# Modèles linéaires - Statistical Analysis System (SAS)
# Chapitre 1 :
# Moindres Carrés Ordinaires (MCO)
## Ordinary Least Squares Estimator (OLS)

Emmanuelle Gautherat[a]

(a) Crieg-Regards, Université de Reims Champagne Ardenne

Second semestre - 4 ECTS

# Outline

## Exemple de S.J. Sheather

Selon Wikipédia : *Un kicker est un joueur de football américain. Il est chargé de frapper les field goals (à trois points), les bottés de transformation (à un point) et les kickoffs. Certaines équipes National Football League (N.F.L.) possèdent des spécialistes séparés, l'un étant chargé des kickoffs, l'autre s'occupant des bottés de transformation et des field goals.*

Dans les colonnes du "Keeping Score" d'Aaron Schatz - édition du dimanche 12 novembre 2006 du New York Times - on trouve l'article intitulé "N.F.L. Kickers Are Judged on the Wrong Criteria" (Les tireurs de la N.F.L. sont jugés sur des critères erronés).

L'auteur affirme : *Il n'y a en effet aucune corrélation entre le pourcentage de réussite d'un botteur une saison et son pourcentage de réussite la saison suivante.*

# Exemple de S.J. Sheather

Selon Wikipédia : *Un kicker est un joueur de football américain. Il est chargé de frapper les field goals (à trois points), les bottés de transformation (à un point) et les kickoffs. Certaines équipes National Football League (N.F.L.) possèdent des spécialistes séparés, l'un étant chargé des kickoffs, l'autre s'occupant des bottés de transformation et des field goals.*

Dans les colonnes du "Keeping Score" d'Aaron Schatz - édition du dimanche 12 novembre 2006 du New York Times - on trouve l'article intitulé "N.F.L. Kickers Are Judged on the Wrong Criteria" (Les tireurs de la N.F.L. sont jugés sur des critères erronés).

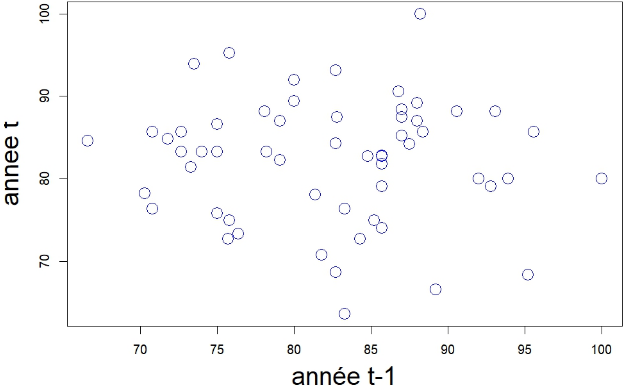
L'auteur affirme : *Il n'y a en effet aucune corrélation entre le pourcentage de réussite d'un botteur une saison et son pourcentage de réussite la saison suivante.*

On s'intéresse aux statistiques relatives aux 19 tireurs de field goal de la NFL qui ont effectué au moins dix tentatives de field goal au cours de chacune des saisons 2002, 2003, 2004 et 2005, ainsi qu'à la fin des matchs du dimanche 12 novembre de la saison 2006.

Les données se trouvent sur *https ://www.nfl.com/stats/player-stats/ ou sur https ://gattonweb.uky.edu>book>docs>datasets> FieldGoals2003to2006.csv*

# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



# Exemple de S.J. Sheather

La corrélation est négative, et vaut -0.0518, ce qui n'est pas significativement différent de 0 avec un test de $H_0$ : la corrélation est nulle, contre $H_1$ : la corrélation n'est pas nulle, sous hypothèse d'un couple gaussien.

Plus précisément :

```
cor.test(df_année[,1], df_année[,2], method="pearson")

Pearson's product-moment correlation

data:  df_année[, 1] and df_année[, 2]
t = -0.38486, df = 55, p-value = 0.7018
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3082306  0.2115999
sample estimates:
       cor
-0.05182524
```

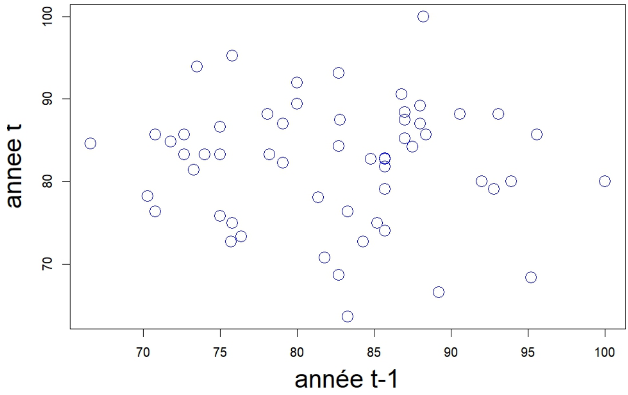
# Exemple de S.J. Sheather

| joueur | annee0 | annee1 |
|---|---|---|
| Adam Vinatieri | 73.5 | 93.9 |
| Adam Vinatieri | 93.9 | 80.0 |
| Adam Vinatieri | 80.0 | 89.4 |
| David Akers | 82.7 | 84.3 |
| David Akers | 84.3 | 72.7 |
| David Akers | 72.7 | 83.3 |
| Jason Elam | 87.0 | 85.2 |
| Jason Elam | 85.2 | 75.0 |
| Jason Elam | 75.0 | 86.6 |

Table – Extrait sur 2 années consécutives

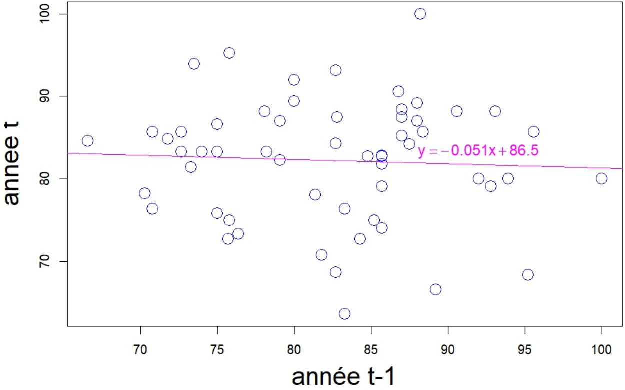
# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



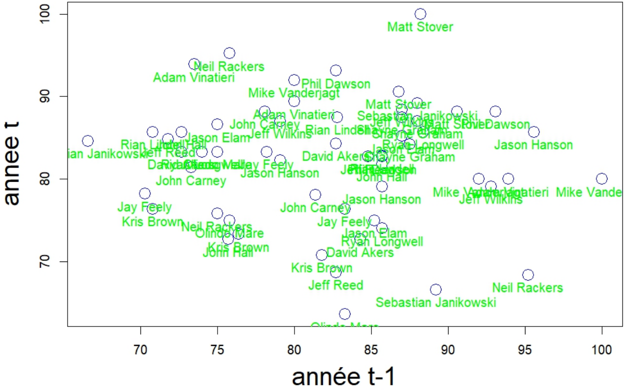
# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



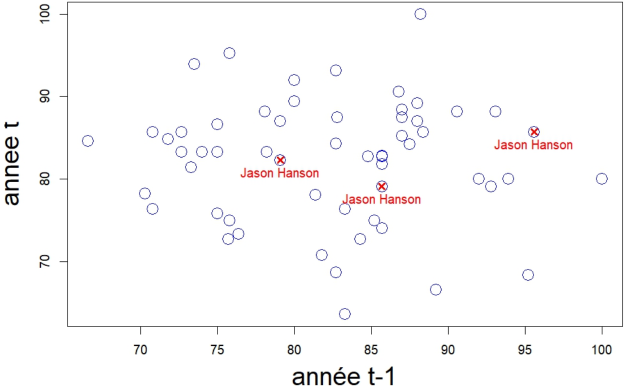
## Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



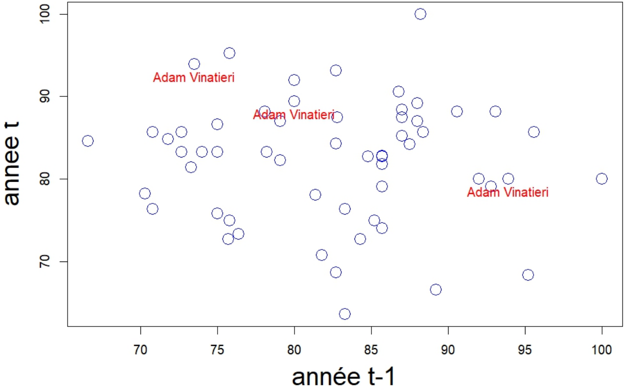
## Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



## Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



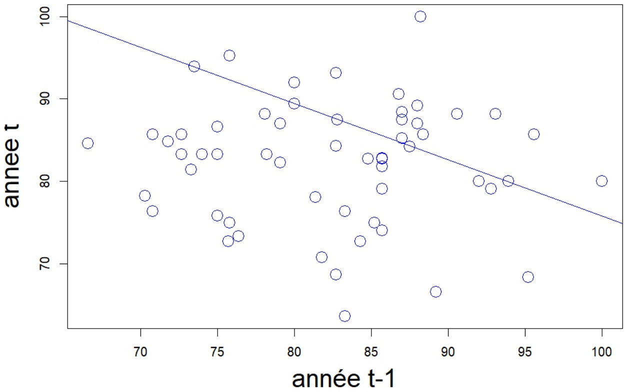
# Exemple de S.J. Sheather

Pour Adam Vinatieri :

```
Call:
lm(formula = annee1[indi] ~ annee0[indi])

Residuals:
       1        2        3
 0.03093  0.01446 -0.04539

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  143.890623   0.319530   450.3  0.00141 **
annee0[indi]  -0.680565   0.003854  -176.6  0.00361 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0568 on 1 degrees of freedom
Multiple R-squared:      1,Adjusted R-squared:  0.9999
F-statistic: 3.118e+04 on 1 and 1 DF,  p-value: 0.003605
```

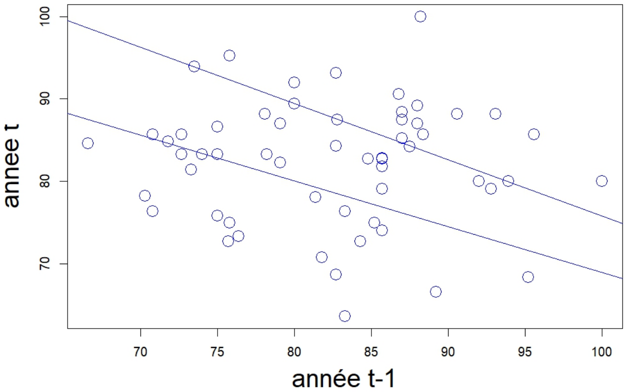
# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



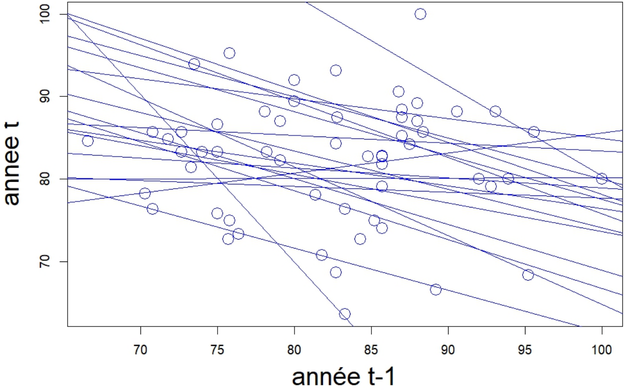
# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



# Exemple de S.J. Sheather

On représente le pourcentage de buts marqués par chaque tireur pour l'année $t$ par rapport à l'année $t-1$ pour 19 joueurs, soit 3 données par joueur.



# Modèle probabiliste

Modèle probabiliste = relation entre deux quantités aléatoires $\mathbb{X}$ et $\tilde{\mathbb{Y}}$ :

$$\tilde{\mathbb{Y}} = g(\mathbb{X}) + \tilde{\epsilon}.$$

On s'intéresse à la relation entre les espérances de ces quantités sachant $\mathbb{X}$ . les variables $\mathbb{Y} = \tilde{\mathbb{Y}}/\mathbb{X}$ et $\mathbb{X}$ sont observables :

$$\begin{aligned}
\mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= \mathbb{E}\left(g(\mathbb{X})/\mathbb{X}\right) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right) \\
\mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= g(\mathbb{X}) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right)
\end{aligned}$$

On suppose :

- $\mathbb{E}(\tilde{\epsilon}/\mathbb{X}) = 0$ et $var(\tilde{\epsilon}/\mathbb{X}) = \sigma^2$

# Modèle probabiliste

Modèle probabiliste = relation entre deux quantités aléatoires $\mathbb{X}$ et $\tilde{\mathbb{Y}}$ :

$$\tilde{\mathbb{Y}} = g(\mathbb{X}) + \tilde{\epsilon}.$$

On s'intéresse à la relation entre les espérances de ces quantités sachant $\mathbb{X}$ . les variables $\mathbb{Y} = \tilde{\mathbb{Y}}/\mathbb{X}$ et $\mathbb{X}$ sont observables :

$$\begin{aligned}
\mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= \mathbb{E}\left(g(\mathbb{X})/\mathbb{X}\right) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right) \\
\mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= g(\mathbb{X}) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right)
\end{aligned}$$

On suppose :

- $\mathbb{E}(\tilde{\epsilon}/\mathbb{X}) = 0$ et $var(\tilde{\epsilon}/\mathbb{X}) = \sigma^2$

Comme on considère le modèle sachant $\mathbb{X} = x$, cela revient à considèrer $\mathbb{X}$ comme déterministe, on le note donc $x$.

# Modèle probabiliste

Modèle probabiliste = relation entre deux quantités aléatoires $\mathbb{X}$ et $\tilde{\mathbb{Y}}$ :

$$\tilde{\mathbb{Y}} = g(\mathbb{X}) + \tilde{\epsilon}.$$

On s'intéresse à la relation entre les espérances de ces quantités sachant $\mathbb{X}$ . les variables $\mathbb{Y} = \tilde{\mathbb{Y}}/\mathbb{X}$ et $\mathbb{X}$ sont observables :

$$\begin{aligned} \mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= \mathbb{E}\left(g(\mathbb{X})/\mathbb{X}\right) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right) \\ \mathbb{E}\left(\tilde{\mathbb{Y}}/\mathbb{X}\right) &= g(\mathbb{X}) + \mathbb{E}\left(\tilde{\epsilon}/\mathbb{X}\right) \end{aligned}$$

On suppose :

- $\mathbb{E}(\tilde{\epsilon}/\mathbb{X}) = 0$ et $var(\tilde{\epsilon}/\mathbb{X}) = \sigma^2$

Comme on considère le modèle sachant $\mathbb{X} = x$, cela revient à considèrer $\mathbb{X}$ comme déterministe, on le note donc $x$. On note $\epsilon = \tilde{\epsilon}/\mathbb{X} = x$.

Modèle probabiliste

$$\mathbb{Y} = g(x) + \epsilon$$

avec $\mathbb{E}(\epsilon) = 0$ et $var(\epsilon) = \sigma^2$.

# Modèle statistique

On se restreint dans un premier temps au cas $x \in \mathbb{R}^{K+1}$ et $\mathbb{Y} \in \mathbb{R}$.

On observe $(x_i, y_i)_{i=,1\ldots,n}$ réalisations du couple $(x_i; Y_i)_{i=1,\ldots,n}$ avec $(Y_i)_{i=1,\ldots,n}$ un $n$-échantillon de même loi que $\mathbb{Y}$.

# Modèle statistique

On se restreint dans un premier temps au cas $x \in \mathbb{R}^{K+1}$ et $\mathbb{Y} \in \mathbb{R}$.

On observe $(x_i, y_i)_{i=,1\dots,n}$ réalisations du couple $(x_i; Y_i)_{i=1,\dots,n}$ avec $(Y_i)_{i=1,\dots,n}$ un $n$-échantillon de même loi que $\mathbb{Y}$.
On suppose que l'on observe le modèle statistique issu du modèle probabiliste observe

$$\forall i = 1, \dots, n, \quad Y_i = g(x_i) + \xi_i,$$

avec
Hypothèse $(\xi_i)_{i=1,\dots,n}$ $n$-échantillon centré, de variance $\sigma^2$ inconnue, de même loi que $\epsilon$.

## Modèle statistique

On se restreint dans un premier temps au cas $x \in \mathbb{R}^{K+1}$ et $\mathbb{Y} \in \mathbb{R}$.

On observe $(x_i, y_i)_{i=,1\dots,n}$ réalisations du couple $(x_i; Y_i)_{i=1,\dots,n}$ avec $(Y_i)_{i=1,\dots,n}$ un $n$-échantillon de même loi que $\mathbb{Y}$.
On suppose que l'on observe le modèle statistique issu du modèle probabiliste observe

$$\forall i = 1, \dots, n, \quad Y_i = g(x_i) + \xi_i,$$

avec
Hypothèse $(\xi_i)_{i=1,\dots,n}$ $n$-échantillon centré, de variance $\sigma^2$ inconnue, de même loi que $\epsilon$.

Modèle non-paramétrique : car $(g; \sigma^2; f)$ inconnu, où $f$ densité de $\xi_1$ et $(g; \sigma^2)$ "paramètre" d'intérêt avec $g \in \mathcal{G}$ espace de dimension infinie.

# Modèle statistique

On se restreint dans un premier temps au cas $x \in \mathbb{R}^{K+1}$ et $\mathbb{Y} \in \mathbb{R}$.

On observe $(x_i, y_i)_{i=,1\dots,n}$ réalisations du couple $(x_i; Y_i)_{i=1,\dots,n}$ avec $(Y_i)_{i=1,\dots,n}$ un $n$-échantillon de même loi que $\mathbb{Y}$.
On suppose que l'on observe le modèle statistique issu du modèle probabiliste observe

$$\forall i = 1, \dots, n, \quad Y_i = g(x_i) + \xi_i,$$

avec
Hypothèse $(\xi_i)_{i=1,\dots,n}$ $n$-échantillon centré, de variance $\sigma^2$ inconnue, de même loi que $\epsilon$.

Modèle non-paramétrique : car $(g; \sigma^2; f)$ inconnu, où $f$ densité de $\xi_1$ et $(g; \sigma^2)$ "paramètre" d'intérêt avec $g \in \mathcal{G}$ espace de dimension infinie.

Cadre plus restrictif des fonctions linéaires.
Modèle statistique
Pour $i = 1, \cdots, n$,

$$\begin{aligned} Y_i &= \beta_0 x_i^{(0)} + \beta_1 x_i^{(1)} + \cdots + \beta_K x_i^{(K)} + \xi_i \\ &= x_i \beta + \xi_i \end{aligned}$$

avec $\beta \in \mathbb{R}^{K+1}$, vecteur colonne et $x_i = (x_i^{(0)}, \dots, x_i^{(K)})$ vecteur ligne de taille $K + 1$.

Modèle statistique semi-paramétrique avec $(\beta; \sigma^2; f)$ inconnu où le paramètre d'intérêt est $\theta = (\beta; \sigma^2) \in$ à un e.v. de dimension finie.

# Hypothèses et expression matricielle du modèle linéaire

## Hypothèses

- $(H_0)$ $\beta$ est un vecteur constant dans la population étudiée
- $(H_1)$ Pour tout $i$, $\mathbb{E}(\xi_i) = 0$
- $(H_2)$ Homoscédasticité : $\forall i, \ \ var(\xi_i) = \sigma^2$
- $(H_3)$ $\forall i \neq j, \ \ \mathbb{E}(\xi_i \xi_j) = 0$

On note

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1^{(0)} & \cdots & x_1^{(K)} \\ x_2^{(0)} & \cdots & x_2^{(K)} \\ \vdots & \vdots & \vdots \\ x_n^{(0)} & \cdots & x_n^{(K)} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}.$$

Modèle statistique matriciel :

$$Y = x\beta + \xi.$$

- $(HM_0) = (H_0)$ $\beta$ ne dépend pas des individus $i$
- $(HM_1) = (H_1)$ $\mathbb{E}(Y) = x\beta$ et $\mathbb{E}(\xi) = 0_n$
- $(HM_2) = (H_2) + (H_3)$ $\Sigma_\xi = \sigma^2 Id_n$, où $Id_n$ désigne la matrice identité de taille $n$.

# Objectifs

On veut estimer ce modèle –> trouver $\widehat{\beta}$.

- Tester l'existence d'un effet d'une variable.
  On s'intéresse à l'effet produit par la variable $x^{(3)}$.
  –> test de $H_0$ : $\beta_3 = 0$ contre $H_1$ : $\beta_3 \neq 0$.
  Répond à la question : l'effet est-il statistiquement significatif pour cette variable, sans considérer les autres ?*(dans le cadre du modèle)*
- Quantifier l'effet.
  On s'intéresse à l'ordre de grandeur de l'effet produit par cette variable.
  –> répond à la question : à quelle hauteur est cet effet ?
- Prévoir.

## Les types de données

- Données temporelles ou longitudinales : $x_t, y_t$ indicées par le temps $t$. Le vecteur $x_t$ est un vecteur ligne, dont chaque coordonnée représente l'état sur chaque variable explicative au temps $t$ $\forall t \in \{1, \cdots, T\}, \quad x_t = (x_t^{(0)}, \cdots, x_t^{(K)})$.

# Les types de données

- Données temporelles ou longitudinales : $x_t, y_t$ indicées par le temps $t$. Le vecteur $x_t$ est un vecteur ligne, dont chaque coordonnée représente l'état sur chaque variable explicative au temps $t$ $\forall t \in \{1, \cdots, T\}, \quad x_t = (x_t^{(0)}, \cdots, x_t^{(K)})$.
- Données en coupe : $x_i, y_i$ indicées par les individus de la population étudiée. Le vecteur $x_i$ est un vecteur ligne, dont chaque coordonnée représente l'état sur chaque variable explicative pour l'individu $i$

$$\forall i =\in \{1, \cdots, n\} \quad x_i = (x_i^{(0)}, \cdots, x_i^{(K)}).$$

# Les types de données

- Données temporelles ou longitudinales : $x_t, y_t$ indicées par le temps $t$. Le vecteur $x_t$ est un vecteur ligne, dont chaque coordonnée représente l'état sur chaque variable explicative au temps $t$ $\forall t \in \{1, \cdots, T\}, \quad x_t = (x_t^{(0)}, \cdots, x_t^{(K)})$.
- Données en coupe : $x_i, y_i$ indicées par les individus de la population étudiée. Le vecteur $x_i$ est un vecteur ligne, dont chaque coordonnée représente l'état sur chaque variable explicative pour l'individu $i$

$$\forall i = \in \{1, \cdots, n\} \quad x_i = (x_i^{(0)}, \cdots, x_i^{(K)}).$$

- Données de panel : $x_{i,t}, y_{i,t}$. On suit les individus $i$ au cours du temps $t$. $x$ est une matrice de taille $nT \times (K+1)$. Chaque bloc de lignes de longueur $n$, représente les individus décrits sur toutes les $K+1$ variables explicatives, on reproduit ces blocs pour chacune des périodes $T$.

# Exemples de Modèles linéaires, données en coupe

Linéarisation éventuelle en $\beta$.

Soient $z_i \in \mathbb{R}$, $(a, b, c, d) \in \mathbb{R}^4$

- Linéaire : $y = az + b$. On pose pour tout $i$, $x_i = (1, z_i)$ et $\beta = (a, b)^t$. On obtient le modèle de régression linéaire

$$\forall i = 1, \cdots, n, \quad Y_i = a + bz_i + \xi_i.$$

- Parabolique : $y = ax^2 + bx + c(\longrightarrow y = a\,(x - x_0)^2 + y_0)$
- Polynomial : $y = ax^3 + bx^2 + c$ Soient $x_i = (1, z_i^2, z_i^3)$ et $\beta = (c, b, a)^t$. On obtient un modèle de régression d'ordre 3

$$\forall i = 1, \cdots, n, \quad Y_i = c + bz_i^2 + cz_i^3 + \xi_i.$$

- Log-linéaire : $y = bx^a \longrightarrow \ln y = a \ln x + \ln b$
- Exponentiel : $y = \exp(ax + b) - > \ln y = ax + b$
- Logarithmique : $y = a \ln x + b$
- Hyperbolique : $y = \frac{a}{x - x_0} + y_0$

# Graphes : exemple du modèle linéaire simple

Soient $x_i^{(1)} \in \mathbb{R}$. On suppose que $x^{(1)} \perp 1_n$. Soit $(a, b) \in \mathbb{R}^2$. On note $x = (1, x^{(1)})$ et $\beta = (a, b)^t$.

On suppose que

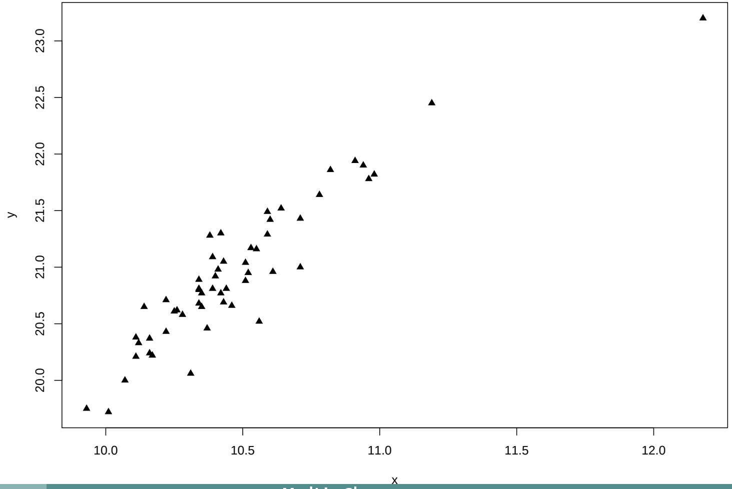
$$\forall i = 1, \cdots, n, \quad Y_i = a + bx_i^{(1)} + \xi_i,$$

avec

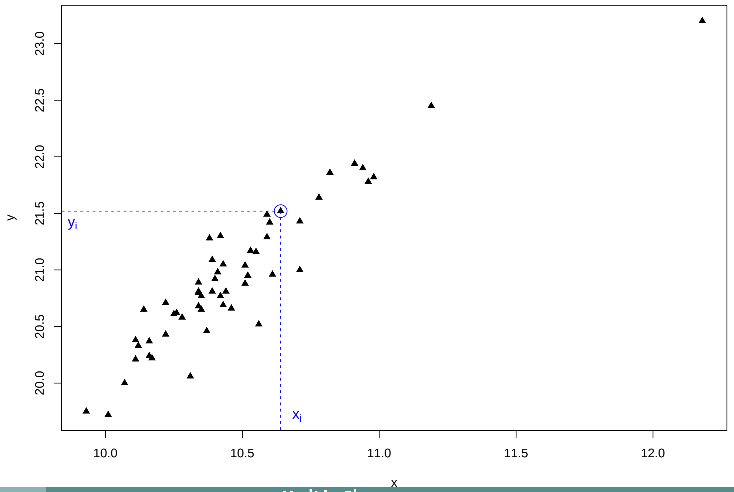
$$\mathbb{E}(\xi_i) = 0.$$

Par abus d'écriture, on notera sur les graphes qui suivent $x$ à la place de $x^1$.
On notera également $1_n$ le vecteur constant de taille $n$ dont chaque coordonnée vaut 1.

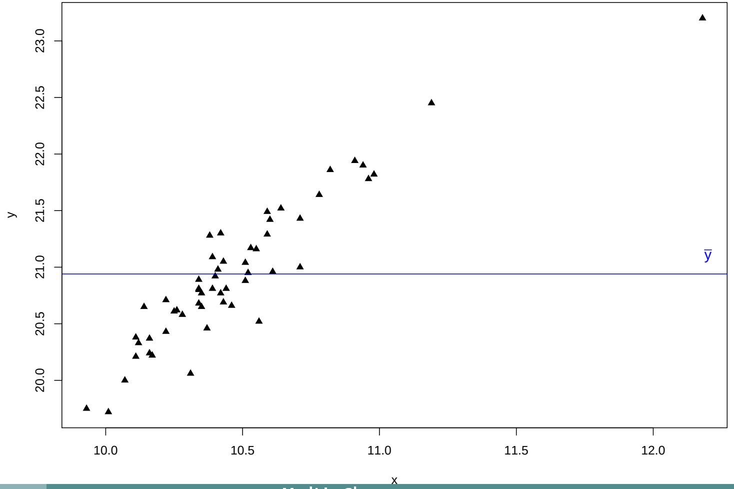
# Espace des variables : quelle droite d'ajustement ?



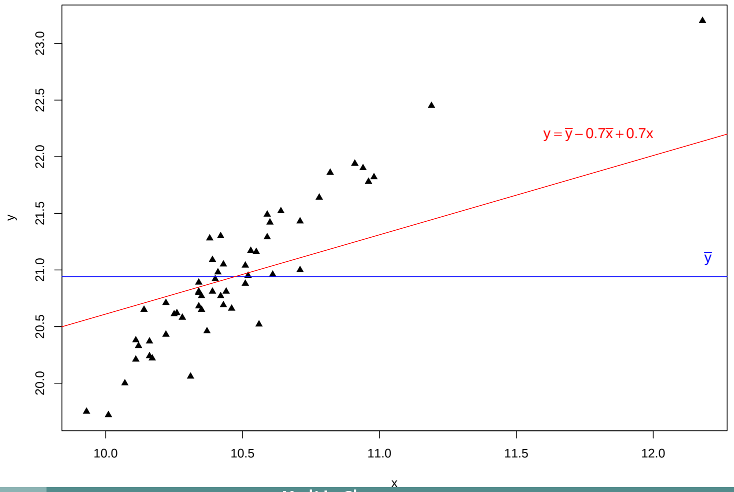
## Espace des variables : quelle droite d'ajustement ?



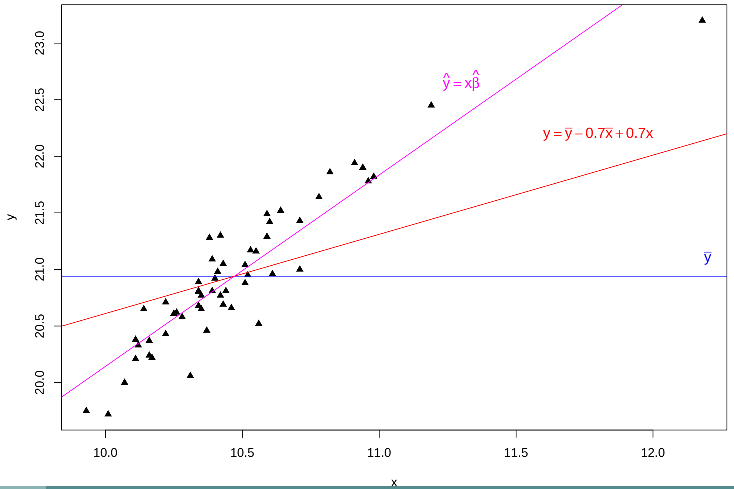
# Espace des variables : quelle droite d'ajustement ?



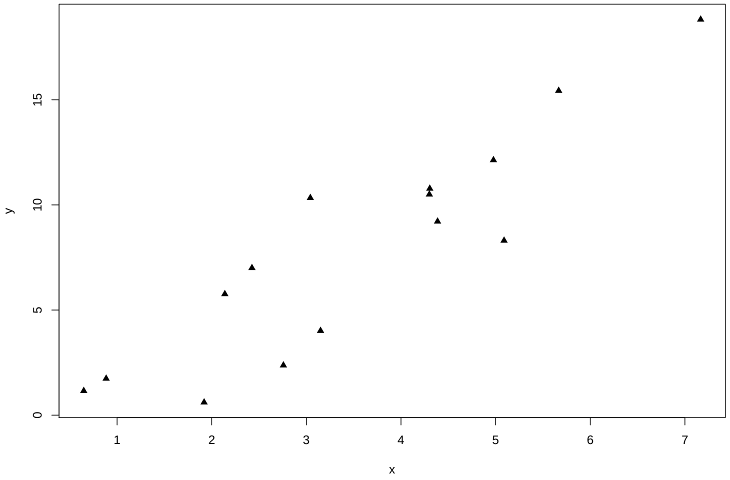
# Espace des variables : quelle droite d'ajustement ?



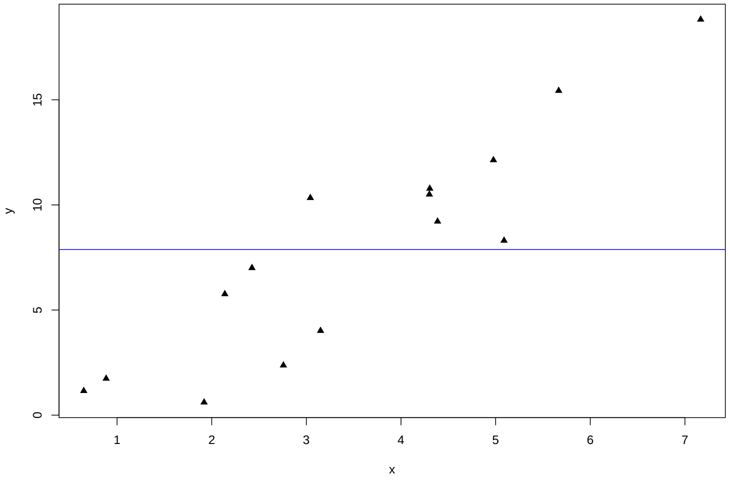
## Espace des variables : quelle droite d'ajustement ?



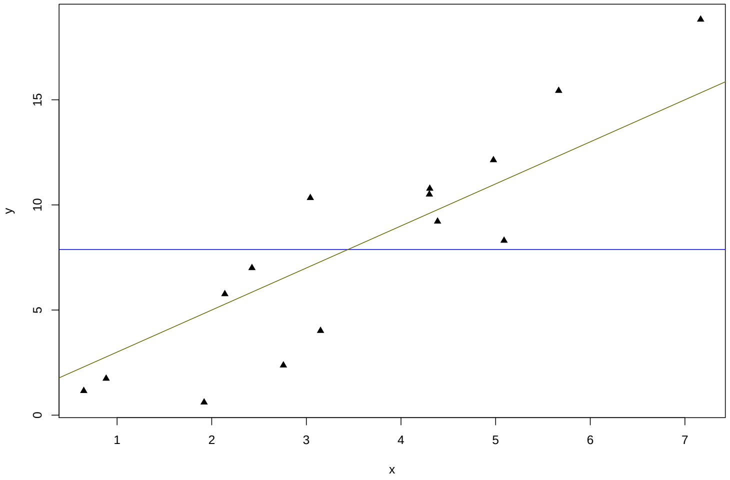
# Espace des variables : modèle et estimation



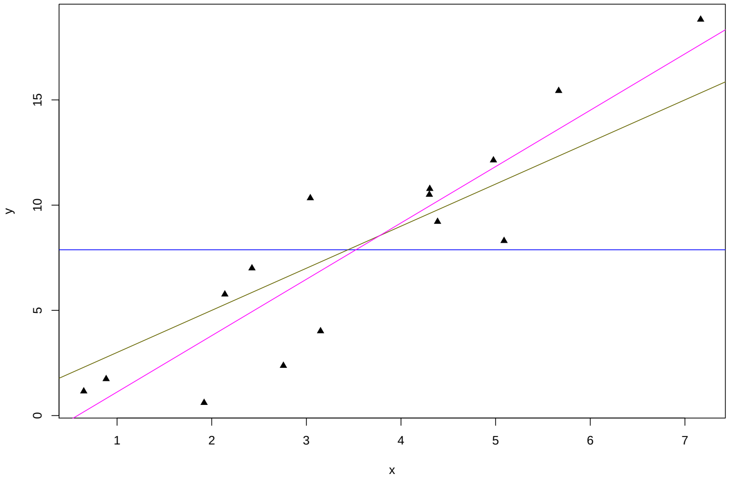
# Espace des variables : modèle et estimation



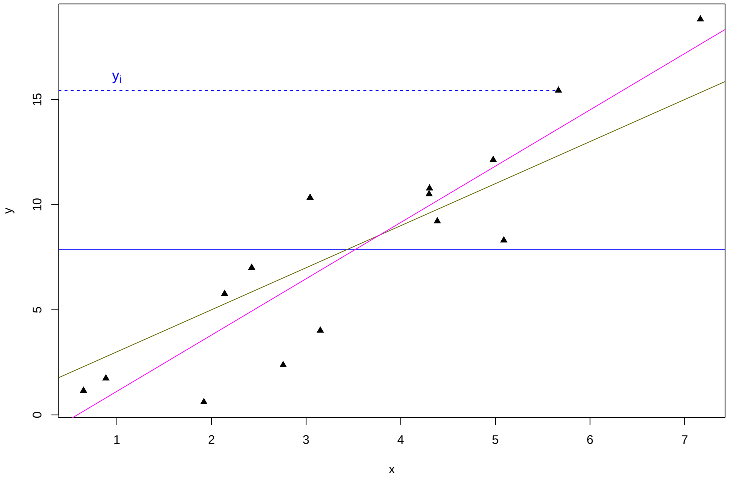
# Espace des variables : modèle et estimation



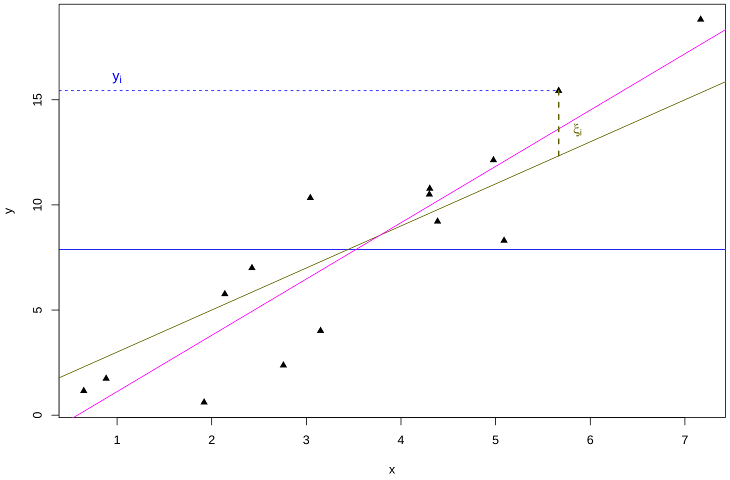
# Espace des variables : modèle et estimation



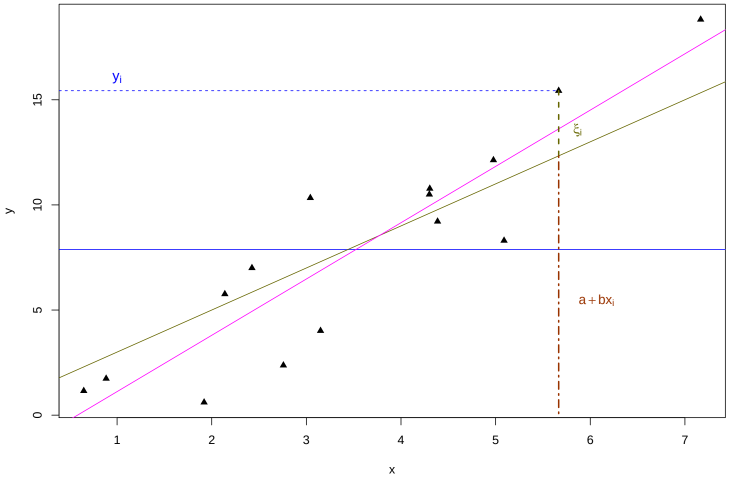
# Espace des variables : modèle et estimation



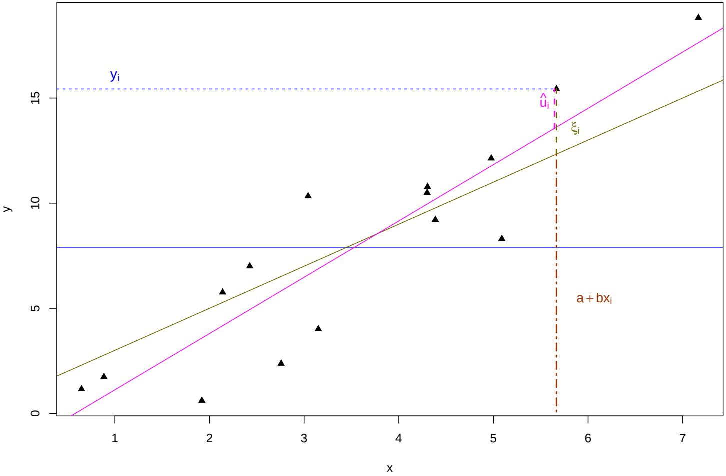
# Espace des variables : modèle et estimation



## Espace des variables : modèle et estimation

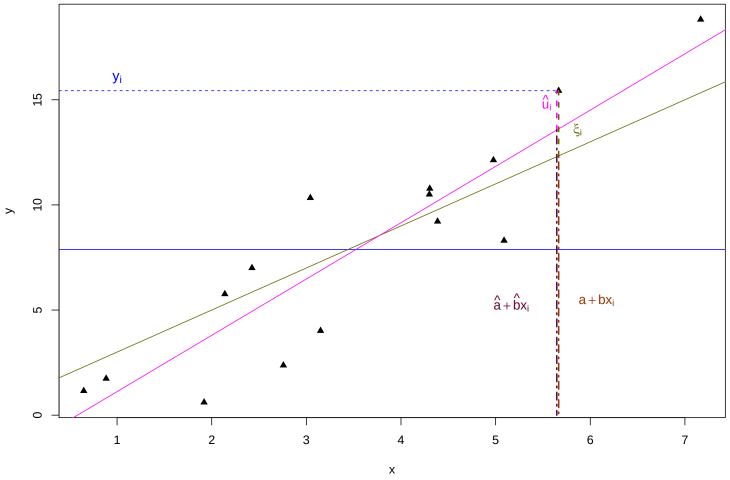


# Espace des variables : modèle et estimation

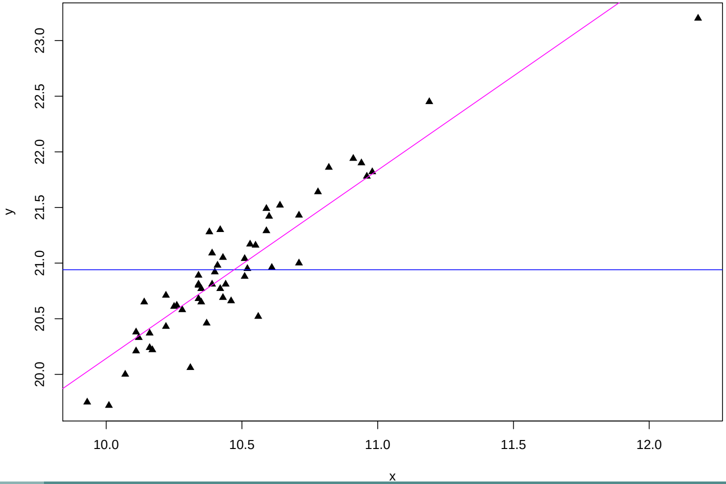


ModLin Chap 1
E. Gautherat
Exemples
Introduction formelle
Contexte
Démarche
$\widehat{\beta}$
Définition
Mat.proj.ortho
Déf $\widehat{\beta}$
Propriétés
Résidus
$R^2$
Premier pas avec R
annexes
Exo. math proj
Vect. gaussien

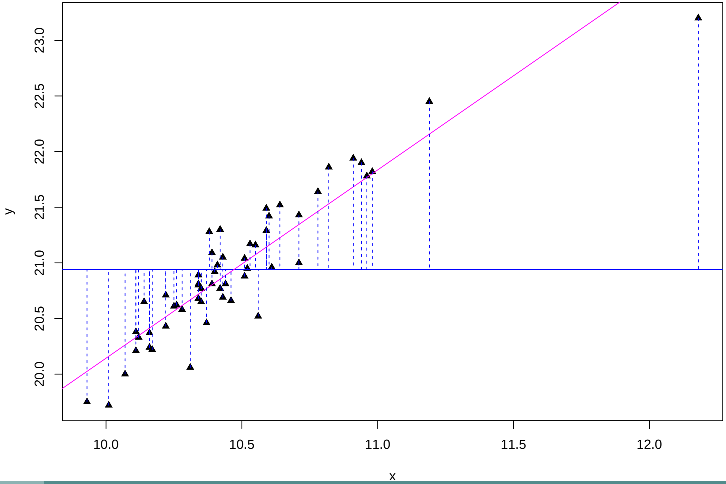
# Espace des variables : modèle et estimation



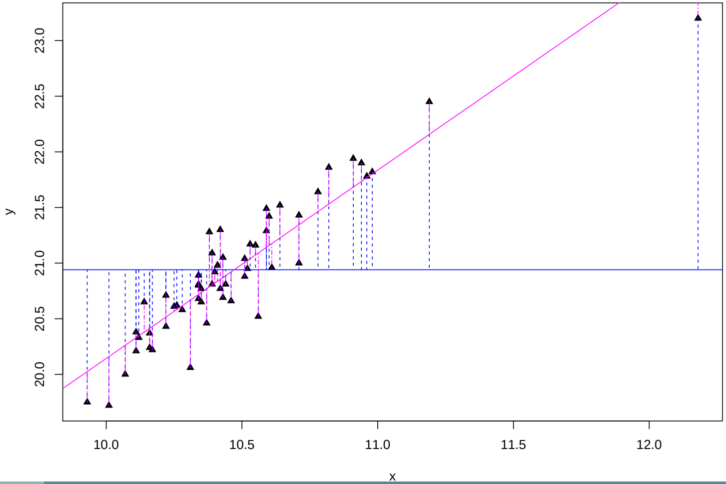
# Décomposition de la variance : graphe régression linéaire simple



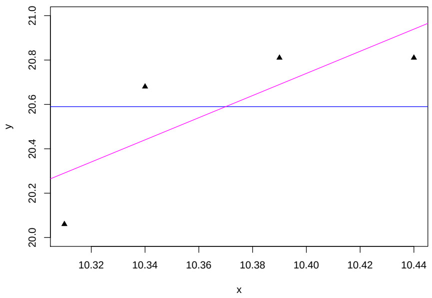
# Décomposition de la variance : graphe régression linéaire simple



# Décomposition de la variance : graphe régression linéaire simple



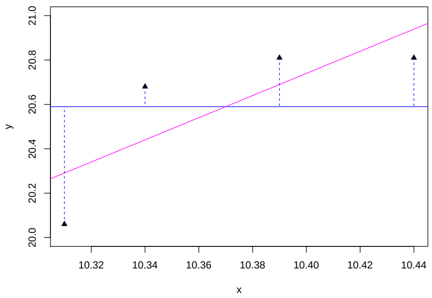
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad Y_i$$

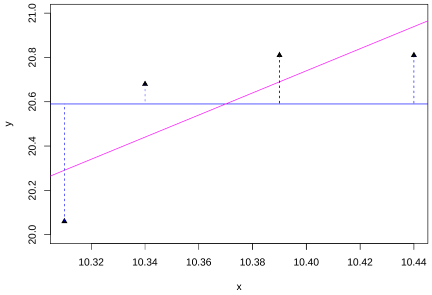
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$\forall i = 1, \dots, n, \qquad Y_i - \overline{Y}$$

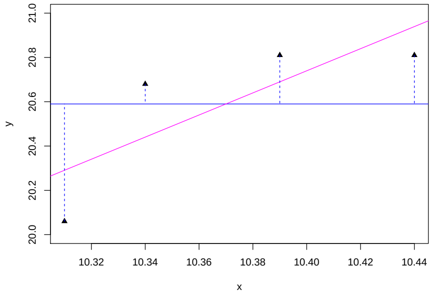
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad |Y_i - \overline{Y}|$$

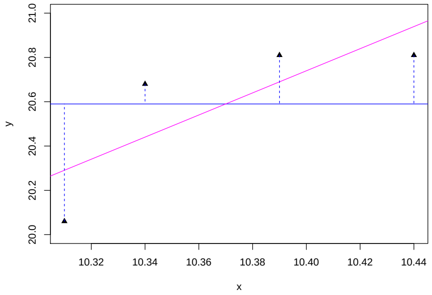
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad ( Y_i - \overline{Y} )^2$$

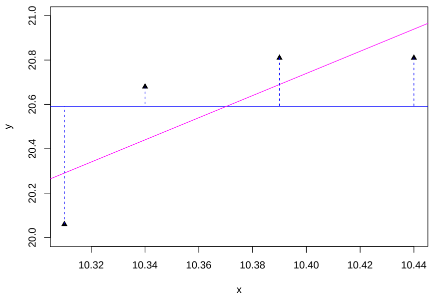
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$\sum_{i=1}^{n} (\ Y_i - \overline{Y}\ )^2$$

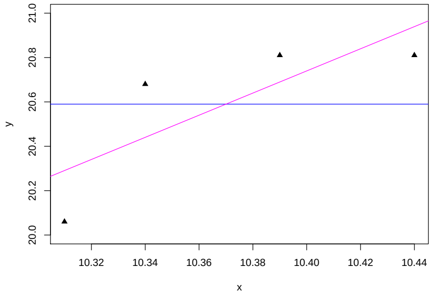
# Somme des carrés totale




Une mesure de la dispersion par rapport à la moyenne est :

$$SCT = \sum_{i=1}^{n} (\; Y_i - \overline{Y} \;)^2$$

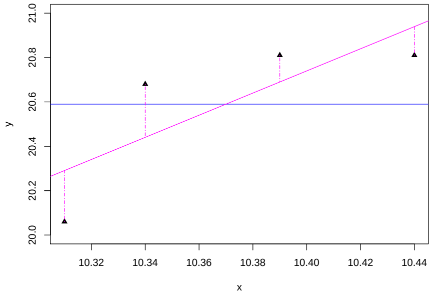
## Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$\forall i = 1, \ldots, n, \qquad Y_i$$

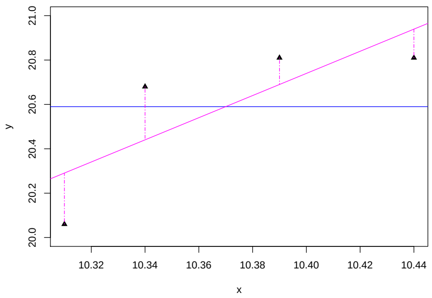
## Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$\forall i = 1, \ldots, n, \qquad Y_i - \widehat{Y}_i$$

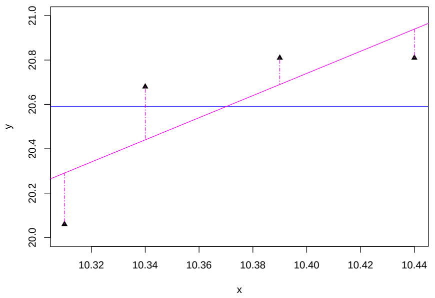
# Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$\forall i = 1, \ldots, n, \qquad |Y_i - \widehat{Y}_i|$$

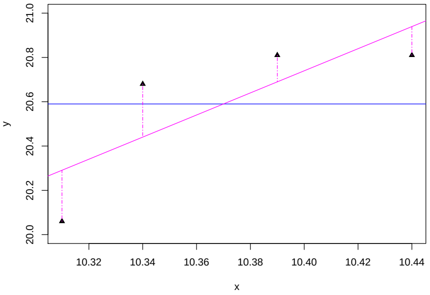
# Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$\forall i = 1, \dots, n, \qquad ( Y_i - \widehat{Y}_i )^2$$

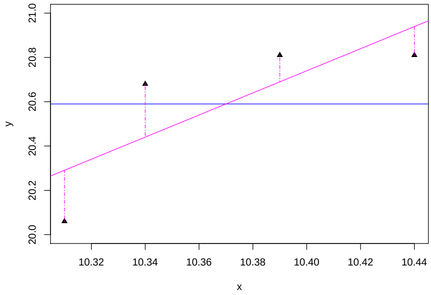
# Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$\sum_{i=1}^{n} (\ Y_i - \widehat{Y}_i\ )^2$$

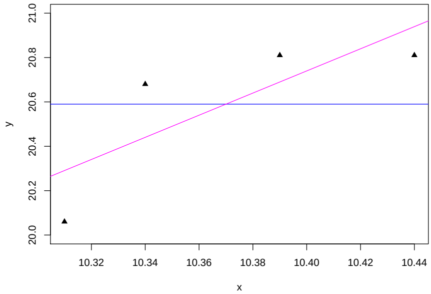
# Somme des carrés des résidus




Une mesure de la dispersion par rapport à la droite de régression est :

$$SCR = \sum_{i=1}^{n} (\ Y_i - \widehat{Y}_i\ )^2$$

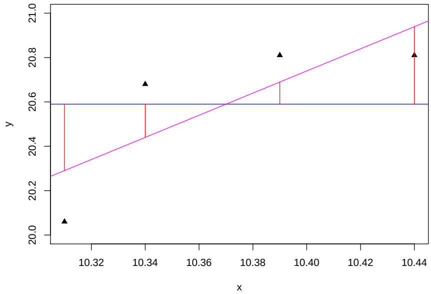
# Somme des carrés expliqués par le modèle




Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad \widehat{Y}_i$$

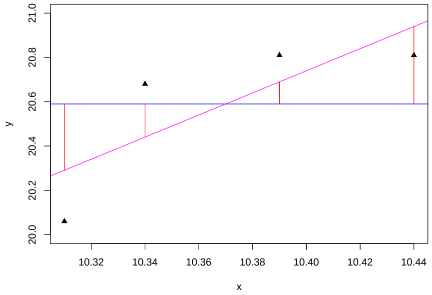
# Somme des carrés expliqués par le modèle




Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad \widehat{Y}_i - \overline{Y}$$

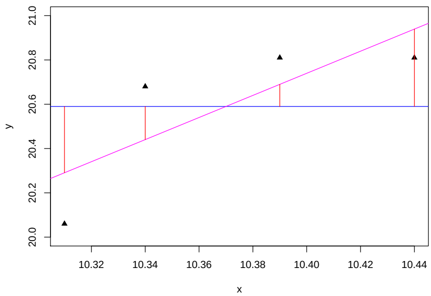
# Somme des carrés expliqués par le modèle




Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad |\widehat{Y}_i - \overline{Y}|$$

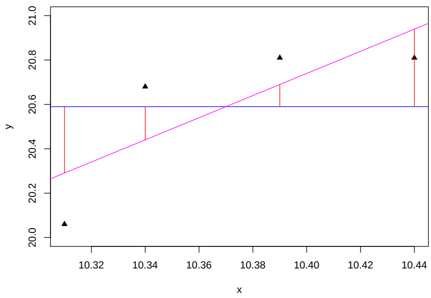
# Somme des carrés expliqués par le modèle




Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$\forall i = 1, \ldots, n, \qquad ( \widehat{Y}_i - \overline{Y} )^2$$

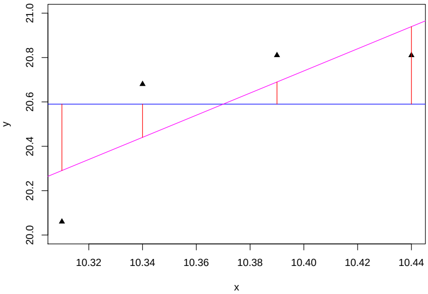
# Somme des carrés expliqués par le modèle




Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$\sum_{i=1}^{n} ( \widehat{Y}_i - \overline{Y} )^2$$

# Somme des carrés expliqués par le modèle



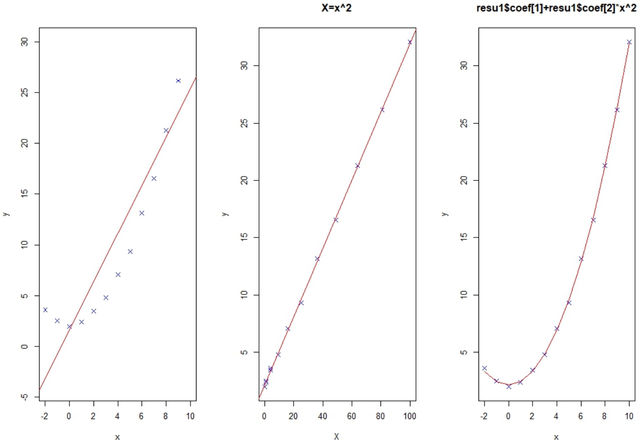

Une mesure de la dispersion du modèle par rapport à la moyenne est :

$$SCE = \sum_{i=1}^{n} (\ \widehat{Y}_i - \overline{Y}\ )^2$$

## Graphes

```
x= -2:10  ; psi= rnorm(length(x), sd=0.2)
beta0=2  ; beta1=0.3

###  modèle 1  ######
y= beta0+beta1*x+psi

###  visualisation modèle 1
plot(x, y, pch=4, col="blue") ;  resu=lm(y~x) ; abline(reg=resu, col="red")

###  modèle 2  ######
y= beta0+beta1*x^2+psi

###  visualisation modèle 2
par(mfrow=c(1,3))
  resu=lm(y~x) ; plot(x, y, pch=4, col="blue", ylim=c(-4,30))
  abline(reg=resu, col="red")

  X=x^2 ; resu1=lm(y~X)
  plot(X, y, pch=4, col="blue", main="X=x^2")
  abline(reg=resu1, col="red")

  plot(x, y, pch=4, col="blue", main="resu1$coef[1]
  +resu1$coef[2]*x^2")
  lines(x, resu1$coef[1]+resu1$coef[2]*x^2, col="red" )
```

# modèle non adapté



# Estimateur de $\widehat{\beta}$

On cherche donc $\widehat{\beta}$, l'estimateur de $\beta$, tel que

$$
\begin{aligned}
\widehat{\beta} &= \underset{\beta \in \mathbb{R}^2}{argmin} \left\{ \sum_{i=1}^{n} (Y_i - x_i\beta)^2 \right\} \\
&= \underset{(a,b) \in \mathbb{R}^2}{argmin} \left\{ \sum_{i=1}^{n} \left(Y_i - (a + bx_i^1)\right)^2 \right\} \\
&= \underset{\beta \in \mathbb{R}^2}{argmin} \left\{ \|Y - x\beta\|_2^2 \right\}
\end{aligned}
$$

# Estimateur de $\widehat{\beta}$

On cherche donc $\widehat{\beta}$, l'estimateur de $\beta$, tel que

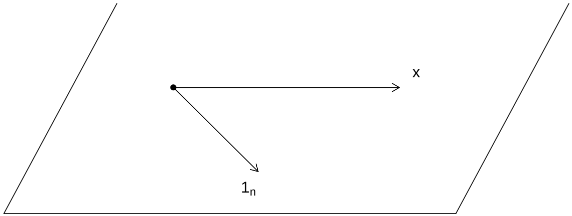
$$\begin{aligned}
\widehat{\beta} &= \underset{\beta \in \mathbb{R}^2}{argmin} \left\{ \sum_{i=1}^{n} (Y_i - x_i\beta)^2 \right\} \\
&= \underset{(a,b) \in \mathbb{R}^2}{argmin} \left\{ \sum_{i=1}^{n} \left(Y_i - (a + bx_i^1)\right)^2 \right\} \\
&= \underset{\beta \in \mathbb{R}^2}{argmin} \left\{ \|Y - x\beta\|_2^2 \right\}
\end{aligned}$$

$\widehat{\beta}$ revient à trouver $\beta$ minimisant la SCR en fonction de $\widehat{\beta}$

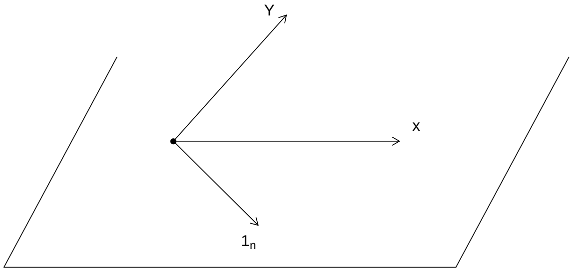
$$\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2.$$

# Qui est $\widehat{\beta}$ ? Espace des individus

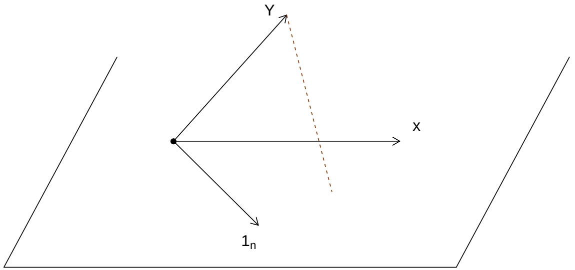
On note $1_n = (1, 1, \ldots, 1)^T$.



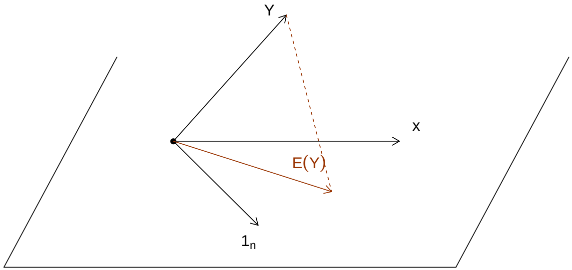
# Qui est $\widehat{\beta}$ ? Espace des individus



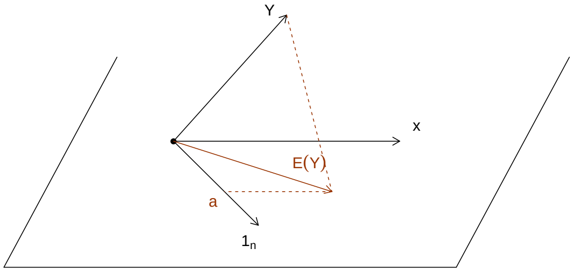
# Qui est $\widehat{\beta}$ ? Espace des individus



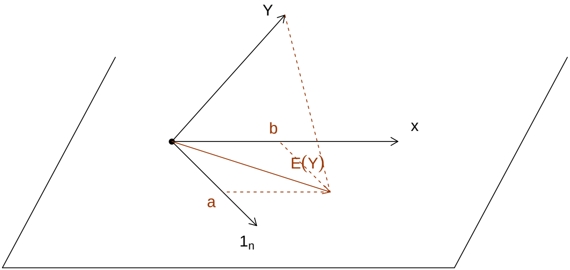
# Qui est $\widehat{\beta}$ ? Espace des individus



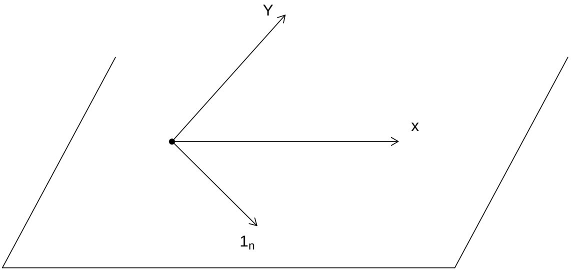
# Qui est $\widehat{\beta}$ ? Espace des individus



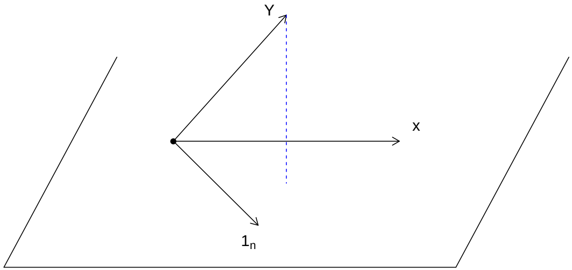
# Qui est $\widehat{\beta}$ ? Espace des individus



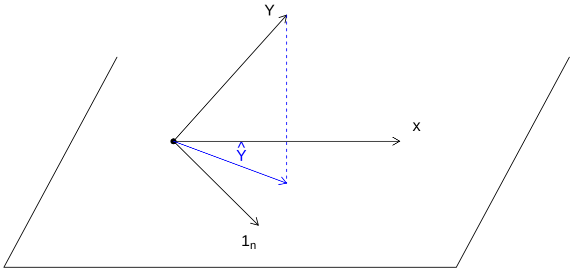
# Qui est $\widehat{\beta}$ ? Espace des individus



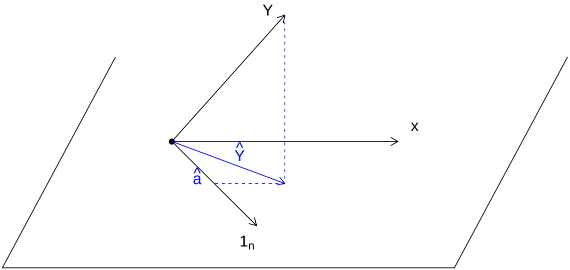
# Qui est $\widehat{\beta}$ ? Espace des individus



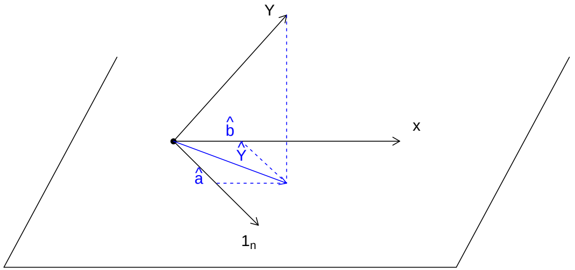
# Qui est $\widehat{\beta}$ ? Espace des individus



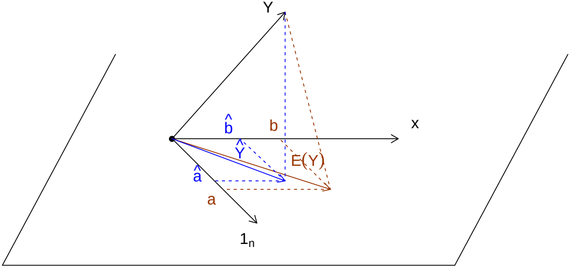
# Qui est $\widehat{\beta}$ ? Espace des individus



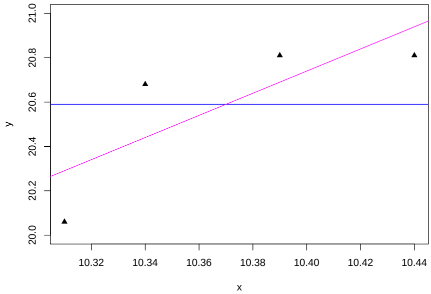
# Qui est $\widehat{\beta}$ ? Espace des individus



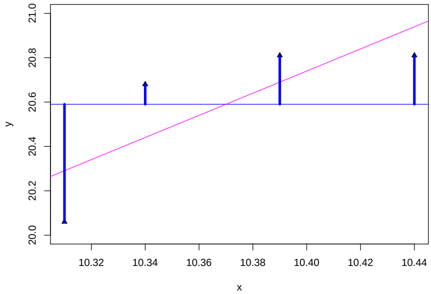
# Qui est $\widehat{\beta}$ ? Espace des individus



# Espace des variables : Décomposition de la variance

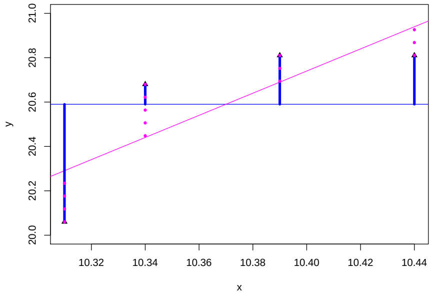


# Espace des variables : Décomposition de la variance



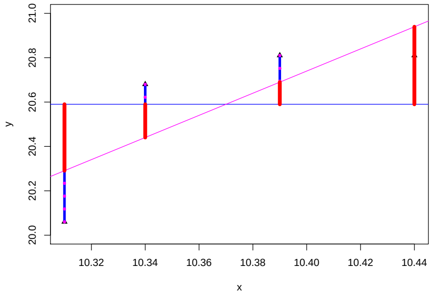

Preuve en image

# Espace des variables : Décomposition de la variance




Preuve en image

# Espace des variables : Décomposition de la variance




On décompose SCT selon SCE et SCR avec :

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \quad = \quad \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$$

Preuve en image

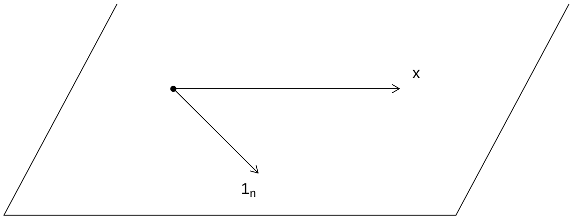
# Espace des variables : Décomposition de la variance

On décompose SCT selon SCE et SCR avec :

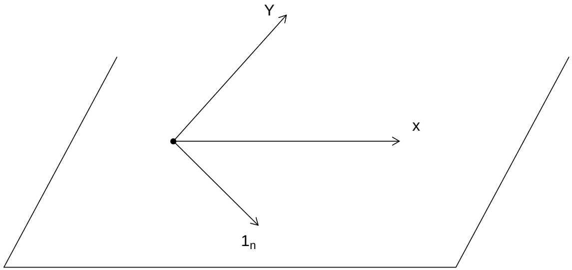
$$\begin{aligned} \sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 \\ SCT &= SCR + SCE \end{aligned}$$

Preuve en image

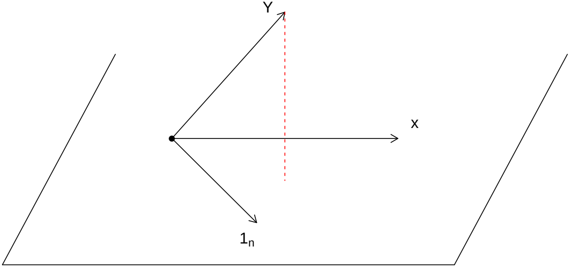
# Espace des individus : Décomposition de la variance



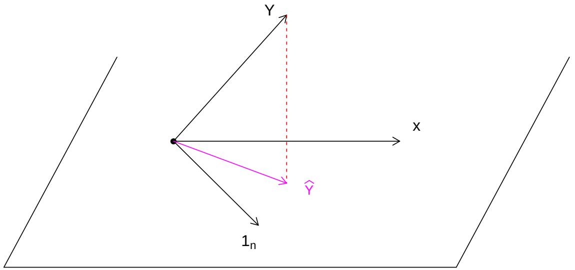
# Espace des individus : Décomposition de la variance



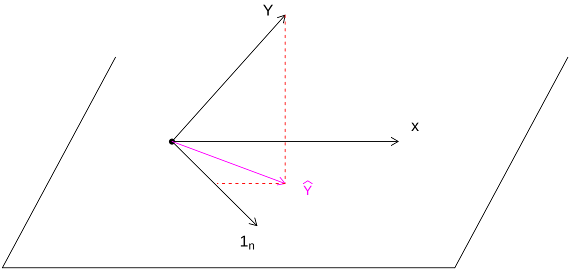
# Espace des individus : Décomposition de la variance



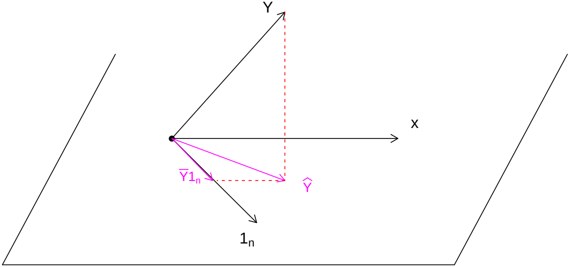
# Espace des individus : Décomposition de la variance



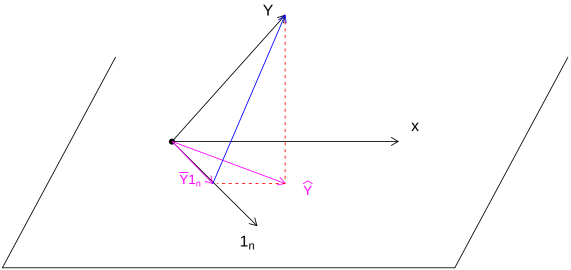
# Espace des individus : Décomposition de la variance



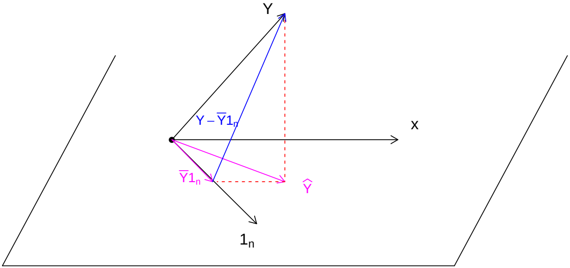
# Espace des individus : Décomposition de la variance



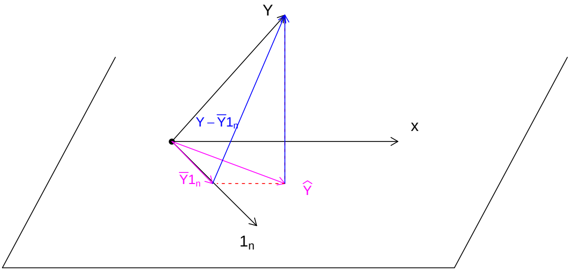
# Espace des individus : Décomposition de la variance



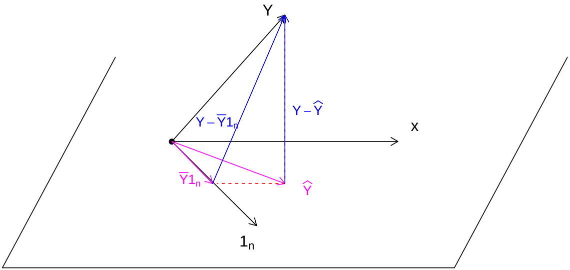
# Espace des individus : Décomposition de la variance



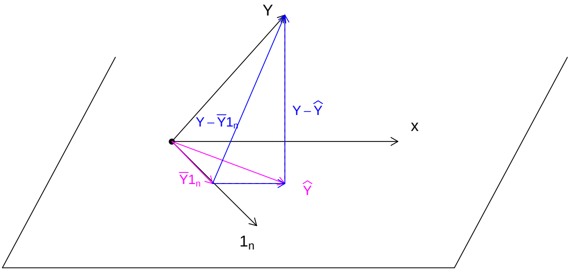
# Espace des individus : Décomposition de la variance



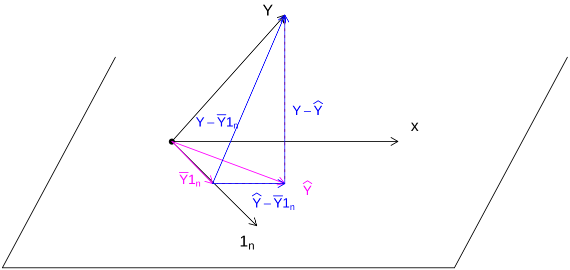
# Espace des individus : Décomposition de la variance



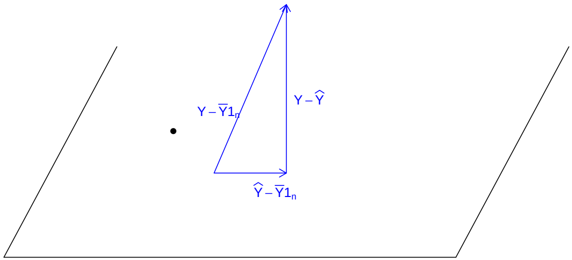
# Espace des individus : Décomposition de la variance



# Espace des individus : Décomposition de la variance

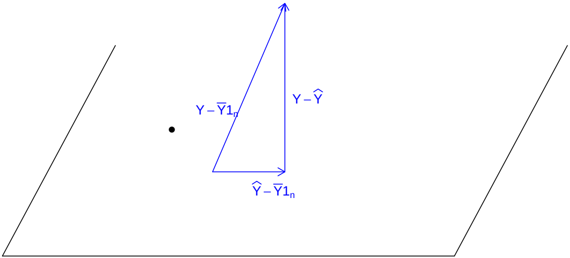


# Espace des individus : Décomposition de la variance



## Espace des individus : Décomposition de la variance

On achève la preuve avec Pythagore.



# Orthogonalité

Soient $v$ et $w$, deux vecteurs appartenant à un e.v. E de dimension $n$, muni du produit scalaire $< . ; . >$, associé à la norme euclidienne $\|.\|_2$.

On confond la notation $v$ et $w$ avec leur représentation matricielle et on note $v = (v_1, \cdots, v_n)^T$ et $w = (w_1, \ldots, w_n)^T$ où $v_i, w_j$ sont des scalaires.

On rappelle alors que :

- $< v ; w >= \sum_{i=1}^{n} v_i w_i$
- $< v ; v >= \|v\|_2^2 = \sum_{i=1}^{n} v_i^2$
- $< v ; w >= v^t w = trace(vw^t)$
- $< v ; w >=< w ; v >$
- $< v ; w >= 0 \iff v \perp w$
- Normalisation du vecteur $v$ : $x = v/\|v\|_2$. On a alors $\|x\|_2 = 1$.
  A noter : Un e.v. engendré par $v_1, \ldots, v_p$ est identique à celui engendré par leur version normalisée.

# Matrice de projection orthogonale

Soit E un e.v. muni d'un produit scalaire $<\ ;\ >$, et W un sous-e.v. (s.e.v) de E
–> on désire projeter orthogonalement un élément $b$ de E sur W. C'est-à-dire trouver l'élément $\tilde{b}$ de $W$ tel que $\tilde{b} = \underset{a \in \mathrm{W}}{\operatorname{argmin}} \|b - a\|_2$.

–> Les mathématiciens ont montré que $\tilde{b}$ est le résultat d'une simple multiplication de $b$ par une matrice.
–> On note $\Pi_{\mathrm{W}}$ cette matrice appelée matrice de projection orthogonale.
D'une manière générale, on notera $\Pi_{\mathrm{W}}$ la matrice de projection orthogonale de E sur W. Ainsi

$$\Pi_{\mathrm{W}} b = \underset{a \in \mathrm{W}}{\operatorname{argmin}} \|b - a\|_2.$$

Exemple :
Soit $\mathrm{E} = \mathbb{R}^3$ muni de sa base canonique : $e_1 = (1,0,0)^T$, $e_2 = (0,1,0)^T$, $e_3 = (0,0,1)^T$.
Soit $\mathrm{W} = vect(e_1, e_2)$, qui désigne le s.e.v. engendré par $e_1$ et $e_2$
La matrice de projection orthogonale $\Pi_{\mathrm{W}}$ de E sur W est

$$\Pi_{\mathrm{W}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(ceci est un cas très simple, ne se construit pas toujours ainsi)

# Matrice de projection orthogonale

Suite exemple :

- Lorsque $b = e_1$, il se projette sur lui même :

$$\begin{aligned} \Pi_{\mathsf{W}} b &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = b. \end{aligned}$$

- Lorsque $b = e_3$, il se projette sur le vecteur nul $(0,0,0)^T$ :

$$\begin{aligned} \Pi_{\mathsf{W}} b &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0_{\mathsf{E}} \end{aligned}$$

# Construction ?

Cadre et notations : Soient $v_1, \ldots, v_p$, $p$ vecteurs de $\mathbb{R}^n$ avec $p \leq n$.
On note W $= vect(v_1, \ldots, v_p)$ ; $V$ la représentation matricielle de $W$ : $V = (v_1, \ldots, v_p)$ ;
et $\Pi_{\mathsf{W}}$, la matrice de projection de $\mathbb{R}^n$ sur W,

Hypothèse $(H_1)$ : $v_1, \ldots, v_p$, forment une base orthogonale de $\mathbb{R}^p$.
On note W $= vect(v_1, \ldots, v_p)$ ; et $V$ la représentation matricielle de $W$ : $V = (v_1, \ldots, v_p)$.
Alors, $\Pi_{\mathsf{W}}$, la matrice de projection de $\mathbb{R}^n$ sur W,

## Propriété : construction de matrice de projection orthogonale

Sous $(H_1)$, $\Pi_{\mathsf{W}} = V(V^tV)^{-1}V^t$.

*Preuve admise.*

## Corollaire

On suppose $(H_1)$, et de plus, on suppose que $(v_1, \cdots, v_p)$ forme une base ortho<u>normale</u> de W, on a alors

$$\Pi_{\mathsf{W}} = VV^t.$$

*Preuve en exercice.*

# Matrice de projection orthogonale

On dit que deux s.e.v. F et G de E sont supplémentaires ssi

1. $F \cap G = \{0_E\}$
2. $F + G = E$

Autrement dit, tout élément $b$ de E se décompose de manière unique en $b_1 \in$ F et $b_2 \in$ G. On écrit $b = b_1 + b_2$.

Exemple (à connaitre) : Soit E muni d'un produit scalaire et F un s.e.v. de E. Alors G $=$ F$^\perp$ est un s.e.v. supplémentaire de F dans E.

Equation aux dimensions :
Et $dim(\mathsf{E}) = n$ et $dim(\mathsf{F}) = p$ entraine $dim(\mathsf{F}^\perp) = n - p$.

# Matrice de projection orthogonale

Soient W et G deux s.e.v. quelconques de E. On note $I_E$ la matrice identité de $E$.

## Propriétés

$$\Pi_{\mathsf{W}}^t = \Pi_{\mathsf{W}} \qquad \Pi_{\mathsf{W}}^2 = \Pi_{\mathsf{W}}$$
$$trace(\Pi_{\mathsf{W}}) = dim(\mathsf{W}) \qquad \Pi_{\mathsf{W}^\perp} = (I_{\mathsf{E}} - \Pi_{\mathsf{W}}).$$

Pour tout $b$ élément de E.

$$\Pi_{\mathsf{W}^\perp} b = b - \Pi_{\mathsf{W}} b.$$

$$\Pi_{\mathsf{W}}\Pi_{\mathsf{G}} = 0 \iff \mathsf{W} \perp \mathsf{G}$$

Si W $\perp$ G, alors :

$$\Pi_{\mathsf{W}+\mathsf{G}} = \Pi_{\mathsf{W}} + \Pi_{\mathsf{G}}.$$

## Théorème de Frisch-Waugh

$$\Pi_{\mathsf{W}+\mathsf{G}} = \Pi_{\mathsf{W}} + \Pi_{(I_{\mathsf{E}} - \Pi_{\mathsf{W}}).\mathsf{G}}$$

Exercices en annexe 1

ModLin Chap 1
E. Gautherat
Exemples
Introduction formelle
Contexte
Démarche
$\widehat{\beta}$
Définition
Mat.proj.ortho
Déf $\widehat{\beta}$
Propriétés
Résidus
$R^2$
Premier pas avec R
annexes
Exo. math proj
Vect. gaussien

# Estimateur MCO

## Définition **Estimateur des MCO**

$$\begin{aligned}\widehat{\beta} &= \underset{\tilde{\beta}\in\mathbb{R}^{K+1}}{argmin}\left\{\sum_{i=1}^{n}\left(Y_i - x_i\tilde{\beta}\right)^2\right\}\\ &= \underset{\tilde{\beta}\in\mathbb{R}^{K+1}}{argmin}\left\{\|Y - x\tilde{\beta}\|_2^2\right\}.\end{aligned}$$

Ainsi

$$x\widehat{\beta} = \Pi_{vect(x)}Y.$$

Ainsi, les coordonnées de $\widehat{\beta}$ - l'estimateur des MCO de $\beta$ - représentent les coefficients de la projection orthogonale de $Y$ sur $vect(x)$, l'e.v. engendré par le vecteur $x$.

## Propriété : **Expression de l'estimateur du MCO**

On suppose $B$ inversible. On a

$$\begin{aligned}\widehat{\beta} &= B^{-1}\sum_{i=1}^{n} x_i^t Y_i,\\ &= (x^t x)^{-1} x^t Y\\ &= B^{-1} x^t Y.\end{aligned}$$

## Propriétés de $\widehat{\beta}$ sans hypothèse supplémentaire

- Il existe toujours une solution $\widehat{\beta}$ par les MCO ;
- Pas toujours unique.

# Propriétés de $\widehat{\beta}$ sans hypothèse supplémentaire

- Il existe toujours une solution $\widehat{\beta}$ par les MCO ;
- Pas toujours unique.

Soit

$$B = \sum_{i=1}^{n} x_i^t x_i = x^t x \in \mathcal{M}_{K+1}(\mathbb{R}).$$

$B$ est symétrique, donc : $B^t = B$.

**Propriété Condition d'existence des MCO**

Si $x^{(0)}, \cdots, x^{(K)}$ sont linéairement indépendants, alors $B$ est inversible.

Preuve :

Si on montre que $B$ est définie positive alors on aura montré que $B$ est inversible (on rappelle qu'une matrice $B$ symétrique à coefficients réels est dite définie positive si pour tout $a \in \mathbb{R}_*^{K+1}$ on a $a^t Ba > 0$)
$B$ est une matrice symétrique à valeur réelles. montrons que $B$ est définie positive. Soit $a \in \mathbb{R}_*^{K+1}$, $a^t Ba = a^t x^t xa = (xa)^t (xa) = \|xa\|_2^2$. Ainsi, l'existence d'un vecteur $a$ tel que $\|xa\|_2^2 = 0$, par définition d'une norme ($\|b\| = 0 \leftrightarrow b = 0$) est équivalent à dire que $xa = 0$.
Or $xa = \sum_{k=0}^{K} a_k x^{(k)}$. Donc $a$ tel que $xa = 0$ signifie que le système $(x^{(0)}, \ldots, x^{(K)})$ est lié. Ce qui est contradictoire avec l'hypothèse d'indépendance linéaire. Ainsi, pour tout $a$ non nul, on a $a^t Ba > 0$. ce qui achève la preuve.

# Propriétés de $\widehat{\beta}$

- Sans hypothèse sur la loi des $\xi_i$.

## Propriété

On suppose $B$ inversible. Alors $\widehat{\beta}$, l'unique estimateur par les moindres carrés ordinaires, est un ESB de $\beta$.

- On suppose maintenant l'hypothèse $(HM_3)$ vérifiée

$$(HM_3) \quad \xi \sim \mathcal{N}_n(0; \sigma^2 Id_n).$$

## Propriété MCO-Normalité

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible. Alors,

$$\forall \beta, \forall \sigma^2, \quad \widehat{\beta} \sim \mathcal{N}_{K+1}\left(\beta; \sigma^2 B^{-1}\right).$$

**Preuve : évident par multiplication matricielle d'un vecteur gaussien.**

# Propriétés de $\widehat{\beta}$

- Sans hypothèse sur la loi des $\xi_i$.

## Propriété

On suppose $B$ inversible. Alors $\widehat{\beta}$, l'unique estimateur par les moindres carrés ordinaires, est un ESB de $\beta$.

- On suppose maintenant l'hypothèse $(HM_3)$ vérifiée

$$(HM_3) \quad \xi \sim \mathcal{N}_n(0; \sigma^2 Id_n).$$

## Propriété MCO-Normalité

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible. Alors,

$$\forall \beta, \forall \sigma^2, \quad \widehat{\beta} \sim \mathcal{N}_{K+1}\left(\beta; \sigma^2 B^{-1}\right).$$

**Preuve : évident par multiplication matricielle d'un vecteur gaussien.**

Remarque : Sous $(HM_3)$, la loi de $\widehat{\beta}$ est connue à $\beta$ et $\sigma^2$ près.

# Propriétés de $\widehat{\beta}$

- Sans hypothèse sur la loi des $\xi_i$.

## Propriété

On suppose $B$ inversible. Alors $\widehat{\beta}$, l'unique estimateur par les moindres carrés ordinaires, est un ESB de $\beta$.

- On suppose maintenant l'hypothèse $(HM_3)$ vérifiée

$$(HM_3) \quad \xi \sim \mathcal{N}_n(0; \sigma^2 Id_n).$$

## Propriété MCO-Normalité

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible. Alors,

$$\forall \beta, \forall \sigma^2, \quad \widehat{\beta} \sim \mathcal{N}_{K+1}\left(\beta; \sigma^2 B^{-1}\right).$$

**Preuve : évident par multiplication matricielle d'un vecteur gaussien.**

Remarque : Sous $(HM_3)$, la loi de $\widehat{\beta}$ est connue à $\beta$ et $\sigma^2$ près. Nécessité d'estimer $\sigma^2$.

# Propriétés de $\widehat{\beta}$

- Sans hypothèse sur la loi des $\xi_i$.

## Propriété

On suppose $B$ inversible. Alors $\widehat{\beta}$, l'unique estimateur par les moindres carrés ordinaires, est un ESB de $\beta$.

- On suppose maintenant l'hypothèse $(HM_3)$ vérifiée

$$(HM_3) \quad \xi \sim \mathcal{N}_n(0; \sigma^2 Id_n).$$

## Propriété MCO-Normalité

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible. Alors,

$$\forall \beta, \forall \sigma^2, \quad \widehat{\beta} \sim \mathcal{N}_{K+1}\left(\beta; \sigma^2 B^{-1}\right).$$

**Preuve : évident par multiplication matricielle d'un vecteur gaussien.**

Remarque : Sous $(HM_3)$, la loi de $\widehat{\beta}$ est connue à $\beta$ et $\sigma^2$ près. Nécessité d'estimer $\sigma^2$. L'estimateur de $\sigma^2$ doit être indépendant de $\widehat{\beta}$ pour ne pas influencer la loi finale obtenue.

# Propriété du MCO avec normalité du bruit

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible. On propose de considérer

$$\widehat{\sigma^2} = \frac{\|Y - x\widehat{\beta}\|_2^2}{n - K - 1}.$$

Il s'agit, à une constante près, d'une mesure de ce qu'il reste de l'observation, après avoir retiré ce que l'on estime avec l'estimateur des MCO.
On remarque que

$$\widehat{\sigma^2} = \frac{\|\Pi_{vect(x)^\perp} Y\|_2^2}{dim(vect(x)^\perp)}.$$

Pour obtenir la loi de $\widehat{\beta}$, on introduit des quantités intermédiaires : le résidu et le prédicteur.

# Résidus, Prédicteur

*ATTENTION : Par souci de simplification des notations, on indiquera $(x)$ au lieu de $vect(x)$.*

- On note $\widehat{Y}$ le vecteur prédit : $\widehat{Y} = x\widehat{\beta} = \Pi_{(x)}Y$.
- On note $\widehat{u}$ le vecteur des résidus : $\widehat{u} = Y - \widehat{Y} = \Pi_{(x)^\perp}Y$.

Ainsi

1. $\widehat{u} \perp (x)$ -> $x^t\widehat{u} = 0$ ;
2. $\widehat{u} \perp \widehat{Y}$ -> $\widehat{Y}^t\widehat{u} = 0$.

## Résidus, Prédicteur

*ATTENTION : Par souci de simplification des notations, on indiquera* $(x)$ *au lieu de* $vect(x)$.

- On note $\widehat{Y}$ le vecteur prédit : $\widehat{Y} = x\widehat{\beta} = \Pi_{(x)}Y$.
- On note $\widehat{u}$ le vecteur des résidus : $\widehat{u} = Y - \widehat{Y} = \Pi_{(x)^\perp}Y$.

Ainsi

1. $\widehat{u} \perp (x)$ -> $x^t\widehat{u} = 0$ ;
2. $\widehat{u} \perp \widehat{Y}$ -> $\widehat{Y}^t\widehat{u} = 0$.

Remarques :

1/ si $x^{(0)} = 1_n$ alors $\bar{\widehat{u}} = 0$

2/

$$\widehat{\sigma}^2 = \frac{\|\widehat{u}\|_2^2}{n-K-1}.$$

Récapitulatif des expressions de $\widehat{Y}$ et $\widehat{u}$ en fonction de $Y$.

$$\begin{aligned} \widehat{Y} &= x(x^tx)^{-1}x^tY = \Pi_{(x)}Y \\ \widehat{u} &= Y - \widehat{Y} = (Id_n - \Pi_{(x)})Y = \Pi_{(x)^\perp}Y = \Pi_{(x)^\perp}\xi, \end{aligned}$$

car $\Pi_{(x)^\perp}x = 0_n$. On note que comme $\Pi_{(x)}$ est une matrice de projection orthogonale sur $x = (x^{(0)}, \dots, x^{(K)})$ on a

$$\Pi_{(x)}^2 = \Pi_{(x)}.$$

# MCO avec normalité du bruit

## Propriété MCO-Normalité *1

On suppose $(HM_0)$ à $(HM_3)$ vérifiées et $B$ inversible.

1. $$(n-K-1)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-K-1);$$
2. $\widehat{\beta}$ indépendant de $\widehat{\sigma}^2$;
3. $\widehat{\sigma^2}$ est un ESB de $\sigma^2$ ;
4. $var(\widehat{\sigma}^2) = 2\sigma^4/(n-K-1)$.

Preuve : On a $\widehat{u} = \Pi_{(x)^\perp} Y = \Pi_{(x)^\perp} x\beta + \Pi_{(x)^\perp}\xi = \Pi_{(x)^\perp}\xi$ et
$\widehat{\beta} - \beta = B^{-1}x^tY - \beta = B^{-1}x^t(x\beta+\xi) - \beta = B^{-1}B\beta + B^{-1}x^t\xi - \beta = B^{-1}x^t\xi$. Ainsi

$$\begin{pmatrix} \widehat{\beta} - \beta \\ \widehat{u} \end{pmatrix} = \begin{pmatrix} B^{-1}x^t \\ \Pi_{(x)^\perp} \end{pmatrix} \xi.$$

On note $A$ la matrice ainsi définie. On a donc un vecteur conjointement gaussien, d'espérance nulle. Il reste à déterminer sa matrice de variance-covariance.

## suite preuve

Calcul de matrice de variance-covariance.

$$\begin{aligned}\Sigma_{(\widehat{\beta}-\beta;\widehat{u})} &= A\Sigma_\xi A^t \\ &= \sigma^2 AA^t \\ &= \sigma^2 \begin{pmatrix} B^{-1} & 0 \\ 0 & \Pi_{(x)^\perp} \end{pmatrix}, \quad \text{car } \Pi_{(x)^\perp} x = 0.\end{aligned}$$

Ainsi $\Sigma_{\widehat{\beta}} = \sigma^2 B^{-1}$ et $\Sigma_{\widehat{u}} = \sigma^2 \Pi_{(x)^\perp}$. La loi jointe étant gaussienne, de matrice de variance covariance diagonale par bloc, ces deux vecteurs sont indépendants.
Donc $\widehat{\sigma}^2 = \|\widehat{u}\|_2^2/(n-K-1)$ est indépendant de $\widehat{\beta}$. On a $\widehat{u} \sim \mathcal{N}_n(0;\sigma^2\Pi_{(x)^\perp})$.
D'après Cochran,

$$\frac{\|\widehat{u}\|_2^2}{\sigma^2} \sim \chi^2(n-K-1).$$

car l'e.v engendré par $x$ est de dimension $n-(K+1)$. Enfin, comme $\|\widehat{u}\|_2^2 = (n-K-1)\widehat{\sigma}^2$, on obtient le résultat voulu.

## preuve, fin

On vérifie que l'on a bien un ESB de $\sigma^2$

$$\mathbb{E}\left((n-K-1)\frac{\widehat{\sigma}^2}{\sigma^2}\right) = n-K-1 \Longrightarrow \forall \sigma^2, \quad \mathbb{E}(\widehat{\sigma}^2) = \sigma^2.$$

Enfin, en utilisant la loi du $\chi 2$ :

$$var(\widehat{\sigma}^2) = 2\sigma^4/(n-K-1).$$

## preuve, fin

On vérifie que l'on a bien un ESB de $\sigma^2$

$$\mathbb{E}\left((n-K-1)\frac{\widehat{\sigma}^2}{\sigma^2}\right) = n-K-1 \Longrightarrow \forall \sigma^2, \quad \mathbb{E}(\widehat{\sigma}^2) = \sigma^2.$$

Enfin, en utilisant la loi du $\chi 2$ :

$$var(\widehat{\sigma}^2) = 2\sigma^4/(n-K-1).$$

Conclusion :

- Plus $n$ augmente, plus la variance de $\widehat{\sigma}^2$ devient petite ;
- Mais, plus la dispersion du bruit est grande, plus l'estimateur est imprécis.

## Décomposition de la variance

Ainsi, par Pythagore, on a

$$\|Y - \overline{Y}1_n\|_2^2 = \|Y - \widehat{Y}\|_2^2 + \|\widehat{Y} - \overline{Y}1_n\|_2^2.$$

rque : si $x^{(0)}$ différent de $1_n$, la même décomposition est valide mais à la place de $\bar{Y}1_n$ on aura $\Pi_{x^{(0)}}Y$.

# Décomposition de la variance

Ainsi, par Pythagore, on a

$$\|Y - \overline{Y}1_n\|_2^2 = \|Y - \widehat{Y}\|_2^2 + \|\widehat{Y} - \overline{Y}1_n\|_2^2.$$

rque : si $x^{(0)}$ différent de $1_n$, la même décomposition est valide mais à la place de $\bar{Y}1_n$ on aura $\Pi_{x^{(0)}}Y$.

$$SCT = SCR + SCE.$$

On veut $SCR$ petit et $SCE$ grand (maximum égal $SCT$).
$x = (x^{(0)}\ x^{(1)}\ \ldots\ x^{(K)})$.
On suppose maintenant que $x^{(0)} = 1_n$ -> présence d'une constante (i.e. $\Pi_{x^{(0)}}Y = \overline{Y}$).

# Décomposition de la variance

Ainsi, par Pythagore, on a

$$\|Y - \overline{Y}1_n\|_2^2 = \|Y - \widehat{Y}\|_2^2 + \|\widehat{Y} - \overline{Y}1_n\|_2^2.$$

rque : si $x^{(0)}$ différent de $1_n$, la même décomposition est valide mais à la place de $\bar{Y}1_n$ on aura $\Pi_{x^{(0)}}Y$.

$$SCT = SCR + SCE.$$

On veut $SCR$ petit et $SCE$ grand (maximum égal $SCT$).
$x = (x^{(0)}\ x^{(1)}\ \ldots\ x^{(K)})$.
On suppose maintenant que $x^{(0)} = 1_n$ -> présence d'une constante (i.e. $\Pi_{x^{(0)}}Y = \overline{Y}$).
Alors,

$$\begin{array}{rll} SCT & = SCE & +SCR \\ \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2 & = \frac{1}{n}\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2 & + \frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 \\ \text{variance totale} & = \text{variance expliquée} & + \text{variance résiduelle} . \end{array}$$

## Décomposition de la variance

Ainsi, par Pythagore, on a

$$\|Y - \overline{Y}1_n\|_2^2 = \|Y - \widehat{Y}\|_2^2 + \|\widehat{Y} - \overline{Y}1_n\|_2^2.$$

rque : si $x^{(0)}$ différent de $1_n$, la même décomposition est valide mais à la place de $\bar{Y}1_n$ on aura $\Pi_{x^{(0)}}Y$.

$$SCT = SCR + SCE.$$

On veut $SCR$ petit et $SCE$ grand (maximum égal $SCT$).
$x = (x^{(0)}\ x^{(1)}\ \ldots\ x^{(K)})$.
On suppose maintenant que $x^{(0)} = 1_n$ –> présence d'une constante (i.e. $\Pi_{x^{(0)}}Y = \overline{Y}$).
Alors,

$$\begin{array}{rll} SCT & = SCE & +SCR \\ \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 & = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 & + \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^2 \\ \text{variance totale} & = \text{variance expliquée} & + \text{variance résiduelle} . \end{array}$$

but –> (variance expliquée / variance totale) grand.

# Décomposition de la variance

$x^{(0)}$ quelconque.
Espace des individus $E_{ind}$ : un élément quelconque évolue dans un espace à $n$ dimensions.
$Y \in E_{ind}$ est bien décrit sur $n$ individus.
Dimension de quelques sous-espaces vectoriels de $E_{ind}$ :

$$
\begin{array}{ll}
dim\left(vect(x^{(0)})\right) = 1 & dim\left(vect(x)\right) = K+1 \\
dim\left(vect(x^{(0)})^{\perp}\right) = n-1 & dim\left(vect(x^{(0)})^{\perp x}\right) = K \\
dim\left(vect(x)^{\perp}\right) = n-K-1 & .
\end{array}
$$

# Décomposition de la variance

$x^{(0)}$ quelconque.
Espace des individus $E_{ind}$ : un élément quelconque évolue dans un espace à $n$ dimensions.
$Y \in E_{ind}$ est bien décrit sur $n$ individus.
Dimension de quelques sous-espaces vectoriels de $E_{ind}$ :

$$\begin{array}{ll} dim\left(vect(x^{(0)})\right) = 1 & dim\left(vect(x)\right) = K+1 \\ dim\left(vect(x^{(0)})^{\perp}\right) = n-1 & dim\left(vect(x^{(0)})^{\perp x}\right) = K \\ dim\left(vect(x)^{\perp}\right) = n-K-1 & . \end{array}$$

| Source | Somme des carrés | $d^0$ liberté | Carré moyen |
|---|---|---|---|
| Expliquée | $SCE = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$ | $K$ | $CME = \frac{SCE}{K}$ |
| Résiduelle | $SCR = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$ | $n-K-1$ | $CMR = \frac{SCR}{n-K-1}$ |
| Totale | $SCT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$ | $n-1$ | $CMT = \frac{SCT}{n-1}$ |

# Coefficient de détermination : $R^2$

$$SCT = SCE + SCR \geq 0$$

Sous hypothèse de présence d'une constante dans le modèle, on note $R^2$, le coefficient de détermination définit par

$$\begin{aligned} R^2 &= \frac{\text{variance expliquée}}{\text{variance totale}} \\ &= \frac{SCE}{SCT} \\ &= 1 - \frac{SCR}{SCT} \end{aligned}$$

Par construction, si présence d'une constante,

$$0 \leq R^2 \leq 1.$$

(preuve évidente) $R$ se nomme également le coefficient de corrélation multiple.

# Coefficient de détermination : $R^2$

$$SCT = SCE + SCR \geq 0$$

Sous hypothèse de présence d'une constante dans le modèle, on note $R^2$, le coefficient de détermination définit par

$$\begin{aligned} R^2 &= \frac{\text{variance expliquée}}{\text{variance totale}} \\ &= \frac{SCE}{SCT} \\ &= 1 - \frac{SCR}{SCT} \end{aligned}$$

Par construction, si présence d'une constante,

$$0 \leq R^2 \leq 1.$$

(preuve évidente) $R$ se nomme également le coefficient de corrélation multiple.
On désire $R^2$ le plus grand possible.

# Premier pas avec R : importation d'un csv

```
A<-read.table("scripts_donnees/sportcourse.csv", header=T, row.names=1, sep=";")

str(A) ; edit(A)

A[1:4,1:4]
          X100m X200m X400m X800m
Argentina 10.39 20.81 46.84  1.81
Australia 10.31 20.06 44.84  1.74
Austria   10.44 20.81 46.82  1.79
Belgium   10.34 20.68 45.04  1.73

A[1:4,5:8]
          X1500m   X5K  X10K Marathon
Argentina   3.70 14.04 29.36   137.72
Australia   3.57 13.28 27.66   128.30
Austria     3.60 13.26 27.72   135.90
Belgium     3.60 13.22 27.45   129.95


variable.names(A)

[1] "X100m"    "X200m"    "X400m"    "X800m"    "X1500m"   "X5K"
 "X10K"     "Marathon"
```

# Premier pas avec R : Visualisation externe

```
X11()         # pour ouvrir une fenêtre graphique en dehors de
              # l'espace de Rstudio

plot(A$x400m, A$x100m)  # graphe

text(A$x400m,A$x100m, labels=rownames(A),
                pos=4, cex=0.8,
                col="blue")             # ajout du texte

dev.off() #fermeture de la fenêtre graphique externe
```

# R et $\widehat{\beta}$

```
>reg<-lm(X100m~X200m+X400m, data=A) ; reg  #avec constante
Call:
lm(formula = X100m ~ X200m + X400m, data = A)
Coefficients:
(Intercept)        X200m        X400m
    -0.3646       0.4086       0.0491

> x<-cbind(1,A$X200m,A$X400m) ;  y<-as.matrix(A$X100m)
> B=t(x)%*%x ; det(B)   ; betaesti<-solve(B)%*%t(x)%*%y ; t(betaesti)
[1] 39088.39
           [,1]     [,2]       [,3]
[1,] -0.3645595 0.408575 0.04909568
>
> t(reg$fitted.values)[,1:5] ; t(reg$residuals)[,1:5]
Argentina Australia   Austria   Belgium   Bermuda
 10.43753  10.03291  10.43655  10.29604  10.29790
Argentina    Australia      Austria      Belgium      Bermuda
-0.047528338  0.277094288  0.003453576  0.043958633 -0.017897102
```

## R et $\widehat{u}$

```
names(reg)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

> names(summary(reg))
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

> names(anova(reg))
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)"
```

# R et $\widehat{\beta}$

```
summary(reg)

Call:
lm(formula = X100m ~ X200m + X400m, data = A)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27705 -0.05780 -0.01790  0.06254  0.46649

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.36456    0.60464  -0.603   0.5492
X200m        0.40858    0.05329   7.666  4.3e-10 ***
X400m        0.04910    0.02358   2.082   0.0423 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1327 on 52 degrees of freedom
Multiple R-squared:  0.8627,    Adjusted R-squared:  0.8574
F-statistic: 163.4 on 2 and 52 DF,  p-value: < 2.2e-16
```

# R et Anova

```
anova(reg)
Analysis of Variance Table

Response: X100m
          Df Sum Sq Mean Sq  F value  Pr(>F)
X200m      1 5.6772  5.6772 322.4193 < 2e-16 ***
X400m      1 0.0763  0.0763   4.3353 0.04227 *
Residuals 52 0.9156  0.0176
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

# Graphes

```
%
X11()
par(mfrow=c(1,2))  # pour partager la fenêtre selon
                                   # le nb de col et lignes de mfrow
plot(X400m, X100m,
        xlim=c(0,60), ylim=c(0,13),
        pch=3, col="blue")
abline(h=mean(X100m), v=mean(X400m), col="red")
abline(coef=c(reg$coef[1],reg$coef[3]),col="Forestgreen")
abline(coef=c(lm(X100m~X400m)$coef[1],lm(X100m~X400m)$coef[2]),col="blue")

plot(X200m, X100m,
        xlim=c(0,30), ylim=c(0,13),
        pch=3, col="blue")
abline(h=mean(X100m), v=mean(X200m), col="red")
abline(coef=c(reg$coef[1],reg$coef[2]),col="Forestgreen")
abline(coef=c(lm(X100m~X200m)$coef[1],lm(X100m~X200m)$coef[2]),col="blue")
dev.off()
```

# Exercice 1 : Matrices de projection, base cano

Soit E $= \mathbb{R}^4$ muni de sa base canonique : $e_1 = (1,0,0,0)^t$, $e_2 = (0,1,0,0)^t$, $e_3 = (0,0,1,0)^t$, $e_4 = (0,0,0,1)^t$.
Soit W $= vect(e_1, e_3)$.

1. Quelle est la dimension de $\mathbb{R}^4$ ?
2. Quelle est la taille du vecteur $e_1$ ?
3. Quelle est la dimension de $W$ ?
4. Expliciter la matrice de projection orthogonale $\Pi_{\mathsf{W}}$ de E sur W.
5. Quelles sont les dimensions de l'espace de départ et de l'espace d'arrivée de $\Pi_{\mathsf{W}}$ ?
6. Quelle est la dimension de cette matrice de projection ?
7. Vérifier que $\Pi_{\mathsf{W}}$ est symétrique.
8. Vérifier que $\Pi_{\mathsf{W}} e_1 = e_1$.
9. Sans effectuer le calcul que vaut $\Pi_{\mathsf{W}} e_2$ ?

# Exercice 2 : matrice de projection et base orthogonale

On rappelle la propriété suivante :
*Soit $(v_1, \ldots, v_p)$, p vecteurs de $\mathbb{R}^n$ formant une base orthogonale de $\mathbb{R}^p$, avec $p \leq n$.*
*On désigne par W le sous-espace vectoriel défini par $\mathsf{W} = vect(v_1, \ldots, v_p)$.*
*On note la représentation matricielle $V = (v_1, \ldots, v_p)$. Alors, $\Pi_{\mathsf{W}}$, la matrice de projection de $\mathbb{R}^n$ sur W, est définie par*

$$\Pi_{\mathsf{W}} = V(V^tV)^{-1}V^t.$$

1. En explicitant $V$ en fonction des $v_j$, indiquer ce que représente $V^tV$ par rapport à des objets de statistique descriptive usuels.
2. Expliciter $(V^tV)^{-1}$ en fonction des $v_j$ puis des objets statistiques usuels.
3. Exprimer formellement $\Pi_{\mathsf{W}}$.
4. Vérifier que la matrice de projection de la partie précédente se construit ainsi.

# Exercice 3 : matrice de projection et base orthonormale

1. Que représente $V^tV$ dans le cas d'une matrice orthonormale ?
2. Quelle conclusion en tirer sur son inverse ?
3. Vérifier que la matrice de projection construite dans la première partie vérifie cette propriété.
4. A l'aide cette observation, peut-on construire la matrice de projection plus rapidement ?

Application immédiate
Soit $\mathsf{E} = \mathbb{R}^3$ muni de sa base canonique $(e_1, e_2, e_3)$ et du produit scalaire usuel. Soit $\mathsf{F} = vect(e_1, e_2)$.

1. A l'aide des documents de cours, rappeler la définition du projeté sur l'orthogonal d'un sous-espace vectoriel
2. Exprimer $\mathsf{F}^\perp$.

# Exercice 4 : matrice de projection, application

Soit $v_1 = (1,2,0)^t$ et $v_2 = (2,-1,0)^t$ deux vecteurs de $\mathbb{R}^3$.

1. Montrer que $v_1$ et $v_2$ sont orthogonaux.
2. Calculer la norme euclidienne de ces vecteurs et les normaliser. On note $w_1$ et $w_2$ la version normalisée de ces vecteurs.
3. Construire $\Pi_{\mathsf{W}}$,la matrice de projection sur $\mathsf{W} = vect(v_1, v_2)$ (on se servira de $w_1$ et $w_2$).
4. Que remarque-t-on ?
5. Soit $X = (1,2,-1)^t$ un vecteur de $\mathbb{R}^3$.
    1. Quel est le projeté de $X$ sur W ?
    2. Que vaut $\Pi^2_{\mathsf{W}}X$ ?
    3. Quel est le projeté de $X$ sur $\mathsf{W}^\perp$ ?
    4. Quel est l'élément de W tel que la distance de cet élément à $X$ soit la plus petite possible ? Exprimer cet élément.

# Vecteur gaussien

## Définition : vecteur gaussien

On dit que $Z = (Z_1, \cdots, Z_n)^t$ est un vecteur gaussien de $\mathbb{R}^n$, et on note $Z \sim \mathcal{N}_n$, si

$$\forall u \in \mathbb{R}^n, \quad u^t Z \sim \mathcal{N}_1.$$

On sait calculer l'espérance et la matrice de variance-covariance de toute combinaison linéaire de vecteur gaussien.

## Propriété : paramètres d'un vecteur gaussien

Soit $Z \sim \mathcal{N}_n(M;\Sigma)$ avec $M$ un vecteur de $\mathbb{R}^n$ et $\Sigma$ une matrice symétrique définie positive de taille $n$ et $A$ une matrice de taille $p \times n$. On a alors

$$A.Z \sim \mathcal{N}_p(AM; A\Sigma A^t).$$

## Lois dérivées de vecteur gaussien : loi du $\chi^2$ et li de Fisher

Loi d $\chi^2$

Soit $Z$ un vecteur gaussien de $\mathbb{R}^n$ centré et réduit. Alors,

$$\begin{aligned} W &= \|Z\|_2^2 = \sum_{i=1}^{n} Z_i^2 \\ &\sim \chi^2(n) \end{aligned}$$

et $\mathbb{E}(W) = n$ et $var(W) = 2n$.

Loi de Fisher

Soient $U \sim \chi^2(s)$ et $W \sim \chi^2(r)$ deux variables indépendantes entre elles, alors

$$\frac{U/s}{W/r} \sim F(s,r)$$

où $F(s;r)$ est la loi de Fisher à $s$ et $r$ degrés de liberté.

# Lois dérivées de vecteur gaussien : loi de Student

Soient $Z \sim \mathcal{N}(0;1)$ et $U \sim \chi^2(s)$ deux variables indépendantes entre elles, alors

$$\frac{Z}{\sqrt{U/s}} \sim T(s)$$

où $T(s)$ est la loi de Student à $s$ degrés de liberté.

On remarque que

$$s.\frac{Z^2}{U} = \left(\frac{Z}{\sqrt{U/s}}\right)^2 \sim F(1;s).$$

# Projection et vecteur gaussien

Soit $Z$ un vecteur gaussien centré et réduit de $\mathbb{R}^n$.

## Cochran

Soit $\mathbb{R}^n = \mathsf{V}_1 \oplus^\perp \cdots \oplus^\perp \mathsf{V}_r$ la somme orthogonale de s.e.v. $\mathsf{V}_i$ tels que $dim(\mathsf{V}_i) = r_i$ et $\sum_{i=1}^r r_i = n$.
Pour tout $i$, on note $\Pi_{\mathsf{V}_i}$ la projection orthogonale sur le s.e.v. $\mathsf{V}_i$.
On a alors

1. $\Pi_{\mathsf{V}_i} Z$ indépendant de $\Pi_{\mathsf{V}_j} Z$ pour tout $i \neq j$.
2. $\|\Pi_{\mathsf{V}_i} Z\|_2^2 \sim \chi^2(r_i)$.

ModLin Chap 1
E. Gautherat
Exemples
Introduction formelle
Contexte
Démarche
$\widehat{\beta}$
Définition
Mat.proj.ortho
Déf $\widehat{\beta}$
Propriétés
Résidus
$R^2$
Premier pas avec R
annexes
Exo. math proj
Vect. gaussien

# Exercice : matrice de projection et vecteur gaussien

- Préliminaires
  Soit $Z$ un vecteur gaussien centré et réduit de $\mathbb{R}^n$ avec $n \geq 2$
  Soit $U$, matrice de projection sur un s.e.v. de dimension $p$ de $\mathbb{R}^n$.

  1. Quelle est l'espérance de $Z$ ? Quelle est la matrice de covariance de $Z$ ? Quelle est la covariance ente $Z_1$ et $Z_2$ les deux premières coordonnées du vecteur $Z$ ?
  2. Quelle est la loi de $Z_1^2$ ?
  3. Quelle est la loi de $\|Z\|_2$ ?
  4. Exprimer $U$.
  5. Simplifier les expressions $U^t$ et $U^2$.
  6. En déduire la loi de $UZ$.

# Exercice suite : matrice de projection et vecteur gaussien

- Application
  Soit W le s.e.v. de $\mathbb{R}^3$ engendré par $v_1 = (1, 1, 0)^t$ et $v_2 = (0, 0, 2)^t$. Soit $X$ un vecteur gaussien de $\mathbb{R}^3$ d'espérance $(1, 2, 3)^t$ et de matrice de variance-covariance

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

1. Exprimer $\Pi_W$ la matrice de projection orthogonale sur $W$.
2. Ecrire formellement la loi suivie par $\Pi_W X$.

# Exercice : matrice de projection et Cochran

Soit $X_1, \cdots, X_n \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(m, \sigma^2)$. On note $X = (X_1, \cdots, X_n)^t$.

1. Préliminaires
   1. Montrer que $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ est le coefficient du projeté de $X$ sur le s.e.v. W engendré par le vecteur $\mathbb{1} = (1, \ldots, 1)^t$.
   2. Exprimer $\overline{X}$ comme produit matriciel entre une matrice ligne et le vecteur $X$.
   3. En déduire la loi de $\overline{X}$.
   4. Exprimer le projeté de $X$ sur W$^\perp$ le s.e.v. orthogonal à W.
   5. En déduire sa loi.
   6. Exprimer la norme euclidienne au carré de ce vecteur.
   7. En supposant que $\sigma^2 = 1$ et $m = 0$, en déduire sa loi.
2. Montrer que $\bar{X} \perp\!\!\!\perp S_n^2$ avec $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.
3. Montrer que $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$.
4. En déduire que $\sqrt{n}\frac{\bar{X}-m}{\sqrt{S_n^2}} \sim T(n-1)$.