Statistiques à deux variables.

On étudie deux caractères d'une population, X et Y.
Voici par exemple une série chronologique représentant le prix au m² des terrains agricoles dans une certaine région. Le premier caractère est la date (année – 1980) appelé rang , X, le deuxième est le prix au m², Y.

Année	1980	1985	1987	1990	1995	1997	2000	Moyenne	Ecart type	Covariance	0,73
Rang xi	0	5	7	10	15	17	20	10,57	7,14	37,34	59,08
Prix au m²	58,8	60,9	62,1	67,5	71,7	73	73,8	66,83	6,23	0,84	= ρ

1) <u>Nuage de points.</u>
Chaque couple (x_i ; y_i) est représenté par un point M_i du plan muni d'un repère $(O;\vec{i};\vec{j})$

2) <u>Point Moyen G</u>.
Le point moyen a pour coordonnées $(\bar{X};\bar{Y})$, moyenne de X et de Y.

3) <u>Approximation par une droite.</u>
Les points sont presque alignés. On trace une droite d'équation y = a x + b qui passe le plus près possible de ces points.
On peut considérer que X et Y sont liés par la relation Y ≈ a X + b et ainsi faire des prévisions.
Le problème est de trouver la droite qui passe le plus près possible. En fait selon les méthodes on trouve différentes droites.

4) <u>Droite de régression linéaire par la méthode des moindres carrés.</u>
C'est la droite d'équation y = a x + b qui minimise les carrés des écarts (et aussi des distances) entre Y_i et (a X_i + b). C'est à dire la quantité $r=\sum_{i=0}^{i=7}(Y_i - a X_i + b)^2$, r est la somme des résidus.
Les nombres a et b sont donnés par la calculatrice. Cette droite passe par le point moyen G.

5) <u>Quelques notions hors programme.</u>
$cov(X;Y)=\frac{1}{8}\sum_{i=0}^{i=7}(X_i-\overline{X})(Y_i-\overline{Y})=\overline{(XY)}-\overline{X}\,\overline{Y}$ donne des indications sur les variations de Y par rapport à X. $cov(X;Y)>0$ indique que X et Y varient dans le même sens, $cov(X;Y)<0$ indique que quand X croît Y décroît. $cov(X;Y)$ proche de 0 indique que les variations de X et Y ne sont pas liées.

Un meilleur indicateur est ρ le coefficient de corrélation linéaire : $\rho=\frac{cov(X;Y)}{\sigma_x \sigma_y}$
Si ρ est proche de 1 ou de -1, X et Y sont liés "linéairement" : Y ≈ a X + b.

Pour la droite de régression linéaire par la méthode des moindres carrés :
$a=\frac{cov(X;Y)}{V(X)}$ et $b=\overline{Y}-a\overline{X}$