**Statistiques**

**I. Introduction**

Il y a deux catégories de données :

* Mesurables (numérique 🡪 âge, taille)
* Non-Mesurables (alphanumériques 🡪 prénom, sexe)

Pour les données de catégorie, on parle aussi de données **qualitatives** tandis que pour les mesures, on parle de données **quantitatives**.

La statistique est donc à l'origine la science qui doit éclairer l'Etat dans ses **décisions** et la gestion de ses ressources, aujourd'hui elle trouve des applications dans **tous les champs techniques et scientifiques**.

Sur le terrain de la collecte de l'information, les choses changent radicalement lorsque les échantillons statistiques deviennent **très gros**. En effet, dans ce cas les lois statistiques s'observent très fidèlement et les phénomènes de masses deviennent **prédictibles** et **quantifiables** (Prescriptive Analytics).

# II. Moyenne, médiane, variance, écart type, interquartiles et barres d'erreur

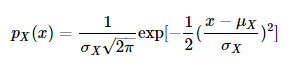
Il existe plusieurs quantités qui permettent de localiser un échantillon de données :

* **la moyenne** (moyenne empirique, moyenne de l’échantillon),
* **la médiane**
* le **mode**
* les **quartiles**

## 1. Moyenne et espérance

* La **moyenne**  renvoie au champ des **statistiques**
* **L'espérance** renvoie au champ des **probabilités**

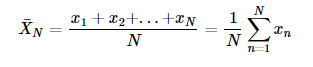
### Si X est une v.a qui résulte de la somme d'un grand nombre d’événements indépendants et de même loi cette v.a tend à se distribuer selon la fonction de Gauss (en forme de cloche) dont l'expression mathématique est :



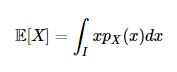
### Moyenne empirique

Soit un échantillon de taille N : 

La moyenne empirique de cet échantillon est donnée par :



### Espérance d'une v.a de loi continue

L’espérance de X est donne par :

Pour une **loi discrète** de densité P(x=k) = pk:

## 

## Exemple pour une loi uniforme sur I = [0,1]

La densité de probabilité d'une loi uniforme sur [0,1] est une constante. C'est le sens qu'il faut donné ici **uniforme**. Il n'y a aucune région de l'intervalle [0,1] qui soit privilégie, la variable aléatoire X peut tomber avec la même probabilité dans deux sous- intervalles de I qui ont même mesure.

## 2. Médiane

La médiane est la valeur qui sépare une population ou un échantillon en deux parties égales

### Médiane d'une loi continue

X variable aléatoire continue (définie sur R) 

### Médiane d'un échantillon de taille N

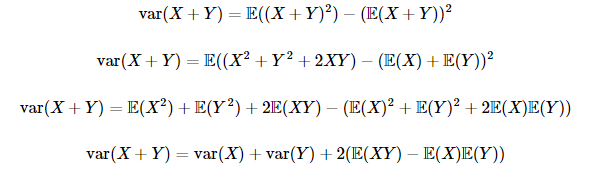
Pour déterminer la médiane d'un ensemble de valeurs, il suffit d'ordonner les valeurs en une liste croissante et de choisir la valeur qui est au centre de cette liste. Dans le cas d'un nombre impair de point la médiane appartient au jeu de données dans le cas d'un nombre pair, elle n'appartient pas au jeu de données.

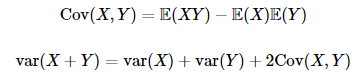
## 3. Dispersion d'une observation

### Variance et écart type

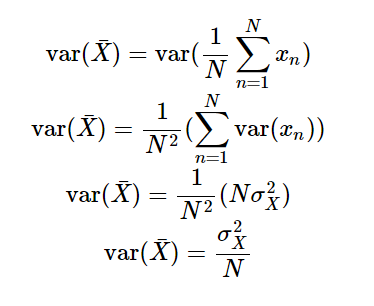
**Propriété de la variance :**

****

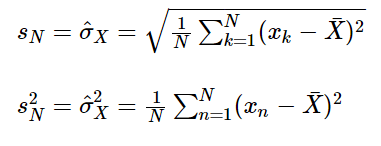
****

****

**Variance de la moyenne empirique :**

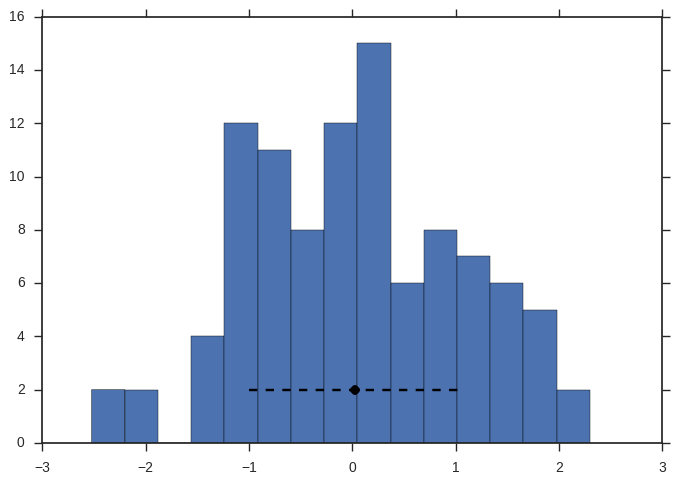
****

**Estimateur empirique de la variance :**

****

**Histogramme :**

Dans chaque sous intervalle, on compte le nombre de valeurs observes, et on reporte cet entier sous la forme d'un rectangle de hauteur proportionnelle cette valeur.

****