Why the sum of residuals equals 0 when we do a sample regression by OLS?

That's my question, I have looking round online and people post a formula by they don't explain the formula. Could anyone please give me a hand with that ? cheers

(statistics) (statistical-inference)

asked Sep 15 '13 at 7:37
Maximilian1988
314 3 8

3 Answers

If the OLS regression contains a constant term, i.e. if in the regressor matrix there is a regressor of a series of ones, then the sum of residuals is *exactly* equal to zero, as a matter of algebra.

I will show this for the simple regression, it is trivial to see that it holds for the multivariate regression.

Specify the regression model

$$y_i = a + bx_i + u_i\ ,\ i = 1, \dots, n$$

Then the OLS estimator $(\hat{a}, \hat{b})$ minimizes the sum of squared residuals, i.e.

$$(\hat{a}, \hat{b}) : \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 = \min$$

For the OLS estimator to be the *argmin* of the objective function, it must be the case as a necessary condition, that the first partial derivatives with respect to a and b, evaluated at $(\hat{a}, \hat{b})$ equal zero. For our result, we need only consider the partial w.r.t. a:

$$\frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \Big|_{(\hat{a}, \hat{b})} = 0 \Rightarrow -2 \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i) = 0$$

But $y_i - \hat{a} - \hat{b}x_i = \hat{u}_i$, i.e. is equal to the residual, so we have that

$$\sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i) = \sum_{i=1}^{n} \hat{u}_i = 0$$

The above also implies that if the regression specification does *not* include a constant term, then the sum of residuals will not, in general, be zero.

answered Sep 17 '13 at 21:40
Alecos Papadopoulos
5,296 1 4 17

Thansk a lot, I didnt see your answer till now, really appreciate it :) – Maximilian1988 Sep 19 '13 at 4:13

Sum of residuals doesn't exactly equal 0. However it is a very reasonable assumption that the expectation of the residuals will be 0. This is similar to the case of unbiased estimation, where we want the bias to be 0. Here the residual $y_i - \beta_0 - \beta_1 x_i$ are sometimes negative sometimes positive, but we hope that their overall sum will be 0, so that the estimation is good enough.

answered Sep 15 '13 at 13:01
Abishanka Saha
1,792 1 9

---

take the estimated values from the line of best fit and use these y values to subtract from the original y values then add them up. if it is a good line of best fit then it should approach zero, but bad lines of best fit will be much less or more than zero

answered Oct 16 '13 at 1:35

kasey
1