

Title: Errors in Microsoft's Excel Data Analysis Tool, Regression Module
Pub Date: July 2006 (updated Oct 2017, to include Excel 2013)

Tecolote Research has formally communicated to Microsoft that the following statistics are reported incorrectly by the Excel data analysis regression module. The following table identifies those statistics that are reported incorrectly across various Excel version numbers.

Excel Version No.	No-Intercept Model Summary Output	No-Intercept Model ANOVA Output	Residual Output
7.0 (Excel 95)	R^2 , Adj R^2 , Multiple R	SSR, SST, F-stat	
8.0 ~ 10.0 (Excel 97~2002)	R^2 , Adj R^2 , Multiple R	SSR, SST, F-stat	Standard Residual
11.0 (Excel 2003)	Adj R^2		Standard Residual
15.0 (Excel 2013)	Adj R^2		Standard Residual

There are three commonly used residual statistics: *standardized residual*, *studentized (deleted) residual*, and *unit normal deviate*. However, the "standard residual" reported by Excel's regression module in the Analysis ToolPak is none of these. In Excel 95 (version 7.0), the standard residual is calculated by dividing the residual by the standard error of estimate. This statistic is commonly called *unit normal deviate* and is correctly reported in version 7.0.

Although Excel 97 (version 8.0) still reports "*unit normal deviate*" in the "standard residual" output, this statistic is different from those reported in Excel 95 and all earlier versions of Excel. According to Excel, *this issue occurs because Excel 97 and later versions standardize on a definition of variance that does not consider bias in the sample set*. In other words, the term "n-1" (not "n-p") is used as the degrees of freedom (DF) for the standard error of estimate when reporting the "standard residual." (p stands for the total number of estimated coefficients.)

Tecolote asked Microsoft to produce a Knowledge Base article to explain their position. After a year of email exchanges, Microsoft published the following Knowledge Base article listed below:

<http://support.microsoft.com/kb/920190/en-us>

We expressed our concerns that the article is too vague to be meaningful and we encourage you to do the same (complete their feedback section at the bottom of the Knowledge article). The change that this Knowledge article describes concerning the standard residual output is inconsistent with all the traditional formulas. The Microsoft's version of the calculation consistently uses "n-1" (number of data points -1) to calculate the degrees of freedom for the standard error of estimate rather than n-p (number of data points – number of estimated coefficients). We have the following observations:

1. We cannot find any statistics books, references, or web sites which suggest using "n-1" as the DF for the standard error of estimate (or variance) in regression analysis.
2. We cannot find a commercial regression tool that matches Excel's "Standard Residual" results. Tools such as SAS, SPSS, Systat, MINITAB, CO\$TAT, all agree with each other, but not with Excel.

As for no-intercept equations, errors in R^2 , Multiple R, SSR, SST, and F-stat in Excel 2002 (version 10.0) and all the earlier versions of Excel appear to have been corrected in Excel 2003 (version 11). However, the adjusted R^2 for a regression equation *without* an intercept (i.e., to force the equation through the origin) is still incorrect in all versions. Request an Excel file from ACEIT Support to compare the adjusted R^2 for a no-intercept equation across three different regression tools (i.e., CO\$TAT, Systat, and Excel's ATP). We have notified Microsoft about this error.

We recommend that you choose a tool like MINITAB, SAS, SPSS, S-Plus, or CO\$TAT to evaluate outlier statistics and to develop the regression equation where the intercept is forced through the origin. All these tools will produce the same results. Excel does not.