

# Chapitre 1: Notions de base

Arnout Van Messem

Bachelier en Sciences informatiques

# Chapitre 1: Notions de base

- La population
- Les variables
- Les observations et les données
- Transformation des données (% , proportion et taux)

# La population

La population représente l'ensemble des unités d'intérêt.

- Etudiants de premier bachelier à l'ULiège
- Les communes de Wallonie
- Les ampoules produites dans une telle usine

Remarque: les éléments ne sont pas toujours des êtres humains

Définitions:

- Eléments = unités statistiques ou individus
- Nombre total d'individus = effectif ( $n$ )

# Les variables

Les variables correspondent à des critères permettant de différencier les individus d'une même population.

- Exemple: parmi les étudiants de premier bachelier, les filières d'étude choisies diffèrent, les lieux de résidence varient,...
- Notation:  $X, Y, \dots$ . On parle des valeurs ou des modalités d'une variable.
- Si les modalités sont **mutuellement exclusives** et **exhaustives**, celles-ci permettent le classement des individus de la population.
- Les variables statistiques sont **qualitatives** (constatées) ou **quantitatives** (mesurées ou énumérées).

# Variables qualitatives

- Les modalités sont appelées attributs ou catégories.
- Exemple: Genre (M ou F), Fumeur (OUI ou NON), Préparation à l'examen (TB, B, M, I), Commune de résidence (*Chaudfontaine, Esneux,...*),...
- Cas particulier: Une variable qualitative avec 2 catégories est une variable dichotomique.
- Classification des modalités sur une échelle **nominale** ou **ordinaire**.

# Echelles nominale et ordinale:

- Echelle nominale: Il n'y pas d'ordre naturel entre les modalités.

**Ex:** couleur des yeux, genre,...

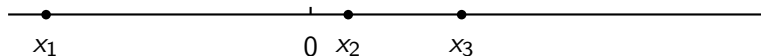
- Echelle ordinale: les catégories peuvent être ordonnées.

**Ex:** Préparation à l'examen, appréciation d'un professeur,...

Mais, il n'y a pas d'indication sur l'**écart** entre deux niveaux successifs de l'échelle.

# Variables quantitatives

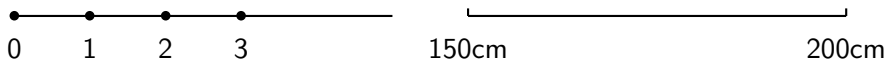
Les modalités des variables quantitatives ont des valeurs numériques  
 $x_1, x_2, \dots$



- Exemples: Points à un examen ( $0, 1, 2, \dots, 20$ ), Taille ( $170\text{cm}, 174\text{cm}, \dots$ ), Nombre d'enfants dans une famille,...
- Choix de l'unité et/ou précision de l'instrument de mesure: m ou cm, dollars ou Euros,...
- Caractère **discret** ou **continu** .

# Caractère discret ou continu

- Caractère discret: les valeurs possibles peuvent être énumérées (procédé de comptage).  
**Ex:** Nombre d'enfants
- Caractère continu: les valeurs sont mesurées.  
**Ex:** Taille, Poids, Revenus,...
- Il y a des sauts entre les valeurs possibles d'une variable discrète tandis que les valeurs d'une variable continue peuvent tomber n'importe où dans un intervalle donné.



- La distinction entre les types discret et continu n'est pas toujours évidente:  
**Ex:** Age, Résultat d'un examen sur 100 points,...



# Echelles d'intervalle ou de rapport

- Echelle d'intervalle:

- ▶ Les valeurs sont ordonnées
- ▶ La différence entre deux valeurs a un sens
- ▶ Les rapports de deux valeurs n'ont pas de sens
- ▶ Le 0 est arbitraire ( $\neq$  absence du phénomène).

**Ex:** Date de naissance

- Echelle de rapport:

Les différences et rapports entre les valeurs ont un sens (la valeur 0 correspond à un vrai 0).

**Ex:** Taille, Revenu, Nombre d'enfants,...

- Toute variable quantitative peut aussi être analysée sur une échelle qualitative ordinale

**Ex:** "Petits", "Moyens" et "Grands"

# Les observations et les données

Les résultats observés d'une ou plusieurs variables sur une population constituent les observations. Si une enquête porte sur  $p$  variables et  $n$  individus, l'ensemble des observations récoltées (les données) se présente sous la forme d'un tableau à  $n$  lignes et  $p$  colonnes.

Individus	Variables				
	1	...	$j$	...	$p$
1	$x_{11}$	...	$x_{1j}$	...	$x_{1p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$x_{i1}$	...	$x_{ij}$	...	$x_{ip}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$x_{n1}$	...	$x_{nj}$	...	$x_{np}$

# Population ↔ échantillon

Lorsqu'une variable est mesurée sur tous les individus de la population, on parle de *recensement* de la population.

Depuis toujours, les états organisent de tels recensements démographiques. Au départ, ceux-ci servaient à compter le nombre d'habitants de chaque commune et à déterminer leurs richesses mais, aujourd'hui, ils sont plutôt considérés comme des outils d'aide à la prévision économique.

En Belgique:



- Premier recensement en 1846 par Adolphe Quetelet
- Puis, un recensement tous les dix ans (1856,... avec parfois des décalages à cause d'une guerre)
- Le dernier recensement a été organisé en 1991 puis la procédure a été abolie suite à une modification de la loi.
- Depuis 2001, un *Census* (= utilisation des BD officielles + enquêtes ponctuelles) est organisé.

# Population ↔ échantillon

Dans la plupart des études cependant, seuls certains individus de la population sont considérés. Ils constituent un *échantillon* que l'on espère *représentatif* de la population visée.

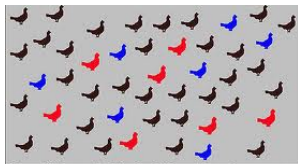


Figure 2. Exemple d'échantillons : les oiseaux en bleu constituent un échantillon, ceux en rouge, un autre



Il y a beaucoup de raisons expliquant pourquoi, en pratique, seul un échantillon est considéré: contraintes de temps ou de budget, destruction de l'individu,...

# Données exploitées dans les notes de cours

## A. Communes de Wallonie:

Sources: [www.lavenir.net](http://www.lavenir.net) et

<http://statbel.fgov.be/fr/statistiques> (site du SPF Economie).

- PartiBourgmestre: le parti du Bourgmestre de la commune (modalités CdH, Ecolo, MR, PS et Autre).
- NombrePartis: nombre de partis se trouvant dans la majorité communale.
- Chomage: le taux de chômage dans la commune
- PrixMaison: prix moyen des maisons vendues sur le sol de la commune lors de l'année 2012.
- IndiceRichesse: indice caractérisant la richesse d'une commune par rapport à une valeur de référence égale à 100

B. Nombre de **déclarations et montants du revenu total net imposable** en Belgique pour les années 2002 (les années 2007 et 2012 sont ajoutées dans les transparents).



# Transformation de données

Les valeurs prises par des individus particuliers ou par un ensemble d'individus sont très souvent exploitées pour effectuer des comparaisons entre différentes populations ou entre différentes époques.

Une base commune de comparaison doit être choisie pour que la comparaison soit honnête. On parle alors en termes de

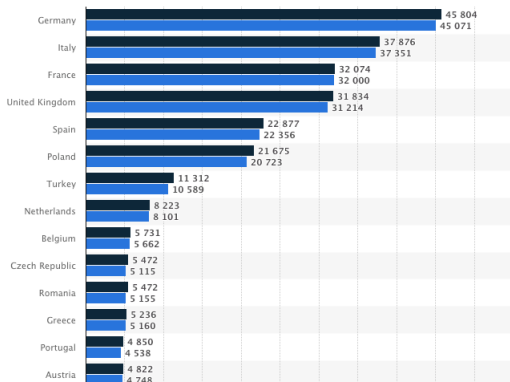
- Proportion (base 1)
- Pourcentage (base 100)
- Taux (base  $10^n$ )

Par ailleurs, des comparaisons temporelles doivent être réalisées en termes relatifs et non absolus.

# Exemple: comparaison de statistiques européennes

Comparaison du parc automobile dans les pays Européens:

**Number of registered passenger cars in Europe from 2015 to 2016, by country (in 1,000 units)**



DOWNLOAD

SETTINGS

SHARE



DESCRIPTION

SOURCE

MORE INFORMATION

This bar chart shows the stock of passenger cars in the European Union and other selected countries in Europe from 2015 to 2016, by country. The total number of registered passenger cars increased in all of the countries listed over this period. In 2016, the total number of registered passenger cars in Europe stood at 259.5 million units.



La comparaison des valeurs brutes n'a pas de sens car les effectifs des populations totales ne sont pas les mêmes.

# En tenant compte des populations:

## Vehicles Per Capita, by Country

The European Union counts 587 passenger cars per 1,000 inhabitants.

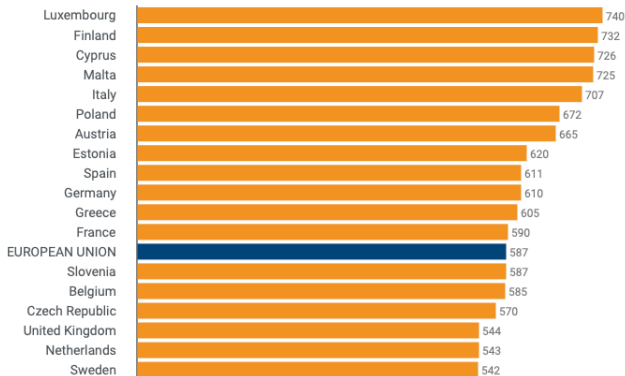
### Motorisation rate in the EU in 2016

Number of vehicles per 1,000 inhabitants



ACEA

European  
Automobile  
Manufacturers  
Association





## Comment passer d'une mesure à l'autre?

Soit une population (au sens large) de  $n$  individus et soit une caractéristique d'intérêt (que vérifient ou pas ces individus).

Par définition, une proportion est la part de la population vérifiant la caractéristique (par rapport à une base 1):

$$\text{Proportion} = \frac{\text{Nombre d'unités ayant la caractéristique}}{\text{Effectif total de la population}}.$$

Le pourcentage s'obtient en multipliant la proportion par la nouvelle base 100

$$\% = 100 \times \text{proportion}$$

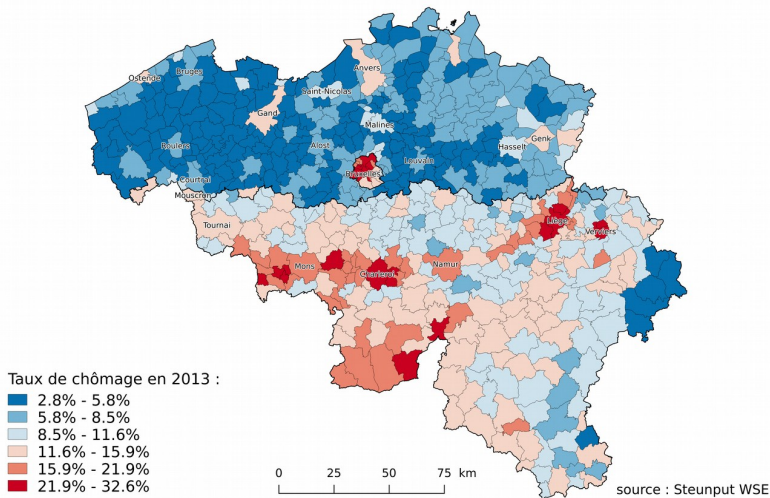
Le taux en base  $10^n$  (1,10,100,1000,...) est donné par:

$$\text{taux} = \text{prop} \times 10^n.$$

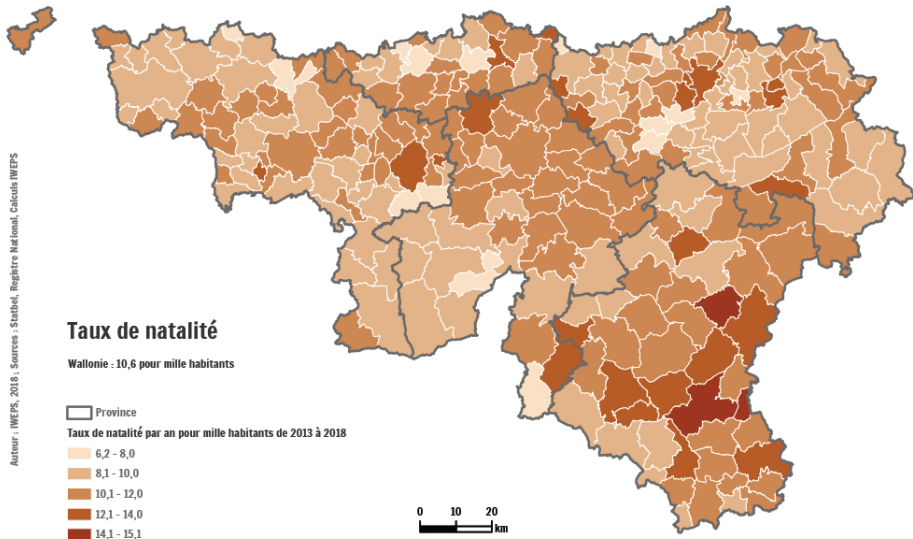
Le choix de la base dépend soit d'une convention, soit de la fréquence d'occurrence de la caractéristique.

# Exemple: statistique de chômage en Belgique

Taux de chômage par commune (2013)

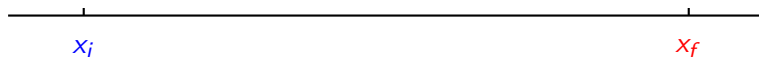


# Exemple: taux de natalité en Belgique (source: Iweps)

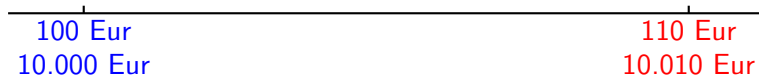


# Pourcentage de variation

Calculer un pourcentage de variation permet de caractériser l'évolution temporelle des valeurs d'une variable



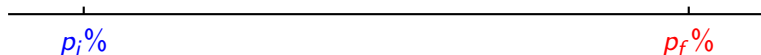
Exemple:



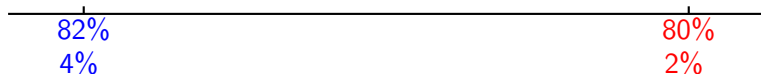
Augmentation brute: 10 Eur

# Pourcentage de variation

Calculer un pourcentage de variation permet de caractériser l'évolution temporelle des valeurs d'une variable (ou des pourcentages d'une modalité particulière de celle-ci).



Exemple:



Diminution de 2%

# Définition du pourcentage de variation

$$\% \text{ de variation} = \frac{\text{valeur finale} - \text{valeur initiale}}{\text{valeur initiale}} \times 100.$$

Il mesure donc le pourcentage d'augmentation ou de diminution qu'une variable a subi dans le temps, en relativisant la variation brute en fonction de la valeur initiale.

## Exemple et interprétation dans le cas d'une variable à valeur positive

Si le pourcentage de variation est positif (resp. négatif), cela signifie qu'il y a eu une augmentation (resp. diminution) de la valeur entre les deux périodes.

- Evolution de la valeur du Bel 20 entre le 30 janvier 2012 et le 31 janvier 2012 à la clôture de la bourse de Bruxelles:

Valeur initiale = 2.197,08 Euro

Valeur finale = 2.206,80 Euro +0,4%

Le 2 février 2011: 2.677,80 -17.5%

- Evolution du pourcentage de ménages wallons habitant dans une maison individuelle entre 1998 et 2001:

Valeur initiale = 40,3%

Valeur finale = 35,2 % -12.65%



[Accueil](#) > [Économie](#) > [Mobilité](#)

## Bruxelles: le nombre de cyclistes a augmenté de 13% en un an, c'est une «excellente nouvelle»

Pour le ministre bruxellois de la Mobilité, Pascal Smet, cela démontre que le vélo est resté crédible aux yeux des nouveaux usagers.



Mis en ligne le 20/01/2018 à 16:53





# Ce qu'il faut retenir de ce chapitre

- Population (ou échantillon):  $n$  individus
- Variables statistiques: caractéristiques des individus
  - ▶ Types de variables (quali, quanti discret ou continu)
  - ▶ Echelles des variables (nominale, ordinale, d'intervalle ou de rapport)
- Comparaison de populations à l'aide de proportions (base =1), de pourcentages (base = 100) ou de taux (base à définir);
- Exploitation des pourcentages
- Evolution temporelle à l'aide de pourcentages de variation.

Pour le 08/02 (deadline 23h30):

- Installer le logiciel R
- Faire le TEST 1 (notions de base, y compris de R)