

UMUC STAT-200 EXCEL Tips

Dr. Brian Killough

EXCEL includes many functions that can easily calculate statistical parameters. This guide provides an explanation and examples of the most common functions you will use in this course. First, remember how to enter functions in an EXCEL cell ... any function must start with an equal sign (=). Then you add the function name and the data values (in parentheses).

(1) Standard Deviation

STDEV: Calculates the standard deviation of a sample set of data. The data can be entered directly into the equation using commas to separate the values, or you can use a range designation to highlight an entire set of data. Examples are shown below for calculating the mean (average), variance and standard deviation of a sample set of data.

NOTE: The EXCEL formula for standard deviation assumes that you divide by (n-1) when calculating the sample variance, where n = number of data values. The reason we use "n-1" rather than "n" is that we do not know the true mean of the population and we are only dealing with a sample set of data. So, the standard deviation equation usually uses "n-1" as the divisor when finding the sample variance and then (by taking the square root) the standard deviation. This may be different than some online tools that calculate standard deviation, so be careful!

$$\text{AVERAGE}(B5:B14) = 74.90$$

$$\text{VARIANCE}(B5:B14) = 25.66$$

$$\text{STDEV}(B5:B14) = 5.07 \text{ or}$$

$$\text{STDEV}(73,73,77,71,70,79,82,75,67,82) = 5.07$$

	A	B	C	D
1				
2		Standard Deviation Example		
3				
4		Data (X)	(X-Mean)^2	
5		73	3.61	
6		73	3.61	
7		77	4.41	
8		71	15.21	
9		70	24.01	
10		79	16.81	
11		82	50.41	
12		75	0.01	
13		67	62.41	
14		82	50.41	
15				
16		sum	230.90	
17		n-1	9	(data values) - 1
18		sum/(n-1)	25.66	variance
19		SQRT(variance)	5.07	standard deviation
20				
21		=AVERAGE	Mean (Average)	74.90
22		=VAR	Variance	25.66
23		=STDEV	Standard Deviation	5.07

(2) Binomial Distribution

BINOMDIST: Calculates the binomial distribution probability. Use this function in problems with a fixed number of tests or trials, when the outcomes of any trial are only success or failure, when trials are independent, and when the probability of success is constant throughout the experiment.

Function: BINOMDIST(number_s, trials, probability_s, cumulative)

number_s = the number of "successes" in the trial

trials = the number of trials in the experiment

probability_s = the probability of a success in a single trial

cumulative = 1=cumulative, 0=individual

Example: We would like to calculate the cumulative probability that up to 4 of 12 students will attend an event based on an 18% probability that any single student will attend the event. So, we will use the following arguments in our function ...

$$\text{BINOMDIST}(4,12,0.18,1) = 0.951 \text{ or } 95.1\%$$

This is exactly the same result if we were add up the individual probabilities of having 1,2,3 or 4 students attend the event. Notice the last argument is 0 in these cases, since we are calculating the individual probabilities of several different numbers of success.

$$\text{BINOMDIST}(4,12,0.18,1) = \text{BINOMDIST}(1,12,0.18,0) + \text{BINOMDIST}(2,12,0.18,0) + \text{BINOMDIST}(3,12,0.18,0) + \text{BINOMDIST}(4,12,0.18,0)$$

(3) Normal Distribution

NORMSDIST: Calculates the standard normal cumulative probability for a given Z-score. For negative values of Z, the distribution will reflect the left-tail probability to the left of the mean. For positive values of Z, the distribution will reflect the cumulative probability from the left tip to an area right of the mean. In order to get the right-tail probability, one must subtract the value from 1.

Function: $\text{NORMSDIST}(\text{z-score}) = \text{probability}$

Example 1: $\text{NORMSDIST}(-1.2) = 0.115$ (left tail)

Example 2: $\text{NORMSDIST}(1.5) = 0.933$ (cumulative probability)

The “right tail” answer for the second example is $1 - 0.933 = 0.067$.

NORMSINV: Calculates the inverse of the standard normal distribution for a given probability.

Function: $\text{NORMSINV}(\text{probability}) = \text{z-score}$

Example 1: $\text{NORMSINV}(0.115) = -1.2$ (notice this is the inverse of example-1 above)

Example 2: $\text{NORMSINV}(0.933) = 1.5$ (notice this is the inverse of example-2 above)

NOTE: In order to find a **confidence interval** we need to consider two tails in the normal distribution. For example, with a 99% confidence interval, there will be a total of 1% of data outside of the distribution. So, 0.5% will be on the left-tail and 0.5% will be on the right tail. In order to find the proper z-score, we need to use the cumulative probability that includes the left-tail and the full distribution.

99% confidence ($a = \text{tail area} = 0.01$), Cumulative Probability = $1 - a/2 = 1 - 0.005 = 0.995$
 $\text{NORMSINV}(0.995) = 2.576$

95% confidence ($a = \text{tail area} = 0.05$), Cumulative Probability = $1 - a/2 = 1 - 0.025 = 0.975$
 $\text{NORMSINV}(0.975) = 1.960$

90% confidence ($a = \text{tail area} = 0.10$), Cumulative Probability = $1 - a/2 = 1 - 0.05 = 0.950$
 $\text{NORMSINV}(0.95) = 1.645$

(4) Student's T-Distribution

TDIST: Calculates the Student's T-distribution, assuming you know the random variable (X), the number of samples, and whether the distribution is one-tailed or two-tailed. Use one-tail distributions when the alternate hypotheses is an inequality (< or >) or use the two-tail distribution when the alternate hypothesis is a not-equal sign (\neq). We typically use this function to find the P-value (probability) to compare with the probability of a Type-1 error (α). Remember that "degrees of freedom" = samples - 1.

Function: TDIST(X,degrees_freedom,tails) = probability where X = test statistic or t-value

Example: TDIST(1.99,80,2) = 0.05 where X=1.99, samples=81 (Degrees of Freedom = 80), and Tails = 2

TINV: Calculates the two-tailed inverse (t-value) of the Student's T-distribution, assuming you know the probability range (or confidence level) and the number of samples (where "degrees of freedom" is samples-1). We typically use this to find a confidence interval with upper and lower bounds.

Function: TINV(probability,degrees_freedom)

Example: TINV(0.05,80) = 1.99 (notice this is the inverse of the example above)

T.INV: Calculates the one-tailed inverse (t-value) of the Student's T-distribution, assuming you know the probability range (or confidence level) and the number of samples (dof=samples-1).

Function: T.INV(probability,degrees_freedom)

Example: T.INV(0.05,19) = -1.73 (this is the one-tailed t-value for a 5% confidence level and 20 samples)

(5) Chi Square Distribution

The Chi Square distribution can be used to test whether observed data differ significantly from theoretical expectations, which is often called "Goodness of Fit" testing. Tests using a Chi Square distribution are always right-tailed, since the distribution is nonsymmetrical and skewed to the right.

CHISQ.DIST.RT: Calculates the right-tail probability of the Chi Square distribution assuming you know the random variable (X) and the degrees of freedom.

Function: CHISQ.DIST.RT(X,degrees_freedom)

Example: CHISQ.DIST.RT(5.333,5) = 0.377

CHISQ.INV.RT: Calculates the inverse of the right-tail probability of the Chi Square distribution assuming you know the probability range (or confidence level) and the number of categories (dof=categories-1).

Function: CHISQ.INV.RT(probability,degrees_freedom)

Example: CHISQ.INV.RT(0.377,5) = 5.33

(6) Linear Regression

SLOPE: Calculates the slope of the best-fit linear regression line.

INTERCEPT: Calculates the intercept of the best-fit linear regression line.

RSQ: Calculates the r-squared correlation of the best-fit linear regression line.

CORREL: Calculates the correlation coefficient (r) of the best-fit linear regression line.

STEYX: Calculates the standard error of the best-fit linear regression line (standard error of the estimate).

Given the following set of X-Y data points, here are some examples for using these functions.

$\text{SLOPE}(\text{y-values}, \text{x-values}) = \text{SLOPE}(\text{B2:B8}, \text{A2:A8}) = 0.9286$

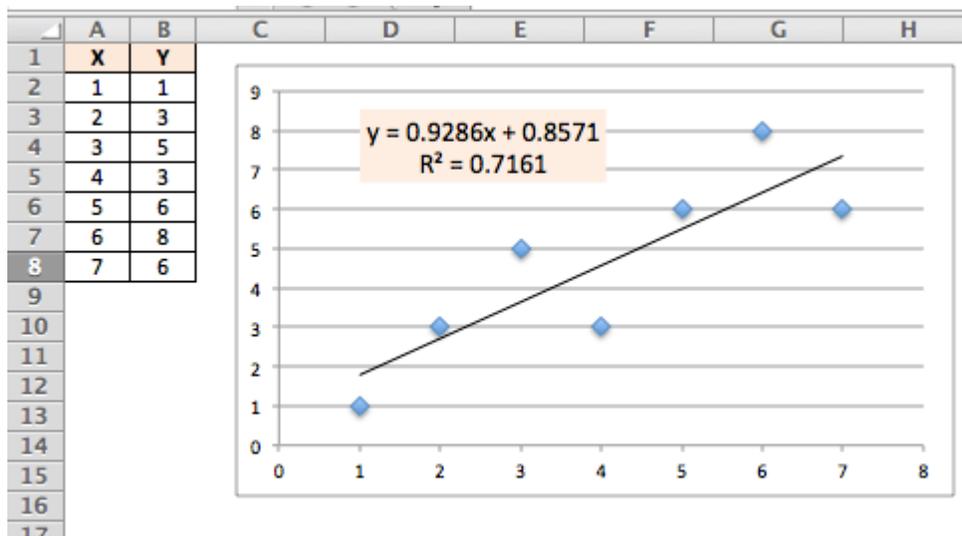
$\text{INTERCEPT}(\text{y-values}, \text{x-values}) = \text{INTERCEPT}(\text{B2:B8}, \text{A2:A8}) = 0.8571$

$\text{RSQ}(\text{y-values}, \text{x-values}) = \text{RSQ}(\text{B2:B8}, \text{A2:A8}) = 0.7161$

* Note that the Correlation (r) is the square root of RSQ, or $\text{SQRT}(0.7161) = 0.8462$

$\text{CORREL}(\text{y-values}, \text{x-values}) = \text{CORREL}(\text{B2:B8}, \text{A2:A8}) = 0.8462$

$\text{STEYX}(\text{y-values}, \text{x-values}) = \text{STEYX}(\text{B2:B8}, \text{A2:A8}) = 1.3836$



(7) F Distribution

FDIST: Calculates The F-distribution (named after Sir Ronald Fisher) which is used for Analysis of Variance (ANOVA) calculations. Assumes you know the following variables:

X = test statistic (this is F in the ANOVA table)

degrees_freedom1 = df (numerator) = (number of groups) – 1

degrees_freedom2 = df (denominator) = (total observations) – (number of groups)

Function: $\text{FDIST}(\text{X}, \text{degrees_freedom1}, \text{degrees_freedom2})$

Example: $\text{FDIST}(3.465, 3, 132) = 0.018$