Cours sur la méthode des moindres carrés

Introduction Ce cours consiste à expliquer comment établir la droite de régression (un ajustement affine)par la méthode des moindres carrés. Pour définir cette méthode, on va faire un petit rappel sur le calcul des dérivées partielles, car par la suite on aura besoin d'effectuer deux types de dérivées par rapport à deux variables différentes.

Rappel sur les sommes en mathématique
$\Sigma_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$.
$\Sigma_{i=1}^{n}(x_i + y_i) = \Sigma_{i=1}^{n} x_i + \Sigma_{i=1}^{n} y_i$.
$\Sigma_{i=1}^{n}(ax_i + b) = \Sigma_{i=1}^{n}(ax_i) + \Sigma_{i=1}^{n} b$
$= (ax_1 + ax_2 + ... + ax_n) + (b + b + ... + b)$
$= a(x_1 + x_2 + ... + x_n) + n \times b$
$= a\Sigma_{i=1}^{n} x_i + nb$

Rappel sur les dérivées partielles
Soit la fonction f(a,b) à deux variables a et b tel que $f(a, b) = (y-(ax+b))^2$, on remarque que cette formule a la forme suivante :
On commence par la dérivée partielle par rapport à "a".
$\frac{\delta f}{\delta a} = 2(y - (ax + b))(-x)$
Car $(U^n)' = nU^{n-1}U'$
On reprend la dérivée tel que le "b" est constant.
$\frac{\delta f}{\delta a} = 2(y - (ax + b))(-x)$
$= -2x(y - (ax + b))$.

Maintenant, on fait la dérivée partielle par rapport à "b".
$\frac{\delta f}{\delta b} = 2(y - (ax + b)) \times (-1)$
$=-2(y - (ax + b))$.

Le principe de la méthode des moindres carrés Soit une population représentée par le couple (x_i, y_i), donc on a deux ensembles $\{\{x_1, x_2, ..., x_n\}; \{y_1, y_2, ..., y_n\}\}$.
La méthode des moindres carrés (pour la régression linéaire) consiste à rechercher une relation entre y et x qui a la forme y = f(x), lorsque cette relation est de type affine donc de la forme y = ax + b donc on parlera de régression linéaire. Dans ce cas, "y" est appelée la variable dépendante ou la variable expliquée et "x" est la variable explicative (on va remarquer qu'il n'existe pas une relation directe entre x et y car on va trouver des erreurs). L'ensemble de couple (x_i, y_i) fait l'objet d'une représentation graphique qui est la suivante :

Les x_i et y_i peuvent être représentées par un nuage de points ssi à pour abscisse x_i et comme ordonnée y_i tel que :

- $D_{y/x}$ est la droite de régression de y en x,

- $D_{y/x}$ a la forme y = ax + b,
- Le point P_i a le même abscisse que M_i.

Si on voit la figure ci-dessous, on remarque qu'on a une forme de dispersion des points autour de la droite $D_{y/x}$.

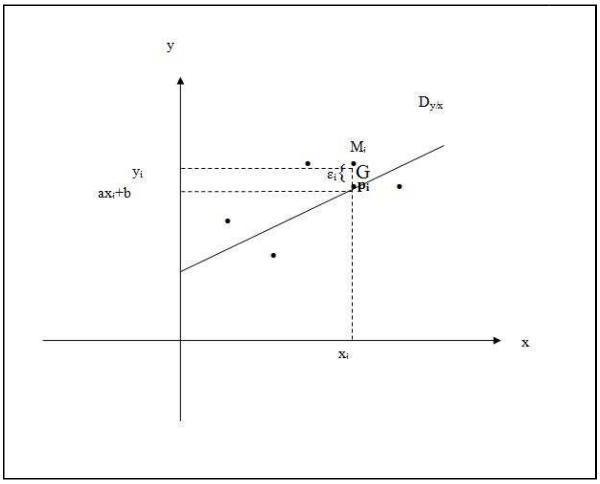

FIGURE 1 – un schéma explicatif

Ce nuage de points est caractérisé par le point moyen $G(\bar{x}, \bar{y})$ tel que $\bar{x}$ c'est la moyenne des x et $\bar{y}$ c'est la moyenne des y sachant que le point G se trouve sur la droite.
Donc l'objectif est de rechercher la droite qui ajuste ce nuage de points, on remarque qu'il y a des écarts appelés résidus qui sont les P_iM_i notés par ϵ_i. Le but de la méthode c'est de déterminer la droite D pour la quelle la somme P_iM_i carrée c'est à dire $\Sigma P_iM_i^2$ soit minimale. Cette somme est la somme de tous les résidus ϵ_i^2.
On fait $\Sigma\epsilon^2$ parce qu'il y a parfois des écarts positifs et parfois des écarts négatifs.
Le P_i vu précédemment a comme ordonnée $ax_i + b$ et le point M_i a comme ordonné y_i et donc le résidu entre P_i et M_i est égale à $\Sigma_{i=1}^n (y_i - (ax_i + b))^2 = \Sigma_{i=1}^n \epsilon_i^2 =$ E (qui représente l'erreur).
Donc l'objectif est de déterminer a et b de telle sorte que cette somme soit minimale.
On cherche l'équation e la droite $D_{y/x} : y = ax + b...(1)$.
Cette droite passera obligatoirement par le point moyen $G(\bar{x}, \bar{y})$, comme le point G appartient à la droite, donc nous avons :
$\bar{y} = a\bar{x} + b.....(2)$
(1)-(2) : $y - \bar{y} = a(x - \bar{x})$
Sachant que $\bar{x} = \frac{\Sigma x_i}{n}$, $\bar{y} = \frac{\Sigma y_i}{n}$ et $\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$ c'est la variance de "x", pour "y" c'est la même chose. On peut aussi mettre $\sigma_x^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2$.
La covariance $\sigma_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$ donc la variance de x est tout simplement la covariance σ_{xx} car :
$\sigma_{xy} = \frac{\Sigma x_i y_i}{n} - \bar{x}\bar{y} \Longrightarrow \sigma_{xx} = \frac{\Sigma x_i^2}{n} - \bar{x}^2$.
Le a= $\frac{\sigma_{xy}}{\sigma_x^2}$
On a E = $\Sigma(y_i - (ax_i + b))^2$, pour minimiser cette somme, il faut rechercher $\frac{\delta E}{\delta b}$ en main-

tenant "a" comme constante.
$\Longrightarrow \frac{\delta E}{\delta b} = 2(-1)\Sigma(y_i - (ax_i + b))$.
Et $\frac{\delta E}{\delta a} = -2\Sigma(y_i - (ax_i + b))x_i$ en maintenant "b" comme constante et comme l'objectif c'est de minimiser E, donc on doit poser $\frac{\delta E}{\delta b} = 0$ et $\frac{\delta E}{\delta a} = 0$.
On va montrer que la droite $D_{y/x}$: $y = ax + b$ passe par G($\bar{x}\bar{y}$) donc on va poser :
$\frac{\delta E}{\delta b} = 0 \Longleftrightarrow -2\Sigma_{i=1}^{n}(y_i - (ax_i + b)) = 0 ondivisepar(-2)$ ce qui donne.
$\Longleftrightarrow \Sigma_{i=1}^{n}(y_i - (ax_i + b)) = 0$
$\Longleftrightarrow \Sigma_{i=1}^{n}y_i - \Sigma_{i=1}^{n}(ax_i + b) = 0$
$\Longleftrightarrow \Sigma_{i=1}^{n}y_i = \Sigma_{i=1}^{n}(ax_i + b)$
$\Longleftrightarrow \Sigma_{i=1}^{n}y_i = a\Sigma_{i=1}^{n}x_i + \Sigma_{i=1}^{n}b$
$\Longleftrightarrow \frac{\Sigma_{i=1}^{n}y_i}{n} = a\frac{\Sigma_{i=1}^{n}x_i}{n} + \frac{nb}{n}$.
On sait que $\frac{\Sigma_{i=1}^{n}y_i}{n} = \bar{y}$ et $\frac{\Sigma_{i=1}^{n}x_i}{n} = \bar{x}$, on obtient $\bar{y} = a\bar{x} + b$, ce résultat traduit que le point G($\bar{x}\bar{y}$) appartient bien à la droite D. G est le centre de gravité du nuage de points.
Si on pose : $\frac{\delta E}{\delta a} = -2\Sigma_{i=1}^{n}(y_i - (ax_i + b))x_i$.
$\frac{\delta E}{\delta a} = 0 \Rightarrow -2\Sigma_{i=1}^{n}(y_i - (ax_i + b))x_i = 0$
$\Rightarrow \Sigma_{i=1}^{n}(x_iy_i - (ax_i^2 + bx_i) = 0$
$\Rightarrow \Sigma_{i=1}^{n}x_iy_i = \Sigma_{i=1}^{n}ax_i^2 + \Sigma_{i=1}^{n}bx_i$
On a b$= \bar{y} - a\bar{x}$
$\Rightarrow \Sigma_{i=1}^{n}x_iy_i = \Sigma_{i=1}^{n}ax_i^2 + \Sigma_{i=1}^{n}(\bar{y} - a\bar{x})x_i$
$\Rightarrow \Sigma_{i=1}^{n}x_iy_i = \Sigma_{i=1}^{n}ax_i^2 + \Sigma_{i=1}^{n}x_i\bar{y} - \Sigma_{i=1}^{n}ax_i\bar{x}$
$\Rightarrow \frac{\Sigma_{i=1}^{n}x_iy_i}{n} = a\frac{\Sigma_{i=1}^{n}x_i^2}{n} + \bar{y}\frac{\Sigma_{i=1}^{n}x_i}{n} - a\bar{x}\frac{\Sigma_{i=1}^{n}x_i}{n}$
$\Rightarrow \frac{\Sigma_{i=1}^{n}x_iy_i}{n} = a(\frac{\Sigma_{i=1}^{n}x_i^2}{n} - \bar{x}^2) + \bar{x}\bar{y}$
$\Rightarrow a(\frac{\Sigma_{i=1}^{n}x_i^2}{n} - \bar{x}^2) = \frac{\Sigma_{i=1}^{n}x_iy_i}{n} - \bar{x}\bar{y}$
On sait que $\frac{\Sigma_{i=1}^{n}x_iy_i}{n} - \bar{x}\bar{y} \Leftrightarrow \sigma_{xy}$ et $\frac{\Sigma_{i=1}^{n}x_i^2}{n} - \bar{x}^2 \Leftrightarrow \sigma_x^2$
$\Rightarrow a = \frac{\sigma_{xy}}{\sigma_x^2}$.

Exemple Soit le couple de données suivant :

x_i	y_i
25	27
28	28
30	28
30	30
35	31
F	F

Sachant que x_i est la variable explicative,
y_i est la variable à expliquer.
La droite de régression à la forme suivante : $y_i = \alpha + \beta x_i + \epsilon$ β (le a) et α (le b) représente les paramètres du modèle, β est appelé la pente et α c'est la constante.
Question : Estimer les paramètres du modèle ou bien estimer les coefficients de régression ou bien trouver l'équation des moindres carrés ou bien trouver l'équation de la droite de régression simple.
Pour estimer α noté $\hat{\alpha}$ et $\hat{\beta}$ pour β on suppose que $\epsilon = 0$.

$\hat{\beta} = \frac{\sigma_{xy}}{\sigma_x^2}$ et $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\bar{x} = \frac{\Sigma_{i=1}^{n} x_i}{n} = (25 + 28 + 30 + 30 + 35)/5 = 148/5 = 29,6$

$\bar{y} = \frac{\Sigma_{i=1}^{n} y_i}{n} = (27 + 28 + 28 + 30 + 31)/5 = 144/5 = 28,8$

$\sigma_x^2 = \frac{\Sigma_{i=1}^{n} (x_i - \bar{x})^2}{n} = 53,2/5 = 10,64$

$\sigma_{xy} = \frac{1}{n}\Sigma_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 21{,}6/5 = 4{,}32$

$\hat{\beta} = \frac{4,32}{10,64} = 0,406$

$\hat{\alpha} = 28,8 - 0,406 \times 29,6 = 16{,}782$

$\hat{y} = 16,782 + 0,406x_i$

Le tableau des valeurs est comme suit :

x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
25	27	-4,6	21,16	-1,8	8,28
28	28	-1,6	2,56	-0,8	1,28
30	28	0,4	0,16	-0,8	0,32
30	30	0,4	0,16	1,2	0,48
35	31	5,4	29,16	2,2	11,88

Deuxième méthode On peut aussi utiliser le tableau suivant :

x_i	y_i	x_i^2	$x_i y_i$
25	27	625	675
28	28	784	784
30	28	900	840
30	30	900	900
35	31	1225	1085
$\bar{x}$ =29,6	$\bar{y}$ =28,8	Σ =4434	Σ =4284

$$\hat{\beta} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\sigma_x^2 = \frac{\Sigma_{i=1}^n x_i^2}{n} - \bar{x}^2$

$\Rightarrow \sigma_x^2 = \frac{4434}{5} - (29,6)^2$

$\Rightarrow \sigma_x^2 = 10,64$

$\sigma_{xy} = \frac{\Sigma_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}$

$\Rightarrow \sigma_{xy} = \frac{4284}{5} - (29,6 \times 28,8)$

$\Rightarrow \sigma_{xy} = 4,32$

$\hat{\beta} = \frac{4,32}{10,64} = 0,406$

$\hat{\alpha} = 28,8 - (0,406 \times 29,6) = 16,782$

L'équation de la droite de régression est donc la suivante :

$$\hat{y} = 16,782 + 0,406x_i$$