# Cours de Statistiques et Probabilités pour Ingénieurs en Génie des Matériaux Présentation du cours

## Objectifs de la séance

— Comprendre ce que sont les statistiques.

— Comprendre ce que sont les probabilités.

— Appliquer ces notions à des exemples concrets.

## C'est quoi une statistique

Tout étudiant a déjà calculé une moyenne. Après trois tests notés 12, 15 et 14, le total est 41. En divisant par 3, on obtient environ 13,67.

Cette note n'a jamais été obtenue. Pourtant elle a un sens précis : c'est la note qui, répétée trois fois, donne exactement le même total que les notes réelles.

La moyenne permet donc de résumer plusieurs valeurs en un seul nombre qui conserve le total (la « masse ») des données. C'est extrêmement utile pour comparer des étudiants, faire des classements ou calculer des moyennes de moyennes.

**La moyenne est un exemple de statistique** : une opération qui transforme un ensemble de données en un résumé numérique.

Mais ce n'est pas le seul résumé possible.

Prenons un second étudiant qui a obtenu 10, 19 et 12. Le total est encore 41, donc la moyenne est identique (13,67). Pourtant les notes sont très différentes : l'un est régulier, l'autre est irrégulier (une très bonne note et deux notes plus faibles).

La moyenne ne voit pas cette différence de variation. Existe-t-il une statistique qui la mesure ? Oui : c'est la **variance**.

### Qu'est-ce que la variance ?

Nous avons vu que la moyenne résume bien le « centre » des notes, mais qu'elle ne dit rien sur leur dispersion. Les deux étudiants ont exactement la même moyenne ($13{,}67$), pourtant leurs notes sont très différentes.

Pour mesurer cette dispersion par un seul nombre, on introduit une nouvelle statistique : la **variance**.

L'idée est simple : on regarde de combien chaque note s'écarte de la moyenne. Cependant, on ne peut pas se contenter d'additionner ces écarts, car leur somme est toujours nulle (les écarts positifs et négatifs s'annulent).

La solution consiste à élever chaque écart au carré (ce qui les rend tous positifs), puis à en calculer la moyenne. C'est précisément ce que mesure la variance.

**Exemple avec les deux étudiants** (moyenne $= 13{,}67$ pour les deux)

**Étudiant 1** : notes $12$, $15$, $14$
Écarts à la moyenne : $12 - 13{,}67 \approx -1{,}67$ $15 - 13{,}67 \approx +1{,}33$ $14 - 13{,}67 \approx +0{,}33$
Somme des écarts : $-1{,}67 + 1{,}33 + 0{,}33 = 0$
Carrés des écarts : $(-1{,}67)^2 \approx 2{,}79$ $(1{,}33)^2 \approx 1{,}77$ $(0{,}33)^2 \approx 0{,}11$
Somme des carrés $\approx 4{,}67 \neq 0$
D'où la variance $\approx$ **1,56**.

**Étudiant 2** : notes $10$, $19$, $12$
Écarts à la moyenne : $10 - 13{,}67 \approx -3{,}67$ $19 - 13{,}67 \approx +5{,}33$ $12 - 13{,}67 \approx -1{,}67$
Somme des écarts : $-3{,}67 + 5{,}33 - 1{,}67 = 0$
Carrés des écarts : $(-3{,}67)^2 \approx 13{,}47$ $(5{,}33)^2 \approx 28{,}41$ $(-1{,}67)^2 \approx 2{,}79$
Somme des carrés $\approx 44{,}67 \neq 0$
D'où la variance $\approx$ **14.89**.

On voit clairement que la somme des écarts est toujours nulle, tandis que la somme des carrés mesure bien la dispersion (beaucoup plus grande pour le second étudiant).

La variance du second étudiant est presque **10 fois plus grande** que celle du premier ($14{,}89$ contre $1{,}56$).

Cela signifie que les notes du second étudiant sont beaucoup plus dispersées autour de la moyenne : il y a une très grande note et deux notes plus faibles. Au contraire, les notes du premier étudiant restent proches de la moyenne, ce qui se traduit par une variance faible.

La variance permet donc de quantifier, par un seul nombre, la régularité (ou l'irrégularité) ; c'est ce qu'on a appelé la **dispersion** des résultats.

# C’est quoi une probabilité

Dans la vie courante, on utilise souvent le mot « probable ». Si un ami arrive presque toujours en retard, on s’attend à ce qu’il soit encore en retard la prochaine fois. On serait même surpris s’il arrivait à l’heure.

Cette surprise révèle quelque chose d’important : à partir d’observations **passées**, on s’est formé une opinion sur un événement **futur**.

C’est exactement l’idée de la probabilité en mathématiques : donner une mesure objective de la chance qu’un événement se produise, en s’appuyant soit sur l’expérience, soit sur le raisonnement.

### Approche empirique, ou expérimentale

Prenons une pièce de monnaie. Je la lance 25 fois et j’obtiens 15 fois face. La fréquence observée de face est $15/25 = 0,6$ (soit 60 %).

On peut alors *estimer* que la probabilité d’obtenir face est d’environ 0,6. Mais ce n’est qu’une estimation : si je relance 25 fois, je peux obtenir 12 faces, ou 18… La fréquence change d’une série de lancers à l’autre.

### Approche théorique, ou raisonnement

On peut aussi regarder la pièce elle-même. Elle a deux faces, et (si elle est bien équilibrée) aucune raison de favoriser pile ou face. Les deux événements sont donc **équiprobables** : on leur attribue la même probabilité $1/2 = 0,5$.

### Pourquoi les deux approches ne donnent pas le même résultat ?

Parce que l’expérience est soumise à des fluctuations aléatoires. Plus on lance la pièce de fois, plus la fréquence observée se rapproche en général de 0,5. C’est ce que garantit une loi mathématique fondamentale : la **loi des grands nombres**.

Elle est la base de toute la théorie des probabilités : elle explique pourquoi, malgré les variations d’une expérience à l’autre, l’approche empirique finit par confirmer l’approche théorique.

# Conclusion – Statistique

Une **statistique** est un nombre qui résume un ensemble de données.

- La **moyenne** indique le centre des valeurs en conservant leur total. - La **variance** mesure la dispersion autour de ce centre (moyenne des carrés des écarts à la moyenne).

Deux séries peuvent avoir la même moyenne mais des variances très différentes : la variance permet alors de quantifier la régularité des données.

# Conclusion – Probabilité

Une **probabilité** mesure objectivement la chance qu’un événement futur se produise, à partir d’observations passées ou d’un raisonnement.

- Approche **empirique** : on estime la probabilité par la fréquence observée (mais celle-ci fluctue d’une expérience à l’autre). - Approche **théorique** : on utilise le raisonnement (par exemple l’équiprobabilité quand aucune issue n’est favorisée).

La **loi des grands nombres** relie les deux approches : plus on répète l'expérience, plus la fréquence se rapproche de la probabilité théorique.

## Discussion

### Sur la statistique

1. Pourquoi dit-on que la moyenne « conserve la masse » des notes ? Que se passerait-il si on utilisait un autre résumé (par exemple la note la plus fréquente) ?
2. Deux étudiants ont la même moyenne mais des variances très différentes. Lequel vous semble le plus « régulier » ? Dans quel contexte (contrôle continu, examen final, sport…) la variance est-elle plus importante que la moyenne ?
3. Peut-on avoir une variance nulle ? Que cela signifierait-il concrètement ?

### Sur la probabilité

4. Si on lance une pièce 10 fois et qu'on obtient 8 faces, peut-on conclure que la pièce est truquée ? Pourquoi la loi des grands nombres nous invite-t-elle à la prudence ?
5. Dans la vie quotidienne, on utilise souvent des probabilités « empiriques » (ex. : « mon bus est en retard 3 fois sur 5 »). Quels sont les risques de se fier uniquement à ce type d'estimation ?
6. Y a-t-il des situations où l'approche théorique (équiprobabilité) est impossible ou trompeuse ? Donnez des exemples.

### Question de synthèse

7. Moyenne, variance et probabilité : en quoi ces trois notions permettent-elles de passer de données brutes (notes, lancers de pièce…) à une information plus utile pour prendre une décision ?

## Exemples de réponses possibles

### Sur la statistique

1. La moyenne conserve la masse car la somme des notes est égale à $n$ fois la moyenne. Si on prenait par exemple la note la plus fréquente (le mode), on perdrait cette information sur le total : deux séries avec le même mode peuvent avoir des totaux très différents.
2. L'étudiant avec la plus petite variance est le plus régulier. La variance est particulièrement importante dans un contrôle continu (on préfère un élève constant) ou dans un sport d'équipe (on préfère un joueur fiable plutôt qu'un joueur irrégulier, même s'il a le même niveau moyen).
3. Oui, une variance nulle signifie que toutes les notes sont identiques (égales à la moyenne). Il n'y a aucune dispersion.

### Sur la probabilité

4. Non, on ne peut pas conclure. Avec seulement 10 lancers, les fluctuations sont encore très importantes. La loi des grands nombres nous dit qu'il faut un grand nombre de répétitions pour que la fréquence se rapproche de la vraie probabilité.
5. Le principal risque est de généraliser trop vite à partir d'un petit nombre d'observations. Les conditions peuvent aussi changer (nouveau chauffeur, travaux, etc.), ce qui rend l'estimation empirique obsolète.
6. Oui. Exemples : - une pièce truquée (les faces ne sont plus équiprobables) ; - un dé pipé ; - des événements qui ne sont pas symétriques (ex. : « il va pleuvoir demain » n'a aucune raison d'avoir une probabilité de $1/2$).

### Question de synthèse

7. Ces trois notions permettent de condenser une grande quantité d'informations brutes en quelques nombres significatifs : - la moyenne donne le niveau global, - la variance indique la régularité, - la probabilité quantifie le risque ou la chance d'un événement futur. On peut alors comparer, classer ou décider de façon plus rationnelle.

# Cours de Statistiques pour Ingénieurs en Génie des Matériaux
# Chapitre 1 : Introduction aux Notions de Base

## Objectifs de la séance

— Comprendre les concepts de population, échantillon, variables et modalités.

— Identifier et distinguer les différents types de variables statistiques : qualitatives, quantitatives, discrètes et continues.

— Appliquer ces notions à des exemples concrets en génie des matériaux.

## Introduction

La statistique descriptive a pour but :

— de dégager les propriétés essentielles que l'on peut déduire d'une accumulation de données ;

— de donner une image concise et simplifiée de la réalité.

Le résultat d'une observation, d'une mesure, n'est pas égale à la valeur théorique calculée ou espérée par l'ingénieur ; la répétition d'une même mesure, réalisée dans des conditions qui semblent identiques, ne conduit pas toujours aux mêmes résultats. Ces fluctuations, dues à des causes nombreuses, connues ou inconnues, contrôlées ou non, créent des difficultés aux ingénieurs et aux scientifiques. Quel résultat doivent-ils prendre ? Quel degré de confiance peuvent-ils accorder à la décision prise ? Les réponses à une enquête varient d'un individu à un autre ; quelles conclusions valables peut-on tirer d'un sondage ? Les méthodes de la statistique descriptive apportent des réponses à ces problèmes.

Pour être soumis à un traitement statistique, un tableau de données doit comporter au moins une variable de nature aléatoire. Une définition simple du caractère aléatoire d'une variable est qu'elle peut prendre au hasard des valeurs différentes.En génie des matériaux, les statistiques sont utilisées pour étudier les propriétés des matériaux (ex. : résistance, dureté), optimiser les procédés de fabrication, ou évaluer la fiabilité des produits.

Ce premier chapitre donne les définitions et les propriétés des principales notions utiles pour comprendre et traiter un problème de statistique.

# 1 Notions de population et d'échantillon

## 1.1 Population et Individu

**Population** : Ensemble complet des individus, objets ou événements étudiés.

On peut aussi dire *ensemble statistique*. Cette terminologie vient de ce que les premiers statisticiens étaient intéressés par les recenssements civiles et militaires.

**Unité statistique** : chaque individu.

Une population doit être correctement définie afin que l'appartenance d'un individu à cette population soit reconnue sans ambiguïté.

**Exemple** :
Une usine fabrique des tiges métalliques utilisées dans l'assemblage de certaines structures. Pour étudier la résistance à la traction de ces tiges, on mesure cette résistance pour un lot de 100 tiges.
Population statistique : l'ensemble des 100 tiges ou des 100 mesures.
Unité statistique : chacune des tiges ou chacune des 100 mesures.
Propriété étudiée : la résistance à la traction de tiges métalliques.

## 1.2 Échantillon

**Échantillon** : groupe restreint, ou sous-ensemble, issu de la population.

Pour définir un tel échantillon, une méthode consiste à prélever, au hasard, un sous-ensemble d'individus, en utilisant, par exemple, des tables de nombres au hasard

— Exemple : 50 échantillons d'aluminium prélevés pour tester leur résistance à la traction.

— Importance : Étudier un échantillon est souvent plus pratique et économique qu'étudier toute la population.

## 1.3 Caractères et Variables Aléatoires

### Caractères

On s'intéresse à certaines particularités ou caractères des individus d'une population statistique :

— un seul caractère étudié, série numérique à une dimension,

— deux caractères étudiés, série numérique à deux dimensions,

— plus de deux caractères, on doit utiliser les techniques de l'analyse multidimensionnelle (qui dépasse le cadre de ce cours.

Les caractères étudiés peuvent être :

— le poids, la taille, le niveau d'études, la catégorie socioprofessionnelle, le lieu d'habitation..., dans le secteur des sciences humaines,

— le poids, la masse, la composition..., dans le secteur des sciences techniques.

### Modalités

Un caractère peut prendre différentes **modalités**. Ces modalités doivent être incompatibles et exhaustives afin que l'appartenance ou la non-appartenance d'un individu à une modalité soit définie sans ambiguïté.

Un caractère peut être :

— quantitatif, les modalités sont mesurables ou repérables,

— qualitatif, les modalités ne sont pas mesurables.

**Exemple** : Pour la variable "type de revêtement", les modalités peuvent être : peinture, galvanisation, anodisation.

### Variables statistiques

**Variable statistique :** un caractère faisant l'objet d'une étude statistique.

Elle peut donc être qualitative ou quantitative.

**Exemple** : Dureté d'un alliage, température de fusion, type de revêtement.

Une variable quantitative est appelée :

— discrète si elle prend un nombre fini de valeurs souvent entières,

— continue si elle prend toutes les valeurs d'un intervalle fini ou infini.

## Discussion

— Pourquoi ne pas tester tous les échantillons ? (Coût, temps, faisabilité)

— Comment choisir un échantillon représentatif ? (Aléatoire, stratification)

## Application pratique

Soit une population de 100 pièces en céramique. On mesure deux variables :

— Taille des pores (en micromètres).

— Type de céramique (zircone, alumine, carbure de silicium).

Question : Identifiez les variables et leurs modalités.

# 2 Types de variables statistiques

## 2.1 Variables qualitatives

**Variables qualitatives** : Décrivent une qualité ou une catégorie, sans valeur numérique intrinsèque.

Elles peuvent être de deux types :

— *Nominales* : Pas d'ordre (ex. : type de matériau : acier, aluminium, titane).

— *Ordinales* : Avec un ordre (ex. : qualité d'un revêtement : faible, moyenne, élevée).

**Exemple** : Classification des alliages par type (ferreux, non ferreux).

## 2.2 Variables quantitatives

**Variables quantitatives** : Représentent une quantité mesurable.

Elles peuvent être de deux types :

— *Discrètes* : Valeurs dénombrables (ex. : nombre de défauts dans un échantillon).

— *Continues* : Valeurs dans un intervalle continu (ex. : résistance à la traction en MPa).

**Exemple** : Mesure de la densité d'un matériau (continue) ou nombre de cycles avant rupture (discrète).

## 2.3 Tableau récapitulatif

| Type | Exemple | Caractéristique |
|---|---|---|
| Qualitative nominale | Type de matériau | Catégories sans ordre |
| Qualitative ordinale | Qualité du matériau | Catégories ordonnées |
| Quantitative discrète | Nombre de défauts | Valeurs dénombrables |
| Quantitative continue | Résistance (MPa) | Valeurs dans un intervalle |

TABLE 1 – Types de variables statistiques

### Application pratique

Pour chaque variable suivante, indiquez si elle est qualitative (nominale ou ordinale) ou quantitative (discrète ou continue) :

1. Dureté Vickers d'un alliage.
2. Type de polymère (ABS, PLA, PET).
3. Nombre de fissures observées.
4. Température de traitement thermique.
5. Classe de résistance (basse, moyenne, haute).

# 3 Effectif, Fréquence et Pourcentage

**Effectif** : Nombre d'observations d'une modalité ou d'une classe pour une variable donnée. On le note $n_i$. Et on note l'effectif total par $N$, c'est-à dire $N = \sum_i n_i$.

**Exemple** : Dans un échantillon de 100 pièces en acier, on observe 20 pièces avec des défauts de surface. L'effectif de la modalité « défaut » est 20.

**Fréquence** Rapport entre l'effectif d'une modalité (ou classe) et l'effectif total, exprimé sous forme décimale.
Formule : $f_i = \frac{n_i}{N}$, où $n_i$ est l'effectif de la modalité $i$ et $N$ est l'effectif total.

**Exemple** : Si 20 pièces sur 100 ont un défaut, la fréquence des défauts est $\frac{20}{100} = 0,2$.
**Remarque :** On vérifie facilement que la somme des fréquences est 1 :
$\sum_i f_i = \sum_i \frac{n_i}{N} = \frac{1}{N} \sum_i n_i = \frac{N}{N} = 1$

**Pourcentage** : Fréquence exprimée en pourcentage ($f_i \times 100$) %.

**Exemple** : La fréquence de 0,2 correspond à 20 %.

## 3.1 Exemple concret

On mesure la qualité d'un revêtement sur 50 échantillons de céramique, avec les modalités suivantes : faible, moyenne, élevée. Résultats :

— Faible : 10 échantillons.

— Moyenne : 25 échantillons.

— Élevée : 15 échantillons.

Calculs :

— Effectif total : $10 + 25 + 15 = 50$.

— Fréquences : Faible $\frac{10}{50} = 0,2$, Moyenne $\frac{25}{50} = 0,5$, Élevée $\frac{15}{50} = 0,3$.

— Pourcentages : Faible 20 %, Moyenne 50 %, Élevée 30 %.

## 3.2 Tableau récapitulatif

| Qualité | Effectif | Fréquence | Pourcentage |
|---|---|---|---|
| Faible | 10 | 0,2 | 20 % |
| Moyenne | 25 | 0,5 | 50 % |
| Élevée | 15 | 0,3 | 30 % |
| **Total** | 50 | 1,0 | 100 % |

TABLE 2 – Série statistique pour la qualité du revêtement

# 4 Effectif cumulé et Fréquence cumulée

**Effectif cumulé** : Somme des effectifs des modalités ou classes jusqu'à une certaine valeur (souvent pour des variables ordinales ou quantitatives).

**Exemple** : Pour la qualité du revêtement (faible, moyenne, élevée), les effectifs cumulés sont :

— Faible : 10.

— Moyenne : $10 + 25 = 35$.

— Élevée : $35 + 15 = 50$.

**Fréquence cumulée** : Somme des fréquences jusqu'à une certaine modalité ou classe, exprimée sous forme décimale ou en pourcentage.
Formule : $F_i = \sum_{k=1}^{i} f_k$, où $f_k$ est la fréquence de la modalité $k$.

**Exemple** : Fréquences cumulées pour l'exemple précédent :

— Faible : 0,2 (20 %).

— Moyenne : $0,2 + 0,5 = 0,7$ (70 %).

— Élevée : $0,7 + 0,3 = 1,0$ (100 %).

## 4.1 Exemple avec une variable quantitative

On mesure la résistance à la traction (en MPa) de 60 échantillons d'un alliage, regroupés en classes :

| Classe (MPa) | Effectif | Fréquence | Effectif cumulé | Fréquence cumulée |
|---|---|---|---|---|
| $[200, 250[$ | 10 | 0,167 | 10 | 0,167 |
| $[250, 300[$ | 20 | 0,333 | 30 | 0,500 |
| $[300, 350[$ | 15 | 0,250 | 45 | 0,750 |
| $[350, 400[$ | 15 | 0,250 | 60 | 1,000 |
| **Total** | 60 | 1,000 | | |

TABLE 3 – Série statistique pour la résistance à la traction

### Interprétation

— 50 % des échantillons ont une résistance inférieure à 300 MPa (fréquence cumulée à [250, 300[).

— Utile pour évaluer la proportion d'échantillons répondant à un seuil de performance (ex. : résistance minimale pour une application).

# 5 Exemple pratique

**Défauts relevés sur une pièce de tissu**

Un fabricant de tissu essaie une nouvelle machine ; il compte le nombre de défauts sur 75 échantillons de 10 mètres. Il a trouvé les résultats suivants :

Nombre d'individus : les 75 échantillons. $N = 75$

Effectif associé à la valeur $k$, le nombre $n_k$ : par exemple, sur les 75 échantillons examinés, 11 présentent $k = 2$ défauts, donc si $k = 2$, $n_k = 11$.

Fréquence associée à la valeur $k$ : le quotient $n_k/n$.

$11/75 = 0,146$ est la fréquence associée à la valeur $k = 2$.

Effectif cumulé associé à la valeur $k$ : le nombre d'échantillons ayant au plus $k$ défauts ($k$ compris).

$38 + 15 + 11 = 64$ est la fréquence cumulée absolue associée à la valeur $k = 2$.

Fréquence cumulée associée à la valeur $k$, le nombre d'échantillons ayant au plus $k$ défauts ($k$ compris) divisé par n.

$64/75 = 0,853$ est la fréquence cumulée relative associée à la valeur $k = 2$.

# 6 Conclusion et discussion

— Les fréquences relatives et les fréquences cumulées relatives peuvent être utilisées pour comparer deux ou plusieurs populations.

— Dans le cas d'une distribution continue, les données sont en général regroupées en classes. Les fréquences absolues, relatives et cumulées sont définies par rapport aux classes et non par rapport aux valeurs de la variable.

— Questions ouvertes : Quelles variables dans vos projets pourraient être analysées avec ces outils ?