

December 2012

## Improving How Microsoft Excel Displays Default Extremely Small Probability Values

David A. Larson

*University of South Alabama*, [dlarson@usouthal.edu](mailto:dlarson@usouthal.edu)

Sylvia E. Rogers Ms

*University of South Alabama*, [ser804@jagmail.southalabama.edu](mailto:ser804@jagmail.southalabama.edu)

Follow this and additional works at: <http://epublications.bond.edu.au/ejsie>



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

---

### Recommended Citation

Larson, David A. and Rogers, Sylvia E. Ms (2012) Improving How Microsoft Excel Displays Default Extremely Small Probability Values, *Spreadsheets in Education (eJSiE)*: Vol. 5: Iss. 3, Article 6.

Available at: <http://epublications.bond.edu.au/ejsie/vol5/iss3/6>

This Regular Article is brought to you by the Bond Business School at [ePublications@bond](mailto:EPublications@bond). It has been accepted for inclusion in *Spreadsheets in Education (eJSiE)* by an authorized administrator of [ePublications@bond](mailto:EPublications@bond). For more information, please contact [Bond University's Repository Coordinator](#).

---

# Improving How Microsoft Excel Displays Default Extremely Small Probability Values

## Abstract

Microsoft Excel™'s default method of displaying probability values observed in the sample, in the case of very small values, is confusing to beginning students in statistics. This is because Excel displays them in engineering notation (3.52955E-.05, as opposed to .0000352955, for example). In addition, high quality statistical software programs, such as SAS/STAT™, never display an actual default probability value as small as .0000352955. For any probability value in a sample less than 1-in-10,000, SAS will display the given probability value as <.0001.

## Keywords

Regression, t-test, probability values, IF Statements

## Distribution License



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

Almost all introductory statistics textbooks in Business use Excel as their software of choice. This means that authors routinely display the default Excel generated probability values in their textbook's problem solutions. Using regression as an illustration, it is easy to show that this makes it difficult for beginning students to correctly understand probability values observed in the sample. This problem is caused by the fact that Excel, unlike higher-quality programs like SAS/STAT® software, does not adhere to the standard that the default probability values displayed should never be less than one in ten thousand ( $<.0001$ ) in terms of the lower limit of probability values. Excel, on the other hand, will print very small probability values in engineering notation. This leads to three possible mistakes for students: (1) misinterpreting the E-notation (interpreting  $3.52955E-.05$  as a probability of 3.53, which, of course, is a logical impossibility); (2) false precision in rounding (display actual probability value of  $.0000352955 <.0001$ ); (3) incorrectly rounding back to zero (0.00). In this latter case, it implies the alpha risk is zero when the null is rejected. This implies certainty or causality or proof has been established, which cannot happen with sampling. This is a very serious error. In regard to incorrectly rounding back to zero, it warrants repeating; the many textbooks using Excel commonly round back very small probability values to zero. This implies there is no chance that an alpha or type 1 error (a true null hypothesis has incorrectly been rejected) has been committed when the null hypothesis is rejected in these cases of very small probability values ('when the p-value is low, the null must go' as the saying goes).

## 2. Textbook Example

An example from Jaggia and Kelly (2013), illustrates this shortcoming<sup>1</sup>.

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	292136.9	1	292136.9	73	1E-.08
4	Residual	96045.6	24	4001.9		
5	Total	388182.5	25	15527.3		

Figure 1: Spreadsheets with P-value displayed in engineering notation

<sup>1</sup>. Excel output display in Figure 1 is from page 435 of the Jaggia and Kelly textbook. See Reference 1.

In Figure 1, the probability value observed in the sample (P-value (ob) in cell F3) corresponding to the F-value observed in the sample of 73 (Fob: E3) is displayed as 1E-.08 in engineering notation. This means the actual probability value is .00000001. As illustrated in Figure 2, statistical software programs like SAS display this as <.0001. This can be easily done in Excel using the IF statement shown in Figure 3.

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	292136.9	1	292136.9	73	<.0001
4	Residual	96045.6	24	4001.9		
5	Total	388182.5	25	15527.3		

Figure 2: Spreadsheet with P-value properly displayed

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	292136.91	1	292136.91	73	=IF(FDIST(E3,C3,C4)<0.0001,"<.0001",FDIST(E3,C3,C4))
4	Residual	96045.55	24	4001.9		
5	Total	388182.46	25	15527.3		

Figure 3: Spreadsheet with IF statement

The IF Statement displayed in Figure 3, uses the FDIST function. It is recommended to use FDIST to calculate the probability of an F statistic. The arguments in FDIST are as follows: FDIST(X, DF1, DF2). The F statistic is Fob in cell E3. The argument, DF1, is the degrees of freedom for Regression, which, in this case, is 1 in cell C3. The argument, DF2, is the degrees of freedom Residual, which is the 24 in cell C4 above.

FDIST, in this case, will calculate the value .00000001, which is displayed as <.0001. We instruct the IF STATEMENT to put <.0001 into cell F3 if the actual P-value (ob) is <.0001, and to otherwise display the actual calculated P-value (ob) value in cell F3 above using the FDIST portion which is furthest to the right in the IF STATEMENT.

### 3. Another Textbook Example

Another very common mistake made in textbook Excel illustrations, given a very small value for the actual probability value, is to improperly round the displayed value back to 0.0000. Using an example from the same textbook, this serious mistake is illustrated below in Figure 4.

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	292136.9	1	292136.9	73	<.0001
4	Residual	96045.6	24	4001.9		
5	Total	388182.5	25	15527.3		
6	Parameters	Coeffs	Std Errs	t(ob)	P-value(ob)	
7	Intercept	210.3	91.3	2.3	0.0303	
8	X=Income	10.4411	1.2220	8.5443	0.0000	

Figure 4: Spreadsheet with incorrectly rounded P-value

In cell E8 above for Figure 4, the P-value (ob) for a t-value observed in the sample of 8.5443 (D8) is incorrectly displayed in the text as 0.0000 (E8). A sample of size 26 is used here and, given sampling; the P-value (ob) is never zero. If it were, the decision to reject the null hypothesis would be made with certainty – no alpha risk at all! The actual P-value (ob) in this case is an extremely small number (4.82686E-.09 or, .0000000048, but greater than zero as it must be. Obviously, the P-value (ob) in this instance should be displayed as <.0001 in cell E8 above. The following IF STATEMENT is therefore used as illustrated in Figure 5.

	A	B	C	D	E	F	G
1	ANOVA						
2	Source Variation	SS	DF	MS	Fob	P-value(ob)	
3	Regression	292136.9	1	292136.9	73	<.0001	
4	Residual	96045.6	24	4001.9			
5	Total	388182.5	25	15527.3			
6	Parameters	Coeffs	Std Errs	t(ob)	P-value(ob)		
7	Intercept	210.3	91.3	2.3	0.0303		
8	X=Income	10.4411	1.2220	8.5443	<.0001		
9	IF(TDIST(D8,\$C\$4,2)/2<0.0001,"<.0001",TDIST(D8,\$C\$4,2)/2)						

Figure 5: Spreadsheet with formula example

The logic of the TDIST IF STATEMENT is analogous to the logic of the FDIST IF STATEMENT, except an extra argument for number of tails (1 or 2) for the t-test used is added (= third argument in TDIST = 2, for 2-tailed t-test employed). The formula displayed in row 9 of Figure 5 is copied from cell E8, where it is used to compute the value < .0001.

#### 4. Example for T-Test in Multiple Regression

With respect to the t-test on each model parameter in multiple regression, Figure 5 shows the t-test formula displayed in row 9 above in its broadest context. That is, it covers all the possibilities with these t-tests and permits “dragging and dropping” the formula productively in Excel. It now takes into account that the t-test may be on the lower tail for one or more parameters and also that using the IF STATEMENT shown above, Excel cannot distinguish between the cases of .0001 and <.0001. (The interested reader can verify for her or himself by leaving out the round argument in the formula shown below and then using the given formula above on this example).

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	5918.32	2	2959.2	39.6633	5.1464E-06
4	Residual	895.28	12	74.61		
5	Total	6813.6	14	486.7		
6	Parameters	Coeffs	Std Errs	tob	P-value(ob)	
7	Intercept	21.5874	6.4249	3.3600	0.005674598	
8	X1	6.1341	0.9622	6.3752	3.52955E-.05	
9	X2	24.4079	4.502	5.4216	0.000154585	

Figure 6: T-Test in Multiple Regression

Figure 6 is an example constructed by authors to illustrate how to make the Excel IF STATEMENT distinguish between a probability value of .0001 and one <.0001 (put Round Function in IF STATEMENT).

The above default Excel solution can be made to conform to the industry standard for the lower limit for probability values observed in the sample (and to eliminate confusion for students in interpreting the values in F3 and E7 through E9) as follows (two-tail t-test is used).

	A	B	C	D	E	F
1	ANOVA					
2	Source Variation	SS	DF	MS	Fob	P-value(ob)
3	Regression	5918.32	2	2959.2	39.6633	<.0001
4	Residual	895.28	12	74.61		
5	Total	6813.6	14	486.7		
6	Parameters	Coeffs	Std Errs	tob	P-value(ob)	
7	Intercept	21.5874	6.4249	3.3600	0.0028	
8	X1	6.1341	0.9622	6.3752	<.0001	
9	X2	24.4079	4.502	5.4216	0.0001	

Figure 7: Result after the formulas are applied

The number .0028 in E7 results with the formula because .0057 (E7 to start) is greater than <.0001, but .0057 is cut in half to show the area in each tail of the two-tail t-test in this case. The value in E8 is now <.0001 because, to start, 3.59255E-.05 is <.0001. The value in E9, however, is now exactly .0001 because, to start, the value in E9 is exactly .0002, which, for a two-tail t-test puts .0001 in each tail. The formulas used in this case are displayed below.

	F
2	P-value(ob)
3	=IF(FDIST(E3,C3,C4)<0.0001,"<.0001",FDIST(E3,C3,C4))

Figure 8: Formula for P-value (ob) of Fob

	E
6	P-value(ob)
7	=IF(ROUND(TDIST(ABS(D7), \$C\$4, 2)/2, 4) < 0.0001, "<.0001", ROUND(TDIST(ABS(D7), \$C\$4, 2)/2, 4))
8	=IF(ROUND(TDIST(ABS(D8), \$C\$4, 2)/2, 4) < 0.0001, "<.0001", ROUND(TDIST(ABS(D8), \$C\$4, 2)/2, 4))
9	=IF(ROUND(TDIST(ABS(D9), \$C\$4, 2)/2, 4) < 0.0001, "<.0001", ROUND(TDIST(ABS(D9), \$C\$4, 2)/2, 4))

Figure 9: Formula for P-values (ob) for tob

In Figure 9, the above formula for the P-value (ob) values in cells E7 through E9 is entered into cell E7 and then dragged down through cells E8 and E9 (it is only entered once in E7). The Round function is required (4-decimals) because, without it, Excel default IF STATEMENT cannot distinguish between <.0001 and .0001 (see E8 and E9 above) – it will simply print them both as <.0001 otherwise. The tob value in D7 above must be entered in absolute value [ABS(D7)], otherwise, when it is dragged down, a negative t-value is encountered (parameter is signed minus), and an error message will appear. And, of course, the degrees of freedom cell for the t-tests must be locked in order to correctly drag the formula down starting from Cell E7.

## 5. Concluding Remarks

Students in beginning statistics courses that use Excel typically do not have difficulty ‘plugging in’ the formulas used in this paper. And, by using them, they significantly increase their understanding of probability values observed in the sample. Most importantly, it clearly illustrated to them that a probability value in the sample can never be zero. In almost all the many textbooks that use Excel for statistics, rounding P-values observed in a sample to zero is common place. With SAS software and other similar programs, this mistake cannot happen because <.0001 is the minimum value ever displayed. However, because Excel’s default output is in a dynamic spreadsheet format, incorrectly displaying P-values rounded to 0.0000 in textbooks is a pervasive mistake that is easily correctible.

## Reference

- [1] Jaggia, S., and Kelly, A., (2013), *Business Statistics: Communicating with Numbers*. McGraw-Hill. New York, New York.