

Module d'Analyse Statistique : Introduction générale

M2 IGAST

Paul Chapron ¹

2021-2022

¹LASTIG-UGE-IGN/ENSG

ENSG
Géomatique

ÉCOLE NATIONALE
DES SCIENCES
GÉOGRAPHIQUES

Introduction Générale

- Cours M2 IGAST 2018 d'Ana-Maria Olteanu-Raimond
- R et espace <https://framabook.org/r-et-espace/>
- Probabilités, analyse de données et statistiques , Gilbert Saporta, Editions TECHNIP, 2011
- Cours de H. Commenges
https://gitlab.huma-num.fr/hcommenges/cours_statcomplet/-/raw/master/cours_statcomplet.pdf
- Nombreuses ressources en ligne, e.g. :
- http://www.foad-mooc.auf.org/IMG/pdf/424B_-Application_des_methodes_statistiques_d_analyse.pdf
- http://www.itse.be/statistique2010/co/Module_statistique_FSP.html

L'analyse spatiale étudie la **répartition** et l'**organisation** d'ensembles d'objets qui sont **localisés**

L'objectif est de :

«décerner en quoi la localisation apporte un élément utile à la connaissances des objets étudiés et peut en expliquer les caractéristiques»

[Pumain, Saint-Julien 97]

Analyse statistique

Méthodes **résumant** et **généralisant** des observations

- Les unités d'analyse sont des éléments indépendants en principe
- On ne s'intéresse pas à leur localisation ni à leur interactions (spatiales)

Analyse **spatiale** statistique

- Les unités d'analyse sont localisables
- On s'intéresse à leur propriétés y compris la localisation
- On fait l'hypothèse que leur localisation peut influencer les valeurs observées

Données spatiales :

Individus restreints spatialement (**selection spatiale**), ou variables de **localisation** géographique (e.g. Lieu de résidence, coordonnées) renseignées pour les individus

quid des **distances** ? → modèle gravitaire , réseau etc.

Analyse géométrique :

approche **géométrique** pour mieux décrire les données: analyse de forme, de réseaux, de proximité, méthodes de création de nouvelles entités à partir de la géométrie des objets

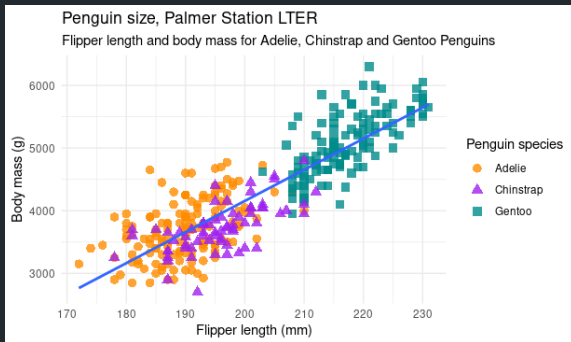
Analyse de données :

approche **statistique** permettant de faire émerger des relations (des groupes, des lois) pour aider l'étude de certains phénomènes

Statistiques Inférentielles vs. Statistiques Descriptives

A partir d'un échantillon , que peut-on attendre (=inférer) de la population ?

- Modèles, estimateurs, ... : régression, estimation, extrapolation
- e.g. sondages, recensement



Penguins data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network.
[<https://github.com/allisonhorst/palmerpenguins>]

Décrire, résumer, synthétiser les propriétés d'une **population** à partir des **variables** qui décrivent ses individus.

- **Graphiques** : nuages de points , histogramme, ...
- **Mesures** (fréquences, distributions, moments) sur des variables
- **Liaisons** statistiques entre variables : corrélation, covariance,...
- **Structure** interne des données : classification , ACP,...

Nous ferons majoritairement de la statistique **descriptive**
(même si, pour bien décrire, il faut parfois inférer)

Vocabulaire

Ensemble d'individus

"données", "corpus", "échantillon", "data"

très souvent **tabulaires**

Unité statistique élémentaire: personnes, logements, ...

→ "les lignes du tableau"

Caractéristiques, propriétés d'un individu, mesurées par des enquêtes, des observations...

→ "les colonnes du tableau"

Qualitatives : facteurs e.g. couleur, genre, CSP, type de pokemon,... → notion de **modalité**

Quantitatives : nombres e.g. taille, masse, revenu, surface, points de vie,... → parfois exprimées avec des **unités** : m, kg, s

Paradoxe de Simpsons

source: wikipedia

Variables quantitatives **continues** : $var \in \mathbb{R}$

Valeurs réelles, toutes les valeurs de l'intervalle de mesures peuvent exister

Variables quantitatives **discrètes** : $var \in \mathbb{N}$

Valeurs entières, pour des attributs **dénombrables** (comptage) ,
parfois utilisées pour encoder une variable qualitative à deux
modalités e.g. présence (1) , absence(0)

les valeurs sont prises dans un ensemble **fini** de valeurs possibles, défini par **extension** (i.e. on donne la liste des valeurs possibles)

→ notion de **modalités**

→ **nominales** (non ordonnées e.g. état civil) ou **ordinales** (ordonnées e.g. échelle de Lickert)

Spécificité de la statistique spatiale : à quelle échelle observer ?

Quel découpage, quelles unités spatiales ?

«Problème insoluble» : le **MAUP** (Modifiable Areal Unit Problem)

Mailles administratives :

agrégation/imbrication d'unités spatiales prédéfinies : comtés, départements, régions, pays

e.g. Comprendre comment le taux de chômage d'un pays est distribué entre les régions pour guider les politiques économiques

Découpages:

identification d'unités spatiales ayant des caractéristiques semblables

e.g. IRIS, carroyage

Désagrégation ou Ventilation

→ Inférer des caractéristiques individuelles à partir de l'analyse de données agrégées (ni facile ni immédiat)

Agrégation

→ Inférer des caractéristiques concernant les unités agrégées d'après les caractéristiques individuelles

Problème d'agrégation spatiale : les résultats d'une analyse statistique spatiale dépendent du choix d'agrégation

biais «systématique et insoluble»

Exemples tirés du rapport ESPON :

https://www.espon.eu/sites/default/files/attachments/espon343_maup_final_version2_nov_2006.pdf


```
<div style="text-align:center"> 
```

```
<div style="text-align:center"> 
```

Représenter/Tracer/Cartographier les variables de la population !

Attention aux seules valeurs chiffrées : exemple du "dinosaur"