# Chapitre 3. Les distributions à deux variables

Jean-François Coeurjolly

http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/

Laboratoire Jean Kuntzmann (LJK), Grenoble University

Notes





Notes

# Tableau de contingence

- = tableau statistique permettant de présenter séries statistiques et de
- exemple : dans une entreprise de 200 salariés, on étudie les variables $X$=âge et $Y$=salaires.

| X=Age \ Y=Salaire | $[800, 1000[$ ($j = 1$) | $[1000, 1200[$ ($j = 2$) | Total |
|---|---|---|---|
| $[20, 22[$ ($i = 1$) | 14 | 6 | 20 |
| $[22, 24[$ ($i = 2$) | 28 | 46 | 74 |
| $[24, 26[$ ($i = 3$) | 20 | 86 | 106 |
| Total | 62 | 138 | 200 |

- $X$ et $Y$ sont des variables continues (regroupées en classes)
- On note $I$ le nombre de modalités de $X$ (ici ) et $J$ le nombre de modalités de $Y$ (ici ).

Notes



# Tableau de contingence (2)

| X=Age \ Y=Salaire | $[800, 1000[$ ($j = 1$) | $[1000, 1200[$ ($j = 2$) | Total |
|---|---|---|---|
| $[20, 22[$ ($i = 1$) | 14 | **6** | 20 |
| $[22, 24[$ ($i = 2$) | 28 | 46 | **74** |
| $[24, 26[$ ($i = 3$) | 20 | 86 | 106 |
| Total | **62** | 138 | 200 |

- $i$ désigne l'indice d'une et $j$ désigne l'indice d'une .
- désigne l' .
  Exemple : $n_{12} = 6$ salariés sont âgés entre 20 et 22 ans **et** ont un salaire compris entre 1000 et 1200 €.
- on note **l'** **de** $X$ (eff. total en lignes) et **l'** **de** $Y$ (effectif total en colonnes).
  Exemple : $n_{2\bullet} = 74$ salariés sont âgés entre 22 et 24 ans ;
  $n_{\bullet 1} = 62$ salariés ont un salaire ente 800 et 1000€.
- correspond à l'effectif total.

Notes

# Tableau de contingence (3)

| X=Age \ Y=Salaire | [800, 1000[ ($j=1$) | [1000, 1200[ ($j=2$) | Total |
|---|---|---|---|
| [20, 22[ ($i=1$) | 14 | **6** | 20 |
| [22, 24[ ($i=2$) | 28 | 46 | **74** |
| [24, 26[ ($i=3$) | 20 | 86 | 106 |
| Total | **62** | 138 | 200 |

Formules : Pour $i = 1, \ldots, I$ et pour $j = 1, \ldots, J$

$n_{i\bullet} =$

$n_{i\bullet} =$

$$n = n_{\bullet\bullet} = \sum_{i=1}^{I} \quad = \sum_{j=1}^{J} \quad = \sum_{i=1}^{I} \sum_{j=1}^{J} \quad .$$

Notes



# Fréquences partielles et marginales

⚠ Les fréquences sont notées entre parenthèses.

| X=Age \ Y=Salaire | [800, 1000[ ($j=1$) | [1000, 1200[ ($j=2$) | Total |
|---|---|---|---|
| [20, 22[ ($i=1$) | 14 ( %) | 6 ( %) | 20 ( %) |
| [22, 24[ ($i=2$) | 28 ( %) | 46 ( %) | 74 ( %) |
| [24, 26[ ($i=3$) | 20 ( %) | 86 ( %) | 106 ( %) |
| Total | 62 ( %) | 138 ( %) | 200 **(100%)** |

- $\square$ désigne la **fréquence** .
  Exemple : $f_{12} = 3\%$ des salariés sont âgés entre 20 et 22 ans **et** ont un salaire compris entre 1000 et 1200 €.
- on note $\square$ la **fréquence** **de** $X$ (fréq. totale en lignes) et $\square$ la **fréquence** **de** $Y$ (fréq. totale en colonnes).
  Exemple : $f_{2\bullet} = 37\%$ des salariés sont âgés entre 22 et 24 ans ; $f_{\bullet 1} = 31\%$ des individus ont un salaire ente 800 et 1000€.

Notes

## Fréquences partielles et marginales (2)

Formules : Pour $i = 1, \ldots, I$ et pour $j = 1, \ldots, J$

$$f_{ij} = \qquad \left(\text{ex : } 3\% = \frac{6}{200}\right)$$

$$f_{i\bullet} = \quad = \sum_{j=1}^{J} \frac{n_{ij}}{n} = \sum_{j=1}^{J} f_{ij} \qquad \left(\text{ex : } 37\% = \frac{74}{200} = 14\% + 23\%\right)$$

$$f_{\bullet j} = \quad = \sum_{i=1}^{I} \frac{n_{ij}}{n} = \sum_{i=1}^{I} f_{ij} \qquad \left(\text{ex : } 31\% = \frac{62}{200} = 7\% + 14\% + 10\%\right)$$

Notes



## Fréquences partielles et marginales (3)

- La distribution marginale de $X$ est représentée par la colonne "total" (fréquences bleues).
- La distribution marginale de $Y$ est représentée par la ligne "total" (fréquences vertes).
- Ce sont bien des **distributions** car lorsque l'on somme les $f_{i\bullet}$ ou les $f_{\bullet j}$, on obtient **100%**.
- $\Rightarrow$ puisqu'on a une distribution, on peut calculer tous les indicateurs du

Notes

## Généralités

- Une distribution conditionnelle est une distribution statistique obtenue en la population à un (une classe par exemple).
- $J = 2 \Rightarrow$ il y a conditionnelles de $X$ par rapport à $Y$.
  1. la distribution de $X$ sachant $Y \in [800, 1000[$.
  2. la distribution de $X$ sachant $Y \in [1000, 1200[$.
- $I = 3 \Rightarrow$ il y a distributions conditionnelles de $Y$ par rapport à $X$
  1. la distribution de $Y$ sachant $X \in [20, 22[$.
  2. la distribution de $Y$ sachant $X \in [22, 24[$.
  3. la distribution de $Y$ sachant $X \in [24, 26[$.

Notes



## Fréquences conditionnelles de $X$ sachant $Y$

| X=Age \ Y=Salaire | $[800, 1000[$ ($j = 1$) | $[1000, 1200[$ ($j = 2$) | Total |
|---|---|---|---|
| $[20, 22[$ ($i = 1$) | 14 ( %) | 6 ( %) | 20 |
| $[22, 24[$ ($i = 2$) | 28 ( %) | 46 ( %) | 74 |
| $[24, 26[$ ($i = 3$) | 20 ( %) | 86 ( %) | 106 |
| Total | 62 (100%) | 138 (100%) | 200 |

- On calcule les fréquences des âges en se restreignant à la sous-population des individus ayant un salaire entre **800 et 1000 €** , puis à la sous-population des individus ayant un salaire entre **1000 et 1200 €** .
- Les fréquences conditionnelles sont en général notées $\boxed{\phantom{00}}$
- Interprétation :
  - **22.6%** des employés ayant un salaire entre 800 et 1000 €sont âgés entre 20 et 22 ans.
  - Parmi les employés ayant un salaire entre 1000 et 1200 €, **62.4%** d'entre eux sont âgés entre 24 et 26 ans.

Notes

## Fréquences conditionnelles de $X$ sachant $Y$ (2)

Notes

Formules : Pour $i = 1, \ldots, I$ et pour $j = 1, \ldots, J$

$$f_{i|j} = \frac{n_{ij}}{n_{\bullet j}} \qquad \left(ex : 22.6\% = \frac{14}{62}\right)$$



## Fréquences conditionnelles de $Y$ sachant $X$

Notes

| X=Age \ Y=Salaire | [800, 1000[ ($j = 1$) | [1000, 1200[ ($j = 2$) | Total |
|---|---|---|---|
| [20, 22[ ($i = 1$) | 14 ( %) | 6 ( %) | 20 **100%** |
| [22, 24[ ($i = 2$) | 28 %) | 46 ( %) | 74 **100%** |
| [24, 26[ ($i = 3$) | 20 ( %) | 86 ( %) | 106 **100%** |
| Total | 62 | 138 | 200 |

- Ces fréquences conditionnelles sont en général notées $\boxed{\phantom{00}}$
- Interprétation :
  - **70%** des employés âgés entre 20 et 22 ans ont un salaire compris entre 800 et 1000 €.
  - Parmi les employés âgés entre 22 et 24 ans, **62.2%** d'entre eux ont un salaire compris entre 1000 et 1200 €.

# Fréquences conditionnelles de $Y$ sachant $X$ et quelques formules

Formules : Pour $i = 1, \ldots, I$ et pour $j = 1, \ldots, J$

$$f_{j|i} = \qquad \left(ex : 30\% = \frac{6}{20}\right)$$

En utilisant les précédentes définitions des fréquences conditionnelles, on peut obtenir

$$f_{ij} = f_{i|j} \times f_{\bullet j}$$

De la même façon on peut obtenir

$$f_{ij} = f_{j|i} \times f_{i\bullet}$$

Notes



| X=Age \ Y=Salaire | $[800, 1000[$ ($j = 1$) | $[1000, 1200[$ ($j = 2$) | **Total** |
|---|---|---|---|
| $[20, 22[$ ($i = 1$) | 14 | 6 | 20 |
| $[22, 24[$ ($i = 2$) | 28 | 46 | 74 |
| $[24, 26[$ ($i = 3$) | 20 | 86 | 106 |
| Total | 62 | 138 | 200 |

Concentrons-nous sur la variable $X$ : on notera $\overline{x}_1$ (ou $\overline{x}_{|Y\in[800,1000[}$) et $\overline{x}_2$ (ou $\overline{x}_{|Y\in[1000,1200[}$) les deux moy. cond. de $X$ sachant $Y$ :

- La moyenne de $X$= la moyenne des moyennes conditionnelles

$$\overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{J} n_{\bullet j} \overline{x}_j.$$

- Vérification :
  - En utilisant la distribution marginale : $\overline{x} \simeq$ **ans** .
  - En utilisant les fréq. conditionnelles, $\overline{x}_1 \simeq$ **ans** et $\overline{x}_2 \simeq$ **ans** .
  - En combinant ans.

Notes

## Décomposition de la variance

- Notons $Var_j(X)$ les variances conditionnelles de $X$ sachant $Y$. Rappelons la formule de décomposition de la variance (qui peut s'exprimer en fonction des variances conditionnelles) :

$$Var(X) = \underbrace{\frac{1}{n}\sum_{j=1}^{J} n_{\bullet j} Var_j(X)}_{} + \underbrace{\frac{1}{n}\sum_{j=1}^{J} n_{\bullet j}(\overline{x}_j - \overline{x})^2}_{}$$

- La vérification sur l'exemple considéré est laissée en exercice.
- Des résultats tout à fait similaires sont bien évidemment valables pour la variable $Y$ (⚠notez que ceci est possible car $Y$ est quantitative).

Notes



## Généralités

Il y a deux extrêmes du niveau de liaison entre deux variables (quelles que soient la ou les natures des variables) :

- l' (ou liaison nulle).
- la (ou liaison fonctionnelle).

Le but de cette section est de mesurer la dépendance, et de quantifier en particulier le niveau de proximité par rapport aux deux cas précédents.

Notes

**Définition**

1. La variable $Y$ est totalement indépendante de la variable $X$ si les variations de $X$ n'entraînent pas de variations de $Y$.
2. La variable $X$ est totalement indépendante de la variable $Y$ si les variations de $Y$ n'entraînent pas de variations de $X$.

**Théorème**

1. $Y$ est totalement indépendante de $X$ si et seulement si

   (c-a-d les fréquences conditionnelles ne dépendent pas des lignes du tableau de contingence et sont égales aux fréquences marginales).
2. $X$ est totalement indépendante de $Y$ si et seulement si
3. L'indépendance est .

Notes



## Indépendance et tableau de contingence

**Théorème**

Les variables $X$ et $Y$ sont indépendantes si et seulement si

**Corollaire**

Un tableau de contingence est associé à deux variables $X$ et $Y$ indépendantes si et seulement si les sont entre elles.

Exemple : tableau associé à deux var. indépendantes

| $X \mid Y$ | $y_1$ | $y_2$ | $y_3$ | Total |
|---|---|---|---|---|
| $x_1$ | 2 | 4 | 12 | 18 |
| $x_2$ | 4 | 8 | 24 | 36 |
| Total | 6 | 12 | 36 | 54 |

On peut par exemple vérifier que

$$\frac{n_{2\bullet} \times n_{\bullet 3}}{n} = \frac{36 \times 36}{54} = 24 = n_{23}.$$

Notes

# Dépendance totale

**Définition**

1. $Y$ est de $X$ (ou fonctionnellement liée à $X$) si à chaque valeur $x_i$ de $X$ correspond une unique valeur $y_j$ de $Y$, autrement dit si chaque ligne du tableau de contingence ne contient qu'un seul effectif $n_{ij}$ non nul.
2. $X$ est de $Y$ (ou fonctionnellement liée à $Y$) si à chaque valeur $y_j$ de $Y$ correspond une unique valeur $x_i$ de $X$, autrement dit si chaque colonne du tableau de contingence ne contient qu'un seul effectif $n_{ij}$ non nul.
3. ⚠La dépendance totale n'est pas une .

Notes



# Application à la notion de dépendance

Exemple 1 :

| $X \mid Y$ | $y_1$ | $y_2$ |
|---|---|---|
| $x_1$ | 2 | 0 |
| $x_2$ | 1 | 0 |
| $x_3$ | 0 | 1 |

$\Rightarrow$ est de et la réciproque est .

Exemple 2 :

| $X \mid Y$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $x_1$ | 2 | 0 | 0 |
| $x_2$ | 0 | 1 | 4 |

$\Rightarrow$ est de et la réciproque est .

Exemple 3 :

| $X \mid Y$ | $y_1$ | $y_2$ |
|---|---|---|
| $x_1$ | 2 | 0 |
| $x_2$ | 0 | 1 |

$\Rightarrow$ est de et la réciproque est .

Notes

# $\chi^2$ et Coefficient de Cramer

**Définition**

Le $\chi^2$ est un nombre mesurant l'écart entre la situation observée et la situation si les variables avaient été théoriquement .

Méthodologie :

1. construction du tableau de contingence sous hypothèse d'indépendance, c-a-d calcul des

2. on calcule ensuite

$$\chi^2 =$$

Notes



# $\chi^2$ et Coefficient de Cramer (2)

**Théorème**

La quantité $\chi^2_{\max}$ est la valeur du $\chi^2$ si la dépendance entre $X$ et $Y$ était totale et réciproque.

**Définition**

Le coefficient de Cramer $C \in [0,1]$ est défini par

- Si $C$ est proche de alors les variables $X$ et $Y$ sont presque .
- Si $C$ est proche de , alors les variables $X$ et $Y$ sont fortement (pas nécessairement liées fonctionnellement)
- Le $C$ de Cramer peut être calculé pour n'importe quel type de variables $X$ et $Y$.

Notes

# $\chi^2$ et Coefficient de Cramer (3)

| X=Age \ Y=Salaire | $[800, 1000[$ $(j=1)$ | $[1000, 1200[$ $(j=2)$ | Total |
|---|---|---|---|
| $[20, 22[$ $(i=1)$ | 14 ( ) | 6 ( ) | **20** |
| $[22, 24[$ $(i=2)$ | 28 ( ) | 46 ( ) | **74** |
| $[24, 26[$ $(i=3)$ | 20 ( ) | 86 ( ) | **106** |
| Total | **62** | **138** | 200 |

1. calcul des effectifs théoriques $n'_{ij}$ .
   Exemple : $n'_{32} = \frac{n_{3\bullet} \times n_{\bullet 2}}{n} = \frac{138 \times 106}{200} \simeq 73.14$.
2. Calcul du $\chi^2$

$$\chi^2 = \frac{(14 - \quad)^2}{\quad} + \frac{(6 - \quad)^2}{\quad} + \ldots + \frac{(86 - \quad)^2}{\quad} \simeq \quad .$$

3. $\chi^2_{max} = 200 \times$ .
4. $C = \sqrt{\quad} \simeq$ % (dépendance modérée).

Notes



# $\chi^2$ et Coefficient de Cramer (4)

**Question**

Quels sont les couples $(x_i, y_j)$ qui contribuent le plus au $\chi^2$ ?

Réponse : il suffit de calculer pour chaque case le rapport

| X=Age \ Y=Salaire | $[800, 1000[$ $(j=1)$ | $[1000, 1200[$ $(j=2)$ | Total |
|---|---|---|---|
| $[20, 22[$ $(i=1)$ | 14 **(42.4%)** | 6 **(19.1%)** | 20 |
| $[22, 24[$ $(i=2)$ | 28 **(4.8%)** | 46 **(2.2%)** | 74 |
| $[24, 26[$ $(i=3)$ | 20 **(21.8%)** | 86 **(9.8%)** | 106 |
| Total | 62 | 138 | 200 |

- Exemple 1ère case : $((6.2 - 14)^2/6.2)/23.13 \simeq 42.4\%$.
- La case des individus s'écarte le plus de l'hypothèse d'indépendance.

Notes

## Généralités

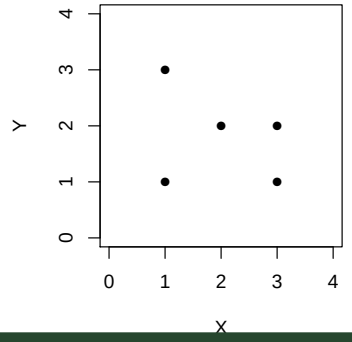
- pour savoir si $X$ et $Y$ sont liées fonctionnellement, on trace le nuage de points $(x_i, y_i)$.
- $\Rightarrow$ section valable uniquement pour $X$ et $Y$
- $\Rightarrow$ il faut disposer des données brutes, autrement dit chaque couple $(x_i, y_i)$ est observée une et une seule fois. Autrement dit, la table de contingence correspondante ne contient que des
- On trace alors le nuage de points $(x_i, y_j)$ et on essaie d'estimer la fonction de lien éventuelle.

Notes



## Exemple et définition

| $X \mid Y$ | 1 | 2 | 3 | Total |
|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 2 |
| 2 | 0 | 1 | 0 | 1 |
| 3 | 1 | 1 | 0 | 2 |
| Total | 2 | 2 | 1 | 5 |




**Définition**

1. est obtenue en faisant correspondre à chaque valeur de $x_i$ de $X$ la moy. conditionnelle de $Y$ sachant $X = x_i$. Cette courbe est notée .
2. est obtenue en faisant correspondre à chaque valeur de $y_j$ de $Y$ la moy. conditionnelle de $X$ sachant $Y = y_j$. Cette courbe est notée

Notes

# Propriétés

**Théorème**

- Si $X$ et $Y$ sont deux variables indépendantes alors $C_{Y/X}$ est parallèle à l'axe des abscisses et la courbe $C_{X/Y}$ est parallèle à l'axe des ordonnées (⚠réciproque fausse).
- Si aucun point ne s'écarte de , $Y$ totalement dépendante de $X$ ( ).
- Si aucun point ne s'écarte de , $X$ totalement dépendante de $Y$ ( ).

Notes



Concept basé sur la formule de décomposition de la variance

**Définition**

1. Le rapport de corrélation de $Y$ en $X$ est défini par

$$\eta^2_{Y/X} = \qquad = \frac{\frac{1}{n}\sum_i n_{i\bullet}(\overline{Y}_i - \overline{Y})^2}{Var(Y)}$$

2. Le rapport de corrélation de $X$ en $Y$ est défini par

$$\eta^2_{X/Y} = \qquad = \frac{\frac{1}{n}\sum_i n_{\bullet j}(\overline{X}_j - \overline{X})^2}{Var(X)}$$

- et
- Plus $\eta^2$ est (resp. ) et plus la liaison fonctionnelle est (resp. )

Notes

| X=Age \ Y=Salaire | $[800, 1000[$ ($j=1$) | $[1000, 1200[$ ($j=2$) | Total |
|---|---|---|---|
| $[20, 22[$ ($i=1$) | **14** | **6** | **20** |
| $[22, 24[$ ($i=2$) | **28** | **46** | **74** |
| $[24, 26[$ ($i=3$) | **20** | **86** | **106** |
| Total | 62 | 138 | 200 |

Démarche pour calculer le rapport de corrélation de $X$ en $Y$ :

- calcul des moyenne et variance marginale de $X$ : $\overline{x} \simeq$ **(ans)** et $Var(X) \simeq$ (ans$^2$).
- calcul des moyennes conditionnelles de $X$ sachant $Y \in [800, 1000[$ et de $X$ sachant $Y \in [1000, 1200[$ : $\overline{x}_1 \simeq$ (ans) et $\overline{x}_2 \simeq$ (ans).
- calcul de la variance interpopulation (var. moy. cond.)

$$Var.Inter = \frac{62 \times (\quad - \quad)^2 + 138 \times (\quad - \quad)^2}{200} \simeq \quad (\text{ans}^2).$$

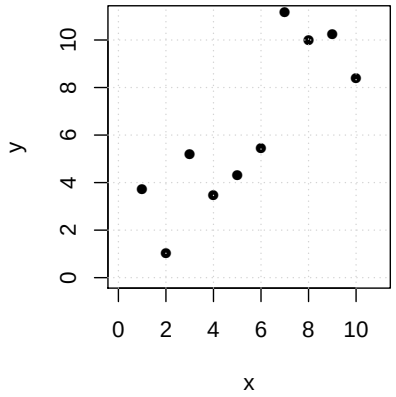
- $\eta^2_{X/Y} \simeq \text{—} \simeq$ %
( % de la variance de $X$ est expliquée par la variable $Y$).

Notes



## Régression linéaire

- Si le nuage de points observé est "presque" linéaire, il y a de fortes chances que la liaison entre $X$ et $Y$ soit linéaire (et que celle de $Y$ à $X$ soit linéaire).
- Exemple : imaginons observer le nuage suivant :




$\Rightarrow$ On peut suspecter une
Pour mesure ceci on utilise le coefficient de

Notes

## Coefficient de corrélation linéaire

Soit $(x_i, y_i)$ pour $i = 1, \ldots, n$ un nuage de points. Ce coefficient est défini par

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

où

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x} \times \overline{y}.$$

- Si $r$ est proche de , $X$ et $Y$ sont (certainement)
- Si la pente de la droite est Si , la pente de la droite est
- Si $r$ est proche de , l'ajustement linéaire n'est pas (⚠ce qui ne signifie pas que $X$ et $Y$ ne puissent pas être liées par une fonction).
- $r^2$ est appelé coefficient de $(0 \leq r^2 \leq 1)$.

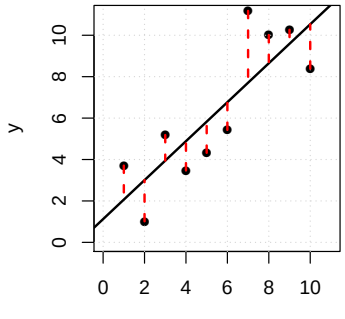
Notes



## Méthode des moindres carrés

Si le coefficient $r$ est jugé acceptable, on peut tenter d'estimer la droite de régression (de $Y$ en $X$) en utilisant la

on se donne une droite d'équation $y = ax + b$, la MMC consiste à minimiser la somme des écarts rouges au carré.




Autrement dit, on va chercher le minimum en $a$ et $b$ de la fonction

$$(a,b) \mapsto \sum_{i=1}^{n}(y_i - (ax_i + b))^2$$

Notes

# Solutions au problème

## La droite de régression ...

- ...de $Y$ en $X$ a pour équation $y = \widehat{a}x + \widehat{b}$ avec

$$\widehat{a} = \qquad \text{et } \widehat{b} =$$

- ...de $X$ en $Y$ a pour équation $x = \widehat{a'}y + \widehat{b'}$ avec

$$\widehat{a'} = \qquad \text{et } \widehat{b'} =$$

- les deux droites de régression passent par le point
- On peut remarquer que

$$\widehat{a} \times \widehat{a'} =$$

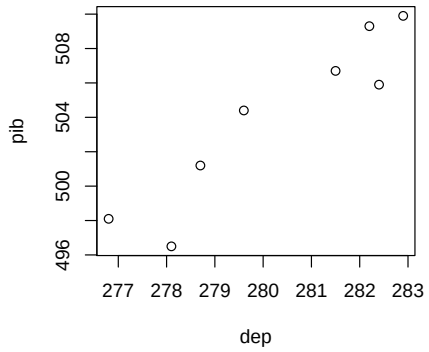
Notes



# Exemple d'application

Le tableau suivant présente les dépenses (`dep`) des ménages et PIB (`pib`) en milliards d'euros pour les 4 trimestres de 2011 et 2012. Peut-on expliquer l'évolution du PIB en fonction des dépenses ?

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| dep | 278.1 | 276.8 | 278.7 | 279.6 | 282.4 | 281.5 | 282.2 | 282.9 |
| pib | 496.5 | 498.1 | 501.2 | 504.4 | 505.9 | 506.7 | 509.3 | 509.9 |




$\Rightarrow$ L'ajustement linéaire semble

Notes

# Exemple d'application (2)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| dep | 278.1 | 276.8 | 278.7 | 279.6 | 282.4 | 281.5 | 282.2 | 282.9 |
| pib | 496.5 | 498.1 | 501.2 | 504.4 | 505.9 | 506.7 | 509.3 | 509.9 |

Démarche

1. Calculez $\overline{dep}$, $\overline{pib}$, $Var(dep)$ et $Var(pib)$

   $\overline{dep} \simeq$ $(Me)$, $\overline{pib} \simeq$ $(Me)$, $Var(dep) \simeq$ $(Me)^2$, $Var(pib) \simeq$ $(Me)^2$

2. Calcul intermédiaire

   $$\overline{dep \times pib} = \frac{1}{8}(278 \times 496 + \ldots + 283 \times 510) = \quad (Me)^2.$$

3. Calcul de la covariance

   $$Cov(dep, pib) = \overline{dep \times pib} - \overline{dep} \times \overline{pib} \simeq \quad (Me)^2.$$

4. Calcul du coefficient de corrélation linéaire

   $$R = \frac{\qquad}{\sqrt{\qquad}} \simeq$$

5. Puisque l'ajustement linéaire est très bon, calculons la droite de régression

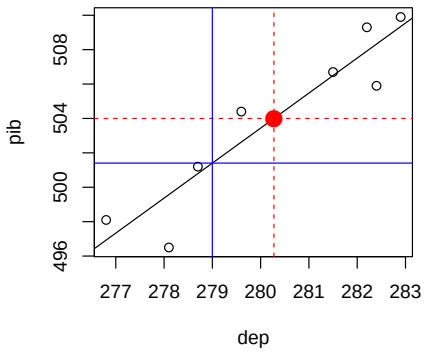
   $\widehat{a} =$ et $\widehat{b} =$ $(Me)$.

Notes



# Exemple d'application (3)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| dep | 278.1 | 276.8 | 278.7 | 279.6 | 282.4 | 281.5 | 282.2 | 282.9 |
| pib | 496.5 | 498.1 | 501.2 | 504.4 | 505.9 | 506.7 | 509.3 | 509.9 |




- La droite de régression

  $$pib = 2.04 \times dep - 67.77.$$

  passe par le point $(\overline{dep}, \overline{pib})$ .
- Quelle estimation du PIB proposer pour une $dep = 279$ **(M€)** ? $\Rightarrow$

  $\widehat{\textbf{pib}} =$
  $=$ $(Me)$.

Notes