STAT 2593

Lecture 034 - Z tests and Confidence Intervals for a Difference of Two Population Means

Dylan Spicker

Z tests and Confidence Intervals for a Difference of Two Population Means

Learning Objectives

1. Understand the mean and variance for the estimator of a difference of means, under normality assumptions.

2. Understand the Z hypothesis test for a difference of means, under normality assumptions.

3. Form confidence intervals for the estimated difference between two different means, under normality assumptions.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.
- Suppose that we have two, independent populations, which we wish to compare.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.
- Suppose that we have two, independent populations, which we wish to compare.
  - Can think of two different manufacturing processes or treatments.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.
- Suppose that we have two, independent populations, which we wish to compare.
  - Can think of two different manufacturing processes or treatments.
  - Take $X_1, \ldots, X_n$ to be the first sample, with mean μ_1 and variance σ_1^2.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.
- Suppose that we have two, independent populations, which we wish to compare.
  - Can think of two different manufacturing processes or treatments.
  - Take $X_1, \ldots, X_n$ to be the first sample, with mean μ_1 and variance σ_1^2.
  - Take $Y_1, \ldots, Y_m$ to be the second sample, with mean μ_2 and variance σ_2^2.

Two Populations in Place of One

- Until now we have assumed that there is a single population of interest.
- Suppose that we have two, independent populations, which we wish to compare.
  - Can think of two different manufacturing processes or treatments.
  - Take $X_1, \ldots, X_n$ to be the first sample, with mean μ_1 and variance σ_1^2.
  - Take $Y_1, \ldots, Y_m$ to be the second sample, with mean μ_2 and variance σ_2^2.
  - Suppose that $X \perp Y$, and we wish to compare μ_1 and μ_2.

Estimating a Difference of Means

- We have seen that μ_1 is estimated well by $\overline{X}$, and μ_2 by $\overline{Y}$.

Estimating a Difference of Means

- We have seen that μ_1 is estimated well by $\overline{X}$, and μ_2 by $\overline{Y}$.
  - By extension, $\overline{X} - \overline{Y}$ is unbiased for $\mu_1 - \mu_2$.

Estimating a Difference of Means

- We have seen that μ_1 is estimated well by $\overline{X}$, and μ_2 by $\overline{Y}$.
  - By extension, $\overline{X} - \overline{Y}$ is unbiased for $\mu_1 - \mu_2$.
  - The variance of the estimator is given by

$$\text{var}\left(\overline{X} - \overline{Y}\right) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

Estimating a Difference of Means

- We have seen that μ_1 is estimated well by $\overline{X}$, and μ_2 by $\overline{Y}$.
  - By extension, $\overline{X} - \overline{Y}$ is unbiased for $\mu_1 - \mu_2$.
  - The variance of the estimator is given by

$$\text{var}\left(\overline{X} - \overline{Y}\right) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

- We can consider the typical standardization, forming

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}.$$

Estimating a Difference of Means

- We have seen that μ_1 is estimated well by $\overline{X}$, and μ_2 by $\overline{Y}$.
  - By extension, $\overline{X} - \overline{Y}$ is unbiased for $\mu_1 - \mu_2$.
  - The variance of the estimator is given by

$$\text{var}\left(\overline{X} - \overline{Y}\right) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

- We can consider the typical standardization, forming

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}.$$

  - If the populations are normal, or if the sample sizes are large enough, this will be approximately distributed as $N(0, 1)$.

Hypothesis Test for Mean Differences

- Once we have computed Z, then the hypothesis test is **no different** than in the standard case.

Hypothesis Test for Mean Differences

- Once we have computed Z, then the hypothesis test is **no different** than in the standard case.
- The null hypothesis will take the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus the alternative of $H_1 : \mu_1 - \mu_2 \neq \Delta_0$.

Hypothesis Test for Mean Differences

- Once we have computed Z, then the hypothesis test is **no different** than in the standard case.
- The null hypothesis will take the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus the alternative of $H_1 : \mu_1 - \mu_2 \neq \Delta_0$.
  - The one-sided hypothesis tests are also available.

Hypothesis Test for Mean Differences

- Once we have computed Z, then the hypothesis test is **no different** than in the standard case.

- The null hypothesis will take the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus the alternative of $H_1 : \mu_1 - \mu_2 \neq \Delta_0$.

  - The one-sided hypothesis tests are also available.

  - Under the null hypothesis, replacing $\mu_1 - \mu_2$ with Δ_0 results in a $N(0, 1)$ sampling distribution.

Hypothesis Test for Mean Differences

- Once we have computed Z, then the hypothesis test is **no different** than in the standard case.
- The null hypothesis will take the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus the alternative of $H_1 : \mu_1 - \mu_2 \neq \Delta_0$.
  - The one-sided hypothesis tests are also available.
  - Under the null hypothesis, replacing $\mu_1 - \mu_2$ with Δ_0 results in a $N(0, 1)$ sampling distribution.
  - The hypothesis test runs exactly as expected.

Confidence Intervals for the Mean Difference

- The equivalence of hypothesis tests and confidence intervals gives the fact that confidence intervals proceed equivalently as well.

Confidence Intervals for the Mean Difference

- The equivalence of hypothesis tests and confidence intervals gives the fact that confidence intervals proceed equivalently as well.

- If we want an α level confidence interval for $\mu_1 - \mu_2$, we can take

$$\overline{X} - \overline{Y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.
  - If the populations are non-normal, but the sample sizes are large, and the variances are known, this will be approximate.

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.
  - If the populations are non-normal, but the sample sizes are large, and the variances are known, this will be approximate.
- If σ_1^2 and σ_2^2 are not known, then this procedure will **not** be exact.

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.
  - If the populations are non-normal, but the sample sizes are large, and the variances are known, this will be approximate.
- If σ_1^2 and σ_2^2 are not known, then this procedure will **not** be exact.
- We can replace these with S_1^2 and S_2^2, as before.

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.
  - If the populations are non-normal, but the sample sizes are large, and the variances are known, this will be approximate.
- If σ_1^2 and σ_2^2 are not known, then this procedure will **not** be exact.
- We can replace these with S_1^2 and S_2^2, as before.
  - If the sample sizes are large enough, these results will hold approximately.

Operating Assumptions

- If the populations are normal with known variances, this procedure is *exact*.
  - If the populations are non-normal, but the sample sizes are large, and the variances are known, this will be approximate.
- If σ_1^2 and σ_2^2 are not known, then this procedure will **not** be exact.
- We can replace these with S_1^2 and S_2^2, as before.
  - If the sample sizes are large enough, these results will hold approximately.
  - If not, we will need to rely on alternative distributions to conduct the test.

Summary

- Two populations can be compared under the assumption of normality by estimating the mean difference by the difference of sample means.
- Under normality assumptions, or approximate normality assumptions, the difference will follow a normal distribution and normal procedures apply.