

STAT 3011 Discussion 015

Week 11: Two-Sample Comparisons

Talha Hamza

University of Minnesota
College of Science and Engineering

Fall 2025

Two-Sample Comparison Framework

Independent Samples	Matched Pairs
Compare μ_1 vs μ_2 $\bar{x}_1 - \bar{x}_2$	Compare μ_D (pair differences) \bar{x}_D (mean of differences)
Example: Test scores from two different classrooms	Example: Twins assigned to different treatments

Key Distinction

- **Independent:** Two completely separate groups with no pairing
- **Paired:** Individuals paired by characteristics (age, weight, etc.)

Note

Matched pairs could be the same subject measured twice

Assumptions

Independent Samples	Matched Pairs
<ol style="list-style-type: none">1. Two independent random samples2. Both populations normal OR $n_1, n_2 \geq 30$	<ol style="list-style-type: none">1. Random sample of pairs2. Differences normally distributed OR large enough sample size

Confidence Intervals for Two-Sample Comparisons

Independent Samples	Matched Pairs
$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\text{df} = \min(n_1 - 1, n_2 - 1)$	$\bar{x}_D \pm t_{\alpha/2} \frac{s_D}{\sqrt{n_D}}$ $\text{df} = n_D - 1$

Intuitive Interpretation

- **CI excludes 0:** The true difference is unlikely to be zero ("statistically significant")
 - Example: If 95% CI (1.2, 3.4) kg for weight loss \implies Effective treatment
- **CI includes 0:** No evidence of difference ("null plausible")
 - Example: If 95% CI (-0.5, 1.5) kg \implies Might just be random variation

Hypothesis Testing for Two Samples

Independent Samples

Hypotheses:

- $H_0 : \mu_1 = \mu_2$
- $H_a : \mu_1 \neq \mu_2$ (two-tailed)
- $H_a : \mu_1 < \mu_2$ (left-tailed)
- $H_a : \mu_1 > \mu_2$ (right-tailed)

Test Statistic:

$$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Matched Pairs

Hypotheses:

- $H_0 : \mu_D = 0$
- $H_a : \mu_D \neq 0$ (two-tailed)
- $H_a : \mu_D < 0$ (left-tailed)
- $H_a : \mu_D > 0$ (right-tailed)

Test Statistic:

$$t^* = \frac{\bar{x}_D}{s_D / \sqrt{n_D}}$$

p-values and Conclusions

Calculating p-values

- Left-tailed: `pt(t*, df, lower.tail=TRUE)`
- Right-tailed: `pt(t*, df, lower.tail=FALSE)`
- Two-tailed: `2*pt(abs(t*), df, lower.tail=FALSE)`

Conclusion

- if $\text{p-value} \leq \alpha \implies$ Reject H_0 and Accept H_a
- if $\text{p-value} > \alpha \implies$ Fail to reject H_0

Remember

- We never **accept** H_0 - we only fail to reject it

R Commands for Two-Sample Tests

Basic Syntax

- **Independent:** `t.test(group1, group2)`
- **Paired:** `t.test(group1, group2, paired = TRUE)`

Customize Using

- `alternative = "two.sided"` \Rightarrow test if means differ (\neq)
- `alternative = "less"` \Rightarrow test if $\mu_1 < \mu_2$
- `alternative = "greater"` \Rightarrow test if $\mu_1 > \mu_2$
- **For CI only:** Add `conf.level = #` AND `alternative = "two.sided"`

Questions?