Initiations aux calculs statistiques

1-Calcul de la moyenne (caractéristique de position) :

On calcule la moyenne d'une variable aléatoire réelle ou statique en additionnant les valeurs de toutes les observations, puis en divisant cette somme par le nombre d'observations. Ce calcul permet d'obtenir la valeur moyenne de toutes les données.

Moyenne = Somme de toutes les valeurs d'observation ÷ nombre d'observations

On a $\overline{X} = \frac{\sum x}{n}$ où **x** représente une valeur observée, n représente le nombre d'observations incluses dans l'ensemble de données, Σ **x** représente la somme de toutes les valeurs x observées, $\overline{X}$ représente la valeur moyenne de **x**.

2-Calcul de la variance(caractéristique de dispersion) :

Soit la série statistique définie dans le tableau suivant :
La variance permet de combiner toutes les valeurs à l'intérieur d'un ensemble de données afin d'obtenir une mesure de dispersion. La variance (symbolisée par **V**) et l'écart-type (la racine carré de la variance, symbolisée par **S**) sont les mesures de dispersion les plus couramment utilisées.

V = moyenne de l'écart au carré de valeurs par rapport à la moyenne

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \quad \text{avec } m = \overline{X} \text{ la moyenne de la variable X}$$

Pour l’écart type il suffit de calculer la racine de la variance,

Écart-type (S) = Racine carrée de la variance

$$s(X) = \sqrt{V(X)}$$

3-Calcul de la corrélation :

La corrélation entre deux variables aléatoires est définie par l'exemple des actifs financiers, elle représente l'intensité de la liaison qu'il existe entre ces deux variables.

$$\rho(X;Y) = \frac{\Sigma(X-\bar{X}).(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}.\sqrt{\Sigma(Y-\bar{Y})^2}} = \frac{Cov(X;Y)}{\sqrt{V(X)}.\sqrt{V(Y)}}$$

X *et* Y *: deux variables aléatoires*

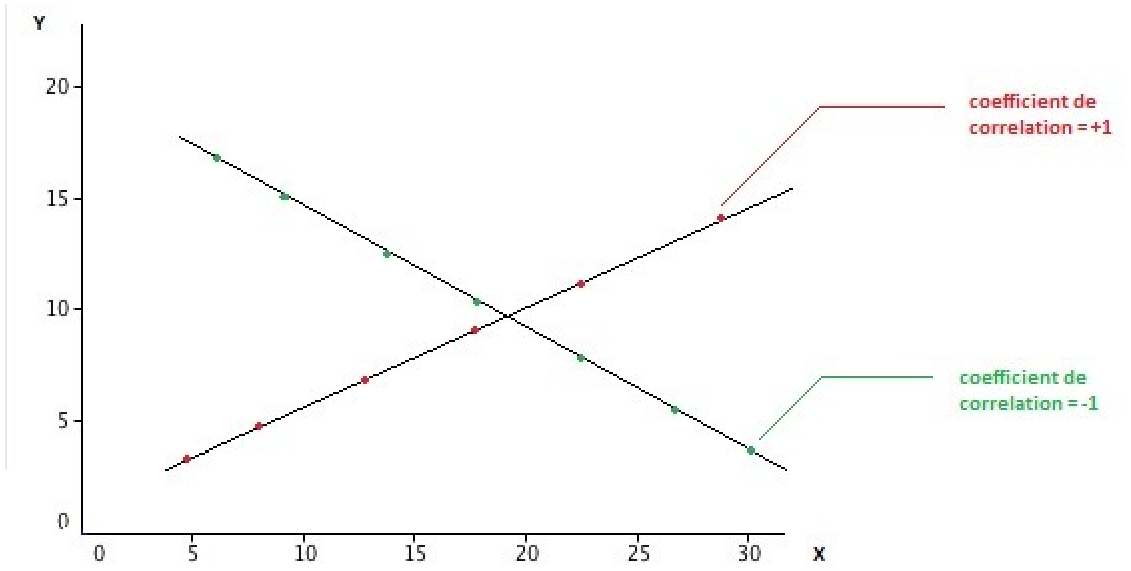
$\bar{X}$ *et* $\bar{Y}$ *: moyenne respective des variables* X *et* Y

$Cov(X;Y)$ *: Covariance entre* X *et* Y

$V(X)$ *et* $V(Y)$ *: Variance respective de* X *et* Y

L'interprétation du résultat est relativement simple. Il est tout toujours compris entre +1 et -1.

Plus la valeur $\rho(X;Y)$ est proche de -1 ou +1 plus les variables sont susceptibles d'être liées par une relation linaire de la forme y=ax+b (a et b sont des réels) ,



4-Calcul du décile et quartile :

Les quartiles permettent de séparer une série statistique en quatre groupes de même effectif (à une unité près).
Un quart des valeurs sont inférieures au premier quartile Q1.Tel que $\#\{Xi \leq Q1\}$=N/4
Un quart des valeurs sont supérieures au troisième quartile Q3.
on calcule la quantité ¼ de N qui vaut N/4
Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non

Cas n°1: le résultat est entier (la division tombe juste) :

-on vérifie que les valeurs sont rangées par ordre croissant

-Q1 est la n ème valeur de la série X où n = N/4

-Q3 est le n' ème valeur où l' entier n'=3×N/4

Cas n°2: le résultat n'est pas entier

-on vérifie que les valeurs sont rangées par ordre croissant

-on arrondit le décimal N/4 à l'entier supérieur: l'entier n = N/4 ; Q1 est la n ème valeur

-on arrondit le décimal 3×N/4 à l'entier supérieur: l'entier n' =3×N/4 ; Q3 est la n' ème valeur

Comment interpréter des quartiles donnés?
Si on connaît les quartiles Q1 et Q3 d'une série, que peut-on en déduire?
Au moins un quart (25%) des valeurs sont inférieures ou égales à Q1.
Au moins trois quarts (75%) des valeurs sont inférieures ou égales à Q3.
L'interquartile Q3-Q1 est une mesure de dispersion.

Les déciles permettent de séparer une série statistique en dix groupes de même effectif (à une unité près).
Un dixième des valeurs sont inférieures au premier décile D1.
Un dixième des valeurs sont supérieures au neuvième décile D9.

On calcule le 1 dixième de N qui vaut N/10

Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non

Cas n°1: le résultat est entier (la division tombe juste):

-on vérifie que les valeurs sont rangées par ordre croissant

-D1 est la n ème valeur où n = N/10

-D9 est le n' ème valeur où l' entier n' =9×N/10

Cas n°2: le résultat n'est pas entier

-on vérifie que les valeurs sont rangées par ordre croissant

-on arrondit le décimal N/10 à l'entier supérieur: l'entier n = N/10 ; D1 est la n ème valeur

-on arrondit le décimal 9×N/10 à l'entier supérieur: l'entier n' =9×N/10 ; D9 est la n' ème valeur.

Dk est K ème valeur telle que $\#\{Xi < Dk\} \leq K*N/10$

5-Calcul de la fréquence cumulée:

On utilise la *fréquence cumulée* pour déterminer le nombre d'observations qui se situent au-dessus (ou au-dessous) d'une valeur particulière dans un ensemble de données. On calcule la fréquence cumulée à l'aide d'un tableau de distribution de fréquences, qu'on peut construire à partir de diagrammes à tiges et à feuilles ou directement à partir des données.

On calcule la fréquence cumulée en ajoutant chaque fréquence tirée d'un tableau de distribution de fréquences à la somme de celles qui précèdent. La dernière valeur sera toujours égale au total des observations, puisque toutes les fréquences auront déjà été ajoutées au total précédent.

Exemple :

On considère les résultats d'un devoir d'une classe de 25 élèves :
11, 13, 8, 16, 11, 7, 13, 12, 11, 12, 12, 8, 11, 8, 12, 11, 11, 7, 16, 8, 11, 8, 12, 8, 16
que l'on range dans un tableau.

Notes	7	8	11	12	13	16	total
Effectifs	2	6	7	5	2	3	25
Fréquences	0,08	0,24	0,28	0,20	0,08	0,12	1
Fréquences cumulées	0,08	0,32	0,60	0,80	0,88	1	