Comparing Two Means

STA 198: Introduction to Health Data Science

Yue Jiang

June 17, 2020

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

Review: hypothesis testing steps

1. State the null and alternative hypotheses. The null hypothesis states "nothing unusual is happening" and the alternative challenges it
2. Collect relevant data and summarize it
3. Assess how surprising it would be to see data like that *if the null hypothesis were really true*
4. Draw conclusions

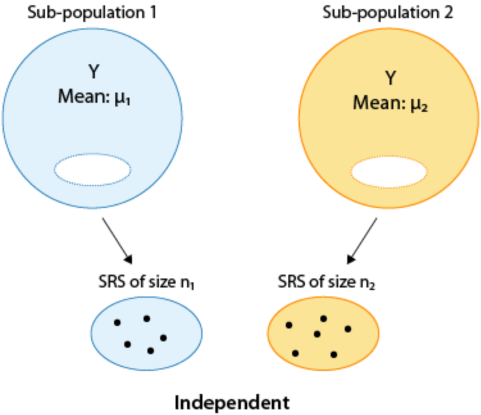
Two-sample t-tests

Last time, we compared one sample to a hypothesized population value.

What if we wanted to compare two samples to each other?

Independent samples

The type of t-test we use to compare two means depends on how the samples were obtained. One approach would be to obtain two independent samples and test the equality of means μ_1 and μ_2



Paired or matched samples

An alternative would be to obtain paired or matched samples and test the equality of means μ_1 and μ_2. Matching could be by person (e.g., before and after measures) or could be a pair of individuals who belong together in another way (e.g., same date of birth in same hospital; husband and wife; etc.)

Paired samples

Samples are often paired for a variety of reasons

- Measurements are taken on a single subject at two distinct points in time (e.g., baseline and follow-up)
- Subjects may be matched so that members of each pair are as much alike as possible with respect to important characteristics like age and gender (e.g., matched case-control study)

Pairing can control for unwanted sources of variation that might otherwise influence the results of a comparison. Matching within subject (e.g., baseline and follow-up) is a powerful way to eliminate subject-specific factors.

Designing a study: impaired driving

The Department of Motor Vehicles wishes to compare impairment of drivers while texting to impairment after being sleep deprived for 24 hours. Describe an independent samples design and a matched pairs design for this question of interest.

Case study: licorice and surgery

Reutzler et al. (2013) performed an experiment among patients having surgery who required intubation. Prior to anesthesia, patients were randomly assigned to gargle either a licorice-based solution or sugar water (as placebo). Sore throat was evaluated 30 minutes, 90 minutes, and 4 hours after conclusion of the surgery on an ordinal pain scale.

On HW 05, you will evaluate whether gargling licorice before surgery led to different mean pain scores at rest.

Case study: licorice and surgery

Let's analyze a related question from their study, which was evaluate whether gargling licorice before surgery led to different mean pain scores when swallowing, at 30 minutes after arrival in the PACU (post-anesthesia care unit).

Pain score was evaluated on an ordinal scale from 0 to 10 (0 = no pain; 10 = worst).

Case study: hypothesis testing step 1

The null hypothesis is that patients receiving licorice gargle and sugar solution placebo have the same mean pain scores relating to swallowing 30 minutes after arrival in the PACU (treatment is unrelated to mean pain), while the alternative is that they do not.

$$H_0 : \mu_L = \mu_S$$

$$H_1 : \mu_L \neq \mu_S$$

Or equivalently,

$$H_0 : \mu_L - \mu_S = 0$$
$$H_1 : \mu_L - \mu_S \neq 0$$

Case study: hypothesis testing step 2

The researcher enrolled 233 subjects, 116 receiving placebo and 117 receiving licorice. Analyzing the data, we obtained $\bar{x}_L = 0.307, \bar{x}_S = 1.379, s_L = 0.825, s_S = 2.287$

Two-sample t-test, independent samples

The two-sample t-test for independent samples is given by

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

The degrees of freedom of the t statistic depend on whether or not $\sigma_1 = \sigma_2$.

Equal or unequal variances?

The choice of *df* depends on whether the independent samples have the same or different variances.

If the variances are equal, then we can use a pooled estimate of s^2, and the degrees of freedom are given by
$(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

If the variances are unequal, the degrees of freedom are difficult to derive, and the Satterthwaite approximation is often used (use software).

Unequal variances should be the default choice, as the t-test assuming equal variances can be quite unreliable if the variances differ, especially when the group sizes differ as well.

Case study: hypothesis testing step 3

Carrying out the two sample t-test for independent samples with unequal variances using software, we get $t = 4.75$ $df \approx 144.21$, with a corresponding p-value < 0.001.

On HW 05, you will learn how to conduct this test.

Case study: hypothesis testing step 4

Based on our observed data, we conclude that there is evidence of a potential difference in mean pain score between the two group. In particular, we have evidence that those receiving the licorice gargle before their surgery reported a lower mean pain score compared to placebo patients.

Case study: Body temperature

You may believe normal body temperature is 98.6^{o}F. A 1992 *JAMA* article examined this assumption, and published temperature data from a large cohort of people. We aim to answer whether the mean body temperature is different for men and women.

Describe the null and alternative hypothesis, explain how you could carry out a specific test of interest and what information you would need, and explain how you would make a conclusion in light of observed data.

The real article showed the sample mean for females was 98.4 and the sample mean for men was 98.1. A t-statistic was calculated (127.51 df), and was found to be 2.285, corresponding to a two-sided p-value of 0.024.

Paired samples

What if you use an unpaired test for paired data?

- You will get the same estimates of the means
- If the pairs have positive correlation (almost always the case if you paired correctly), the unpaired estimate of the variance will be larger than the paired estimate. This means the paired test has greater power.

Paired sample t-test

The paired sample t-test is pretty easy to carry out. All we do is to create a new outcome variable, d, that contains the differences in outcomes between members of a pair. Then we analyze the differences d using the usual one-sample t-test.

Case study: athletic training

Subject	Before	After
1	12.9	12.0
2	13.5	12.2
3	12.8	11.2
4	15.6	13.0
5	17.2	15.0
6	19.2	15.8
7	12.6	12.2
8	15.3	13.4
9	14.4	12.9
10	11.3	11.0

A school athletics department has taken a new instructor and wants to test the effectiveness of the new type of training proposed by comparing the average times of 10 runners in the 100 meters in seconds before and after the new training is implemented.

- What are H_0 and H_1?
- Which test should we use?