

 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 $Q \rightarrow Q$ > Case $Q \rightarrow Q$ > Case $Q \rightarrow Q$: Overview

 Bookmark this page

Case $Q \rightarrow Q$: Overview

Learning Objective: Choose the appropriate inferential method for examining the relationship between two variables and justify the choice.

Learning Objective: In a given context, carry out the appropriate inferential method for comparing relationships and draw the appropriate conclusions.

Inference for the Linear Relationships between Two Quantitative Variables

Overview

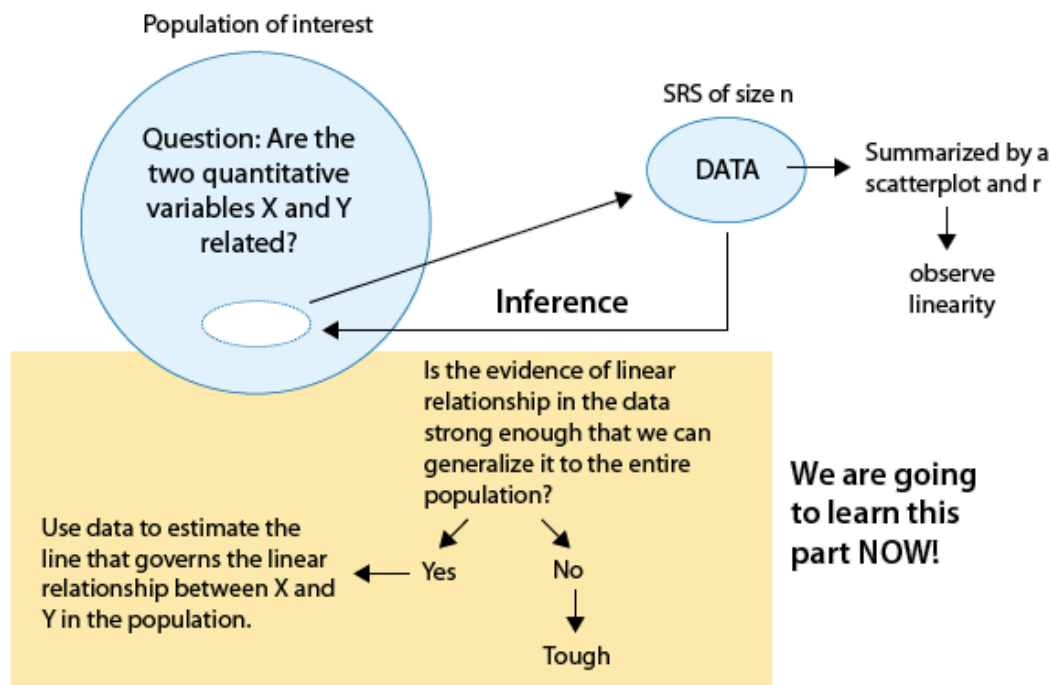
In inference for relationships, so far we have learned inference procedures for both cases $C \rightarrow Q$ and $C \rightarrow C$ from the role/type classification table below. The last case to be considered in this course is case $Q \rightarrow Q$, where both the explanatory and response variables are quantitative. (Case $Q \rightarrow C$ requires statistical methods that go beyond the scope of this course, but might be part of extension modules in the future).

		Response	
		Categorical	Quantitative
Explanatory	Categorical	✓ C → C	✓ C → Q
	Quantitative	Q → C	Q → Q

In the Exploratory Data Analysis section, we examined the relationship between sample values for two quantitative variables by looking at a scatterplot and focused on the linear relationship by supplementing the scatterplot with the correlation coefficient r .

There was no attempt made to claim that whatever relationship was observed in the sample necessarily held for the larger population from which the sample originated. Now that we have a better understanding of the process of statistical inference, we will present the method for inferring something about the relationship between two quantitative variables in an entire population, based on the relationship seen in the sample. In particular, the method will focus on **linear** relationships and will answer the following question: Is the observed linear relationship due to a true linear relationship between two variables in the population, or could it be that we obtained this kind of pattern in the data just by chance?

If we conclude that we can generalize the observed linear relationship to the entire population, we will then use the data to estimate the line that governs the linear relationship between the two variables in the population, and use it to make predictions. The following figure summarizes this process:



Note that the figure summarizes the whole process. Let's review it again.

- We start by asking whether the two quantitative variables are related (in any way).
- We collect data, and when we summarize them with a scatterplot and the correlation r , we observe a linear relationship.

Then we get to the inference part of the process, which we are going to learn here:

- We will carry out a test that will tell us whether the observed linear relationship is significant (i.e., can be generalized to the entire population).
 - If the observed linear relationship is not significant—too bad.
 - If the observed linear relationship is significant, we can use the data to estimate the line that governs the linear relationship between X and Y in the population, and can use it to make predictions (see comment 1 below).

Comments

1. We estimate the line that governs the linear relationship between X and Y in the population by the line that best fits the linear pattern in our observed data. Recall that in the Exploratory Data Analysis unit we've actually already learned how to find the least squares regression line—the line that best fits the observed data. You can now see that finding the least squares regression line actually belongs to the inference unit, and while it is true that it is the line that best fits (in some sense) the observed data, it is really an estimate of the true linear relationship that exists in the population. The good thing is that we already learned how to obtain this line, so we'll only need to review it.

2. This section on regression will be very qualitative in nature and will rely mostly on conceptual ideas and on output. An extension module to this course, which will go deeper into the inferential processes of regression, will exist in the near future.

3. This section will be organized around a leading example. At some stages along the way, you'll be directed to an activity, where you'll get to have hands-on practice with a different example.

Open Learning Initiative [↗](#)



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

© All Rights Reserved