Phil Sundin
February 6, 2020

Proof of Regression Sum of Squares, Orthogonal Predictors and No Intercept

The purpose of this document is to derive the regression sum of squares in a OLS setting, when the predictors are orthogonal and there is no intercept included in the model. It was mentioned in the paper "Orthogonal decomposition of interaction effect in analysis of variance" by Xiaofeng Steven Liu. The exact sentence is "As the orthogonal coding is used, the regression sum of squares is the squared regression coefficient times the sum of squared values of the orthogonal predictor." What he failed to mention is that this convenient representation of the regression sum of squares only holds if an intercept is not included in the model.

Assume we have a standard linear regression setting with n $(i = 1 \ldots n)$ observations and p $(j = 1 \ldots p)$ predictors. Denote $\boldsymbol{Y}$ as our $n \times 1$ response vector, $\boldsymbol{X}$ is our $n \times p$ design matrix, and $\hat{\boldsymbol{\beta}} = \boldsymbol{(X'X)^{-1}X'Y}$ is our $p \times 1$ vector of estimated regression coefficients. In our case, our design matrix does not include a vector of 1's for the intercept, and the predictors are also all orthogonal to one another. Because of the orthogonal columns, $\boldsymbol{(X'X)}$ is a diagonal matrix. Since we do not have an intercept, we have our uncorrected total sum of squares (SST) as

$$SST = \sum_{i=1}^{n} y_i^2 = \boldsymbol{Y}'\boldsymbol{Y},$$

where y_i is the i^{th} response in the vector $\boldsymbol{Y}$. In the setting without an intercept, partitioning the total sum of squares into a regression sum of squares (SSR) and an error sum of squares (SSE) can be less straightforward than in the case with an intercept. In particular, the sum of squares due to regression becomes less intuitive. We still define our SSE in the same way as with an intercept, that is,

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}),$$

where $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is our $n \times 1$ vector of predicted responses.

So to get a formula for SSR, let's subtract the SSE from the SST. Writing this in matrix form, we can solve for SSR = SST - SSE:

$$= \boldsymbol{Y}'\boldsymbol{Y} - (\boldsymbol{Y} - \hat{\boldsymbol{Y}})'(\boldsymbol{Y} - \hat{\boldsymbol{Y}})$$

$$= \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{Y} + \hat{\boldsymbol{Y}}'\boldsymbol{Y} + \boldsymbol{Y}'\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}$$

$$= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

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

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

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

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

$$= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

So as mentioned above, the orthogonal structure of the design matrix means $\boldsymbol{X}'\boldsymbol{X}$ is a diagonal matrix, with the j^{th} diagonal elements as $\sum_{i=1}^{n} x_{ij}^2$. So now, our matrix multiplication looks like:

$$= \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_p \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & 0 & \dots & 0 \\ 0 & \sum_{i=1}^{n} x_{i2}^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \sum_{i=1}^{n} x_{ip^2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

This yields our desired result that SSR $= \sum_{j=1}^{p} \left(\beta_j^2 \sum_{i=1}^{n} x_{ij}^2 \right)$. And more specifically, the sum of squares due to any particular predictor can be represented as $\beta_j^2 \sum_{i=1}^{n} x_{ij}^2$.