

Formation en Analyse des Données

Présenter par:

GUINDO SIDIKI,

**Ingénieur Statisticien Economiste,
Enseignant à l'ENSEA – Sénégal.**

Émail: guindosidiki@yahoo.fr

Cél: +221 77 168 95 60

PLAN DE LA FORMATION

1. Présentation des participants
2. Présentation du SESRIC
3. Introduction et objectif de la formation
4. Notions élémentaires de traitement statistique des données
5. Principe d'analyse des données multidimensionnelles,
6. Analyse en Composantes Principales.
7. Analyse des correspondances multiples
8. Méthodes de classification
9. Construction d'un indicateur composite par ACP.

Présentation des participants

- ❖ **Nom et Prénom, Profil de formation ou poste occupé, etc.**

Présentation du SESRIC

Voir diapo SESRIC

INTRODUCTION ET OBJECTIF

Les points abordés dans l'introduction

1. **Objectif de la formation**
2. **Durée**
3. **Difficultés**
4. **Logiciels utilisés**

1. Objectif de la formation

Introduction ou révision des différentes méthodes d'analyse des données.

Voir cette formation comme un cadre d'échange.

finalité: apprendre à mettre en œuvre l'essentiel des méthodes d'analyse des données.

2. Durée

La formation s'étale sur 4 jours.

La contrainte de durée fait qu'elle sera surtout pratique.

3. Difficultés

- ❖ **Contraintes de temps**
- ❖ **Hétérogénéité des participants (en termes de formation, de domaine d'activité, etc.)**

4. Logiciels utilisés

Nous utiliserons:

SPAD ET EXCEL.

Fin introduction

Notions élémentaires de traitement statistique des données

Les points abordés

- 1. Domaines d'application du traitement des données.**
- 2. Démarche dans le travail du statisticien.**
- 3. Quelques éléments de la statistique descriptive.**

1. Domaines d'application du traitement des données.

- ❖ **Démographie , Économie, Études de marché**
- ❖ **Assurances, Agriculture, Finance,**
- ❖ **Transport , Communications – etc.**

2. Démarche dans le travail du statisticien.

- **Conception: définition de la population, des unités, des caractères (variables), questionnaires**
- **Collecte des données: exécution de l'enquête**
- **Apurement: vérification, contrôle, redressement...**
- **Analyse: statistique descriptive, analyse des données multidimensionnelle, méthodes économétriques, etc.**
- **Publication des résultats**

3. Quelques éléments de la statistique descriptive.

La description des données par la statistique descriptive utilise des indicateurs simples:

3. Quelques éléments de la statistique descriptive.

- **Tendance centrale ou position**
- **Dispersion**
- **Forme**
- **Concentration**

Tendance centrale ou position ou localisation

- Médiane
- Mode
- Moyenne

Dispersion

- Étendue
- Écart-type
- Coefficient de variation

Forme

- Coefficient d'asymétrie
- Coefficient d'aplatissement

Concentration

- Courbe de concentration
- Coefficient de Gini
- Médiale

FIN de quelques éléments de la statistique descriptive.

Principe d'ADD

Les points abordés

- ❖ Les limites de la statistique descriptive
- ❖ Le principe d'ADD

Limites de la statistique descriptive

La statistique descriptive étudie une, deux ou trois variables.

On s'intéresse aux caractéristiques de tendance centrale, de dispersion, de forme, les liaisons entre deux variables.

Limites de la statistique descriptive

Cependant, le statisticien peut se trouver devant un tableau contenant plusieurs variables et individus.

Dans ce tableau, il cherche à dégager par exemple la tendance globale des données.

Limites de la statistique descriptive

**Par exemple: les variables qui sont liées,
les individus qui se ressemblent.**

**Par exemple: regrouper les individus
suivant leur proximité au vue des
variables.**

Limite de la statistique descriptive

Dans ces situations, la statistique descriptive reste limitée.

On passe donc aux méthodes d'analyse des données multidimensionnelles: c'est la « grande statistique descriptive ».

Limite de la statistique descriptive

Les méthodes ADD sont donc pour la plupart des temps des méthodes descriptives (il existe quelques méthodes ayant en partie des objectifs explicatifs).

Fin limite de la statistique descriptive

2. Principe générale d'ADD

Exemple introductif:

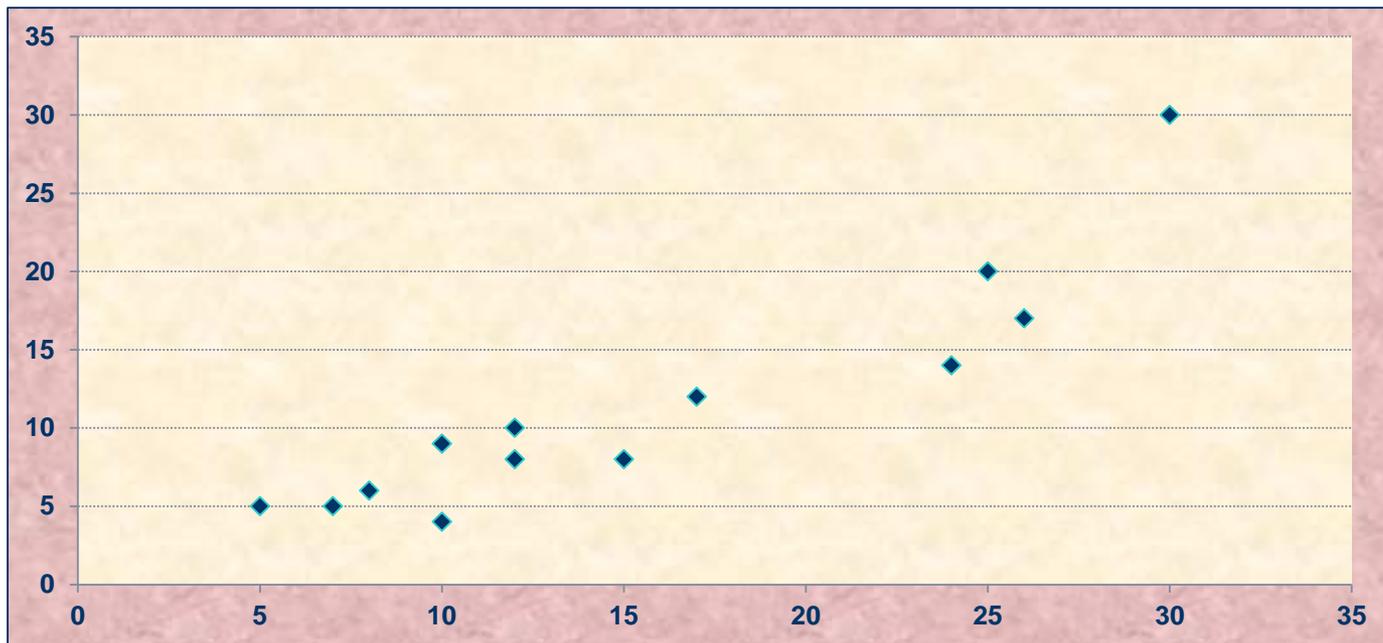
On dispose de deux variables: revenu et consommation sur 100 ménages.

Voir tableau

2. Principe générale d'ADD

MENAGE	REVENU	CONSOMMATION
1	10	9
2	25	20
3	12	10
4	7	5
5	26	17
6	5	5
.	30	30
.	24	14
.	10	4
.	8	6
.	15	8
99	12	8
100	17	12

2. Principe générale d'ADD



2. Principe générale d'ADD

Si nous avons trois variables : Revenu, Consommation et nombre personnes dans le ménage.

On peut faire un graphique à trois dimensions.

2. Principe générale d'ADD

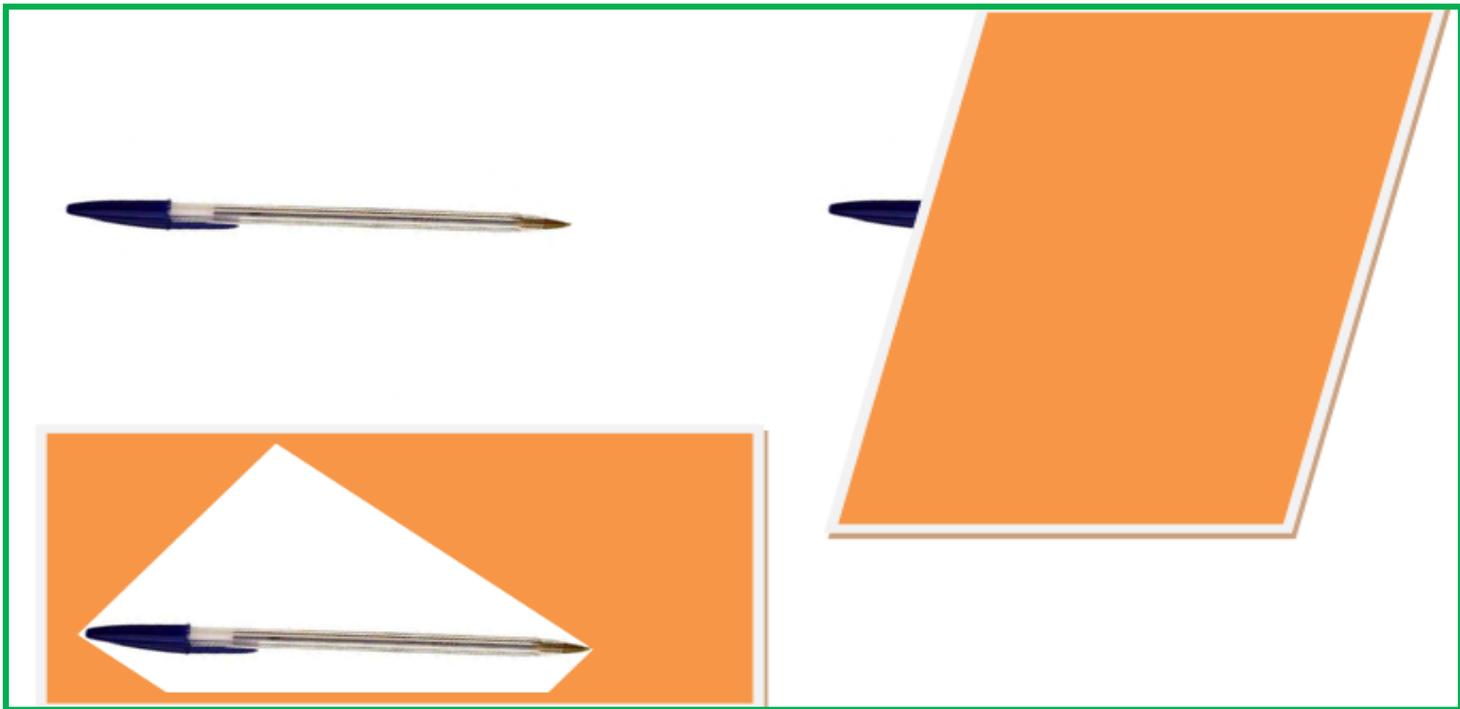
Si nous avons plusieurs variables (par exemple plus de 15) sur plusieurs individus alors on ne peut plus faire des graphique à 15 dimensions.

D'où l'utilisation des méthodes de projection.

2. Principe générale d'ADD

**Projeter: → c'est faire de bonne photo.
On se pose donc la question de savoir sur
quelle direction projeter.**

2. Principe générale d'ADD



2. Principe générale d'ADD

Les différents types de tableaux:

Tableau de variables quantitatives décrit sur des individus.

Tableau croisant deux variables qualitatives.

Tableau de variables qualitatives décrit sur les individus.

Autre type de tableaux (tableau de paquets de variables).

2. Principe générale d'ADD

Les méthodes d'analyse des données suivant le type de tableau:

Tableau de variables quantitatives décrit sur des individus (ACP)

Tableau croisant deux variables qualitatives (AFC).

Tableau de variables qualitatives décrit sur les individus (ACM).

Autre type de tableaux (tableau de paquets de variables):
les méthodes AFM

2. Principe générale d'ADD

De l'image à la réalité: les outils d'interprétation.

Ce que nous observons sur les photos peuvent être trompeuse. Il nous faut des outils d'aide à interprétation.

2. Principe générale d'ADD

Les outils:

Les Cosinus carré: (CO2), qualité de la représentation.

Le contribution (CTR): permet de mesurer la part des variables ou individus dans la formation des axes.

Disto: distance d'un individu à l'individu moyen.

2. Principe générale d'ADD

Cette formation va consister l'application des méthodes : ACP, ACM, CLASSIFICATION, EXEMPLE DE CONSTRUCTION D'INDICATEUR COMPOSITE.

Fin principe générale d'ADD

Analyse en composantes principales (ACP)

Les points abordés

- ❖ Objectif de l'ACP
- ❖ Nuage des individus, nuage des variables
- ❖ Présentation du tableau de l'exercice
- ❖ Première mise en pratique de l'ACP.
- ❖ Éléments supplémentaires, éléments actifs.

Objectif de l'ACP

On est supposé être devant un tableau de variables quantitatives décrit par des individus.

Sur ce tableau, on veut savoir quelles sont les variables qui sont liées entre elles, quels sont les individus qui se ressemblent.

Nuage des individus, nuage des variables:

On présentera deux nuages, celui des variables et celui des individus.

Présentation du tableau de l'exercice

Voir fichier Excel.
Nécessité d'une normalisation.

Première mise en pratique de l'ACP

- **Présentation sommaire de SPAD.**
- **Importation des données**
- **Lancer une première analyse.**

Éléments supplémentaires, éléments actifs.

- **Mettre certains individus en supplémentaires.**
- **Interpréter les résultats.**

Analyse des correspondances multiples (ACM)

Les points abordés

- ❖ Objectif de l'ACM
- ❖ Présentation du tableau de l'exercice
- ❖ Lien entre ACM et ACP.
- ❖ Interprétation des résultats.

Objectif de l'ACM

On est supposé être devant un tableau de variables qualitatives décrit par des individus.

Sur ce tableau, on veut savoir les différents regroupement suivant certains aspects.

Présentation du tableau de l'exercice

Voir fichier Excel.

Le tableau disjonctif complet, le tableau de BURT.

Lien entre ACM et ACP.

- **Même principe de base (projection).**
- **ACM = 2ACP sur le tableau disjonctif complet.**

Interprétation des résultats.

- **Mise en œuvre de la méthode**
- **Interprétation des résultats.**

Les méthodes de classification

Les points abordés

- ❖ Objectif de la classification
- ❖ explosion combinatoire
- ❖ Les deux grandes méthodes de classification
- ❖ Description des classes

Objectif de la classification

On est supposé être devant un tableau de variables quantitatives ou qualitatives mélangé (en termes de la nature des variables).

On cherche à regrouper les individus qui se ressemblent.

On se pose deux questions: combien de groupes former, où mettre les coupures.

Explosion combinatoire

Idéalement, il s'agit de former toutes les partitions fin de choisir la meilleures.

Cette situation aboutie à une explosion combinatoire.

D'où les méthodes hiérarchiques ou non hiérarchiques.

Les deux grandes méthodes de classification

- **La CAH, La CDH.**
- **Les méthodes de partitionnement**

Description des classes.

- **Les variables quantitatives.**
- **Les variances qualitatives**
- **Les axes d'une autre analyse.**
- **Les individus types.**

Mise en œuvre de la méthode.

- **Application des méthodes classification sur le tableau de l'exercice.**