

⚠ 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 > Case C→Q: Overview

🔖 Bookmark this page

Case C→Q: Overview

Learning Objective: Identify and distinguish among cases where use of calculations specific to independent samples, matched pairs, and ANOVA are appropriate.

Recall the role-type classification table framing our discussion on inference about the relationship between two variables.

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

We start with case C→Q, where the explanatory variable is categorical and the response variable is quantitative. Recall that in the Exploratory Data Analysis unit, examining the relationship between X and Y in this case amounts, in practice, to comparing the distributions of the (quantitative) response Y for each value (category) of the explanatory X. To do that, we used side-by-side boxplots (each representing the distribution of Y in one of the groups defined by X), and supplemented the display with the corresponding descriptive statistics.

What will we do in inference? To understand the logic, we'll start with an example and then generalize.

Example: GPA and Year in College

Say that our variable of interest is the GPA of college students in the United States. From the previous module we know that since GPA is quantitative, we do inference on μ , the (population) mean GPA among all U.S. college students. Since this module is about relationships, let's assume that what we are really interested in is not simply GPA, but the relationship between:

X : year in college (1 = freshmen, 2 = sophomore, 3 = junior, 4 = senior) and

Y : GPA

In other words, we want to explore whether GPA is related to year in college. The way to think about this is that the population of U.S. college students is now broken into **4 sub-populations**: freshmen, sophomores, juniors and seniors. Within each of these four groups, we are interested in the GPA.

The inference must therefore involve the 4 sub-population means:

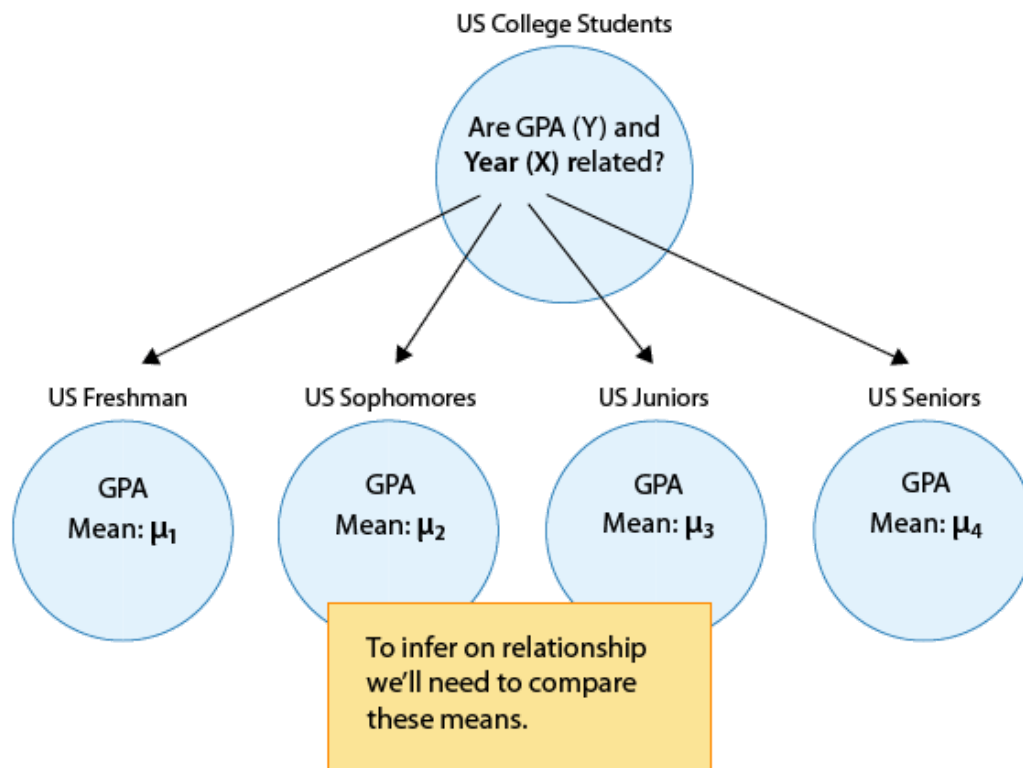
μ_1 : mean GPA among freshmen in the United States.

μ_2 : mean GPA among sophomores in the United States

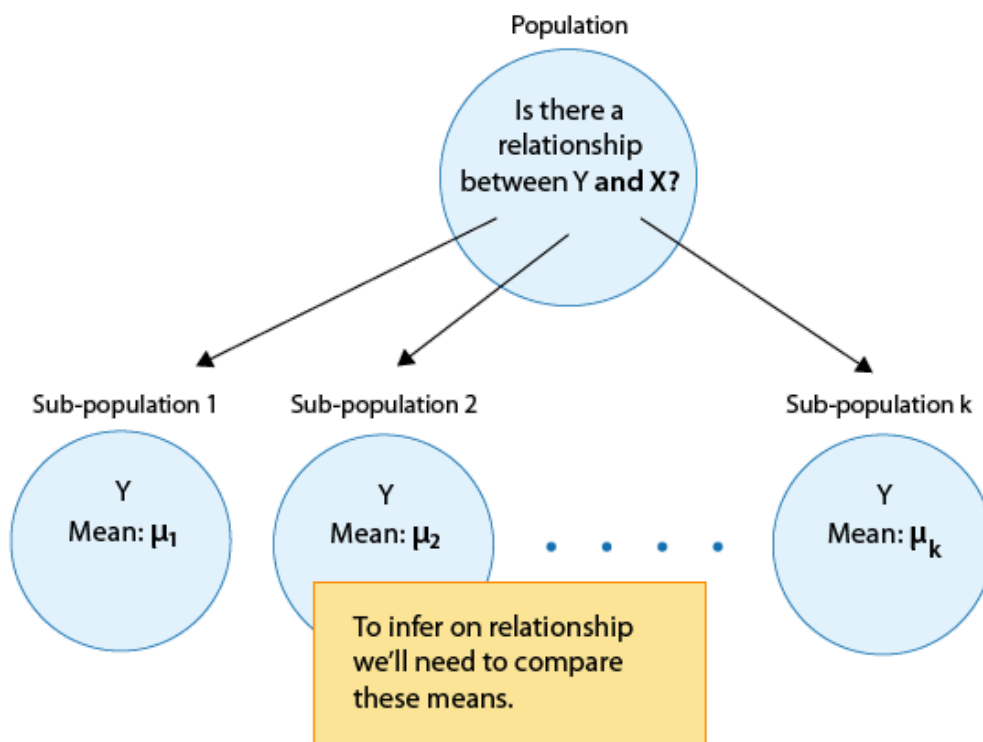
μ_3 : mean GPA among juniors in the United States

μ_4 : mean GPA among seniors in the United States

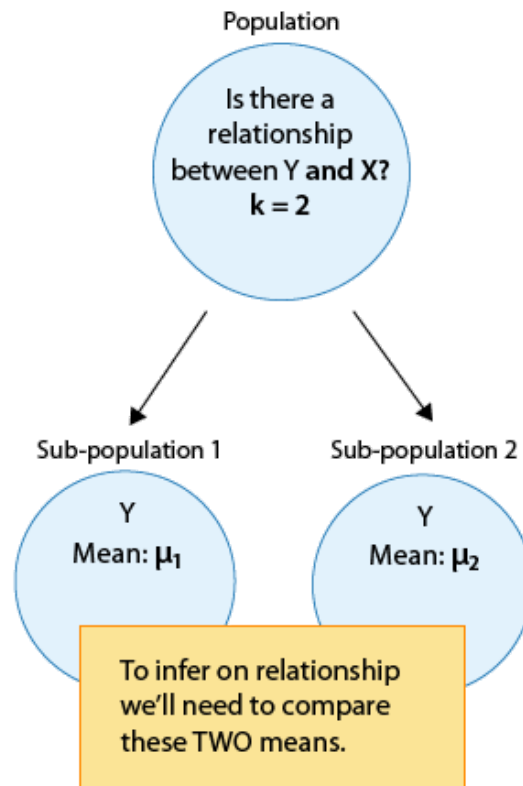
It makes sense that the inference about the relationship between year and GPA has to be based on some kind of comparison of these four means. If we infer that these four means are not all equal (i.e., that there are some differences in GPA across years in college) then that's equivalent to saying GPA is related to year in college. Let's summarize this example with a figure:



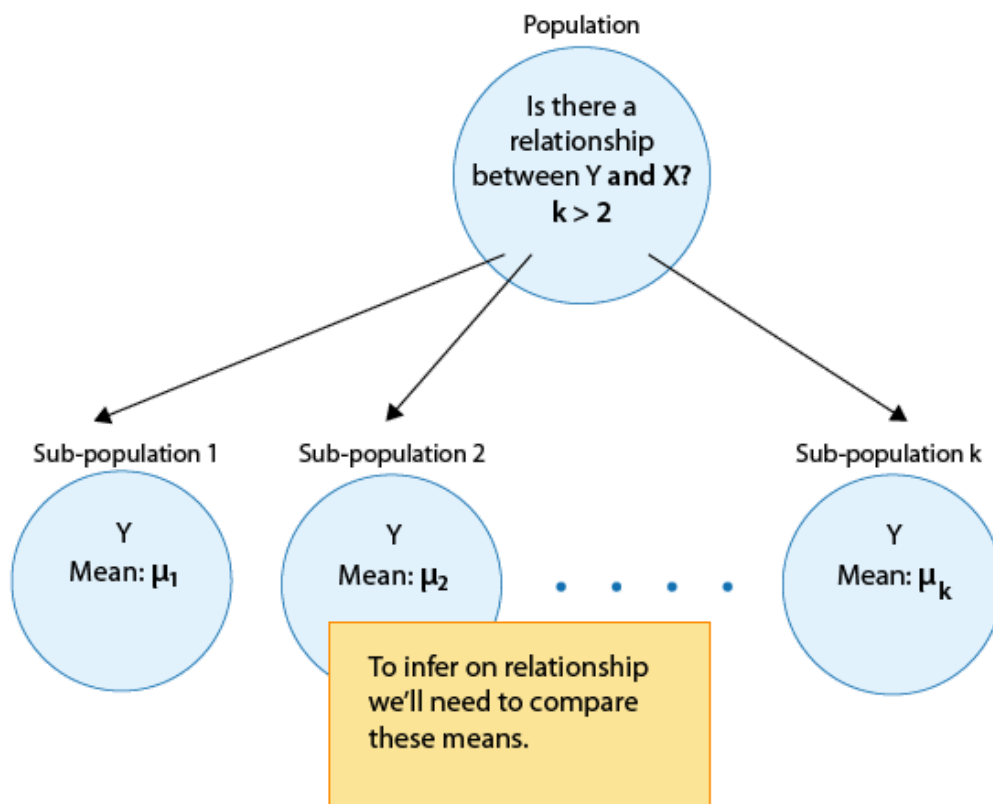
In general, then, making inferences about the relationship between X and Y in Case C→Q boils down to comparing the means of Y in the sub-populations, which are created by the categories defined in X (say k categories). The following figure summarizes this:



As the introduction to this module mentioned, we will learn three inferential methods in Case C→Q, corresponding to a sub-division of this case. First we will distinguish between cases where the explanatory X has only two categories ($k = 2$), and cases where X has more than two categories ($k > 2$). In other words, we will look separately at cases where we are comparing two sub-population means:



and cases where we are comparing more than 2 sub-population means:



For example, if we are interested in whether GPA (Y) is related to gender (X), this is a case where $k = 2$ (since gender has only two categories: M, F), and the inference will boil down to comparing the mean GPA in the sub-population of males to that in the sub-population of females. On the other hand, in the example we looked at earlier, the relationship between GPA (Y) and year (X) is a case where $k > 2$ or more specifically, $k = 4$ (since year has four categories). In terms of inference, these two examples will be treated differently!

Open Learning Initiative [↗](#)



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

© All Rights Reserved