Prove $SST = SSE + SSR$

Prove

$$SST = SSE + SSR$$

I start with

$$SST = \Sigma(y_i - \bar{y})^2 = \ldots = SSE + SSR + \Sigma 2(y_i - y_i^*)(y_i^* - \bar{y})$$

and I don't know how to prove that $\Sigma 2(y_i - y_i^*)(y_i^* - \bar{y}) = 0$

a note on notation: the residuals e_i is $e_i = y_i - y_i^*$. A more common notation is $\hat{y}$.

(statistics) (regression)

edited Mar 12 '14 at 12:33

asked Mar 12 '14 at 12:07
jacob
977 1 3 17

What is y_i^*? And I assume that $\bar{y}$ is the average of the observations $y_1, \ldots, y_n$, but please write such things explicitly in your post. – Stefan Hansen Mar 12 '14 at 12:16

y^* is my notation of the often used $\hat{y}$ – jacob Mar 12 '14 at 12:26

In a nutshell, you have to use the fact that $\sum e_i = 0$ and $\sum \hat{y}_i e_i = 0$ (see lectures 3 and 6 at robots.ox.ac.uk/~fwood/w4315_fall2010/Lectures) – Brad S. Mar 15 '14 at 19:14

@BradS. Can't see on what slide# – jacob Mar 16 '14 at 12:57

@jacob Sorry, I should have been more specific. In lecture 3 (robots.ox.ac.uk/~fwood/w4315_fall2010/Lectures/lecture-3/…), he derives the equations for the parameter estimates in simple linear regression and then in lecture 6 (robots.ox.ac.uk/~fwood/w4315_fall2010/Lectures/lecture-6/…) he directly addresses SST = SSR + SSE . In essence, the two fact I mentioned in my previous comment fall out of the minimization he does on the very first (non-title) page of lecture 3. – Brad S. Mar 16 '14 at 16:22

2 Answers

The principle underlying least squares regression is that the sum of the squares of the errors is minimized. We can use calculus to find equations for the parameters β_0 and β_1 that minimize the sum of the squared errors.

Let $S = \sum_{i=1}^{n} (e_i)^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \beta_0 - \beta_1 x_i)^2$

We want to find β_0 and β_1 that minimize the sum, S. We start by taking the partial derivative of S with respect to β_0 and setting it to zero.

$$\frac{\partial S}{\partial \beta_0} = \sum 2(y_i - \beta_0 - \beta_1 x_i)^1 (-1) = 0$$

notice that this says,

$$\sum (y_i - \beta_0 - \beta_1 x_i) = 0$$
$$\sum (y_i - \hat{y}_i) = 0 \qquad (eqn. 1)$$

Hence, **the sum of the residuals is zero** (as expected). Rearranging and solving for β_0 we arrive at,

$$\sum \beta_0 = \sum y_i - \beta_1 \sum x_i$$
$$n\beta_0 = \sum y_i - \beta_1 \sum x_i$$
$$\beta_0 = \frac{1}{n} \sum y_i - \beta_1 \frac{1}{n} \sum x_i$$

now taking the partial of S with respect to β_1 and setting it to zero we have,

$$\frac{\partial S}{\partial \beta_1} = \sum 2(y_i - \beta_0 - \beta_1 x_i)^1 (-x_i) = 0$$

and dividing through by -2 and rearranging we have,

$$\sum x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum x_i (y_i - \hat{y}_i) = 0$$

but, again we know that $\hat{y}_i = \beta_0 + \beta_1 x_i$. Thus, $x_i = \frac{1}{\beta_1}(\hat{y}_i - \beta_0) = \frac{1}{\beta_1}\hat{y}_i - \frac{\beta_0}{\beta_1}$. Substituting this into the equation above gives the desired result.

$$\sum x_i (y_i - \hat{y}_i) = 0$$

$$\sum \left(\frac{1}{\beta_1}\hat{y}_i - \frac{\beta_0}{\beta_1}\right)(y_i - \hat{y}_i) = 0$$

$$\frac{1}{\beta_1}\sum \hat{y}_i (y_i - \hat{y}_i) - \frac{\beta_0}{\beta_1}\sum (y_i - \hat{y}_i) = 0$$

Now, the second term is zero (by eqn. 1) and so, we arrive immediately at the desired result:

$$\sum \hat{y}_i (y_i - \hat{y}_i) = 0 \qquad (eqn. 2)$$

Now, let's use *eqn. 1* and *eqn. 2* to show that $\sum (\hat{y}_i - \bar{y}_i)(y_i - \hat{y}_i) = 0$ - which was your original question.

$$\sum (\hat{y}_i - \bar{y}_i)(y_i - \hat{y}_i) = \sum \hat{y}_i (y_i - \hat{y}_i) - \bar{y}_i \sum (y_i - \hat{y}_i) = 0$$

edited Mar 18 '14 at 16:07

answered Mar 17 '14 at 23:10
Brad S.
863 3 11

Thank you for a detailed answer! A small error: you say "with respect to β_1 and setting it to zero we have..." but you write β_0 in the partial. I stopped reading there since your answer was very long and the risk for small errors is big. Is the rest correct? – jacob Mar 18 '14 at 14:38

@jacob - good catch. I've corrected the typo. The rest is/was correct. – Brad S. Mar 18 '14 at 16:07

Perfect answer! – jacob Mar 18 '14 at 16:18

$$2\sum(y_i - y_i^*)(y_i^* - \bar{y})$$

$$= 2\sum[y_i(y_i^* - \bar{y}) - y_i^*(y_i^* - \bar{y})]$$

$$= 2\sum Ye_i - 2\bar{Y}\sum e_i$$

$$= 0$$

answered Mar 12 '14 at 17:14
Yilun Zhang
341 1 8

I realised I do not get why y_i turns into a random variable Y. Also, how could y_i^* turn into *the very same* random variable Y? – jacob Mar 14 '14 at 19:35