



## L'analyse des données statistiques

### Public :

Les cadres devant analyser des données quantitatives et qualitatives

### Objectif :

Apprendre, en utilisant principalement Excel :

- à traiter des données provenant notamment des données de statistiques publiques (Exemple : données issues du recensement, de la base SITADEL, d'un fichier commun de demande de logements sociaux)
- en utilisant les techniques les plus communes d'analyse statistique descriptive, d'analyse multidimensionnelle des données et d'outils de prévision.

### Méthodes pédagogiques:

Notre méthode de travail est l'alternance d'apports théoriques et d'exercices pratiques bâtis sur des exemples non confidentiels tirés de notre expérience.

### Support logiciel :

Excel, StatBox

*Certaines méthodes pourront être démontrées sur d'autres supports.*

### Formateur : Nicolas Paris

Le déroulement de la formation est susceptible d'être adapté en cours de formation.

## 1<sup>ère</sup> journée

### Objectifs :

- ✓ acquérir ou revoir les notions fondamentales
- ✓ adopter le vocabulaire normalisé ou d'usage
- ✓ comprendre les principes de la statistique descriptive
- ✓ produire les résultats dans Excel
- ✓ savoir interpréter les résultats

### Statistiques descriptives :

- Individus et variables ou la construction du tableau de données :
  - comment repérer les individus ?
  - définir les variables,
  - problème des données hétérogènes quant :
    - aux niveaux d'agrégats différents
    - aux périodicités non identiques
    - aux sources différentes
- Classement des données :
  - données quantitatives : continues, discrètes,
  - données qualitatives : nominales, ordinales,
- Distribution de fréquence :
  - notion de variable aléatoire,
  - lois de probabilité usuelles : le cas de la loi normale,
  - fréquence absolue,
  - fréquence relative,
  - mise en œuvre dans Excel.
- Mesures de tendance centrale :
  - la moyenne : arithmétique, géométrique, quadratique...
  - la médiane, le mode
  - mise en œuvre dans Excel
- Mesures de dispersion :
  - la variance,
  - l'écart-type,
  - les percentiles,
  - mise en œuvre dans Excel.
- Mesure des données temporelles :
  - les choix des indices
  - calcul des évolutions
  - lecture associée d'indice

- Les représentations graphiques associées :
  - o Histogramme,
  - o diagramme en bâton,
  - o diagrammes en secteurs,
  - o nuages de points,
  - o boîtes à moustaches,
  - o mise en œuvre dans Excel
- Les tableaux croisés dynamiques
- La cartographie statistique :
  - o les fonds de cartes et les variables d'identification
  - o les modes de représentation
  - o le choix des variables à représenter

### **Bonnes pratiques dans la conduite d'analyse statistique**

- découpage des variables numériques en classes,
- prise en compte des valeurs manquantes,
- moyenne et écart-type,
- problèmes de généralisation des résultats obtenus.

## 2ème journée

### Objectifs :

- ✓ fixer les connaissances acquises en journée 1
- ✓ comprendre et mettre en œuvre : statistique inférentielle et modèle de régression
- ✓ savoir interpréter les méthodes

Rappel de la première journée, questions.

### Statistique inférentielle

- Problèmes d'estimation
  - o distribution d'échantillonnage,
  - o intervalle de confiance,
- Tests d'hypothèses :
  - o principe,
  - o hypothèses et risques,
  - o uni et bilatéralité,
  - o spécificités des tests de normalité et des tests non paramétriques,
  - o parallèle entre les tests usuels paramétriques et non paramétriques.
- Mise en œuvre dans Excel
  - o test-t de Student et test-t sur séries appariées,
  - o test du chi<sup>2</sup>,
  - o analyse de la variance à un facteur sur séries indépendantes,
- Avec pour chaque test :
  - o but, hypothèses et conditions de validité,
  - o mode de calcul,
  - o interprétation des résultats.

### Outils de prévision : expliquer

- Notion de corrélation :
  - o mesures de dépendance,
  - o mesures de dépendance monotone,
  - o mesures de concordance,
  - o la notion de causalité,
  - o mise en œuvre dans Excel

- Régression linéaire :
  - o le modèle,
  - o l'estimation des coefficients du modèle,
  - o ajustement du modèle,
  - o hypothèses sous jacentes à vérifier,
  - o régression linéaire et corrélation : les pièges à éviter,
  - o mise en oeuvre dans Excel.
- Régression linéaire multiple :
  - o le modèle,
  - o une généralisation de la régression linéaire simple,
  - o problèmes de colinéarité,
  - o choix des variables à inclure,
  - o exemple
  - o mise en oeuvre dans Excel.

### 3ème journée :

#### Objectifs :

- ✓ fixer les connaissances acquises en journée 2
- ✓ comprendre et mettre en œuvre : ACP, AFC et classification hiérarchique
- ✓ savoir interpréter les résultats

#### Synthétiser

- l'analyse en composante principale (ACP)
  - les dimensions
  - les composantes principales
  - le tableau des données
  - l'espace des individus
  - l'espace des variables
  - une représentation graphique optimisée : qualité de la représentation
  - de l'intérêt à normaliser les données
  - en pratique : expliquer la variabilité
  - en pratique : lire la représentation graphique
  - ACP et ajustement
  - mise en œuvre dans StatBox.
- l'analyse factorielle des correspondances (AFC)
  - sur quelles variables ?
  - le tableau des données
  - l'espace des modalités lignes et des modalités colonnes
  - représentation des modalités lignes et colonnes dans le même sous espace
  - qualité de la représentation
  - en pratique : expliquer la variabilité
  - en pratique : lire la représentation graphique
  - mise en œuvre dans StatBox
- l'analyse factorielle des correspondances multiples :
  - de l'AFC à l'AFCM
  - le tableau des données
  - analyse et interprétation par l'exemple
  - mise en œuvre dans StatBox

## Classifier

- la classification ascendante hiérarchique :
  - o principe de fonctionnement
  - o choix des variables
  - o indice de dissimilarité
  - o indice d'agrégation et algorithme de classification
  - o représentation graphique : le dendrogramme
  - o déterminer le nombre de groupe
  - o déterminer la césure du dendrogramme
  - o caractériser les groupes
  - o exemple
  - o mise en œuvre dans StatBox

## Synthèse : organigramme général de décision

- les grandes familles de méthodes étudiées
- savoir formaliser sa problématique
- savoir choisir la bonne méthode pour la bonne question
- 

## Evaluation écrite et orale