Fiche P3.1 : Séries statistiques à deux variables

1 – Série statistique double

● On étudie simultanément deux **caractères** au sein d'une même **population**.

● Une **série statistique à deux variables** est constituée de deux listes de n **valeurs** $x_1, \dots ,x_n$ et $y_1, \dots ,y_n$:

Valeurs de x_i	x_1	x_2	x_3	…	x_n
Valeurs de y_i	y_1	y_2	y_3	…	y_n

● Dans certaines situations, il semble exister un lien entre les deux caractères : On parle alors de **corrélation.** On ne peut cependant pas affirmer que ce lien est lien de cause à effet (**causalité**).

Exemple 1 : Une étude statistique française donne une correspondance entre le prix moyen x_i d'un paquet de 20 cigarettes (en €) et le nombre y_i de cigarettes vendues chaque année (en dizaines de milliards). Les données sont regroupées dans le tableau ci-dessous :

Années	2001	2002	2003	2006	2012	2013	2016	2018
Prix : (x_i)	3.4	3.6	4	5	6.3	6.7	7	8
Nombre total : (y_i)	8.3	8.1	7	5.5	5.2	4.8	4.5	4

2 – Nuage de points, Point moyen

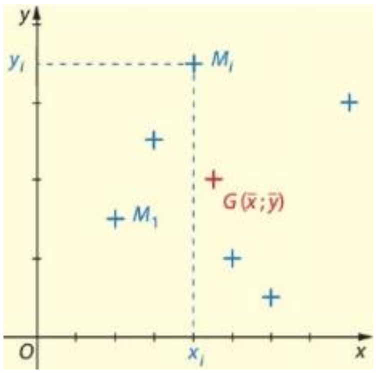
Définition 1 : On se place dans un repère du plan

● Une série statistique à deux variables se représente par un **nuage de points** de coordonnées $M_i(x_i; y_i)$

● On appelle **point moyen** du nuage le point $G(\bar{x}; \bar{y})$ où :

. $\bar{x}$ est la moyenne de la 1e variable : $\bar{x} = \frac{x_1+\dots+x_n}{n}$

. $\bar{y}$ est la moyenne de la 2e variable : $\bar{y} = \frac{y_1+\dots+y_n}{n}$



Exemple 1 (Suite) :

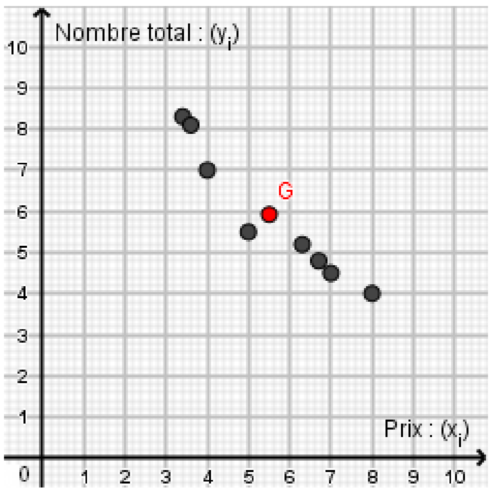
1) Représenter dans le repère ci-dessous, le nuage de points $M_i(x_i; y_i)$ associé à la série ci-dessus.

2) Calculer les coordonnées du point moyen G du nuage, puis placer le dans le repère.

$\bar{x} = \frac{3.4+3.6+\dots+8}{8} = 5.5$

$\bar{y} = \frac{8.3+8.1+\dots+4}{8} = 5.93$

Le point moyen est donc $G(5.5; 5.93)$.



3 – Covariance

Commençons par rappeler la définition de la variance (et de l'écart type) pour une série simple :

Définition 1 : On considère une série statistique $x = (x_1; \dots ; x_n)$ de moyenne $\bar{x}$.

● La **variance** est la moyenne des écarts au carré entre chaque valeur de la série et sa moyenne $\bar{x}$:

$$V(x) = \frac{(x_1-\bar{x})^2+\dots+(x_1-\bar{x})^2}{n}.$$

● **L'écart-type** est alors le nombre $\sigma(x) = \sqrt{V(x)}$

Remarque :

● Dans la formule de la variance, le carré sert à obtenir un nombre positif qui mesure la distance entre chaque valeur et la moyenne. La racine carrée dans la formule de l'écart-type permet alors de revenir dans la même unité que les valeurs de la série. En pratique, on utilise donc plutôt l'écart-type que la variance.

● Ces deux paramètres mesurent la **dispersion** des valeurs de la série autour de leur moyenne : Plus l'écart-type est grand, plus les écarts avec la moyenne sont importants et plus les valeurs seront éloignées les unes des autres.

Définition 2 : On considère une série statistique à deux variables $x = (x_1; \dots ; x_n)$ et $y = (y_1; \dots ; y_n)$ de moyenne respective $\bar{x}$ et $\bar{y}$. La **covariance** de x et y notée $cov(x; y)$ est le nombre défini par :

$$cov(x; y) = \frac{(x_1-\bar{x})(y_1-\bar{y})+(x_2-\bar{x})(y_2-\bar{y})+\dots+(x_n-\bar{x})(y_n-\bar{y})}{n} = \frac{x_1y_1+x_2y_2+\dots+x_ny_n}{n} - \bar{x}\bar{y}$$

Remarque : La covariance mesure si ces 2 variables « covarient » ensemble par rapport à leurs moyennes : Si les écarts à la moyenne ont tendance à être dans le même sens pour les deux variables alors la covariance sera positive ; s'ils ont tendance à être dans le sens opposé alors la covariance sera négative. Plus la valeur absolue de la covariance est importante, plus il y a une **corrélation** entre ses deux variables.

Calculatrice : En pratique on utilise la calculatrice pour déterminer la variance, l'écart-type et la covariance.

Casio : Touche MENU puis menu STAT

- Entrer la série statistique (EXE pour valider) :
  . $\boldsymbol{List\ 1}$: Valeurs de la 1e variable.
  . $\boldsymbol{List\ 2}$: Valeurs de la 2e variable.
- Configuration : F2 (CALC) puis F6 (SET) :
  . $\boldsymbol{2Var\ XList}$: $\boldsymbol{List1}$ F2 (LIST) / 1
  . $\boldsymbol{2Var\ YList}$: $\boldsymbol{List2}$ F2 (LIST) / 2
  . Appuyer sur EXIT puis F2 (2VAR)
- Affichage :
  . $\boldsymbol{\bar{x}}$ et $\boldsymbol{\bar{y}}$: Moyenne de $\boldsymbol{x}$ et de $\boldsymbol{y}$.
  . $\boldsymbol{x\sigma n}$ et $\boldsymbol{y\sigma n}$: Ecart-type de $\boldsymbol{x}$ et de $\boldsymbol{y}$.
  . $\boldsymbol{\Sigma xy}$: Sommes des produits $\boldsymbol{x_iy_i}$

TI : Touche STAT puis touche 1 : $\boldsymbol{1}$: $\boldsymbol{Editer}$.

- Entrer la série statistique (ENTER pour valider) :
  . $\boldsymbol{List\ 1}$: Valeurs de la 1e variable.
  . $\boldsymbol{List\ 2}$: Valeurs de la 2e variable.
- Configuration : Touche STAT, puis touche ▶ ($CALC$)
  . Dans l'onglet $\boldsymbol{CALC}$ sélectionner $\boldsymbol{2}$: $\boldsymbol{Stats\ 2\ Var}$.
  . $\boldsymbol{XListe}$: $\boldsymbol{L_1}$ et $\boldsymbol{YListe}$: $\boldsymbol{L_2}$ puis $\boldsymbol{Calculer}$
- Affichage :
  . $\boldsymbol{\bar{x}}$ et $\boldsymbol{\bar{y}}$: Moyenne de $\boldsymbol{x}$ et de $\boldsymbol{y}$.
  . $\boldsymbol{\sigma x}$ et $\boldsymbol{\sigma y}$: Ecart-type de $\boldsymbol{x}$ et de $\boldsymbol{y}$.
  . $\boldsymbol{\Sigma xy}$: **Sommes des produits** $\boldsymbol{x_iy_i}$

Exemple 1 (Suite) : Déterminer les écarts-types de x et y ainsi que la covariance de la série.

$\sigma(x) \approx 1.62$ et $\sigma(y) \approx 1.55$. $cov(x; y) = \frac{\Sigma xy}{n} - \bar{x}\bar{y} = \frac{241,3}{8} - 5.5 \times 5.93 = -2.425$