Ajustement linéaire par les moindres carrés

Clément Rau
Laboratoire de Mathématiques de Toulouse
Université Paul Sabatier-IUT GEA Ponsan

Module: Outils maths pour la gestion (S2)

Introduction

Motivations :

Introduction

Motivations :

- Examiner la dépendance entre deux variables.

Introduction

Motivations :

- Examiner la dépendance entre deux variables.
  - Trouver un critère.

Introduction

Motivations :

- Examiner la dépendance entre deux variables.
  - Trouver un critère.

Introduction

Motivations :

- Examiner la dépendance entre deux variables.
  - Trouver un critère.
- Si il y a une dépendance, trouver une relation linéaire la plus "fidèle".

Introduction

Motivations :

- Examiner la dépendance entre deux variables.
  - Trouver un critère.
- Si il y a une dépendance, trouver une relation linéaire la plus "fidèle".
- Intérets :

Introduction

Motivations :

- Examiner la dépendance entre deux variables.
  - Trouver un critère.
- Si il y a une dépendance, trouver une relation linéaire la plus "fidèle".
- Intérets : Prévisions, pronostics...

Exemple 1

Exemple 1

Un responsable du tourisme d'une station balnéaire fait le bilan de la fréquentation touristique de 1994 à 2001 :

Année x_i	1994	1995	1996	1997	1998	1999	2000	2001
Nbr de touristes y_i en milliers	24,4	26,3	27,8	29,5	30,7	32,8	34,4	35,7

Exemple 1

Un responsable du tourisme d'une station balnéaire fait le bilan de la fréquentation touristique de 1994 à 2001 :

Année x_i	1994	1995	1996	1997	1998	1999	2000	2001
Nbr de touristes y_i en milliers	24,4	26,3	27,8	29,5	30,7	32,8	34,4	35,7

1. Est il légitime de supposer une relation linéaire entre le nombre de touristes et les années ?

Exemple 1

Un responsable du tourisme d'une station balnéaire fait le bilan de la fréquentation touristique de 1994 à 2001 :

Année x_i	1994	1995	1996	1997	1998	1999	2000	2001
Nbr de touristes y_i en milliers	24,4	26,3	27,8	29,5	30,7	32,8	34,4	35,7

1. Est il légitime de supposer une relation linéaire entre le nombre de touristes et les années ?
2. Déterminer le nombre de touristes que l'on peut prévoir en 2004 ?

Exemple 1

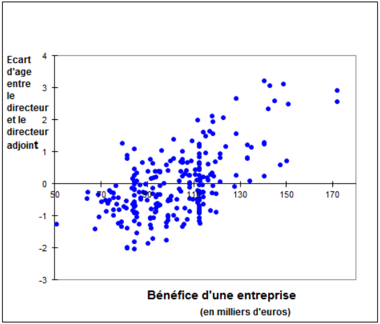
Un responsable du tourisme d'une station balnéaire fait le bilan de la fréquentation touristique de 1994 à 2001 :

Année x_i	1994	1995	1996	1997	1998	1999	2000	2001
Nbr de touristes y_i en milliers	24,4	26,3	27,8	29,5	30,7	32,8	34,4	35,7

1. Est il légitime de supposer une relation linéaire entre le nombre de touristes et les années ?
2. Déterminer le nombre de touristes que l'on peut prévoir en 2004 ?
3. À partir de quelle année la fréquentation touristique dépassera 45000 touristes ?

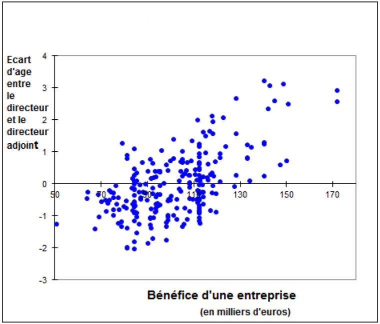
Exemple 2

On a demandé à 100 entreprises, le bénéfice annuel et l'écart d'âge entre le directeur et le sous directeur. On a obtenu les résultats suivants :



Exemple 2

On a demandé à 100 entreprises, le bénéfice annuel et l'écart d'âge entre le directeur et le sous directeur. On a obtenu les résultats suivants :



Est il légitime de supposer une relation linéaire entre ces 2 variables ?

Buts du chapitre

- Ces problèmes de dépendance entrent dans la théorie de la correlation.

Buts du chapitre

- Ces problèmes de dépendance entrent dans la théorie de la correlation.
  - Dans le premier exemple, un dépendance semble plausible.

Buts du chapitre

- Ces problèmes de dépendance entrent dans la théorie de la correlation.
  - Dans le premier exemple, un dépendance semble plausible.
  - Dans le second exemple, une dépendance semble illusoire.

Buts du chapitre

- Ces problèmes de dépendance entrent dans la théorie de la correlation.
  - Dans le premier exemple, un dépendance semble plausible.
  - Dans le second exemple, une dépendance semble illusoire.

Buts du chapitre

- Ces problèmes de dépendance entrent dans la théorie de la correlation.
  - Dans le premier exemple, un dépendance semble plausible.
  - Dans le second exemple, une dépendance semble illusoire.
- Fabriquer un nombre qui "mesure" cette dépendance (linéaire).

Buts du chapitre

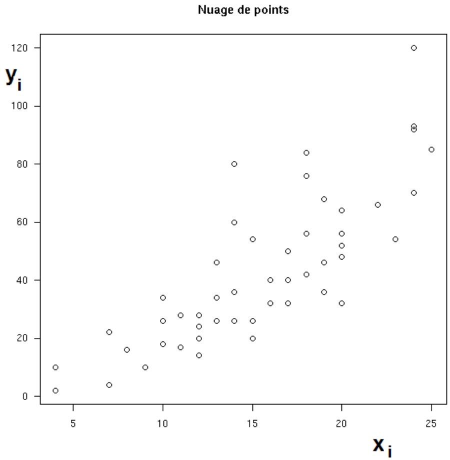
- Ces problèmes de dépendance entrent dans la théorie de la correlation.
  - Dans le premier exemple, un dépendance semble plausible.
  - Dans le second exemple, une dépendance semble illusoire.
- Fabriquer un nombre qui "mesure" cette dépendance (linéaire).
- Si ce nombre est "correct", trouver la "meilleure" relation linéaire qui lie les deux grandeurs.
  - Application du chapitre précédent : pb d'optimisation !

- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments.

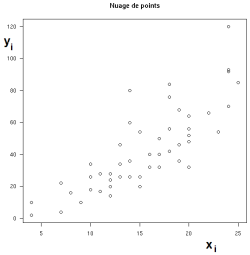
- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.

- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.
- On peut dans un premier temps, représenter ces points sur un graphe pour se faire une idée. Ce graphe s'appelle nuage de points

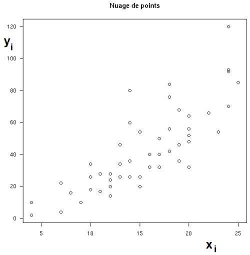
- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.
- On peut dans un premier temps, représenter ces points sur un graphe pour se faire une idée. Ce graphe s'appelle nuage de points


Nuage de points
120
100
80
60
40
20
0
y_i
5
10
15
20
25
x_i

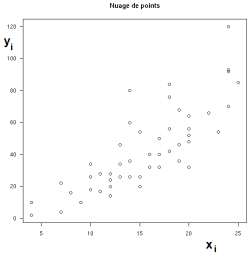
- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.
- On peut dans un premier temps, représenter ces points sur un graphe pour se faire une idée. Ce graphe s'appelle nuage de points



- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.
- On peut dans un premier temps, représenter ces points sur un graphe pour se faire une idée. Ce graphe s'appelle nuage de points

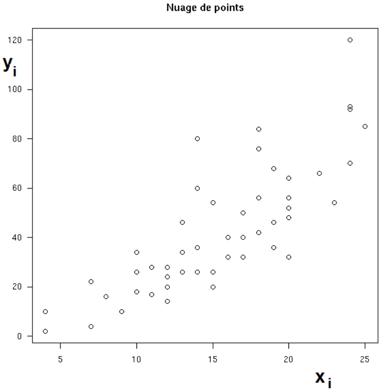


- On se donne deux séries $X = (x_i)_i$ et $Y = (y_i)_i$ ayant le même nombre d'éléments. But : établir un lien (si il existe) entre X et Y.
- On peut dans un premier temps, représenter ces points sur un graphe pour se faire une idée. Ce graphe s'appelle nuage de points

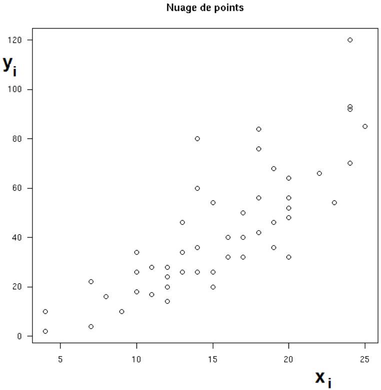


- Selon l'allure du nuage, on a envie de remplacer ce nuage par le graphe d'une fonction f. (la courbe $y = f(x)$) Cette opération s'appelle un ajustement

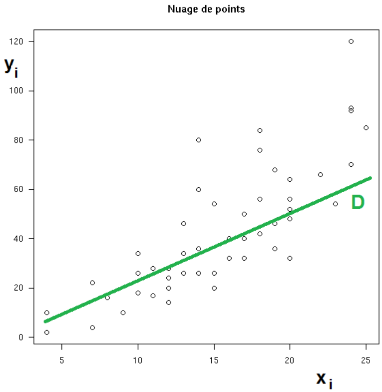
- La nature de l'ajustement dépend de la forme du nuage de points. Essentiellement, on étudiera ici les ajustements linéaires (ou ceux qui s'y raménent). On approchera donc le nuage par une droite...



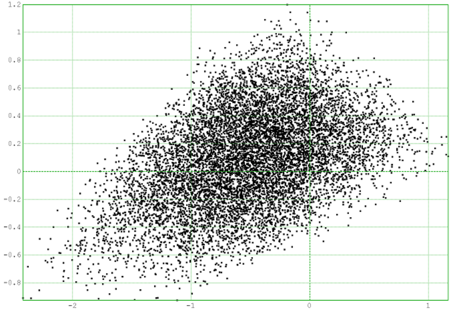
- La nature de l'ajustement dépend de la forme du nuage de points. Essentiellement, on étudiera ici les ajustements linéaires (ou ceux qui s'y raménent). On approchera donc le nuage par une droite...



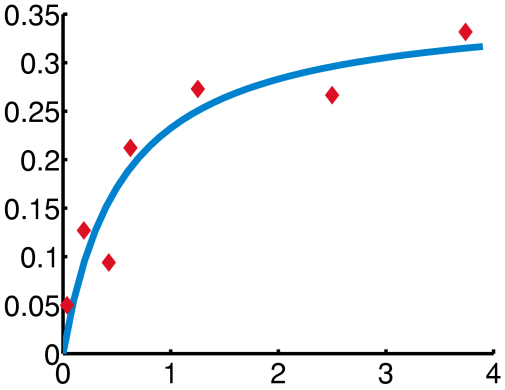
- La nature de l'ajustement dépend de la forme du nuage de points. Essentiellement, on étudiera ici les ajustements linéaires (ou ceux qui s'y raménent). On approchera donc le nuage par une droite...



- La nature de l'ajustement dépend de la forme du nuage de points. Essentiellement, on étudiera ici les ajustements linéaires (ou ceux qui s'y raménent). On approchera donc le nuage par une droite...



- La nature de l'ajustement dépend de la forme du nuage de points. Essentiellement, on étudiera ici les ajustements linéaires (ou ceux qui s'y raménent). On approchera donc le nuage par une droite...



Divers ajustements

On peut penser à beaucoup d'ajustements.

Divers ajustements

On peut penser à beaucoup d'ajustements.

- Ajustement linéaire $Y = aX + b$

Divers ajustements

On peut penser à beaucoup d'ajustements.

- Ajustement linéaire $Y = aX + b$
- Ajustement non linéaire :

Divers ajustements

On peut penser à beaucoup d'ajustements.

- Ajustement linéaire $Y = aX + b$
- Ajustement non linéaire :
  - ajustement exponentiel : $Y = \lambda e^{\alpha X}$

Divers ajustements

On peut penser à beaucoup d'ajustements.

- Ajustement linéaire $Y = aX + b$
- Ajustement non linéaire :
  - ajustement exponentiel : $Y = \lambda e^{\alpha X}$
  - ajustement logarithmique : $Y = a \ln(X) + b$

Divers ajustements

On peut penser à beaucoup d'ajustements.

- Ajustement linéaire $Y = aX + b$
- Ajustement non linéaire :
  - ajustement exponentiel : $Y = \lambda e^{\alpha X}$
  - ajustement logarithmique : $Y = a\ln(X) + b$
  - ajustement polynomial, par ex : $Y = aX^2 + bX + c$.

Principe

- Historiquement, l’idée est manuelle. On tracait le nuage de points et on faisait "pivoter" une règle qui passait au "mieux" entre les points du nuage.

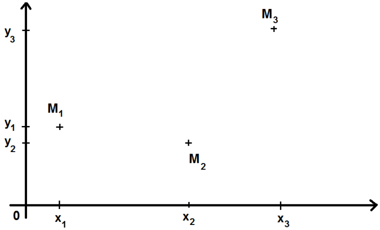
Principe

- Historiquement, l'idée est manuelle. On tracait le nuage de points et on faisait "pivoter" une règle qui passait au "mieux" entre les points du nuage.
- Il faut avoir conscience que l'on pourrait ajuster de diverses manières, en fonction du choix de la distance on veut minimiser.

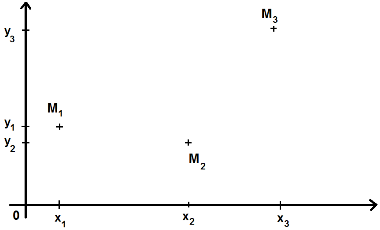
Principe

- Historiquement, l'idée est manuelle. On tracait le nuage de points et on faisait "pivoter" une règle qui passait au "mieux" entre les points du nuage.
- Il faut avoir conscience que l'on pourrait ajuster de diverses manières, en fonction du choix de la distance on veut minimiser.
- La méthode des moindres carrés s'interesse à une *certaine* distance. Précisons les choses...

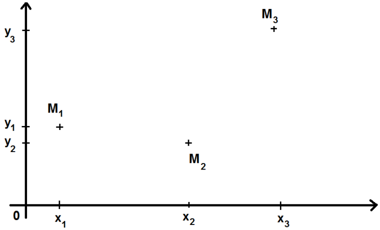
- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)



- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)

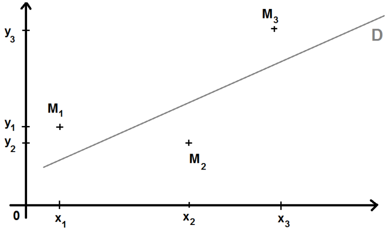


- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)



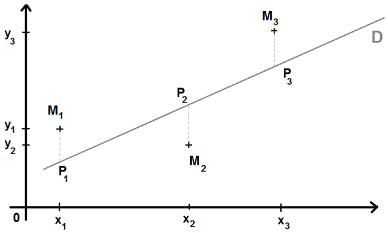
- Soit D une droite quelconque, d'équation $y = ax + b$.

- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)



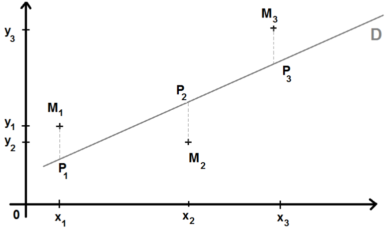
- Soit D une droite quelconque, d'équation $y = ax + b$.

- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)



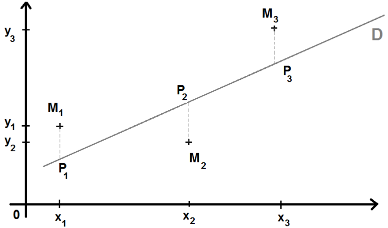
- Soit D une droite quelconque, d'équation $y = ax + b$.
- Soient P_i la projection de M_i sur D parallèlement à (Oy).

- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)

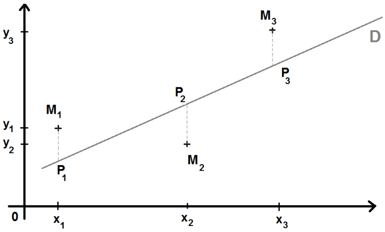


- Soit D une droite quelconque, d'équation $y = ax + b$.
- Soient P_i la projection de M_i sur D parallèlement à (Oy). Donc les coordonnées de P_i sont de la forme (x_i, y_i').

- Soient $X = (x_i)_i$ et $Y = (y_i)_i$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i)

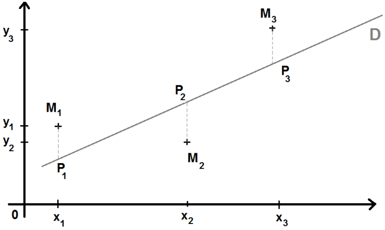


- Soit D une droite quelconque, d'équation $y = ax + b$.
- Soient P_i la projection de M_i sur D parallèlement à (Oy). Donc les coordonnées de P_i sont de la forme (x_i, y_i'). Notez que $P_i \in D$, donc $y_i' = ax_i + b$.



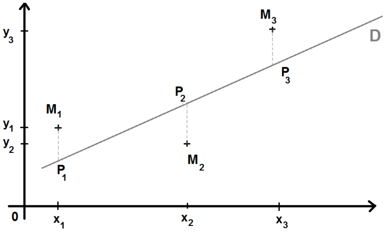
ie : la quantité

$$Z = \sum_i P_iM_i^2 = \sum_i (y_i - y_i')^2 = \sum_i (y_i - ax_i - b)^2.$$



Definition

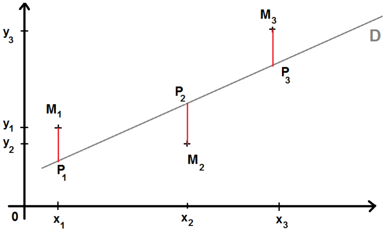
Ajuster par la méthode des moindres carrés consiste à rechercher parmi toutes les droites D, celle qui rend minimale la somme des carrés des longueurs P_iM_i.



Definition

Ajuster par la méthode des moindres carrés consiste à rechercher parmi toutes les droites D, celle qui rend minimale la somme des carrés des longueurs P_iM_i.
ie : la quantité

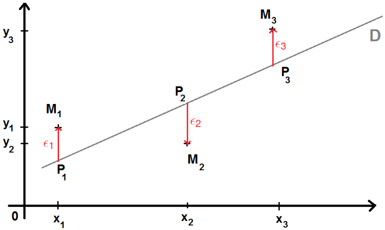
$$Z = \sum_i P_iM_i^2 = \sum_i (y_i - y_i')^2 = \sum_i (y_i - ax_i - b)^2.$$



Definition

Ajuster par la méthode des moindres carrés consiste à rechercher parmi toutes les droites D, celle qui rend mimimal la somme des carrés des longueur P_iM_i.
ie : la quantité

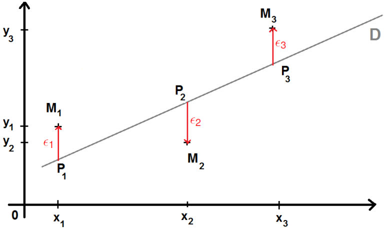
$$Z = \sum_i P_iM_i^2 = \sum_i (y_i - y_i')^2 = \sum_i (y_i - ax_i - b)^2.$$



On notera

$$\epsilon_i = \overline{P_i M_i},$$

(mesure algèbrique de $P_i M_i$).



On notera

$$\epsilon_i = \overline{P_i M_i},$$

(mesure algèbrique de $P_i M_i$).
Ainsi on cherche la (ou les) droite(s) D qui rende(nt) minimal

$$\begin{aligned} Z &= \sum_i \epsilon_i^2 \\ &= \sum_i (y_i - ax_i - b)^2 \end{aligned}$$

$\Rightarrow$ Probléme d'optimisation sans contraintes.

Détermination du min

Détermination du min

Le nuage $(x_i, y_i)_{i=1..n}$ est fixé. On a donc l'application :

$$\begin{array}{ccc} \{Droites\} & \rightarrow & \mathbb{R}_+ \\ D & \mapsto & Z = Z_D \end{array}$$

que l'on cherche à minimiser.

Détermination du min

Le nuage $(x_i, y_i)_{i=1..n}$ est fixé. On a donc l'application :

$$\begin{array}{ccc} \{Droites\} & \rightarrow & \mathbb{R}_+ \\ D & \mapsto & Z = Z_D \end{array}$$

que l'on cherche à minimiser.
Une droite D non parallèle à l'axe (Oy) étant déterminée par son coeff directeur a et son ordonnée à l'origine b, on peut plutôt considérer l'application suivante

$$\begin{array}{cccl} \phi : & \mathbb{R}^2 & \rightarrow & \mathbb{R}_+ \\ & (a, b) & \mapsto & Z = \sum_i (y_i - ax_i - b)^2 \end{array}$$

Point(s) critique(s) de ϕ

$$\begin{array}{rccl} \phi : & \mathbb{R}^2 & \to & \mathbb{R}_+ \\ & (a,b) & \mapsto & Z = \sum_i (y_i - ax_i - b)^2 \end{array}$$

Determinons le (ou les) point(s) critique(s) de ϕ.

Point(s) critique(s) de ϕ

$$\begin{array}{rccl} \phi : & \mathbb{R}^2 & \to & \mathbb{R}_+ \\ & (a,b) & \mapsto & Z = \sum_i (y_i - ax_i - b)^2 \end{array}$$

Determinons le (ou les) point(s) critique(s) de ϕ.

On résoud donc : $\begin{cases} \frac{\partial \phi}{\partial a}(a,b) = 0 \\ \frac{\partial \phi}{\partial b}(a,b) = 0 \end{cases}$

Point(s) critique(s) de ϕ

$$\begin{array}{rccl} \phi : & \mathbb{R}^2 & \rightarrow & \mathbb{R}_+ \\ & (a,b) & \mapsto & Z = \sum_i (y_i - ax_i - b)^2 \end{array}$$

Determinons le (ou les) point(s) critique(s) de ϕ.

On résoud donc : $\begin{cases} \frac{\partial \phi}{\partial a}(a,b) = 0 \\ \frac{\partial \phi}{\partial b}(a,b) = 0 \end{cases}$

$$ie : \begin{cases} -2\sum_i x_i(y_i - ax_i - b) = 0 \\ -2\sum_i (y_i - ax_i - b) = 0 \end{cases}$$

Point(s) critique(s) de ϕ

$$\begin{array}{rccl} \phi : & \mathbb{R}^2 & \to & \mathbb{R}_+ \\ & (a,b) & \mapsto & Z = \sum_i (y_i - ax_i - b)^2 \end{array}$$

Determinons le (ou les) point(s) critique(s) de ϕ.

On résoud donc : $\left\{ \begin{array}{l} \frac{\partial \phi}{\partial a}(a,b) = 0 \\ \frac{\partial \phi}{\partial b}(a,b) = 0 \end{array} \right.$

$$ie : \left\{ \begin{array}{l} -2\sum_i x_i(y_i - ax_i - b) = 0 \\ -2\sum_i (y_i - ax_i - b) = 0 \end{array} \right.$$

D'où

$$\left\{ \begin{array}{l} \sum_i x_i y_i - a\sum_i x_i^2 - b\sum_i x_i = 0 \\ \sum_i y_i - a\sum_i x_i - bn = 0 \end{array} \right.$$

En notant, $\bar{x} = \frac{1}{n}\sum_i x_i$ et $\bar{y} = \frac{1}{n}\sum_i y_i$, le systéme se réécrit :

$$\begin{cases} \sum_i x_i y_i - a\sum_i x_i^2 - bn\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases}$$

En notant, $\bar{x} = \frac{1}{n}\sum_i x_i$ et $\bar{y} = \frac{1}{n}\sum_i y_i$, le systéme se réécrit :

$$\begin{cases} \sum_i x_i y_i - a\sum_i x_i^2 - bn\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases}$$

De la ligne 2, on tire que $b = \bar{y} - a\bar{x}$, que l'on reporte dans la ligne 1.

En notant, $\bar{x} = \frac{1}{n}\sum_i x_i$ et $\bar{y} = \frac{1}{n}\sum_i y_i$, le systéme se réécrit :

$$\begin{cases} \sum_i x_i y_i - a\sum_i x_i^2 - bn\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases}$$

De la ligne 2, on tire que $b = \bar{y} - a\bar{x}$, que l'on reporte dans la ligne 1.On trouve alors :

$$\sum_i x_i y_i - a\sum_i x_i^2 - (\bar{y} - a\bar{x})n\bar{x} = 0.$$

En notant, $\bar{x} = \frac{1}{n}\sum_i x_i$ et $\bar{y} = \frac{1}{n}\sum_i y_i$, le systéme se réécrit :

$$\left\{ \begin{array}{l} \sum_i x_i y_i - a\sum_i x_i^2 - bn\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{array} \right.$$

De la ligne 2, on tire que $b = \bar{y} - a\bar{x}$, que l'on reporte dans la ligne 1.On trouve alors :

$$\sum_i x_i y_i - a\sum_i x_i^2 - (\bar{y} - a\bar{x})n\bar{x} = 0.$$

Puis,

$$\begin{aligned} a &= \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} \\ &= \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i x_i^2 - \bar{x}^2} = \frac{cov(x,y)}{var(x)} \end{aligned}$$

En notant, $\bar{x} = \frac{1}{n}\sum_i x_i$ et $\bar{y} = \frac{1}{n}\sum_i y_i$, le systéme se réécrit :

$$\begin{cases} \sum_i x_i y_i - a\sum_i x_i^2 - bn\bar{x} = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases}$$

De la ligne 2, on tire que $b = \bar{y} - a\bar{x}$, que l'on reporte dans la ligne 1.On trouve alors :

$$\sum_i x_i y_i - a\sum_i x_i^2 - (\bar{y} - a\bar{x})n\bar{x} = 0.$$

Puis,

$$\begin{aligned} a &= \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} \\ &= \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i x_i^2 - \bar{x}^2} = \frac{cov(x,y)}{var(x)} \end{aligned}$$

On a donc un point critique (a, b) définit par les formules précédentes. Reste à voir si c'est bien un extremum...

Hessienne de ϕ

$$H_{(a,b)} = \begin{pmatrix} \frac{\partial^2 \phi}{\partial^2 a}(a,b) & \frac{\partial^2 \phi}{\partial a \partial b}(a,b) \\ \frac{\partial^2 \phi}{\partial a \partial b}(a,b) & \frac{\partial^2 \phi}{\partial^2 b}(a,b) \end{pmatrix} = \begin{pmatrix} 2\sum_i x_i^2 & 2\sum_i x_i \\ 2\sum_i x_i & 2n \end{pmatrix}$$

Hessienne de ϕ

$$H_{(a,b)} = \begin{pmatrix} \frac{\partial^2 \phi}{\partial^2 a}(a,b) & \frac{\partial^2 \phi}{\partial a \partial b}(a,b) \\ \frac{\partial^2 \phi}{\partial a \partial b}(a,b) & \frac{\partial^2 \phi}{\partial^2 b}(a,b) \end{pmatrix} = \begin{pmatrix} 2\sum_i x_i^2 & 2\sum_i x_i \\ 2\sum_i x_i & 2n \end{pmatrix}$$

On a :

$$\begin{cases} Tr(H_{(a,b)}) = 2n + 2\sum_i x_i^2 > 0 \\ et \\ det(H_{(a,b)}) = 4n\sum_i x_i^2 - 4(\sum_i x_i)^2 = 4n^2 Var(x) > 0 \end{cases}$$

Hessienne de ϕ

$$H_{(a,b)} = \begin{pmatrix} \frac{\partial^2\phi}{\partial^2 a}(a,b) & \frac{\partial^2\phi}{\partial a\partial b}(a,b) \\ \frac{\partial^2\phi}{\partial a\partial b}(a,b) & \frac{\partial^2\phi}{\partial^2 b}(a,b) \end{pmatrix} = \begin{pmatrix} 2\sum_i x_i^2 & 2\sum_i x_i \\ 2\sum_i x_i & 2n \end{pmatrix}$$

On a :

$$\begin{cases} Tr(H_{(a,b)}) = 2n + 2\sum_i x_i^2 > 0 \\ et \\ det(H_{(a,b)}) = 4n\sum_i x_i^2 - 4(\sum_i x_i)^2 = 4n^2 Var(x) > 0 \end{cases}$$

Ainsi, le (a,b) trouvé est bien un min. Et la droite d'équation $y = ax + b$ minimise bien $\sum_i \epsilon_i^2$

Synthèse - à retenir !

Proposition

La droite d'ajustement de y en fonction de x a une équation du type $y = ax + b$,

$$\left\{\begin{array}{l} a = \frac{cov(x,y)}{var(x)} \\ et \\ b = \bar{y} - a\bar{x} \end{array}\right.$$

où

$$\begin{aligned} cov(x,y) &= \frac{1}{n}\sum_i x_i y_i - \overline{x}\,\overline{y} \\ &= \overline{xy} - \overline{x}\,\overline{y}. \end{aligned}$$

Remarques

- $\frac{\partial \phi}{\partial b}(a,b) = 0$, donne que $\sum_i \epsilon_i = 0$

Remarques

- $\frac{\partial \phi}{\partial b}(a, b) = 0$, donne que $\sum_i \epsilon_i = 0$ La droite "passe bien au milieu" du nuage...

Remarques

- $\frac{\partial \phi}{\partial b}(a, b) = 0$, donne que $\sum_i \epsilon_i = 0$ La droite "passe bien au milieu" du nuage...
- Autre expression du coeff directeur a, on a :

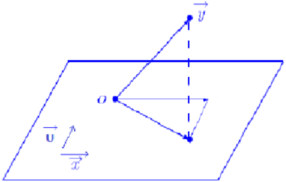
$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}.$$

Remarques

- $\frac{\partial \phi}{\partial b}(a,b) = 0$, donne que $\sum_i \epsilon_i = 0$ La droite "passe bien au milieu" du nuage...
- Autre expression du coeff directeur a, on a :
$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}.$$
- Géométriquement toutes ces manipulations, correspondent à une projection

Remarques

- $\frac{\partial\phi}{\partial b}(a,b) = 0$, donne que $\sum_i \epsilon_i = 0$. La droite "passe bien au milieu" du nuage...
- Autre expression du coeff directeur a, on a :
$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}.$$
- Géométriquement toutes ces manipulations, correspondent à une projection



Exemple Taille - Poids

On donne :

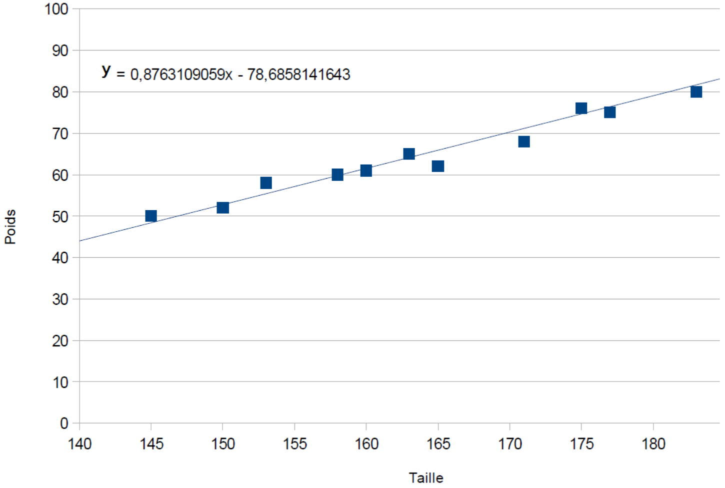
Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

Exemple Taille - Poids

On donne :

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

L'ajustement linéaire fournit $y = 0,876x - 78,7$


y = 0,8763109059x - 78,6858141643
Poids
Taille
100
90
80
70
60
50
40
30
20
10
0
140
145
150
155
160
165
170
175
180

Attention aux erreurs d'approximation. Conserver un nombre de décimales suffisant dans vos calculs à la main pour ne pas altérer les valeurs de a et b.

Exemple non linéaire

Exemple non linéaire

On a relevé le bénéfice en milliers d'euros sur 24 mois :

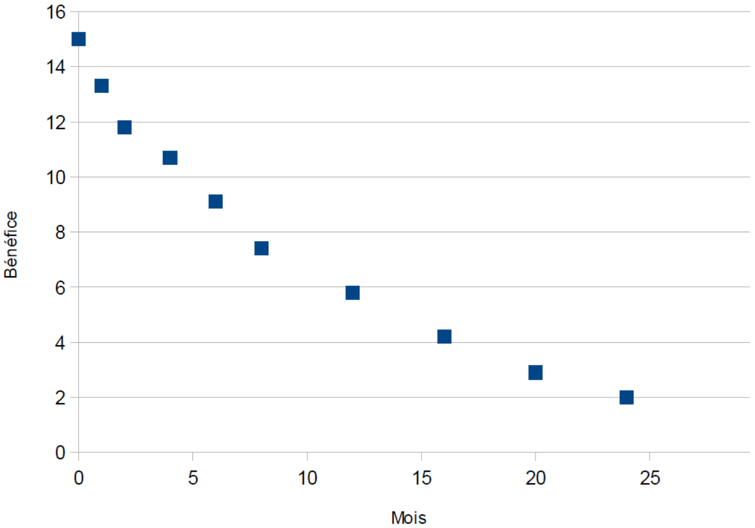
Mois	0	1	2	4	6	8	12	16	20	24
Bénéfice	15	$13,3$	$11,8$	$10,7$	$9,1$	$7,4$	$5,8$	$4,2$	$2,9$	2

Exemple non linéaire

On a relevé le bénéfice en milliers d'euros sur 24 mois :

Mois	0	1	2	4	6	8	12	16	20	24
Bénéfice	15	$13,3$	$11,8$	$10,7$	$9,1$	$7,4$	$5,8$	$4,2$	$2,9$	2

Traçons le nuage.


16
14
12
10
8
6
4
2
0
Bénéfice
0
5
10
15
20
25
Mois

- Un ajustement linéaire "brut" semble peu adapté.

- Un ajustement linéaire "brut" semble peu adapté.
- On cherche un ajustement de la forme

$$B(t) = B_0 e^{-\alpha t},$$

où B représente le bénéfice au temps t. (B_0 et α sont à déterminer)

- Un ajustement linéaire "brut" semble peu adapté.
- On cherche un ajustement de la forme

$$B(t) = B_0 e^{-\alpha t},$$

où B représente le bénéfice au temps t. (B_0 et α sont à déterminer)
- En passant au ln, on obtient

$$\ln(B(t)) = \ln(B_0) - \alpha t.$$

- Un ajustement linéaire "brut" semble peu adapté.
- On cherche un ajustement de la forme

$$B(t) = B_0 e^{-\alpha t},$$

  où B représente le bénéfice au temps t. (B_0 et α sont à déterminer)
- En passant au ln, on obtient

$$\ln(B(t)) = \ln(B_0) - \alpha t.$$

  Ainsi $\ln(B)$ est une fonction linéaire de t.

- Un ajustement linéaire "brut" semble peu adapté.
- On cherche un ajustement de la forme

$$B(t) = B_0 e^{-\alpha t},$$

où B représente le bénéfice au temps t. (B_0 et α sont à déterminer)

- En passant au $\ln$, on obtient

$$\ln(B(t)) = \ln(B_0) - \alpha t.$$

Ainsi $\ln(B)$ est une fonction linéaire de t. On réalise ainsi un ajustement linéaire avec les variables $\ln(B)$ et t.

- Un ajustement linéaire "brut" semble peu adapté.
- On cherche un ajustement de la forme

$$B(t) = B_0 e^{-\alpha t},$$

où B représente le bénéfice au temps t. (B_0 et α sont à déterminer)

- En passant au ln, on obtient

$$\ln(B(t)) = \ln(B_0) - \alpha t.$$

Ainsi $\ln(B)$ est une fonction linéaire de t. On réalise ainsi un ajustement linéaire avec les variables $\ln(B)$ et t.

- On obtient

$$\ln(B) \approx -0,081t + 2,68.$$

D'où

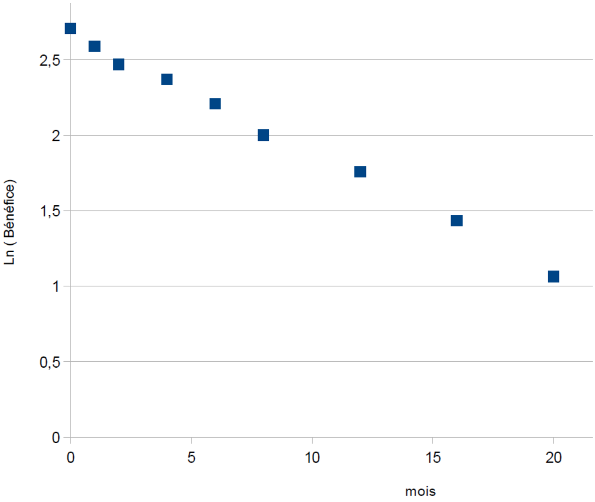
$$B(t) \approx 14,6\, e^{-0,081t}.$$

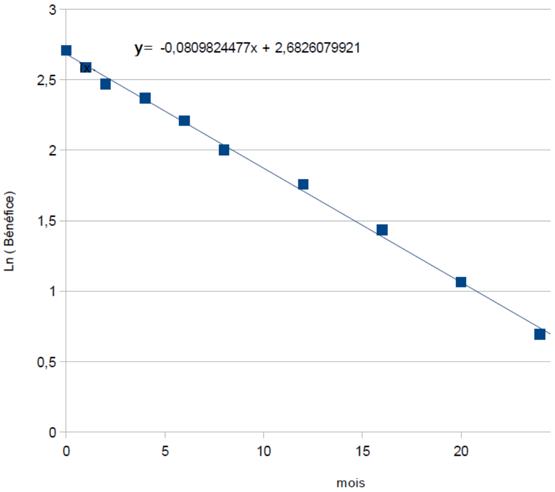
nouveau x

Mois	0	1	2	4	6	8	12	16	20	24
Bénéfice	15	13,3	11,8	10,7	9,1	7,4	5,8	4,2	2,9	2

ln(B)	2,708	2,588	2,468	2,370	2,208	2,001	1,758	1,435	1,065	0,693

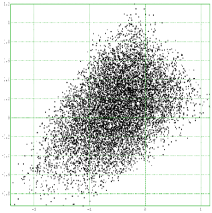
nouveau y


Ln (Bénéfice)
2,5
2
1,5
1
0,5
0
0
5
10
15
20
mois

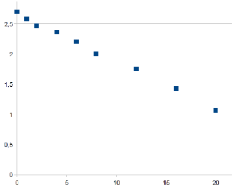

y= -0,0809824477x + 2,6826079921
Ln (Bénéfice)
mois
3
2,5
2
1,5
1
0,5
0
0
5
10
15
20

1. Ajustement
   - Position du problème
   - Ajustement par la méthode des moindres carrés
   - Exemples

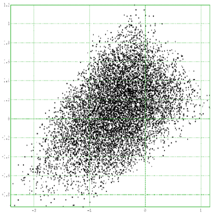
2. Correlation
   - Position du probléme
   - Droites de régression
   - Coefficient de correlation
   - Interpretation du coefficient de correlation



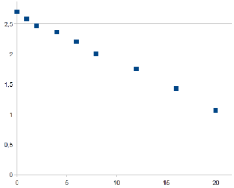
S_1



S_2

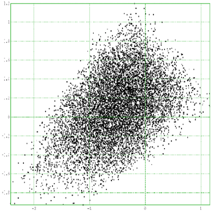


S_1

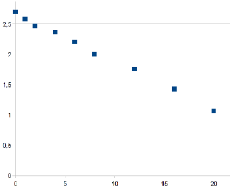


S_2

Ajuster linéairement S_2 semble réaliste alors que ajuster linéairement S_1 semble illusoire. Pourtant tous les calculs effectués en partie 1, peuvent être pratiqués...Les résultats seront de mauvaise "qualité".



S_1



S_2

Ajuster linéairement S_2 semble réaliste alors que ajuster linéairement S_1 semble illusoire. Pourtant tous les calculs effectués en partie 1, peuvent être pratiqués...Les résultats seront de mauvaise "qualité".

$\rightarrow$ Mesurer la qualité de l'ajustement.

outil : coefficient de correlation.

1. Ajustement
   - Position du problème
   - Ajustement par la méthode des moindres carrés
   - Exemples

2. Correlation
   - Position du probléme
   - **Droites de régression**
   - Coefficient de correlation
   - Interpretation du coefficient de correlation

Introduction

Reprenons l'exemple Taille/Poids.

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

Introduction

Reprenons l'exemple Taille/Poids.

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

On avait exprimé Y en fonction de X et on avait obtenu $y = 0,876x - 78,7$.

Introduction

Reprenons l'exemple Taille/Poids.

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

On avait exprimé Y en fonction de X et on avait obtenu $y = 0,876x - 78,7$.
On dit qu'on a effectué une régression de Y en fonction X.

Introduction

Reprenons l'exemple Taille/Poids.

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Taille X (en cm)	145	150	153	158	160	163	165	171	175	177	183	187
Poids Y (en kg)	50	52	58	60	61	65	62	68	76	75	80	90

On avait exprimé Y en fonction de X et on avait obtenu $y = 0,876x - 78,7$.
On dit qu'on a effectué une régression de Y en fonction X.

On aurait tres bien pu effectuer une régression de X en fonction Y ! ! !

Intervertissons les rôles de X et Y. On refait l'ajustement linéaire. On obtient un a' et un b', donné par :

$$\begin{cases} a' = \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i y_i^2 - \bar{y}^2} = \frac{cov(Y,X)}{var(Y)} \\ b' = \bar{x} - a'\bar{y} \end{cases}$$

Intervertissons les rôles de X et Y. On refait l'ajustement linéaire. On obtient un a' et un b', donné par :

$$\begin{cases} a' = \dfrac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i y_i^2 - \bar{y}^2} = \dfrac{cov(Y,X)}{var(Y)} \\ b' = \bar{x} - a'\bar{y} \end{cases}$$

La droite D' de régression de X par rapport à Y a donc pour équation $x = a'y + b'$.

Intervertissons les rôles de X et Y. On refait l'ajustement linéaire. On obtient un a' et un b', donné par :

$$\begin{cases} a' = \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i y_i^2 - \bar{y}^2} = \frac{cov(Y,X)}{var(Y)} \\ b' = \bar{x} - a'\bar{y} \end{cases}$$

La droite D' de régression de X par rapport à Y a donc pour équation $x = a'y + b'$.

$$\text{ie : } y = \frac{1}{a'}x - \frac{b'}{a'}.$$

Intervertissons les rôles de X et Y. On refait l'ajustement linéaire. On obtient un a' et un b', donné par :

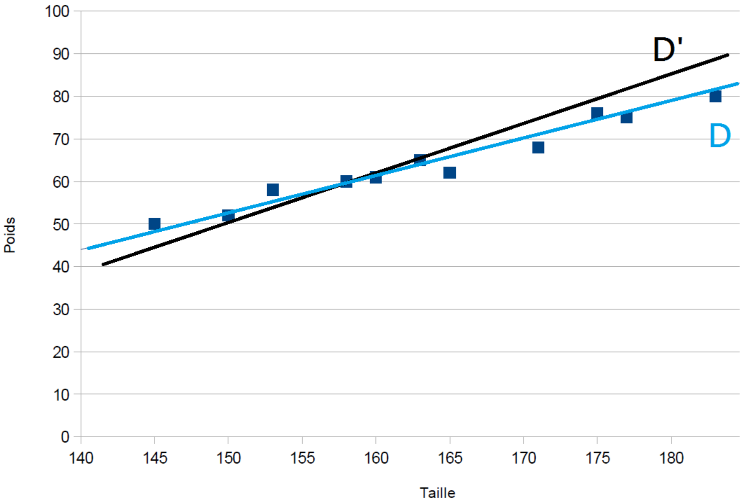
$$\begin{cases} a' = \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i y_i^2 - \bar{y}^2} = \frac{cov(Y,X)}{var(Y)} \\ b' = \bar{x} - a'\bar{y} \end{cases}$$

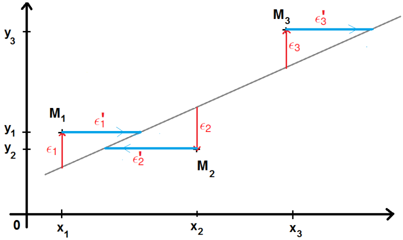
La droite D' de régression de X par rapport à Y a donc pour équation $x = a'y + b'$.

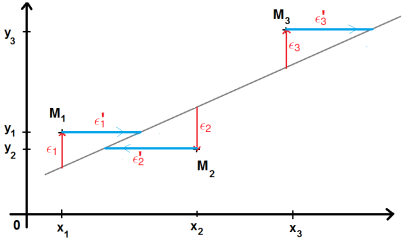
$$\text{ie : } y = \frac{1}{a'}x - \frac{b'}{a'}.$$

Dans notre exemple Taille/Poids, on trouve :

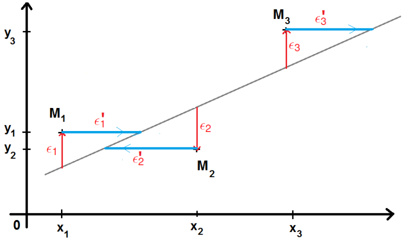
$$\begin{cases} (D) & y = 0,876x - 78,7 \\ (D') & x = 1,092y + 93,1, \quad ie: \ y = 0,916x - 85,2 \end{cases}$$


100
90
80
70
60
50
40
30
20
10
0
Poids
140
145
150
155
160
165
170
175
180
Taille
D'
D


y3
y1
y2
0
M1
M2
M3
x1
x2
x3
ε1
ε2
ε3
ε'1
ε'2
ε'3



D minimise la somme des écarts au carrés verticaux



D minimise la somme des écarts au carrés verticaux alors que (D') minimise la somme des écarts au carrés horizontaux

Proposition

- *(D) et (D') sont d'autant plus distinctes que le nuage de point est dispersé.*
- *Plus le nuage de points est "aligné", plus les deux droites de régression sont "proches" (les v.a X et Y sont dépendantes.)*

Proposition

- *(D) et (D') sont d'autant plus distinctes que le nuage de point est dispersé.*
- *Plus le nuage de points est "aligné", plus les deux droites de régression sont "proches" (les v.a X et Y sont dépendantes.)*

Cas particulier : Si le nuage est parfaitement aligné, on a $D = D'$.

Proposition

- *(D) et (D') sont d'autant plus distinctes que le nuage de point est dispersé.*
- *Plus le nuage de points est "aligné", plus les deux droites de régression sont "proches" (les v.a X et Y sont dépendantes.)*

Cas particulier : Si le nuage est parfaitement aligné, on a $D = D'$. Ainsi, les 2 droites ont les mêmes équations, ce qui se traduit par :

$$a = \frac{1}{a'} \quad et \quad b = -\frac{b'}{a'}$$

[Rappel : (D) a pour équation $y = ax + b$ et (D') a pour équation $y = \frac{1}{a'}x - \frac{b'}{a'}$.]

Proposition

- *(D) et (D') sont d'autant plus distinctes que le nuage de point est dispersé.*
- *Plus le nuage de points est "aligné", plus les deux droites de régression sont "proches" (les v.a X et Y sont dépendantes.)*

Cas particulier : Si le nuage est parfaitement aligné, on a $D = D'$. Ainsi, les 2 droites ont les mêmes équations, ce qui se traduit par :

$$a = \frac{1}{a'} \quad et \quad b = -\frac{b'}{a'}$$

[Rappel : (D) a pour équation $y = ax + b$ et (D') a pour équation $y = \frac{1}{a'}x - \frac{b'}{a'}$.]

Dans ce cas le produit $aa' = 1$.

Introduction

Ce qui précéde, incite donc à considérer le nombre aa', puis à le comparer à 1.

Introduction

Ce qui précéde, incite donc à considérer le nombre aa', puis à le comparer à 1.
Remarque : a *et* a' sont du même signe.

Introduction

Ce qui précéde, incite donc à considérer le nombre aa', puis à le comparer à 1.
Remarque : a et a' sont du même signe. En effet, les formules donnent :

$$a = \frac{cov(X,Y)}{var(X)} \quad et \quad a' = \frac{cov(Y,X)}{var(Y)}.$$

Introduction

Ce qui précéde, incite donc à considérer le nombre aa', puis à le comparer à 1.

Remarque : a *et* a' sont du même signe. En effet, les formules donnent :

$$a = \frac{cov(X,Y)}{var(X)} \quad et \quad a' = \frac{cov(Y,X)}{var(Y)}.$$

La positivité des variances et le fait que $cov(X,Y) = cov(Y,X)$, donnent le résultat.

Introduction

Ce qui précéde, incite donc à considérer le nombre aa', puis à le comparer à 1.

Remarque : a *et* a' sont du même signe. En effet, les formules donnent :

$$a = \frac{cov(X,Y)}{var(X)} \quad et \quad a' = \frac{cov(Y,X)}{var(Y)}.$$

La positivité des variances et le fait que $cov(X,Y) = cov(Y,X)$, donnent le résultat.

Definition

On appelle coefficient de correlation le nombre r tel que $r^2 = aa'$ et tel que r soit du signe de a et a'.

Definition

On appelle coefficient de correlation le nombre r tel que $r^2 = aa'$ et tel que r soit du signe de a et a'.

- *Si a et a' >0, alors $r = \sqrt{aa'}$,*
- *Si a et a' <0, alors $r = -\sqrt{aa'}$,*

On peut retenir $r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

Definition

On appelle coefficient de correlation le nombre r tel que $r^2 = aa'$ et tel que r soit du signe de a et a'.

- *Si a et a' >0, alors $r = \sqrt{aa'}$,*
- *Si a et a' <0, alors $r = -\sqrt{aa'}$,*

On peut retenir $r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

- On a tjrs $r \in [-1;1]$

Definition

On appelle coefficient de correlation le nombre r tel que $r^2 = aa'$ et tel que r soit du signe de a et a'.

- *Si a et a' >0, alors $r = \sqrt{aa'}$,*
- *Si a et a' <0, alors $r = -\sqrt{aa'}$,*

On peut retenir $r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

- On a tjrs $r \in [-1;1]$
- $cov(X,Y)$ est du même signe que a et a'.

Definition

On appelle coefficient de correlation le nombre r tel que $r^2 = aa'$ et tel que r soit du signe de a et a'.

- *Si a et a' >0, alors $r = \sqrt{aa'}$,*
- *Si a et a' <0, alors $r = -\sqrt{aa'}$,*

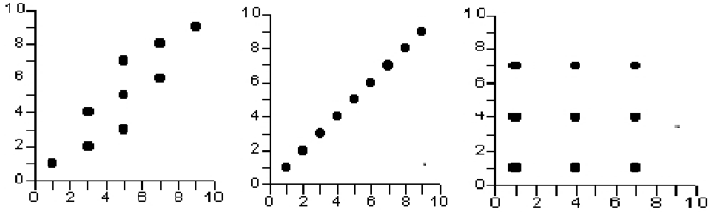
On peut retenir $r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

- On a tjrs $r \in [-1;1]$
- $cov(X, Y)$ est du même signe que a et a'.
- $var(X) = cov(X, X)$.

1. Ajustement
   - Position du problème
   - Ajustement par la méthode des moindres carrés
   - Exemples

2. Correlation
   - Position du probléme
   - Droites de régression
   - Coefficient de correlation
   - Interpretation du coefficient de correlation

Coeff de correlation	Droites de régression (D) et (D')	Liaison entre X et Y	Nuage de points
$\|r\| = 1$	confondues	dépendance linéaire parfaite	aligné
$\|r\| \approx 1$	voisine	dépendance linéaire d'autant plus prononcée que r est proche de 1	d'autant plus aligné que $\|r\|$ est proche de 1
$\|r\| \approx 0$	(D) et (D') sont quasi orthogonales	indépendant	dispersé



Relation faible

Relation forte
r proche de 1

Absence de relation
r proche de 0

Explication du contrôle de la dépendance par r

Vocabulaire :

Explication du contrôle de la dépendance par r

Vocabulaire :

- Variance résiduelle : $V_r = \frac{1}{n}\sum_i \epsilon_i^2 = \frac{1}{n}\sum_i (y_i - ax_i - b)^2$.

Explication du contrôle de la dépendance par r

Vocabulaire :

- Variance résiduelle : $V_r = \frac{1}{n}\sum_i \epsilon_i^2 = \frac{1}{n}\sum_i (y_i - ax_i - b)^2$.
- Variance expliquée : $V_e = \frac{1}{n}\sum_i (y_i' - \bar{y})^2$

Explication du contrôle de la dépendance par r

Vocabulaire :

- Variance résiduelle : $V_r = \frac{1}{n}\sum_i \epsilon_i^2 = \frac{1}{n}\sum_i (y_i - ax_i - b)^2$.
- Variance expliquée : $V_e = \frac{1}{n}\sum_i (y_i' - \bar{y})^2$

Or $\sum_i \epsilon_i = 0$ et $y_i' = y_i + \epsilon_i$, donc $\bar{y} = \bar{y'}$.

Explication du contrôle de la dépendance par r

Vocabulaire :

- Variance résiduelle : $V_r = \frac{1}{n}\sum_i \epsilon_i^2 = \frac{1}{n}\sum_i (y_i - ax_i - b)^2$.
- Variance expliquée : $V_e = \frac{1}{n}\sum_i (y_i' - \bar{y})^2$

Or $\sum_i \epsilon_i = 0$ et $y_i' = y_i + \epsilon_i$, donc $\bar{y} = \bar{y'}$. Ainsi,

$$V_e = \frac{1}{n}\sum_i (y_i' - \bar{y'})^2 = var(Y') = var(aX + b) = a^2 var(X).$$

Explication du contrôle de la dépendance par r

Vocabulaire :

- Variance résiduelle : $V_r = \frac{1}{n}\sum_i \epsilon_i^2 = \frac{1}{n}\sum_i (y_i - ax_i - b)^2$.
- Variance expliquée : $V_e = \frac{1}{n}\sum_i (y_i' - \bar{y})^2$

Or $\sum_i \epsilon_i = 0$ et $y_i' = y_i + \epsilon_i$, donc $\bar{y} = \bar{y'}$. Ainsi,

$$V_e = \frac{1}{n}\sum_i (y_i' - \bar{y'})^2 = var(Y') = var(aX + b) = a^2 var(X).$$

$$\boxed{\text{Fait 1 : } V(Y) = V_e + V_r}$$

$\boxed{\text{Fait 1 : } V(Y) = V_e + V_r}$

$\boxed{\text{Fait 1 : } V(Y) = V_e + V_r}$ Preuve : On a successivement,

$$
\begin{aligned}
V(Y) &= \frac{1}{n}\sum_i (y_i - \bar{y})^2 \\
&= \frac{1}{n}\sum_i (y_i - y_i' + y_i' - \bar{y})^2 \\
&= \frac{1}{n}\sum_i (y_i - y_i')^2 + \frac{1}{n}\sum_i (y_i' - \bar{y})^2 + \frac{2}{n}\sum_i (y_i - y_i')(y_i' - \bar{y}) \\
&= V_r + V_e + \frac{2}{n}\sum_i (y_i - y_i')(y_i' - \bar{y})
\end{aligned}
$$

$\boxed{\text{Fait 1 : } V(Y) = V_e + V_r}$ Preuve : On a successivement,

$$\begin{aligned}
V(Y) &= \frac{1}{n}\sum_i (y_i - \bar{y})^2 \\
&= \frac{1}{n}\sum_i (y_i - y_i' + y_i' - \bar{y})^2 \\
&= \frac{1}{n}\sum_i (y_i - y_i')^2 + \frac{1}{n}\sum_i (y_i' - \bar{y})^2 + \frac{2}{n}\sum_i (y_i - y_i')(y_i' - \bar{y}) \\
&= V_r + V_e + \frac{2}{n}\sum_i (y_i - y_i')(y_i' - \bar{y})
\end{aligned}$$

Or,

$$\begin{aligned}
\sum_i (y_i - y_i')(y_i' - \bar{y}) &= \sum_i y_i'(y_i - y_i') - \bar{y}\sum_i (y_i - y_i') \\
&= \sum_i (ax_i + b)\epsilon_i - \bar{y}\sum_i \epsilon_i \\
&= 0 \quad [car\ \frac{\partial \phi}{\partial a} = 0\ et\ \frac{\partial \phi}{\partial b} = 0]
\end{aligned}$$

On peut maintenant faire le lien entre le coefficient de corrélation r et "l'alignement du nuage".

On peut maintenant faire le lien entre le coefficient de corrélation r et "l'alignement du nuage". On a :

$$\begin{aligned}
\frac{V_r}{var(Y)} &= \frac{var(Y) - V_e}{var(Y)} \\
&= 1 - \frac{V_e}{var(Y)} \\
&= 1 - a^2 \frac{var(X)}{var(Y)} \\
&= 1 - \left(\frac{cov(X, Y)}{var(X)}\right)^2 \frac{var(X)}{var(Y)} \\
&= 1 - r^2
\end{aligned}$$

On peut maintenant faire le lien entre le coefficient de corrélation r et "l'alignement du nuage". On a :

$$\begin{aligned}
\frac{V_r}{var(Y)} &= \frac{var(Y) - V_e}{var(Y)} \\
&= 1 - \frac{V_e}{var(Y)} \\
&= 1 - a^2 \frac{var(X)}{var(Y)} \\
&= 1 - \left(\frac{cov(X,Y)}{var(X)}\right)^2 \frac{var(X)}{var(Y)} \\
&= 1 - r^2
\end{aligned}$$

Ainsi, si r^2 proche de 1, alors $Vr = \frac{1}{n} \sum_i \epsilon_i^2$ est petit !