BOSTON UNIVERSITY

PS 211: Introduction to Experimental Design

Fall 2025 · Section C1

Lecture 13: Reporting results in APA style, One-way ANOVA

Updates & Reminders

- Homework 3 is due on **Friday, October 31** at 11:59 PM.
  - Submit via Blackboard as a PDF file.
  - Late submissions will incur a penalty of 3% per day.
  - We are going to try to post grades and answers by Monday night.
- Exam 3 will be on **Thursday, Nov. 6th** during our regular class time.
  - It will focus on material from Lectures 10-13 (today).
  - It will build on material from earlier in the course, so be sure to review those topics as well.
  - The format will be the same as Exam 1 (33 multiple choice questions).
  - A review sheet will be posted by the end of the day tomorrow.

Review: Confidence Intervals

- Remember, a *z* or *t* statistic corresponds to a percentile in the sampling distribution.
- For example, if we have $t = 2.45$ with df = 28, we can look up the corresponding p-value: $p = .022$.
- This means that if the null hypothesis were true, we would expect to see a *t* value this extreme (or more extreme) only 2.2% of the time.
- Confidence intervals rely on this logic **in reverse**.
- Instead of asking "Given the null hypothesis, what is the probability of observing this data?",
- We ask "Given the observed data, what range of population values are plausible?"

Review: Confidence Intervals (Continued)

- Instead of starting with the *null* distribution, we start with the *sampling* distribution.
- We assume our sample mean was drawn from a population with some unknown mean μ.
- Our best estimate of μ is the sample mean M.
- But because of sampling variability, we know that M might not equal μ exactly.
- So we construct a range of values around M that are plausible values for μ.
- This range is called a **confidence interval (CI)**.

Review: Confidence Intervals (Continued)

- To find a **95% confidence interval**, we identify the range of values that would fall within the middle 95% of the sampling distribution.
- This corresponds to the values between the 2.5th and 97.5th percentiles.
- We can use the critical t value for our sample size to find these cutoffs.
- Our critical t value tells us how far from M we need to go to capture 95% of the sampling distribution.

Review: Confidence Intervals (Continued)

- Remember, our t value is in units of standard errors (SE).
- A critical t of 2.57 means that 95% of the distribution falls within 2.57 SEs of the mean.

Review: Confidence Intervals (Continued)

- Remember, our *t* value is in units of standard errors (SE).
- A critical *t* of 2.57 means that 95% of the distribution falls within 2.57 SEs of the mean.
- We can use this to calculate the CI around our sample mean:

$$\text{CI} = M \pm (t_{critical} \times SE)$$

Where:

- *M* = sample mean
- $t_{critical}$ = critical *t* value for desired confidence level
- *SE* = standard error of the mean

Review: Confidence Intervals (Continued)

One more time…

- Our best estimate of μ is the sample mean *M*.
- But because of sampling variability, we know that *M* might not equal μ exactly.
- So we construct a range of values around *M* that are plausible values for μ.
- We do this by determining where 95% (or 99% or 99.9%, etc.) of the sampling distribution lies.
- We compute *this* by finding the critical *t* value and multiplying it by the standard error.
- This range is called a **confidence interval (CI)**.
- Lower standard error -> narrower CI.
- Higher standard error -> wider CI.

Review: Confidence Intervals (Continued)

Practice question:

How does increasing the sample size affect the width of a confidence interval?

A. Increases the width because there is more data.

B. Increases the width because the critical t-value gets larger.

C. Decreases the width because the standard error is smaller and the estimate is more precise.

D. No effect on the width because n does not factor into CI calculations.

Answer: C. Decreases the width because the standard error is smaller and the estimate is more precise.

Review: Confidence Intervals (Continued)

Practice question:

How does increasing the estimate of the sample mean affect the width of a confidence interval?

A. Increases the width because the values are higher.

B. Increases the width because the critical t-value gets larger.

C. Decreases the width because the standard error will also increase.

D. No effect on the width because the sample mean does not factor into CI calculations.

Answer: D. No effect on the width because the sample mean does not factor into CI calculations.

Recap: Worksheet example

- Let’s go through the worksheet together.
- I will attempt to fill it out in real time.
- Please correct me if I make a mistake!

Reporting Results in APA Style

APA style tells us **how to clearly report statistics** so others can understand and replicate our work.
It ensures clarity and consistency across psychology and related sciences.

What is APA Style?

- **APA** = *American Psychological Association*
- A standardized format for writing and reporting scientific results
- Ensures clarity, transparency, and consistency across research papers

APA style guides how we:
- write numbers and statistics
- report results of tests (t, z, F, etc.)
- format p-values, CIs, and effect sizes

Why Use APA Style?

- Helps readers quickly interpret results.
- Makes research **replicable** and **comparable**.
- Prevents ambiguity — everyone reports in the same format.
- Required by most psychology journals and conferences.

Think of APA as the “grammar rules” of scientific writing.

Always Include These Elements

Element	Description	Example
Test statistic	Which test you ran (t, z, F)	t
Degrees of freedom (df)	Indicates sample size info	$t(28)$
Test value	The computed statistic	$t(28) = 2.45$
p value	Probability under H_0	$p = .004$ or $p < .001$
Descriptive stats	Means, SDs for each group	$M = 5.3$, $SD = 1.2$

✅ These should appear in **every** statistical report.

Technically optional (but strongly recommended)

Element	Why Include It	Example
Effect size	Shows *magnitude* of difference	$d = 0.68$
Confidence interval (CI)	Range of plausible values	95% CI [0.12, 1.24]
Exact p value	More informative than $p < .05$	$p = .037$

- Including these helps readers understand *how strong* and *how precise* your findings are.
- Many journals now require effect sizes, CIs, and exact p-values.
- There is no reason *not* to include them!

Reporting a *z* Test

Template

$$M = X,\ SD = Y,\ z = Z,\ p = P,\ d = D,\ 95\%\ \text{CI}\ [\text{LL, UL}]$$

Example

Students' average rent (*M* = \$920, *SD* = \$192) was significantly higher than the population mean of \$500, *z* = 9.39, *p* < .001, *d* = 4.20, 95% CI [\$832, \$1008].

✅ No *degrees of freedom* are reported for *z* tests. *Z* tests are based on known population parameters, not sample estimates, so df are not applicable.

Reporting a One-Sample *t* Test

Template

$$t(df) = X.XX,\ p = P,\ d = D,\ 95\%\ \mathrm{CI}\ [\mathrm{LL},\ \mathrm{UL}]$$

Example

The sample's mean reaction time (M = 312 ms, SD = 28) was significantly faster than the population mean of 350 ms, $t(14) = -5.68$, $p < .001$, $d = 1.47$, 95% CI [–50.1, –21.9].

✅ State the **known or hypothesized** population mean and the sample mean and SD.

Reporting a Paired-Samples *t* Test

Template

$$t(df) = X.XX,\ p = P,\ d = D,\ 95\%\ \mathrm{CI}\ [\mathrm{LL},\ \mathrm{UL}]$$

Example

Participants recalled more words after caffeine (M = 12.3, SD = 2.1) than without caffeine (M = 9.8, SD = 2.5), $t(19) = 3.42$, $p = .003$, $d = 0.76$, 95% CI [0.92, 4.11].

✅ State the means and SDs for **both** conditions.

Reporting an Independent-Samples *t* Test

Template

$$t(df) = X.XX,\ p = P,\ d = D,\ 95\%\ \text{CI}\ [\text{LL, UL}]$$

Example

Participants in the study-group condition had higher quiz scores (M = 84.6, SD = 4.2) than participants in the individual condition (M = 78.3, SD = 5.0), $t(38) = 4.11$, $p < .001$, $d = 1.30$, 95% CI [3.21, 9.59].

✅ State the means and SDs for **both** groups.

Practice question: How many participants were in this study?

A. 38

B. 40

C. 38 per group (76 total)

D. 36

Formatting Rules (APA 7th Edition)

- Italicize *t*, *z*, *p*, *M*, *SD*, *F*, *r*
- Use **two decimal places** for statistics (except *p*)
- Report *p* values as:
  - $p = .042$ (no leading zero)
  - $p < .001$ (for very small values)
- Match decimal precision across CIs and descriptive stats

Putting It All Together

Always Include

- Test type and statistic (t, z, etc.)
- df (for t tests)
- p value
- Descriptive stats (M, SD)

Usually Include

- Effect size (d, r, etc.)
- Confidence interval
- Directional phrasing (“significantly greater than…”)

APA Reporting: Practice 🧠

Rewrite each in APA style:

1. The Caffeine group got an average of 14.2 right answers (standard deviation = 3), and the Placebo group got 11.1 right answers (SD 2.8). The difference was statistically significant with t = 3.2 and p=0.004, df = 28. The Effect size was D= 0.79 and 95 percent CI = (0.50 TO 4.80).

2. Participants slept on average 7.6 Hours (s.d.=0.8) which was not significantly different from the national Mean of 7.5. The T test was not significant (P = .66; t = .45; DF=19). Cohen's D = 0.06 and the 95% confidence interval was (-0.42 , 0.62).

Practice: Answers

Below are the **APA-corrected versions** of those two messy examples.

☕ Correct Example 1 — Independent-Samples *t*

> The caffeine group (M = 14.2, SD = 3.0) scored higher than the placebo group (M = 11.1, SD = 2.8), $t(28)$ = 3.20, p = .004, d = 0.79, 95% CI [0.5, 4.8].

What was fixed?

- Italicize statistical symbols (*t*, *p*, *d*, *M*, *SD*).
- Report *t*(df) = value, *p*, *d*, and 95% CI in this order.
- Use brackets for the CI and **no leading zeros** for *p* values less than 1.

Practice: Answers (Continued)

😴 Correct Example 2 — One-Sample *t*

> Participants slept an average of 7.6 hours (*SD* = 0.8), which did not differ significantly from the national mean of 7.5, *t*(19) = 0.45, *p* = .66, *d* = 0.06, 95% CI [–0.42, 0.62].

What was fixed?

- Write numbers and units clearly (“7.6 hours”).
- Keep *p* = .66 (not *P* = 0.66).
- Report CI in brackets.

Moving to our next topic...

One-Way ANOVA

Are We Worse Drivers When on the Phone?

A simple *t*-test compares driving ability when talking vs. not talking on the phone.

But what if we wanted to compare **more than two conditions**?

💭 Possible conditions:
– Driving alone
– With a passenger
– On a cell phone
– On a video call

Can I Use a *t* Test to Compare >2 Groups?

If you used a *t*-test for every possible combination, you'd run many tests!

- Driving alone vs. passenger
- Driving alone vs. video call
- Driving alone vs. cell phone
- Passenger vs. video call
- Passenger vs. cell phone
- Video call vs. cell phone

That's **6 *t*-tests!**

Is there any problem with that? 🤔
Yes — it **increases the chance of a Type I error** (false positive).

Why You Can't Do Many *t* Tests

- Each test carries a 5% chance (α = .05) of a **Type I error** — falsely rejecting the null.
  - There is a 5% chance of concluding there is a difference when there really isn't one.
  - There is a 95% chance of correctly *failing* to reject the null.
- The more *t*-tests you do, the higher your overall risk of error.
- This problem is called **alpha inflation**.

With one test → 95% chance of correctly retaining the null & 5% chance of a false positive

With two tests -> 95% chance of correctly retaining the null **twice** → $(0.95)^2 = 0.903$ → 10% chance of error

With three tests → 95% chance of correctly retaining the null **three** times → $(0.95)^3 = 0.857$ → 14% chance of error

Why You Can't Do Many *t* Tests (continued)

So if you did 6 *t*-tests (like in our example), you'd have **a 54% chance of a Type I error!**

> **Conclusion:** Multiple t-tests inflate the error rate. We need a single test for 3+ groups.

The Solution: Using the *F* Statistic

When we want to compare **3 or more means**, we use the **F distribution** — the foundation of ANOVA.

Like *z* and *t* tests, the *F* statistic is a **ratio**:

$$z = \frac{\text{Difference Between Means}}{\text{Standard Error}}$$

$$t = \frac{\text{Difference Between Means}}{\text{Standard Error}}$$

$$F = \frac{\text{Between-Groups Variance}}{\text{Within-Groups Variance}}$$

That's all for today!

Tuesday: Exam 3 Review Session featuring lots of practice questions