

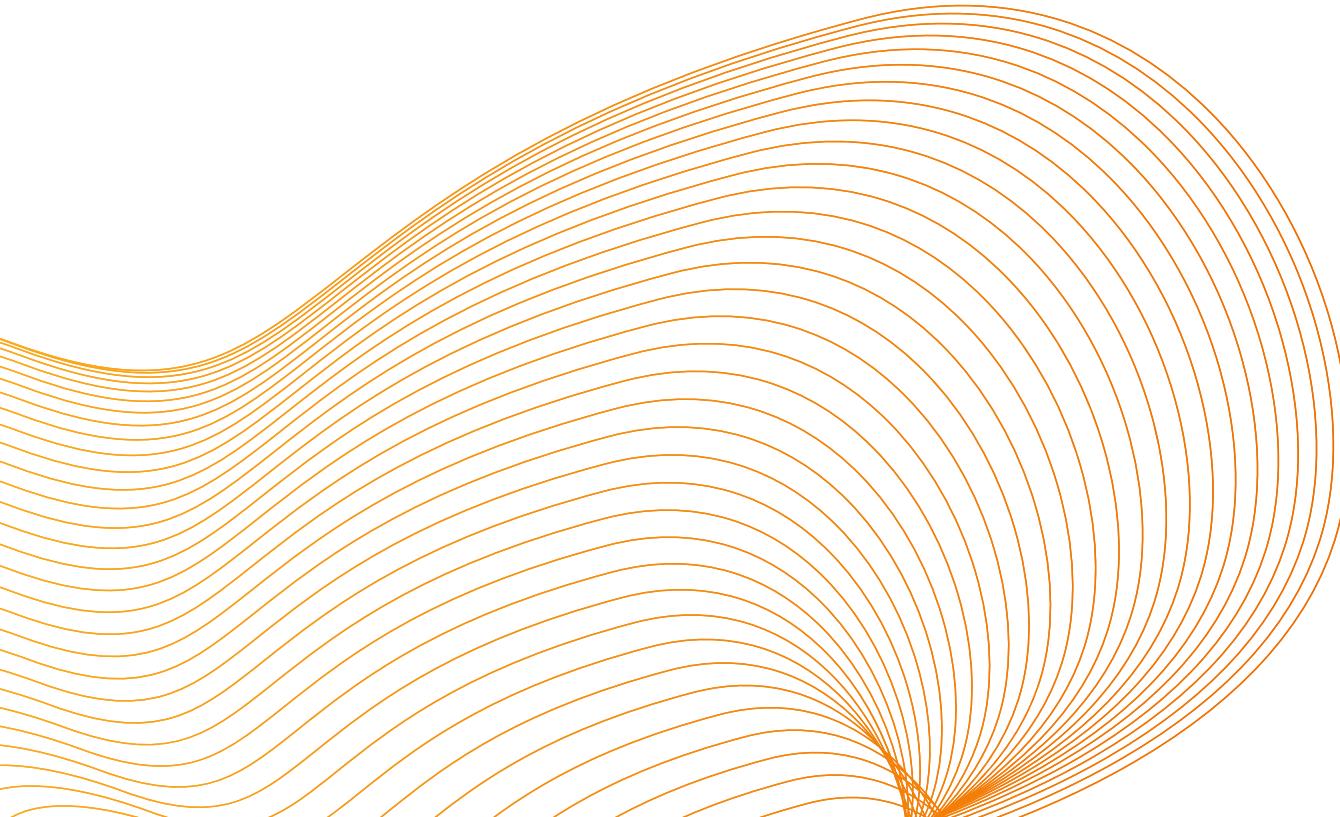


# STATISTIQUES

Présenté par :  
Aissatou DIOP  
Moussa DIOUF  
Serigne Modou DIOP  
Salimata Ba  
Daour TOP



# **PLAN**

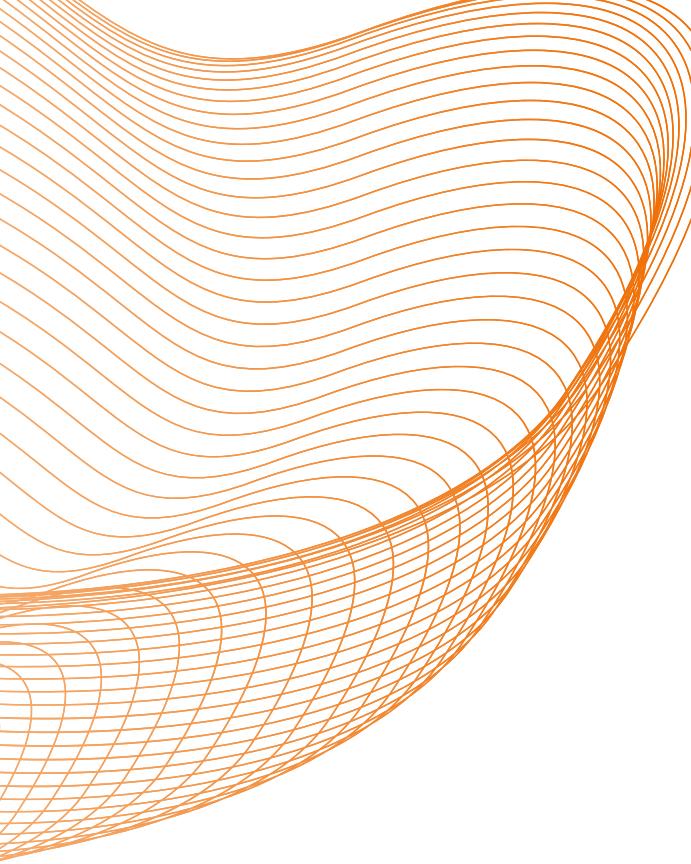


- STATISTIQUES DESCRIPTIVES
- CORRELATION
- VARIABILITE
- REGRESSION
- DISTRIBUTION
- ESTIMATIONS ET BIAIS

# LES STATISTIQUES

- Étude de la collecte de données
- Analyse et leur traitement
- Interprétation des résultats
- Présentation des données d'une manière facile à comprendre pour tout le monde.

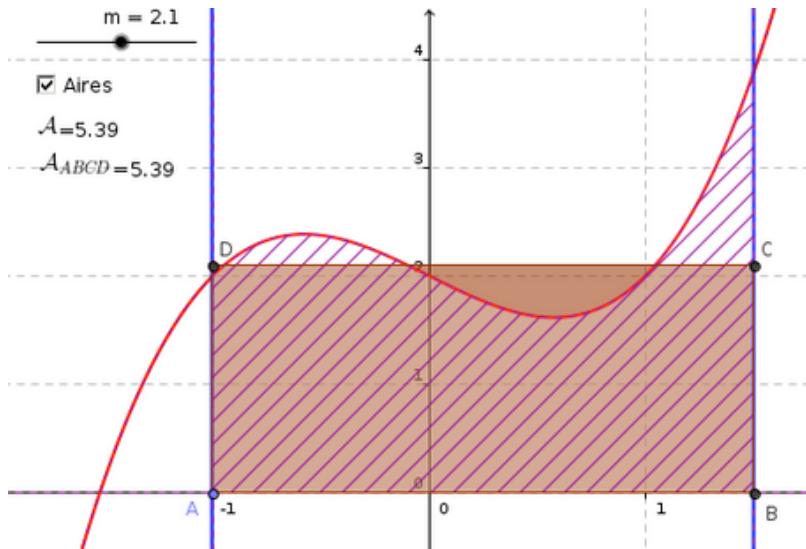




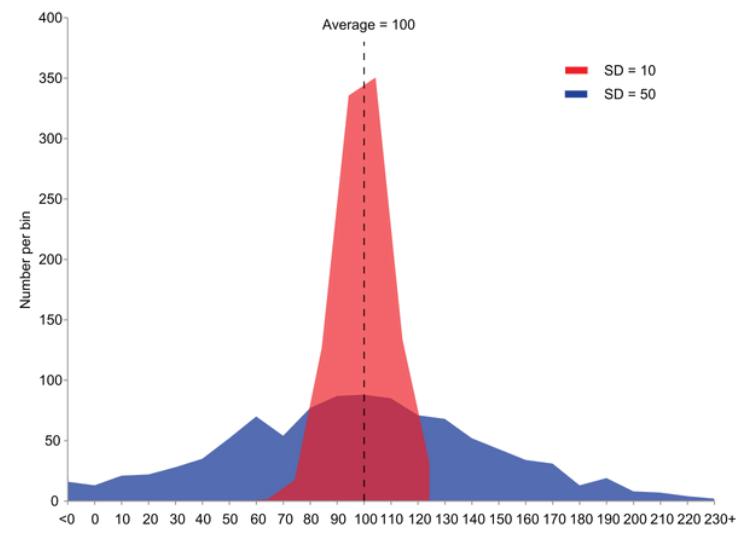
# STATISTIQUES DESCRIPTIVES

- Utiliser des données pour étudier des phénomènes.
  - Présentation les données (Synthèse)
  - Création de graphique
  - Production d' un résumé numérique.
- 

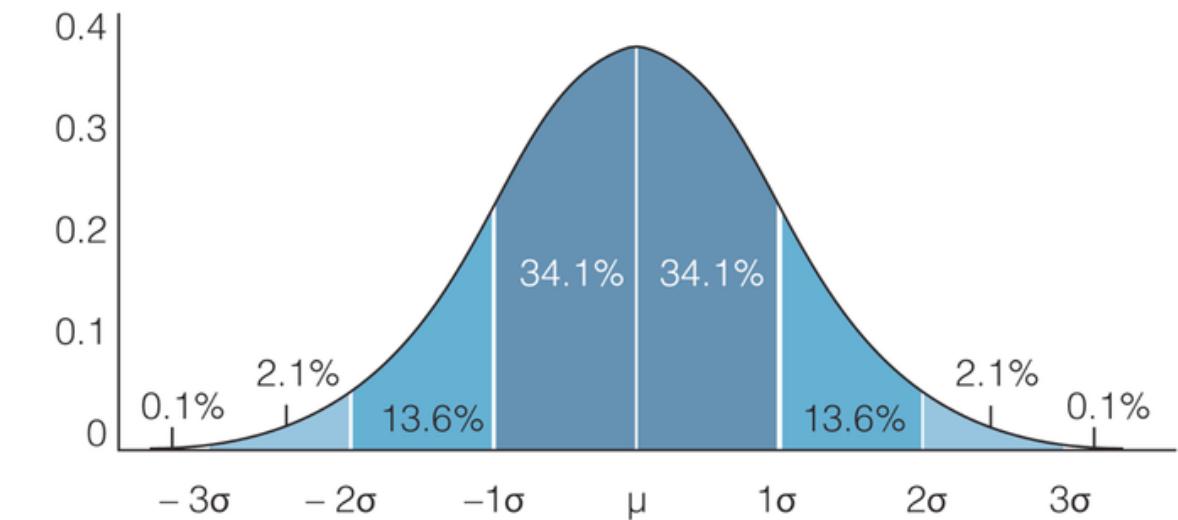
# ➤ Statistiques descriptives les plus couramment utilisées



**MOYENNE**



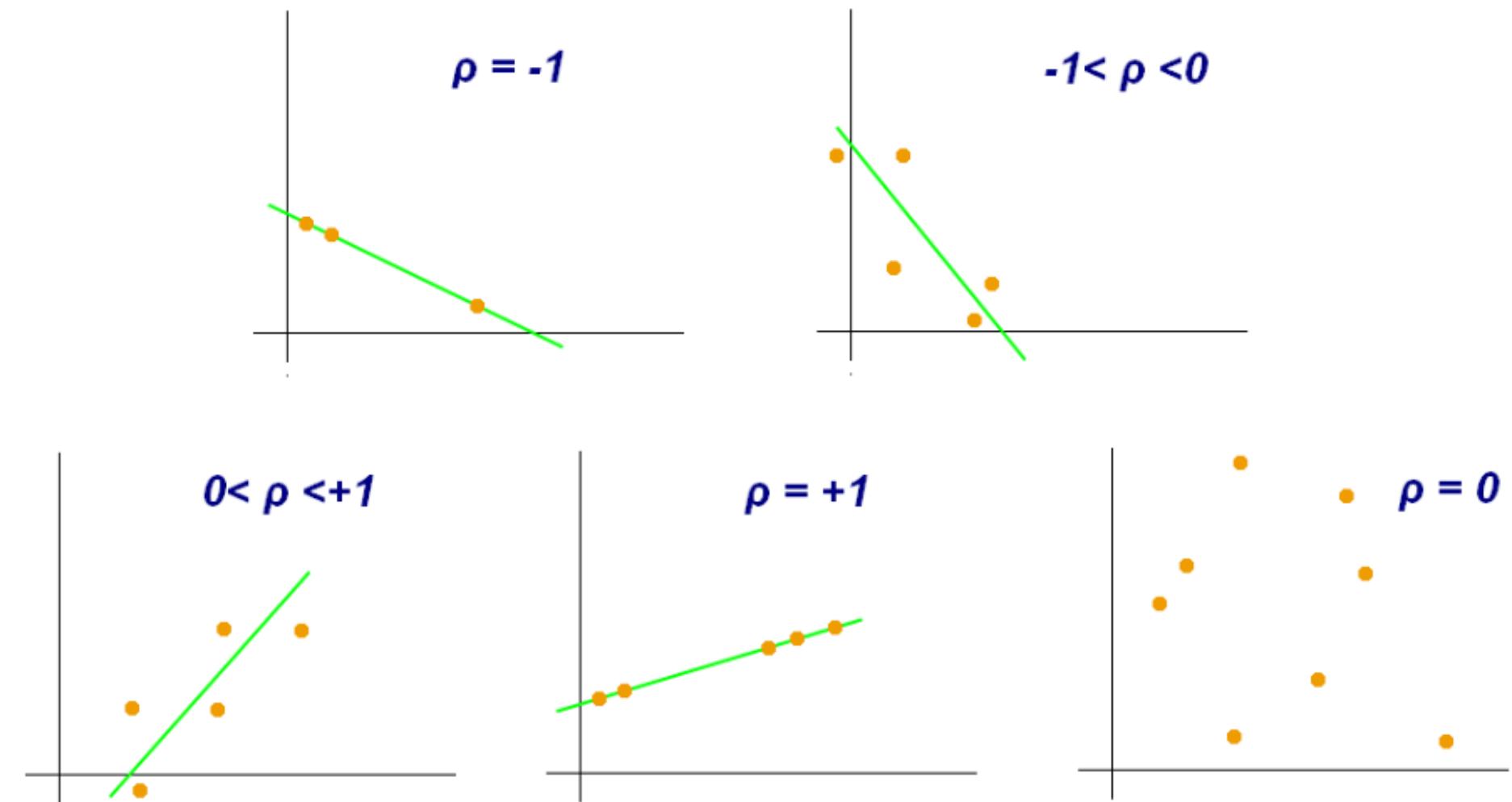
**MODE**



**MEDIANE**

# La corrélation

Le coefficient de corrélation ( $\rho_{XY}$ ) fournit une mesure de la force de la relation entre deux caractères et de leur orientation lorsque cette relation est monotone.



# VARIABILITÉ

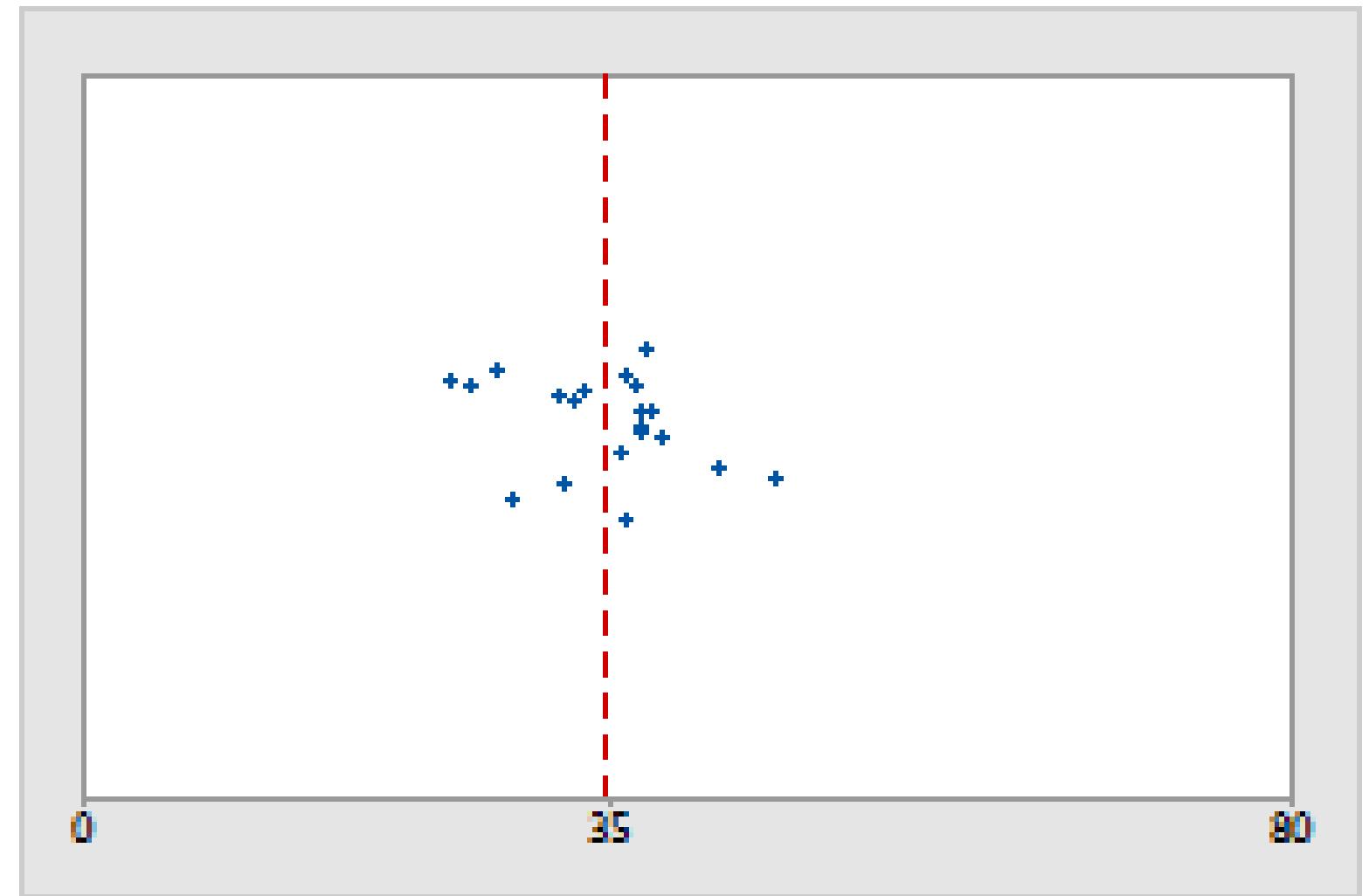
- La variabilité est une mesure de la dispersion ou de la différence entre les valeurs d'un ensemble de données numériques.
- Décrit la manière dont les valeurs de l'ensemble de données sont dispersées ou regroupées autour de la tendance centrale, ce qui peut avoir des implications importantes pour la prise de décision ou l'analyse des données.

# Mesures de variabilité

- Écart-type
- Variance
- Plage
- Percentile
- Quartile
- Intervalle interquartile

# ECART-TYPE

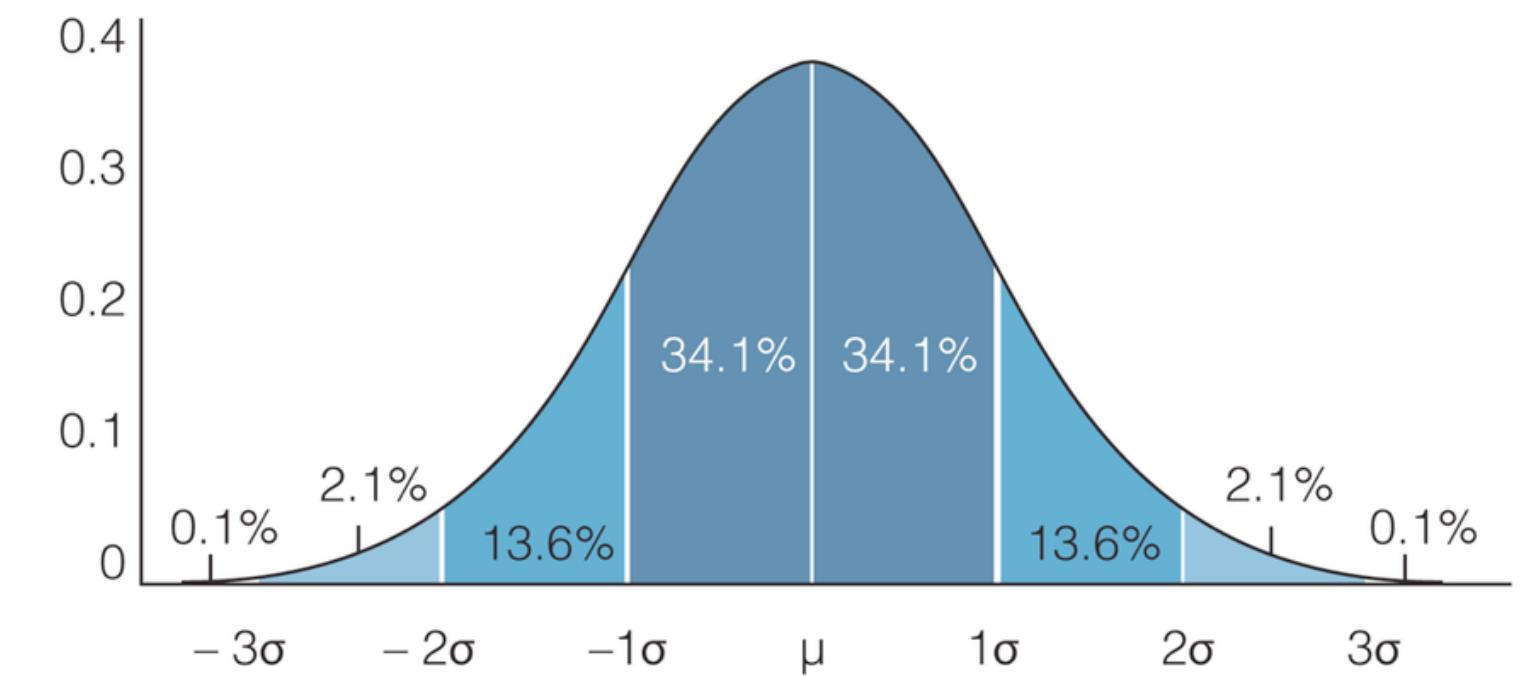
La distance moyenne de chaque observation  
par rapport à la moyenne



$$\sqrt{\frac{1}{n} \sum_i (x_i - \bar{x})^2}$$

# VARIANCE

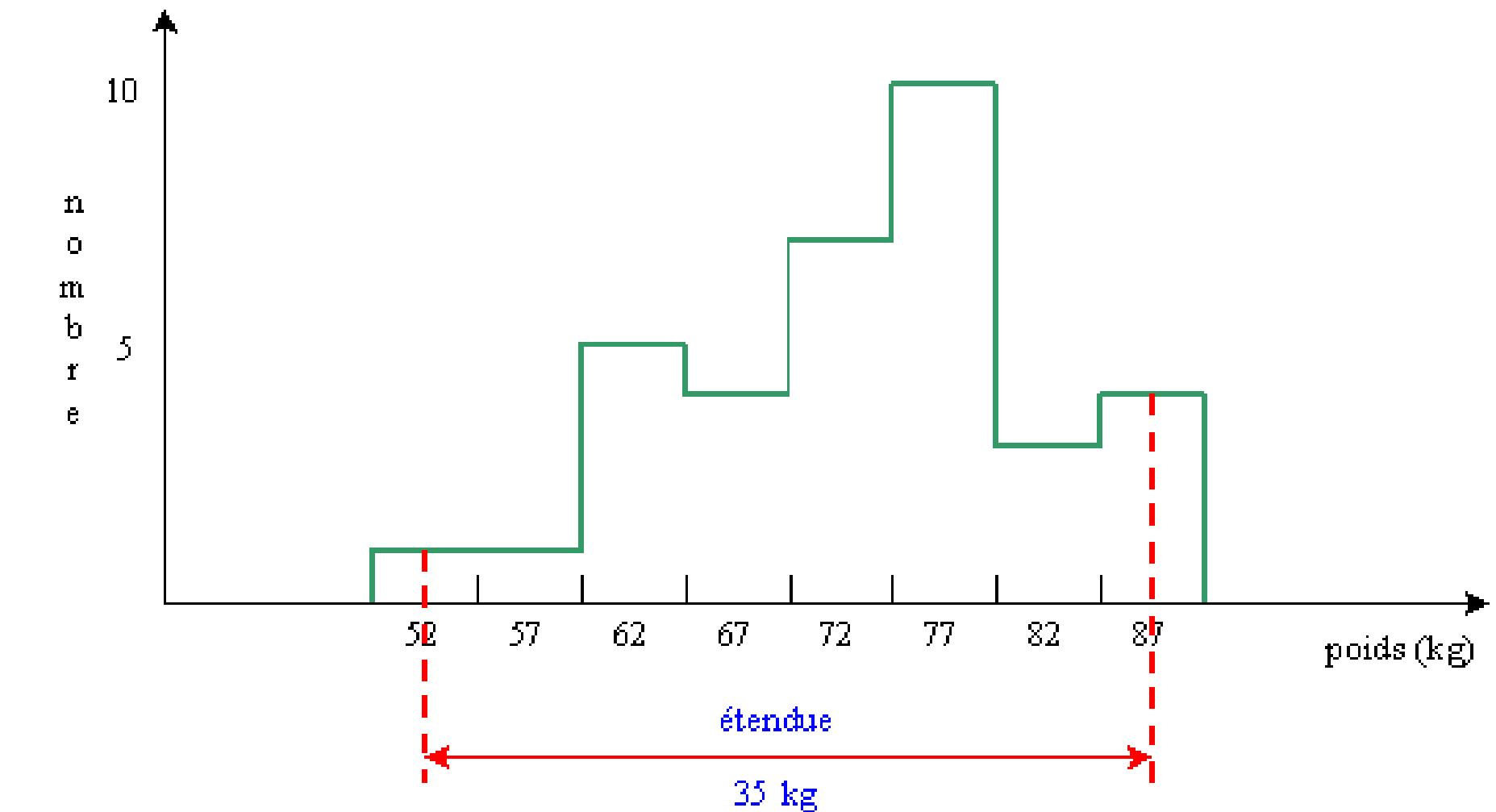
**La différence moyenne au carré de chaque observation par rapport à la moyenne, elle est le carré de l'écart-type**



$$\sigma^2 = \frac{\sum(\chi - \mu)^2}{N}$$

# PLAGE

La plage est la différence entre la plus grande valeur observée et la plus petite valeur observée dans un ensemble de données



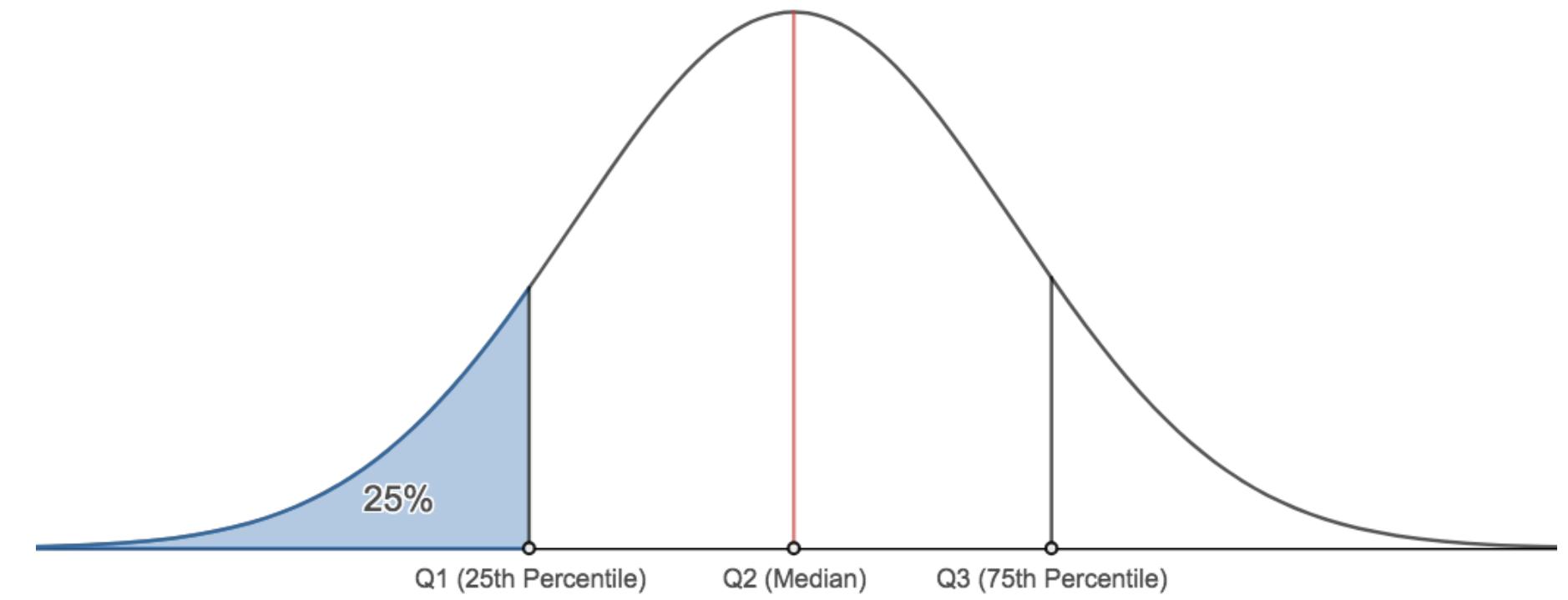
-5, -4, -4, -3, -3, -2, -1, 0, 0, 1, 2, 3, 3, 4, 4, 6, 7, 8, 9, 10, 10, 11, 11, 12

$x_{\min}$

$x_{\max}$

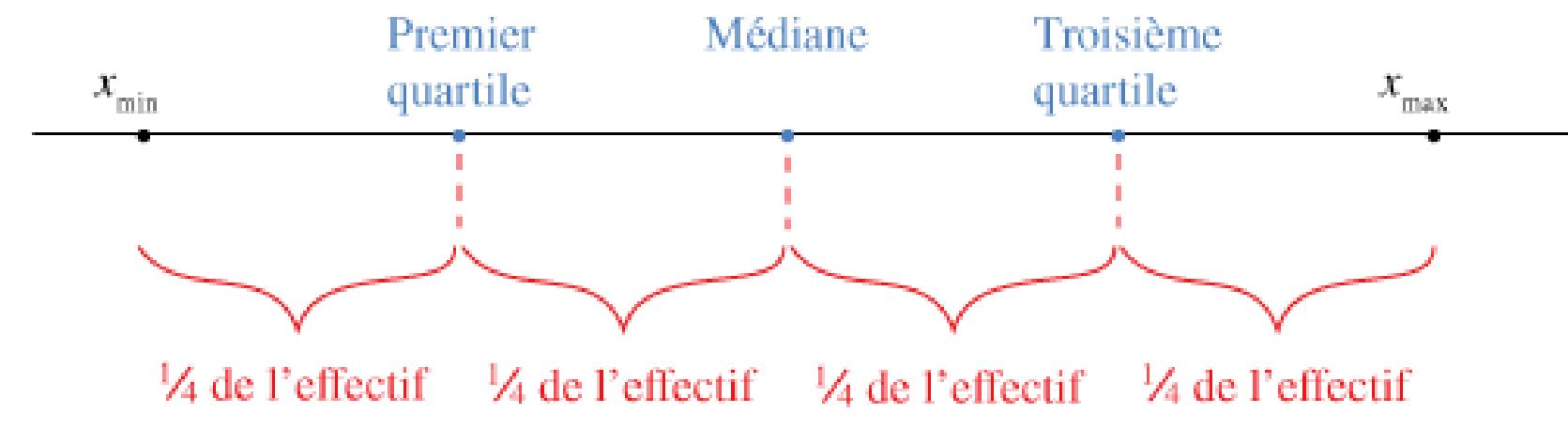
# PERCENTILE

Un percentile est un score qui divise une distribution en 100 parties égales



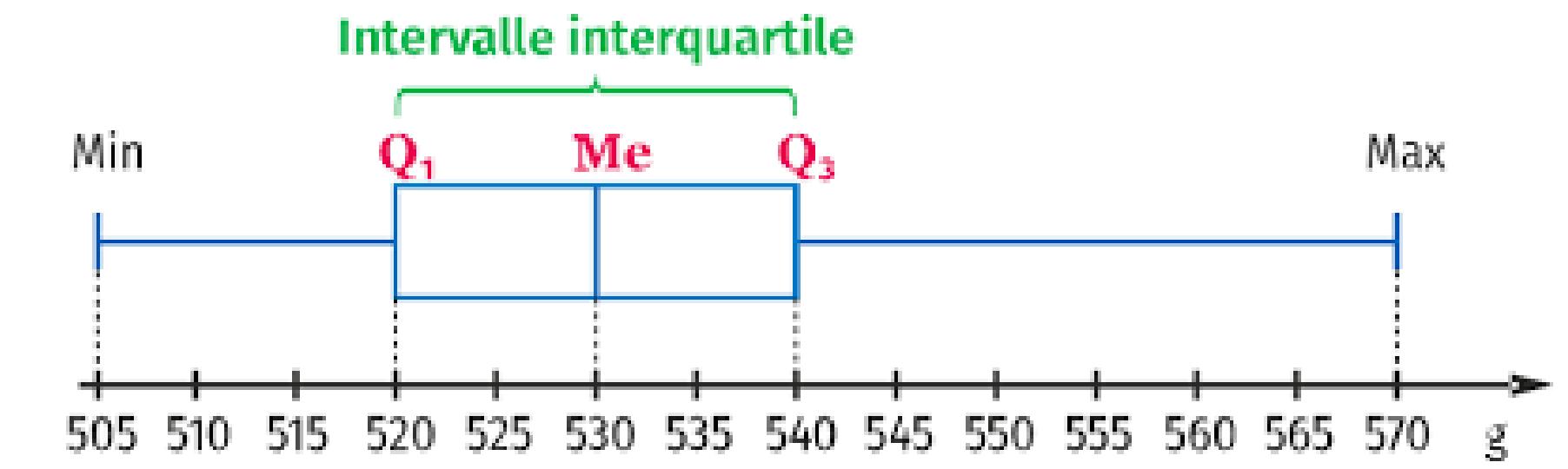
# Quartile

Les quartiles sont des mesures qui divisent l'ensemble de données en quatre parties égales



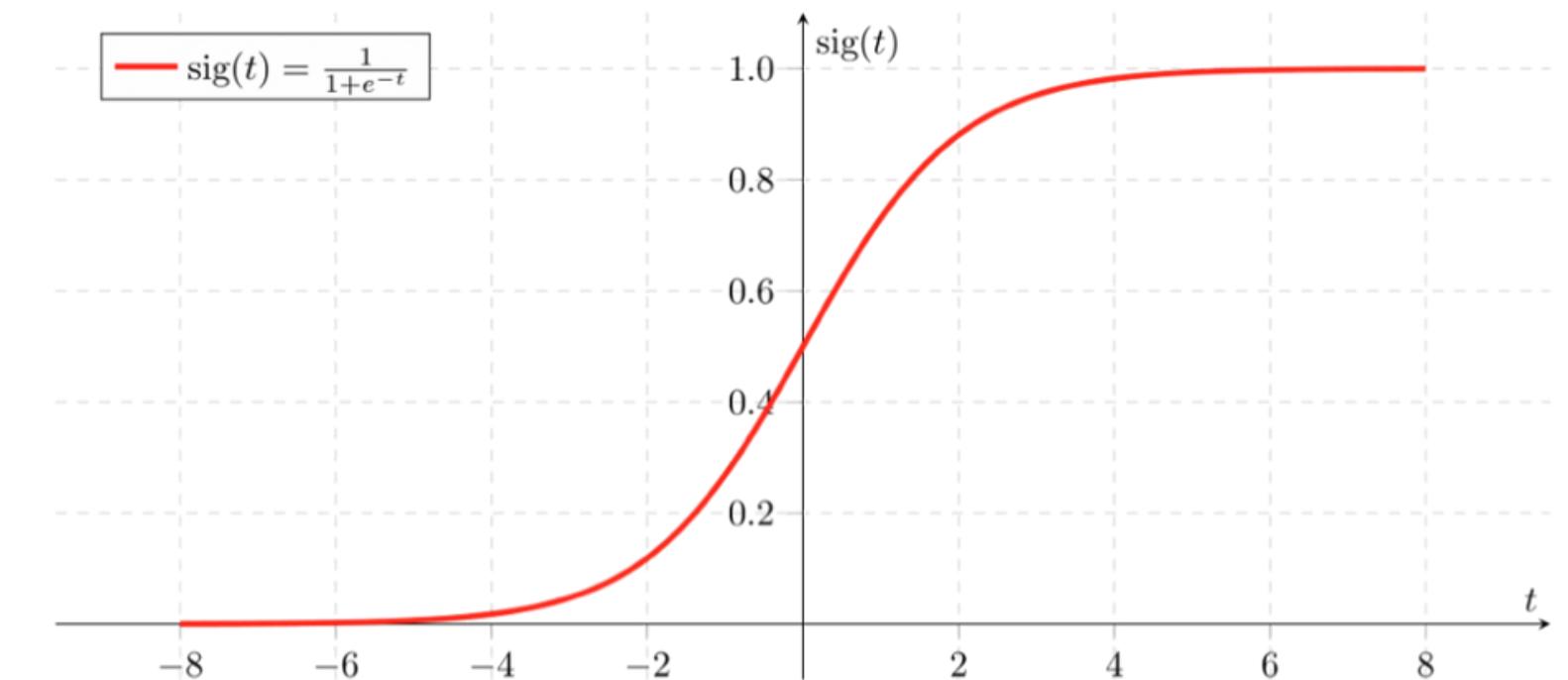
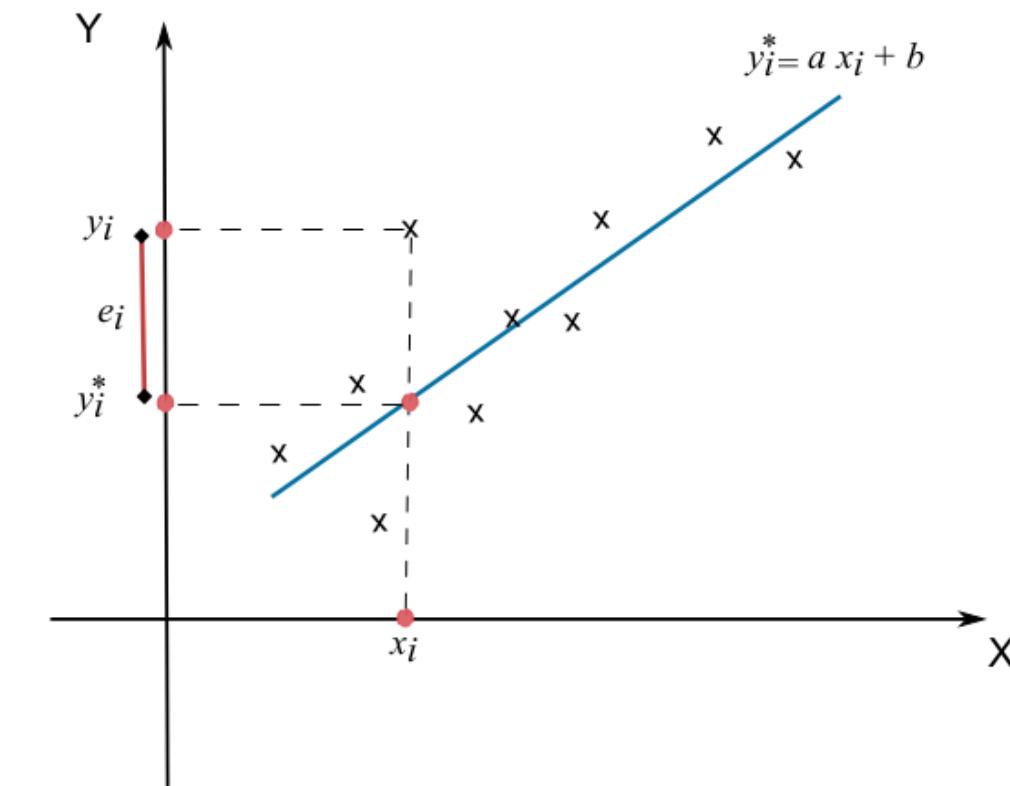
# Intervalle interquartile

Représente la plage de valeurs qui contient la moitié centrale de l'ensemble de données. Autrement dit, il englobe 50% des données les plus centrales.



# Régression

- La régression est basée sur l'idée que la variable dépendante est déterminée par une ou plusieurs variables indépendantes.
- La régression linéaire permet de prédire la valeur de la variable dépendante en fonction des valeurs des variables explicatives
- La régression logistique: il s'agit d'expliquer au mieux une variable binaire par des observations réelles nombreuses, grâce à un modèle mathématique.



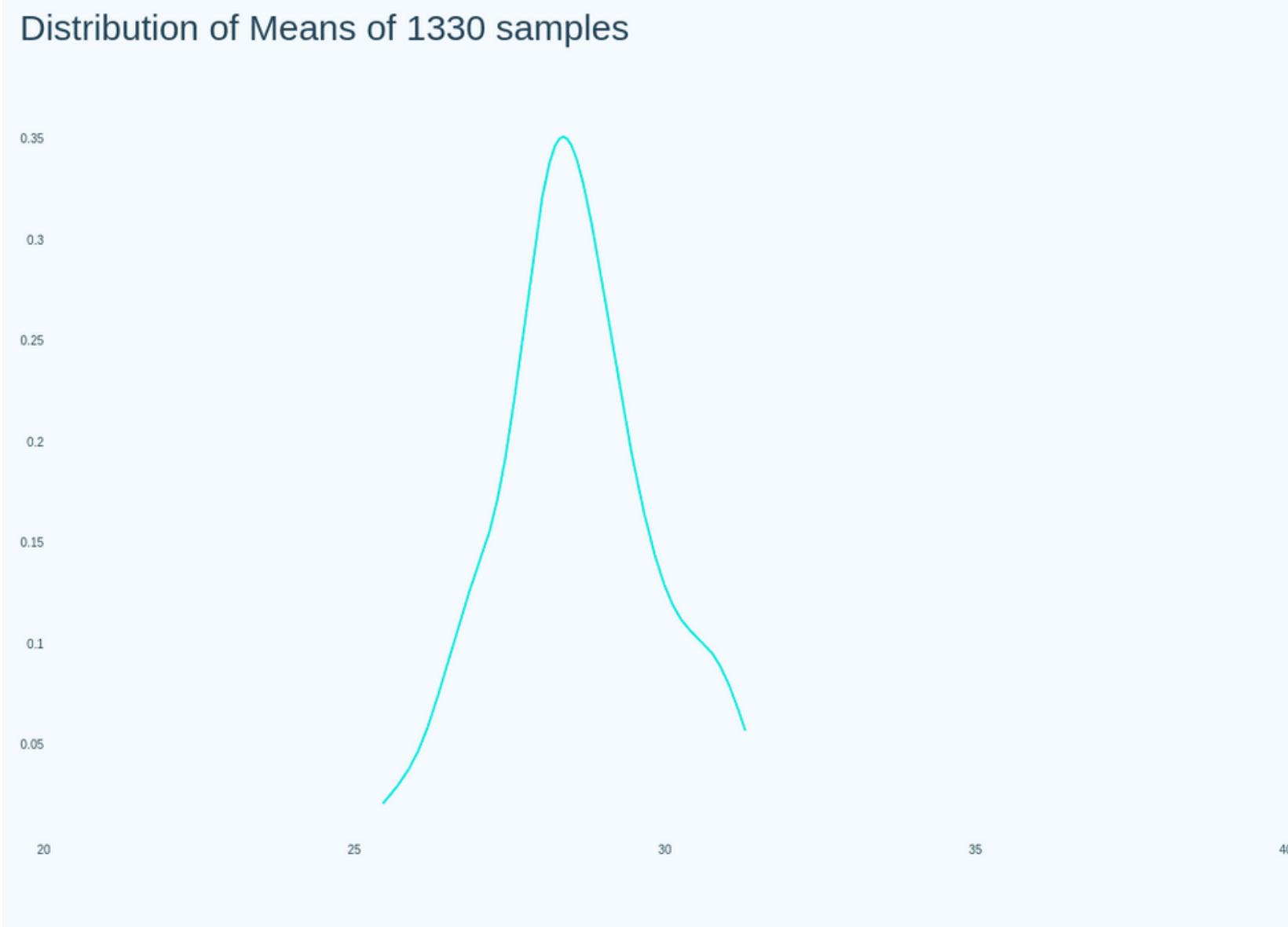
# **Distribution des probabilités**

- La distribution des probabilités est utilisée pour décrire la probabilité d'un événement ou d'une série d'événements.
- Elle permet de calculer la probabilité d'un résultat ou d'un ensemble de résultats possibles, en fonction de la distribution des valeurs possibles..

# **LES EVENEMENTS**

- Un évènement est dit indépendant s'il n'affecte pas la probabilité qu'un autre évènement se produise.
- Un événement est considéré comme dépendant lorsque la probabilité d'occurrence(fréquence ou nombre de fois ) de cet événement dépend de la probabilité d'occurrence d'autres événements

# Theoreme Centrale Limite



- Permet de modéliser beaucoup de phénomène continu que nous observons dans la vie
- Sa symétrie facilite le traitement statistique
- Le résultat le plus important c'est la convergence des lois usuelles vers la loi normal lorsque la taille de l'échantillon est très grande

# BIAIS

- Les biais ou Erreurs systématiques sont des sources d'erreur lors des analyses statistiques liées à la méthodologie d'une étude.
- Le biais de sélection se produit lorsqu'il y a une erreur dans le choix de l'échantillon
- Le biais d'intervalle de temps se produit lorsque les mesures sont prises à des moments différents pour différents groupes,
- Compromis biais/Variance est un compromis important en statistiques lors de la modélisation des données, qui consiste à trouver un modèle qui minimise à la fois le biais et la variance des prédictions.

