Lesson 034
Difference of Two Population Means

Monday, November 27

A 95% confidence interval for μ is formed as $[-2, 3]$. Suppose that we wish to test $H_0 : \mu = 4$. Which of the following conclusions hold?

0

We reject H_0 at a 5% level of significance.

0%

We fail to reject H_0 at a 5% level of significance.

0%

We reject H_0 at a 10% level of significance.

0%

We fail to reject H_0 at a 10% level of significance.

0%

More than one of the above.

0%

Suppose that the hypothesis test of $H_0 : \mu = 0$ has a compute p-value of 0.03. Which of the following confidence intervals are plausible?

1

A 95% confidence interval of $[-1, 1]$.

0%

A 90% confidence interval of $[-1, 1]$.

0%

A 99% confidence interval of $[-1, 1]$.

0%

More than one of the above.

100%

A 95% confidence interval for μ is formed as $[-2, 3]$. Suppose that we wish to test $H_0 : \mu = -1$. Which of the following conclusions hold?

0

We reject H_0 at a 5% level of significance.

0%

We fail to reject H_0 at a 5% level of significance.

0%

We reject H_0 at a 1% level of significance.

0%

We fail to reject H_0 at a 1% level of significance.

0%

More than one of the above.

0%

Suppose that $H_0 : \mu \geq \mu_0$ is tested against $H_1 : \mu < \mu_0$, where $\hat{\mu} < \mu_0$. The p-value is computed as 0.03. What would be the conclusion of testing $H_0 : \mu = \mu_0$ at $\alpha = 0.05$?

0

We fail to reject H_0.

0%

We reject H_0.

0%

We cannot test this from the given information.

0%

Suppose that the p-value for a (symmetric) test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is 0.04. What would the p-value be for the same procedure of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.

0

- 0.02 — 0%
- 0.04 — 0%
- 0.96 — 0%
- 0.98 — 0%
- We need more information. — 0%

Suppose that the p-value for a (symmetric) test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is 0.04. What would the p-value be for the same procedure of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. Suppose $\widehat{\mu} < \mu_0$.

0

0.02

0%

0.04

0%

0.96

0%

0.98

0%

We need more information.

0%

Suppose that the p-value for a (symmetric) test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is 0.04. What would the p-value be for the same procedure of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. Suppose $\widehat{\mu} > \mu_0$.

0

0.02

0%

0.04

0%

0.96

0%

0.98

0%

We need more information.

0%

Comparing Populations

- We have assumed that there is only a single population of interest.
- What if we have two independent populations, and we wish to compare them?
  - $X_1, \ldots, X_n$ from a population with $E[X_j] = \mu_1$ and $\mathrm{var}(X_j) = \sigma_1^2$.
  - $Y_1, \ldots, Y_m$ from a population with $E[Y_j] = \mu_2$ and $\mathrm{var}(Y_j) = \sigma_2^2$.
  - $X_j \perp Y_\ell$ for all j, ℓ.

Comparing Populations

- We want to compare μ_1 and μ_2.
  - Does one type of component last longer than another?
  - Is one treatment more effective than another?
  - Are the two facilities equivalent in terms of produced yield?
- The same techniques we have already seen will produce confidence intervals and hypothesis tests for the difference.
- We only need to understand the sampling distribution.

Estimating the Difference

- Recall that $\overline{X}$ is an estimator for μ_1 and $\overline{Y}$ for μ_2.
- It seems reasonable to use $\overline{X} - \overline{Y}$ as an estimator for $\mu_1 - \mu_2$.
- This is an unbiased estimator.
- The variance of $\text{var}(\overline{X} - \overline{Y}) = \text{var}(\overline{X}) + \text{var}(\overline{Y}) = \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}$.
- If the populations are normally distributed **or** if both n and m are large, we know that $\overline{X} - \overline{Y}$ is approximately normal.

Test Statistic for the Differences

- As a result, we can consider our standardized test statistic

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/n}} \dot{\sim} N(0,1)$$

- This quantity can then be used to form confidence intervals or test for hypotheses **exactly** as we have seen for the one sample case.

Confidence Intervals for the Differences

- Based on the test statistic we can take an $100(1-\alpha)\,\%$ CI to be given by

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

- The same procedure works for $\bar{Y}-\bar{X}$.

Hypothesis Tests for the Differences

- If we wish to test $H_0 : \mu_1 - \mu_2 = \Delta_0$ versus the alternative that $H_1 : \mu_1 - \mu_2 \neq \Delta_0$ this proceeds as expected.

- Assuming the null hypothesis is true ...

$$T(\widehat{\Delta}, \Delta_0) = \frac{(\overline{X} - \overline{Y}) - \Delta_0}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \dot{\sim} N(0,1)$$

- P-values or critical values are exactly as before.

Assumptions and Alternatives

- If both populations are normal with known variances, this procedure is exact.
- If at least one population is non-normal, with known variances, then this is approximately accurate for large n and m.
- If the variances are not known, can be replaced by S_1^2 and S_2^2, in the above cases, for an approximately correct result.
- Otherwise, different distributions are required.