Statistiques à deux variables, cours, terminale STMG

F.Gaudon

30 mars 2021

Table des matières

1 Statistiques à deux variables **2**
1.1 Vocabulaire 2
1.2 Ajustement d'un nuage de points 3
1.3 Détermination d'une équation de droite d'ajustement affine 3

1 Statistiques à deux variables

1.1 Vocabulaire

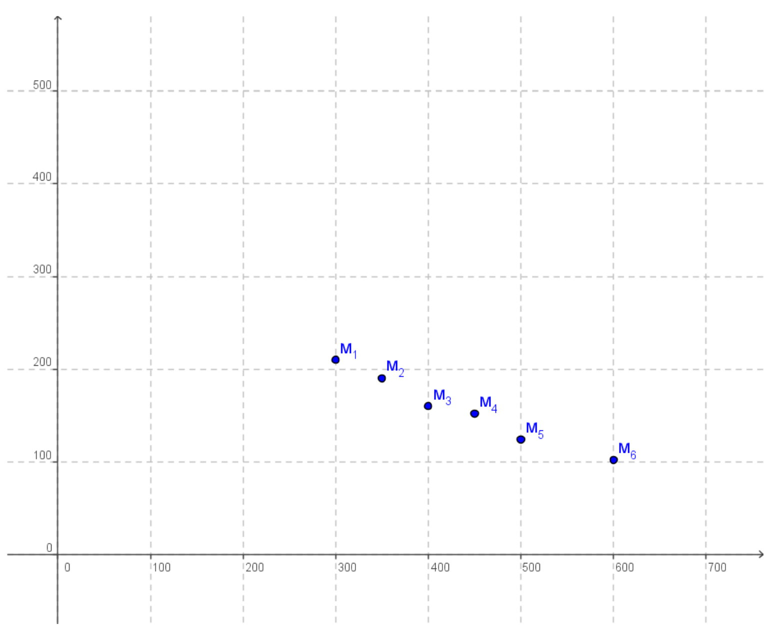
Définition :

- Soient x et y deux caractères quantitatifs d'une même population. A chaque individu de la population on associe un couple $(x_i; y_i)$ où x_i et y_i pour $i \in \{1; \ldots; n\}$ avec n entier naturel sont les valeurs prises respectivement par x et y. L'ensemble de ces couples constitue une *série statistique à deux variables* x et y.
- Dans un repère $(O; \vec{i}; \vec{j})$, l'ensemble des points M_i de coordonnées $(x_i; y_i)$ est appelé *nuage de points* associé à la série statistique.
- On appelle *point moyen* du nuage de points de cette série statistique le point G de coordonnées $(\bar{x}; \bar{y})$ où $\bar{x}$ est la moyenne pondérée de la série des x_i et $\bar{y}$ est la moyenne de la série des y_i, c'est à dire :
$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$
et
$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

Exemple :
Un magasin réalise une étude sur l'influence du prix de vente sur le nombre de machines à laver vendues au cours d'une année. Le tableau suivant donne les résultats de cette étude :

Prix x_i en euros	300	350	400	450	500	600
Nombre de machines vendues	210	190	160	152	124	102

Le nuage de points associé à cette série est constitué des points M_i pour i allant de 1 à 6 dont les coordonnées sont $(300; 210)$, $(350; 190)$,...,$(600; 102)$.
$\bar{x} = \frac{300+350+400+450+500+600}{6} = \frac{2600}{6} \approx 433,33$ et $\bar{y} = \frac{210+190+160+152+124+102}{6} = \frac{938}{6} \approx 15,33$ donc le point moyen est $G(\frac{2600}{6}; \frac{938}{6})$.



1.2 Ajustement d'un nuage de points

Définition :

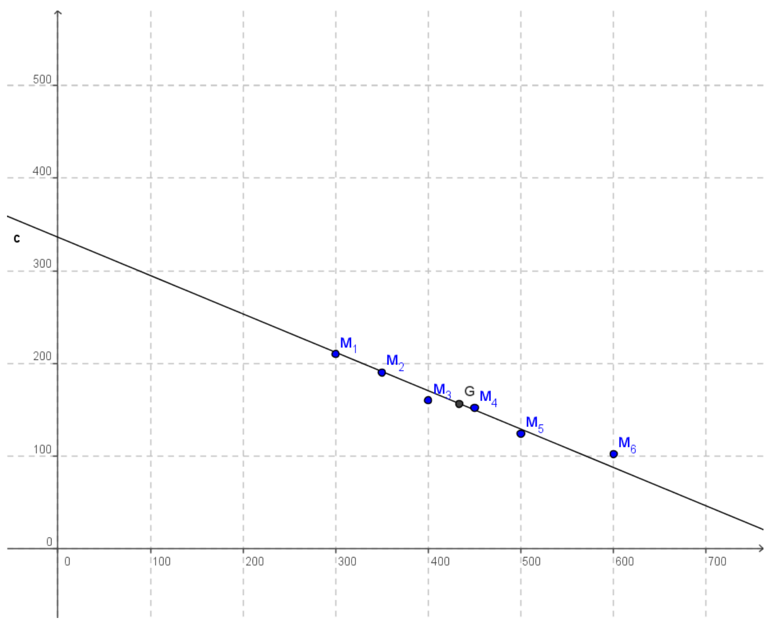
Toute droite "résumant approximativement" le nuage est appelée *droite d'ajustement affine* du nuage de points.

Remarque :
Il existe d'autres types d'ajustement : dans certains cas, on peut observer que visiblement une droite ne convient pas mais que le nuage de points semble être approché par un autre type de courbe, parabole par exemple. En outre, certains nuages peuvent ne pas sembler être approchables par une quelconque courbe auquel cas les deux variables ne sont pas reliées entre elles.

1.3 Détermination d'une équation de droite d'ajustement affine

Méthode graphique au jugé :

On trace « au jugé » une droite qui « semble résumer » le nuage de points. C'est une méthode simple mais qui dépend de la droite tracée.



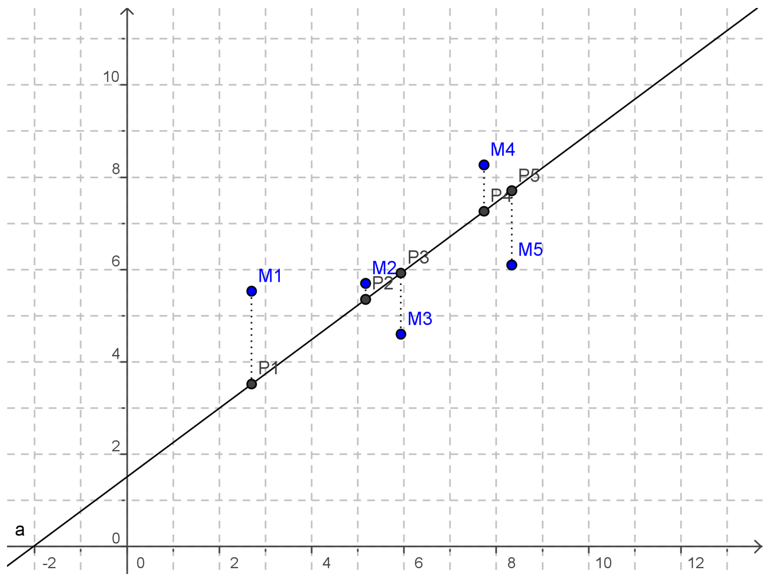
Méthode des moindres carrés :

Avec les notations de la figure ci-dessous, étant donné un nuage de n points M_i, il existe une droite passant par le point moyen G et telle que la somme des carrés des écarts (ou *résidus*) $P_1M_1^2 + P_2M_2^2 + \ldots + P_nM_n^2$ soit minimale. Cette droite est appelée *droite de régression de y en x*. On peut montrer que son équation réduite est $y = mx + p$ avec :

$$m = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \ldots + (x_p - \bar{x})(y_p - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_p - \bar{x})^2}$$

et

$$p = \bar{y} - m\bar{x}$$



Exemple :

On reprend l'exemple précédent.

- Recherche de l'équation réduite à l'aide des formules :

$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
	Total		

  D'où $m \approx -0,37$

  et $p \approx 315,06$

- Recherche de l'équation réduite avec la calculatrice :
  * TI 82 et plus :
    Aller dans le menu $\boxed{\text{STAT}}$ puis $\boxed{\text{EDIT}}$. Entrer les valeurs x_i dans la colonne L_1 et les valeurs y_i dans la colonne L_2. Quitter ($\boxed{\text{2nde}}$ $\boxed{\text{QUIT}}$) puis menu $\boxed{\text{STAT}}$ et $\boxed{\text{CALC}}$. Choisir $\boxed{\text{LinReg(ax+b)}}$ puis $\boxed{\text{2nd}}\boxed{\text{L1}}\boxed{,}\boxed{\text{2nd}}\boxed{\text{L2}}$ pour indiquer les deux colonnes à utiliser. Valider ensuite $\boxed{\text{ENTER}}$.
  * CASIO Graph 25 et plus :
    Aller dans le menu $\boxed{\text{STAT}}$ puis entrer les valeurs x_i dans la colonne 1 et les valeurs y_i dans la colonne 2. Sélectionner ensuite $\boxed{\text{CALC}}$. Choisir $\boxed{\text{SET}}$ et vérifier que la ligne « 2Var XList » est mise à « List1 » et que la ligne « 2Var YList » est mise à « List2 », sinon choisir le menu $\boxed{\text{LIST}}$ pour indiquer les numéros de liste adaptés. Taper ensuite sur $\boxed{\text{EXIT}}$ puis choisir $\boxed{\text{REG}}$ puis $\boxed{\text{X}}$.
- Obtention de l'équation réduite à l'aide d'un programme :

```
def moindreCarres(L1,L2):
        mL1=0
        for k in range(len(L1)):
                mL1=mL1+L1[k]
        mL1=mL1/len(L1)
        mL2=0
```

```
    for k in range(len(L2)):
            mL2=mL2+L2[k]
    mL2=mL2/len(L2)
    numerateur=0
    denominateur=0
    for k in range(len(L1)):
            numerateur=numerateur+(L1[k]-mL1)*(L2[k]-mL2)
            denominateur=denominateur+(L1[k]-mL1)**2
    m=numerateur/denominateur
    p=mL2-a*mL1
    return m,p
```