Linear Regression Model

Marco Corneli

Ex. 1. We consider the linear regression model $Y = X\beta + \epsilon$, where

$$Y = (Y_1, \ldots, Y_n)^T$$

are the n observed dependent variables, $X \in \mathbb{R}^{n \times p}$ is the matrix whose rows are observations of the p features and it is assumed that

$$\text{rank}(X) = p \qquad (\Rightarrow X^T X \quad \text{invertible})$$

Moreover, $\beta \in \mathbb{R}^p$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ are i.i.d.residuals such that

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Denoting by $\hat{\beta}$ the OLS estimator of β and by $\hat{\epsilon}$ the OLS residuals, the aim of this exercise is to show that

$$\begin{pmatrix} \hat{\beta} \\ \hat{\epsilon} \end{pmatrix}$$

is a Gaussian vector with mean $(\beta, 0)^T$ and variance covariance matrix

$$\Sigma = \sigma^2 \begin{pmatrix} (X^T X)^{-1} & 0 \\ 0 & I_n - X(X^T X)^{-1} X^T \end{pmatrix}.$$

(a) Recall how the OLS estimator

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

is computed and show that the fitted values $\hat{Y}$ are obtained as

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

(b) Show that that $\hat{\beta}$ is an unbiased estimator of β and compute its variance.

(c) We now introduce the matrix $P_1 := X(X^TX)^{-1}X^T$ and show that it is a **projection** matrix[1]. In order to do that
   i. Show that P_1 is symmetric.
   ii. Show that $P_1^2 = P_1$.

(d) Prove that $X\hat{\beta} = X\beta + P_1\epsilon$.

(e) Show that the OLS residuals $\hat{\epsilon} := Y - X\hat{\beta}$ are obtained as[2]

$$\hat{\epsilon} = (I_n - P_1)Y$$

(f) Prove that the matrix $P_2 := (I_n - P_1)$ is also a projection matrix and show that $P_1P_2 = 0_n$.

(g) Show that $\hat{\epsilon} = P_2\epsilon$.

We now focus on the two vectors $X\hat{\beta}$ and $\hat{\epsilon}$ (points (d) and (g) respectively). Since they are linear combinations of the errors $\epsilon_1, \dots, \epsilon_n$ they are both Gaussian.

(a) By using point (f), show that $X\hat{\beta}$ and $\hat{\epsilon}$ are independent.

(b) Compute the expected values of $\hat{\epsilon}$ and show that its variance is P_2. Conclude.

Ex.2 **Reminder**: consider a sequence $Z_1, \dots, Z_K$ of i.i.d. standard Gaussian random variables. The random variable

$$Q := \sum_{k=1}^{K} Z_k^2$$

follows a **chi-squared** $\chi^2(K)$ distribution with K degrees of freedom.

We know from the previous exercise that the OLS residuals $\hat{\epsilon}$ is a Gaussian vector with zero mean and variance-covariance matrix $\sigma^2 P_2$. In this section we derive an estimator of σ^2. It can be shown that[3]

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = \|\hat{\epsilon}\|^2 = \sum_{k=1}^{n-p} Z_k^2, \tag{1}$$

where $Z_k \sim \mathcal{N}(0, \sigma^2)$ and are all independent.

[1] A projection matrix is any symmetric matrix M such that $M^2 = M$.

[2] I_n denotes the square identity matrix of order n. In the following 0_n denotes the null square matrix.

[3] See the appendix to this document for a proof.

(a) Modify Eq. (1) is such a way to obtain a sum of independent *standard* (unit variance) Gaussian random variables.
(b) What is the distribution of $\|\hat{\epsilon}\|^2$?
(c) Compute $\mathbf{E}(\|\hat{\epsilon}\|^2)$ and deduce an estimator of σ^2.

Ex.3 With the same notations of the previous exercises, the aim of this exercise is to build a <u>statistical test of level α</u> to test[4]

$$\mathcal{H}_0 : \beta_j = 0, \tag{2}$$

where $\beta = (\beta_1, \ldots, \beta_p)^T$.

Part one: σ^2 is assumed to be known.

(a) Consider first a **known** vector $a \in \mathbb{R}^p$ and focus on the random variable $a^T\hat{\beta}$. What is its distribution? Compute its mean and its variance.
(b) Provide a 95% level confidence interval for $a^T\beta$.
(c) It is now assumed that

$$a = (0, \ldots, 1)^T.$$

Who is now $a^T\hat{\beta}$? What about its distribution?
(d) You can now compute the following statistic test

$$T =: \frac{\hat{\beta}_p - \mathbf{E}_0[\hat{\beta}_p]}{\sqrt{\mathrm{Var}(\hat{\beta}_p)}}, \tag{3}$$

where $\mathbf{E}_0[\cdot]$ denotes the expectation under $\mathcal{H}_0$ being true. What is the distribution of T? Compute the probability

$$\mathbf{P}_0\{|T| > z_{1-\frac{\alpha}{2}}\}.$$

(e) Assume that, on a real dataset, you observe $T = 3.72$. Would you reject H_0? And what if $T = 1.975$?

Part two: σ^2 is unknown.

[4]Notice that this is equivalent to test the significance of including in the model the j-th feature (why?).

Remainder: Consider two *independent* random variables $Z \sim \mathcal{N}(0,1)$ and $X \sim \chi^2(p)$. The random variable

$$\overline{T} = \frac{Z}{\sqrt{\frac{X}{p}}} \tag{4}$$

follows a **t-student** distribution with p degrees of freedom.

(a) Write a test statistic $\overline{T}$ obtained by replacing σ^2 in Eq. (3) with $\hat{\sigma}^2$ obtained in Exercise 2:

$$\hat{\sigma}^2 = \frac{\|\hat{\epsilon}\|^2}{n-p}$$

(b) Multiply and divide the denominator of $\overline{T}$ by $\sqrt{\sigma^2}$ and write $\overline{T}$ as a function of T.

(c) What is the distribution of $\frac{\hat{\sigma}^2}{\sigma^2}$? Conclude.

(d) If $t_{n-p,1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of a t-student distribution with $n-p$ degrees of freedom, what is the probability

$$\mathbf{P}_0\{|\overline{T}| > t_{n-p,1-\alpha/2}\}$$

equal to? If you observed $\overline{T} > t_{n-p,1-\alpha/2} + \delta$, for some positive δ, would you reject $\mathcal{H}_0$?

Appendix

Proof. From Ex.1, we known that

$$\|\hat{\epsilon}\|^2 = \|P_2 Y\| = Y^T P_2 Y, \tag{5}$$

since P_2 is a projection matrix onto $V^{\perp}$, where $V = \text{Span}(X^1, \ldots, X^p)$. Moreover, $\text{rank}(P_2) = n - p$ and

$$P_2 = M \Lambda M^T$$

where M is orthogonal and $\Lambda = \text{diag}(\underbrace{1, \ldots, 1}_{n-p \quad \text{times}}, 0, \ldots, 0)$. Thus, one can introduce the vector

$$Z = M^T Y$$

and reformulate Eq.(5) as

$$\|\hat{\epsilon}\|^2 = Z^T \Lambda Z = \sum_{i=1}^{n-p} Z_i^2.$$

It is immediate to check that Z is a Gaussian vector of zero mean and variance $\sigma^2 I_n$ and this concludes the proof. □