# Confirmatory Data Analysis in Excel: Takeaways



## Syntax

- The `RANDARRAY()` function can be used to create an array of random numbers. This can be helpful to simulate basic experiments using Excel.

```
=RANDARRAY([rows],[columns],[min],[max],[whole_number])
```

### Where:

- `[rows]` : The number of rows to be returned
- `[columns]` : The number of columns to be returned
- `[min]` : The minimum number you would like returned
- `[max]` : The maximum number you would like returned
- `[whole_number]` : Return a whole number or a decimal value (True or False)

## Concepts

- The **law of large numbers** tells us that as our sample size grows, our sample average should get closer and closer to the population average.
- The **central limit theorem** tells us that, given a sufficiently large sample, the distribution of *any* variable's sample means will be normally distributed . . . not just those that are normally distributed themselves.
- Using a 95% confidence interval, we can derive a range of values within which we'd expect to find the population's value 95% of the time.
- The **independent samples t-test** is used to check for a significant difference in the means of two variables.
- The **R-square** of a linear regression model tells us what percentage of variance in the dependent variable can be explained by the independent variable. This will always be a number between 0 and 1.

## Resources

- Read Conrad Carlberg's *Regression Analysis Microsoft Excel* for an in-depth look at regression in Excel.
- Read George Mount's *Advancing into Analytics* for an overview of data analysis in Excel.

As you continue building your analytics skills in Excel, you may consider branching into Python and R. Be sure to check Dataquest's learning options on those resources!