

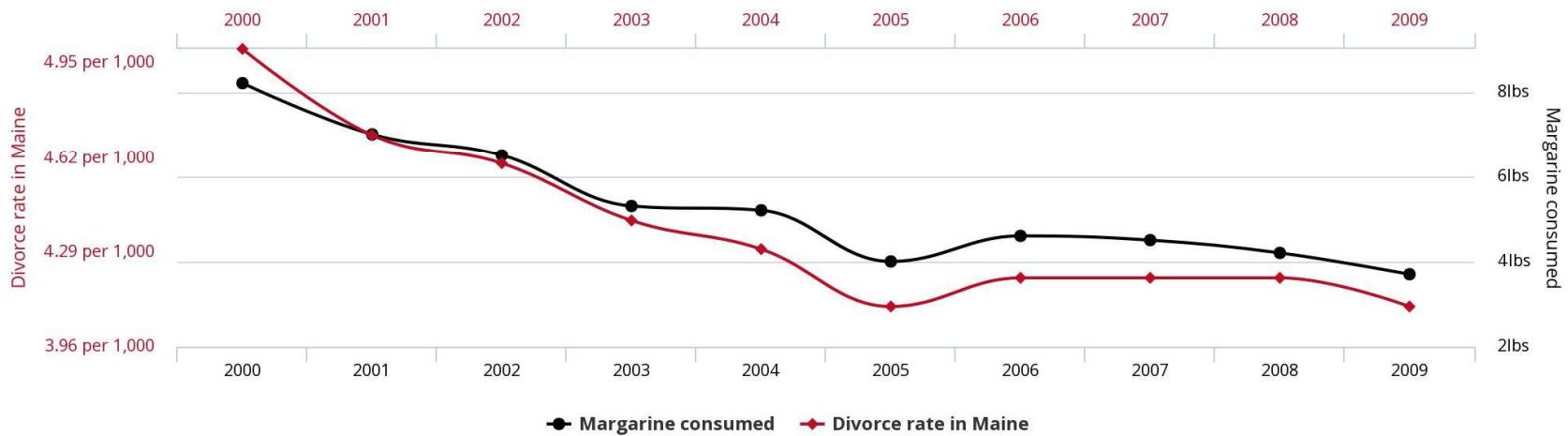
# Plan du cours – partie I

1. Vocabulaire
2. Variables ou caractères
  1. Vocabulaire
  2. Notion de distribution
3. Grandeurs statistiques usuelles
  1. Paramètres de position
  2. Paramètres de dispersion
4. Représentations graphiques
5. Lois de distribution usuelles
6. Statistiques bivariées
  1. Représentation graphique
  2. Covariance
  3. Régression linéaire



y'a t-il un lien entre les variables ?  
est ce qu'elles sont corrélées ?  
est ce qu'il y a un lien de cause à effet  
= même nature

**Divorce rate in Maine**  
correlates with  
**Per capita consumption of margarine**

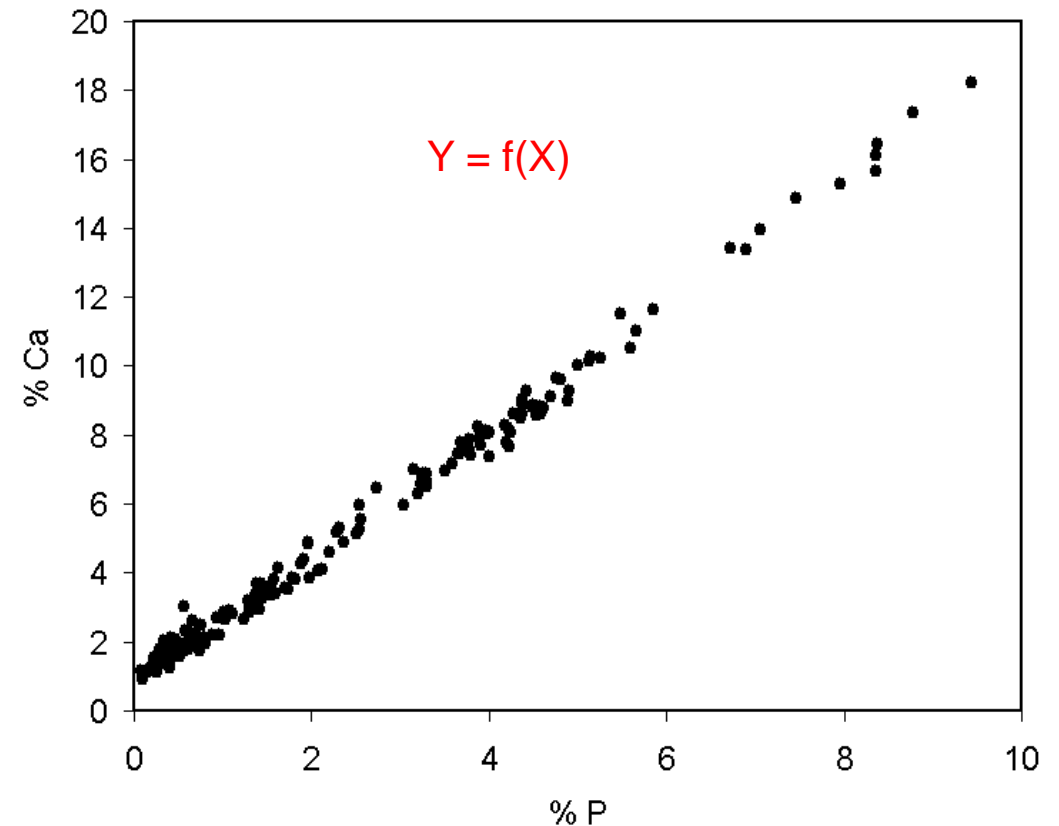


$R^2 > 99 \%$

tylervigen.com

partir d'un tableau de données avec plusieurs variables  
la représentation graphique le plus simple est le nuage de point.  
calcul une covariance pour voir le lien entre les deux variables

échantillon	Variable X	Variable Y
1	$X_1$	$Y_1$
2	$X_2$	$Y_2$
3	$X_3$	$Y_3$
4	$X_4$	$Y_4$
5	$X_5$	$Y_5$
...	...	...
n	$X_n$	$Y_n$



covariance = tendance à varier ensemble pour 2 variables.  
si x et y ne sont pas corrélées  $\rightarrow \text{cov}(x,y)=0$   
(réciproque pas toujours vraie)  
si x et y varient dans le même sens  $\rightarrow \text{cov}(x,y)>0$   
dans le sens opposé  $\rightarrow \text{cov}(x,y)<0$

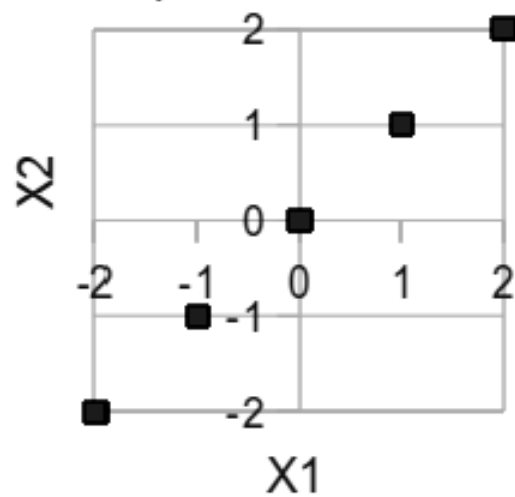
$$\begin{aligned}\text{cov}(x,x) &= \text{var}(x) \\ \text{cov}(x,y) &= \text{cov}(y,x) \\ \text{cov}(ax+b;cy+d) &= ab \text{ cov}(x,y)\end{aligned}$$

$$\text{coefficient de corrélation} = \frac{\text{cov}(x,y)}{\sqrt{\text{ET}(x) \cdot \text{ET}(y)}} \quad 3$$

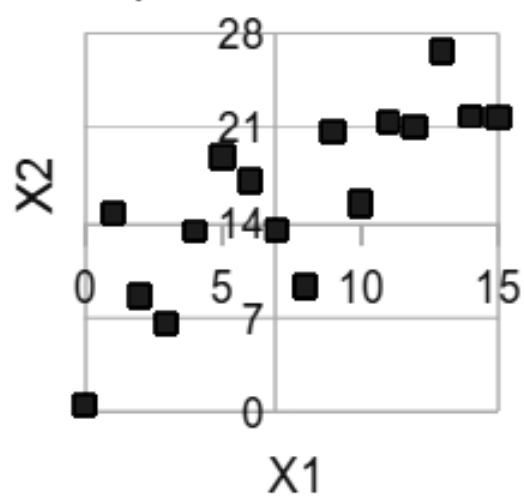
plus x,y tendent vers 1  $\rightarrow$  corrélation  
plus x,y tendent vers -1  $\rightarrow$  anti-corrélation  
0  $\rightarrow$  pas corrélés

$R^2$  coefficient de détermination entre 0 et 1

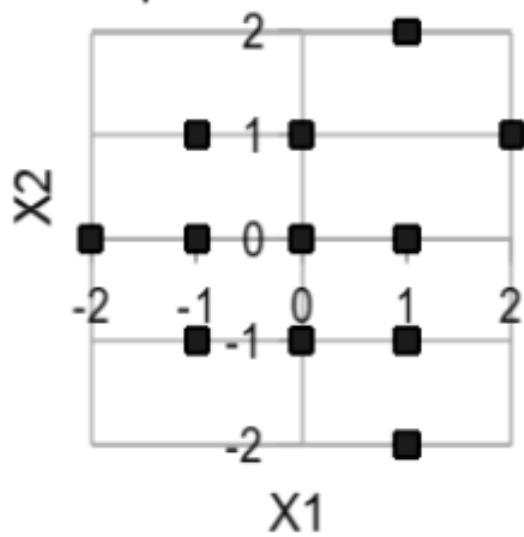
Exemple 1 :  $r=1$



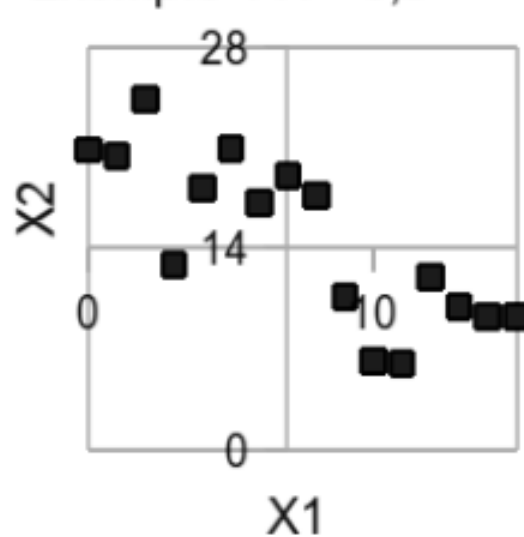
Exemple 2 :  $r=0,8$



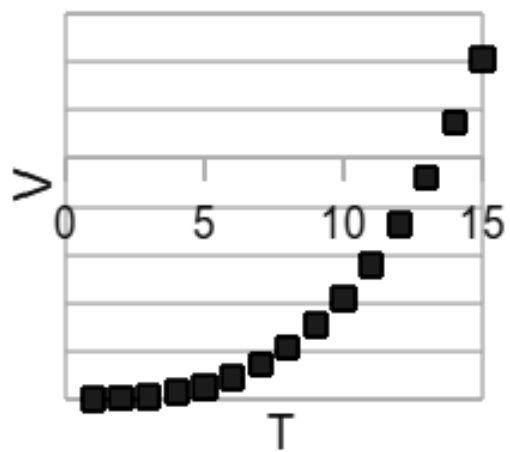
Exemple 3 :  $r=0$



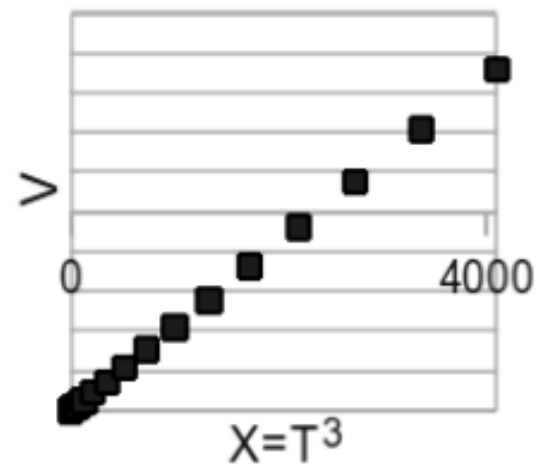
Exemple 4 :  $r=-0,8$



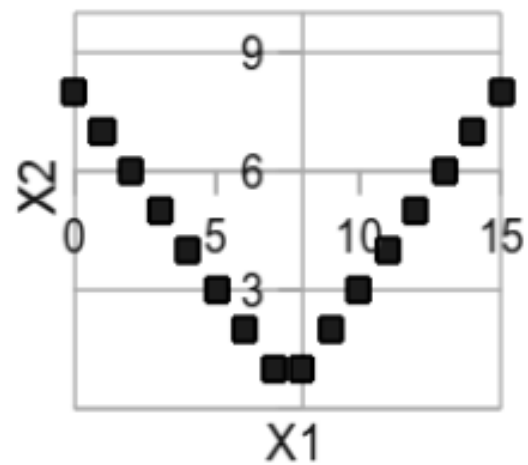
Exemple 7 :  $r=0,92$



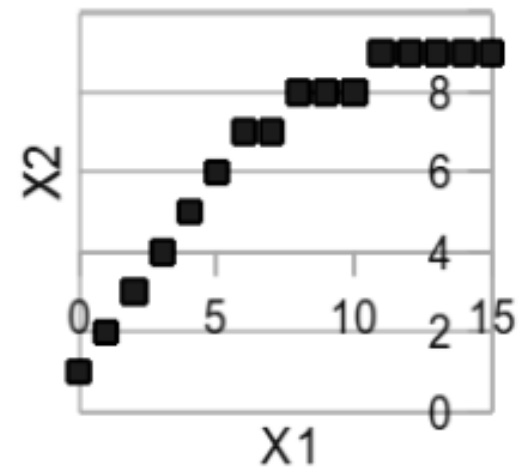
Exemple 8 :  $r=1$



Exemple 9 :  $r=0$



Exemple 10 :  $r=0,94$



## Formules EXCEL pour la covariance

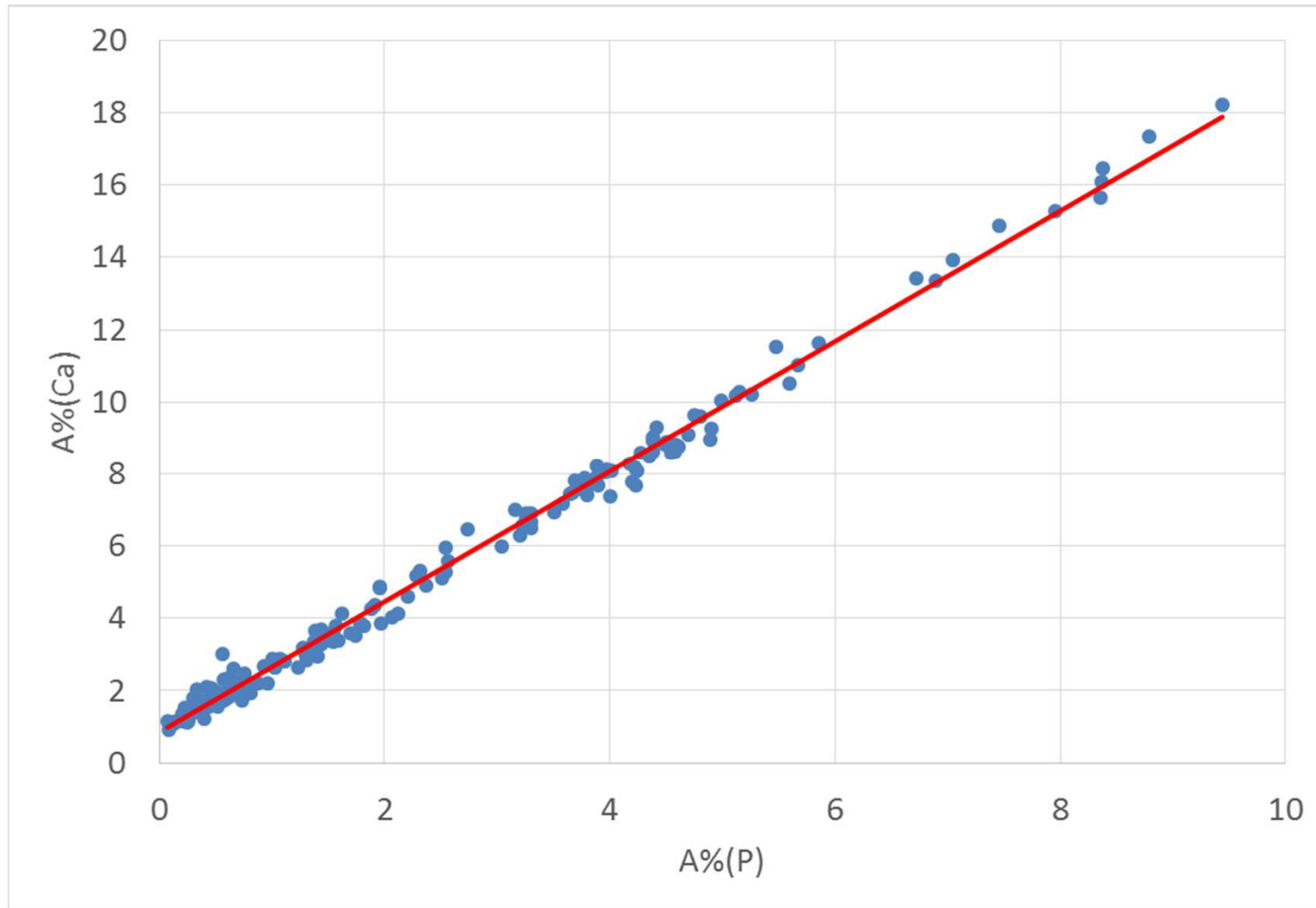
Paramètre	Formule	Arguments
Covariance	= COVARIANCE.STANDARD (matrice Y; matrice X)	Tableau de données
Coefficient de corrélation	= COEFFICIENT.CORRELATION (matrice X; matrice Y)	Tableau de données
Coefficient de détermination	= COEFFICIENT.DETERMINATION (matrice X; matrice Y)	Tableau de données

## Fonctions R pour la covariance

Paramètre	Fonction	Arguments
Covariance	= cov(x,y)	Tableau de données
Coefficient de corrélation	= cor(x,y)	Tableau de données

régression linéaire : consiste à rechercher une fonction permettant d'expliquer le comportement de la variable y en fonction de la variable x  
 $Y=ax+b$  a: pente b:ordonnée à l'origine

La meilleure solution est de calculer par un ajustement aux moindres carrés ordinaires qui minimise la somme quadratique des résidus (epsilon)

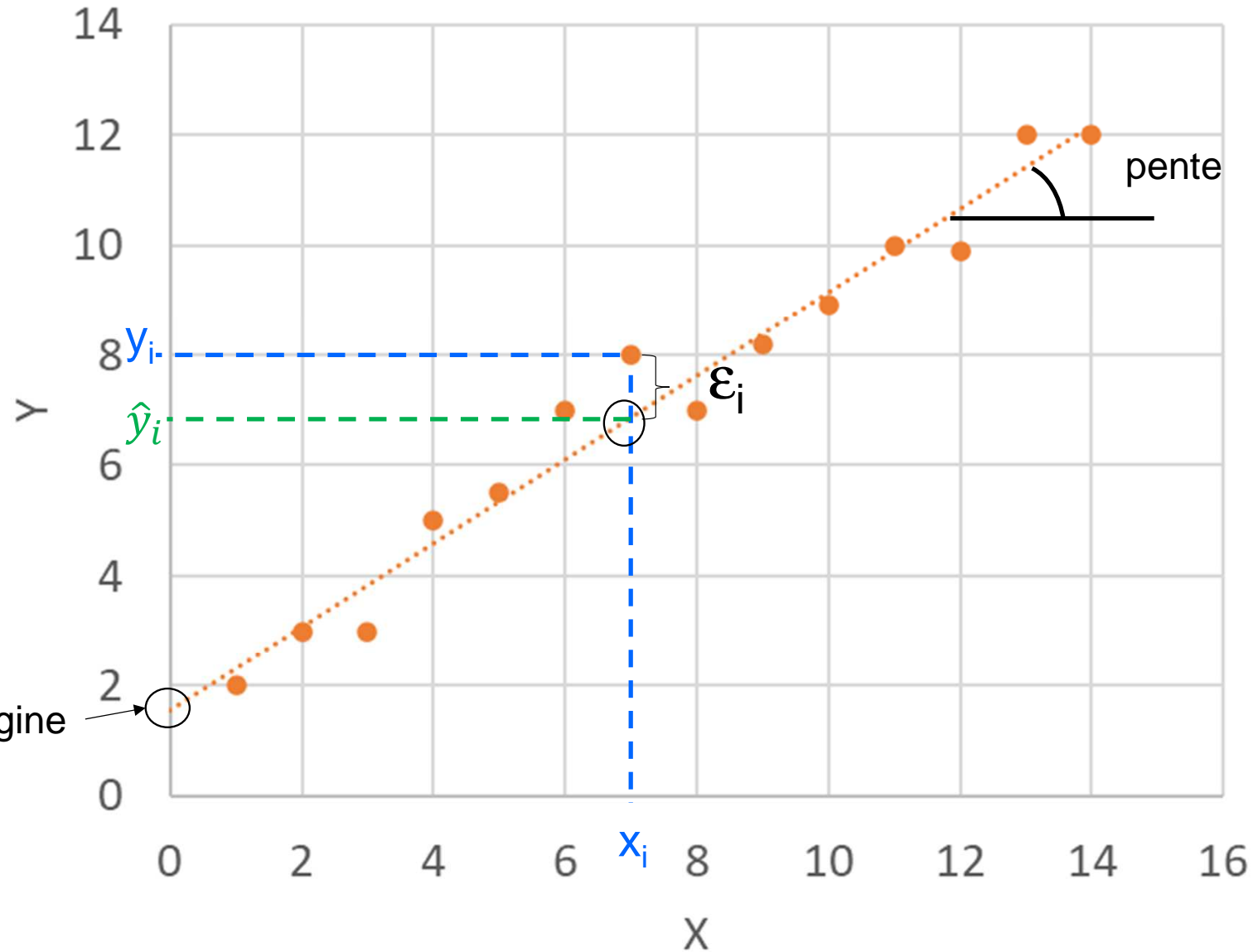


$$a = \text{cov}(x,y)/\text{ET}^2(X)$$

$$b = \text{moyenne de } y - a \times \text{moyenne de } x$$

interprétation de la régression  
 1. coefficient de détermination  $R^2$   
 2. analyse de résidus  
 3. incertitude sur les coefficients -> intervalles de confiance

coefficient de détermination = part de la variation de Y expliquée par la régression (=variation de x)

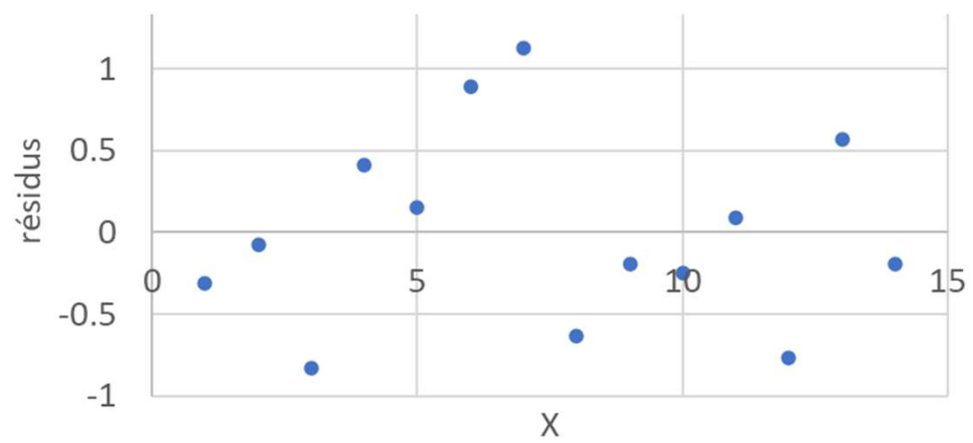
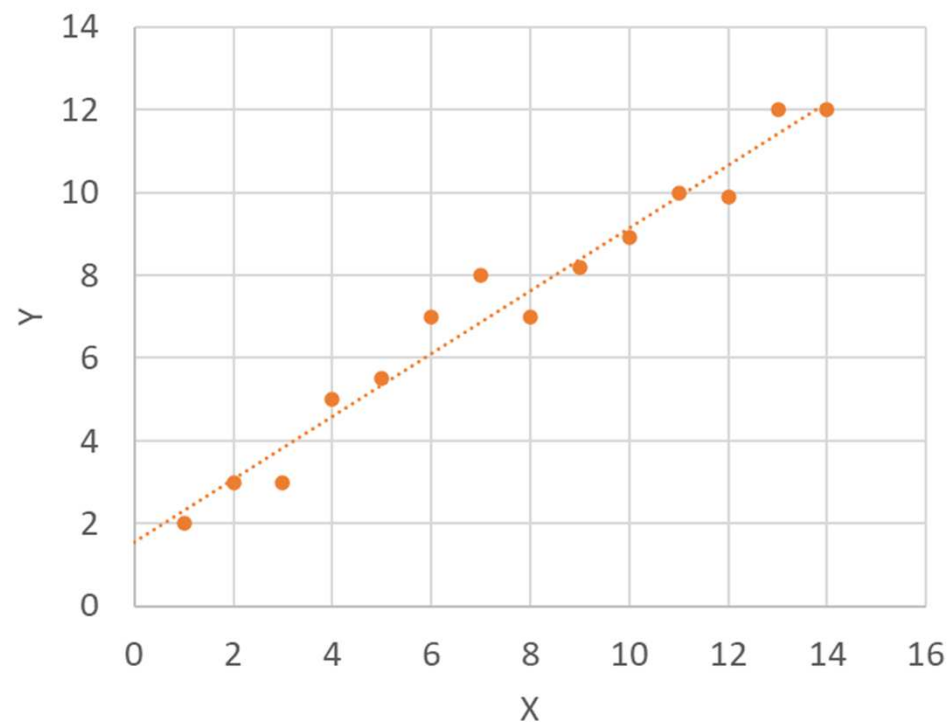


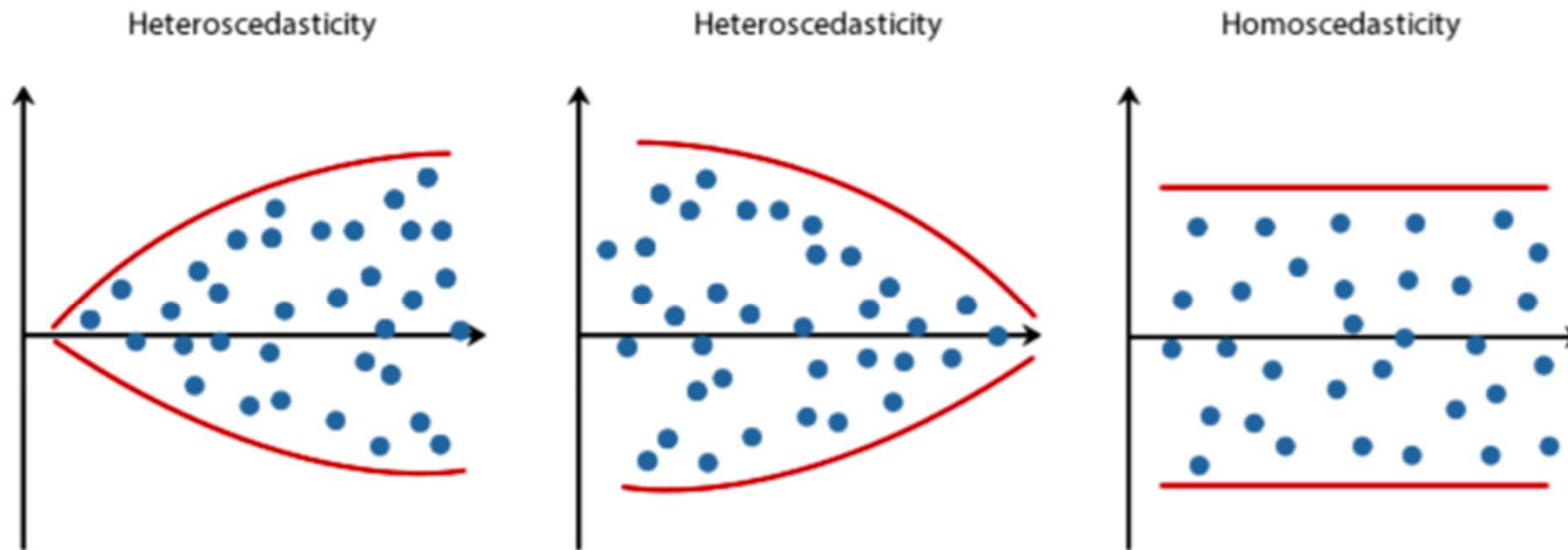
somme de carrés totaux = variance (Y) à la division près  
 somme des carrés résiduels : variance (E)  
 somme des carrés expliqués = variance ( $\hat{y}$ )

$$SCT = SCE + SCR$$

$$R^2 = SCE/SCT = 1 - (SCR/SCT) (\%)$$



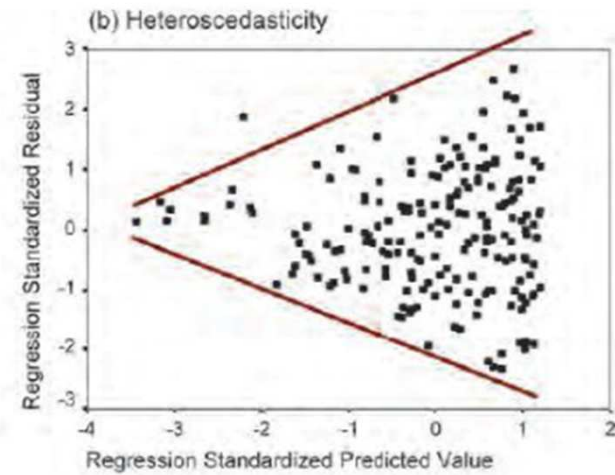
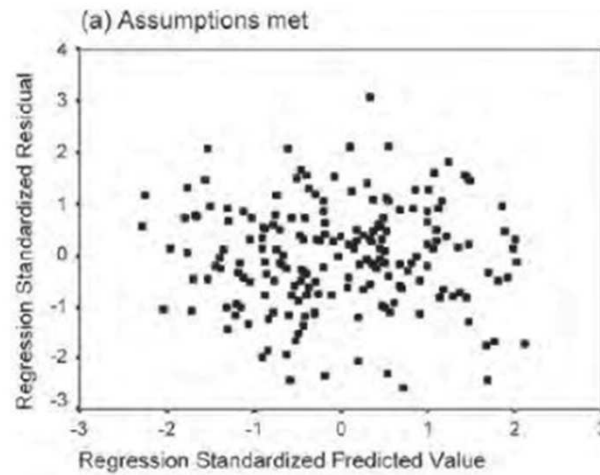




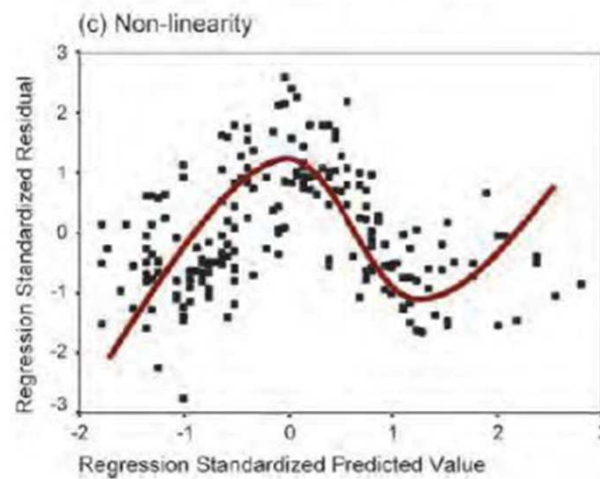
Copyright 2014. Laerd Statistics.

homoscédasticité : individus dépendant de  $x \rightarrow \text{elipsum} = 0$   
 variance de  $\text{epsilum} = \text{constante}$   
 $\text{cov}(x, \text{elipsum}) = 0$

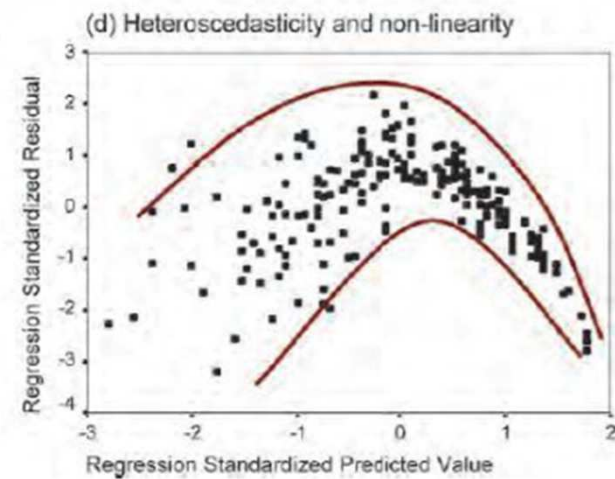
OK



Hétéroscédasticité



Modèle inadapté



Hétéroscédasticité  
+ modèle inadapté

## Formules EXCEL pour la régression linéaire

Paramètre	Formule	Arguments
Régression	= DROITEREG (matrice Y; matrice X)	Tableau de données Formule matricielle
Pente	= PENTE (matrice Y; matrice X)	Tableau de données
Ordonnée à l'origine	=ORDONNEE.ORIGINE (matrice Y; matrice X)	Tableau de données
Coefficient de détermination	= COEFFICIENT.DETERMINATION (matrice $\hat{Y}$ ; matrice Y)	Tableau de données

- Outil graphique (clic droit)
- Utilitaire d'analyse

## 6. Statistiques bivariées

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SCR = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SCT = SCE + SCR$$

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$