

# Exploratory Data Analysis in the Classroom

## Essential Spreadsheet Functions

Count the number of numerical variables in a distribution of scores.

**=COUNT(range)**

**range** the selection of cells, usually a column, that contains the numerical information that you want to count.

Count the number of alphabetic variables in a distribution of scores.

**=COUNTA(range)**

**range** the selection of cells, usually a column, that contains the information that you want to count.

Count the number of variables in a distribution based on a filtering criterion.

**=COUNTIF(range, criteria)**

**range** the selection of cells, usually a column, that contains the information that you want to count.

**criteria** the specific criteria you want to count.

Count the number of variables in a distribution based on multiple filtering criteria.

**=COUNTIFS(range, criteria, range, criteria...)**

**range** the selection of cells, usually a column, that contains the information that you want to count.

**criteria** the specific criteria you want to count.

Calculate the percentage.

**=count/total**

**count** the number of units in a subpopulation.

**total** the total number of units in a population.

Calculate the average of a distribution of scores.

**=AVERAGE(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

Calculate the average of a distribution of scores based on a sorting criteria.

**=AVERAGEIF(range, criteria, average\_range)**

<b>range</b>	the selection of cells, usually a column, that contains the criteria you wish to examine.
<b>criteria</b>	the specific criteria you want to summarize.
<b>average_range</b>	the selection of cells, usually a column, that contains the data you want to find the average for.

Calculate the average of a distribution of scores based on multiple sorting criteria.

**=AVERAGEIFS(average\_range, range, criteria, range, criteria...)**

<b>average_range</b>	the selection of cells, usually a column, that contains the data you want to find the average for.
<b>range</b>	the selection of cells, usually a column, that contains the data that you want to summarize.
<b>criteria</b>	the specific criteria you want to summarize.

Identify the median of a distribution of scores.

**=MEDIAN(range)**

<b>range</b>	the selection of cells, usually a column, that contains the data that you want to summarize.
--------------	--

Identify the mode of a distribution of scores.

**=MODE(range)**

<b>range</b>	the selection of cells, usually a column, that contains the data that you want to summarize.
--------------	--

Identify a single mode in a distribution of scores.

**=MODE.SNGL(range)**

<b>range</b>	the selection of cells, usually a column, that contains the data that you want to summarize.
--------------	--

Identify multiple modes in a distribution of scores.

**=MODE.MULTI(range)**

<b>range</b>	the selection of cells, usually a column, that contains the data that you want to summarize.
--------------	--

## Exploratory Data Analysis in the Classroom: Essential Spreadsheet Functions

Identify the minimum score in a distribution of scores.

**=MIN(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

Identify the maximum score in a distribution of scores.

**=MAX(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

Identify the minimum score in a distribution of scores based on multiple sorting criteria.

**=MINIFS(min\_range, criteria\_range, criteria...)**

**min\_range** the selection of cells, usually a column, that contains the data that you want to summarize.

**criteria\_range** the selection of cells, usually a column, that contains the criteria you want to examine.

**criteria** The specific data point you want to examine.

Identify the maximum score in a distribution of scores based on multiple sorting criteria.

**=MAXIFS(max\_range, criteria\_range, criteria...)**

**max\_range** the selection of cells, usually a column, that contains the data that you want to summarize.

**criteria\_range** the selection of cells, usually a column, that contains the criteria you want to examine.

**criteria** The specific data point you want to examine.

Calculate the range of a distribution of scores.

**=MAX(range) - MIN(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

Calculate the standard deviation of a distribution of scores representing a whole population.

**=STDEV.P(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

## Exploratory Data Analysis in the Classroom: Essential Spreadsheet Functions

Calculate the standard deviation of a distribution of scores representing a sample of a population.

### **=STDEV.S(range)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

Identify the quartile divisions of a distribution of scores.

### **=QUARTILE.INC (range, quart)**

**range** the selection of cells, usually a column, that contains the data that you want to summarize.

**quart** a number representing the quartile you wish to report (0, 1, 2, 3, or 4).

Calculate the correlation between two distributions of scores.

### **=PEARSON(range1, range2)**

**range1** the first selection of cells, usually a column, that contains the data that you want to examine.

**range2** the second selection of cells, usually a column, that contains the data that you want to examine.

Perform a t-test on two distributions of scores.

### **=T.TEST(range1, range2, tails, type)**

**range1** the first selection of cells, usually a column, that contains the data that you want to examine.

**range2** the second selection of cells, usually a column, that contains the data that you want to examine.

**tails** set to "1" if the t-test uses the one-tailed distribution or "2" if the t-test uses the two-tailed distribution.

**type** Set to "1" if you want to calculate a paired-sample t-test, "2" if you want to calculate a two-sample with equal variance t-test, or "3" if you want to calculate a two-sample with unequal variance t-test.

Calculate the pooled standard deviation of two distributions of scores.

### **=SQRT(((n1-1)\*SD1^2+(n2-1)\*SD2^2)/(n1+n2-2))**

**n<sub>1</sub>** the count of range 1.

**SD<sub>1</sub>** the standard deviation of range 1.

**n<sub>2</sub>** the count of range 2.

**SD<sub>2</sub>** the standard deviation of range 2.

## Exploratory Data Analysis in the Classroom: Essential Spreadsheet Functions

Calculate the effect size between two distributions of scores.

$$=(M1-M2)/SDp$$

**M<sub>1</sub>** the mean (or average) of range 1.  
**M<sub>2</sub>** the mean (or average) of range 2.  
**SDp** the pooled standard deviation.