The two-sample t test

Carol Howald

version 0.2 2019-03-14

## Introduction

The two-sample t-test is a staple of introductory statistics. In many courses, it is the most advanced topic taught. In others, it is followed by linear regression and/or ANOVA. At Howard Community College, the t-test is taught in the last 2 weeks of our course.

A t-test is part of the apparatus of *statistical inference*. The purpose of statistical inference is to guide valid conclusions about the *population* from a *sample*. The specific question underlying the t-test is how the mean value of some quantitative variable differs between two groups … in the population. Using samples from each of the two groups, the t apparatus provides a way to calculate two closely related quantities:

1. A *confidence interval* on the difference between the two sample means. The purpose of a confidence interval is to provide a reasonable statement of how the difference in means in the population relates to the difference of means of the two samples.
2. A *p-value* which quantifies how plausible is a claim that in the population the two groups are the same and that the observed difference in sample means comes about by the chance variability stemming from random sampling.

For simplicity, we’ll refer to both of these as the “t-test.”

## Objectives for studying the t-test

Typical course objectives relating to the t-based test and confidence interval are:

* Confidence interval:
  + Compute the confidence interval from data.
  + Interpret the confidence interval in the context of the question intended to be addressed with the data.
* p-value:
  + State an appropriate null hypothesis
  + Draw a sketch relating the sampling distribution under the null to the observed difference between sample means, and marking the region of the distribution corresponding to the p-value. [DO WE WANT THIS TO BE PART OF THE APP? I CAN ADD A TAB SHOWING THIS.] [TO POINT OUT AT WORKSHOP: We can add pedagogical displays as needed.]
  + FOR THE NEW DIAGRAM: Parametric diagram. Show the t-distribution. Variable versus probability density.
  + Compute the numerical p-value
  + Appropriately frame the result as “reject the null” or “fail to reject the null” (the only two valid outcomes of the test).
  + State a valid interpretation of result in the context of the question intended to be addressed with the data.
* Additional objectives:
  + Be able to translate a confidence interval into a simple, approximate statement about the p-value (e.g. or .
  + Identify situations such as outliers that may call into question whether the results can be taken at face value. Know how to deal with such situtations.
* NOTE: One-tailed versus two-tailed tests. ARTICULATION AGREEMENT: State of MD affinity group but no mechanism for organizing them or making sure they are consistent.

## Student background

Students will need some background statistical knowledge to be able to follow lessons on the t-test.

* Basic:
  + Know the difference between a quantitative variable and a categorical variable. For a categorical variable, know the number of *levels* of the variable. Resources: [Little App on jitter plots](https://dtkaplan.shinyapps.io/LA_center_spread/) and class lesson NOT YET PUBLISHED
  + Be comfortable with graphical presentations showing a quantitative variable versus a two-level categorical variable. In this lesson, we use jitter plots. Helpful resources: [Little App on jitter plots](https://dtkaplan.shinyapps.io/LA_center_spread/) and class lesson NOT YET PUBLISHED
  + Understand the distinction between “center” and “spread” of a distribution of values. Resources: [Little App on center and spread](https://dtkaplan.shinyapps.io/LA_center_spread/) and class lesson NOT YET PUBLISHED
  + Understand the process of sampling and the distinction between a population and a sample, and, correspondingly, a “parameter” and a “statistic”.
  + Understand how a descriptive statistic is a summary of a *group* and combines many individual observations.
* Intermediate
  + Be aware central purpose of *statistical inference*, namely to draw valid conclusions about the *population* from a *sample*.
  + Understand that confidence intervals describe the uncertainty in a sample statistic due to sampling variation. Resources: [Little App on resampling](https://dtkaplan.shinyapps.io/LA_bootstrap)
  + Be familiar with the basic nomenclature and logic of hypothesis testing: null-hypothesis, test-statistic, sampling distribution under the null, observed value from the sample, p-value

## Background data knowledge

In this document, we will use data from the National Health and Nutrition Evaluation Survey (NHANES).

* Have students do a web search on NHANES to understand what it is.
* Review some of the variables included in the NHANES data frame. You can do this *within* the Little App by going to the “codebook” tab in the app. Before starting the lessons, make sure the students understand what the “unit of observation” is (an individual person surveyed as part of the NHANES project) and any variables that will be used:
  + HomeOwn - whether the person owns or rents the home they live in. (Possible questions: Across the US, what fraction of households own their home?)
  + Poverty - the *income* of the person’s family, stated as a multiple of the government’s official poverty level. (Possible questions to ask students: What is the current “poverty level?” How poor would a family be that is at the poverty level? At twice the poverty level? For what government services is the poverty level used as a eligibility qualification?)

You might decide to use other data, but the same principles apply of understanding how the data are organized and what the variables stand for.

## (If we had) the Population

The data set we have is large: . Let’s look at the whole data set, imagining that it stands for the actual population.

INDIVIDUAL VS GROUPS

PREDICTION OF AN INDIVIDUAL

PREDICTION OF A GROUP

## Student tasks and activity

### First student task (see student handout, page 1)

Before handing out student task sheets, I would set the stage with the context of a friend and I discussing whether home ownership was a decent predictor of wealth or poverty. I would ask students to share with the class what they thought and why or why not.

After pulling up the Little App: The t-test, to generate some samples, I would provide the handout for students to start recording their work. I would give them a couple minutes to write their own answers to the Think, Pair, Share question on page 1 and then share with a neighbor. I would circulate around the room to interact with individuals and pairs before pulling the class together. Based on responses I see around the room, I would then have a couple pairs share their statements. The think, pair, share task allows me to assess and provide feedback on individuals’ understanding of previous concepts and vocabulary. As students move onto discussing Q1, I can assess if they can use the previous concepts to investigate a new concept, the difference in the mean for 2 independent groups.

### Second student task (see student handout, page 2-3)

After sharing their methods of looking for the difference in the mean poverty level between the two independent groups, I introduce the t interval click box on the Little App. Since a new sample is generated. I ask students to refer back to the example on the sheet. They can then use their estimation technique on an example they can look at up-close and conjecture how it relates to the new interval generated.

With questions 2-4, I hope students can relate the 2 sample t context to the 1 sample t interval and 1 sample t test that would have preceded this lesson.

### Third student task (see student handout, page 4)

This activity allows the students to explore a new set of variables. I will be able to assess if they can apply the ideas from task 1 and 2 to a new context. This gives me the chance to circulate among the groups to provide feedback. After letting each group explore and analyze 15-20 minutes, I will give the groups a few minutes each to present their results. Their goal will be to incorporate the language correctly as they present their results.

NOTE: Students often want to spend too much time just choosing variables. I need to give them a signal when it is time to commit and move on! I also need to assure them that finding out that the variables do not relate in the way they thought is still a valid investigation.

### Fourth classroom task (see student handout, page 5)

At this point, students should conceptually understand ways we can assess the difference in the means for two independent groups. We now need to tie together the relationship between the visualization and the generation of the 2 Independent Sample T-Intervals and Test statistic with p-value and how this is reported with typical statistical software packages where the interval is typically centered around 0.

At this point, the traditional approach used within your curriculum can be linked, hopefully in a condensed manner based on the conceptual understanding developed. I provided a table of data to use within that development so that it would build directly from this lesson. The Little App generates the data list, so our students could capture that as well.

### One-tailed and two-tailed (do we need this?)

## Creating an active classroom

WE WOULD NEED TO Describe some of the ways that you can structure this to be different from a lecture style.

PURPOSE IS TO help people who have very little experience in a classroom setting.

Small groups and class responses

## Assessment Items

* If you have an active classroom, you’ll be able to gain feedback about student accomplishment by witnessing their small-group work and class responses.

ADD THINGS IN AS WE TALK TO WORKSHOP PARTICIPANTS. HAVE A VARIETY TO SUIT Favorite examples and Different teaching styles. DO THIS AS A WORKSHOP ACTIVITY.

* Homework:
* Let students choose another pair of variables and use the Little App to tell a story.
* Assign a traditional problem asking for confidence interval for difference. Most texts have some problems with small data sets. Have students jitter plot by hand and relate the results to the relative position of the means and confidence intervals of the individual variables.
* Exams: An exam question in a format of task one or two would allow students demonstrate their understanding of the graphs, intervals and p-values.

## Pushing the envelope

PROVIDE LINKS TO ESSAYS and other resources.

* Useful approximations
  + Checking whether the 95% confidence intervals on the individual means overlap with each other is a valid equivalent. When the intervals don’t overlap, the p-value will be .
  + For data with, say, and , the t distribution doesn’t add much to the test.
* Pedagogical innovations:
  + The t-test is a special case of “one-way” ANOVA. The equal-variance t-test is mathematically identical to ANOVA. The unequal-variance t-test generally gives a result very similar to ANOVA.
  + The t-test and ANOVA are forms of regression, so it may be more effective to start with regression and then move on to ANOVA and the t-test.
* Streamlining the curriculum:
  + Focus on the confidence interval.
  + Forget about one- and two-tailed tests.
  + For several reasons, there’s never much reason to use an unequal-variance t-test: mathematical complexity, failure to add much power to the test, alternatives (such as rank transforms), philosophical quandries (if you know the variances are unequal, why do you need to look at the means to see if the groups are different).
  + Using regression to set up the t-test

## Author Info

Carol Howald is an Associate Professor of mathematics at Howard Community College. She is also a StatPREP Hub Leader.

Contact info:

* Email: chowald@howardcc.edu
* Location: Howard Community College, Columbia, Maryland, USA