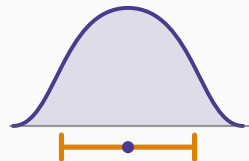


# Statistique inférentielle

## Chapitre 1 : La troisième face



---

Jaouad Madkour

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

Prérequis

Où nous sommes

Le vocabulaire

La démarche

Le fil rouge du cours

Ce qu'il faut retenir

## Prérequis

---

## Le cours qui précède

*Probabilités*, dont ce cours utilise en permanence les lois et le théorème central limite.

## Les notions mobilisées

Variable aléatoire, espérance, variance. Les lois **normale**, de **Student**, du **khi-deux** et de **Fisher**. Le théorème central limite, qui justifie à lui seul la moitié du cours.

### Repris depuis le début

Échantillonnage, estimation, intervalle de confiance et test d'hypothèse sont construits ici.

**Le cours ferme la boucle du bloc statistique** : la descriptive décrivait un échantillon, les probabilités modélisaient le hasard, l'inférence conclut de l'un à l'autre.

Où nous sommes

---

### Ce que nous savons faire

La **statistique descriptive** a résumé cinq cents salaires : moyenne 7 158,26, écart-type 2 598,10, une droite qui relie salaire et ancienneté.

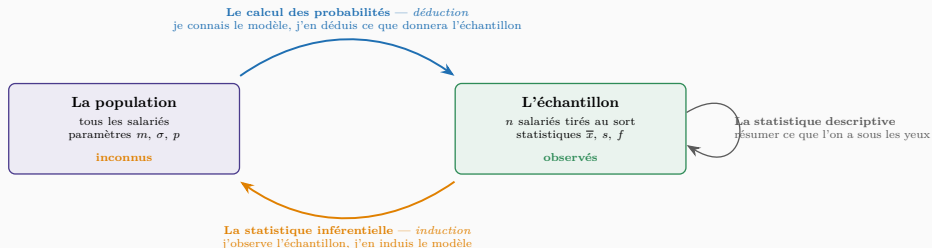
Le **calcul des probabilités** a construit le langage du hasard : lois, espérance, variance, et deux théorèmes de convergence.

### Ce que nous ne savons pas encore faire

Un institut interroge 50 salariés et trouve une moyenne de 7 005,80 dirhams.

Que peut-il en dire de la **population entière** ? Rien, pour l'instant — et c'est précisément l'objet de ce cours.

# Les trois faces d'une même discipline



Les trois cours ne sont pas trois matières : ce sont trois sens de circulation entre les deux mêmes boîtes.

La *descriptive* reste dans l'échantillon et le résumé. Les *probabilités* descendent de la population vers l'échantillon : le modèle est donné, on en calcule les conséquences. L'*inférence* remonte en sens inverse, et c'est le seul trajet incertain des trois — d'où la nécessité de mesurer cette incertitude.

L'*économétrie* prolongera la flèche orange : au lieu d'un seul paramètre  $m$ , elle remontera vers toute une relation entre variables.



## Ce que la figure établit

### Deux boîtes, trois trajets

| Discipline                | Trajet                   | Nature           |
|---------------------------|--------------------------|------------------|
| Statistique descriptive   | dans l'échantillon       | résumer          |
| Calcul des probabilités   | population → échantillon | <b>déduction</b> |
| Statistique inférentielle | échantillon → population | <b>induction</b> |

### Le seul trajet incertain est le nôtre

Descendre est sûr : si l'on sait que  $p = 0,04$ , on calcule sans risque la probabilité de 14 défauts.

Remonter ne l'est pas : d'un échantillon, plusieurs populations différentes auraient pu naître.

**L'inférence ne supprime jamais cette incertitude — elle la mesure. C'est là toute sa valeur.**

## Le vocabulaire

---

## Deux mots qu'il ne faut jamais confondre

Un **paramètre** décrit la population. Il est **fixe** et **inconnu** :  $m, \sigma, p$ .

Une **statistique** se calcule sur l'échantillon. Elle est **connue** et **aléatoire** :  $\bar{x}, s, f$ .

## La convention de notation, à respecter dès maintenant

Lettres **grecques** ou majuscules latines pour les paramètres; lettres **latines minuscules** pour les statistiques.

$$m \longleftrightarrow \bar{x} \quad \sigma \longleftrightarrow s \quad p \longleftrightarrow f \quad \beta_1 \longleftrightarrow \hat{\beta}_1$$

Une copie qui écrit « la moyenne de la population vaut 7 005,80 » commet, en une phrase, l'erreur fondamentale du cours.

## Le mot le plus important est « aléatoire »

### Sur nos données

Prélevons trois échantillons de 40 salariés dans les mêmes 500. Les moyennes obtenues sont

6 863      7 001      7 895

Trois nombres différents, calculés sur la même population, par la même méthode.

### La conséquence

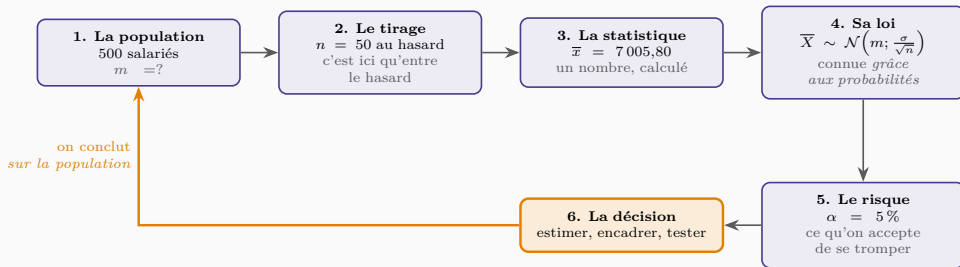
$\bar{X}$  n'est pas un nombre : c'est une **variable aléatoire**, avec une loi, une espérance et un écart-type.

Toute la statistique inférentielle consiste à connaître cette loi, puis à s'en servir.

## La démarche

---

# Les six étapes, une fois pour toutes



La quatrième étape est la seule qui demande un cours entier de probabilités, et c'est elle qui rend l'inférence possible : sans connaître la loi de  $\bar{X}$ , un écart entre 7 005,80 et une valeur annoncée ne voudrait rien dire.

Les six étapes ne changent jamais. Seule change la statistique calculée à l'étape 3 — une moyenne, une proportion, une variance, une pente de régression — et donc la loi de l'étape 4.

# Les trois questions que l'on saura traiter

## Le programme du cours

| Question                       | Réponse                        | Chapitres |
|--------------------------------|--------------------------------|-----------|
| Quelle valeur retenir ?        | <b>estimation ponctuelle</b>   | 6 à 8     |
| Avec quelle précision ?        | <b>intervalle de confiance</b> | 9 à 12    |
| Cette affirmation tient-elle ? | <b>test d'hypothèse</b>        | 13 à 19   |

## Trois questions, un seul outil

Les trois reposent sur la même chose : la **loi de la statistique calculée**.

Estimer, encadrer et tester ne sont pas trois méthodes; ce sont trois façons de lire une même distribution d'échantillonnage.

### Trois décisions, trois inférences

- Un **contrôle qualité** ne teste pas tous les articles : il conclut sur le lot à partir de quelques pièces.
- Un **sondage** interroge mille personnes et annonce un résultat national.
- Une **étude de marché** teste un produit sur un quartier et décide d'un lancement.

Dans les trois cas, la décision porte sur ce qu'on n'a **pas** observé.

### Pourquoi on ne mesure pas tout

Trois raisons, et elles suffisent : le **coût**, le **délai**, et la **destruction** — on ne peut pas tester la durée de vie de toutes les ampoules produites, il n'en resterait aucune à vendre.

Quand un recensement complet est possible et peu coûteux, l'inférence est inutile. Elle n'est pas une commodité intellectuelle : elle est une **nécessité économique**.



## Le fil rouge du cours

---

# Les mêmes cinq cents salariés

## Le jeu de données

Le fichier `salaires.csv` du cours de statistique descriptive : 500 salariés, six colonnes — identifiant, salaire, ancienneté, diplôme, région, sexe.

| Grandeur            | Valeur   |
|---------------------|----------|
| Effectif $N$        | 500      |
| Moyenne $m$         | 7 158,26 |
| Écart-type $\sigma$ | 2 598,10 |

## Un dispositif inhabituel, et volontaire

Ces cinq cents salariés seront traités comme **la population**, dont nous connaissons donc les paramètres.

Nous y tirerons des échantillons, nous ferons semblant d'ignorer  $m$ , et nous **vérifierons** à chaque étape que la méthode retrouve bien la vraie valeur. C'est la seule façon de voir fonctionner l'inférence de l'intérieur.

## Un salarié sur dix

Nous retiendrons les individus dont l'identifiant est un multiple de dix :  $n = 50$  observations.

| Grandeur              | Symbole      | Valeur   |
|-----------------------|--------------|----------|
| Moyenne d'échantillon | $\bar{x}$    | 7 005,80 |
| Écart-type corrigé    | $s$          | 2 250,98 |
| Erreur-type           | $s/\sqrt{n}$ | 318,34   |

## Un choix reproductible

Ce tirage ne dépend d'aucune graine aléatoire : n'importe quel étudiant, sous R, sous Python ou sur tableur, retrouvera exactement ces trois nombres.

C'est une exigence de méthode : **un résultat statistique qu'on ne peut pas reproduire n'est pas un résultat.**

Ce qu'il faut retenir

---

# Le résumé du chapitre

## Six points

1. La descriptive résume l'échantillon; les probabilités descendent vers lui; l'inférence **remonte** vers la population.
2. Le trajet inférentiel est le seul **incertain** — et le cours sert à mesurer cette incertitude.
3. Un **paramètre** est fixe et inconnu; une **statistique** est connue et aléatoire.
4.  $\bar{X}$  change d'un échantillon à l'autre : c'est une **variable aléatoire**.
5. Trois questions — estimer, encadrer, tester — reposent sur un seul objet : la **distribution d'échantillonnage**.
6. Le fil rouge : 500 salariés de moyenne 7 158,26, dont on tire 50 individus de moyenne 7 005,80.

## La suite

Tout ce qui précède suppose une chose que nous n'avons pas encore examinée : que l'échantillon **représente** la population.

Cette condition n'est pas automatique, et aucune formule ne la remplace. Le chapitre 2 lui est entièrement consacré.