Exercice 1: **Croisement Quantitatif × Quantitatif**

$X : \begin{cases} \bar{X} \\ S_X^2 \end{cases}$ et $Y : \begin{cases} \bar{Y} \\ S_Y^2 \end{cases}$

$$\langle X, Y \rangle = \frac{1}{n} \sum_{k=1}^{n} X_k . Y_k .$$

1°) Covariance en fonction de produit scalaire :

on a: $$C_{XY} = \frac{1}{n} \sum_k (X_k - \bar{X})(Y_k - \bar{Y})$$

on note : $\begin{cases} \tilde{X}_k = X_k - \bar{X} \\ \tilde{Y}_k = Y_k - \bar{Y} \end{cases}$

donc $$C_{XY} = \frac{1}{n} \sum \tilde{X}_k . \tilde{Y}_k = \langle \tilde{X}, \tilde{Y} \rangle .$$

$$\boxed{C_{XY} = \langle X - \bar{X}, Y - \bar{Y} \rangle}$$

Produit scalaire en fct de C_{XY} et les moyens:

$$C_{XY} = \frac{1}{n} \sum (X_k - \bar{X})(Y_k - \bar{Y})$$

$$= \frac{1}{n} \sum X_k . Y_k - \frac{1}{n} \sum X_k . \bar{Y} - \frac{1}{n} \sum \bar{X} . Y_k$$
$$+ \frac{1}{n} \sum \bar{X} . \bar{Y}$$

$$= \frac{1}{n} \sum X_k . Y_k - \bar{X} . \bar{Y} - \cancel{\bar{X}\bar{Y}} + \cancel{\bar{X}\bar{Y}}$$

$$= \frac{1}{n} \sum X_k . Y_k - \bar{X}\bar{Y}$$

donc $$C_{XY} = \langle X, Y \rangle - \bar{X}\bar{Y}$$
$$\Leftrightarrow \langle X, Y \rangle = C_{XY} + \bar{X}\bar{Y} .$$

) Norme du vecteur $X - \bar{X}$:

Rappel : $\|X\|^2 = \langle X, X \rangle$
$= \frac{1}{n} \sum X_K . X_K$

$$\|X - \bar{X}\|^2 = \langle X - \bar{X}, X - \bar{X} \rangle$$

$$= \frac{1}{n} \sum_K (X - \bar{X})^2 = S_X^2 .$$

$$\Leftrightarrow \boxed{\|X - \bar{X}\|^2 = S_X^2}$$

$\|X\|^2$ en fonction de variance et de moyenne de X :

$$S_X^2 = \frac{1}{n} \sum X_K^2 - (\bar{X})^2 .$$

$$\Rightarrow S_X^2 = \langle X, X \rangle - (\bar{X})^2 = \|X\|^2 - (\bar{X})^2$$

$$\Leftrightarrow \boxed{\|X\|^2 = S_X^2 + (\bar{X})^2}$$

3°) Moyenne $\bar{X}$ avec produit scalaire :

$$\bar{X} = \frac{1}{n} \sum X_K . = \frac{1}{n} \sum X_K . 1_K = \langle X, 1 \rangle$$

4°) $r_{XY} = \frac{C_{XY}}{S_X . S_Y} = \frac{\langle X - \bar{X}, Y - \bar{Y} \rangle}{\|X - \bar{X}\| . \|Y - \bar{Y}\|}$ par 2°

$$= \cos (X - \bar{X}, Y - \bar{Y}) .$$

Scanned with CamScanner

5) Si $r_{xy} = \cos(x-\bar{x}, y-\bar{y}) = 0$, alors il y a un angle droit entre les variables donc elles sont orthogonales

Si $r_{xy} = \pm 1$, il y a un angle plat ou nul entre les 2 variables, donc elles sont liées.

6) a) $\bar{\hat{y}} = \bar{y}$??

$\boxed{y = ax + b}$

$$\bar{\hat{y}} = \langle \hat{y}, 1 \rangle$$
$$= \langle a x + b, 1 \rangle$$
$$= \hat{a} \langle x, 1 \rangle + \langle \hat{b}, 1 \rangle$$
$$= \hat{a}\bar{x} + \hat{b}$$
$$= \cancel{\hat{a}}\bar{x} + (\bar{y} - \cancel{\hat{a}}\bar{x})$$
$$= \bar{y}$$

$$\langle \hat{b}, 1 \rangle = \frac{1}{n} \sum_k \hat{b}_k \cdot 1_k$$
$$= \frac{1}{n} \sum \hat{b}_k = \frac{1}{\cancel{n}} (n \hat{b})$$
$$= \hat{b}$$

$$y = \hat{a}x + \hat{b}$$
$$\Rightarrow y_i = \hat{a}x_i + \hat{b}$$
$$\Rightarrow \frac{1}{n}\sum y_i = \frac{1}{n}\sum \hat{a}x_i + \frac{1}{n}\sum \hat{b}$$
$$\Rightarrow \bar{y} = \hat{a}.\bar{x} + \hat{b}$$

b) $\bar{e} = 0$?

$$\bar{e} = \langle e, 1 \rangle = \langle y - \hat{y}, 1 \rangle$$
$$= \langle y, 1 \rangle - \langle \hat{y}, 1 \rangle$$
$$= \bar{y} - \bar{\hat{y}}$$
$$= 0$$

Exercice 2 :

$X =$ nb. d'offres d'emploi

$Y =$ demandes d'emploi.

1°) coeff de corrélation

$$r_{xy} = \frac{C_{xy}}{S_x . S_y} = \frac{-639,90}{\sqrt{97,49} \times \sqrt{4329,14}} = -0,98$$

- $|r_{xy}| = |-0.98| = 0,98 \simeq 1$ donc X et Y sont très liés entre eux par une droite affine.

- $r_{xy} < 0$, alors l'offre et la demande varie en sens inverse, plus l'offre augmente et plus la demande diminue.

2°) $y = \hat{a}x + \hat{b}$; $\hat{a} = \frac{C_{xy}}{S_X^2} = \frac{-639,90}{97,49} = -6.56$

$$\hat{b} = \bar{y} - \hat{a}\bar{X} = 1947,15 - (-6.56)(73,35)$$
$$= 2441,73$$

$$\Rightarrow \boxed{y = -6,56X + 2441,73}$$

3°) prévision ? offre, $X = 61$

$$\hat{y} = -6{,}56(61) + 2441{,}73 = 2041{,}34$$

. La demande réele (Y) pour l'offre $X = 61$, c'était $Y = 2034$, donc un peu inferieur de La valeur prédite $\hat{y} = 2041{,}34$

4°) . $S^2_{\hat{y}} = r^2_{xy} \cdot S^2_y = (-0{,}98)^2 \times (4329{,}14)$

$= 4157.706$

. $S^2_e = S^2_y \cdot (1 - r^2_{xy})$

$= (1 - (-0{,}98)^2) \times (4329{,}14)$

$= 171.43$

. $S^2_y = 4329{,}14$ [donner dans l'enoncé, si c'est pas fait, on peut le calculer, avec

$$S^2_y = \frac{1}{n} \sum_{k=1}^{n} (y_k - \bar{y})^2.$$

vérification : $S^2_{\hat{y}} + S^2_e = 4157.706 + 171.43 = 4329{,}14 = S^2_y$

. Le coeff de détermination :

$$R^2 = \frac{S^2_{\hat{y}}}{S^2_y} = \frac{4157.706}{4329{,}14} = 0.96 \quad (\text{ou } r^2_{xy})$$

donc 96% de La variabilité des demandes observées est expliquée par la droite de régression.