UNIVERSITÉ LAVAL
École d'actuariat

ACT 2003
Notes de cours
Modèles linéaires en actuariat

David Beauchemin

Automne 2017

Résumé

abstrat

Remerciements

blah blah

Table des matières

Chapitre 1

Introduction

L'établissement de prévisions joue un rôle central dans notre vie de tous les jours (prévisions météorologique, horoscope, etc.), et plus particulièrement dans celle des actuaires.

Objectifs de la régression

Régulièrement en actuariat, on se questionne sur les effets de différentes variables sur d'autres. Par exemple,

- Quel est l'effet de l'âge sur la fréquence des sinistres automobiles ?
- Quel est l'effet du sexe sur la mortalité ?

On cherche à étuider et déterminer les relations entre des variables mesurables à partir de données.

Deux grandes classes de variables mesurables :

- Qualitatives : basées sur des opinions et/ou des intuitions.
- Quantitatives : basées sur des observations, un modèle et des arguments mathématiques.

Deux *grandes étapes* pour établir des prévisions quantitatives

1. Bâtir le modèle et estimer les paramètres :
   ex : $F = M \times a$ Qui représente un modèle déterministe
   ex : $Y = 3 \times X + 6 + \epsilon_t$;où$\epsilon_t \sim N(0, 10)$ Qui représente un modèle probabiliste
2. Calculer les prévisions à partir du modèle.

Dans le cadre du cours, seulement les modèles probabilistes linéaires seront étudiez.

Chapitre 2

Régression linéaire simple

2.1 Introduction

De façon générale, en régression, nous avons :

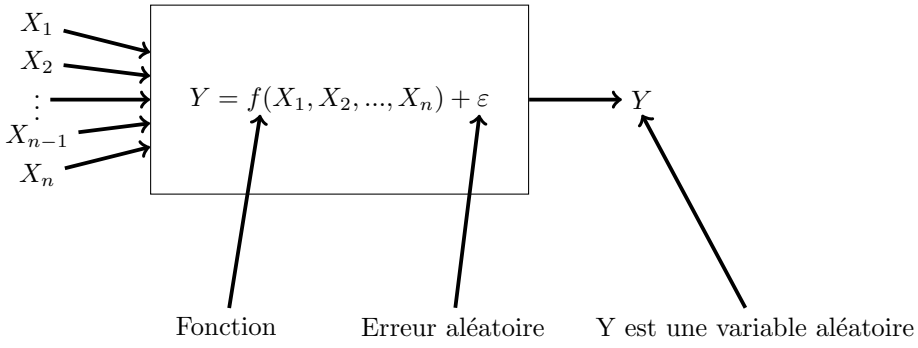
Y	Variable dépendante, ou de réponse	Output
$X_1, X_2, ..., X_n$	Soit n variables indépendantes ou explicatives, ou exogènes [1]	Input
$\beta_0, \beta_1, ...\beta_n$	Les paramètres à estimer	

Voici une illustration du concept de régression linéaire

Étape 1
Observation des X_i

Étape 2
Modèle de régression

Étape 3
Prévision de Y



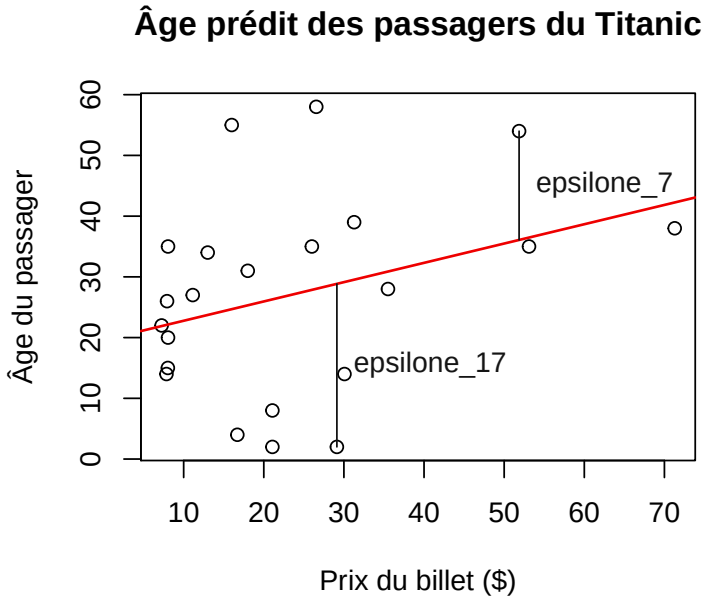
1. Les variables X_i sont indépendante par rapport à y, mais pas nécessairement entre elles.

2.1.1 Regression linéaire simple

On cherche à prédire l'âge des passagers du Titanic selon le prix du billet à l'aide du modèle linéaire suivant,

$$Y = \beta_0 + \beta_1 \times X + \varepsilon$$

Âge du passager — Prix du billet — Erreur aléatoire

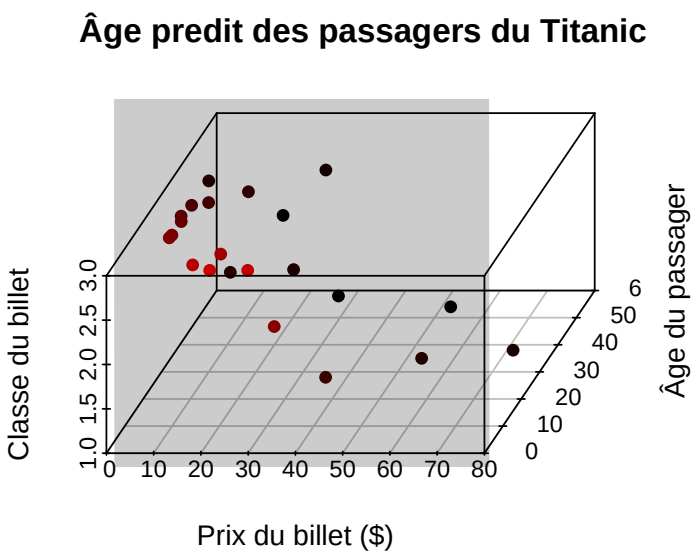


2.1.2 Regression linéaire multiple

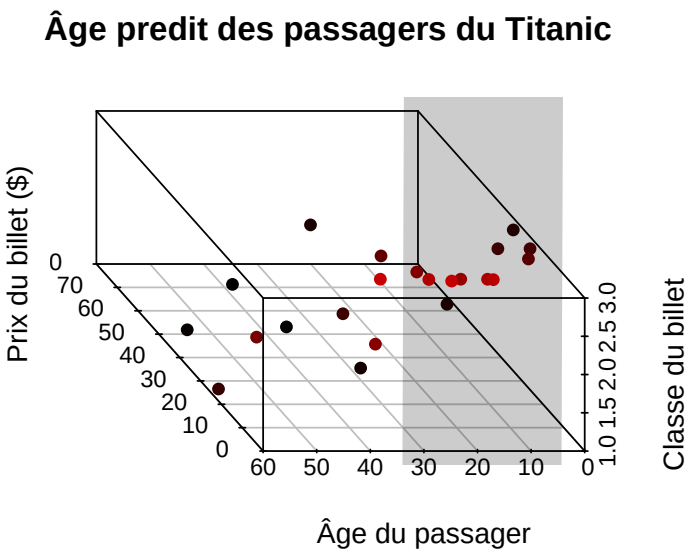
On cherche à prédire l'âge des passagers du Titanic selon le prix du billet et son sexe à l'aide du modèle linéaire suivant,

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \varepsilon$$

Âge du passager — Prix du billet — Sexe du passager — Erreur aléatoire



Voici la régression sous un autre angle, on voit la surface plane de régression.

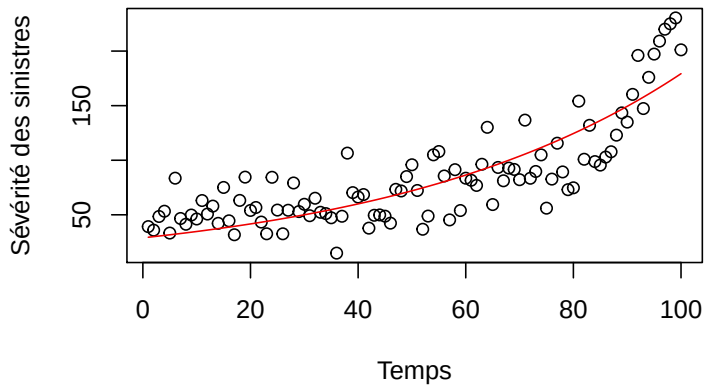


2.1.3 Régression exponentielle

On cherche à prédire la sévérité d'un sinistre automobile en fonction du temps à l'aide du modèle exponentielle suivant,

$$Y = \beta_0 \times e^{\beta_1 \times X} \times \varepsilon$$

Sévérité du sinistre — Temps — Erreur aléatoire



Note

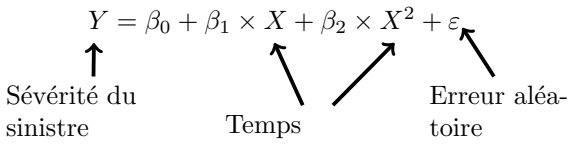
On remarque que la régression exponentielle est similaire à une régression linéaire simple.

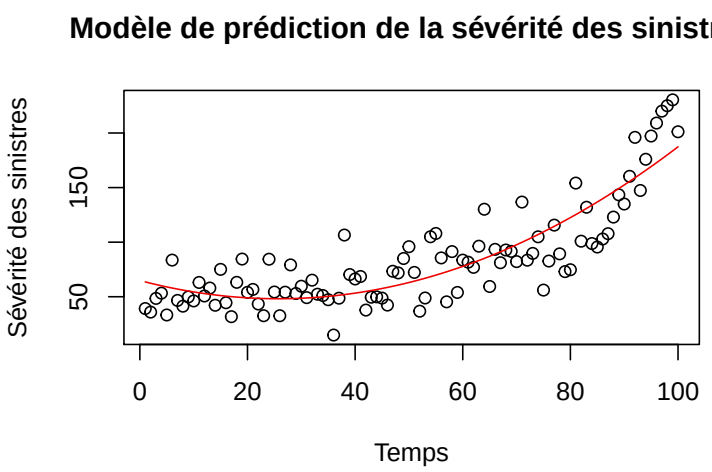
$$\ln(Y) = \ln(\beta_0) + \beta_1 \times X + \ln(\varepsilon)$$
$$Y^* = \beta_0^* + \beta_1 \times X + \varepsilon^*$$

Qu'on appel aussi une régression multiplicative ou log-linéaire.

2.1.4 Régression quadratique

On cherche à prédire la sévérité d'un sinistre automobile en fonction du temps et du temps au carré à l'aide du modèle quadratique suivant,





Note

On remarque que la régression quadratique est similaire à une régression linéaire multiple. En posant $X_1 = X$ et $X_2 = X^2$

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \varepsilon$$

Soit une régression linéaire multiple.
Dans le cadre du cours, seulement les modèles linéaires seront à l'étude car,

- Plus simples
- Plusieurs modèles peuvent se ramener à un modèle linéaire simple ou multiple. (voir 2.1.3 et 2.1.4)
- Constituent souvent une très bonne approximation de la réalité qui peut être très complexe, tel que l'assurance.
- Se généralisent faciment, tel que les *Generalized Linear Models*.

Le principale problème de la modélisation linéaire est de trouver les différents paramètres $\beta_0, \beta_1, ..., \beta_n$ de tel sorte que

$$\varepsilon = Y - f(X_1, ..., X_n; \beta_0, \beta_1, ..., \beta_n) \tag{2.1}$$

soit minimiser.
Il existe plusieurs méthode pour calcul l'erreur. Soit les erreurs suivants :

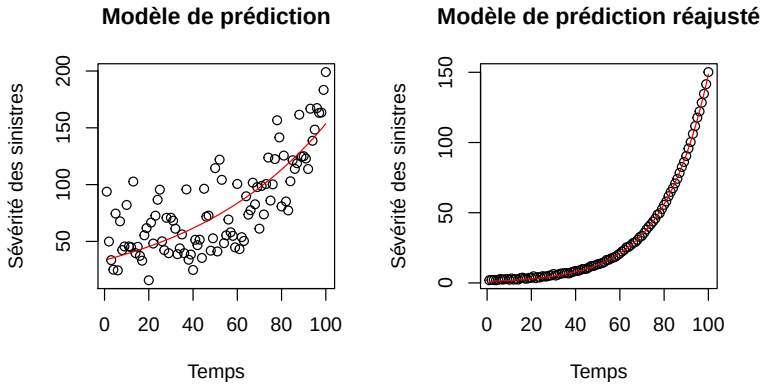
- Erreur totale
- Erreur absolue
- Erreur quadratique

Quel type d'erreur est suffisante pour déterminer ε ?

2.1.4.1 Erreur totale

$$\sum_{t=1}^{n} \varepsilon_t = \sum_{t=1}^{n} \Big(Y_t - (\beta_0 + \beta_1 \times X_t) \Big) \tag{2.2}$$

- Facile à mettre à 0
- Pas fiable à cause de la mise à zéro



2.1.4.2 Erreur absolue

$$\sum_{t=1}^{n} |\varepsilon_t| = \sum_{t=1}^{n} \Big| Y_t - (\beta_0 + \beta_1 \times X_t) \Big| \tag{2.3}$$

- Très robuste
- Très compliqué mathématiquement, pour minimiser $\sum_{t=1}^{n} |\varepsilon_t|$ cela implique de dériver la fonction.

2.1.4.3 Erreur quadratique

$$\sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} \Big[Y_t - (\beta_0 + \beta_1 \times X_t) \Big]^2 \tag{2.4}$$

- Mathématiquement plus simple que l'erreur quadratique.
- Donne beaucoup de poids aux grandes erreurs

L'erreur quadratique semble donc l'option la plus simple dû à la facilité mathématique et ça fiabilité.

2.2 Le modèle de régression linéaire simple

Le modèle de régression linéaire simple tente d'expliquer le mieux possible la variable dépendante [2] Y à l'aide d'une variable indépendante [3] X .
Si on dispose de n paires d'observations $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ alors, le modèle s'exprime comme suit :

$$Y_i = \beta_0 + \beta_1 \times X_i + \varepsilon_i, \text{ i} = 1, ...,\text{n}. \tag{2.5}$$

Où β_0 est le paramètre associé à l'ordonnée à l'origine du modèle ; β_1 est le paramètre associé à la pende de la droite ; et ε est le terme d'erreur.

Quelques remarques sur le modèle

Dans l'équation 2.5 du modèle, on remarque que

- Les observations de Y_i sont tiré d'une varaible aléatoire ;
- Les observations de X_i sont considérées comme des valeurs connues et non aléatoires ;
- Les paramètres β_0 et β_1 sont inconnus au départ et doivent être estimer ;
- ε_i sont des réalisations inconnues d'une variable aléatoire.

Exemple d'un modèle de régression

X_t : Nombre d'années de scolarité de l'actuaire$t$
Y_t : Salaire de l'actuaire$t$

Comment résoudre le modèle pour prédire les salaires des actuaires en fonction du nombre d'années de scolarité ?
Raisonnement :

- Pour $X_t = 0$; on a $Y_t = \beta_0$. Autrement dit, le salaire avec un nombre d'année de scolarité est *en moyenne* de β_0. Par exemple, β_0 serait le salaire moyen d'un stagiaire.
- Par la suite, pour chaque année additionnelle de scolarité, le salaire augmente *en moyenne* de β_1 unitées.

2. On appel parfois la variable dépendante une variable endogène. Qui s'interpréte comme étant une variable qui est dû à une cause interne.

3. On appel parfois les variables dépendantes des variables exogène. Qui s'intrepréte comme étant extérieur à un système.

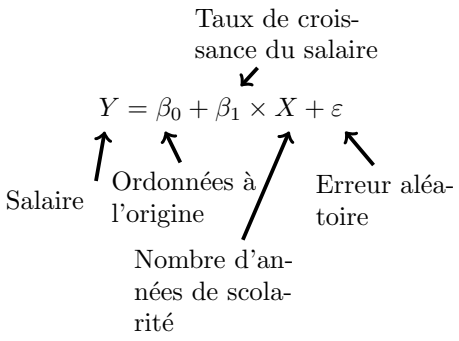
Ainsi, *en moyenne* on a

$$E[Y_t|X_t] = \beta_0 + \beta_1 \times X_t$$

Habituellement, la relation n'est pas parfaitement exacte dans la réalité. On se retrouve ainsi avec une *différence* dans notre variable exogène prédite. L'erreur est notée ε_t et est tel que mentionnée plus tôt, assumée aléatoire.

$$\begin{aligned}\varepsilon_t &= Y_t - E[Y_t|X_t] \\ &= Y_t - (\beta_0 + \beta_1 \times X_t)\end{aligned}$$

En réoganisant, on retrouve l'équation 2.5.



On doit maintenant trouver les paramètres β_0 et β_1 de manière à minimiser l'erreur ε_t.
Si ε_t est minimal, cela veut dire que $Y_t \approx \beta_0 + \beta_1 \times X_t$. Ce qui signifie que la droite de régression est une bonne approximation de Y_t.

> ***En résumé***
>
> En résumé, on cherche à minimiser nos résidus en optimisant les paramètres β_i.

2.2.1 Coefficients de régression

Les paramètres β_0 et β_1 sont déterminés en minimisant l'erreur quadratique à l'aide de la méthode des moindres carrées.

$$
\begin{aligned}
S(\beta_0, \beta_1) &= \sum_{t=1}^{n} \varepsilon_t^2 \\
&= \sum_{t=1}^{n} \big(Y_t - (\beta_0 + \beta_1 \times X_t)\big)^2 \\
&= \sum_{t=1}^{n} \big(Y_t - \beta_0 - \beta_1 \times X_t\big)^2
\end{aligned}
$$

Où $S(\psi)$ peut être considéré comme une mesure de la *distance* entre les données observées et le modèle théorique qui prédit ces données [4].

Afin de minimiser la fonction $S(\beta_0, \beta_1)$ on dérive la fonction partiellement en fonction de chacun des paramètres.

Minimisation de β_0

$$
\begin{aligned}
\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_0} &= 0 \\
\frac{\partial}{\partial \beta_0} \sum_{t=1}^{n} \big(Y_t - \hat{\beta}_0 - \hat{\beta}_1 \times X_t\big)^2 &= 0 \\
-2 \sum_{t=1}^{n} \big(Y_t - \hat{\beta}_0 - \hat{\beta}_1 \times X_t\big) &= 0
\end{aligned}
$$

$$
\sum_{t=1}^{n} Y_t - n \times \hat{\beta}_0 - \hat{\beta}_1 \sum_{t=1}^{n} X_t = 0 \tag{2.6}
$$

4. Pour de plus ample information sur la méthode des moindres carrées et la fonction de *distance*, la page Wikipédia contient des bonnes explications sur le sujet.

Minimisation de β_1

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \beta_1} = 0$$
$$\frac{\partial}{\partial \beta_1} \sum_{t=1}^{n} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 \times X_t\right)^2 = 0$$
$$-2 \sum_{t=1}^{n} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 \times X_t\right) \times X_t = 0$$

$$\sum_{t=1}^{n} Y_t \times X_t - \hat{\beta}_0 \sum_{t=1}^{n} X_t - \hat{\beta}_1 \sum_{t=1}^{n} X_t^2 = 0 \tag{2.7}$$

À l'aide des équations 2.6 et 2.7, on peut trouver les deux inconnus β_0 et β_1.
À partir de 2.6 :

$$\sum_{t=1}^{n} Y_t - n \times \hat{\beta}_0 - \hat{\beta}_1 \sum_{t=1}^{n} X_t = 0$$
$$\sum_{t=1}^{n} Y_t - \hat{\beta}_1 \sum_{t=1}^{n} X_t = n \times \hat{\beta}_0$$
$$\frac{\sum_{t=1}^{n} Y_t}{n} - \hat{\beta}_1 \frac{\sum_{t=1}^{n} X_t}{n} = \hat{\beta}_0$$

$$\boxed{\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}} \tag{2.8}$$

Et à partir de 2.7 :

$$\sum_{t=1}^{n} Y_t \times X_t - \hat{\beta}_0 \sum_{t=1}^{n} X_t - \hat{\beta}_1 \sum_{t=1}^{n} X_t^2 = 0$$
$$\sum_{t=1}^{n} Y_t \times X_t - \hat{\beta}_0 \sum_{t=1}^{n} X_t = \hat{\beta}_1 \sum_{t=1}^{n} X_t^2$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} Y_t \times X_t - \hat{\beta}_0 \sum_{t=1}^{n} X_t}{\sum_{t=1}^{n} X_t^2} \tag{2.9}$$

On utilise l'équation 2.8 de $\hat{\beta}_0$ avec l'équation 2.9 de $\hat{\beta}_1$, on développe l'équation résultante afin d'isoler $\hat{\beta}_1$.

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^{n} Y_t \times X_t - (\overline{Y} - \hat{\beta}_1 \overline{X}) \sum_{t=1}^{n} X_t}{\sum_{t=1}^{n} X_t^2} \\ &= \frac{\sum_{t=1}^{n} Y_t \times X_t - (\overline{Y} - \hat{\beta}_1 \overline{X}) \times n\overline{X}}{\sum_{t=1}^{n} X_t^2} \\ &= \frac{\sum_{t=1}^{n} Y_t X_t - n\overline{YX} + \hat{\beta}_1 \times \overline{X}^2 \times n}{\sum_{t=1}^{n} X_t^2}\end{aligned}$$

En isolant $\hat{\beta}_1$, on obtient la définition suivante

$$\boxed{\hat{\beta}_1 = \frac{\sum_{t=1}^{n} Y_t X_t - n\overline{YX}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2}} \tag{2.10}$$

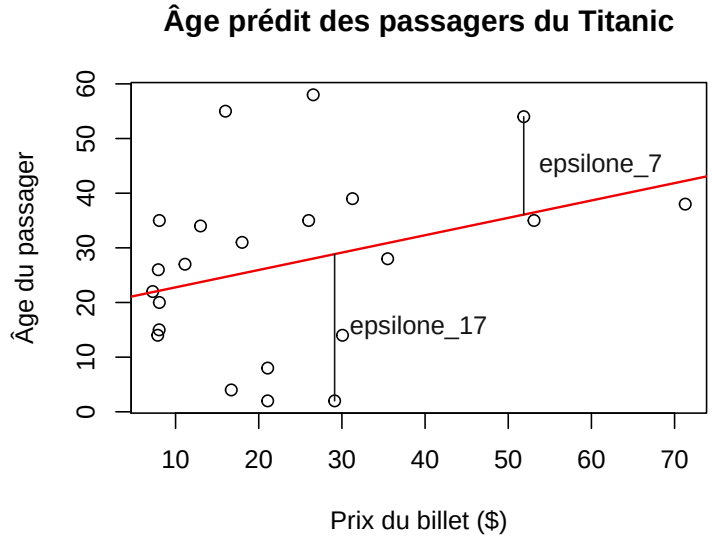
Remarques

1. On note $\hat{\varepsilon}_t$ les résidus générés par le modèle estimé :

$$\begin{aligned}\hat{\varepsilon}_t &= Y_t - \hat{Y}_t \\ \hat{\varepsilon}_t &= Y_t - (\hat{\beta}_0 - \hat{\beta}_1 X_t) \text{ ; pour t = 1,2,...,n}\end{aligned}$$

Si on illustre graphiquement les résidus, il s'agit du segment le plus court entre la droite de régression et la donnée observée.

Si on reprend le graphique de la section 2.1.1, on observe facilement les résidus sur cette représentation graphique :



2. Le *centre de gravité*[5] des données $(\overline{X}, \overline{Y})$ se trouve exactement sur la droite de régression.

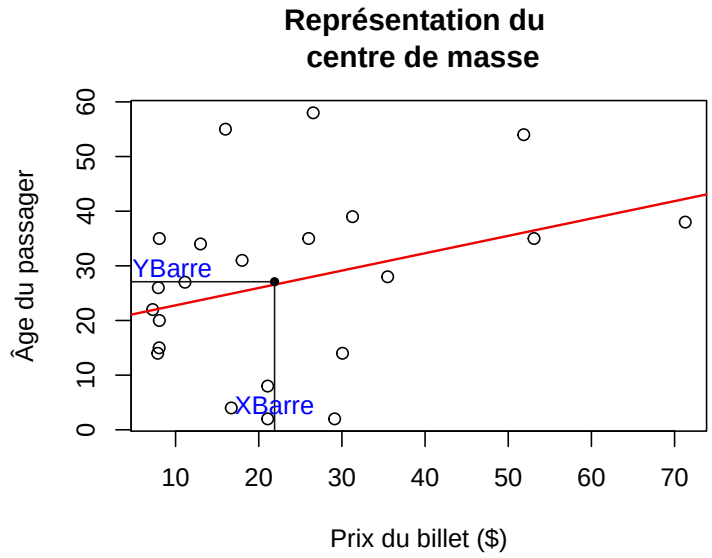
   On peut facilement effectuer cette preuve à partir de l'équation 2.8,

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} + 0$$

   On note ainsi une absence de résidus pour le centre de masse.

   Si on reprend (encore) le graphique de la section 2.1.1, on observe facilement le centre de masse sur le graphiqe.

5. Qu'on appelle parfois centre de masse.



3. La somme des résidus de tout modèle de régression linéaire est nulle.

$$\begin{aligned}
\sum_{t=1}^{n} \hat{\varepsilon}_t &= \sum_{t=1}^{n} \left(Y_t - (\hat{\beta}_0 + \hat{\beta}_1 X_t)\right) \\
&\overset{2.8}{=} \sum_{t=1}^{n} \left(Y_t - (\overline{Y} - \hat{\beta}_1 \overline{X})\right) \\
&= \sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} \overline{Y} + \hat{\beta}_1 \sum_{t=1}^{n} \overline{X} - \hat{\beta}_1 \sum_{t=1}^{n} X_t \\
&= n\overline{Y} - n\overline{Y} + \hat{\beta}_1 + n\overline{X} - \hat{\beta}_1 + n\overline{X} \\
&= 0
\end{aligned}$$

Notation

Afin de faciliter l'écriture, on intégre la notation suivante, S_{xx} et S_{xy}. Qui sont appelés respectivement la somme des carrées corrigée de x et la somme des produits

croisés corrigée de x et de y. Voici le développement pour S_{xx},

$$\begin{aligned}
S_{xx} &= \sum_{t=1}^{n}(X_t - \overline{X})^2 \\
&= \sum_{t=1}^{n}(X_t - \overline{X})^2 \\
&= \sum_{t=1}^{n}(X_t^2 - 2X_t\overline{X} + \overline{X}^2) \\
&= \sum_{t=1}^{n} X_t^2 - 2\overline{X}\sum_{t=1}^{n} X_t + n\overline{X}^2 \\
&= \sum_{t=1}^{n} X_t^2 - 2\overline{X}n\overline{X} + n\overline{X}^2 \\
&= \sum_{t=1}^{n} X_t^2 - n\overline{X}^2
\end{aligned}$$

On effectue le même type de développement pour S_{xy},

$$\begin{aligned}
S_{xy} &= \sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y}) \\
&\vdots \\
&= \sum_{t=1}^{n} X_t Y_t - n\overline{XY}
\end{aligned}$$

À l'aide des sommes de carrés corrigés, on peut réécrire la définition de $\hat{\beta}_1$

$$\boxed{\hat{\beta}_1 = \frac{S_{xy}}{S_{xy}}} \tag{2.11}$$

Exemple

On poursuit avec un exemple pour assimiler l'information.

- On dispose des cinq observations suivantes du couple (X_t, Y_t) dans le tableau de gauche ainsi que les éléments calculer nécessaire pour trouver les paramètres dans le tableau de droite.

t	X_t	Y_t
1	2	2
2	3	5
3	6	3
4	9	6
5	12	5
Totaux :	32	21

t	X_t^2	X_tY_t
1	4	4
2	9	15
3	36	18
4	81	54
5	144	60
Totaux :	274	151

À partir des définitions 2.8 et 2.10, on trouve facilement la valeur de $\hat{\beta}_0$ et de $\hat{\beta}_1$.

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^{n} Y_t X_t - n\overline{YX}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} \\ &= \frac{151 - (5)(\frac{21}{5})(\frac{32}{5})}{274 - (5)(\frac{32}{5})^2} \\ &= \frac{83}{346} \\ &\approx 0.2399\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1\overline{X} \\ &= \frac{21}{5} - (\frac{83}{346}) \times (\frac{32}{5}) \\ &\approx 2.6647\end{aligned}$$

On obtient ainsi le modèle de régression suivant :

$$Y_t = 2.6647 + 0.2399X_t + \varepsilon_t$$

t	$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$	$\hat{\varepsilon}_t$
1	3.1445	-1.1445
2	3.3844	1.6156
3	4.1041	-1.1041
4	4.8238	1.1762
5	5.5435	-0.5435

$$\sum_{t=1}^{5} \varepsilon_t \approx -0.0003$$

Execution en R

```
> # dataset
> x <- c(2,3,6,9,12); y <- c(2,5,3,6,5)
> # Estimations des parametres
> reg <- lm(y ~ x)
> # Resume de l'estimation
> summary(reg)
> # Valeurs de Yt
> fitted(reg)
> # Residus
> residuals(reg)
```

Listing 2.1 – Code source en R pour l'exemple

> ***Astuce calculatrice***
>
> La calculatrice TI-30XZ Multiview permet de créé un tableau de donnée et de sortir rapidement et facilement différentes informations sur une régression à partir des données. Tel que :
>
> - $\overline{X}$ et $\overline{Y}$;
> - $\sum_{t=1}^{n} X_t$, $\sum_{t=1}^{n} X_t^2$, $\sum_{t=1}^{n} Y_t$, $\sum_{t=1}^{n} Y_t^2$ et $\sum_{t=1}^{n} X_t Y_t$;
> - $\hat{\beta}_0$ et $\hat{\beta}_1$
>
> Pour de plus ample information, consulter le guide sur les calculatrices.

2.2.2 Caractéristiques du terme d'erreur

On rappel que l'équation du modèle de régression correspond à

$$Y_t = \beta_0 + \beta_1 \times X_t + \varepsilon_t \quad (2.5)$$

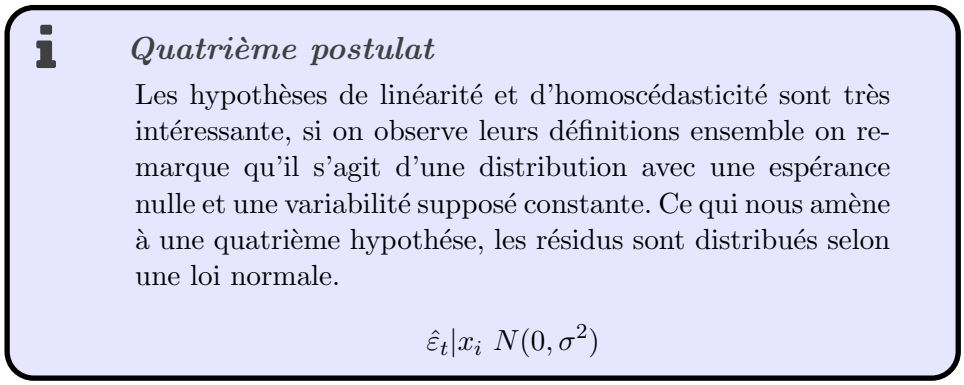
De plus, on sait qu'il s'agit des valeurs moyennes de Y_t en sachat X_t, soit

$$Y_t = E[Y_t|X_t] + \varepsilon_t$$

On peut ainsi formuler les trois postulats [6] suivants,

6. Le postulat est un principe non démontré mais utilisé dans la construction d'une théorie mathématique.

1. $E[\varepsilon_t] = 0$, par définition pour que $E[Y_t] = E[Y_t|X_t]$. Il s'agit de l'hypothèse de linéarité ou d'exogénéité de la variable explicative. On dit qu'elle est exogène si elle n'est pas corrélée au terme d'erreur.
2. $Var(\varepsilon_t) = \sigma^2$, la variance des termes d'erreurs est supposés constante. Il s'agit de l'hypothèse d'homoscédasticité.
3. $Cov(\varepsilon_t, \varepsilon_s) = 0$, pour $t \neq s$,il n'y a pas de corrélation entre les termes d'erreurs. Il s'agit de l'hypothèse d'indépendance des erreurs.



Quatrième postulat

Les hypothèses de linéarité et d'homoscédasticité sont très intéressante, si on observe leurs définitions ensemble on remarque qu'il s'agit d'une distribution avec une espérance nulle et une variabilité supposé constante. Ce qui nous amène à une quatrième hypothése, les résidus sont distribués selon une loi normale.

$$\hat{\varepsilon}_t|x_i \ N(0, \sigma^2)$$

2.3 Propriétés de l'estimateur des moindres carrées (EMC)

2.3.1 Estimateur sans biais

On rappel qu'un estimateur est dit sans biais lorsque son espérance est égale à la valeur vrai du paramètre [7].

$$\begin{aligned} E[\hat{\beta}_1] &= E\left[\frac{\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}\right] \\ &= \frac{\sum_{t=1}^{n}(X_t - \overline{X})E[Y_t - \overline{Y}]}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \\ &= \frac{\sum_{t=1}^{n}(X_t - \overline{X})(E[Y_t] - E[\overline{Y}])}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \end{aligned}$$

7. note thomas

De l'équation 2.5, et avec le postulat 1, on sait que

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 \times X_t + \varepsilon_t \\ E[Y_t] &= E[\beta_0 + \beta_1 \times X_t] + E[\varepsilon_t] \\ &\overset{1}{=} \beta_0 + \beta_1 \times X_t + 0 \end{aligned}$$

On applique le même raisonement pour l'espérance de $\overline{Y}$.

$$\begin{aligned} E[\hat{\beta}_1] &= \frac{\sum_{t=1}^{n}(X_t - \overline{X})(E[Y_t] - E[\overline{Y}])}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \\ &= \frac{\sum_{t=1}^{n}(X_t - \overline{X})(\beta_0 + \beta_1 \times X_t - \beta_0 - \beta_1\overline{X_t})}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \\ &= \frac{\sum_{t=1}^{n}(X_t - \overline{X})\beta_1(X_t - \overline{X_t})}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \\ &= \beta_1 \frac{\sum_{t=1}^{n}(X_t - \overline{X})^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2} \\ E[\hat{\beta}_1] &= \beta_1 \end{aligned}$$

Par conséquent,

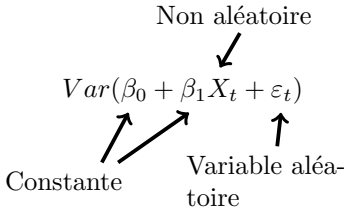
$$\begin{aligned} E[\hat{\beta}_0] &= E[\overline{Y} - \hat{\beta}_1\overline{X}] \\ &= E[\overline{Y}] - \overline{X}E[\hat{\beta}_1] \\ &= \beta_0 + \beta_1\overline{X} - \overline{X}\beta_1 \\ E[\hat{\beta}_0] &= \beta_0 \end{aligned}$$

On peut ainsi conclure que les deux estimateurs des paramètres sont sans biais.

2.3.2 Variances et covariances des estimateurs

On s'intéresse aux variances et aux covariances des estimateurs, cette deuxième propriété ainsi que la première nous permetteras de déduire une conclusion en lien avec le quatrième postulat.

$$
\begin{aligned}
Var(\hat{\beta}_1) &= Var\left(\frac{\sum_{t=1}^{n}(X_t-\overline{X})(Y_t-\overline{Y})}{\sum_{t=1}^{n}(X_t-\overline{X})^2}\right) \\
&= \frac{Var\big(\sum_{t=1}^{n}(X_t-\overline{X})Y_t - \sum_{t=1}^{n}(X_t-\overline{X})\overline{Y}\big)}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2} \\
&= \frac{Var\big(\sum_{t=1}^{n}(X_t-\overline{X})Y_t\big) + Var\big(\sum_{t=1}^{n}(X_t-\overline{X})\overline{Y}\big)}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2} \\
&= \frac{\sum_{t=1}^{n}(X_t-\overline{X})^2 Var(Y_t) + Var\big(\overline{Y}(n\overline{X}-n\overline{X})\big)}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2} \\
&= \frac{\sum_{t=1}^{n}(X_t-\overline{X})^2 Var(\beta_0+\beta_1 X_t+\varepsilon_t)+0}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2}
\end{aligned}
$$



$$
\begin{aligned}
&= \frac{\sum_{t=1}^{n}(X_t-\overline{X})^2 Var(\varepsilon_t)}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2} \\
&\overset{2}{=} \frac{\sum_{t=1}^{n}(X_t-\overline{X})^2\sigma^2}{\big(\sum_{t=1}^{n}(X_t-\overline{X})^2\big)^2}
\end{aligned}
$$

$$
\boxed{Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{t=1}^{n}(X_t-\overline{X})^2}} \tag{2.12}
$$

$$
\begin{aligned}
Var(\hat{\beta}_0) &= Var(\overline{Y} - \hat{\beta}_1\overline{X}) \\
&= Var(\overline{Y}) + Var(\hat{\beta}_1\overline{X}) - 2Cov(\overline{Y}, \hat{\beta}_1\overline{X}) \\
&= Var\Big(\frac{\sum_{t=1}^{n} Y_t}{n}\Big) + \overline{X}^2 Var(\hat{\beta}_1) - 2\overline{X}Cov(\overline{Y}, \hat{\beta}_1) \\
&= \frac{n \times Var(Y_t)}{n^2} + \overline{X}^2\left(\frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}\right) - 2\overline{X}Cov(\overline{Y}, \hat{\beta}_1)
\end{aligned}
$$

$$
\begin{aligned}
Cov(\overline{Y}, \hat{\beta}_1) &= Cov\left(\frac{\sum_{t=1}^{n} Y_t}{n}, \frac{\sum_{s=1}^{n}(X_s - \overline{X})(Y_s - \overline{Y})}{\sum_{s=1}^{n}(X_s - \overline{X})^2}\right) \\
&= \frac{1}{n}\frac{1}{\sum_{s=1}^{n}(X_s - \overline{X})^2} Cov\left(\sum_{t=1}^{n} Y_t, \sum_{s=1}^{n}(X_s - \overline{X})Y_s - \overline{Y}\sum_{s=1}^{n}(X_s - \overline{X})\right) \\
&= \frac{1}{n}\frac{1}{\sum_{s=1}^{n}(X_s - \overline{X})^2} \sum_{t=1}^{n}\sum_{s=1}^{n}(X_s - \overline{X})Cov(Y_t, Y_s) \\
&= \frac{1}{n}\frac{1}{\sum_{s=1}^{n}(X_s - \overline{X})^2}\left(\sum_{\substack{t=1 \\ :t\neq s}}^{n}\sum_{s=1}^{n}(X_s - \overline{X}) \times 0 + \sum_{\substack{t=1 \\ :t=s}}^{n}\sum_{s=1}^{n}\sigma^2\right) \\
&= \frac{1}{n}\frac{1}{\sum_{s=1}^{n}(X_s - \overline{X})^2}\sigma^2\left(\sum_{t=1}^{n}(X_t - \overline{X})\right) \\
&= \frac{1}{n}\frac{1}{\sum_{s=1}^{n}(X_s - \overline{X})^2}\sigma^2\left(\sum_{t=1}^{n}(n\overline{X} - n\overline{X})\right) \\
&= 0
\end{aligned}
$$

$$
\boxed{Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \overline{X}^2\left(\frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}\right)} \tag{2.13}
$$

Finalement, pour la covariance entre $\hat{\beta}_0$ et $\hat{\beta}_1$

$$
\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\overline{Y} - \hat{\beta}_1\overline{X}, \hat{\beta}_1) \\
&= Cov(\overline{Y}, \hat{\beta}_1) - \overline{X}Var(\hat{\beta}_1) \\
&= 0 - \overline{X}\frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}
\end{aligned}
$$

$$\boxed{Cov(\hat{\beta}_0, \hat{\beta}_1) = -\overline{X}\frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}} \tag{2.14}$$

Résumé des propriétés des estimateurs

Les équations 2.13 et 2.12 ainsi que le postulat 4 à la section 2.2.2 nous permettent de conclure que

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{n} + \overline{X}^2\left(\frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}\right)\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2})\right)$$

2.3.3 Optimalité

Le théorème de Gauss-Markor établit que l'estimateur des moindres carrés est l'estimateur non biaisé à variance minimale.

Notions importantes de la preuve :

1. Considérer l'estimateur $\Theta^* = \sum_{t=1}^{n} C_t \times Y_t$
2. Minimiser $Var(\Theta^*)$ sous la contrainte que $E[\Theta^*] = \beta$; où

$$\beta = \begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix}$$

2.4 Régression passant par l'origine

Dans certaines situations, il est possible que l'on souhaite forcer la droite de régression à passer par l'origine. Voici un exemple de situation ou il est plus logique de forcer le modèle,

X_t : Nombre de Km parcouru$t$
Y_t : Consommation d'essence en L d'une voiture$t$

Il est plus logique d'avoir une consommation de 0 L pour une distance de 0 Km.

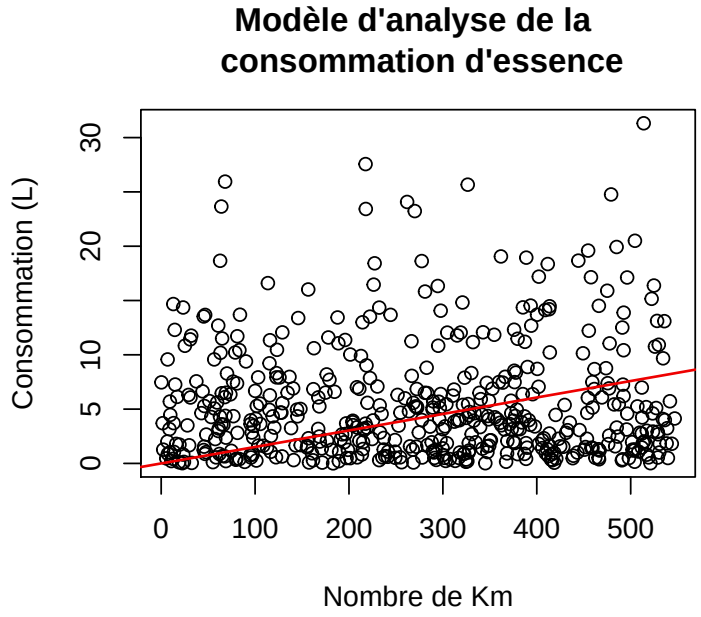
Dans ce cas, on peut postuler le modèle suivant :

$$Y_t = \beta \times X_t + \varepsilon_t \tag{2.15}$$

On peut démontrer par le même raisonnement qu'à la section 2.2.1 que de minimisation du paramètre β correspond à

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2}} \tag{2.16}$$

On reprent l'exemple énoncer plus haut, voici le modèle représenter graphiquement :



Code R

Voici le code R permettant de créé un modèle linéaire simple avec une droite passant par l'origine.

```
> # dataset
> # X Km parcourus
> # Y consommation essence en L
> simul <- 500
> alpha <- 1
> beta <- alpha/5.1
> y <- rgamma(simul, alpha, beta)
> x <- runif(simul, 0, 550)
```

```
> # Estimation de beta
> reg <- lm(y ~ x - 1)
> plot(x, y, xlab = "Nombre de Km", ylab = "Consommation (L)",
    main= "Modele d'analyse de la \n consommation d'essence")
> abline(reg, col="red2", lwd = 1.5)
```

Listing 2.2 – Code source en R pour l'exemple

2.5 Analyse de la variance