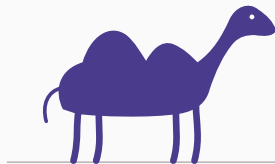


Introduction à l'analyse des données

Chapitre 2 : De la statistique descriptive à l'analyse des données



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Une continuité, non une rupture

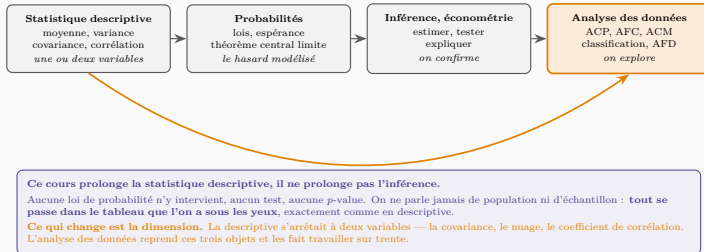
Ce qui distingue descriptive et inférentielle

Le plan du cours

Ce qu'il faut retenir

Une continuité, non une rupture

Ce que vous savez déjà faire



Le cours de statistique descriptive, rappelé

Nombre de variables

Ce qu'on y a appris

Une

tableau, graphiques, tendance centrale, dispersion, forme, concentration

Deux

tableau croisé, nuage de points, covariance, corrélation, droite d'ajustement

Intuition

La statistique descriptive s'arrête à **deux** variables, parce que deux variables se dessinent sur une feuille. Rien dans ses définitions ne l'exigeait pourtant : une moyenne, une covariance se calculent aussi bien sur trente variables.

Ce qui manque n'est pas la mesure : c'est la représentation.

Le nom exact de ce cours

Trois appellations, un même objet

Nom	Ce qu'il souligne
Analyse des données	l'usage français courant
Statistique descriptive multivariée	la filiation exacte
Statistique exploratoire	la démarche : faire émerger, non tester

Multivariée, et rien de plus

Ce cours **n'introduit aucune notion statistique nouvelle.**

La moyenne, la variance, la covariance, la corrélation — tout y est déjà. Ce qui change est le **nombre** de variables traitées ensemble, et l'outil géométrique qu'il faut pour les regarder.

C'est pourquoi les trois chapitres qui suivent sont des **rappels**. Il faut les prendre au sérieux : tout le reste du cours en découle mécaniquement.

Ce qui distingue descriptive et
inférentielle

Deux démarches, deux moments

La distinction

La **statistique inférentielle** part d'une hypothèse posée d'avance et demande aux données si elles la démentent. Elle conclut avec un risque d'erreur mesuré.

La **statistique exploratoire** ne pose aucune hypothèse. Elle demande aux données quelles structures elles contiennent.

Ni loi de probabilité, ni test, ni p -value

Vous ne rencontrerez dans ce cours **aucune** loi normale, **aucun** intervalle de confiance, **aucune** p -value.

Ce n'est pas un oubli : ces objets n'ont pas de rôle ici. On ne généralise pas d'un échantillon à une population — on **décrit** le tableau qu'on a sous les yeux, le plus lisiblement possible.

Laquelle vient en premier

L'exploration vient **avant** : on repère des groupes, des liaisons, des valeurs aberrantes, puis on formule des hypothèses qu'on ira tester sur d'autres données.

L'ordre inverse — tester une hypothèse suggérée par les mêmes données — produit des conclusions qui ne se reproduisent pas. C'est une faute de méthode, et elle est fréquente.

Le plan du cours

Quatre mouvements

Ce que traitent les quatorze chapitres

Chapitres	Objet
1–5	Le problème de la dimension, et les rappels indispensables
6–7	Le tableau vu comme un nuage , et l' inertie qui mesure son information
8–12	L' analyse en composantes principales : variables quantitatives
13–14	L' analyse des correspondances : variables qualitatives

Le fil rouge chiffré

Un exemple minuscule traverse tout le cours : **cinq clients**, notés sur deux critères — l'accueil et la fidélité.

Ce cours est une introduction

Il traite les deux méthodes fondatrices — composantes principales et correspondances — sur des cas simples.

Le cours d'**analyse des données** du niveau Master reprend l'ensemble depuis le début et l'approfondit :

Prolongement	Contenu
Correspondances multiples	plus de deux variables qualitatives
Classification	constituer des groupes de clients
Analyse discriminante	affecter un nouveau client à un groupe
Positionnement multidimensionnel	cartographier des ressemblances
Mise en œuvre	sur R et sur Python, cas réels

Ce qu'il faut retenir

Cinq points

1. Ce cours **prolonge** la statistique descriptive : c'est sa version **multivariée**.
2. Il n'introduit **aucune notion statistique nouvelle** — seulement plus de variables à la fois.
3. Il est **exploratoire** : ni loi, ni test, ni p -value.
4. L'exploration vient **avant** l'inférence, jamais après sur les mêmes données.
5. Quatre parties : le problème et les rappels, le nuage et l'inertie, l'**ACP**, l'**AFC**.

La suite

Les rappels commencent. Le chapitre 3 reprend ce qu'il faut savoir sur **une** variable — et s'arrête longuement sur une opération qui paraîtra anodine et qui commande tout le cours : centrer et réduire.