

# NOTE F: HELP SCREEN ERRORS

## EXCEL 2000 AND EARLIER EDITIONS

### CONFIDENCE FUNCTION

#### THE FAULTS

The Confidence Function Help Screen as actually shown. Several commentators state that this is in complete error. At best the narrative confuses the reader about confidence intervals. At the worst, it should be completely ignored except for the equations defining the calculations. The numbers in brackets refer to the comments and discussion below

---

CONFIDENCE {1}

[See Also](#)

Returns the confidence interval for a population mean. {2} The confidence interval is a range on either side of a sample mean. {3}. For example, if you order a product through the mail, you can determine, with a particular level of confidence, the earliest and latest the product will arrive. {4}

#### **Syntax**

**CONFIDENCE(alpha,standard\_dev,size)**

**Alpha** is the significance level used to compute the confidence level. The confidence level equals 100\*(1 - alpha)%, or in other words, an alpha of 0.05 indicates a 95 percent confidence level. {5}

**Standard\_dev** is the population standard deviation for the data range and is assumed to be known.

**Size** is the sample size.

#### **Remarks**

- If any argument is nonnumeric, CONFIDENCE returns the #VALUE! Error value.
- If  $\alpha \leq 0$  or  $\alpha \geq 1$ , CONFIDENCE returns the #NUM! Error value.
- If  $\text{standard\_dev} \leq 0$ , CONFIDENCE returns the #NUM! error value.
- If size is not an integer, it is truncated.
- If  $\text{size} < 1$ , CONFIDENCE returns the #NUM! error value.
- If we assume alpha equals 0.05, we need to calculate the area under the standard normal curve that equals (1 - alpha), or 95 percent. This value is  $\pm 1.96$ . The confidence interval is therefore: {6}

$$\bar{x} \pm 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$$

#### **Example**

Suppose we observe that, in our sample of 50 commuters, the average length of travel to work is 30 minutes with a population standard deviation of 2.5. We can be 95 percent confident that the population mean is in the interval: {7}

$$30 \pm 1.96 \left( \frac{2.5}{\sqrt{50}} \right)$$

or:

CONFIDENCE ( 0 . 05 , 2 . 5 , 50 ) equals 0.692951. In other words, the average length of travel to work equals  $30 \pm 0.692951$  minutes, or 29.3 to 30.7 minutes.

---

### **CHANGES**

(The numbers refer to the inserted reference numbers in the help screen above)

(1) Although in a sense this definition has validity, it does not correspond to contemporary definitions of what a confidence interval is. It should be completely rewritten.

I would suggest the following, which fits most of the current long explanatory descriptions in the current literature. The contemporary interpretation avoids the use of probability expressions, which were the main arguments against the use of confidence intervals.

“A confidence interval is characterized by a location value and a width value. The interval is a statement about the likelihood of finding a value of some unknown population parameter within the interval, based on the results of a random sample from the population. The likelihood is associated with an alpha value or significance, and means that  $100*(1-\alpha)\%$  of all repeated random samples drawn from the population will have confidence intervals that include the true value of the unknown population parameter.” (*After text in Blume and Royall (2003) and text in Henderson and Meyer (2001).*)

“The confidence interval either includes the value of the population parameter (100 - alpha% of the time), or it does not. (alpha% of the time).”

For a statement about a normal population mean, the interval location is derived from the sample mean and the width derived from the size of the sample and the known population variance (Blume and Royall (2003)). For a normal population, the interval is symmetric.

(2) There is a profound error here, the value (width) of the interval does not depend on or have anything to do for or with the mean. Contemporary usage now refers to the interval as the full width rather than the half interval (Statistical Services center (2000), Dawson (2003), Blume and Royall (2003), etc.). This usage matches the suggested definition above by Blume and Royall (2003). The definition of alpha is based on the full interval width.

“The function CONFIDENCE returns the normal distribution interval width given an alpha value, the known population standard deviation and the size of the sample. The sample mean value lies at the center of this interval.”

(3) When located on a finite scale, the lower bound on the interval is the mean minus the half-interval and the upper bound on the interval is the mean plus the half-interval. Contemporary statistics uses bounds rather than ranges.

(4) Delete this sentence. Cryer (2001) strongly objects to this since it is a “misstatement of the meaning of a confidence interval”. There is always an uncertainty about the actual “earliest” and “latest” that one would encounter in an application.

{5} and {6}. Fails to indicate that this is a two-tailed interval. For the example given the lower tail and upper tails have equal probabilities of 2.5%. The overall probability is 5%.

(6) Incorrect. The area is 0.95; you need to find a z value corresponding to this area. Should be “If we assume alpha equals 0.05, then we need to calculate the z value corresponding to the area under the standard normal curve that equals (1 - alpha), or 95 percent. This z value is  $\pm 1.96$ . The confidence interval is therefore:”

{7} Henderson and Meyer (2001) say “Although it is tempting to interpret this as ‘the probability that the true parameter  $p$  is in (0.083, 0.317) is 0.95’, this interpretation is incorrect because  $p$  is not a random variable. The correct interpretation is ‘if we repeat many times the process of taking a random sample of size  $n$  from this population, and compute a 95% confidence interval from each sample, then 95% of these confidence intervals will contain the true parameter  $p$ .’ This is the big difference between probability and likelihood concepts.

## **CONFIDENCE LEVEL FOR THE MEAN**

Help on Descriptive Statistics says:

### ***“CONFIDENCE LEVEL FOR MEAN***

Select if you want to include a row in the output table for the confidence level of the mean. In the box, enter the confidence level you want to use. For example, a value of 95 percent calculates the confidence level of the mean at a significance of 5 percent.”

### **COMMENTS:**

The Statistical Service Center (2000) says this is completely wrong, and I agree with their conclusions.

At the very least it should read:

### ***“CONFIDENCE INTERVAL FOR THE MEAN***

Select if you want to include a row in the output table for the confidence interval of the mean. In the box, enter the confidence level you want to use. For example, a value of 95 percent calculates the confidence interval of the mean at a significance of 5 percent.”

To be correct Excel should return the full interval value, not the half-interval value.

Excel uses the inverse of the t distribution (TINV) with (100-“entered percent”)/100 and “Count” – 1 as the degrees of freedom for the function inputs. The central t distribution approach here is correct for normal populations.

With “Standard Deviation” being the square root of the unbiased variance, the form used by Excel ( “Standard Deviation”/”Square Root”(“Count)) is the correct interval.

Excel now only gives one-half the interval width, leaving it up to the user to calculate the lower and upper ends of the interval. The lower and upper limits of the interval should be output from the routine as two cells.

Different textbooks give different apparent intervals (i.e. Hogg 1978), but as long as the interval about  $\bar{x}$  reduces to

$$= \pm ( t\{p, n-1\} * \text{sqrt}(\text{sum of squares}) ) / ( \text{sqrt}(n) * \text{sqrt}(n-1) )$$

It is correct.

There is no direct Excel function for the interval.

Select and format a cell for percentage. Enter the p value as described in para 2. This is cell a.

In the cell for the interval, enter:

$$= \text{TINV}( (1-a) , (\text{COUNT}(\text{rng}) - 1) ) * (\text{SQRT}(\text{DEVSQ}(\text{rng})) / ( \text{SQRT}(\text{COUNT}(\text{rng})) * \text{SQRT}(\text{COUNT}(\text{rng}) - 1) )$$

The upper and lower ends of the interval are

$$=\text{AVERAGE}(\text{rng}) + \text{the interval cell}$$

$$=\text{AVERAGE}(\text{rng}) - \text{the interval cell}$$

Where rng is the range of the data.

If the data set defines a population, then the above is incorrect, and should not appear. It is unlikely that the entered population values are normally distributed. Re-sampling statistics is the only way to construct confidence intervals about the population mean to include probable sample means. Excel does not do re-sampling.

## **CORREL**

The equation CORREL actually uses is different from the one shown. CORREL actually returns a value corresponding to

$$\rho_{x,y} = \frac{(n/n-1) * \text{Cov}(X,Y)}{\sigma_x * \sigma_y}$$

## **COVAR**

Under the syntax section, Array1 and Array2 identify the ranges as integers. This is wrong. The range should identify values, to be consistent with other Help definitions.

The covariance equation is wrong. The summation symbol at the bottom of the sigma symbol, shown as “j - 1” should be “j = 1”

---

## **T-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES ANALYSIS TOOL**

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the means of both data sets are equal; it is referred to as a homoscedastic t-test. {1} You can use t-tests to determine whether two sample means are equal. For more information about options in the **t-Test: Two-Sample Assuming Equal Variances** dialog box, click –

{1} Error, “means” should be “variances”. Reported by several sources. Should read:

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the variances of both data sets are equal; it is referred to as a homoscedastic t-test. {1} You can use t-tests to determine whether two sample means are equal. For more information about options in the **t-Test: Two-Sample Assuming Equal Variances** dialog box, click –

---

## **STDEV**

[See Also](#)

Estimates standard deviation based on a sample. The standard deviation is a measure of how widely values are dispersed from the average value (the mean).

### **Syntax**

**STDEV(number1,number2,...)**

Number1, number2,... are 1 to 30 number arguments corresponding to a sample of a population. You can also use a single array or a reference to an array instead of arguments separated by commas.

Logical values such as TRUE and FALSE and text are ignored. If logical values and text must not be ignored, use the STDEVA worksheet function. {1}

### **Remarks**

- STDEV assumes that its arguments are a sample of the population. If your data represents the entire population, then compute the standard deviation using STDEVP.
- The standard deviation is calculated using the "nonbiased" or "n-1" method.

- STDEV uses the following formula:

- 

### **Example**

Suppose 10 tools stamped from the same machine during a production run are collected as a random sample and measured for breaking strength. The sample values (1345, 1301, 1368, 1322, 1310, 1370, 1318, 1350, 1303, 1299) are stored in A2:E3, respectively. STDEV estimates the standard deviation of breaking strengths for all the tools.

STDEV ( A2 : E3 ) equals 27.46

---

{1} The reference to STDEVA should be left out, since it is confusing, and may lead to erroneous output values. STDEVA will count in every non-blank cell, but if the processor determines that the entry is to be treated as text, a numerical value of zero is assigned. STDEV will properly skip any cell not identified as containing a number.

## **EXCEL 2003**

Microsoft has completely rewritten the help text for excel 2003. There has not yet been a body of comments about the new material.

## **EXCEL 2007**

The help section of Excel 2007 has been revised, leaving a bare minimum on the Excel 2007/Office 2007 software CD's. The body of help is now accessible via the internet and has been expanded. Microsoft has decided to do this to allow for more rapid responses to Help problems.

## **THE FAULTS AND ERRORS IN THE CURRENT 2007 HELP HAVE NOT BEEN IDENTIFIED.**