

Statistique descriptive

Introduction : Qu'est-ce que la statistique ?

Jaouad Madkour

12 octobre 2026

Plan de l'introduction

Prérequis	2
Ce que le cours suppose, ce qu'il établit	2
Le mot et la chose	2
D'où vient le mot	2
Le problème qu'elle résout	2
Ce que voit le directeur	3
Le fil conducteur de tout le cours	3
À quoi elle sert en économie	3
Mesurer : des grandeurs construites	3
Comparer et décider	4
Rendre comparable : un exemple chiffré	4
Une mise en garde, dès l'introduction	4
Ce que vous devez savoir faire à la fin du semestre	5
Descriptive et inférentielle	5
La distinction fondamentale	5
Les deux définitions	6
La phrase qui les sépare	6
Le tableau de comparaison	6
Lettres latines, lettres grecques	6
L'ordre des deux cours	7
Le piège symétrique	7
Le fil rouge du cours	7
Un seul jeu de données, du premier au dernier chapitre	7
Les premières lignes du fichier	8
Un premier coup d'œil, sans aucun calcul	8
S'entraîner	9
Exercice 1 : le premier coup d'œil d'un gérant	9
Exercice 1 : solution	9
Exercice 2 : comparer deux régions	9

Exercice 2 : solution	10
Ce qu'il faut retenir	10
Le résumé de l'introduction	10

Prérequis

Ce que le cours suppose, ce qu'il établit

Aucun cours préalable

Le programme du baccalauréat : manipuler une somme, un pourcentage, une fraction, et lire un graphique. Rien de plus.
La notation $\sum$ est introduite au premier chapitre ; il suffit de savoir additionner. Aucune notion de probabilité n'est employée.

Tout est repris depuis le début

Moyenne, médiane, variance, corrélation : **tout est défini ici**, et c'est ce cours qui sert de socle à trois autres — probabilités, analyse des données, économétrie.
Il décrit un ensemble de données observées. **Il n'en tire aucune conclusion sur une population** : c'est le rôle de la statistique inférentielle, et la distinction est faite dès cette introduction.

Le mot et la chose

D'où vient le mot

Statistique

La **statistique** est l'ensemble des méthodes qui permettent de **recueillir, résumer, représenter** et **interpréter** des données chiffrées portant sur un ensemble d'individus.

L'étymologie n'est pas anodine

Le mot vient de l'italien *statista*, l'homme d'État. La statistique naît au XVII^e siècle comme la « science de l'État » : recenser la population, mesurer les récoltes, évaluer la matière imposable.
Elle est née d'un besoin de **gouverner**, c'est-à-dire de décider sans pouvoir tout observer. C'est encore exactement sa fonction.

Le problème qu'elle résout

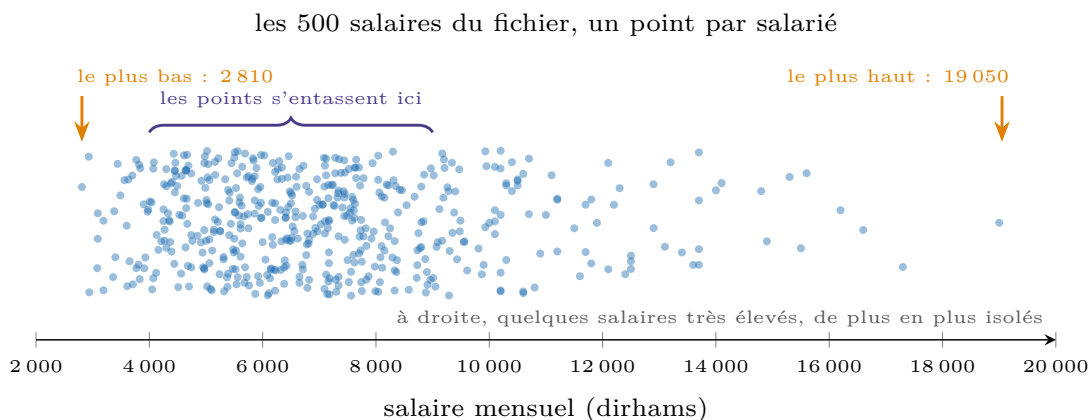
Un problème très concret

Une entreprise emploie 500 salariés. Le directeur des ressources humaines dispose d'un fichier de 500 lignes et de six colonnes.
Question du directeur général : « Alors, où en est-on sur les salaires ? »
Il est impossible de répondre en lisant les 500 lignes. Il faut **résumer**.

Le geste fondamental

Résumer, c'est **perdre volontairement de l'information** pour rendre le reste utilisable. 500 nombres deviennent un tableau de six lignes, puis un graphique, puis quatre indicateurs, puis un seul.

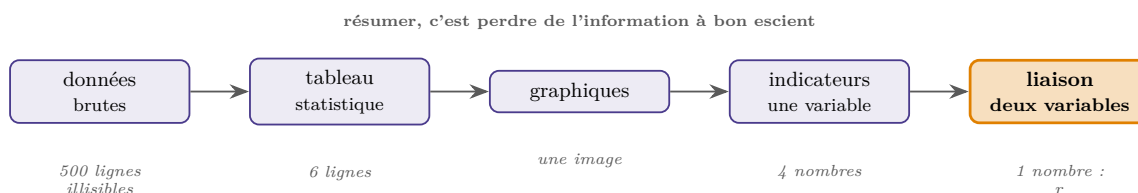
Ce que voit le directeur



Une image, pas encore un résumé

Un gros paquet entre 4 000 et 9 000 dirhams, une longue traîne vers le haut : on le voit, mais on ne peut ni le chiffrer, ni le comparer. **Le cours sert à passer de cette impression à des nombres.**

Le fil conducteur de tout le cours



À quoi elle sert en économie

Mesurer : des grandeurs construites

Mesurer

Aucune grandeur économique ne s'observe directement. Le **PIB**, l'**inflation**, le **chômage** n'existent pas dans la nature : ce sont des **constructions statistiques**.

L'indice des prix à la consommation résulte d'un panier de biens, de pondérations, d'un relevé mensuel de milliers de prix. Le taux de chômage dépend d'une définition — celle du Bureau international du travail — qui décide qui est « chômeur » et qui ne l'est pas.

Changer la méthode change le chiffre. C'est pourquoi un économiste doit savoir comment un indicateur est fabriqué avant de le commenter.

Comparer et décider

Comparer

Deux régions, deux entreprises, deux périodes. Comparer exige des grandeurs **rendues comparables** : par tête, en volume, à prix constants, en parité de pouvoir d'achat.

C'est un travail statistique, et il précède toute analyse économique.

Décider

Fixer un prix, dimensionner un stock, accorder un crédit, calibrer une politique publique.

Toute décision suppose une estimation, et toute estimation suppose une méthode.

Rendre comparable : un exemple chiffré

Deux régions fictives

	Région C	Région D
Production (milliards de dirhams)	120	84
Population (millions)	5,0	2,8
Par habitant (dirhams)	24 000	30 000

Pour C : $120 \times 10^9 / (5,0 \times 10^6) = 24\,000$ dirhams par habitant.

Le classement s'inverse

C produit plus, mais D produit plus **par habitant**. Les deux phrases sont vraies ; elles ne répondent pas à la même question. **Diviser par la bonne grandeur est déjà un choix statistique.**

Une mise en garde, dès l'introduction

Le chiffre n'est jamais neutre

Un même jeu de données peut donner deux messages opposés selon l'indicateur choisi.

Nos 500 salariées : le salaire **moyen** est de 7 158 dirhams, le salaire **médian** de 6 770. Un syndicat citera le second, une direction le premier. Aucun des deux ne ment.

Femmes et hommes : égaux ou non ?

	Femmes (215)	Hommes (285)
Salaire moyen	7 158	7 158
Salaire médian	6 580	6 820

Les moyennes disent « égalité parfaite », les médianes disent « 240 dirhams d'écart ». **Savoir lequel des deux chiffres répond à la question posée est précisément l'objet de ce cours.**

Ce que vous devez savoir faire à la fin du semestre

Cinq compétences

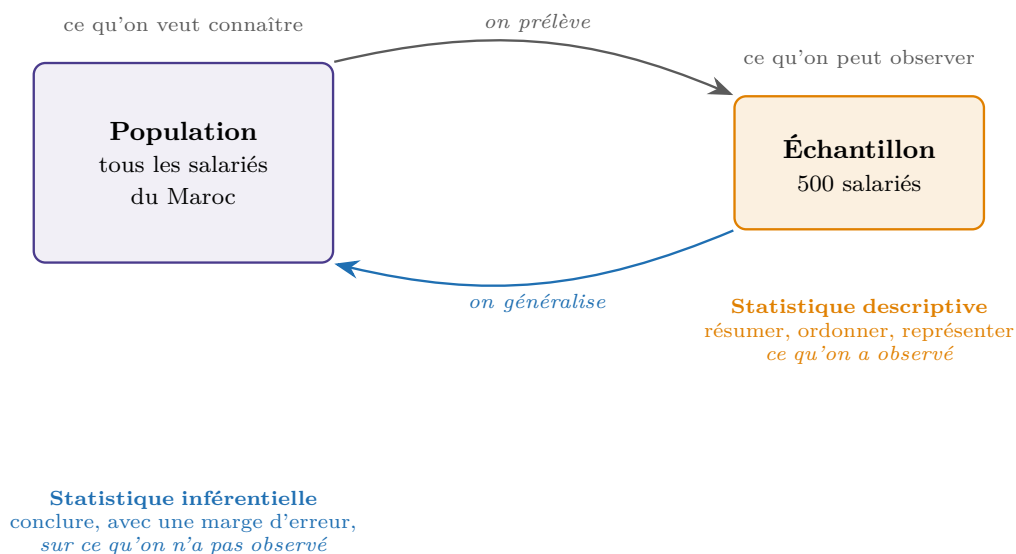
1. Construire un tableau statistique à partir de données brutes.
2. Choisir la représentation graphique qui convient au type de variable.
3. Calculer et **interpréter** les indicateurs de tendance centrale, de dispersion, de forme et de concentration — sur données brutes comme sur données groupées.
4. Étudier la liaison entre deux variables et l'ajuster par une droite.
5. **Reconnaître un chiffre mal construit** — et c'est peut-être le plus utile.

Et sur ordinateur

Les deux derniers chapitres refont toute la chaîne sous R, puis sous Python, sur le même fichier : chaque nombre du cours y est retrouvé à l'identique.

Descriptive et inférentielle

La distinction fondamentale



Les deux définitions

Statistique descriptive

Elle **résume et représente** les données dont on dispose. Elle ne prétend rien dire au-delà. Ses conclusions sont **certaines** : si l'on a bien calculé, le salaire moyen de ces 500 salariés est de 7 158 dirhams. Il n'y a pas de marge d'erreur.

Statistique inférentielle

Elle **généralise** de l'échantillon observé à la population dont il est issu. Ses conclusions sont **probables**, jamais certaines, et elles s'accompagnent toujours d'une marge d'erreur et d'un niveau de confiance.

La phrase qui les sépare

Deux phrases, deux démarches

- Descriptive : « Dans **cette** entreprise, le salaire moyen est de 7 158 dirhams. »
- Inférentielle : « Le salaire moyen dans **le secteur** est compris entre 6 900 et 7 400 dirhams, avec 95 % de confiance. »

La première décrit ce qu'on a vu. La seconde parie sur ce qu'on n'a pas vu.

Le test du mot « ce »

Une phrase descriptive porte sur **ces** données-là : ces 500 salariés, cette année, cette entreprise. Dès qu'une phrase parle du secteur, du pays ou de l'avenir à partir des seules données observées, elle est inférentielle — et elle doit le dire, avec sa marge d'erreur.

Le tableau de comparaison

Les deux démarches

	Descriptive	Inférentielle
Sur quoi elle porte	les données observées	la population entière
Ce qu'elle produit	des constats	des estimations et des tests
Nature des conclusions	certaines	probables
Outil central	moyenne, écart-type, graphique	intervalle de confiance, test
Suppose	rien	un échantillon aléatoire
Notation	$\bar{x}$, s	μ , σ

Lettres latines, lettres grecques

La notation, à retenir dès maintenant

Les grandeurs calculées **sur l'échantillon** se notent en lettres latines : $\bar{x}$, s , r .

Les grandeurs **de la population**, inconnues, se notent en lettres grecques : μ , σ , ρ .

Cette convention vaut dans toute la statistique. Elle rappelle en permanence ce qu'on a

mesuré et ce qu'on cherche à atteindre.

Une même grandeur, deux notations

Grandeur	Sur les données observées	Dans la population
Moyenne	$\bar{x}$	μ
Écart-type	s	σ
Corrélation	r	ρ

L'ordre des deux cours

Pourquoi la descriptive vient en premier

Trois raisons, dans cet ordre d'importance :

1. **L'inférentielle repose sur elle.** Un intervalle de confiance est construit à partir d'une moyenne et d'un écart-type : il faut savoir les calculer et les lire.
2. **Elle se passe d'hypothèses.** Aucune loi de probabilité, aucun échantillonnage aléatoire. Elle est vraie par construction.
3. **Elle vient toujours en premier dans la pratique.** Aucun statisticien ne teste avant d'avoir regardé ses données. Le premier réflexe est toujours de décrire.

Le piège symétrique

Décrire un échantillon, parler de la population

Attention à la tentation inverse : **décrire un échantillon et parler comme si l'on décrivait la population.**

« Le salaire moyen au Maroc est de 7 158 dirhams » — non : le salaire moyen **de ces 500 salariés** l'est. Passer de l'un à l'autre demande l'inférentielle, et une justification.

Ce que fait le statisticien public

Une enquête nationale sur l'emploi interroge quelques dizaines de milliers de ménages, pas tout le pays. Le taux de chômage qu'elle publie est donc une **estimation** : la description de l'échantillon est l'étape descriptive, son extension au pays l'étape inférentielle.

Le fil rouge du cours

Un seul jeu de données, du premier au dernier chapitre

Le fichier salaires

500 salariés d'une entreprise, six colonnes :

Variable	Nature	Valeurs
id	repère	1 à 500
salaire	quantitative continue	de 2 810 à 19 050 dirhams

anciennete	quantitative discrète	de 0 à 35 ans
diplome	qualitative ordinale	Bac, Bac+2, Bac+5
region	qualitative nominale	Tanger-Tétouan, Casablanca, Rabat, Autres
sexe	qualitative nominale	F, H

Les premières lignes du fichier

Cinq salariés sur 500

id	salaire	anciennete	diplome	region	sexe
1	8 540	10	Bac+5	Rabat	H
2	7 100	15	Bac	Rabat	F
3	4 380	4	Bac	Autres	F
4	11 790	19	Bac+2	Casablanca	F
5	13 430	19	Bac+5	Tanger-Tétouan	H

Pourquoi un seul jeu de données

Chaque indicateur du cours sera calculé sur **ces mêmes** 500 salariés. Vous verrez donc, à chaque chapitre, le même fichier révéler une propriété de plus — et vous pourrez retrouver chaque nombre à la main, sous R et sous Python.

Un premier coup d'œil, sans aucun calcul

Trois nombres lus directement

Question	Réponse
Combien de salariés ?	500
Salaire le plus bas	2 810 dirhams
Salaire le plus haut	19 050 dirhams
Écart entre les deux	16 240 dirhams, soit un facteur 6,8

Une impression à préciser

Trois nombres, et l'on sait déjà que la distribution est **très étalée**. Mais le plus haut salaire est-il un cas isolé ou le sommet d'un groupe ? Tout le reste du cours va préciser cette impression.

S'entraîner

Exercice 1 : le premier coup d'œil d'un gérant

Les douze salariés d'une PME de Tétouan

Salaires mensuels, en dirhams, rangés par ordre croissant :

3 200 ; 3 500 ; 3 800 ; 4 100 ; 4 300 ; 4 600 ; 5 000 ; 5 400 ; 6 200 ; 7 500 ; 9 800 ; 15 600.

1. Donner le salaire le plus bas, le plus haut, leur écart et leur rapport.
2. Quelle proportion des salariés gagne moins de 5 000 dirhams ?
3. Calculer le salaire moyen. Combien de salariés gagnent moins que lui ?
4. Le gérant affirme : « Chez nous, on gagne plus de 6 000 dirhams en moyenne. » Un salarié répond : « La moitié d'entre nous gagne moins de 5 000 dirhams. » Qui a raison ?
5. « Dans les PME de Tétouan, le salaire moyen dépasse 6 000 dirhams. » Cette phrase est-elle descriptive ?

Exercice 1 : solution

Les calculs

1. Plus bas 3 200, plus haut 15 600 : écart 12 400 dirhams, rapport $15\,600/3\,200 = 4,9$.
2. Six salaires sur douze sont inférieurs à 5 000 : **la moitié**, 50 %.
3. La somme vaut 73 000 dirhams, donc $\bar{x} = 73\,000/12 = 6\,083$ dirhams. **Huit salariés sur douze** gagnent moins que la moyenne.

L'interprétation

4. **Les deux ont raison.** Le seul salaire de 15 600 dirhams tire la moyenne vers le haut : elle décrit un point d'équilibre que les deux tiers des salariés n'atteignent pas.
5. **Non.** Elle parle de toutes les PME de Tétouan à partir d'une seule : c'est une généralisation, donc une phrase inférentielle — et, avec un seul établissement observé, une généralisation injustifiée.

Exercice 2 : comparer deux régions

Production, population et prix

Deux régions fictives :

	A en 2015	A en 2023	B en 2015	B en 2023
Production (milliards de dh)	60	90	48	72
Population (millions)	2,8	3,0	1,6	2,0

Production en dirhams courants.

Entre 2015 et 2023, les prix ont augmenté de 20 % dans les deux régions.

1. Calculer la production par habitant de chaque région en 2023. Laquelle est la plus riche ?
2. De combien la production de A a-t-elle augmenté en valeur ? En volume, c'est-à-dire à prix constants ?
3. La production par habitant de B a-t-elle augmenté en volume ?

Exercice 2 : solution

Par habitant, puis à prix constants

1. A : $90 \times 10^9 / (3,0 \times 10^6) = 30\,000$ dirhams ; B : $72 \times 10^9 / (2,0 \times 10^6) = 36\,000$ dirhams. **B est la plus riche par habitant**, bien que A produise davantage.
2. En valeur : $90/60 = 1,50$, soit +50 %. En volume, on retire la hausse des prix : $1,50/1,20 = 1,25$, soit +25 % seulement.
3. En 2015, B produisait $48 \times 10^9 / (1,6 \times 10^6) = 30\,000$ dirhams par habitant. En volume, $36\,000/1,20 = 30\,000$: **aucune progression**. La production a crû, les prix et la population aussi, exactement autant.

La leçon

Trois chiffres vrais — +50 %, +25 %, 0 % — pour la même économie. Avant de commenter un chiffre, il faut savoir **par quoi il a été divisé et en quels prix il est exprimé**.

Ce qu'il faut retenir

Le résumé de l'introduction

Six points

1. La statistique **recueille, résume, représente et interprète** des données.
2. Résumer, c'est **perdre de l'information à bon escient**.
3. En économie, elle sert à **mesurer**, à **comparer** et à **décider**.
4. Comparer exige des grandeurs **rendues comparables** : par habitant, à prix constants.
5. La **descriptive** décrit : conclusions certaines. L'**inférentielle** généralise : conclusions probables.
6. Lettres latines pour l'échantillon, lettres grecques pour la population.

La suite

Avant de calculer, il faut nommer ce sur quoi l'on travaille : population, individu, variable, modalité. C'est le chapitre 1, et ce vocabulaire commande tout le reste — notamment le choix des graphiques et des indicateurs autorisés.