

# Classe Statistique

Etude d'une classe métier qui rends des services statistiques et qui vérifie la loi de répartition Normale d'un jeu de données

## CONTENU

presentation d'une classe de calcul statistique .....	1
création de la Classe statistique .....	2
1) Travail à faire : créer une classe métier qui a partir d'une collection générique de données (ARRAYLIST ou LINKEDLIST) de type DOUBLE PERMET de calculer la moyenne, La mediane, l'écart-type et les quartiles de la distribution statistique. ....	2
2) Tester cette classe avec le jeu de données sur les salaires qui vous EST COMMUNIQUE en 2 <sup>nd</sup> partie et conclure si la répartition des salaires dans l'informatique suis une loi ( distribution) normale ou pas ? .....	2

## PRESENTATION D'UNE CLASSE DE CALCUL STATISTIQUE

La **loi normale** (ou distribution de Gauss) est la star des statistiques. On l'appelle souvent la "courbe en cloche" car elle représente des phénomènes naturels où la majorité des individus se situe autour d'une moyenne, avec des exceptions de plus en plus rares à mesure que l'on s'en éloigne.

## 1. Les indicateurs centraux

Dans une distribution normale parfaite, la symétrie est totale. Cela simplifie énormément les choses :

- **La Moyenne  $\mu$**  : C'est le centre de gravité de la courbe.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

- **La Médiane** : C'est la valeur qui coupe la population en deux. Pour une loi normale :

### Moyenne = Médiane

La médiane est la valeur qui partage la population en deux groupes de 50 %.

1. La courbe de Gauss ( loi normale) est une "cloche" centrée sur  $\mu$ .

- **L'Écart-type ( $\sigma$ )** : Il mesure la dispersion (la "largeur" de la cloche). Plus il est grand, plus la cloche est aplatie.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

## 2. Les Quartiles

Les quartiles divisent vos données en quatre parts égales (25 % chacune). Dans une loi normale, ils sont directement liés à la moyenne et à l'écart-type via des coefficients fixes :

Quartile	Position	Formule de calcul
Premier Quartile (Q1)	25 % des données	$Q1 = \mu - 0,674 \sigma$
Deuxième Quartile (Q2)	50 % (Médiane)	$Q2 = \mu$
Troisième Quartile (Q3)	75 % des données	$Q3 = \mu + 0,674 \sigma$

**Note** : L'écart interquartile ( $QR = Q3 - Q1$ ) représente la zone où se trouvent les 50 % "centraux" de votre population.

----- Partie 2 -----

J'ai généré un jeu de données qui suit une **distribution réaliste** : une base solide de profils techniques (développeurs), quelques experts et une direction.

Voici la répartition par fonction avec les salaires nets mensuels associés.

### 1. Structure de l'entreprise (100 salariés)

Fonction	Nombre	Fourchette de salaire	Rôle type
Support / Admin	15	2 100€ - 2 800€	Assistant, RH Junior, Office Manager
Développeurs Junior	25	2 900€ - 3 400€	Dev Fullstack, Frontend (< 2 ans)
Développeurs Senior / Ops	35	3 800€ - 5 200€	Lead Dev, DevOps, Cloud Engineer
Data / Product Managers	15	4 500€ - 6 200€	Data Scientist, Product Owner Senior
Top Management / Experts	10	6 500€ - 9 500€	CTO, VP Engineering, Architecte

### 2. Les données brutes (triées par ordre croissant)

Pour faciliter vos calculs, voici les 100 valeurs regroupées par tranches :

#### Tranche 2000€-3000€ (22 salaries) :

2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2900, 2950, 2950, 3000, 3000, 3000, 3000.

#### Tranche 3000€-4000€ (28 salaries) :

3100, 3100, 3150, 3150, 3200, 3200, 3250, 3250, 3300, 3300, 3350, 3350, 3400, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 3900, 3950, 3950, 4000.

#### Tranche 4000€-5500€ (30 salaries) :

4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 4950, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, 5500, 5500.

#### Tranche 5500€-9500€ (20 salaries) :

5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6700, 6900, 7100, 7400, 7700, 8000, 8300, 8700, 9100, 9500.

## CREATION DE LA CLASSE STATISTIQUE

- 1) TRAVAIL A FAIRE : CREER UNE CLASSE METIER QUI A PARTIR D'UNE COLLECTION GENERIQUE DE DONNEES (ARRAYLIST OU LINKEDLIST) DE TYPE DOUBLE PERMET DE CALCULER LA MOYENNE, LA MEDIANE, L'ECART-TYPE ET LES QUARTILES DE LA DISTRIBUTION STATISTIQUE.

Demander au formateur la validation de votre classe « Statistique »

- 2) TESTER CETTE CLASSE AVEC LE JEU DE DONNEES SUR LES SALAIRES QUI VOUS EST COMMUNIQUE EN 2<sup>ND</sup> PARTIE ET CONCLURE SI LA REPARTITION DES SALAIRES DANS L'INFORMATIQUE SUIV UNE LOI ( DISTRIBUTION) NORMALE OU PAS ?

--- FIN DU DOCUMENT ---