

TOOL

Selecting Techniques to Quantify Uncertainty





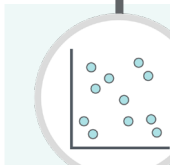
As a data scientist, whenever you generalize the results from your sample to a larger population, you should quantify the uncertainty associated with your results and report it along with the output of your summarization and visualization analyses. The techniques for quantifying uncertainty can help you test hypotheses about your data by determining whether the data you collected from your sample indicates a new discovery, or whether your result was due to random chance. Choosing the correct technique is critical to accurately quantifying uncertainty.

Use the steps and the table below to help you determine which methods to use:

1. Identify the **question type and variable type** appropriate for your data set and question in the first section of the table below. Note the column corresponding to your selections.
2. Continuing down the column you selected, do the following:
 - a. Identify the appropriate **summarization method**. This determines how to summarize the data you will use in the distribution to assess uncertainty.
 - b. Identify the **method for assessing uncertainty**. This determines the technique you will use to create the distribution with which to assess uncertainty.
3. If you need more help using these methods, refer to the **tool** listed in the appropriate column.



1. Choose your question and variable type

Question type	Univariate		Bivariate		
	1	2	3	4	5
Variable type	Categorical	Numerical	Categorical Categorical	Categorical Numerical	Numerical Numerical
					
	Barplot	Histogram	Stacked barplot	Side-by-side boxplot	Scatterplot
Summarization method	Proportion	Mean/Median	Difference of two proportions	Difference of two means	Correlation coefficient
	Method to assess uncertainty	Method to assess uncertainty	Method to assess uncertainty	Method to assess uncertainty	Method to assess uncertainty
	Sample proportion distribution across many simulated data sets	Sample mean or median distribution across many bootstrapped data sets	Distribution of sample difference of two proportions across many simulated data sets	Distribution of sample difference of two means across many permuted data sets	Sample correlation coefficient distribution across many bootstrapped data sets
Tool	Simulation tool (Module 1)	Bootstrap tool (Module 2)	Hypothesis test tool (Module 1)	Permutation tool (Module 2)	Regression uncertainty tool (Module 2)

2. Identify the appropriate methods

3. Get more help

