
My sample book

The Jupyter Book Community

Aug 18, 2022

CONTENTS

This is a small sample book to give you a feel for how book content is structured. It shows off a few of the major file types, as well as some sample content. It does not go in-depth into any particular topic - check out [the Jupyter Book documentation](#) for more information.

Check out the content pages bundled with this sample book to see more.

- À propos du cours
 - *Introduction*
- Statistique Exploratoire
 - *Types de données*
 - *Graphiques*
 - *Synthèses numériques*
 - *Boxplot*
 - *Loi normale*
 - *Quelques principes généraux*
- Probabilité
 - *Concepts de base*
 - *Arrangements et combinaisons*
 - *Probabilité conditionnelle et indépendance*
 - *Probabilités totales et théorème de Bayes*
 - *Variables aléatoires discrètes*
 - *Variables aléatoires continues*
 - *Variables aléatoires conjointes*
 - *Valeurs caractéristiques*
 - *Théorème fondamentaux*
 - *Test live code*
- Statistique inférentielle
 - *Introduction*
 - *Estimation de paramètres*
 - *Propriétés d'un estimateur*
 - *Estimation par intervalle*
 - *Tests d'hypothèses statistiques*
 - *Tests et IC*
 - *Test du chi2*
 - *Puissance d'un test*
- Régression linéaire
 - *Introduction*
 - *Régression linéaire: cas général*
 - *Régression linéaire: tests*

- *Régression linéaire: hypothèses et diagnostics*

Part I

À propos du cours

INTRODUCTION

1.1 Statistiques: une définition ?

Les mathématiques (du grec *Mathema* \approx apprendre) sont une manière:

1. d'exprimer une grande variété de notions complexes avec précision et cohérence.
2. de « *légitimer les conquêtes de notre intuition* » (Jacques Hadamard)¹ : raisonner rigoureusement à partir d'hypothèses, tirer les conclusions correctes d'une observation, d'une expérience, etc.

La statistique souffre d'un problème de nomenclature. On a deux « statistiques » (définitions TLF):

1. « Recueil de données numériques concernant des faits économiques et sociaux »; par exemple des données démographiques (répartition en âge, métiers, etc. de la population), ou économiques (taux de chômage, salaire médian, etc.). « L'ensemble des connaissances que doit posséder un homme d'État », introduit en allemand *Statistik* par l'économiste G. Achenwall (1719-1772) (de l'italien *statista*, homme d'État).
2. « Branche des mathématiques ayant pour objet l'analyse et l'interprétation de données quantifiables. »

Utiliser les maths
pour
extraire des informations
à partir de
données
en présence
d'incertitudes

Les données sont absolument partout de nos jours.

1. OFS: démographie, chômage \rightarrow décisions politiques;
2. science: expérience \rightarrow données \rightarrow conclusion;
3. utilisation d'internet: « cookies » \rightarrow publicités ciblées.

¹ La citation exacte est « La rigueur n'a d'autre objet que de sanctionner et de légitimer les conquêtes de l'intuition ».

1.2 Et les probabilités ?

Les probabilités nous aident à appréhender **l'incertitude**; elles permettent de la transcrire en un formalisme mathématique.

1. C'est la discipline qui étudie les phénomènes aléatoires (ou *stochastiques*).
2. C'est la base indispensable à toute étude mathématiquement rigoureuse de ces phénomènes.

Les probabilités nous donnent **le formalisme** dans lequel on peut comprendre et quantifier l'effet que la présence d'incertitude dans les données a sur notre analyse de ces données.

1.3 Le but de la statistique

L'expérience montre que de nombreuses expériences sont *intrinsèquement* aléatoire: jet de dé, un tirage au sort, une campagne de vaccination. D'autres ne le sont peut-être pas intrinsèquement, mais il s'avère impossible de les reproduire exactement: tir au panier, mesure physique etc. Ainsi, *le hasard est une composante essentielle de bien des expériences*. La plupart du temps, le but de la statistique est de **comprendre ce hasard**.

On peut identifier quatre étapes majeures de la démarche statistique:

1. planification de l'expérience; (développement théorique du problème, élaboration du plan expérimental);
2. collecte des données;
3. **analyse des données**;
4. présentation des résultats et conclusions / actions;

Ce cours se concentre sur **l'analyse des données**. Je conseille fortement la référence [suivante](#): Cox, D. R. and Donnelly, C. A. (2011) *Principles of applied statistics*, Cambridge, UK: Cambridge University Press.

- **L'analyse exploratoire des données**: consiste en l'utilisation de méthodes simples, intuitives, essentiellement graphiques. Son objectif est l'identification informelle de la structure d'un jeu de données (tendances, formes, observation atypiques). Elle permet donc de se familiariser avec les données.

L'analyse exploratoire suggère des hypothèses de travail et des modèles, qui sont formalisés et vérifiés dans le second pôle:

- **L'analyse confirmatoire des données**: elle conduit à des conclusions statistiques à partir de données en utilisant des notions de la théorie des probabilités. Cette partie plus formelle concerne notamment des méthodes de test, d'estimation et de prévision.

On distingue en général deux grands types d'études: expérimentales et observationnelles. La démarche est fondamentalement différente, ainsi que les conclusions que l'on peut en tirer.

	Etude expérimentale	Etude observationnelle
Situation	Sous contrôle: les paramètres que vous souhaitez sont ceux que vous obtenez	Donnée : ce que vous observez est tout ce que vous avez
Analyse	Aisée et planifiée	Potentiellement difficile
Interprétation	causale (si fait correctement)	associative, uniquement

- **L'analyse exploratoire des données**: consiste en l'utilisation de méthodes simples, intuitives, essentiellement graphiques. Son objectif est l'identification informelle de la structure d'un jeu de données (tendances, formes, observation atypiques). Elle permet donc de se familiariser avec les données.

L'analyse exploratoire suggère des hypothèses de travail et des modèles, qui sont formalisés et vérifiés dans le second pôle:

- **L'analyse confirmatoire des données:** elle conduit à des conclusions statistiques à partir de données en utilisant des notions de la théorie des probabilités. Cette partie plus formelle concerne notamment des méthodes de test, d'estimation et de prévision.

On distingue en général deux grands types d'études: expérimentales et observationnelles. La démarche est fondamentalement différente, ainsi que les conclusions que l'on peut en tirer.

	Etude expérimentale	Etude observationnelle
Situation	Sous contrôle: les paramètres que vous souhaitez sont ceux que vous obtenez	Donnée : ce que vous observez est tout ce que vous avez
Analyse	Aisée et planifiée	Potentiellement difficile
Interprétation	causale (si fait correctement)	associative, uniquement

Fig. 1.1: Here is my figure caption!

Part II

Statistique Exploratoire

TYPES DE DONNÉES

GRAPHIQUES

SYNTHÈSES NUMÉRIQUES

BOXPLOT

LOI NORMALE

QUELQUES PRINCIPES GÉNÉRAUX

Part III

Probabilité

