Régression Multiple.

Introduction:

Le modèle de régression linéaire multiple est l'outil statistique le plus habituellement mise en oeuvre pour l'étude de données multidimensionnelles, cas particulier du modèle linéaire il consistue la généralisation naturelle de la régression simple.

Modèle: Une variable quantitative Y dite à expliquer (ou encore, réponse, exogène, dépendante) est mise en relation avec p variables quantitatives $X^1, X^2, \ldots, X^p$ dites explicatives (ou encore de contrôle, endogènes, indépendantes, régressents).

les données sont supposées provenir de l'observation d'un échantillon statistique de taille n $(n > p+1)$ de $\mathbb{R}^{p+1}$: $(x_i^1, x_i^2, \ldots, x_i^p, y_i)$. $i = 1, \ldots, n$.

l'écriture du modèle linéaire dans cette situation conduit à supposer que l'espérance de Y appartient au sous-espace de $\mathbb{R}^n$ engendré par $\{\mathbb{1}, X^1, X^2, \ldots, X^p\}$ où $\mathbb{1}$ désigne le vecteur de $\mathbb{R}^n$ constitué de "1". c'est-à-dire que les $(p+1)$ variables aléatoires vérifient :

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \ldots \beta_p x_i^p + \varepsilon_i \quad i = 1, 2, \ldots n$$

avec les Hypothèses suivantes:

1- les ε_i sont des termes d'erreur observés, indépendantes et identiquement distribués:

$$E(\varepsilon_i) = 0 \qquad Var(\varepsilon) = \sigma^2 \mathbb{I}.$$

2- les termes x^j sont supposés déterministes (facteurs contrôlés), ou bien l'erreur ε est indépendantes de la distribution conjointe de $X^1, X^2, \dots, X^p$. On écrit dans ce dernier cas que:

$$E(Y/X^1, X^2, \dots, X^p) = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \dots + \beta_p X^p. \text{ et}$$

$$Var(Y/X^1 X^2 \dots X^p) = \sigma^2.$$

3- Les paramètres inconnus $\beta_0, \dots, \beta_p$ sont supposés constants.

4- On option, pour l'étude spécifique des lois des estimateurs, une quatrième hypothèse considère la normalité de la variable d'erreur ε. $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$. Les ε_i sont alors i.i.d de loi $\mathcal{N}(0, \sigma^2)$.

Les données sont rangées une matrice. $X(n \times (p+1))$ de terme général x_i^j, dont la première colonne contient le vecteur $\mathbb{1}$ ($x_0^j = 1$), et dans un vecteur Y de terme général y_i. En notant les vecteurs $\varepsilon = [\varepsilon_1, \dots, \varepsilon_p]'$ et $\beta = [\beta_0, \beta_1 \dots, \beta_p]'$. le modèle s'écrit matriciellement

$$Y = X\beta + \varepsilon.$$

Estimation :

1 Estimation par MC.

l'expression à minimiser sur $\beta \in \mathbb{R}^{p+1}$ s'écrit :

$$\sum_{i=0}^{n} (y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2 - \dots - \beta_p x_i^p)^2 = \|Y - X\beta\|^2$$

$$= (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - 2\beta' X'Y + \beta' X'X\beta.$$

Par dérivation matricielle de la dernière équation on obtient les équations normales.

$$X'Y - X'X\beta = 0.$$

Dont la solution correspond bien à minimum car la matrice hessienne $2X'X$ est semi diffinie-positive

Nous faisons l'hypothèse supplémentaire que la matrice $X'X$ est inversible, c'est-à-dire que la matric X est de rang $(p+1)$ et donc qu'il n'existe pas de colinéarité entre ses colonnes.

Alors. l'estimation des paramètres β_j est donnée par :

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

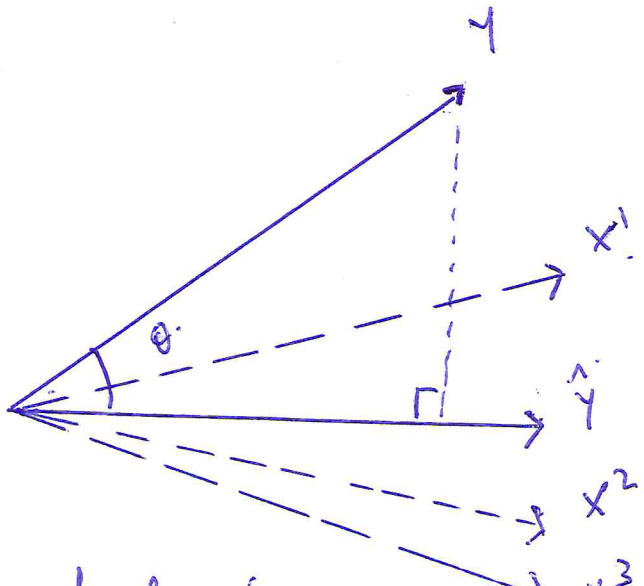
Et les valeurs ajustées (ou estimées, prédites) de y pour expression $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$.

Où $H = X(X'X)^{-1}X'$ est appelé "hat matrix" ; elle met un chapeau Y. Géométriquement, c'est la matrice de projection.

Orthogonale dans $\mathbb{R}^n$ sur le sous espace vect(x) engendré par les vecteurs colonnes de X.

On note $e = y - \hat{y} = y - X\hat{\beta} = (I - H)\,y$.

le vecteur des résidus; c'est la projection de y sur le sous-espace orthogonal de vect(x) dans $\mathbb{R}^n$



Géométriquement la régression est la projection $\hat{y}$ de y sur l'espace vect $\{1, x^1, x^2, \ldots, x^p\}$; de plus $R^2 = \cos^2\theta$.

.2 Propriétés :

Les estimateurs des MC $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_p$ sont des estimateurs sans biais : $E(\hat{\beta}) = \beta$

On montre que la matrice de covariance des estimateurs se met sous la forme :

$$E\left((\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right) = \sigma^2 (X'X)^{-1}.$$

celle des prédicteurs est

$$E\left((\hat{y}-X\beta)(\hat{y}-X\beta)'\right) = \sigma^2 H.$$

et celle des estimateurs des résidus est :

$$E((e-\varepsilon)(e-\varepsilon)') = \sigma^2 (I - H).$$

Tandis qu'un estimateur sans biais de σ^2 est fourni par

$$S^2 = \frac{\|e\|^2}{n-p-1} = \frac{\|Y - X\hat{\beta}\|^2}{n-p-1} = \frac{SS_E}{n-p-1}$$

3. Sommes des carrés :

SS_E est la somme des carrés résidus :

$$SS_E = \|Y - \hat{Y}\|^2 = \|e\|^2.$$

On définit également la somme totale des carrés :

$$SS_T = \|Y - \bar{Y}\mathbb{1}\|^2 = Y'Y - n\bar{y}^2.$$

et la somme des carrés de régression :

$$SS_R = \|\hat{Y} - \bar{Y}\mathbb{1}\|^2 = \hat{Y}'\hat{Y} - n\bar{y}^2 = Y'HY - n\bar{y}^2.$$

$$= \hat{\beta}'X'Y - n\bar{y}^2.$$

On vérifier alors : $SS_T = SS_R + SS_E$.

4. Coefficient de détermination

On appelle coefficient de détermination le rapport.

$$R^2 = \frac{SS_R}{SS_T}$$

qui est donc la part de variation de Y expliquée par le modèle de régression. Géométriquement, c'est un rapport de carrés de longueur de deux vecteurs.

c'est donc le cosinus carré de l'angle entre ces vecteurs Y et sa projection $\hat{Y}$ sur Vect (X).

Attention: Dans le cas extrême où $n = p+1$, c'est-à-dire si le nombre de variables explicatives est grand comparativement au nombre d'observation $R^2 = 1$.

4. Inférence dans le cas gaussien:

En principe, l'hypothèse optionnelle "4" de normalité des erreurs est nécessaire pour cette section, En pratique, des résultats asymptotiques, donc valides pour de grands échantillons, Ainsi que des études de simulation, montrent que cette hypothèse n'est pas celle dont la violation est la plus pénalisante pour la fiabilité des modèles.

4.1. Inférence sur les coefficients:

Pour chaque coefficient β_j on montre que la statistique $\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}}$. Où $\sigma^2_{\hat{\beta}_j}$ variance de $\hat{\beta}_j$ est le j-iéme terme diagonal de la matrice $s^2(x'x)^{-1}$ suit une loi de student à $(n-p-1)$ ddl. cette statistique est donc utilisée pour tester une hypothèse $H_0: \beta_j = a$. Ou pour construire un intervalle de confiance de niveau $100(1-\alpha)\%$

$$\hat{\beta}_j \pm t_{\alpha/2,\, n-p-1} \times \hat{\sigma}_{\hat{\beta}_j}.$$

A.2. Inférence sur le modèle:

Le modèle peut être testé globalement, sous l'hypothèse nulle $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$, la statistique $\frac{SS_R / p}{SS_E / (n-p-1)} = \frac{MS_R}{MS_E}$ suit une loi de Fisher avec p et $(n-p-1)$ degrés de liberté.

- Les résultats sont présentés dans un tableau "d'analyse de la variance" sous la forme suivante:

Sources de variation	d.d.L.	Somme des carrés	Variance	F.
Régression.	p.	SS_R	$MS_R = \frac{SS_R}{p}$	$\frac{MS_R}{MS_E}$.
Erreur	$n-p-1$.	SS_E	$MS_E = \frac{SS_E}{n-p-1}$.	
Total	$n-1$.	SS_T	/	

3. Inférence sur un modèle réduit :

Le testé précédent amène à rejeter H_0 dès que l'une des variables x^j est liée à Y. Il est donc d'un intérêt limité.

Il est souvent plus utile de tester un modèle réduit c'est-à-dire dans le quel certains coefficients sont nuls (à l'exception du terme constant) contre le modèle complet avec toutes les variables. En ayant éventuellement réordonné les variables. On considère l'hypothèse nulle

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_q = 0 \qquad q < p$$

• Notons respectivement SS_{R_q}, R_q^2 les sommes carrés et le coefficient de détermination du modèle réduit à $(p-q)$ variables.

Sous H_0, la statistique :

$$\frac{SS_R - SS_{R_q}}{SS_E / n-p-1} = \frac{(R^2 - R_q^2)/q}{(1-R^2)/n-p-1}.$$

Suit une loi de Fisher à q et $(n-p-1)$ d.d.L.

• Dans le cas particulier où $q=1$ $(\beta_j = 0)$, la F-statistique est alors le carré de la t-statistique de l'inférence sur un paramètre et conduit au même test.

4. Prévision par un intervalle de Confiance:

Pour x_0 : $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0^1 + \cdots + \hat{\beta}_p x_0^p$

$$\hat{y}_0 \pm t_{\alpha/2;\, n-p-1} \times S \left(\mathbb{1} + X_0' (X'X)^{-1} X_0 \right)^{1/2}.$$