

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Relationships C→Q > Case C→Q > Inference for Relationships Introduction

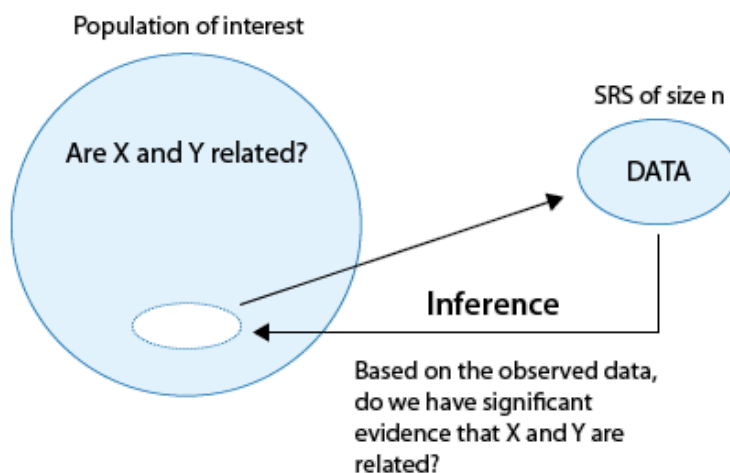
🔖 Bookmark this page

Inference for Relationships Introduction

In the previous Inference sections we performed inference for one variable. More specifically, we learned about inference for the population proportion p (when the variable of interest is categorical) and inference for the population mean μ (when the variable of interest is quantitative). In the previous Inference sections we were also exposed to the following three forms of inference which will continue to be central as we move forward in the course:

- **Point estimation**—estimating an unknown parameter with a single value that is computed from the sample.
- **Interval estimation**—estimating an unknown parameter by an interval of plausible values. To each such interval we attach a level of confidence that indeed the interval captures the value of the unknown parameter and hence the name confidence intervals.
- **Hypothesis testing**—a four-step process in which we are assessing evidence provided by the data in favor or against some claim about the population parameter.

Our next (and final) goal for this course is to perform inference about relationships between two variables in a population, based on an observed relationship between variables in a sample. Here is what the process looks like:



We are interested in studying whether a relationship exists between the variables X and Y in a population of interest. We choose a random sample and collect data on both variables from the subjects. Our goal is to determine whether these data provide strong enough evidence for us to generalize the observed relationship in the sample and conclude (with some acceptable and agreed-upon level of uncertainty) that a relationship between X and Y exists in the entire population.

The primary inference form that we will use in this module, then, is hypothesis testing. Conceptually, across all the inferential methods that we will learn, we'll test some form of:

H_0 : *There is no relationship between X and Y*

H_A : *There is a significant relationship between X and Y*

(We will also discuss point and interval estimation, but our discussion about these forms of inference will be framed around the test.)

Recall that in the module about examining the relationship between two variables in the Exploratory Data Analysis unit, our discussion was framed around the role-type classification table. This part of the course will be structured exactly in the same way.

In other words, we will go through 3 sections corresponding to cases $C \rightarrow Q$, $C \rightarrow C$, and $Q \rightarrow Q$ in the table below.

		Response	
		Categorical	Quantitative
Explanatory	Categorical	$C \rightarrow C$	$C \rightarrow Q$
	Quantitative	$Q \rightarrow C$	$Q \rightarrow Q$

(Recall that case $Q \rightarrow C$ is not discussed in this course.)

In total, we will introduce 5 inferential methods: three in case $C \rightarrow Q$ (corresponding to a division of this case into 3 sub-cases) and one in each of the cases $C \rightarrow C$ and $Q \rightarrow Q$.

Unlike the previous part of the course on Inference for One Variable, where we discussed in some detail the theory behind the machinery of the test (such as the null distribution of the test statistic, under which the p-values are calculated), in the 5 inferential procedures that we will introduce in Inference for Relationships, we will discuss much less of that kind of detail. The principles are the same, but the

details behind the null distribution of the test statistic (under which the p-value is calculated) become more complicated and require knowledge of theoretical results that are definitely beyond the scope of this course.

Instead, within each of the five inferential methods we will focus on:

- When the inferential method is appropriate for use.
- Under what conditions the procedure can safely be used.
- The conceptual idea behind the test (as it is usually captured by the test statistic).
- How to use software to carry out the procedure in order to get the p-value of the test.
- Interpreting the results in the context of the problem.

Also, we will continue to introduce each test according to the four-step process of hypothesis testing. We are now ready to start with Case $C \rightarrow Q$.

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved