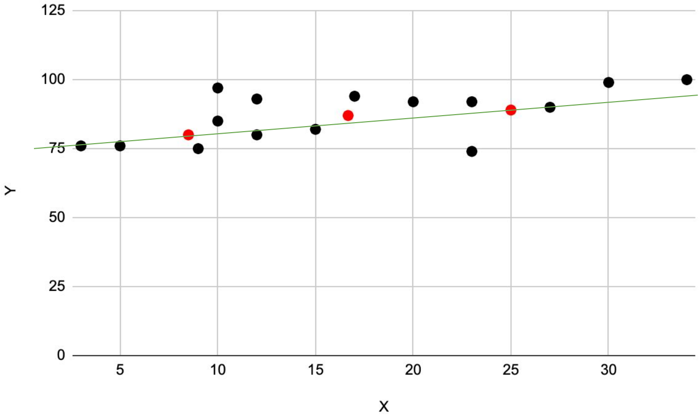
Statistique à deux variables

Covariance, coefficient de corrélation linéaire et droites de régression

Dernière séance :

- Tableau élémentaire
- Nuage statistique
- Barycentre
- Types de relation entre deux variables :
  - Intensité (forte, faible, nulle)
  - Forme (linéaire, non-linéaire)
  - Sens (positive = croissante, négative = décroissante)
- La droite de régression de Mayer

X	Y
12	80
30	99
27	90
9	75
20	92
3	76
12	93
15	82
5	76
10	85
23	74
34	100
23	92
10	97
17	94



X, Y : couple de variables statistiques

La **covariance** est donnée par la formule :

$$Cov(X, Y) = m(XY) - m(X)m(Y)$$

$m(X)$: la moyenne de la variable X
$m(Y)$: la moyenne de la variable Y
$m(XY)$: la moyenne de la variable XY

Exemple :

$m(X) = \frac{12+30+...}{15} \sim 16,67$

$m(Y) = \frac{80+99+...}{15} \sim 87$

$m(XY) = \frac{12 \cdot 80+30 \cdot 99+...}{15} \sim 1497$

donc la covariance est :

$$Cov(X,Y) = m(XY) - m(X)m(Y) \sim 47,67$$

Le **coefficient de corrélation linéaire** est donné par la formule :

$$r(X,Y) = \frac{Cov(X,Y)}{s(X)s(Y)}$$

$Cov(X,Y)$: covariance de X et Y
$s(X)$: écart-type de la variable X
$s(Y)$: écart-type de la variable Y

Propriétés du coefficient de corrélation linéaire :

- $r(X, Y)$ est compris entre -1 et $+1$
- si $r(X, Y)$ est proche de $+1 \implies$ forte relation linéaire positive entre X et Y
- si $r(X, Y)$ est proche de $-1 \implies$ forte relation linéaire négative entre X et Y

$$\text{"proche de } \pm 1 \text{ "} \quad \sim \quad |r(X, Y)| \geqslant 0,75$$

Exemple :

Il faut calculer l'écart-type de la variable X :

$m(X^2) = \frac{12^2+30^2+...}{15} \sim 357,33$

$V(X) = m(X^2) - (m(X))^2 \sim 357,33 - (16,67)^2 \sim 79,56$

$s(X) = \sqrt{V(X)} \sim \sqrt{79,56} \sim 8,92$

dans la même façon on calcule l'écart-type de la variable Y :

$m(Y^2) \sim 7649,3$

$V(Y) \sim 79,33$

$s(Y) \sim 8,91$

finalement on trouve :

$$r(X,Y) = \frac{47,67}{8,92 \cdot 8,91} \sim 0,6$$

X : variable **indépendante**
Y : variable **dépendante**

La **droite de régression** $D_{Y|X}$ est donnée par :

$$Y = a \cdot X + b$$

avec coefficients :

$$a = \frac{Cov(X,Y)}{V(X)}$$

$$b = m(Y) - a \cdot m(X)$$

Exemple :

$a = \frac{Cov(X,Y)}{V(X)} \sim \frac{47{,}667}{79{,}56} \sim 0,599$

$b = m(Y) - a \cdot m(X) \sim 87 - 0,599 \cdot 16,667 \sim 77,016$

donc la droite $D_{Y|X}$ est donnée par

$$Y = 0,599 \cdot X + 77$$

On peut utiliser la droite pour **faire des prévisions** :

si $X = 20$ alors $Y = 0,599 \cdot 20 + 77 \sim 89$

X : variable **dépendante**
Y : variable **indépendante**

La **droite de régression** $D_{X|_Y}$ est donnée par :

$$X = \hat{a} \cdot Y + \hat{b}$$

avec coefficients :

$$\hat{a} = \frac{Cov(X, Y)}{V(Y)}$$

$$\hat{b} = m(X) - \hat{a} \cdot m(Y)$$