H.M. James Hung

Magnificent Dummy (or indicator) Variable

Without loss of generality, consider a simple linear model $y = \beta_0 + \beta_1 x + \varepsilon$, where ε has zero expectation and constant variance σ^2. Suppose that there are n statistically independent pairs of data, denoted by $(y_i\ ,\ x_i)$, $i = 1, \ldots, n$.

Problem: Predict y value at a given value x_0 based on the model above.

From Slide #6 of Module 2 Lecture 2C, we know how to generate the predictor $\hat{y}_0$. Let us create a y-value labeled as y_{n+1} and set it to zero. In addition, we create a dummy variable z as follows:

$$z_i = 0\ , \quad i = 1, \ldots., n$$

$$z_{n+1} = -1$$

and set $x_{n+1} = x_0$.

Now we have two regressors x and z, and $(n + 1)$ data values $(y_i\ ,\ x_i\ ,\ z_i)$, $i = 1, \ldots, n + 1$. Let us fit the multiple linear regression model $y_i = \beta_0 + \beta_1 x_i + \gamma z_i + \varepsilon_i$ and obtain the OLS estimator of γ.

Amazingly, the OLS estimator of γ is mathematically equal to the predictor $\hat{y}_0$ of y at $x = x_0$. The indicator variable z is truly magnificent.

Let us prove this amazing finding mathematically. Then derive the estimated variance of the OLS estimator of γ and prove that this estimated variance is equal to the estimated variance of $\hat{y}_0$.

Here is a proof.

Denote by y_0 the y value to be predicted at a given value x_0, following the simple regression model $y = \beta_0 + \beta_1 x + \varepsilon$ with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. That is,

$$y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 \quad , \qquad (*)$$

where ε_0 is the random error for y_0, $E(\varepsilon_0) = 0$, $Var(\varepsilon_0) = \sigma^2$, and y_0 is statistically independent with $y_1, \ldots, y_n$. Let us treat the unknown y_0 as a parameter. By moving it to the right of the equation (*), we have

$$0 = \beta_0 + \beta_1 x_0 - y_0 + \varepsilon_0 = \beta_0 + \beta_1 x_0 + y_0 \times (-1) + \varepsilon_0$$

For the n observations $(y_i \ , \ x_i)$, $i = 1, \ldots, n$, we have

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \beta_0 + \beta_1 x_i + y_0 \times 0 + \varepsilon_i$$

Now we can play the magnificent dummy and create the data matrix using the table below.

Observation number	**y**	**Intercept (for β_0)**	**x (for β_1)**	**z (for γ)**	**Random error**
1	y_1	1	x_1	0	ε_1
2	y_2	1	x_2	0	ε_2
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
n	y_n	1	x_n	0	ε_n
n+1 (appended to predict y_0)	0	1	x_0	-1	ε_0

Now label each part of this matrix.

y_1	1	x_1	0	ε_1
y_2	1	x_2	0	ε_2
.	.	.		.
.	.	.		.
.	.	.		.
y_n	1	x_n	0	ε_n
0	1	x_0	-1	ε_0

The new observation vector is $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{y} \\ 0 \end{bmatrix}$ where $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

The new design matrix is $\boldsymbol{D} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{W} & -1 \end{bmatrix}$

where $\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ $\boldsymbol{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ $\boldsymbol{W} = [1 \quad x_0]$.

Thus,

$$\boldsymbol{D}'\boldsymbol{D} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{W}'\boldsymbol{W} & \boldsymbol{W}' \\ \boldsymbol{W} & -1 \end{bmatrix}$$

The OLS estimator of $(\beta_0 \, , \beta_1 \, , \gamma)$ from the multiple regression model $y_i = \beta_0 + \beta_1 x_i + \gamma z_i + \varepsilon_i$ is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\gamma} \end{pmatrix} = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{U} \quad .$$

We can prove: $(\boldsymbol{D}'\boldsymbol{D})^{-1} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{W}'\boldsymbol{W} & \boldsymbol{W}' \\ \boldsymbol{W} & -1 \end{bmatrix}^{-1}$

$$= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{W} & 1 \end{bmatrix} \begin{bmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{W}' \\ \boldsymbol{0} & 1 \end{bmatrix} \quad ,$$

using the following matrix algebra:

For a symmetric positive definite matrix $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$,

$$A^{-1} = \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix} .$$

Hence, $(\boldsymbol{D}'\boldsymbol{D})^{-1} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{W}'\boldsymbol{W} & \boldsymbol{W}' \\ \boldsymbol{W} & -1 \end{bmatrix}^{-1}$

$$= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{W} & 1 \end{bmatrix} \begin{bmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W}' \\ \boldsymbol{0} & 1 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} & (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{W}' \\ \boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{W}' + \boldsymbol{1} \end{bmatrix}$$

Next,

$$\boldsymbol{D}'\boldsymbol{U} = \begin{bmatrix} \boldsymbol{X}' & \boldsymbol{W}' \\ \boldsymbol{0} & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\gamma} \end{pmatrix} = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{U} = \begin{bmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} & (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{W}' \\ \boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{W}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \end{bmatrix}$$

Now we see the magnificent findings:

The OLE estimator of (β_0, β_1) from this multiple linear regression is exactly equal to the OLS estimator of (β_0, β_1) from that simple linear regression to start with; that is, it is $(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$. Furthermore, the predictor of y_0 derived from this multiple linear regression is exactly equal to the predictor from that simple linear regression; that is, $\hat{\gamma} = \boldsymbol{W}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} = [1 \quad x_0]\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Clearly, the variances of $\hat{\beta}$ and $\hat{\gamma}$

$$Var\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\gamma}\end{pmatrix} = \boldsymbol{\sigma}^2(\boldsymbol{D'D})^{-1} = \boldsymbol{\sigma}^2\begin{bmatrix}(\boldsymbol{X'X})^{-1} & (\boldsymbol{X'X})^{-1}\boldsymbol{W'}\\ \boldsymbol{W}(\boldsymbol{X'X})^{-1} & \mathbf{1} + \boldsymbol{W}(\boldsymbol{X'X})^{-1}\boldsymbol{W'}\end{bmatrix} .$$

Thus,

$$\boldsymbol{Var}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_0\\ \widehat{\boldsymbol{\beta}}_1\end{pmatrix} = \boldsymbol{\sigma}^2(\boldsymbol{X'X})^{-1} \ .$$

$$\boldsymbol{Var}(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\sigma}^2(\mathbf{1} + \boldsymbol{W}(\boldsymbol{X'X})^{-1}\boldsymbol{W'}) = \boldsymbol{\sigma}^2\{\mathbf{1} + [\mathbf{1} \quad \boldsymbol{x}_0](\boldsymbol{X'X})^{-1}\begin{bmatrix}\mathbf{1}\\ \boldsymbol{x}_0\end{bmatrix}\} .$$

As usual, the σ^2 is estimated by

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n+1}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i - \widehat{\gamma} z_i)^2}{n+1-3} = \frac{\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 + (0 - \widehat{\beta}_0 - \widehat{\beta}_1 x_0 + \widehat{\gamma})^2}{n-2}$$

Note that $\hat{\gamma} = \hat{\beta}_0 + \hat{\beta}_1 x_0$; thus, the term $(0 - \hat{\beta}_0 - \hat{\beta}_1 x_0 + \hat{\gamma})^2 = 0$.

What does this mean? It means that adding the “Magnificent Dummy” does not change the estimator of σ^2 . That is,

$$\tilde{\boldsymbol{\sigma}}^2 = \frac{\sum_{i=1}^{n+1}(\boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}_0 - \widehat{\boldsymbol{\beta}}_1 \boldsymbol{x}_i - \widehat{\boldsymbol{\gamma}} \boldsymbol{z}_i)^2}{\boldsymbol{n}+\mathbf{1}-\mathbf{3}} = \widehat{\boldsymbol{\sigma}}^2$$ which is the estimator of σ^2 from those n observations before prediction.

Lastly, the estimated variance of $\hat{\gamma}$ is equal to the estimated variance of $\hat{y}_0$; that is,

$$\hat{\sigma}^2\{1 + [1 \quad x_0](X'X)^{-1}\begin{bmatrix} 1 \\ x_0 \end{bmatrix}\} \ .$$

Exercise Problems:

1) Create a numerical example to confirm the results shown above.

2) Prove mathematically that the "Magnificent Dummy" method can generate an unbiased estimator of $E(y \mid x_0)$ and an unbiased estimator of the variance of the unbiased estimator of $E(y \mid x_0)$.