

Statistiques

EPITA

Mars Avril 2020

Probabilités vs. Statistiques

- **Probabilités : modélisation mathématique de phénomènes aléatoires**
- **Propriétés, théorèmes, etc...**

- **Statistiques : Étude sur des données, aujourd'hui de tailles pouvant être considérables**
- **Prise de décisions au vu de ces données**
- **en fonction de ce qui est observé**
- **en fonction de ce qui est accepté**

Comment vous aurez à importer des données

- **Travail statistique :**
 - Travail sur des données
- **Accorder une attention particulière aux points suivants**
 - La qualité des données
 - Les données manquantes

Diversité des variables rencontrées

- Variables qualitatives
 - Ex 1 : des noms ~ nom d'un individu, nom de rue, ...
 - Ex 2 : des catégories ~ enquête client : très satisfait, moyen, pas satisfait,
- Variables quantitatives
 - Entiers
 - Nombres réels résultats de mesures
 - Nombres ordinaux déterminant un rang ...

Présentation de R : toute cette séance se termine en R

- On va commencer par se familiariser avec l'outil
 - Téléchargement, installation
 - Commandes élémentaires
 - Lecture de code source en R
 - Installation / chargement de packages
 - Lecture / écriture de tableaux dans un fichier (ici .csv français)
 - Prendre l'habitude de consulter la doc 😊

Les principaux type d'objets

- **Les nombres :**
 - numeric
 - integer
- **Les chaînes de caractères : character**
 - En R de très nombreuses primitives sur les chaînes de caractères car de nombreuses données statistiques sont qualitatives
- **Tous ces éléments sont à priori des vecteurs**
- **Les matrices**
- **Les data frames**

Visualisation (cf. démo)

- **Tableaux statistiques**
- **Camemberts**
- **Barres**
- **Histogrammes**
- **Boîtes à moustache**

Les données qui résument un tableau

- Cf. la fonction R summary
- En plus des min, max, moyennes, quantiles, ...
 - L'asymétrie (skewness)
 - L'applatissage (kurtosis)

Savoir reconnaître une distribution qui a l'apparence d'une distribution normale

- Cf. démo