Bookmark this page

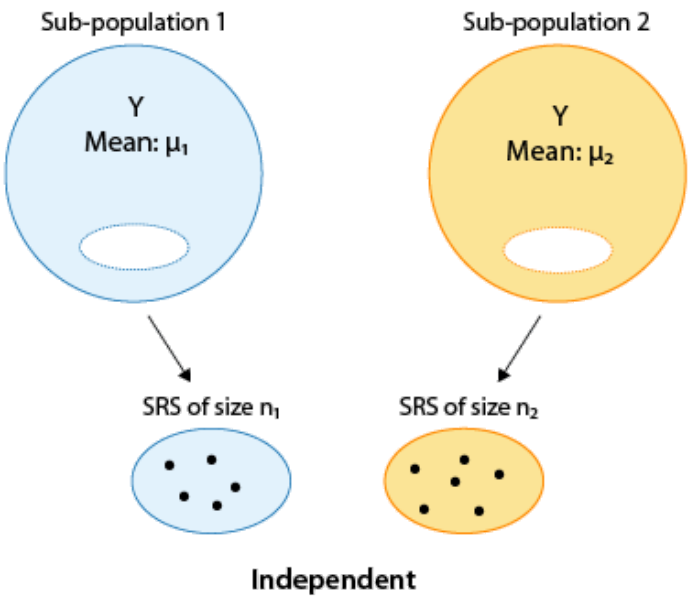
Two Independent Samples: Overview

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

Comparing Two Means—Two Independent Samples (The Two-Sample t-Test)

Overview

As we mentioned in the summary of the introduction to Case C→Q, the first case that we will deal with is comparing two means when the two samples are independent:



Recall that here we are interested in the effect of a two-valued (k = 2) categorical variable (X) on a quantitative response (Y). Samples are drawn independently from the two sub-populations (defined by the two categories of X), and we need to evaluate whether or not the data provide enough evidence for

us to believe that the two sub-population means are different.

In other words, our goal is to test whether the means μ_1 and μ_2 (which are the means of the variable of interest in the two sub-populations) are equal or not, and in order to do that we have two samples, one from each sub-population, which were chosen independently of each other. As the title of this part suggests, the test that we will learn here is commonly known as the **two-sample t-test**. As the name suggests, this is a t-test, which as we know means that the p-values for this test are calculated under some t distribution. Here is how this part is organized.

We first introduce our leading example, and then go in detail through the four steps of the two-sample t-test, illustrating each step using our example.

NOTE...

Up until now, we have been dividing our population into **sub-populations**, then sampling from these sub-populations.

From now on, instead of calling them sub-populations, we will usually call the groups we wish to compare **population 1, population 2,** and so on. These two descriptions of the groups we are comparing can be used interchangeably.

Example

What is more important to you — personality or looks?

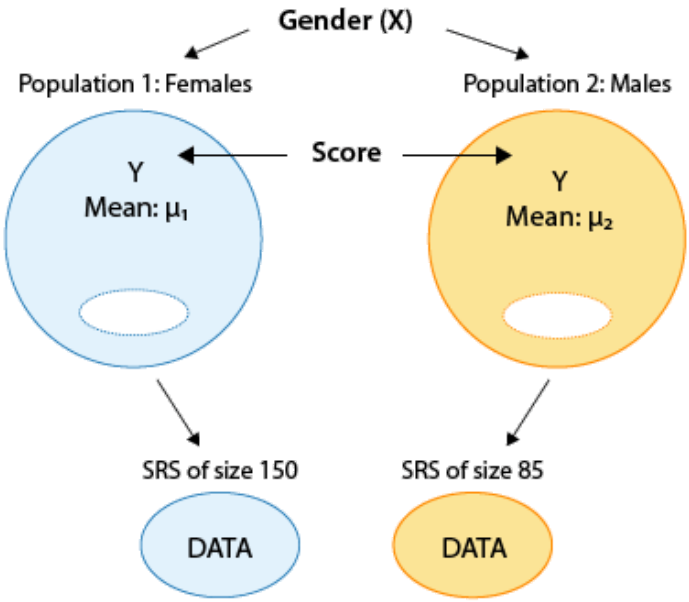
This question was asked of a random sample of 239 college students, who were to answer on a scale of 1 to 25. An answer of 1 means personality has maximum importance and looks no importance at all, whereas an answer of 25 means looks have maximum importance and personality no importance at all. The purpose of this survey was to examine whether males and females differ with respect to the importance of looks vs. personality.

The data have the following format:

Score (Y)	Gender (X)
15	Male
13	Female
10	Female
12	Male
14	Female

14	Male
6	Male
17	Male
etc.	

The format of the data reminds us that we are essentially examining the relationship between the two-valued categorical variable, gender, and the quantitative response, score. The two values of the categorical explanatory variable define the two populations that we are comparing — males and females. The comparison is with respect to the response variable score. Here is a figure that summarizes the example:



Comments:

1. Note that this figure emphasizes how the fact that our explanatory is a two-valued categorical variable means that in practice we are comparing two populations (defined by these two values) with respect to our response Y.

2. Note that even though the problem description just says that we had 239 students, the figure tells us that there were 85 males in the sample, and 150 females.

3. Following up on comment 2, note that 85 + 150 = 235 and not 239. In these data (which are real) there are four "missing observations"—4 students for which we do not have the value of the response variable, "importance." This could be due to a number of reasons, such as recording error or nonresponse. The bottom line is that even though data were collected from 239 students, effectively we have data from only 235. (Recommended: Go through the data file and note that there are 4 cases of missing observations: students 34, 138, 179, and 183).

We will now introduce the two-sample t-test by going through its four steps.

Many Students Wonder ...

Question: Does it matter which population we label as population 1 and which as population 2?

Answer: No, it does not matter as long as you are consistent, meaning that you do not switch labels in the middle. Keeping track of how you label the populations is important in stating the hypotheses and in the interpretation of the results. In our example, it would have been fine to make population 1 be the males, and population 2 be the females. (We'll say more about this in the Learn By Doing activity).