Exercise 3: Matrices and vectors for regression with R

Murshid Saqlain
Deadline February 23; msq@du.se

1. Basic Matrix operations

   (a) Create the following matrix with cbind(), rbind() and matrix().

   $$\mathbf{X} = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 4 & 5 \\ 1 & 5 & 3 \\ 1 & 6 & 1 \end{pmatrix}$$

   Not only with one of the functions, do it three times with all three functions.

   (b) Compute $\mathbf{X}'\mathbf{X}$ with R

   (c) Compute the inverse of $\mathbf{X}'\mathbf{X}$ with R. And control this is the inverse (Hint: **Check the definition of an inverse matrix**).

2. This exercise goes through how statistical packages conduct least squares estimation with matrix algebra.

   The variable $\mathbf{Y} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 5 \end{pmatrix}$ is the dependent variable in the following regression model

   $$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

   where $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{X}$ is the matrix from the previous exercise.

   (a) What is the sample size, n?

   (b) How many are the regressors, K?

(c) Compute the least squares estimator of β with matrix operations in R.

(d) We have a very small sample size but if we assume we have a random sample and the population random error variable is normally distributed $U \sim N(0, \sigma^2)$, then

$$T = \frac{\hat{\beta}_k - \beta_k}{\sqrt{S^2 \left((\mathbf{X}'\mathbf{X})^{-1}\right)_{(k+1),(k+1)}}} \sim t_{n-K-1},$$

where $\left((\mathbf{X}'\mathbf{X})^{-1}\right)_{(k+1),(k+1)}$ is the $(k+1)^{th}$ element on the diagonal of $(\mathbf{X}'\mathbf{X})^{-1}$. The T-statistic has a t-distribution with $n-K-1$ degrees of freedom. For $H_0 : \beta_k = 0$ compute the T-statistics for all three betas in the vector β. You need to calculate $S^2 = \frac{\sum \hat{U}_i^2}{n-K-1}$, under the null hypothesis $\beta_k = 0$ and you have calculated $\hat{\beta}_k$ in 2c.

(e) For the following hypotheses

$$H_0 : \beta_k = 0 \; H_a : \beta_k \neq 0$$

compute the appropriate P-values $2P(T > |t|)$ for two-sided alternatives (for all betas).
(Hint: The probabilities $P(T > |t|)$ are obtained from the t-distribution with $n - K - 1$ degrees of freedom. In R pt() can be used to obtain these probabilities. Read the help file for pt(). t is the t-value you have computed for a coefficient in 2d , so you will have three different t.)

(f) Control that you have computed the t-values and the P-values correctly by using lm() to obtain these values. The t-values and P-values you have computed should be what you get from $summary()$ on the regression object (could be some minor rounding-off errors).