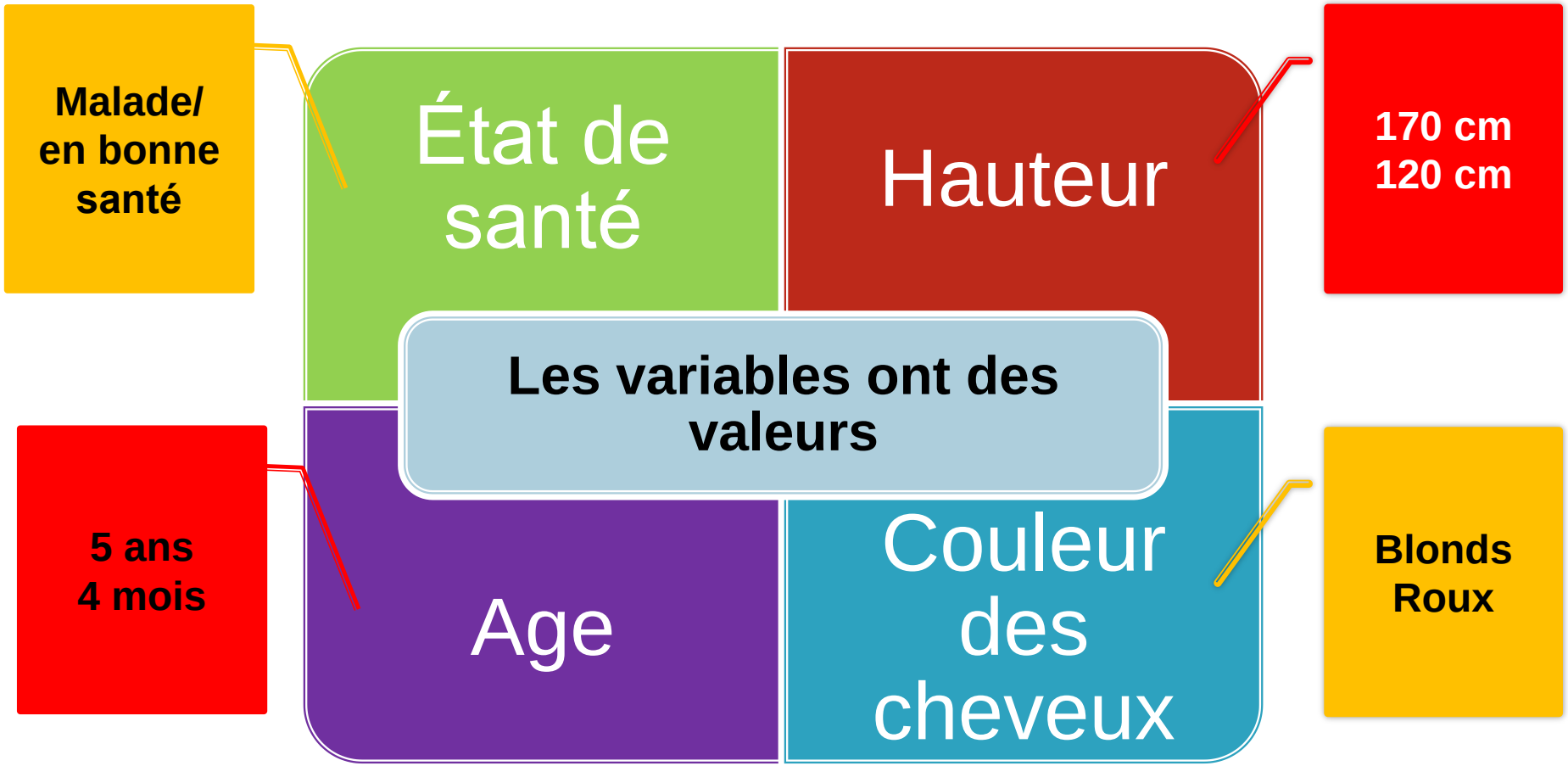
Hauteur

Variables décrivant des personnes

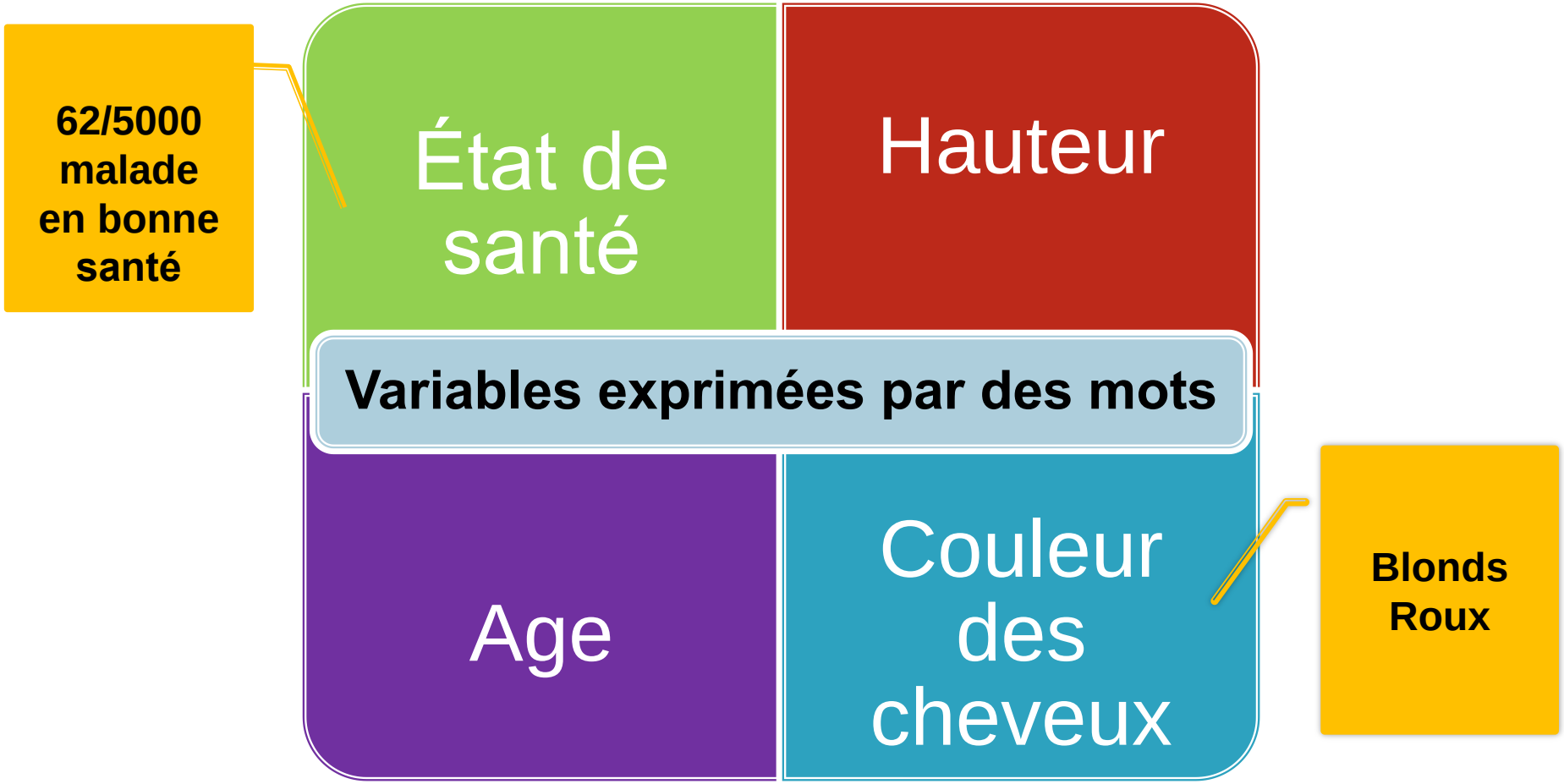
L'unité statistique est la personne

Age

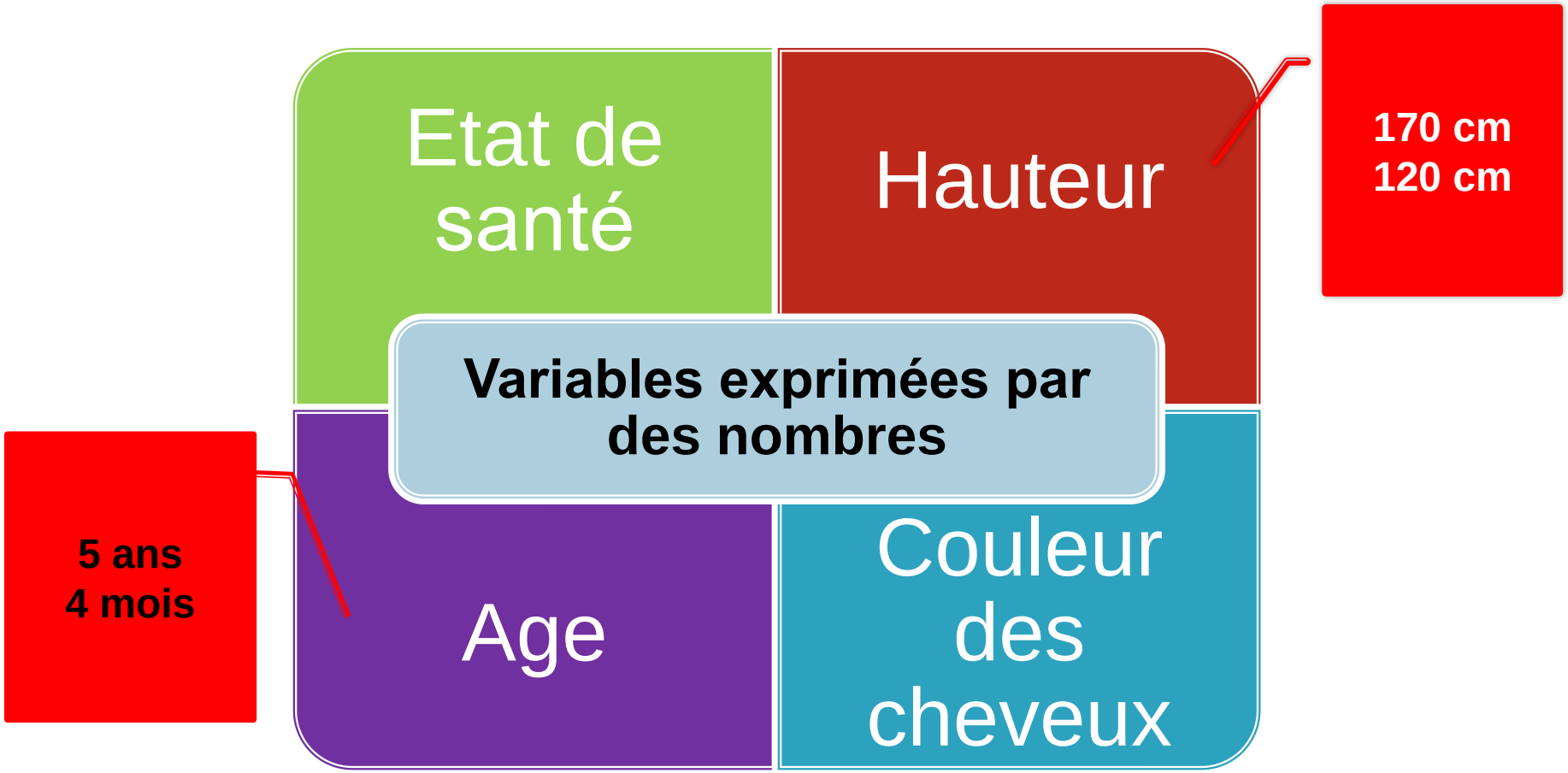
Couleur des cheveux



Les valeurs des variables sont appelées **données**.
Données: malade, en bonne santé, blonde, rouge, 170 cm, 12 cm, 5 ans, 4 ans

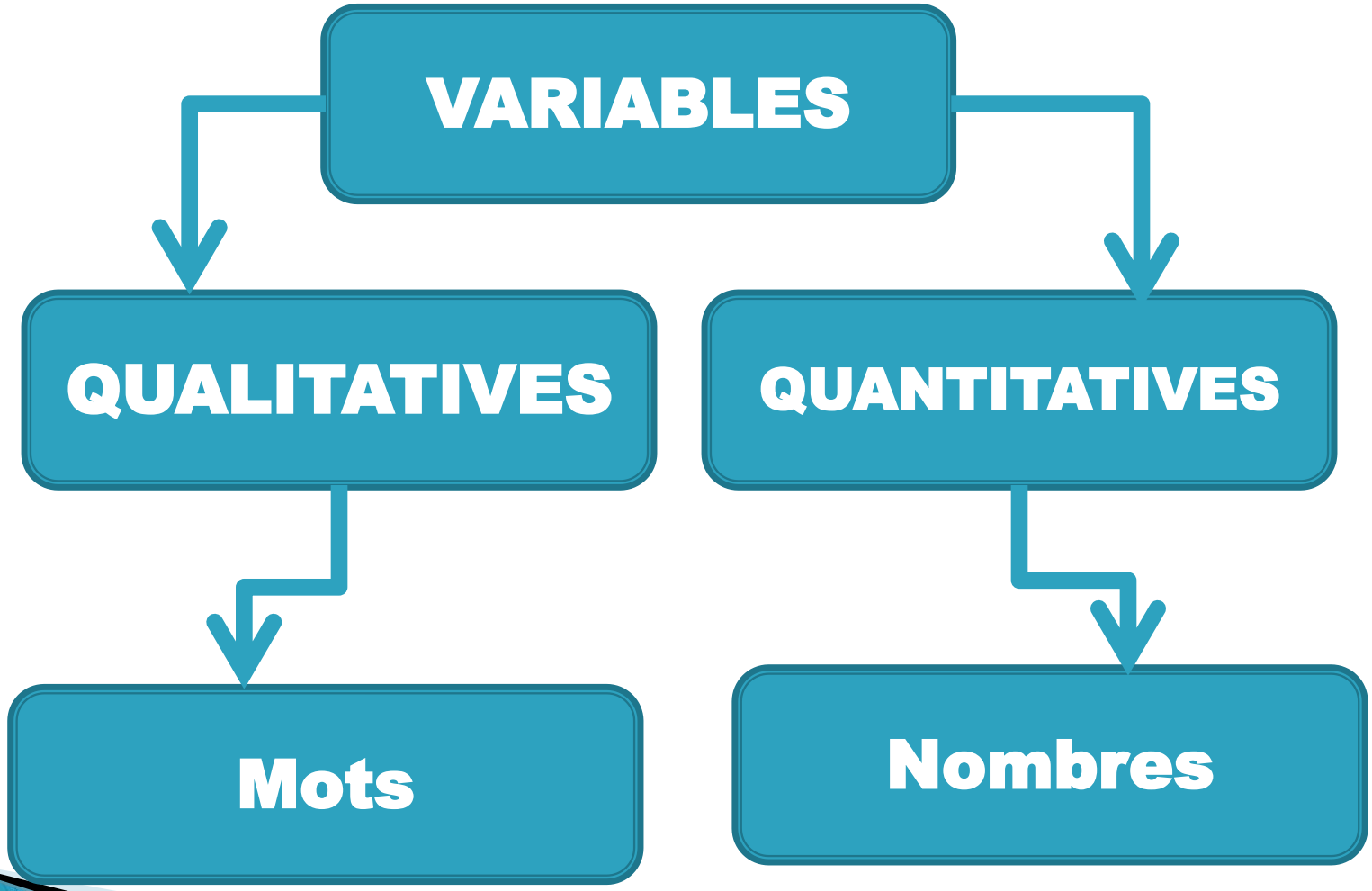


Caractéristiques qualitatives.



Caractéristiques quantitatives

Types de variables



Variables qualitatives ou quantitatives?

Variable

Qualitative

Quantitative

La hauteur

Grand

Petit

Variables qualitatives ou quantitatives?

Variable

Qualitative

Quantitative

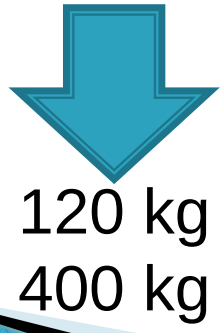
La hauteur

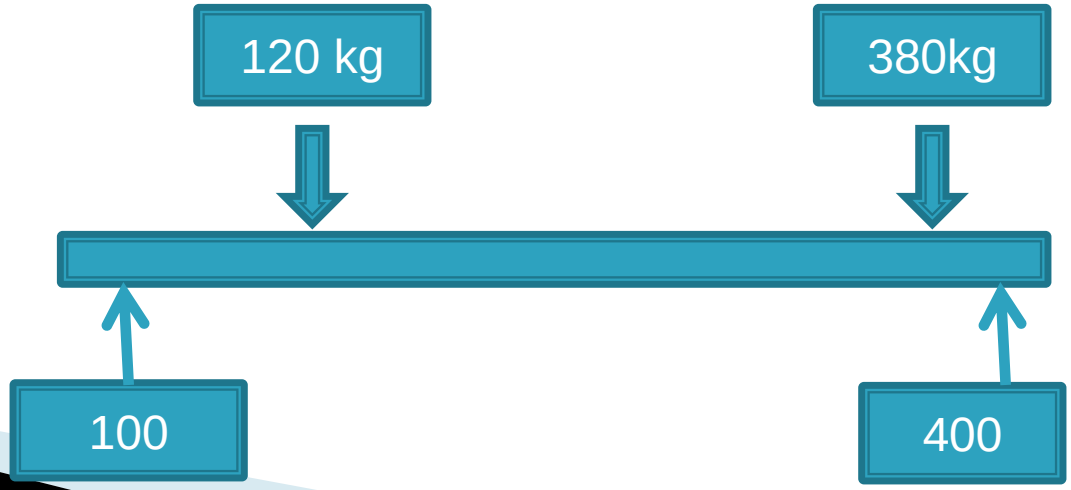
170 cm

120 cm

Variables continues ou discrètes?

Variable	Continue	Discrète
Poids	☑	☐

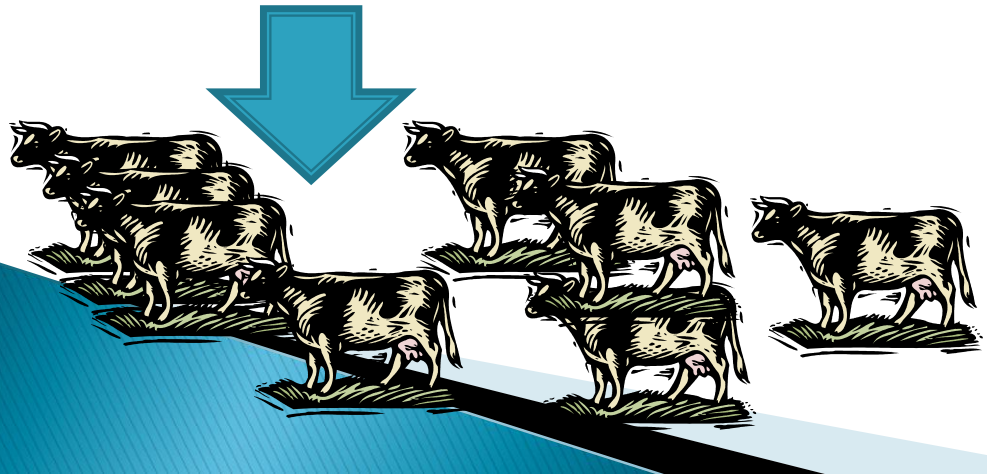




Variables continues ou discrètes?

Variable	Continue	Discrète

Nombre d'animaux

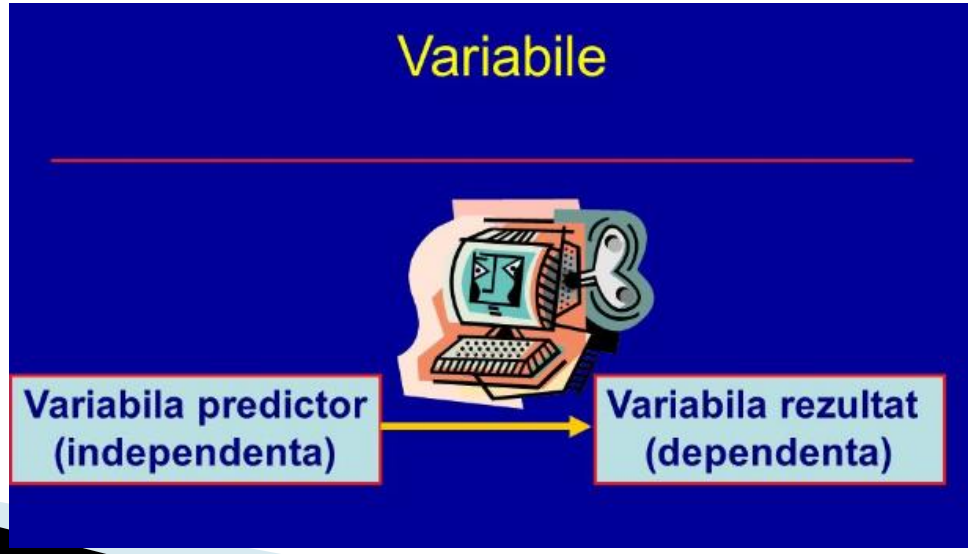


Variables dépendantes et indépendantes

Lorsque deux variables sont liées, elles sont appelées variables associées ou dépendantes.

- L'association peut être positive ou négative.

Si deux variables ne sont pas associées, elles ne se correspondent pas, on les appelle indépendantes.



Variable indépendante et dépendante

La variable indépendante est la variable manipulée par le chercheur, qui est explicitement définie, mesurée ou mesurable.

La variable indépendante est utilisée comme variable d'influence dont les effets éventuels sur la variable dépendante doivent être mis en évidence.

La variable dépendante est la réponse du ou des sujet (s) observée et enregistrée par l'expérimentateur

La variable dépendante (varie en fonction des valeurs d'une ou d'autres variables) fait l'objet d'une mesure dans le but d'être soumise aux conclusions.

En résumé

Unité statistique = élément constitutif de la population statistique

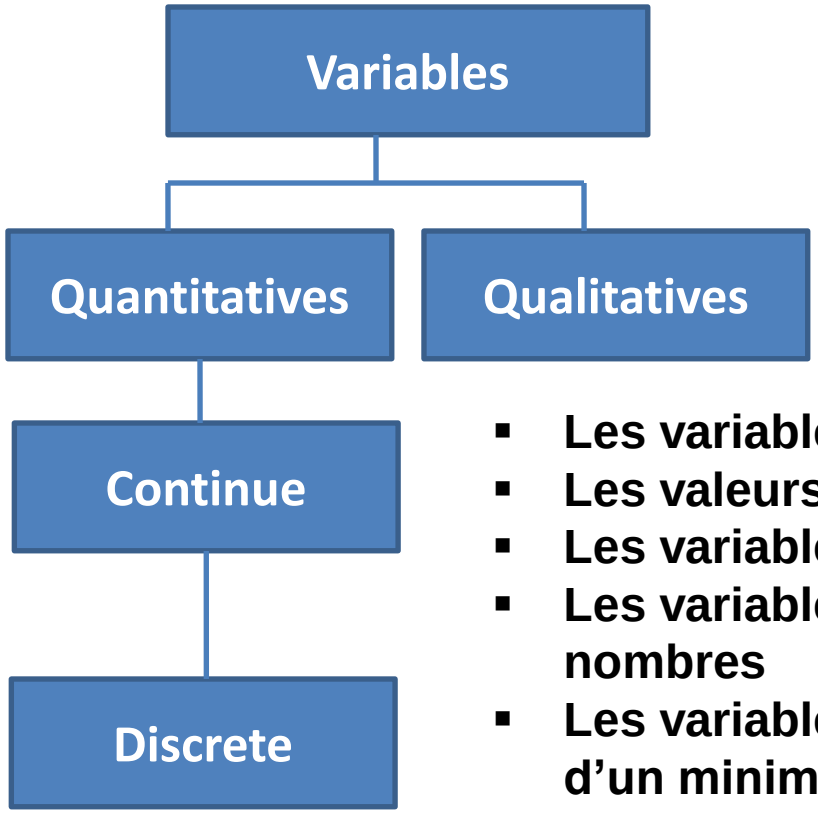
Variable = entité pouvant présenter des variations d'un élément à l'autre dans une collectivité

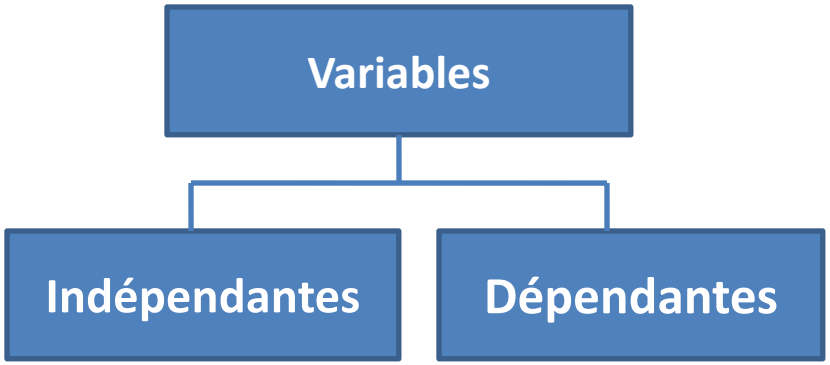
Données statistiques = valeurs observées et enregistrées d'une variable statistique

Informations statistiques = résultats du traitement et de l'analyse des données statistiques

Statistiques descriptives = branche des statistiques traitant de la présentation d'un ensemble de données. La présentation permet de comprendre facilement les caractéristiques des événements étudiés.

En résumé





- Les variables sont des attributs de choses
- Les valeurs des variables s'appellent données
- Les variables qualitatives sont exprimées par des mots
- Les variables quantitatives sont exprimées par des nombres
- Les variables continues peuvent prendre toute valeur, d'un minimum à un maximum
- Les variables discrètes ont seulement des valeurs entières entre une valeur minimale et une autre maximale
- Les variables peuvent être indépendantes ou dépendantes

Les valeurs enregistrées des variables sont le résultat d'une activité de *mesurage*.

Le mesurage **représente un processus systématique de quantification de l'aspect des unités statistiques. Dans le processus de quantification, on utilise un set de règles d'attribution de nombres ou de symboles des unités statistiques.**

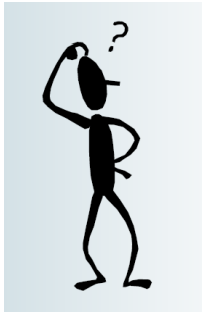
La manière dont on fait l'attribution de ces nombres / symboles s'appelle niveau ou échelle de mesurage.

Parmi les affirmations suivantes, lesquelles sont vraies?

1. Les variables peuvent être classées en variables quantitatives et qualitatives.
2. Les variables qualitatives peuvent être des variables continues.
3. Les variables quantitatives peuvent être des variables discrètes.

La réponse correcte est:

a) Seulement 1
b) Seulement 2.
c) Seulement 3.
d) Correct 1 et 2.
e) Correct 1 et 3.



Lesquelles des variables suivantes sont des variables qualitatives et lesquelles sont des variables quantitatives?

- la hauteur mesurée en centimètres
- le poids mesuré en kg
- le nombre de jours de travail par mois
- nombre de jours ensoleillés
- couleur des cheveux
- sexe
- la valeur moyenne de la température

Dans l'étude des maladies coronariennes, de nombreux facteurs sont associés au développement de cette maladie. Ces facteurs sont connus et sont appelés facteurs de risque.

Lesquelles des variables suivantes sont des variables dépendantes:

a) La maladie cardiovasculaire
b) Le tabagisme
c) L'obésité
d) Le sédentarisme,
e) Le diabète,
f) L'hypercholestérolémie,
g) L'hypertension

Lesquelles des variables suivantes sont indépendantes:

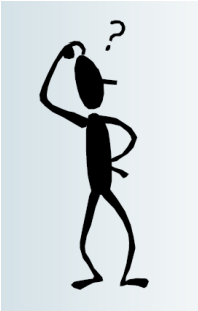
a) La maladie cardiovasculaire
b) Le tabagisme
c) L'obésité
d) Le sédentarisme,
e) Le diabète,
f) L'hypercholestérolémie,
g) L'hypertension

Dans une étude de l'effet de la dextérité sur la capacité sportive, trois groupes ont été formés: droitiers, gauchers et ambidextres.

La capacité athlétique a été mesurée sur une échelle de 12 niveaux.

La variable indépendante est .... et le nombre de niveaux de la variable indépendante est .....

La variable dépendante est .... et le nombre de niveaux de la variable dépendante est .....



a) aptitude athlétique; 3;
b) aptitude athlétique; 12
c) dextérité; 3
d) dextérité; 12

Dans une étude de l'effet de la dextérité sur la capacité sportive, trois groupes ont été formés: droitiers, gauchers et ambidextres.

La capacité athlétique a été mesurée sur une échelle de 12 niveaux.

Lesquelles des affirmations suivantes sont correctes?

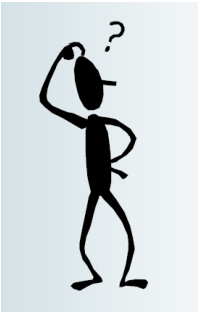
La variable capacité athlétique est:

1. discrète
2. qualitative
3. continue
4. quantitatif
5. Variable dépendante
6. Variable indépendante

Dans une expérience où on étudie l'effet du sommeil sur la mémoire

la variable indépendante est:

1. Le nombre d'heures de sommeil
2. Le résultat obtenu au test de mémoire
3. Le genre des sujets
4. Le genre de l’expérience



Dans une expérience où on étudie l'effet du sommeil sur la mémoire

la variable dépendante est:

1. Le nombre d'heures de sommeil
2. Le résultat obtenu au test de mémoire
3. Le genre des sujets
4. Le genre de l’expérience

variables

quantitatives

- Sont mesurées

qualitatives

- Sont observées

Qu'est-ce que la mesure?

"le processus d'attribution des numéros aux qualités représentées" (N.R. Campell, 1920-267)

- Les valeurs enregistrées des variables sont le résultat d'une activité de **mesure**.
- **La mesure** est un processus systématique de quantification de l'apparence des unités statistiques. Dans le processus de quantification, on utilise un ensemble de règles d'attribution des numéros ou des symboles aux unités statistiques.
- La manière dont ces chiffres / symboles sont attribués s'appelle un **niveau** ou une **échelle de mesure**.

Qu’est-ce que la mesure?

- **Concours de modèles**
  - **hauteur**
  - **buste**
  - **taille**
  - **hanches**
- **beauté**
  - **laides**
  - **sympathiques**
  - **belles**
  - **très belles**

S.S. Stevens (1946) “Sur la théorie d’échelle de mesure”

- mesurer signifie « attribuer des valeurs numériques aux objets et phénomènes, selon certaines règles »
- toutes les mesures ne sont pas numériques
- le genre (sexe) peut être symbolisé par les caractères "m" et "f"
- Une règle d’attribution des nombres a un caractère de mesure seulement si elle établie une certaine signification concernant la caractéristique mesurée.

Niveaux de mesure

- **nominal**
- **ordinal**
- **d'intervalle**
- **de rapport**

Mesure non numérique

Mesure numérique

- **Ils sont ordonnés hiérarchiquement**
- **Les échelles supérieures incluent les propriétés des inférieures**

Le niveau "nominal"

Russell H. Bernard (1995): « La mesure nominale est qualitative »

- Les nombres associés aux objets n'ont pas leur signification quantitative, ils n'expriment pas des tailles: ce ne sont que des signes.
- La règle d'attribution des numéros pour les mesures nominales prévoit uniquement que le même numéro sera attribué à tous les objets ayant les mêmes caractéristiques et uniquement à eux.

Le niveau "nominal" - Exemples

- **Le groupe sanguin, les symboles 0, A, B, AB sont des catégories exhaustives (chaque individu appartient exactement à l'un d'entre eux) et mutuellement exclusifs (un individu ne peut pas entrer simultanément dans deux catégories) et sont de type nominal;**
- **Les mesures nominales consistent à assigner des patients à des groupes ou à des catégories.**
- **Exemple: race, sexe, régime alimentaire, type de tempérament, couleur des yeux, cheveux, etc.**

Le niveau "nominal"

Postulats:

- **Identité [(a = b) ou (a ≠ b), mais pas les deux]**
- **Symétrie [si (a = b), alors et (b = a)]**
- **La transitivité [si (a = b) et (b = c) alors (a = c)**

Opérations mathématiques:

- **Calcul des fréquences (chiffres bruts et pourcentages)**
- **Trouver la médiane**
- **Trouver des quarks**
- **Classement**

Niveau “ordinal”

- Les nombres associés aux objets et aux phénomènes n'ont qu'une signification de rang, n'indiquent pas des tailles absolues.
- La règle d’assignation des nombres spécifie uniquement l’association à des numéros d’ordre croissant ou décroissant d’attributs eux-mêmes classés par ordre croissant ou décroissant.

Niveau “ordinal” – exemples

- **stade évolutif de certaines tumeurs malignes (stades 0, 1, 2, 3 et 4).**
- **est la place occupée par quelqu'un lors d'une certaine compétition (première, deuxième, … dernière).**

Niveau “ordinal”

Postulats:

- **Transitivité de l'inégalité [si (a> b) et (b> c) alors (a> c)]**

Statistiques non paramétriques:

Les coefficients de signification de la différence entre deux moyennes

- **Mann-Whitney**
- **Wilcoxon**
- **Kolmogorov-Smirnov**
- **Kroskal**
- **Wallis**
- **Le test du signe médian**

Niveau “d’intervalle”

- La plage de mesure est assimilée au niveau métrique ou à la mesure d'intervalle.
- Mesure la distance entre les stimuli étudiés

Par exemple:

La valeur de 0 °C correspond à 32 degrés sur l'échelle Fahrenheit.

Les différences comparées sont significatives.

Par exemple, s'il y a une différence de 5 ans entre 1990 et 1995, la même distance est aussi entre 2000 et 2005.

Le niveau “d’intervalle” – exemples

- **TEMPERATURE mesurée en degrés Celsius (les plages entre les valeurs sont égales, mais le point 0 est conventionnellement choisi comme température à laquelle l'eau gèle)**
- **LE COEFFICIENT D'INTELLIGENCE - QI - (si deux personnes ont des scores respectifs de 100 et 150, on peut dire que la différence entre les deux est de 50 points, mais on ne peut pas dire que la seconde est 1/2 plus intelligente que la première ou que le score 0 signifie l'absence d'intelligence).**

Le niveau “d’intervalle”

Postulats:

Les intervalles peuvent être additionnées et soustraites:

[(a-b) = (b-c)], comme (5-3) = (3-1)

Il n'y a pas de zéro absolu exprimant l'absence de la caractéristique mesurée.

Les intervalles peuvent être déplacées ou étendues

Opérations mathématiques:

Les opérations de multiplication ou de division ne sont pas autorisées (une température de 30 °C n'est pas le double d'une température de 15 °C)

Statistiques paramétriques:

- **corrélation**
- **régression**
- **tests de signification**

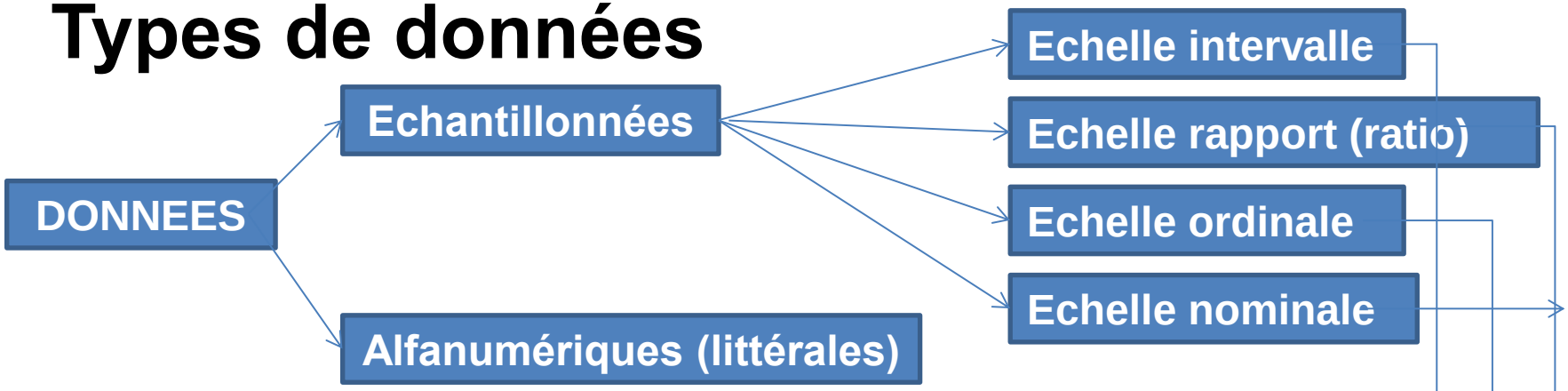
Types de données

Pour enregistrer les données on peut utiliser:

- **Des nombres (on enregistre des données numériques),**
- **Des symboles consacrés ou convenus (on enregistre des données nominales ou des données ordinales),**
- **Des symboles mélangés aux nombres (cas dans lequel on enregistre des données alfanumériques).**

Les données enregistrées comme nombres ou par symboles consacrés ou convenus s'appellent échantillonnées, pendant que le reste sont nommées alfanumériques.

Types de données



Echelle **intervalle** permet à celui qui enregistre des données d'écrire des nombres réels ou entiers.

Echelle **rapport** permet d' enregistrer toujours par nombres réels ou entiers, mais exprime des rapports.

Echelle **ordinale** et échelle **nominale** se ressemblent , permettent d' enregistrer d'après des codes et des abréviations**.**

Par exemple, le groupe sanguin, le stade évolutif d'une affection maligne, le type d' affection, le type de tumeur, etc sont d'habitude standardisés par des conventions internationales à l'échelle continentale ou mondiale.

Echelle nominale

Caractéristiques:

- **Le niveau le plus réduit de mesurage**
- **Les codes des valeurs sont arbitraires**
- **Les valeurs n'ont qu'une signification qualitative**
- **Elle ne supporte pas d' opérations arithmétiques, sauf l'addition**
- **Peuvent être groupées ou raffinées**

Echelle ordinale

Décrit une caractéristique pour l'enregistrement de laquelle on utilise toujours des symboles qui définissent des catégories exhaustives et mutuellement exclusives, mais qui sont ordonnées graduellement, d'habitude croissant.

Exemple: stade évolutif dans certaines affections comme celles malignes (stades 0, 1, 2, 3 et 4).

Les mesurages sur l'échelle ordinale sont ordonnés, ainsi les nombres plus grands représentent des valeurs plus grandes. Pourtant, les intervalles entre les nombres ne sont pas nécessairement égaux.

Echelle ordinale

Caractéristiques:

- **Les valeurs ont une signification quantitative (plus grand/plus petit; plus/moins), mais… elles précisent seulement le rapport de dimension entre les valeurs.**
- **Les intervalles entre valeurs ne sont pas précisés**
  - **que signifie “plus grand”?, que signifie “plus petit”?**
  - **premier admis=9.90; deuxième=9.70; troisième=8.15 !**
- **Les codes des valeurs peuvent être accordés aussi arbitrairement, mais ils doivent exprimer l’idée d’ordre.**

Echelle intervalle

- **L'échelle numérique utilise des constantes comme unités de mesure, et la distance entre ces constantes (niveaux) est égale et permet, part l' ordination des données, le *calcul de la distance* entre niveaux .**
- **L'échelle intervalle *n'a pas de point d' origine unique*.**

EXEMPLES:

- **TEMPERATURE mesurée en degrés Celsius (les intervalles entre valeurs sont égaux, mais le point 0 est conventionnellement choisi – température du gel de l'eau),**

- **COEFFICIENT D' INTELLIGENCE - QI - (si deux personnes ont des scores respectifs de 100 et 150, on peut dire que la différence entre les deux est de 50 points, mais on ne peut pas dire que le second est de 1/2 plus intelligent que le premier ou que le score 0 signifie l'absence d'intelligence).**

Echelle intervalle

Caractéristiques:

- **Les intervalles sont égaux, mais le zéro absolu manque**
- **0°C ne signifie pas l' absence de la température (0°K)**
- **-10°C par rapport à 5°C est "plus grand avec 5°"**
- **-... mais non pas "deux fois plus chaud"!**
- **Supporte toutes les transformations mathématiques possibles**

L'exemple typique est l'échelle pour mesurer la température. L'origine sera différente à l'échelle Fahrenheit par rapport à l'échelle Celsius. Par exemple, 0 degré Celsius est égal à 32 degrés Fahrenheit, ce qui signifie que la température n' a pas un point unique d' origine 0, car l'échelle est de type intervalle.

Echelle rapport

Les données enregistrées sur l'échelle rapport décrivent les caractéristiques du patient qui sont par excellence quantitatives et s'expriment par nombres entiers ou décimaux.

Les mesurages sur l'échelle rapport sont de telle nature qu'une unité sur l'échelle a la même magnitude sur toute l'échelle.

Par exemple, l'enregistrement du poids suppose qu'un poids de 40 kg, est deux fois plus réduit qu'un poids de 80 kg.

Sur l'échelle rapport sont représentées des caractéristiques quantitatives continues et des caractéristiques quantitatives discrètes.

Par exemple, le nombre d'accouchements ou le nombre d' avortements chez une patiente est enregistré numériquement**, discrètement, c'est-à-dire de 1 en 1, les valeurs décimales n'ayant pas de sens. La plupart des analyses de laboratoire au contraire, sont enregistrées numériquement, par un ou deux décimales, d'habitude.**

Echelle rapport (numérique)

Caractéristiques:

-le plus haut niveau de mesurage (les valeurs ont plus d' information)

-supporte toutes les transformations mathématiques possibles

En pratique, la distinction entre les données mesurées sur échelle numérique ou intervalle (ratio) n' est pas relevante, les deux supportant les mêmes procédés statistiques.

En résumé, sur les niveaux de mesurage:

Le niveau le plus réduit

Le plus haut niveau

Echelle	Caractéristiques
Nominale	Dénomination
Ordinale	Ordre de la dénomination
Intervalle	Intervalles égaux sans zéro absolu
Rapport	Intervalles égaux avec zéro absolu

Dénomination de l'échelle	Caractéristiques de l'échelle	EXEMPLES D' UTILISATION
Nominale	Absence de relation d' ordre	Sexe, état civil, service militaire, profession, couleur des cheveux,
Ordinale	Relation d' ordre	stade évolutif dans certaines affections
Intervalle	Relation d' ordre Différences (intervalles significatives) Origine arbitraire Valeur 0 ne signifie pas absence	Température Temps selon le calendrier Pointage obtenu à un test de vérification des connaissances
Rapport	Relation d' ordre Différences (intervalles) significatives, Origine fixe, valeur 0 signifie absence, Opération de multiplication / division	Age, poids corporel

Variables catégorielles

Une variable s'appelle catégorielle lorsque ses valeurs décrivent des catégories.
De manière habituelle, les variables mesurées à échelle nominale sont catégorielles.
-Les variables ordinales peuvent être, elles-aussi, catégorielles.
-Catégories d'âge:

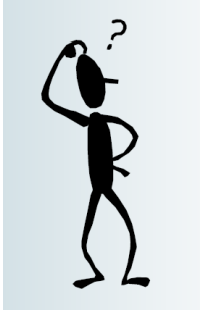
- **1 (21-30 ans)**
- **2 (31-40 ans)**
- **3 (41-50 ans)**

-Les variables quantitatives peuvent être, elles-aussi catégorielles:

- **Catégories de moteurs: 1100 cmc; 1400cmc; 1600 cmc ...**

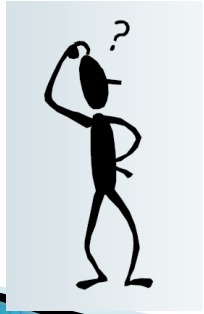
-Il y a des procédés statistiques spéciaux pour les variables catégorielles, donc il faut les retenir!

Identifiez l’échelle de mesurage pour l’épaisseur de la couche de graisse mesuré en mm.



a) **Nominale**
b) **Ordinale**
c) **Intervalle**
d) **Rapport**

Identifiez l’échelle de mesurage pour les grades militaires suivants : capitaine, chef de bataillon (major), colonel, général.



a) **Nominale**
b) **Ordinale**
c) **Intervalle**
d) **Rapport**

POPULATION STATISTIQUE

Les masses sur lesquelles se réalise une étude statistique s'appellent population statistique.

Definition:

LA POPULATION STATISTIQUE EST UNE MASSE D' ÉLÉMENTS QUI:

- PRÉSENTE UN INTÉRÊT À UN MOMENT DONNÉ,
- A UN NOMBRE DE CARACTÉRISTIQUES ESSENTIELLES COMMUNES
- EST SOUMISE À UNE ÉTUDE STATISTIQUE

L'étude de la population se réalise à partir d'une caractéristique commune à tous les individus; ce qui nous intéresse en fait, c'est leur caractéristique commune qu'on étudie, et non pas les individus concrets.

Lorsqu'on parle de populations, on se rapporte, en fait, aux masses de nombres qui représentent des mesurages effectués sur individus humains ou d'autre nature, plantes, objets, évènements, etc.

On la note avec une majuscule du début de l'alphabet: A, B, C etc

Elle peut être finie ou infinie.

Individu statistique

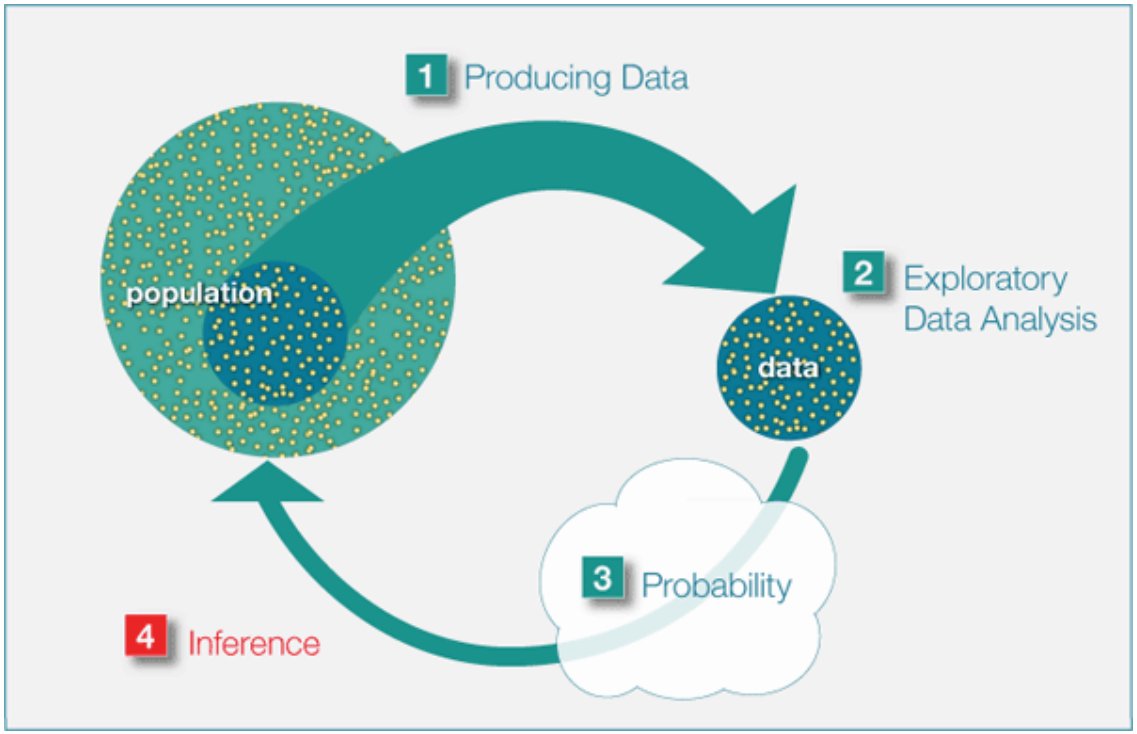
Définition:

On appelera individu statistique un élément d' une population statistique indifféremment de sa nature.

Un individu statistique peut être:

- **Patient (humain, animal, etc.)**
- **Un groupe (un hôpital, un abri, une localité, etc.)**
- **Un évènement**
- **un mesurage, etc.**

Population et échantillon



Dans la pratique statistique on ne travaille pas directement avec la population, car elle contient un grand nombre d'individus (éléments).
Dans certaines recherches on dispose d'une quantité limitée de matériel d'analyse

On extrait une sous-collectivité de la population, elle est connue sous les dénominations: sélection, lot, épreuve ou échantillon

Echantillon

La totalité des "unités d' information" qui constituent l'objectif d' intérêt d'une investigation.
"population" = extension maximale possible, sous l'aspect du volume, de la respective "unité d' information".

"unités d' information" sélectionnées pour être effectivement étudiées.

Les recherches basées sur des échantillons, permettent des appréciations sur toute une population, dans certaines conditions, à base des caractéristiques mesurées sur une de celle-ci.

Population vs. Echantillon

- **Les attributs de la population sont nommés paramètres**
- **Les attributs de l'échantillon sont nommés statistiques**

Exemples:
Une caractéristique mesurable d'une population, par exemple une déviation moyenne ou standard, s'appelle paramètre;
Une caractéristique mesurable d'un échantillon est appelée statistique.

Notation:
Pour les paramètres: lettres grecques (par exemple σ, μ)
Pour les statistiques: lettres latines (par exemple, **s, m**)

Selection de l'echantillon de la population

METHODE DE SELECTION:

- **échantillonnage aléatoire** (probabiliste);
- **échantillonnage non-aléatoire** (empirique, au hasard) ou à base de raisonnement.

Echantillonnage (sélection) = set d' opérations avec l'aide desquelles, de l'ensemble de la population visée par la recherche, on choisit une partie échantillon, qui sera soumise à l'investigation.

Echantillonnage aléatoire

- chaque unité de la population recherchée a une certaine probabilité d'être incluse dans l'échantillon, probabilité différente de zéro, qui peut être connue à priori.
- chaque unité de la population doit avoir la même chance que toutes les autres de figurer en échantillon.

Cela mène à la constitution de l'échantillon par tirage au sort.

Echantillonnage non-aléatoire

- **Technique d'établir l'échantillon qui suppose la méconnaissance de la probabilité d'inclure en échantillon les individus de la collectivité .**
- **La sélection a donc un caractère arbitraire, subjectif, et se base, tout d'abord, sur le jugement personnel du chercheur, supposant un "choix raisonnable".**

Représentativité de l'échantillon

L'échantillon doit posséder une qualité essentielle = représentativité (capacité de reproduire le plus fidèlement les structures et les caractéristiques de la population dont il est extrait).

Un échantillon est représentatif lorsque les scores dans l'échantillon ont les mêmes caractéristiques que celles de la population dont il a été extrait.

Détermination du volume de l'échantillon

Le volume de l'échantillon dépend de trois éléments:

1. **Erreur maximale admise** = Précision (Dx);
2. **Probabilité de garantir les résultats** = Coefficient de confiance en réponses (p);
3. **Dispersion** = Niveau de la dispersion de la population étudiée (σ^2).

On lit : "avec une probabilité p on garantit les résultats à une erreur maximale admise de +/-e%".

Le volume de l'échantillon doit être établi apriori, et non pas en fonction des possibilités de moment ou des contraintes de temps et d' espace.

On observe que le volume de l'échantillon ne dépend pas du volume de la collectivité générale (N), lorsque celui-ci prend des valeur supérieures.

Prob./Er.max.	+/-1%	1,5%	2%	2,5%	3%	3,5%	4%	4,5%	5%
95%	9603	4268	2400	1536	1067	784	600	474	384
99%	16641	7395	4160	2262	1849	1359	1040	821	665

Etablir la dimension de l'échantillon (combien sélectionner ?)

1. **Le principe fondamental – nombre de participants considéré acceptable pour former un échantillon représentatif dépend du type de recherche.**
2. **Pour les recherches correlationnelles 30 participants suffisent pour former un échantillon représentatif (on accepte que, à partir de 30 sujets, la distribution est normale).**

Dimensions proposées de l'échantillon
Pour la probabilité de garantir les résultats à 95%

Erreur maximale admise	Volume échantillon
1	9604
1,5	4268
2,5	1537
3	1067
3,5	784
4	600
4,5	474
5	384
5,5	317
6	267
6,5	227
7	196
7,5	171
8	150
8,5	133
9	119
9,5	106
10	96

“There are three kinds of lies: lies, damned lies, and statistics.”

Il y a trois catégories de mensonges: Mensonges, mensonges insolents et statistiques

Benjamin Disraeli (1804 - 1881)

“Statistics: The only science that enables different experts using the same figures to draw different conclusions.”
La statistique: c’est la seule science qui permet aux experts d’utiliser les mêmes chiffres pour tirer des conclusions différentes.

Evan Esar (1899 - 1995)

“Medical students may not like statistics, but as doctors they will.”
Martin Bland, Letter to the Editor, 1998. BMJ; 316:1674.
“Les étudiants en médecine n’aiment pas la statistique, mais en tant que docteurs, ils sont obligés de la faire”