UE 4

COURS 3A:
LA METHODE STATISTIQUE EN MEDECINE

2 TYPES D'ANALYSE EN STATISTIQUES :

- *Analyse descriptive*

  ⟶ description d'une situation avec des paramètres

- *Analyse déductive* *(explicative ou inductive)*

  ⟶ conclusions à partir d'observations et de mesures

Le Tutorat est gratuit, toute reproduction ou vente est interdite

DEFINITIONS

- *Statistiques* ⟶ art de collecter et d'interpréter des données

- *Biostatistiques* ⟶ applications de la statistique au domaine biologique

- *Donnée* ⟶ résultat de l'observation d'un individu par un instrument (poids, taille) ou les sens de l'observateur (couleur des yeux)

Le Tutorat est gratuit, toute reproduction ou vente est interdite

→ Une donnée s'observe sur plusieurs individus pour lesquels elle n'est pas strictement équivalente

On parle donc de VARIABLE

→ Elle prend telle valeur pour un individu, telle valeur pour un autre : Une variable … varie !

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- *Paramètre* → Grandeur apportant une information résumée sur la variable étudiée (seulement pour des variables quantitatives !)

- *Variabilité* → Entre différentes données, peut être due :
  - au hasard (→ *variabilité intrinsèque*)
  - à la physiologie (ou une autre origine)

→ Intra-individuelle : la donnée n'est pas équivalente d'un instant à l'autre pour un même individu

→ Inter-individuelle : la donnée n'est pas équivalente d'un individu à l'autre pour un instant T

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- ***Toute observation est soumise à une variabilité intrinsèque (= Hasard) !!***

→ Pour plusieurs observations, le résultat est très souvent variable

- ***L'observation d'une différence ne permet pas en soi d'en préciser la cause !!***

→ Constater une différence statistiquement significative ne donne pas la clé de son interprétation

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- *Série statistique* ⟶ Collection d'objets de même nature présentant des caractéristiques différentes (variables)

→ Variables quantitatives : *mesurables* (par un instrument de mesure)

→ Variables qualitatives : *non mesurables* (binaire, nominale)

UNE VARIABLE NUMERIQUE N'EST PAS FORCEMENT UNE VARIABLE QUANTITATIVE

→ Ex : numero étudiant / code postal / numéro de téléphone

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- *Population* ⟶ Série exhaustive de TOUS les individus que l'on veut étudier
  → **une population est une série statistique !!**

- *Echantillon* ⟶ Ensemble fini et d'effectif limité, extrait de la population

Pourquoi échantillonner ? → Population inaccessible en entier
→ Etude sur l'échantillon, pari sur l'extrapolation des résultats à la population

Le Tutorat est gratuit, toute reproduction ou vente est interdite

→ L'échantillon doit être <u>représentatif de la population</u> !

→ Pour cela, <u>une seule solution</u> → **LA RANDOMISATION !**
= TIRAGE AU SORT (TAS)

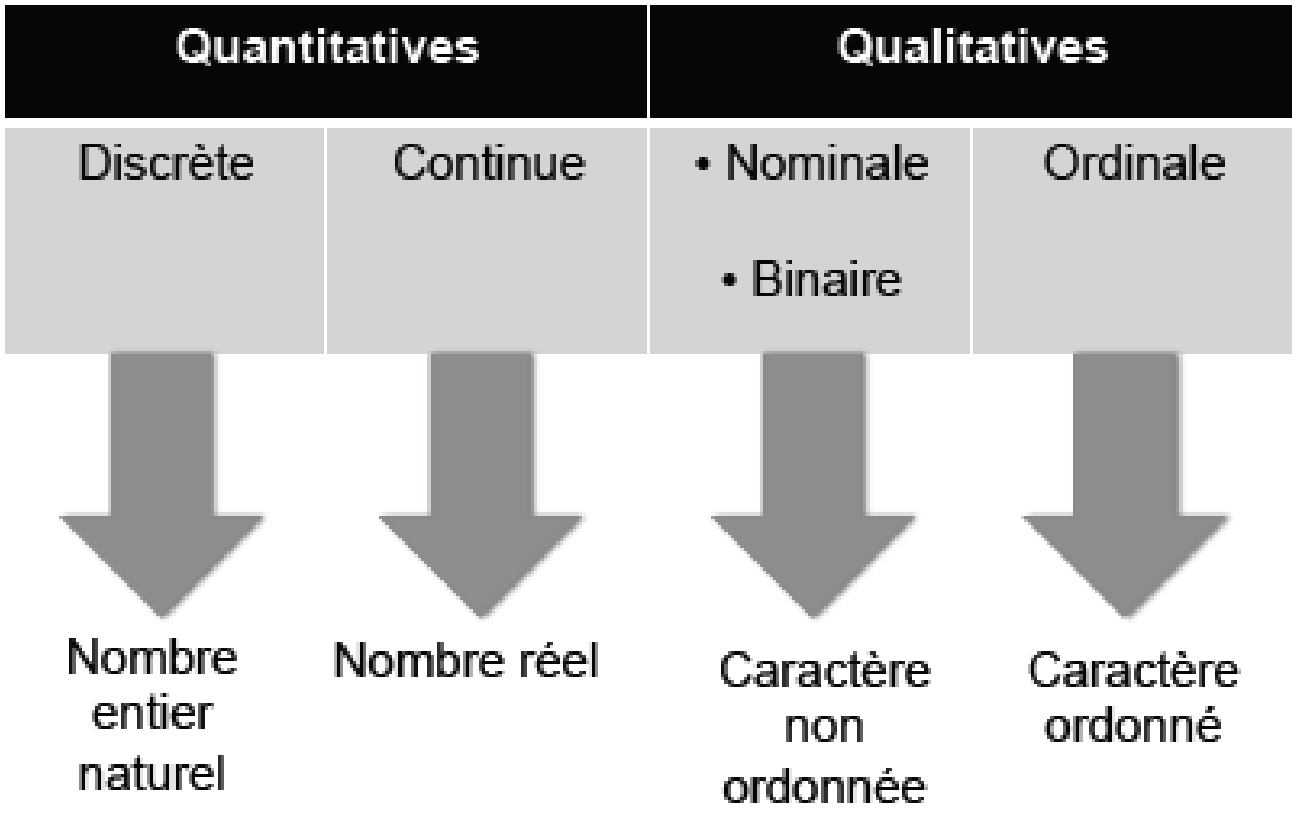
→ **Echantillon = Connu**
Population = Inconnue

→ **On travaille sur l'échantillon !**

→ On extrapole à la population si tout a été respecté
(représentativité , TAS, bonne conduite de l'étude …)

Le Tutorat est gratuit, toute reproduction ou vente est interdite

LES DIFFÉRENTS TYPES DE VARIABLES



Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

QCM 1) A propos des variables. Donnez les vraies.

A) Le numéro étudiant est une variable quantitative

B) Le temps en seconde est une variable quantitative discrète

C) Le nombre d'enfants d'un couple est une donnée quantitative discrète

D) La mention d'un élève au bac est une variable qualitative ordinale

E) Toutes les propositions sont fausses

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

Réponse: CD

QCM 1) A propos des variables. Donnez les vraies.

A) Le numéro étudiant est une variable ~~quantitative~~ QUALITATIVE

B) Le temps en seconde est une variable quantitative ~~discrète~~ CONTINUE

C) Le nombre d'enfants d'un couple est une donnée quantitative discrète

D) La mention d'un élève au bac est une variable qualitative ordinale

E) Toutes les propositions sont fausses

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

QCM 2) On demande à un patient l'intensité de sa douleur sur une échelle de 0 à 10. La réponse attendue sera :

A) Une variable quantitative discrète

B) Une variable qualitative continue

C) Une variable qualitative nominale

D) Une variable quantitative continue

E) Toutes les propositions sont fausses

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

Réponse : E → **QUALITATIVE ORDINALE !!!!!!**

QCM 2) On demande à un patient l'intensité de sa douleur sur une échelle de 0 à 10. La réponse attendue sera :

A) Une variable ~~quantitative discrète~~

B) Une variable qualitative ~~continue~~

C) Une variable qualitative ~~nominale~~

D) Une variable ~~quantitative continue~~

E) Toutes les propositions sont fausses

Le Tutorat est gratuit, toute reproduction ou vente est interdite

→ Une douleur moyenne correspondra à :

Evaluez votre douleur en plaçant un point sur l'échelle suivante :

Pas de douleur I----------●----------I Douleur intense

REVIENT au même que :

Evaluez votre douleur en entourant un chiffre :

0 1 2 3 4 (5) 6 7 8 9 10

⟶ La <u>numérisation</u> permet d'exploiter les résultats plus facilement !

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

QCM 3) Parmi les propositions suivantes, donnez les vraies :

A) Les P1 qui m'écoutent parler dans l'amphi forment une série statistique (sûrement très petite…)

B) Tous les PACES forment un échantillon représentatif des étudiants français

C) Pour faire une étude sur tous les PACES de France (qui étudie ça franchement ?), je peux travailler sur la promo 2014/2015 de Nice

D) Que nenni ! Pour faire cette étude je dois tirer au sort des PACES dans la promo 2014/2015 de Nice

E) Vive la Biostat !!! (avec une majuscule s'il vous plaît !)

Le Tutorat est gratuit, toute reproduction ou vente est interdite

QCM TIME

Réponse : AE

QCM 3) Parmi les propositions suivantes, donnez les vraies:

A) Les P1 qui m'écoutent parler dans l'amphi forment une série statistique (sûrement très petite …)

B) Tous les PACES forment un échantillon ~~représentatif~~ des étudiants français

C) Pour faire une étude sur tous les PACES de France (qui étudie ça franchement ?) , ~~je peux travailler sur la promo 2014/2015 de Nice~~

D) Que nenni ! Pour faire cette étude ~~je dois tirer au sort des PACES dans la promo 2014/2015 de Nice~~

E) Vive la Biostat !!! (avec une majuscule s'il vous plaît !)

Le Tutorat est gratuit, toute reproduction ou vente est interdite

STATISTIQUES DESCRIPTIVES

VARIABLES QUALITATIVES

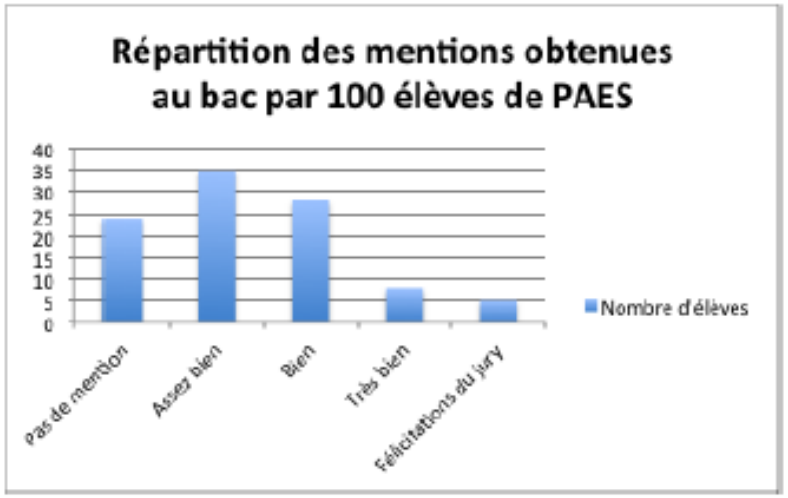
2 manières de les représenter →

- **tableau**

Mentions	Nombre d'élèves
Pas de mention	24
Assez bien	35
Bien	28
Très bien	8
Félicitations du jury	5

- **histogramme**

(normalisé ou non)



Le Tutorat est gratuit, toute reproduction ou vente est interdite

VARIABLES QUANTITATIVES

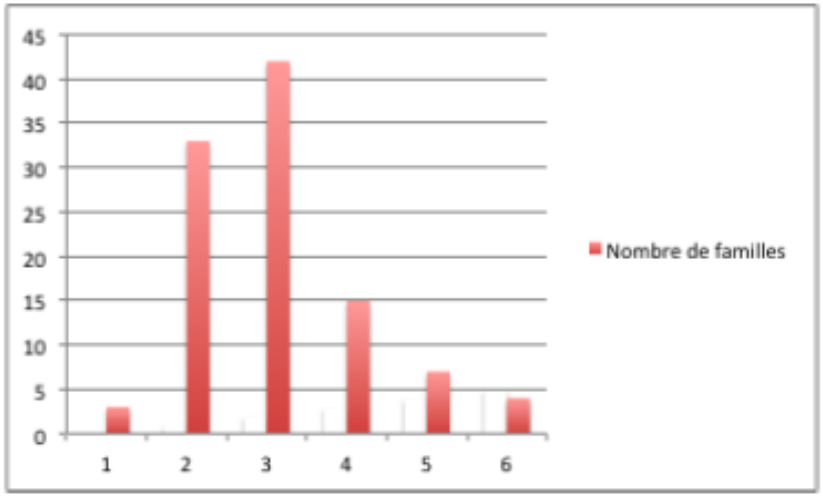
2 manières de les représenter →

- **tableau**

Nombre d'enfants par famille	Nombre de familles
0	3
1	33
2	42
3	15
4	7
5	4

- **histogramme**

(normalisé ou non)



Le Tutorat est gratuit, toute reproduction ou vente est interdite

Seules les variables quantitatives peuvent être résumées par des paramètres :

Indicateurs de position →
- Moyenne
- Médiane
- Quartiles

Indicateurs de dispersion →
- Variance
- Ecart-type

Le Tutorat est gratuit, toute reproduction ou vente est interdite

La moyenne

- Indicateur de position
- Adaptée aux calculs statistiques
- Pour n données : X_1 ; Xi ; … ; Xn

$$m = \frac{\Sigma xi}{n}$$

Le Tutorat est gratuit, toute reproduction ou vente est interdite

LA MEDIANE

- Indicateur de position

- *Valeur centrale d'une liste ordonnée par ordre croissant*

  Ex : Les notes d'une classe sont 7/10/12/12/13/14/18

  ⟶ La médiane est ?

- Elle sépare la liste en 2 (50% en dessous / 50% au-dessus)

Avec n données relevées par ordre croissant :

- **Si n pair →** $M = (X_{n/2} + X_{(n/2)+1}) / 2$

- **Si n impair →** $M = X_{(n+1)} / 2$

Le Tutorat est gratuit, toute reproduction ou vente est interdite

LES QUARTILES

- Indicateurs de position

- *Valeurs partageant une série ordonnée en 4 groupes de mêmes effectifs*

→ **Q1 (premier quartile) sépare les premiers 25% de la série**

→ **Q2 (deuxième quartile) sépare les premiers 50% de la série**

→ **Q3 (troisième quartile) sépare les premiers 75% de la série**

Le Tutorat est gratuit, toute reproduction ou vente est interdite

2e quartile = Médiane !!!!!!!!!

En effet → Médiane ET 2e quartile séparent les premiers 50% de la série

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- Soit n données relevées par ordre croissant (X_1 ; X_i ; … ; X_n)

Pour n multiple de 4 ➡

- $Q1 = X_{n/4}$
- $Q3 = X_{3n/4}$

Pour n non multiple de 4 ➡

- $Q1 = (X_i + X_j) / 2$

→ *Avec i et j les deux valeurs les plus proches de n/4 : i < n/4 < j*

- $Q3 = (X_i + X_j) / 2$

→ *Avec i et j les deux valeurs les plus proches de 3n/4 : i < 3n/4 < j*

Le Tutorat est gratuit, toute reproduction ou vente est interdite

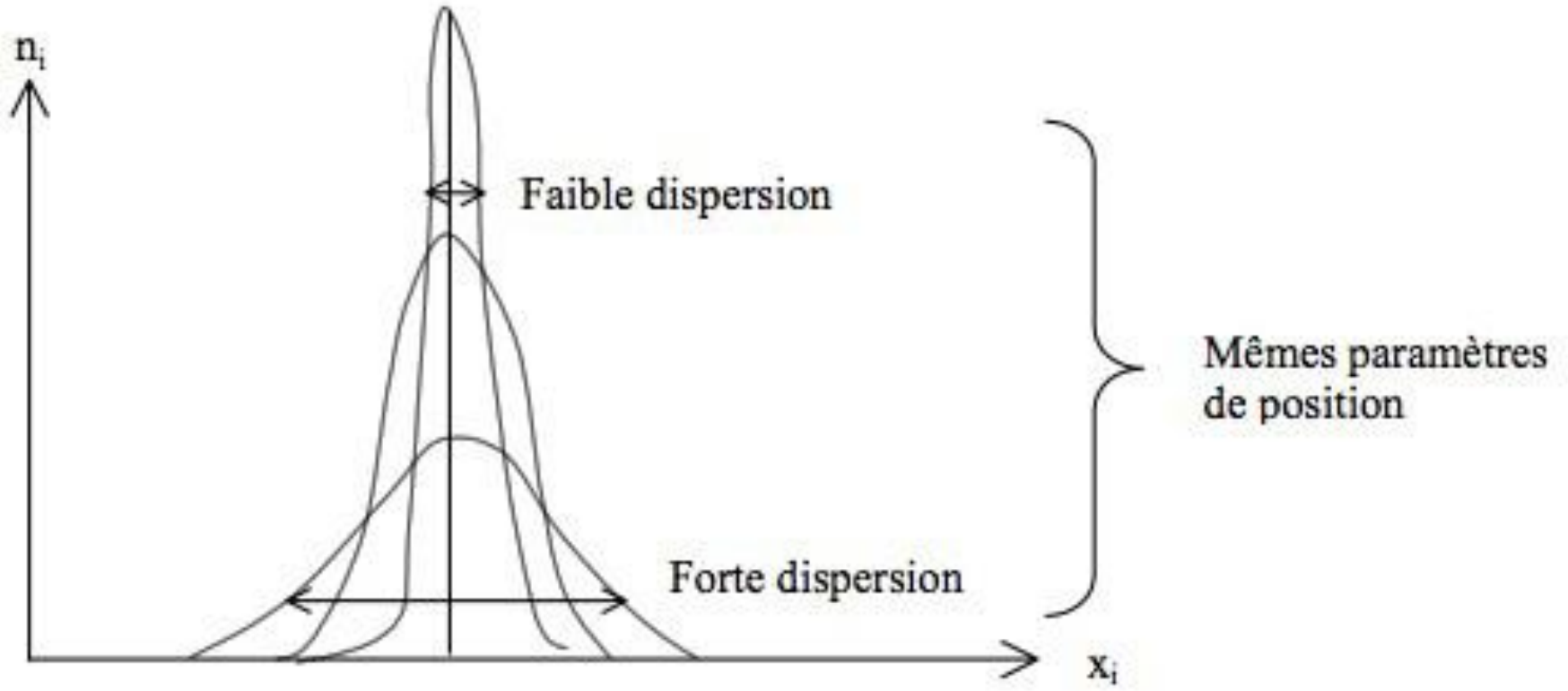
LA VARIANCE

- Indicateur de dispersion → dispersion des données autour de la moyenne
- (Ecart type)^2

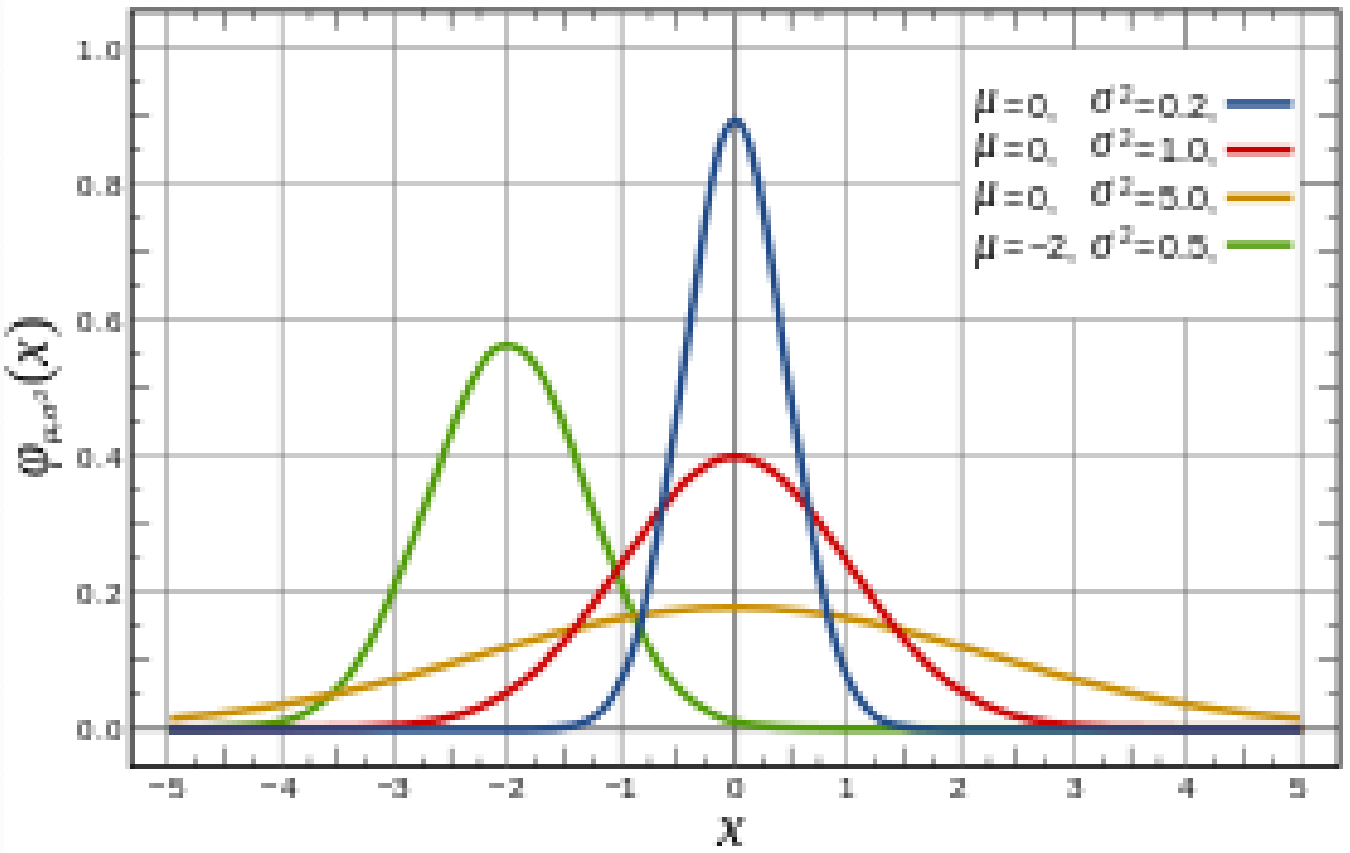
L'ECART TYPE σ

- « Moyenne de l'écart à la moyenne »
- Indicateur de dispersion

Le Tutorat est gratuit, toute reproduction ou vente est interdite



Le Tutorat est gratuit, toute reproduction ou vente est interdite



Le Tutorat est gratuit, toute reproduction ou vente est interdite

Nombre de degrés de liberté

- *Nombre des écarts indépendants (Xi – m)*
- Le nombre de degré de liberté (ou ddl) se traduit par le nombre minimal de données qu'il est nécessaire de connaître afin de pouvoir déduire toutes les données manquantes.
- Quand on veut remplir un tableau à N lignes et n colonnes, il faut connaître au minimum (N-1) x (n-1) données afin d'avoir toutes les données de ce tableau
- Il y a n écarts (Xi – m)
- Leur somme est égale à 0
- Il suffit d'en connaître (n-1) pour tous les connaître → n-1 degrés de liberté

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → ?

Médiane → ?

Q1 → ?

Q2 → ?

Q3 → ?

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → 10.85

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → 10.85

Médiane → 11

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → 10.85

Médiane → 11

Q1 → 7/4 = 1.75 → on fait la moyenne de la 1e et 2e valeur → 6.5

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → 10.85

Médiane → 11

Q1 → 7/4 = 1.75 → on fait la moyenne de la 1e et 2e valeur → 6.5

Q2 → = Médiane = 11

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Exercice

Numéro de l'élève	1	2	3	4	5	6	7
Notes	6	7	7	11	13	14	18

Moyenne → 10.85

Médiane → 11

Q1 → 7/4 = 1.75 → on fait la moyenne de la 1e et 2e valeur → 6.5

Q2 → = Médiane = 11

Q3 → 3*7/4 = 5.25 → on fait la moyenne de la 5e et 6e valeur → 13.5

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Moyenne Médiane
Le clash

	Avantages	Inconvénients
Moyenne	→ Facile à calculer → Adaptée aux calculs statistiques → Significative si : - répartition des données symétrique - dispersion faible (= faible écart type)	→ Sensible aux valeurs anormales ++
Médiane	→ Facile à calculer → Peu sensible aux valeurs anormales → Utilisable pour les valeurs ordinales	→ Peu adaptée aux calculs statistiques …

L'ESTIMATION STATISTIQUE

Déterminer une grandeur définie sur une population *à partir d'observations* effectuées *sur un échantillon représentatif* de cette population

Exemple → Combien de temps dure un séjour à l'hôpital en moyenne en France pour une pathologie donnée ?

Le Tutorat est gratuit, toute reproduction ou vente est interdite

2 TYPES D'ESTIMATION

- **Estimation ponctuelle**

→ Valeur jugée la meilleure à un instant T

→ Peu fiable…

- **Estimation par intervalle**

→ Intervalle de valeurs contenant la valeur recherchée

→ On admet un risque d'erreur α

→ On l'appelle « Intervalle de confiance » (IC)

ou « Intervalle au risque α » (avec α = 5% souvent)

→ Beaucoup plus fiable !

Le Tutorat est gratuit, toute reproduction ou vente est interdite

MÉTHODOLOGIE

- Déterminer précisément la population à étudier
  = Population cible
- Tirage au sort (TAS) d'un échantillon représentatif
- Etude de l'échantillon
- Extrapolation des résultats à la population

Estimation souvent par intervalle

Le Tutorat est gratuit, toute reproduction ou vente est interdite

- Soient A et B deux échantillons représentatifs d'une population :

- Deux **estimations ponctuelles** d'une même variable réalisées sur les échantillons A et B donneront des **valeurs ponctuelles voisines**, mais pas nécessairement les mêmes valeurs.

- Deux **estimations par intervalles** d'une même variable réalisées sur les échantillons A et B donneront des **Intervalles de confiance (IC) qui se recouvrent**, mais pas nécessairement les mêmes.

Le Tutorat est gratuit, toute reproduction ou vente est interdite

EXEMPLE : GLYCÉMIE MOYENNE SUR LA POPULATION FRANCAISE

→ Après TAS, constitution de 2 échantillons représentatifs A et B

	Echantillon A	Echantillon B
Estimation ponctuelle	0,95 g/L	1,03 g/L
Estimation par intervalle (95%)	[0,90 g/L ; 1,04 g/L]	[0,95 g/L ; 1,10 g/L]

⇒ Les estimations ponctuelles sont proches

⇒ Les intervalles de confiance se recouvrent

→ La valeur VRAIE de la glycémie moyenne a de fortes chances de se trouver dans un des intervalles de confiance ou dans celui recoupé [0.90 g/L ; 1.10 g/L]

Le Tutorat est gratuit, toute reproduction ou vente est interdite

NOTION D'INTERVALLE DE CONFIANCE

$$\mu \in \left[m \pm \frac{\varepsilon s}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

α = Probabilité de se tromper dans l'estimation de la moyenne μ

→ ε = Ecart réduit (différent pour chaque risque α choisi)

→ **Si α diminue, ε augmente !** → α = 5% → ε = 1.96
→ α = 1% → ε = 2.6

→ **Plus α est petit, plus l'intervalle est grand** (car ε augmente) !
On réussit plus souvent mais on prend un plus grand risque de se tromper…

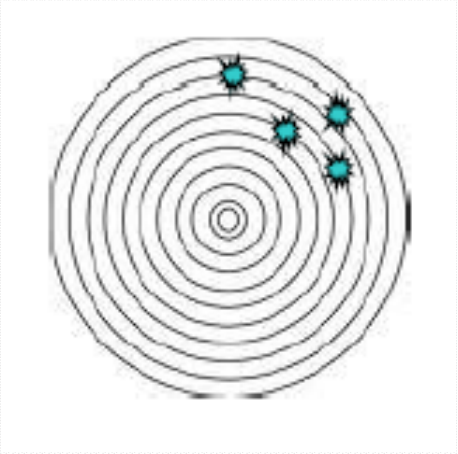
Le Tutorat est gratuit, toute reproduction ou vente est interdite

NOTION D'INTERVALLE DE CONFIANCE

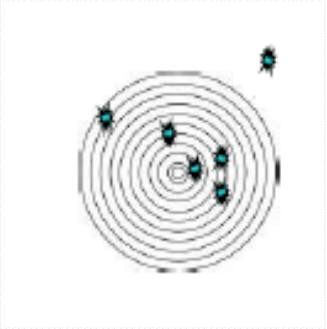
- Différents échantillons → Différentes estimations
- **Taille de l'échantillon augmente → Précision augmente**

- **Plus l'IC est large, moins il est précis !**

Large = plus de chances de l'atteindre, mauvaise précision de l'estimation



Resserré = meilleure précision de l'estimation



Le Tutorat est gratuit, toute reproduction ou vente est interdite

$$i = \varepsilon \frac{s}{\sqrt{n}}$$

Indice de précision i

→ permet de calculer la précision de l'estimation de m

→ c'est la largeur de l'intervalle de confiance !

→ i diminue = précision augmente

$$n = \varepsilon^2 \frac{s^2}{i^2}$$

Nombre de sujets nécessaires

→ pour une précision donnée

Le Tutorat est gratuit, toute reproduction ou vente est interdite

La souffrance sera présente mais la victoire n'en sera que plus belle…

« You never know how strong you are untill being strong is the only choice you have. »

Bob Marley

Le Tutorat est gratuit, toute reproduction ou vente est interdite

Place à Skinii !!!

Bon courage à tous

Tom_C pour vous servir

Le Tutorat est gratuit, toute reproduction ou vente est interdite