Régression linéaire multiple

- Fondements et interprétation géométrique
- Estimation
- Loi des estimateurs
- Pratique en R
- Sélection de modèles

Généralités sur la régression

- Comment exprimer le fait que *Y* dépende de *X* à travers son espérance ?
- $E[Y|X=x]=f(X)$
- Si on note $\varepsilon = Y - E[Y|X=x]$

  (« erreur » supposée indépendante de *X*), alors $E[\varepsilon]=0$

- Du coup, $var(\varepsilon)=var(Y|X=x)\overset{\text{notation}}{=}\sigma^2$

  à supposer que var(*Y*) existe.

- Synthèse : $Y=f(X)+\varepsilon$ où $E[\varepsilon]=0$ et $var(\varepsilon)=\sigma^2$
- Même en admettant que ce modèle soit pertinent, comment trouver *f* ?

Approximation de Y par $f(X)$

- $E[Y|X]$ est la meilleure approximation de Y par une fonction $f(X)$:
  - au sens de la distance² $E[(A-B)^2]$
  - pour les fonctions f telles que $E[(Y-f(X))^2]$ existe

  ($\Leftrightarrow E[f(X)^2]$ existe $\Leftrightarrow var[f(X)^2]$ existe)
- Rappel : $E[Y|X=x]$ est une certaine fonction $f(x)$

  $E[Y|X]$ est $f(X)$ (variable aléatoire)
- Comment trouver f ?
  - chercher dans une famille de fonctions, par exemple fonctions affines de X.

Interprétation cas X déterministe

Réécrire le résultat précédent dans le cas où X est déterministe.

- Interprétez ce résultat. Le connaissiez-vous déjà ? Faire un lien avec var(Y)
- À faire par groupes de deux étudiants. Durée : 3 min

Propriété de $E[Y|X]$: décomposition de var(Y)

- Montrer que

$$\underbrace{E[(Y-E[Y])^2]}_{\mathrm{var}(Y)}=\underbrace{E[(Y-E[Y|X])^2]}_{E[\mathrm{var}(Y|X)]}+\underbrace{E[(E[Y|X]-E[Y])^2]}_{\mathrm{var}(E[Y|X])}$$

- Interpréter le résultat en prenant :
  - X et Y indépendants
  - $Y = f(X)$
  - $Y=f(\underbrace{X)+\varepsilon}_{\text{indépendants}}$

Interprétation : décomposition de la variance

- Que donne $\mathrm{var}(Y) = E[\mathrm{var}(Y|X)] + \mathrm{var}(E[Y|X])$?
  - pour X et Y indépendants :
  - pour $Y = f(X)$
  - pour

Fondements probabilistes de la régression multiple

- Les résultats précédents se généralisent à l'approximation de *Y* par $f(X_1,\dots,X_p)$

indépendant des $X_1,\dots,X_p$

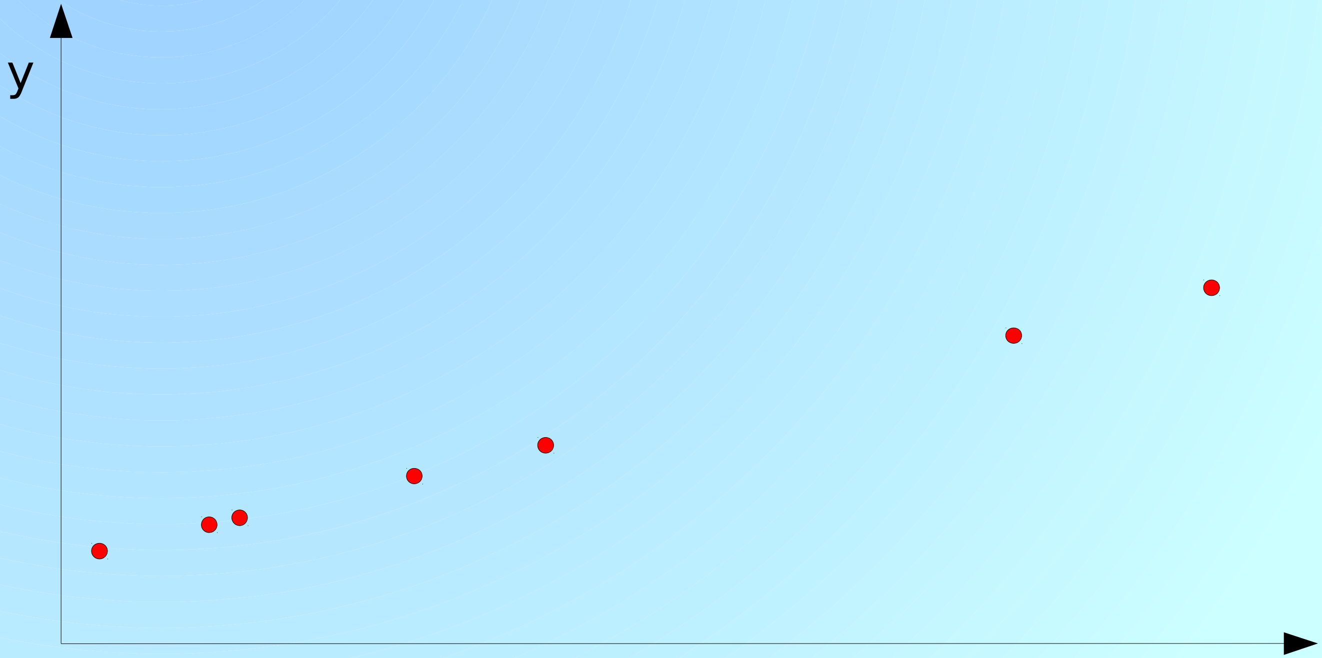
$$Y = f(X_1,\dots,X_p) + \varepsilon \qquad \text{où} \quad E[\varepsilon]=0 \qquad \text{et } var(\varepsilon)=\sigma^2$$

- Une fois de plus, meilleure approximation de *Y* par une fonction $f(X_1,\dots,X_p)$ de variance finie :

$$\underset{f(X_1,\dots,X_p)}{\operatorname{argmin}} \quad E[(Y-f(X_1,\dots,X_p))^2] = E[Y|X_1,\dots,X_p]$$

Régression linéaire simple

- Je dispose des données suivantes :



Est-ce que $E[Y|X=x]$ vaut $ax+b$?

Que signifie $Y = \beta_1 x + \beta_0 + \varepsilon$?

- À comprendre comme :

« je choisis d'approximer *Y* par une fonction affine de *x* »

qui est un plan de l'espace des fonctions de *x*.

Je vais donc chercher la meilleure fonction affine de *X* qui approxime *Y* :

en principe c'est $\frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)} X + E[Y] - \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)} E[X]$

mais on ne connaît pas la loi de (*X*, *Y*)

- En pratique : on va estimer β_1 et β_0 à partir d'un échantillon $(x_i, y_i)_{i=1,\ldots,n}$

Régression linéaire multiple

- Modèle statistique de régression linéaire multiple :

$$Y_i = \beta_p x_{p,i} + \ldots + \beta_1 x_{1,i} + \beta_0 + \varepsilon_i \quad \text{où } E[\varepsilon_i] = 0 \text{ et } var(\varepsilon_i) = \sigma^2$$

- Modèle paramétrique à p+2 paramètres $\beta_p, \ldots, \beta_1, \beta_0$ et σ^2 à estimer à partir de n observations y_i et régresseurs $x_{p,i}$,... , $x_{1,i}$ (approche statistique)

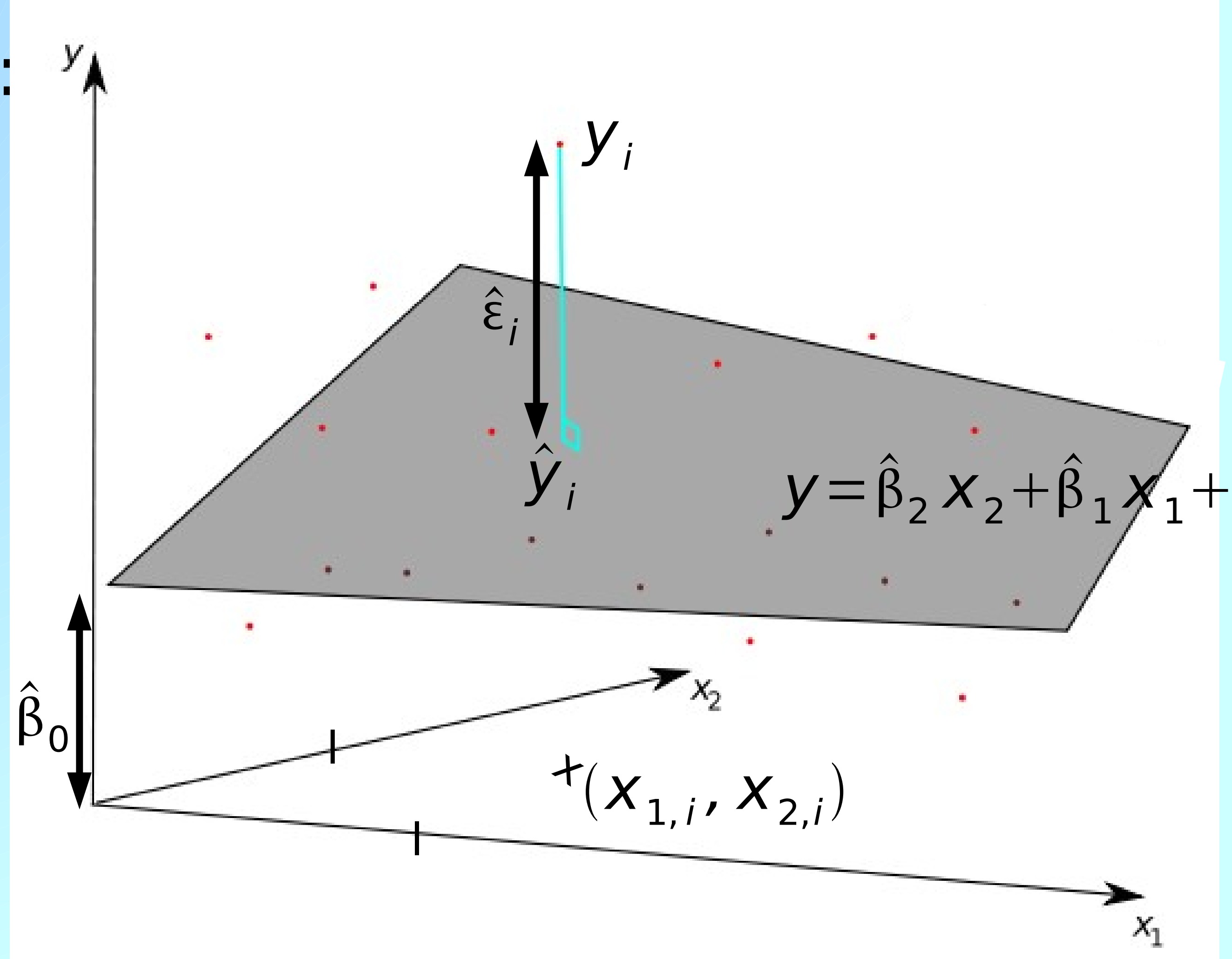
- Notations :

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p+1} \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{p,1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & \cdots & x_{p,n} \end{pmatrix}$$

$$\boxed{\Rightarrow Y = X\beta + \varepsilon}$$

Interprétation géométrique

- On essaie d'ajuster un hyperplan à un nuage de points dans $\mathbb{R}^{p+1}$
  - Ici pour p=2 :



ε écart entre l'observation et l'hyperplan (espérance conditionnelle)
dispersion de cet écart caractérisée par σ^2

Estimation par moindre carrés pour β

- En l'absence d'hypothèse sur la loi des résidus (autre que sur leur espérance et variance-covariance)
- Minimisation de la moyenne des écarts entre les données y_i et l'hyperplan affine d'équation

$$\left\{ y = \beta_p x_p + \ldots + \beta_1 x_1 + \beta_0 \,\middle|\, (y, x_1, \ldots, x_p) \in \mathbb{R}^{p+1} \right\}$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \frac{1}{n} \| Y - X\beta \|^2$$

- $X\hat{\beta} = \Pi_{E_X}(Y)$ où $E_X = \{ X\beta \mid \beta \in \mathbb{R}^{p+1} \}$; Π_{E_X} projection orthogonale sur E_X
- Écrire que $Y - \Pi_{E_X}(Y)$ est orthogonal à tout élément de E_X.
- En déduire ${}^t X Y = {}^t X X \hat{\beta}$.
- En expliciter toutes les conséquences sur l'estimation de β.
  - À faire par groupes de deux étudiants. Durée : 10 min

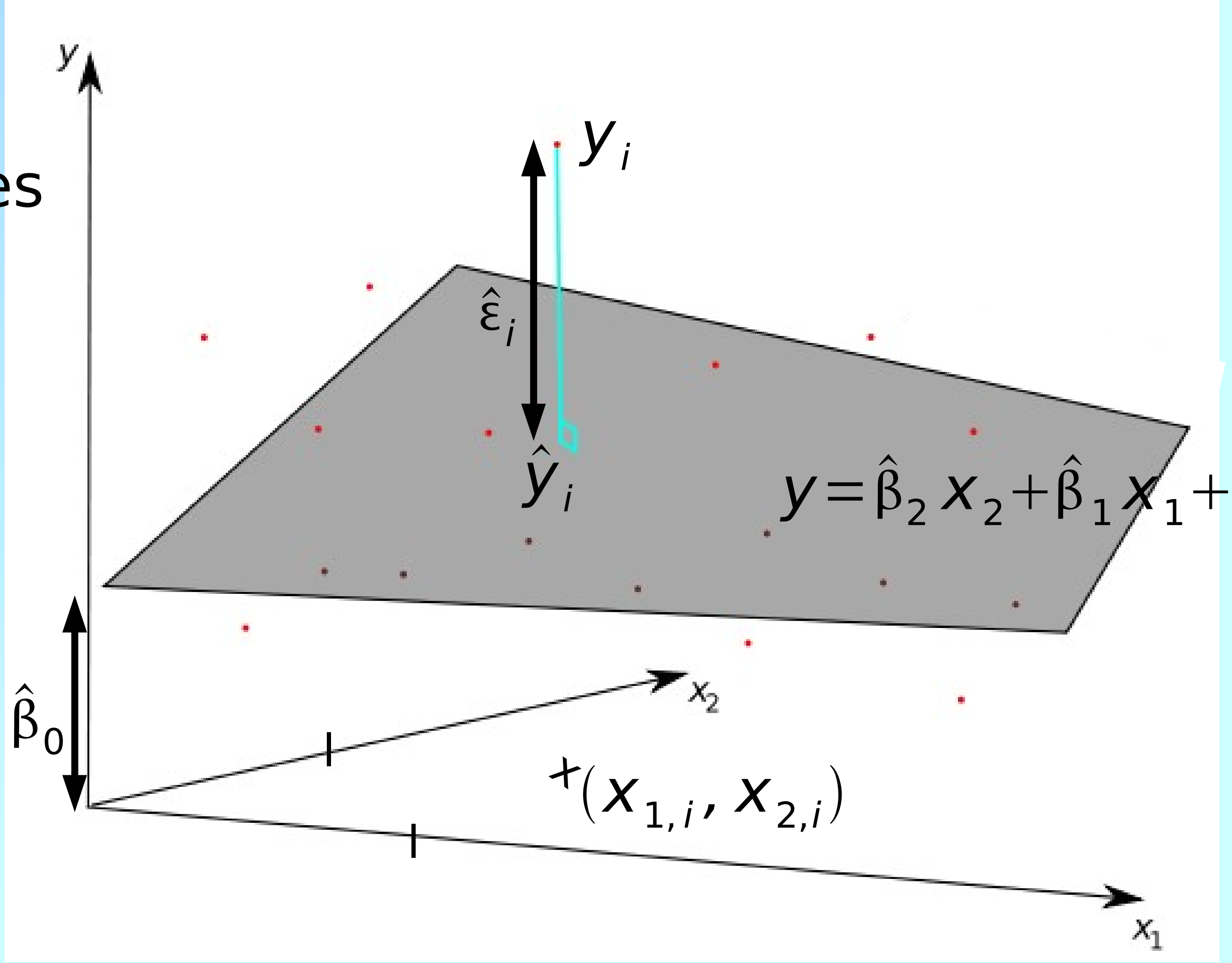
Estimation de la variance

- Idée intuitive :

$$\varepsilon_i \approx \hat{\varepsilon}_i = Y_i - \underbrace{\hat{B}_p x_{pi} - \ldots - \hat{B}_1 x_{1i} - \hat{B}_0}_{(X\hat{B})_i = \hat{Y}_i} \quad \text{donc} \quad \mathrm{var}(\varepsilon_i) = \sigma^2 \approx \mathrm{var}(\hat{\varepsilon}_i)$$

- On estime σ^2 par la variance empirique des résidus empiriques, notée $S^2_{Y|x}$ (variance résiduelle)

$$S^2_{Y|x} = \frac{1}{n}\|Y - X\hat{B}\|^2 = \frac{1}{n}\|\hat{\varepsilon}\|^2$$



«Débiaisage» de l'estimateur de σ²

- $S_{Y|x}^2 = \frac{1}{n} \|Y - X\hat{B}\|^2$ est un estimateur biaisé de σ²
- Propriété : $E[{}^tY\,A\,Y] = tr(A\,cov(Y)) + {}^tE[Y]\,A\,E[Y]$
- $E[\|Y - X\hat{B}\|^2] = E[\|\Pi_{E_x^\perp} Y\|^2] = \sigma^2 \mathrm{tr}(\Pi_{E_x^\perp}) + {}^tE[Y]\,\Pi_{E_x^\perp} \underbrace{E[Y]}_{\in E_x}$
- donc $\hat{\sigma}^2 = \frac{1}{n-(p+1)} \|Y - X\hat{B}\|^2$ est un estimateur sans biais de σ²
- Résultat de nature essentiellement géométrique
- Remarquer que p+1 est la dimension de β

« Rappels » : vecteurs gaussiens

- Loi normale multivariée $\mathcal{N}(\mu, \Sigma)$
- ???

Hypothèse gaussienne sur la loi de ε

- ε est un *vecteur gaussien*

$\Rightarrow Y = X\beta + \varepsilon$ est également un vecteur gaussien

$$\mathcal{N}_n(X\beta, \sigma^2 I_n)$$

- On obtient la loi des estimateurs.

En particulier : $\hat{B} \sim \mathcal{N}_{p+1}\left(\beta, \sigma^2 ({}^t X X)^{-1}\right)$

- Possibilité de faire des intervalles de confiance, tests, validation du modèle

Interpréter les sorties de R

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4)

Residuals:
    Min       1Q   Median       3Q      Max
-1092.9   -263.2   -130.2    197.3   1883.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
x1             27.641      5.429   5.091  0.00224 **
x2          12529.768    400.067  31.319 7.04e-08 ***
x3           2553.211    530.669   4.811  0.00297 **
x4           -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 and 6 DF,  p-value: 1.372e-07
```

commande / formule définissant le modèle et les données

Ici : $Y = \beta_1 x_1 + \ldots + \beta_4 x_4 + \beta_0 + \varepsilon$

Interpréter les sorties de R

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4)

Residuals:
    Min      1Q  Median      3Q     Max
-1092.9  -263.2  -130.2   197.3  1883.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
x1             27.641      5.429   5.091  0.00224 **
x2          12529.768    400.067  31.319 7.04e-08 ***
x3           2553.211    530.669   4.811  0.00297 **
x4           -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 and 6 DF,  p-value: 1.372e-07
```

Résidus empiriques : statistique descriptive (médiane et quartiles empiriques)

Quid de la moyenne empirique des résidus ?

- Pourquoi n'est-elle pas indiquée dans R ?
- À $\frac{1}{n}$ près c'est $< \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} ; \Pi_{E_X^\perp} Y >$
- En général, $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in E_X$ (c'est la 1$^{\text{ère}}$ colonne de X)
- Les résidus empiriques sont de moyenne empirique nulle par construction !

Interpréter les sorties de R

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4)


Residuals:
    Min       1Q  Median       3Q      Max
-1092.9   -263.2  -130.2   197.3   1883.1


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
x1             27.641      5.429   5.091  0.00224 **
x2          12529.768    400.067  31.319 7.04e-08 ***
x3           2553.211    530.669   4.811  0.00297 **
x4           -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 and 6 DF,  p-value: 1.372e-07
```

Pour chaque régresseur : estimation, estimation de l'écart-type de sa loi, statistique de test de nullité et p-valeur de ce test

Interpréter les sorties de R

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4)


Residuals:
    Min      1Q  Median      3Q     Max
-1092.9  -263.2  -130.2   197.3  1883.1


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
x1             27.641      5.429   5.091  0.00224 **
x2          12529.768    400.067  31.319 7.04e-08 ***
x3           2553.211    530.669   4.811  0.00297 **
x4           -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 and 6 DF,  p-value: 1.372e-07
```

le lien n'est pas évident !

p-valeur du test de pertinence globale (prochain cours)

R : cas particulier de prédicteurs qualitatifs

```
> myVar = factor(myVar)
```

Transforme myVar en variable catégorielle : modalités 1, 2, 3

```
Call:
lm(formula = y ~ myVar + x)
```

Modèle $Y_i = \mu_{myVar_i} + \beta x_i + \varepsilon_i$

réécrit $Y_i = \beta_0 + \nu_{myVar_i} + \beta x_i + \varepsilon_i$

avec $myVar_i \neq 1 ; \nu_1 = 0$

voir ANOVA pour les explications

```
Residuals:
    Min      1Q  Median      3Q     Max
-1092.9  -263.2  -130.2   197.3  1883.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
myVar2          2.641    235.429  55.09    0.9224
myVar3      12529.768    400.067  31.319 7.04e-08 ***
x            -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 ...
```

pas de ligne myVar1 :

$\hat{\mu}_1 = \hat{\beta}_0 ; \nu_1 = 0$

ligne myVar3 : $\hat{\nu}_3$

$\hat{\mu}_3 = \hat{\beta}_0 + \hat{\nu}_3$

ligne x : $\hat{\beta}$

(pas de subtilité)

Indépendance de $\hat{B}^2$ et $\hat{\sigma}^2$

- Découle du fait que $X\hat{B}=\Pi_{E_X}Y$ et $Y-X\hat{B}=\Pi_{E_X^\perp}Y$
- Si U est un vecteur gaussien $\mathcal{N}(0,\sigma^2 I)$ et Π un projecteur orthogonal, la matrice de covariance de $\begin{pmatrix}\Pi U\\(I-\Pi)U\end{pmatrix}$ est $\begin{bmatrix}\Pi^t\Pi & \Pi^t(I-\Pi)\\(I-\Pi)^t\Pi & (I-\Pi)^t(I-\Pi)\end{bmatrix}=\begin{bmatrix}\Pi & 0\\0 & I-\Pi\end{bmatrix}$

donc ΠU et U-ΠU sont non-corrélés, d'où

indépendants

- Voir que $\hat{B}=\left[\left({}^tX X\right)^{-1}{}^tX\right]X\hat{B}$: indépendant de $\dfrac{\|Y-X\hat{B}\|^2}{n-p-1}$
- Indépendance essentielle pour construire des intervalles de confiance pour les β_i

Loi de $\hat{\sigma}^2$

- À des constantes près, revient à déterminer la loi de
$$\|Y - X\hat{B}\|^2 = \|\Pi_{E_X^\perp} Y\|^2$$
- Soit Π un projecteur orthogonal sur un sous-espace vectoriel de dimension l et U de loi $\mathcal{N}(0, I)$:
  - $\Pi U \sim \mathcal{N}(0, \underbrace{\Pi\,{}^t\Pi}_{\Pi})$
  - Π diagonalisable dans une B.O.N. : $C\,\Pi\,{}^tC = \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix}$ où seuls l (= rang(Π)= n - p -1 ici) V_i sont non nuls p.s.

  (ces V_i non nuls constituent un échantillon $\mathcal{N}(0,1)$)

  C : matrice d'isométrie. $C\,{}^tC$ = I

$$V = C\,\Pi\,U \sim \mathcal{N}(0, C\,\Pi\,{}^tC) \qquad \|\Pi U\|^2 = \|C\,\Pi\,U\|^2 = \sum_i V_i^2$$

$$\|\Pi U\|^2 \sim \chi_l^2$$

Loi de $\hat{\sigma}^2$: en un mot

- $\dfrac{n-p-1}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p-1}$

Intervalles de confiance pour les β_i

- Si on arrive à établir la pertinence de la régression, alors on peut interpréter les paramètres.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 52317.831  12237.362   4.275  0.00523 **
x1             27.641      5.429   5.091  0.00224 **
x2          12529.768    400.067  31.319 7.04e-08 ***
x3           2553.211    530.669   4.811  0.00297 **
x4           -234.237     13.268 -17.654 2.12e-06 ***
---
Residual standard error: 970.6 on 6 degrees of freedom
Multiple R-Squared: 0.9967,     Adjusted R-squared: 0.9946
F-statistic: 459.8 on 4 and 6 DF,  p-value: 1.372e-07
```

- Formule d'un intervalle de confiance pour β_1 au seuil 5%
- Valeur numérique
- Test de H0 : «$\beta_1 = 0$» contre H1 : «$\beta_1 \neq 0$» au seuil 5%
- Interprétation du résultat du test

  - Cf. Exercice 1 et TP 1

  - À faire par groupes de deux étudiants. Durée : 8 min.
  - Rappel : cas i.i.d. gaussien

Expression des intervalles de confiance

Test d'hypothèses

- Test de H_0 : « $\beta_1 = 0$ » contre H1 : « $\beta_1 \neq 0$ »
- Rappel : $\dfrac{\hat{B}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 ({}^t X X)^{-1}_{22}}} \sim \mathcal{S}t(n-p-1)$
- Par dualité entre intervalle de confiance et test, on sait déjà qu'on rejette H_0 au seuil 5%
- Région critique $W = \left\{ \dfrac{|\hat{B}_1|}{\sqrt{\hat{\sigma}^2 ({}^t X X)^{-1}_{22}}} > F^{-1}_{\mathcal{S}t(n-p-1)}\left(1 - \dfrac{\alpha}{2}\right) \right\}$
- On retrouve la p-valeur du test

$$2 \times \left(1 - F_{\mathcal{S}t(n-p-1)} \left(\frac{|\hat{B}_1|}{\sqrt{\hat{\sigma}^2 ({}^t X X)^{-1}_{22}}} \right) \right)$$

$$\begin{cases} H_0 & Y = \cancel{\beta_1 x_1} + \ldots + \beta_4 x_4 + \beta_0 + \varepsilon \\ H_1 & Y = \beta_1 x_1 + \ldots + \beta_4 x_4 + \beta_0 + \varepsilon \end{cases}$$

- Ici : 0,00224

Choix des régresseurs

- Cf. Exercice

Comprendre la formule de décomposition de la variance = Pythagore

- Version probabiliste (pas besoin de données !)

$$\mathrm{var}(Y) = \underbrace{E[\mathrm{var}(Y|X)]}_{\sigma^2} + \underbrace{\mathrm{var}(E[Y|X])}_{\mathrm{var}(f(X))}$$

$Y = f(X) + \varepsilon$

- Version empirique (géométrie dans $\mathbb{R}^n$)

$$\frac{1}{n}\sum_i (y_i - \bar{y}_n)^2 = \frac{1}{n}\sum_i (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_n)^2 + \frac{1}{n}\sum_i (\hat{Y}_i - \bar{\hat{Y}}_n)^2$$

variance empirique des observations s_y^2	variance empirique des résidus empiriques	variance empirique des valeurs prédites $\hat{f}(x_i)$
• indépendante du modèle de régression choisi	• doit être «relativement faible» si la régression est pertinente	• doit être «relativement élevée» si la régression est pertinente • liée à R^2_{xy} : vaut $R^2_{xy}\, s^2_y$

E.M.V. de β

- Fonction de vraisemblance :

$$\mathcal{L}_{y_1,\dots,y_n}(\beta,\sigma^2)=\prod_i f_{\mathcal{N}(<x_i;\beta>,\sigma^2)}(y_i)\Rightarrow$$

$y_1,\dots,y_n$ est « comme un échantillon gaussien d'espérance $<x_i; \beta>$ »

$$\log \mathcal{L}_{y_1,\dots,y_n}(\beta,\sigma^2)=\sum_i \left\{-\frac{1}{2}\log(2\pi\sigma^2)-\frac{(y_i-<x_i;\beta>)^2}{(2\sigma^2)}\right\}$$

y_i - $<x_i; \hat{\beta}> \hat{=} \varepsilon_i$

$$=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{\|Y-X\beta\|^2}{(2\sigma^2)}$$

indépendant de β

maximum pour $\beta = \Pi_{EX}(Y)$

- Donc E.M.V. et E.M.C. de β coïncident dans le cas gaussien : même notation $\hat{B}$
- $\dfrac{\partial \mathcal{L}_{y_1,\dots,y_n}(\hat{B},\sigma^2)}{\partial\sigma^2}$ est nulle en $\sigma^2=\dfrac{1}{n}\|Y-X\hat{B}\|^2$

(estimateur biaisé, débiaisé en $\hat{\sigma}^2=\dfrac{1}{n-p-1}\|Y-X\hat{B}\|^2$)

Efficacité de $\hat{B}^2$

- On montre que $\hat{B}$ est un estimateur sans biais de β efficace

$$\nabla_\beta \log \mathcal{L}_{y_1,\dots,y_n}(\beta,\sigma^2) = -\frac{1}{2\sigma^2}\nabla_\beta \|Y - X\beta\|^2 = \frac{1}{\sigma^2}\,{}^tX(Y - X\beta)$$

de matrice de covariance $\frac{1}{\sigma^2}\,{}^tX\,X = cov\left(\hat{B}\right)^{-1}$

- $\hat{B}$ est donc l'estimateur sans biais de variance minimale de β

Coefficient de corrélation linéaire multiple

- Définition : $$R = \sup_{(a_1, \ldots, a_p) \in \mathbb{R}^p} R_{Y, \sum_j a_j X_j}$$
- On se ramène à des régressions linéaires simples (en dimension 1)
- Propriété : $R^2 = \frac{S_{\hat{Y}}^2}{S_Y^2}$ - d'autant plus proche de 1 que la variance expliquée est proche de la variance totale (ratio de variance expliquée)

Test de pertinence de la régression multiple

- Test de H0 : «$\beta_1 = \ldots = \beta_p = 0$» contre H1 : « $\exists i \geq 1, \ \beta_i \neq 0$ »

ne concerne pas β_0 qui n'est lié à aucun régresseur

- H1 s'interprète comme : il y a au moins un facteur qui a un effet sur (l'espérance de) Y

- Région critique :

$$W = \left\{ \frac{n-p-1}{p} \frac{R^2}{1-R^2} > F^{-1}_{\mathcal{F}(p,n-p-1)}(1-\alpha) \right\}$$

- Preuve : cas particulier de test d'hypothèse linéaire

Qualité de la régression multiple

- Contexte :
  - valider un modèle de régression linéaire multiple
  - choisir les régresseurs

Quand on augmente le nombre de régresseurs

- le coefficient de corrélation linéaire multiple augmente forcément

oui	non	autres réponses

- la p-valeur du test de pertinence de la régression diminue forcément

oui	non	autres réponses

À faire par groupes de 2 étudiants.
Durée : 7 minutes

Illustration

A

```
Call:
lm(formula = y1 ~ x1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -20.2786     7.6296  -2.658  0.01510 *
x1            1.9029     0.5809   3.276  0.00378 **

Residual standard error: 17.29 on 20 degrees of freedom
Multiple R-Squared: 0.3492
F-statistic: 10.73 on 1 and 20 DF,  p-value: 0.003781
```

B

```
Call :
lm(formula = y1 ~ x1 + x2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.4834     7.7985  -2.498  0.02182 *
x1            1.8830     0.5885   3.199  0.00472 **
x2            1.5596     2.1468   0.726  0.47641
Residual standard error: 17.49 on 19 degrees of freedom
Multiple R-Squared: 0.3668
F-statistic: 5.502 on 2 and 19 DF,  p-value: 0.01303
```

Script :

```
n <- 22
x1 <- seq(1,n)
x2 <- 2*rnorm(n)
y1 <- x1 + 1 + 20 * rnorm(n)
```

Devrait conduire, sauf accident dû au hasard, à préférer le modèle A au modèle B

Comparaison de l'effet des régresseurs

- Un statisticien, fumeur et buveur invétéré s'intéresse à un modèle de prévision de l'espérance de vie publié dans un journal médical : $Y = \beta_3 x_3 + \beta_2 x_2 + \beta_1 x_1 + \beta_0 + \underbrace{\varepsilon}_{\mathcal{N}(0,\sigma^2)}$
  où x_1 représente la consommation annuelle de graisses animales, x_2 la consommation de tabac, x_3 celle d'alcool et Y la durée de vie.

- Le journal donne : $\hat{\beta}_3 = -0.0018$, $\hat{\beta}_2 = -0.0008$, $\hat{\beta}_1 = -0.25$, $\hat{\beta}_0 = 79$

  ans / *l* de bière — ans / paquet de cigarette — ans / *kg*

- Cette personne est prête à réduire sa consommation de bière à 0,5 l / jour et à arrêter de fumer, ou de réduire sa consommation de cigarettes à 1 paquet / jour et à arrêter de boire, mais certainement pas les deux à la fois. À choisir, il préférerait arrêter de fumer.

Pour sa santé, que lui conseillez-vous d'arrêter ?

tabac	bière	autres réponses

À faire par groupes de 2 étudiants.
Durée : 3 minutes

Modélisation

- En arrêtant de fumer, son espérance de vie moyenne estimée devient $365\times(\underbrace{\hat\beta_3\times 0{,}5}_{-0.009}+\hat\beta_2\times 0+\hat\beta_1 x_1)+\hat\beta_0$
- En arrêtant de boire, elle devient $365\times(\hat\beta_3\times 0+\underbrace{\hat\beta_2\times 1}_{-0.008}+\hat\beta_1 x_1)+\hat\beta_0$
  - Il semble qu'arrêter de boire soit une meilleure solution. Cependant même en simulant un modèle avec 0,5 β_3 = β_2 = -0,008 on peut très bien obtenir $\hat\beta_3=-0.0018,\ \hat\beta_2=-0.0008$
  - Dans ce cas la personne arrêterait de boire plutôt que de fumer, contre ses préférences, ce qui serait désolant.

Il faut tester si 0,5 β_3 = β_2

- Sous cette hypothèse : $\beta=\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix}=\beta_0\begin{pmatrix}1\\0\\0\\0\end{pmatrix}+\beta_1\begin{pmatrix}0\\1\\0\\0\end{pmatrix}+\beta_2\begin{pmatrix}0\\0\\1\\2\end{pmatrix}$ est dans un sous-espace vectoriel ξ_0 de dimension 3 de $\mathbb{R}^4$

Pour sa santé, que lui conseillez-vous d'arrêter ?

Autres réponses (??)

Tests d'hypothèses linéaires

- Formulation générale : H0 : «$\beta \in \xi_0$» contre H1 : « $\beta \notin \xi_0$»
- Construction du test : $\mathbb{R}^n = E_0 \overset{\perp}{\oplus} E_X^{\perp} \overset{\perp}{\oplus} E_X \cap E_0^{\perp}$

où $E_0 = \left\{ X\beta \mid \beta \in \xi_O \right\}$

- Sous H0, $X\beta \in E_0 \Rightarrow \Pi_{E_X \cap E_0^{\perp}} X\beta = 0$
- Le test est basé sur la comparaison de

$$\|\Pi_{E_X \cap E_0^{\perp}} X\hat{B}\|^2 = \|\Pi_{E_X \cap E_0^{\perp}} Y\|^2 \qquad \underbrace{\|\Pi_{E_X^{\perp}} Y\|^2}_{n \text{ x variance résiduelle}} \quad \text{et de}$$

(en fait le rapport des deux est une fonction

pivotale : statistique indépendante de β et σ^2)

- Il suffit de déterminer la loi du carré de la norme de projections orthogonales sur des sous-espaces orthogonaux

Théorème de Cochran

- Généralisation des théorèmes d'indépendance de $X\hat{B}$ et $Y - X\hat{B}$ et de la loi du χ^2 pour $\frac{1}{\sigma^2}\|Y - X\hat{B}\|^2$
- Hypothèses :
  - $\mathbb{R}^n = E_1 \overset{\perp}{\oplus} \ldots \overset{\perp}{\oplus} E_r$, $\dim(E_i) = d_i$
  - $U \sim \mathcal{N}(0 ; \sigma^2 I)$

- Résultat : les $\frac{1}{\sigma^2}\|\Pi_{E_i} U\|^2$ sont des variables aléatoires indépendantes de loi $\chi^2_{d_i}$
- Corollaire : les rapports $\frac{d_j}{d_i}\frac{\|\Pi_{E_i} U\|^2}{\|\Pi_{E_j} U\|^2}$ sont de loi $\mathcal{F}(d_i, d_j)$

Retour aux tests d'hypothèses linéaires

- $\mathbb{R}^n = \underbrace{E_0}_{\dim : q} \overset{\perp}{\oplus} \underbrace{E_X^\perp}_{n-p-1} \overset{\perp}{\oplus} \underbrace{E_X \cap E_0^\perp}_{p+1-q}$ *N.B.* $\dim(E_0) = \dim(\{X\beta | \beta \in \xi_O\})$ $= \dim(\xi_O)$
- Sous H0, $\frac{n-p-1}{p+1-q} \frac{\|\Pi_{E_X \cap E_0^\perp} Y\|^2}{\|\Pi_{E_X^\perp} Y\|^2} \sim \mathcal{F}(p+1-q, n-p-1)$
- D'où la région critique

$$W = \left\{ \frac{n-p-1}{p+1-q} \frac{\|\Pi_{E_X \cap E_0^\perp} Y\|^2}{\|\Pi_{E_X^\perp} Y\|^2} > F^{-1}_{\mathcal{F}(p+1-q, n-p-1)}(1-\alpha) \right\}$$

- Cas particulier :

test de pertinence

de la régression

$$\xi_O = \mathrm{vect}\left(\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\} \right)$$