BOSTON UNIVERSITY



# PS 211: Introduction to Experimental Design

## Fall 2025 · Section C1

## Lecture 15: One-Way Repeated-Measures ANOVA & Two-Way ANOVA

# Updates & Reminders

- Data Write-Up due Monday (11/17) by 11:59pm.
- Homework 4 will be posted early next week (due Dec. 2).
  - Only your top 3 out of 4 homework grades will count toward your final course grade.
- Exam 3 went pretty well!
  - Median: 25/30
  - Everyone earned 2 bonus percentage points, which means the median is > 85%.
  - A few concepts are still challenging for people.

> Remember, you can use the grade calculator (pinned on Slack) to see how future assignments will affect your final grade!

# Part 1: One-Way Repeated-Measures ANOVA (Within-Groups)

# From Between-Groups to Repeated-Measures ANOVA

- Recall: The **one-way between-groups ANOVA** compares means from **independent samples**.
- What if the **same participants** complete **multiple conditions**?
- We use a **repeated-measures ANOVA**, also called a **within-groups ANOVA.**
- Just like with *t*-tests, we have a between-groups and a within-groups version of ANOVA.

# When to Use a Repeated-Measures ANOVA

Use when:

- The **same participants** are measured across **three or more conditions.**
- There is **one independent variable (IV)** with 3+ levels.
- The dependent variable (DV) is **numeric** (interval or ratio).

In which of these scenarios would a Repeated-Measures ANOVA be appropriate?

1. Testing memory performance after 0, 1, and 2 cups of coffee with the same participants.
2. Comparing test scores of students from three different schools.
3. Measuring reaction times of the same participants under three different lighting conditions.
4. Evaluating how number of hours of sleep per night relates to GPA.

# When to Use a Repeated-Measures ANOVA

Use when:

- The **same participants** are measured across **three or more conditions.**
- There is **one independent variable (IV)** with 3+ levels.
- The dependent variable (DV) is **numeric** (interval or ratio).

In which of these scenarios would a repeated-measures ANOVA be appropriate?

1. Testing memory performance after 0, 1, and 2 cups of coffee with the same participants. ✅
2. Comparing test scores of students from three different schools. ❌
3. Measuring reaction times of the same participants under three different lighting conditions. ✅
4. Evaluating how number of hours of sleep per night relates to GPA. ❌

# Advantages of Repeated-Measures ANOVA

- **Accounts for individual differences** – the same people are in every condition.
- **Increases statistical power** – less error variance.
- **Fewer participants** needed.

Essentially, every participant acts as their **own control group.**

**This should sound familiar from within-groups t-tests!**

# Disadvantages of Within-Groups Designs

- **Carryover** effects:
  - Once participants have experienced one condition, it may influence their behavior in subsequent conditions.
- **Order effects**:
  - Practice effects: Improvement due to familiarity with the task.
  - Fatigue effects: Decline in performance due to tiredness or boredom.
  - Counterbalancing can help mitigate this issue, but it may not eliminate it entirely.

# Within- vs. Between-Groups Designs: Practice

*You run an experiment where participants memorize a list of words after drinking 0, 1, or 2 cups of coffee. Participants complete all three conditions across three separate days. All participants show improvement over time, regardless of coffee amount.*

Which disadvantage of within-groups designs does this illustrate?

A. This shows how individual differences can impact performance across conditions.

B. This shows how practice effects can impact performance.

C. This shows how fatigue effects can impact performance.

D. This shows that within-groups designs are typically underpowered.

**Answer: B** – Participants improved over time due to practice, regardless of the coffee condition.

# Within- vs. Between-Groups Designs: Practice

*You run an experiment where participants memorize a list of words after drinking 0, 1, or 2 cups of coffee. Participants complete all three conditions across three separate days. All participants show improvement over time, regardless of coffee amount.*

What is one way you could mitigate the influence of practice effects in your experimental design?

A. Increase your sample size to reduce the influence of practice effects.

B. Counterbalance the order of conditions across participants to ensure that the influence of practice equally affects all conditions.

C. Ask participants to promise not to practice between sessions.

D. All of the above.

**Answer: B** – Counterbalancing helps distribute practice effects evenly across conditions, reducing their confounding influence.

# Between- vs. Repeated-Measures ANOVA

- In both cases, we are still using the *F* ratio: Between-Groups Variance / Within-Groups Variance.
  - We still want to know whether the variability between condition means is larger than the variability within conditions.
- The key difference: **how we calculate the Within-Groups Variance.**

# Calculations in a Repeated-Measures ANOVA

- Remember, for between-groups ANOVA, we partition total variability into:

$$SS_{Total} = SS_{Between} + SS_{Within}$$

- **$SS_{Between}$:** variability due to condition means.

- **$SS_{Within}$:** variability due to random noise.

Importantly, in a between-groups design, this random noise includes all sources of variability within conditions, including individual differences between participants.

- The logic is similar to between-groups ANOVA, **but we add one more term:**

$$SS_{Total} = SS_{Between} + SS_{Within} + SS_{Subjects}$$

- **$SS_{Subjects}$:** variability due to stable individual differences between participants.

# Calculating the Subjects Sum of Squares

To calculate **$SS_{Subjects}$**:

1. Compute each participant's overall mean across all conditions ($M_{participant}$).
2. For each participant, calculate the squared difference between their overall mean and the grand mean ($GM$).
3. Sum these squared differences across all $n$ participants.

> Remember, the **grand mean** is the mean of all scores across all participants and conditions.

In equation form:

$$SS_{Subjects} = \sum_{1}^{n} (M_{participant} - GM)^2$$

# Hypothesis Testing Steps (Repeated-Measures ANOVA)

- **Step 1: Identify populations, distribution, and assumptions.**
  - Numeric DV
  - Normally distributed
  - Homogeneity of variance
- **Step 2: Determine null and alternative hypotheses.**
  - $H_0$: $\mu_1 = \mu_2 = \mu_3$
  - $H_1$: At least one $\mu$ differs

# Hypothesis Testing Steps (Repeated-Measures ANOVA)

- **Step 3: Determine comparison distribution.**
  - $F$ distribution with $\text{df}_{\text{Between}}$ and $\text{df}_{\text{Within}}$.

# Degrees of Freedom (df) for Repeated-Measures ANOVA

- $df_{Between}$ = $k - 1$, *where k = number of conditions*
  - when calculating variability between conditions, we lose one degree of freedom from our calculation of the overall mean.
- $df_{Subjects}$ = $n - 1$, *where n = number of subjects*
  - when calculating variability due to individual differences, we lose one degree of freedom from our calculation of the overall mean.
- $df_{Within}$ = $(k-1)*(n-1)$
  - when calculating variability within conditions, we lose (k-1) degrees of freedom for conditions and (n-1) degrees of freedom for subjects.
- Total df = $N - 1$, *where N = (n*k), total number of observations (scores)*

# Hypothesis Testing Steps (Repeated-Measures ANOVA)

- **Step 4: Find critical value** using $df_{Between}$ and $df_{Within}$
  - From *F* table or software

Why do we often have *two* critical values for a *t* test but only one critical value for an *F* test?

**Answer:** Because *t* tests can be one-tailed or two-tailed, while *F* tests are always one-tailed, because we are always testing whether the variance between groups is greater than the variance within groups.

# Hypothesis Testing Steps (Repeated-Measures ANOVA)

- **Step 5: Compute F**:

$$F = \frac{MS_{Between}}{MS_{Within}}$$

Remember, for a repeated-measures ANOVA, we calculate $MS_{\text{Within}}$ after removing $SS_{\text{Subjects}}$:

$$MS_{Within} = \frac{SS_{Within}}{df_{Within}}$$

$$SS_{Within} = SS_{Total} - SS_{Between} - SS_{Subjects}$$

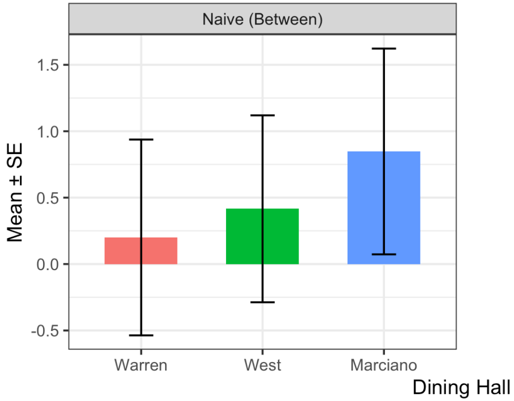
# Hypothesis Testing Steps (Repeated-Measures ANOVA)

- **Step 6: Make a decision.**
  - If *F* calculated ≥ *F* critical, reject $H_0$.
  - If *F* calculated < *F* critical, fail to reject $H_0$.

# Accounting for individual variance increases power

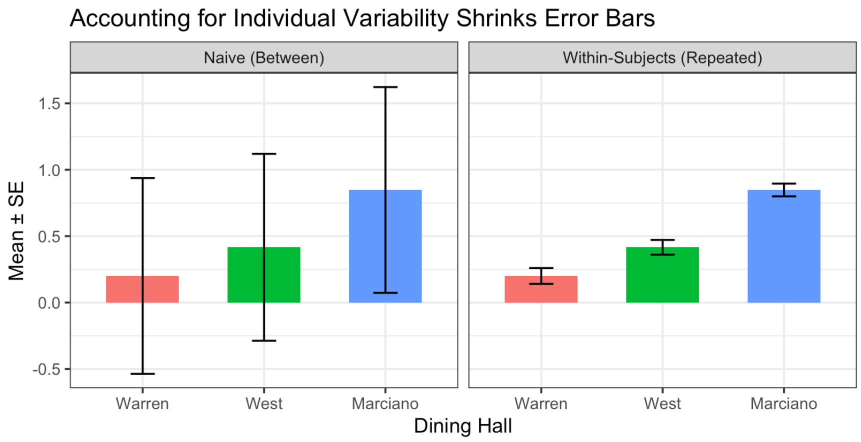
## Example:

- Imagine BU asks students to rate their satisfaction (from 0 - 1) with three different dining halls.
- They ask 10 different students to rate three dining halls.
- They plot the means responses and include error bars that show +/- 1 SE from each mean:



# Accounting for individual variance increases power

- However, some of this variability is due to individual differences among students.
- To account for this, we can subtract each student's overall mean rating from their dining hall ratings. This gives us each student's rating of each dining hall after accounting for individual differences.
- We can then re-compute the standard error of each mean rating and re-plot.



# Effect Size for Repeated-Measures ANOVA

Effect size is measured using **$R^2$ (or $\eta^2$)**:

$$\eta^2 = \frac{SS_{Between}}{SS_{Total} - SS_{Subjects}}$$

This shows the **proportion of variance explained** by the IV after removing variance due to individual differences.

Example interpretation: $\eta^2 = .40$ → 40% of variance in performance is explained by the condition.

# Repeated-Measures ANOVA in R

Use the `aov()` function with a within-subjects design:

```
# Example dataset
data <- data.frame(
  subject = rep(1:10, each = 3),
  condition = rep(c("A", "B", "C"), times = 10),
  score = c(5, 6, 7, 4, 5, 6, 6, 7, 8, 5, 6, 7, 7, 8, 9,
            6, 7, 8, 5, 6, 7, 8, 9, 10, 7, 8, 9)
)

# Run the repeated-measures ANOVA
anova_result <- aov(score ~ condition + Error(subject/condition), data = data)
summary(anova_result)
```

- The `Error(subject/condition)` term specifies the within-subjects design.
- This indicates that each `subject` experiences all levels of `condition`.

# Next Steps After a Significant Repeated-Measures ANOVA

- A significant ANOVA indicates that at least one condition mean differs, but it doesn’t tell us which ones.
- Follow up with post-hoc tests (e.g., pairwise t-tests, Tukey’s HSD) to explore specific group differences.
- Can use corrections (e.g., Bonferroni) to control for Type I error.

# Recap: Repeated-Measures ANOVA

- A repeated-measures ANOVA is used when the **same participants** complete **3+ conditions.**
- Remember, the *F* statistic compares variability between condition means to variability within conditions.
- When the *same participants* complete all conditions, some of that within-condition variability can be attributed to individual differences.
    - Mathematically, accounts for **individual differences** by computing $SS_{\text{Subjects}}$ and subtracting that from $SS_{\text{Within}}$.
- Increases **statistical power**: $SS_{\text{Within}}$ reflects only *random* error.

# Recap: Statistical Tests We've Covered So Far

| Test | # of IVs | IV Type | # of Levels | DV Type | Use Case |
|---|---|---|---|---|---|
| *z-test* | 0 | — | — | Numeric | Compare sample mean to population mean (known SD) |
| *One-Sample t-test* | 0 | — | — | Numeric | Compare sample mean to population mean (unknown SD) |
| *Independent t-test* | 1 | Categorical | 2 | Numeric | Compare two groups (e.g., School A vs. School B) |
| *Paired t-test* | 1 | Categorical | 2 | Numeric | Compare same group before vs. after intervention or in two conditions |
| *One-Way (Between-Groups) ANOVA* | 1 | Categorical | 3+ | Numeric | Compare 3+ groups (e.g., Drug A, B, C) |
| *Repeated-Measures (Within-Subjects) ANOVA* | 1 | Categorical | 3+ | Numeric | Compare same participants across 3+ conditions |

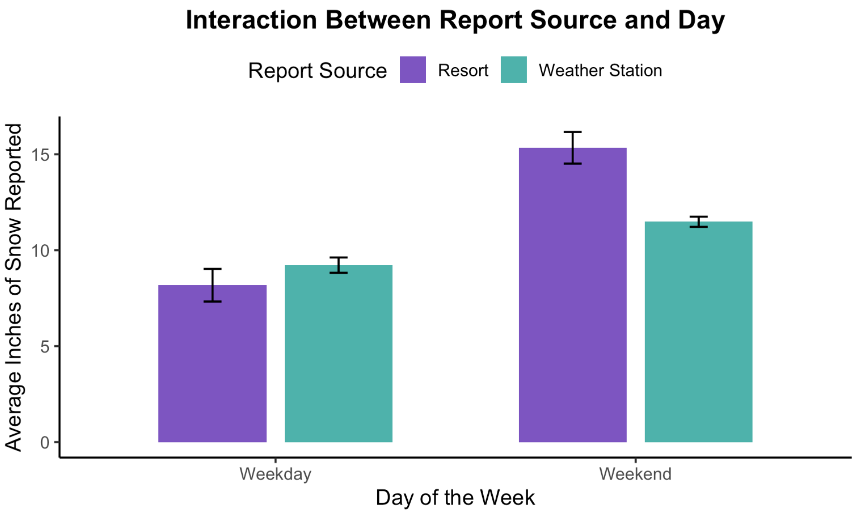
# Part 2: Two-Way ANOVA (Factorial Designs)

What if we have **two** independent variables?

# Why Use a Two-Way ANOVA?

A **Two-Way ANOVA** allows us to test **two independent variables simultaneously.**

Example: ⛷️ Do ski resorts exaggerate weekend snow reports?



# Why Use a Two-Way ANOVA?

# Structure of a Two-Way ANOVA

- A two-way ANOVA involves **two independent variables (IVs).**
- These IVs can each have **two or more levels.**
- In a **2 × 2 design**, there are two independent variables, each with two levels.
- Each unique combination = a **cell** in the design.

| Source | Weekday | Weekend |
|---|---|---|
| Resort Report | $Mean_1$ | $Mean_2$ |
| Weather Station | $Mean_3$ | $Mean_4$ |

- Each cell's mean reflects one condition combination (e.g., Resort–Weekend).

What would a 2 x 3 design look like?

# What Does a Two-Way ANOVA Test?

A Two-Way ANOVA produces **three *F* statistics:**

1️⃣ $F_1$: Main effect of $IV_1$
2️⃣ $F_2$: Main effect of $IV_2$
3️⃣ $F_3$: Interaction between $IV_1 \times IV_2$

- Each *F* statistic tests a specific hypothesis about the effects of the independent variables on the dependent variable.
- Each *F* statistic is computed similarly to a one-way ANOVA:

$$F = \frac{MS_{Between}}{MS_{Within}}$$

- This is because we are still computing the ratio of variability between condition means (e.g., 'signal') to variability within conditions (e.g., 'noise').

# Computing *F* statistics for Two-Way ANOVAs

- We compute separate $SS_{Between}$, $MS_{Between}$, $df$, and critical values for all of our effects of interest:

  1️⃣ Main effect of $IV_1$ (e.g., Report Source)

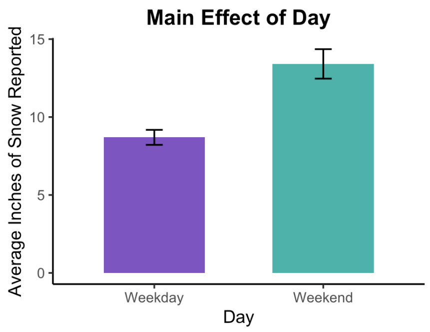
  2️⃣ Main effect of $IV_2$ (e.g., Day)

  3️⃣ Interaction between $IV_1 \times IV_2$ (e.g., Report Source × Day)

- However, they all share the same $SS_{Within}$ and $MS_{Within}$, which reflect variability within "cells."
  - This is because the within-groups variability is the same regardless of which effect we are testing.
  - Remember, when calculating $SS_{Within}$, we are looking at variability within each cell (condition combination), summed across all cells.

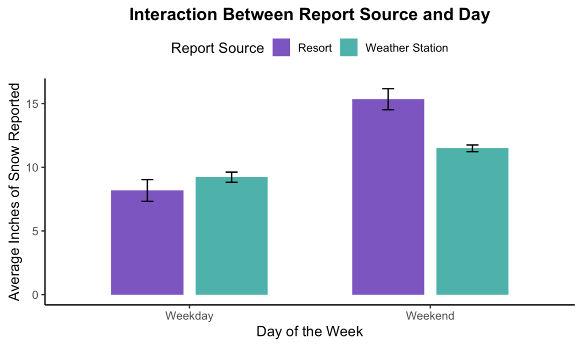
# Interpreting Main Effects vs. Interactions

- **Main Effect:** One IV influences the DV regardless of the other.



# Interpreting Main Effects vs. Interactions

- **Interaction:** The effect of one IV changes depending on the other IV.




- For example: 🎿 Ski resort exaggeration might only happen on **weekends**, not weekdays.
- This would be an example of an **interaction**.
- The effect of Source (Resort vs. Weather Station) depends on Day (Weekday vs. Weekend).

# Marginal Means in Two-Way ANOVA

- **Marginal means** are the means for each level of one IV, averaged across levels of the other IV.
- For example, in our ski resort example:
  - Marginal mean for Resort Report = Average of (Resort–Weekday + Resort–Weekend)
  - Marginal mean for Weather Station = Average of (Weather Station–Weekday + Weather Station–Weekend)
- We use marginal means to assess **main effects** in a two-way ANOVA.

## Means Table

| Source | Weekday | Weekend | Marginal Mean |
|---|---|---|---|
| **Resort Report** | 8.2 | 15.3 | 11.8 |
| **Weather Station** | 9.2 | 11.5 | 10.8 |
| **Marginal Mean** | 8.7 | 13.4 | — |

# Marginal Means in Two-Way ANOVA

## Means Table

| Source | Weekday | Weekend | Marginal Mean |
|---|---|---|---|
| **Resort Report** | 8.2 | 15.3 | 11.8 |
| **Weather Station** | 9.2 | 11.5 | 10.8 |
| **Marginal Mean** | 8.7 | 13.4 | — |

- We can use these marginal means to assess main effects:
  - **Main Effect of Source:** Resort Report (11.8) vs. Weather Station (10.8)
  - **Main Effect of Day:** Weekday (8.7) vs. Weekend (13.4)

# Means and Interactions in Two-Way ANOVA

## Means Table

| Source | Weekday | Weekend | Marginal Mean |
|---|---|---|---|
| **Resort Report** | 8.2 | 15.3 | **11.8** |
| **Weather Station** | 9.2 | 11.5 | **10.8** |
| **Marginal Mean** | **8.7** | **13.4** | — |

- To assess interactions, we look at the cell means directly:
  - Does the difference between Resort Report and Weather Station change from Weekday to Weekend?

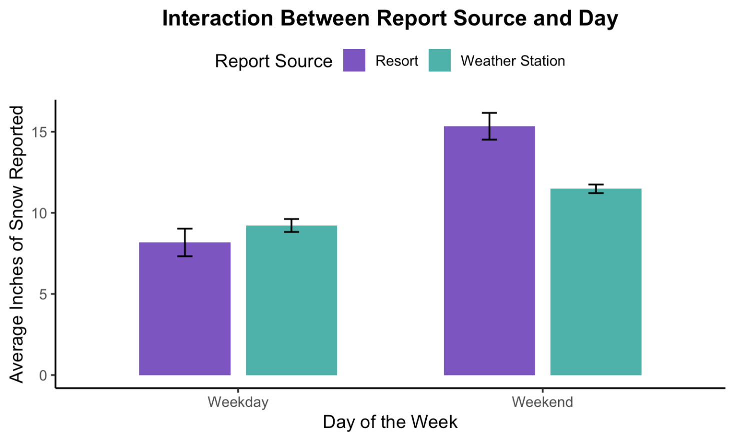
# Understanding interactions in ANOVA

- An interaction occurs when the effect of one independent variable on the dependent variable depends on the level of the other independent variable.
- In our ski resort example, the effect of Report Source (Resort vs. Weather Station) on reported snow levels changes depending on whether it’s a Weekday or Weekend.
- Often, it is difficult to understand interactions just from the output of statistical tests.

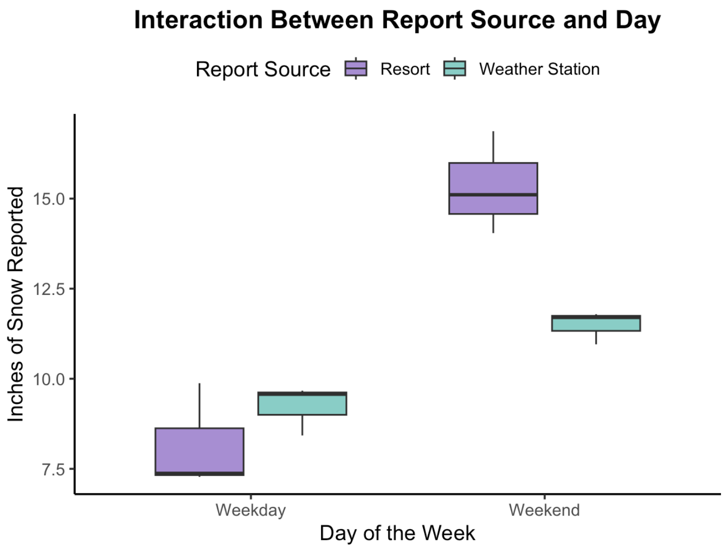
  –> **Visualization** is key!

# Understanding interactions in ANOVA: Visualization

## Interaction Bar Graph




## Interaction Boxplot



# Two-Way ANOVA: Interpreting statistical output

## Run Two-Way ANOVA in R

```r
#Run two-way ANOVA
anova_result <- aov(SnowReported ~ Source * Day, data = snow_data)
summary(anova_result)
```

## Example Output:

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
Source        1   5.93    5.93   4.836  0.05907 .  
Day           1  66.62   66.62  54.374 7.81e-05 ***
Source:Day    1  18.04   18.04  14.720  0.00497 ** 
Residuals     8   9.80    1.23                     
```

- There is one row for each main effect and one row for the interaction.
- Each row includes: Df, Sum Sq, Mean Sq, F value, and p-value (Pr(>F)).

# Two-Way ANOVA: Interpreting statistical output

```
              Df Sum Sq Mean Sq F value   Pr(>F)    
Source         1   5.93    5.93   4.836  0.05907 .  
Day            1  66.62   66.62  54.374 7.81e-05 ***
Source:Day     1  18.04   18.04  14.720  0.00497 ** 
Residuals      8   9.80    1.23                     
```

Which effects are statistically significant at α = .05?

**Answer:** Day (p < .001) and Source:Day interaction ($p$ = .00497) are significant. Source ($p$ = .05907) is not significant at α = .05.

What is the interpretation of these results in words?

**Answer:** The results suggest that the day of the week has a significant effect on reported snow levels, with weekends generally having higher reports. Additionally, the interaction between report source and day indicates that the difference between the two sources varies by day, with resorts exaggerating reports more on weekends compared to weekdays.

# Why use a two-way ANOVA?

- A two-way ANOVA allows us to compare:
  1. How levels from TWO independent variables affect the dependent variable
  2. How the combined (interaction) effects of those two independent variables affect the dependent variable
- A two-way ANOVA is a hypothesis test that includes:
  - Two nominal (or ordinal) independent variables
    - The number of levels of each IV doesn’t matter
  - One numeric (interval or ratio) dependent variable

# Why design a study with two IVs?

- It is often more efficient to evaluate more than one independent variable in a single study.
- It allows us to explore interactions between independent variables!
- Any ANOVA with two or more independent variables is called a **factorial ANOVA**.

# What if one IV is within-groups and the other is between-groups?

- This is called a mixed-design ANOVA. The calculations are more complex (and outside the scope of this course), but the logic is the same.

Which of the following studies would likely use a mixed-design ANOVA?

A. Testing how day of the week and report source affect snow reports.

B. Testing how students in different sections of PS 211 perform on exams 1, 2, and 3.

C. Testing how first-years vs. sophomores rate their satisfaction with three different dining halls.

D. All of the above.

**Answer: D** – All of the above!

# What if we have *lots* of independent variables?

- Multifactorial ANOVA can handle **more than two independent variables.**
- The logic is the same: we can test main effects for each IV and interactions between any combination of IVs.
- However, as the number of IVs increases, the complexity of the design and the required sample size also increase.
- A three-way ANOVA would involve:
  - Three main effects (one for each IV)
  - Three two-way interactions ($IV_1 \times IV_2$, $IV_1 \times IV_3$, $IV_2 \times IV_3$)
  - One three-way interaction ($IV_1 \times IV_2 \times IV_3$)
  - 7 *F* statistics in total!

# What if we have *lots* of independent variables?

- A four-way ANOVA would involve:
  - Four main effects
  - Six two-way interactions
  - Four three-way interactions
  - One four-way interaction
  - 15 *F* statistics in total!

**Example:**

- A researcher may want to examine how teaching method, class size, time of day, and student class year affect student performance on exams.

# Two-Way ANOVA in Action:

- In a 2012 study (Moss-Rascusin et al.), researchers investigated how applicant gender influenced science faculty members' evaluations of lab manager job applications.
- Faculty evaluated identical job applications labeled "John" or "Jennifer."
- Researchers ran a two-way ANOVA to analyze how labeled gender (male vs. female) and faculty participant gender (here, male vs. female) influenced ratings of the applicants' competence.
- **$IV_1$:** Applicant Gender
- **$IV_2$:** Faculty Gender
- **DV:** Competence Rating

**Finding:**

- Both male and female faculty rated "John" as more competent than "Jennifer."

How would you interpret this finding in terms of main effects and interactions?

A. There is a main effect of Applicant Gender.

B. There is a main effect of Faculty Gender.

C. There is an interaction between Applicant Gender and Faculty Gender.

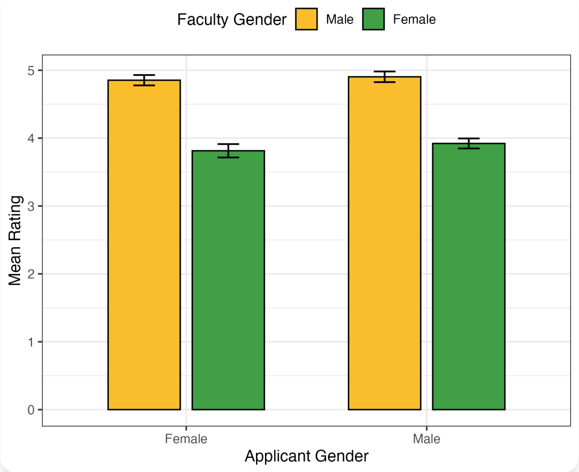
**Answer: A** – There is a main effect of Applicant Gender: both male and female faculty rated "John" as more competent than "Jennifer." There is no main effect of Faculty Gender, and no interaction

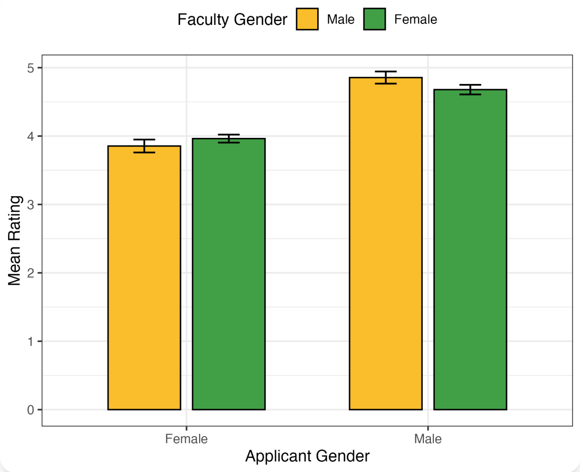
# Two-Way ANOVA in Action: Visualization Practice

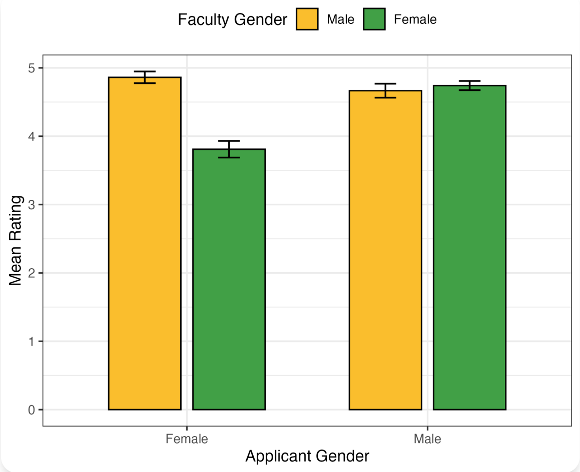
- In a 2012 study (Moss-Rascusin et al.), researchers investigated how applicant gender influenced science faculty members' evaluations of lab manager job applications.

- Both male and female faculty rated "John" as more competent than "Jennifer."

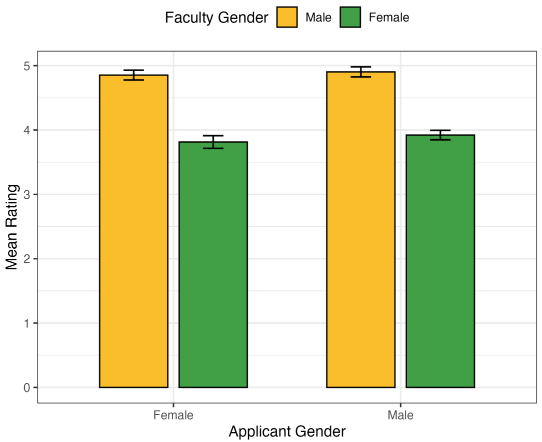
Which of the plots below best represents this finding?



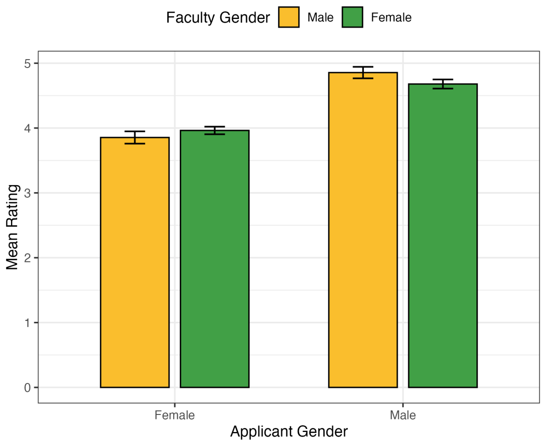



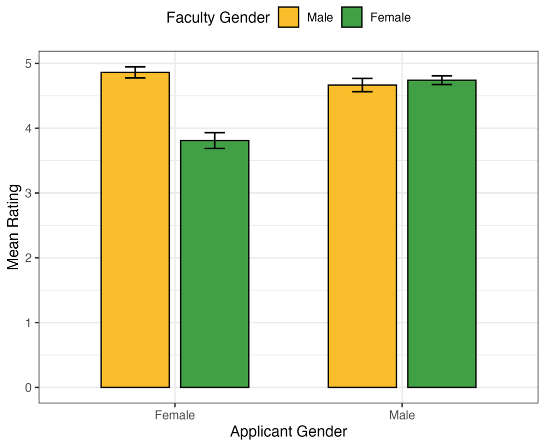
# Two-Way ANOVA in Action: Visualization Practice




Main effect of faculty gender




Main effect of applicant gender




Faculty gender x applicant gender interaction

# Reporting Two-Way ANOVA Results (APA Style)

When reporting:

- Mention the type of ANOVA (e.g., 2 × 2 between-groups)
- Include $F$, df, $p$, and $\eta^2$ for each main effect and interaction

Example:

> A 2 × 2 ANOVA revealed a significant main effect of applicant gender, $F(1, 120) = 5.12$, $p = .026$, $\eta^2 = .04$, on competence ratings, with male applicants rated more competent than female applicants. There was no main effect of faculty gender, nor was there an interaction between applicant gender and faculty gender (*ps* > .05).

# That's all for today!

Next class: Correlation & Regression