STAT 2593

Lecture 035 - The Two Sample t Test and Confidence Interval

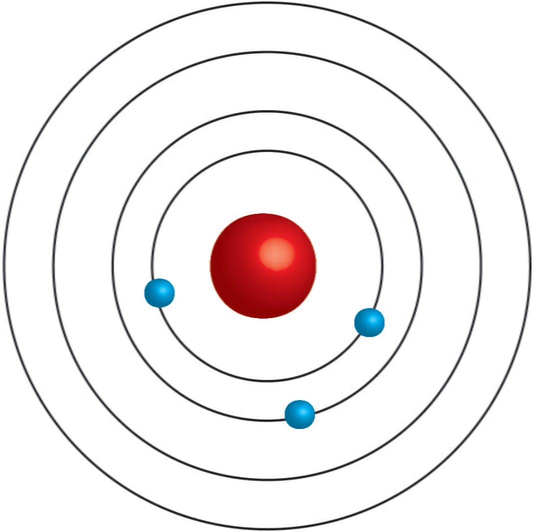
Dylan Spicker

The Two Sample t Test and Confidence Interval

Learning Objectives

1. Construct hypothesis tests and confidence intervals for two sample populations with small sample sizes and unknown variances.

2. Understand the pooled variance estimator under the assumption of equal variances.



Two Populations with Small Sample Sizes

- Previously, we assumed that the variance was known or that the sample sizes were large enough to use normal approximations.

Two Populations with Small Sample Sizes

- Previously, we assumed that the variance was known or that the sample sizes were large enough to use normal approximations.
  - What if we wish to conduct inference regarding two populations, with small sample sizes, and unknown variances.

Two Populations with Small Sample Sizes

- Previously, we assumed that the variance was known or that the sample sizes were large enough to use normal approximations.
  - What if we wish to conduct inference regarding two populations, with small sample sizes, and unknown variances.
  - We form the standard estimator, replacing the variances with the variance estimates,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}.$$

Distribution of the Test Statistic

- This will follow be follow an approximate t_ν distribution with

$$\nu = \frac{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)}{m-1}^2}.$$

Distribution of the Test Statistic

- This will follow be follow an approximate t_ν distribution with

$$\nu = \frac{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)^2}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)^2}{m-1}}.$$

  - In practice, we use $\nu = \min\{n-1, m-1\}$.

Hypothesis Test and Confidence Intervals

- Once the test statistic is formed, we can use critical values from the t distribution as before.

Hypothesis Test and Confidence Intervals

- Once the test statistic is formed, we can use critical values from the t distribution as before.
- Hypothesis tests of the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus a (one- or two-sided) alternative can be tested in the usual way with the corresponding t distribution.

Hypothesis Test and Confidence Intervals

- Once the test statistic is formed, we can use critical values from the t distribution as before.
- Hypothesis tests of the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus a (one- or two-sided) alternative can be tested in the usual way with the corresponding t distribution.
- This applies for any $m, n > 1$.

Hypothesis Test and Confidence Intervals

- Once the test statistic is formed, we can use critical values from the t distribution as before.
- Hypothesis tests of the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus a (one- or two-sided) alternative can be tested in the usual way with the corresponding t distribution.
- This applies for any $m, n > 1$.
  - The procedure we previously saw, using normal approximations, applies only when σ_1^2 and σ_2^2 are known or if m and n are both large enough.

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
  - When this assumption is made, we can use a pooled variance estimator.

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
  - When this assumption is made, we can use a pooled variance estimator.
  - In this case, the mean difference $\overline{X} - \overline{Y}$ will have variance $\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)$.

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
  - When this assumption is made, we can use a pooled variance estimator.
  - In this case, the mean difference $\overline{X} - \overline{Y}$ will have variance $\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)$.
  - An estimator for σ^2 is given by

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}.$$

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
  - When this assumption is made, we can use a pooled variance estimator.
  - In this case, the mean difference $\overline{X} - \overline{Y}$ will have variance $\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)$.
  - An estimator for σ^2 is given by

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}.$$

- The previous estimator can then be scaled by $S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$.

Pooled Variance Estimation

- In some settings it is reasonable to assume a common variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
  - When this assumption is made, we can use a pooled variance estimator.
  - In this case, the mean difference $\overline{X} - \overline{Y}$ will have variance $\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)$.
  - An estimator for σ^2 is given by

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}.$$

- The previous estimator can then be scaled by $S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$.
- This will have a t distribution with $m + n - 2$ degrees of freedom.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.
- Sometimes, you may see advice allowing you to test this from your data.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.
- Sometimes, you may see advice allowing you to test this from your data.
- As a general rule of thumb, you can consider the ratio of S_1^2 to S_2^2.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.
- Sometimes, you may see advice allowing you to test this from your data.
- As a general rule of thumb, you can consider the ratio of S_1^2 to S_2^2.
  - If this ratio falls between 0.5 and 2, a pooled estimator may be appropriate.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.
- Sometimes, you may see advice allowing you to test this from your data.
- As a general rule of thumb, you can consider the ratio of S_1^2 to S_2^2.
  - If this ratio falls between 0.5 and 2, a pooled estimator may be appropriate.
  - Outside of this range is generally not advisable to use this estimator.

Assessing the Pooled Variance Estimator

- If there is a good subject matter reason to believe that variances are equal, it is acceptable to make the assumption.
- Sometimes, you may see advice allowing you to test this from your data.
- As a general rule of thumb, you can consider the ratio of S_1^2 to S_2^2.
  - If this ratio falls between 0.5 and 2, a pooled estimator may be appropriate.
  - Outside of this range is generally not advisable to use this estimator.
- If there is *no good reason* to use a pooled estimator, don't.

Summary

- When testing a difference of means on two populations which have unknown variance and small sample size, we can use an approximate t distribution.

- If the variances between the two populations are known to be equal, then a pooled variance estimator can be used instead.