GCB140 – Statistiques en ingénierie

Chapitre 12

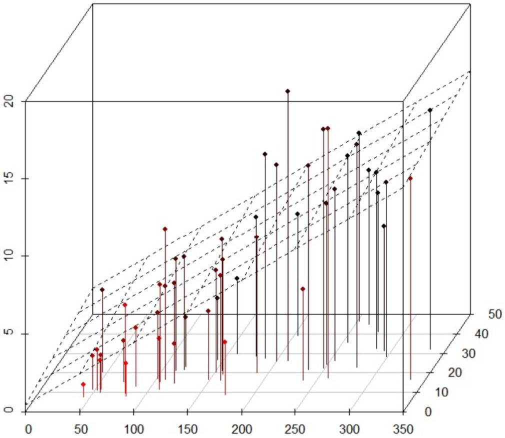
Nasser Sadeghkhani[1]

[1]a.sadeghkhani@usherbrooke.ca

Mars 2017

Outline

A la suite de la régression linéaire simple, cette vignette introduit le modèle linéaire multidimensionnel dans lequel une variable quantitative continue Y est expliquée, modélise, par plusieurs variables quantitatives X_j pour $j = 1; \quad p$.



Cas particulier de modèle linéaire, il constitue la généralisation naturelle d
la régression simple.

Exemple 1

Première écriture du modèle

$$Y = {}_0 + {}_1 X_1 + {}_2 X_2 + \quad ; X_k + \; : \tag{1}$$

L'objectif : estimer les $p = k + 1$ paramètres ${}_0 ; {}_1 ; \quad {}_k$.

Remarque : ^{2} est un autre paramètre inconnu à estimer.

Ici, dans Exemple 1, Pour estimer les parametres, on utilisera la methode des moindres carres, qui consiste a minimiser l'erreur au carre.

Pour le moment, nous acceptons que (avec le logiciel)

Revision de l'algebre lineaire. Pages 1{22

Modèle matriciel

Le Modèle (1) s'écrit :

on peut bien voir que

Donc, l'estimateur des moindres carrés ordinaires (emco) de β (c'est une matrice $p = k + 1$ par 1) est :

$$\hat{\beta} = (X'X)^{-1}X'Y; \tag{2}$$

alors

$$\hat{Y} = X\hat{\beta}. \tag{3}$$

Vérification la régression linéaire sous forme de matrice

Exemple 2 (La droite de regression utilisant la forme matricielle)

Note	10	5	8	1	3	5	9	7	2	4
Heures detude	10	3	7	1	2	4	4	5	1	2

AVOVA, Régression linéaire Multiple

$SSE = Y'Y - \hat{\beta}'X'Y$; $SST = \sum Y_i^2 - n\bar{Y}^2$, SSR=SST-SSE.

On conclut qu'une régression est significative si (on rejette H_0)

$$F_0 > F(k; n-k-1; \alpha)$$

Exemple

L'objectif de cette étude est de modéliser la note obtenue par des échantillons de fromage (Cheddar) lors de tests gustatifs opérés par un jury. Ce test concerne $n = 30$ échantillons de fromage dont la note moyenne doit être modélisée par $p = 3$ variables explicatives :

- `GoutM` : note moyenne de juges
- `Acetic` : log concentration en acide acétique
- `H2S` : log concentration en H2S
- `Lactic` : concentration en acide lactique

```
Analyse de régression : GoutM en fonction de Acetic; H2S; Lactic

Analyse de variance

Source          DL       SC       CM       F       P
Régression       3   4994,5   1664,8   16,22   0,000
Erreur résid    26   2668,4    102,6
Total           29   7662,9

L'équation de régression est
GoutM = - 28,9 + 0,33 Acetic + 3,91 H2S + 19,7 Lactic

Régresseur       Coef   Er-T coef        T       P
Constante      -28,88       19,74    -1,46   0,155
Acetic          0,328       4,460     0,07   0,942
H2S             3,912       1,248     3,13   0,004
Lactic         19,671       8,629     2,28   0,031

S = 10,13             R-carré = 65,2%      R-carré (ajust) = 61,2%
```

Exercices

12.4 An experiment was conducted to determine if the weight of an animal can be predicted after a given period of time on the basis of the initial weight of the animal and the amount of feed that was eaten. The following data, measured in kilograms, were recorded:

Final Weight, y	**Initial Weight, x_1**	**Feed Weight, x_2**
95	42	272
77	33	226
80	33	259
100	45	292
97	39	311
70	36	183
50	32	173
80	41	236
92	40	230
84	38	235

(a) Fit a multiple regression equation of the form

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

(b) Predict the final weight of an animal having an initial weight of 35 kilograms that is given 250 kilograms of feed.

Given the data

x	0	1	2	3	4	5	6	7	8	9
y	9.1	7.3	3.2	4.6	4.8	2.9	5.7	7.1	8.8	10.2

fit a regression curve of the form $\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$ and then estimate $\mu_{Y|2}$.