

Statistique

Concepts de base

Population et échantillon

Deux concepts essentiels:

- **Population:** l'ensemble de tous les objets sur lesquels porte une étude statistique
- **Échantillon:** un sous-ensemble des objets de la population

Exemple:

- La population: $P = \{\text{tous les pays du monde en 2015}\}$
- L'échantillon $E = \{\text{pays du G7}\} \subset P$

Population

C'est un concept très général, par exemple:

- Tous les canadiens
- Tous les pays
- Toutes les voitures immatriculées au Québec

Dans certains cas, la population change dans le temps!

Le choix de la population est intimement lié au but de l'étude statistique. Il faut être précis, sinon on peut porter des jugements ambiguës ou erronés.

Variable aléatoire

En statistique, on étudie une population via les valeurs prises par une variable qui caractérise un aspect

Puisque la variable est sujette aux effets du hasard, on la nomme **variable aléatoire (v.a.)**

On utilise souvent un symbole pour représenter une v.a., par exemple:

- g : le PIB d'un pays
- t : le taux d'endettement d'un pays en % du PIB
- v : la valeur d'un actif financier

Exemple - taux d'endettement

On s'intéresse au taux d'endettement des pays

Population (P): les pays du monde

Échantillon (E): les pays du G7 en 2014

$E = \{\text{Allemagne, Canada, E-U, France, Italie, Japon, UK}\}$

Remarques:

- P est finie et peut changer dans le temps
- L'échantillon n'est pas aléatoire

Exemple - suite

Deux variables aléatoires pertinentes:

- t: “endettement en % du PIB à la fin du 2Q14”
- c: “% de changement de la dette de 2007 à 2014”

Chacune des v.a. prends sept valeurs sur l'échantillon:

pour x: 188, 221, 233, 280, 259, 400, 252

pour c: 8, 39, 16, 66, 55, 64, 30

Remarque: ne pas confondre l'échantillon, qui est un sous-ensemble de la population, avec l'ensemble des valeurs prises par une v.a sur les éléments de l'échantillon

Exemple - suite

Pays du G7	Dette en % du PIB	Changement 2007 - 2014
Allemagne	188	8
Canada	221	39
E-U	233	16
France	280	66
Italie	259	55
Japon	400	64
U-K	252	30

Référence: Debt and deleveraging, MGI, 2015.

Valeurs d'une v.a.

Rappelons que les éléments d'un ensemble n'ont pas d'ordre et ne sont pas répétés

Pour une v.a.:

- une valeur peut apparaître plusieurs fois
- l'ordre de présentation est important

C'est donc une suite et non un ensemble

On présente souvent les suites de valeur des v.a. conjointement avec les objets de l'échantillon dans un tableau (comme dans l'exemple précédent)

Choix des variables aléatoires

Une v.a. peut être

- directement liée à une observation
- obtenue par un modèle combinant d'autres v.a.

Une v.a. peut-être vue comme une fonction **x** évaluée **sur** les objets d'un échantillon **E** et prenant des valeurs **dans** un ensemble de référence **K**

$$\mathbf{x} : \mathbf{E} \rightarrow \mathbf{K} : \mathbf{e} \mapsto \mathbf{x}(\mathbf{e})$$

où **E** est un sous-ensemble de la population **P**

Étude statistique

Une étude statistique porte sur des aspects d'une population P vue au travers des v.a.

Le plus souvent, on veut:

- porter un jugement probabiliste: “on a tant de chances que tel événement se produise”
- porter un jugement général: “cet événement est anormal”
- identifier des liens: “si x change, alors y change”
- prédire le changement: “si x augmente de 10%, alors y va diminuer de 5%”

Attention danger!

On entend régulièrement “on peut faire dire n’importe-quoi aux statistiques...” C’est souvent lié à un manque de rigueur dans la définition de x , E , P , K et le but d’une étude

Exemple: étude statistique sur la chance du débutant

x : “aviez-vous plus de chance à vos début au casino?”

K : { oui , non }

E : les joueurs au Casino de Montréal présent à l’une des dates d_1 , d_2 , d_3

Résultat : 527 oui, 402 non

Conclusion : la chance du débutant est une réalité!

Hummm... Est-ce convainquant? Justification?

Aide: invisible évidence!

Outils mathématiques

Une étude statistique exploite des outils mathématiques qu'on classe en 3 groupes:

1. la **statistique descriptive**: on caractérise la distribution des valeurs prise par une v.a. au sein d'un échantillon provenant d'une population
2. l'**inférence statistique**: on exploite la distribution des valeurs d'une v.a. au sein d'un échantillon pour caractériser la distribution de cette même v.a. au sein de toute la population (généralisation)
3. **probabilité**: on utilise cette théorie mathématique pour porter des jugements à partir de la distribution des valeurs d'une v.a. au sein d'une population

Outil mathématique vs monde réel

