Chapter 3: Multiple Linear Regression

Dr. Abbas Rammal

Bachelor's degree in Mathematics
Option: DATA SCIENCE
September 2023

Plan

1. Introduction
2. Regression hyperplane
3. Matrix writing
4. Vector estimation
5. Properties of least squares
6. Model assumptions
7. Properties of estimators
8. Testing the significance of the model
9. Significance test of a parameter β_j
10. Regression without constant
11. Partial Fisher Test

Introduction

- In practice, the explained variable does not depend only on an explanatory variable.
- In this chapter, we introduce multiple regression models with several explanatory variables.
- The regression line is replaced by a regression hyperplane.
- **Objective**: The objective of this chapter is to find the model estimates and to apply inferential statistics on these parameters.

Regression hyperplane

- **Objective:** Fit several explanatory variables $X_1, X_2, \ldots, X_p$ to a dependent variable Y in a linear manner.
- We seek to construct a relation of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon$$

- For all $i = 1, \ldots . n$ we have:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- The model equation is sometimes expressed in the form:

$$\mu_Y(x_1, x_2, \cdots, x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

- where $\mu_y(x_1, x_2, \ldots, x_p)$ denotes the average of the measured variable Yon all individuals for which the variable X_1 is worth x_1, the variable X_2 is worth x_2, ..., the variable X_p is worth x_p
- This equation is the equation of a hyperplane called the regression hyperplane.
- This hyperplane is however unknown and the multiple regression problem consists of estimating the p +1 parameters from the sample data.
- This involves determining the hyperplane:

$$\hat{y}(x_1, x_2, \cdots, x_p) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

which best approximates these data.

- We note $\hat{y}_i$ the estimated values:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip}$$

- The unobservable quantities e_i can be estimated by the observable quantities:

$$e_i = y_i - \hat{y}_i$$

- The problem will be: Find $\beta_0, \beta_1, \dots, \beta_p$ which minimizes

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots , \beta_p x_{ip}))^2$$

Matrix writing

- For each observation i, ($i = 1, \dots . n$) we have:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- The multiple regression model can be written in matrix form as follows:

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

$$\text{avec: } Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \text{ et } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Vector estimation $\boldsymbol{\beta}$

- The idea is always to minimize the sum of the squares of the residues by the method of least squares.

$$(P) : min \sum_{i=1}^{n} e_i^2 = \mathbf{e'e}$$

- The resolution of this optimization problem gives:

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

Properties of least squares

1) We have $\hat{y}'\hat{y} = y'\hat{y}$
2) The least squares hyperplane contains the point $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p, \bar{y})$
3) Moreover $\sum \hat{y}_i = \sum y_i$
4) Also, $\sum e_i = 0$
5) The decomposition of variation:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)$$

$$SC_{tot} = SC_{reg} + SC_{res}$$

Model assumptions

- The random vector e follows a multinormal law with
- $$\mathbb{E}(\varepsilon) = \mathbf{0} \text{ et } \mathbb{V}(\varepsilon) = \sigma^2 \mathbf{I}_n$$
- Homoscedasticity hypothesis: $\mathbb{E}(\varepsilon_i) = 0 \text{ et } \mathbb{V}(\varepsilon_i) = \sigma^2$
- The error is independent of X_j i.e. $cov(x_{ij}, \varepsilon_i) = 0$
- Normality of residues: $\varepsilon_i \hookrightarrow \mathcal{N}(0, \sigma^2)$
- The ε_i are mutually independent (absence of autocorrelation of the residues) i.e.

$$cov(\varepsilon_i, \varepsilon_j) = 0 \text{ si } i \neq j$$

- These hypotheses amount to saying that the random vector follows a multinormal law with

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta \text{ et } \mathbb{V}(\mathbf{y}) = \sigma^2 \mathbf{I}_n$$

Properties of estimators

$\mathbb{E}(\hat{\beta}) = \beta$

$\mathbb{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}$

$\mathbb{V}(\hat{\beta})$ is called the variance-covariance matrix of the coefficients:

$$\mathbb{V}(\hat{\beta}) = \begin{pmatrix} s^2(\hat{\beta}_0) & cov(\hat{\beta}_0, \hat{\beta}_1) & \cdots & cov(\hat{\beta}_0, \hat{\beta}_p) \\ cov(\hat{\beta}_0, \hat{\beta}_1) & s^2(\hat{\beta}_1) & \cdots & cov(\hat{\beta}_1, \hat{\beta}_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\hat{\beta}_0, \hat{\beta}_p) & cov(\hat{\beta}_1, \hat{\beta}_p) & \cdots & s^2(\hat{\beta}_p) \end{pmatrix}$$

- The variance-covariance matrix involves the error variance σ^2.
- This variance is unknown
- It is estimated by

$$s^2 = \frac{SC_{res}}{n-p-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1}$$

- We estimate $\mathbb{V}(\hat{\beta})$ by $s^2(\hat{\beta}) = s^2(X'X)^{-1}$

Model significance test

- We start by testing:

$$\begin{aligned} H_0 &: \quad \beta_1 = \beta_2 = \cdots = \beta_p = 0 \\ H_1 &: \quad \exists j \in \{1, \ldots, p\}, \beta_j \neq 0 \end{aligned}$$

- We use the following statistic:

$$F = \frac{\left(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2\right) / p}{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right) / (n-p-1)} = \frac{SC_{reg} / p}{SC_{res} / (n-p-1)}$$

which is distributed under H_0 according to a Fisher law at p and n-p-1 degrees of freedom.

- We reject H_0 with an α risk if

$$F > f_{1-\alpha}(p, n-p-1)$$

ANOVA Table

Source of variation	df	Sum of squares	Mean of squares	F
Regression	p	$SCE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_n)^2$	$\frac{1}{p}\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_n)^2$	$\frac{SCE/p}{SCR/(n-p-1)}$
Residual	n-(p+1)	$SCR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$\frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	
Total	n-1	$SCT = \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$	$\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$	

Significance test of a parameter $\boldsymbol{\beta_j}$

- The objective is to test the absence of a linear connection between X_j and Y. So we must test the nullity of β_j for $j = 1 \dots p$.

$$H_0 : \beta_j = 0 \text{ contre } H_1 : \beta_j \neq 0$$

- Test statistic:

$$T = \frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \rightsquigarrow t(n-p-1)$$

- Under H_0:

$$T = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \rightsquigarrow t(n-p-1)$$

- Rejection region:

$$\left]-\infty; -t_{\alpha/2,(n-p-1)}\right[\cup \left]t_{\alpha/2,(n-p-1)}, +\infty\right[$$

- Confidence interval for β_j at confidence level 1-α

$$IC(\beta_j) = \left[\hat{\beta}_j \pm t_{\alpha/2,(n-p-1)} s(\hat{\beta}_j)\right]$$

- We reject H_0 if 0 does not belong to this interval.

Regression without constant

- The model without constant is written as:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- The multiple regression model can be written in matrix form as follows:

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

avec: $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$ et

$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

- The least squares solution is always given by:

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

- Due to the absence of the column of 1 in the matrix X, certain properties of the estimators are lost
  - The hyperplane does not pass through the point $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p, \bar{y})$.
  - We do not have $\sum y_i = \sum \hat{y}_i$
  - The sum of the residues is therefore not zero.
  - It follows that we no longer have: $SC_{tot} = SC_{reg} + SC_{res}$
  - We always have $SC_{res} = \sum (y_i - \hat{y}_i)^2 = \sum y_i^2 - \sum \hat{y}_i^2$

- We define $\mathbf{SC_{reg}}$ and $\mathbf{SC_{total}}$ in this form:

$$SC_{reg} = \sum \hat{y}_i^2 \text{ et } SC_{tot} = \sum y_i^2$$

- The determination coefficient is given by:

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2}$$

- R^2 no longer has the correlation coefficient.

- An unbiased estimator of σ^2 is:

$$s^2 = \frac{SC_{res}}{n - p}$$

- In order to test the hypothesis that all coefficients β_j are zero:

$$\mathcal{H}_0 : \beta_1 = \cdots = \beta_p = 0$$

- The Test statistic is:

$$F_c = \frac{SC_{reg}}{SC_{res}} \cdot \frac{n-p}{p}$$

- We reject H_0 if

$$F_c > f_\alpha(p; n-p)$$

Partial Fisher Test

- Objective: This test makes it possible to test the nullity of a certain number r of parameters in a model with p + 1 parameters.
  - The null hypothesis is that of a reduced model with (p + 1 - r) parameters
  - The alternative hypothesis is that of a complete model with p + 1 parameters.

- **Test procedure:**

- Calculate the estimated values $\hat{y}_i$ using the least squares method for each of the two models defined by H_0 and H_1. We notice:
  - The estimated values by: $\hat{y}_i(\mathcal{H}_0)$ et $\hat{y}_i(\mathcal{H}_1)$
  - The sums of squares of the residues by: $SC_{res}(\mathcal{H}_0)$ et $SC_{res}(\mathcal{H}_1)$

- We calculate the statistic:

$$F_c = \frac{\sum \hat{y}_i^2(\mathcal{H}_0) - \sum \hat{y}_i^2(\mathcal{H}_1)}{\sum y_i^2 - \sum \hat{y}_i^2(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

$$F_c = \frac{SC_{res}(\mathcal{H}_0) - SC_{res}(\mathcal{H}_1)}{SC_{res}(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

- We reject H_0 if

$$F_c > f_\alpha(r; n-p-1)$$

- **When the two models defined by H_0 and H_1 admit a constant:**

- We define:

$$SC_{reg}(\mathcal{H}_0) = \sum(\hat{y}_i(\mathcal{H}_0) - \bar{y})^2$$

$$SC_{reg}(\mathcal{H}_1) = \sum(\hat{y}_i(\mathcal{H}_1) - \bar{y})^2$$

$$R^2(\mathcal{H}_0) = \frac{SC_{reg}(\mathcal{H}_0)}{\sum(y_i - \bar{y})^2}$$

$$R^2(\mathcal{H}_1) = \frac{SC_{reg}(\mathcal{H}_1)}{\sum(y_i - \bar{y})^2}$$

- We then have:

$$F_c = \frac{SC_{reg}(\mathcal{H}_1) - SC_{reg}(\mathcal{H}_0)}{SC_{reg}(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

$$F_c = \frac{R^2(\mathcal{H}_1) - R^2(\mathcal{H}_0)}{1 - R^2(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

- When the two models defined by H_0 and H_1 do not admit a constant:
- We define:

$$SC_{reg}(\mathcal{H}_0) = \sum \hat{y}_i^2(\mathcal{H}_0)$$

$$SC_{reg}(\mathcal{H}_1) = \sum \hat{y}_i^2(\mathcal{H}_1)$$

$$R^2(\mathcal{H}_0) = \frac{SC_{reg}(\mathcal{H}_0)}{\sum y_i^2}$$

$$R^2(\mathcal{H}_1) = \frac{SC_{reg}(\mathcal{H}_1)}{\sum y_i^2}$$

- We then have:

$$F_c = \frac{SC_{reg}(\mathcal{H}_1) - SC_{reg}(\mathcal{H}_0)}{SC_{reg}(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

$$F_c = \frac{R^2(\mathcal{H}_1) - R^2(\mathcal{H}_0)}{1 - R^2(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

- **When the model defined by H_1 admits a constant and the model defined by H_0 does not admit one:**
- We calculate the statistic:

$$F_c = \frac{\sum \hat{y}_i^2(\mathcal{H}_0) - \sum \hat{y}_i^2(\mathcal{H}_1)}{\sum y_i^2 - \sum \hat{y}_i^2(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

$$F_c = \frac{SC_{res}(\mathcal{H}_0) - SC_{res}(\mathcal{H}_1)}{SC_{res}(\mathcal{H}_1)} \cdot \frac{n-p-1}{r}$$

- We reject H_0 if

$$F_c > f_\alpha(r; n-p-1)$$