Identifying Statistical Test Types

It's been a few days since Jeremy began teaching himself to learn R, and he's beginning to feel more confident. He knows there's still a lot to learn about data analysis, statistics, and reporting, but he feels ready to start cutting his teeth on some real data.

Jeremy asked Colleen to show him some of the datasets they were working on, but he didn't expect each file to be so different! Although Jeremy has worked with a variety of datasets during his time at AutosRUs, these aggregated data tables are an entirely new format. Fortunately, Jeremy knows that with some time and patience, he can use R to explore and demystify these data tables and prepare himself for some statistical testing!

When presented with a new or foreign dataset, Jeremy knows it's good practice to familiarize yourself with each data column, the dimensions of the data, and the overall characteristics of the dataset. He's going to start this section of his analysis by exploring the different characteristics to look out for in a dataset, and what each characteristic means when it comes to analysis.

For the remainder of this module, we'll focus on data analytics using hypothesis testing and statistics. Although the discipline of statistics has many specializations and nearly limitless applications, there are only a few concepts required to get started. In this introduction, we'll cover some core statistical concepts such as:

- mathematical data types

- null and alternative hypothesis

- p-values and hypothesis testing

- t-test of the means

- correlation and linear regression tests

- comparing frequency distribution using chi-squared test

Due to the structured nature of statistical testing, if we can determine any two components of a statistical analysis (input variables, analytical question, or statistical test), we can infer the third. To simplify the process of inferring what component is needed, you can use a statistical test lookup table such as the following:

Statistical Test	Input Variable Type		Analytical Question
	Independent	**Dependent**	

	No. of Variables	Data Type	No. of Variables	Data Type	
One-sample t-test	1	Dichotomous (Population or Sample)	1	Continuous	Is there a statistical difference between the mean of the sample distribution and the mean of the population distribution?
Two-sample t-test	1	Dichotomous (Sample A versus Sample B)	1	Continuous	Is there a statistical difference between the distribution means from two samples?
ANOVA	1+	Categorical	1	Continuous	Is there a statistical difference between the distribution means from multiple samples?
Simple linear regression	1	Continuous	1	Continuous	Can we predict values for a dependent variable using a linear model and values from the independent variable?
Multiple linear regression	2+	Continuous	1	Continuous	How much variance in the dependent variable is accounted for in a linear combination of independent variables?
Chi-squared	1	Categorical	1+	Categorical	Is there a difference in categorical frequencies

					between groups?

There are several helpful lookup tables that can reinforce statistical concepts. For your convenience, we've aggregated many of these lookup tables into a single cheat sheet you can download and use for the remainder of the module.

[Download Stats_Cheat_Sheet.pdf](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_15/Stats_Cheat_Sheet.pdf)   (https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_15/Stats_Cheat_Sheet.pdf)

In addition, the remaining sections will provide supplementary notes on the in-depth statistical concepts. These will focus on common statistical interview questions, analytical "gotchas," and other pertinent contexts to bolster your understanding of each statistical concept. If this is your first time learning statistical concepts, or if you're still brushing off the mental cobwebs, we suggest you delay studying these optional notes until you're comfortable with this module's core statistical concepts. Take your time with each section, and refer to the statistics cheat sheet for help.

IMPORTANT

Once you're more familiar with statistical analysis, start searching for more elaborate statistical lookup tables with non-normal, generalized statistical tests. As you practice characterizing datasets, asking questions, determining a hypothesis, and testing using statistics, the process will become easier.