PSY 503: Foundations of Statistical Methods in Psychological Science

ANOVA, connections to lm()

Suyog Chandramouli

311 PSH (Princeton University)

Dec 1, 2025

Dec 1st

- Project presentations

- Lecture – this week
  - One-way (within subjects/repeated measures ANOVA)
  - Two-way ANOVA & interactions
  - ANCOVA
  - Factorial ANOVA

Recap

One-way ANOVA

- **Use Case:** Comparing means across three or more independent groups

- **As a Linear Model:**
  - $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + \varepsilon$
    - where X's are dummy coded for k-1 groups

  - NHST
    - Traditional Form: $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ vs H_1: not all μ_i equal
    - lm() equivalent: $H_0: b_1 = b_2 = \ldots = 0$ vs H_1: not all $b_i = 0$

One-way ANOVA

Variance due to manipulation / random error

You know these:

- k group means – $\mu_1, \mu_2, \ldots, \mu_k$
- k samples
  - k sample mean – $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$
  - Sample size – $n_1, n_2, \ldots, n_k$
  - Within group variation

Question: **Are the group means different?**

- "Is there a difference *somewhere*?"

Test-statistic

- Earlier
  $t = (\bar{X} - \mu) / (s/\sqrt{N})$

Now, comparing k means

$$F = \frac{MS_{between}}{MS_{within}}$$

"How much do groups vary compared to typical within-group noise?"

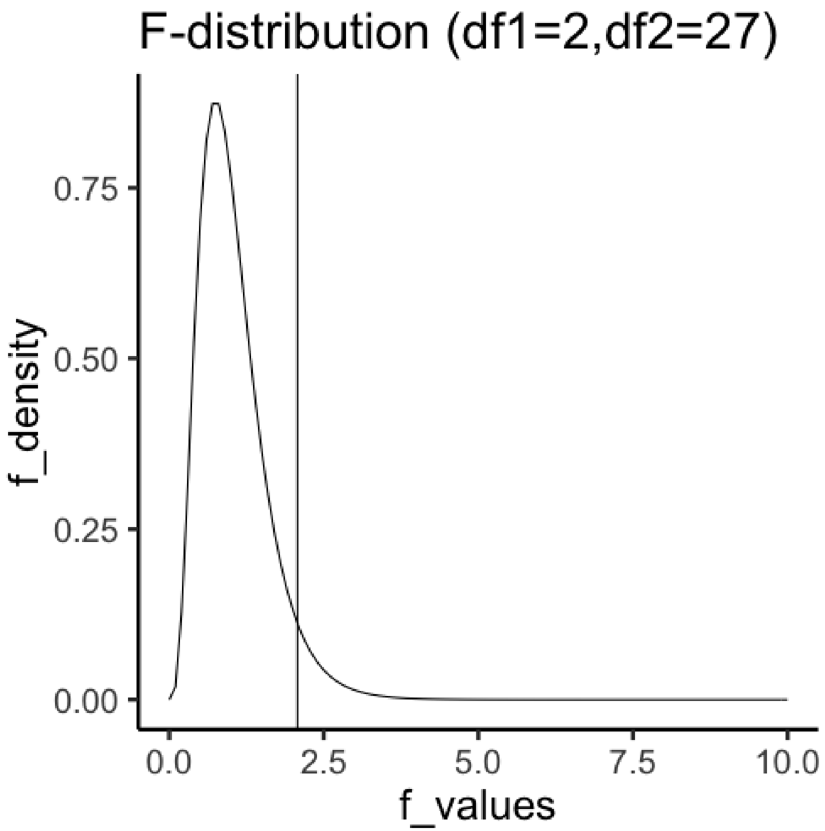
Observed F is computed directly from the data:

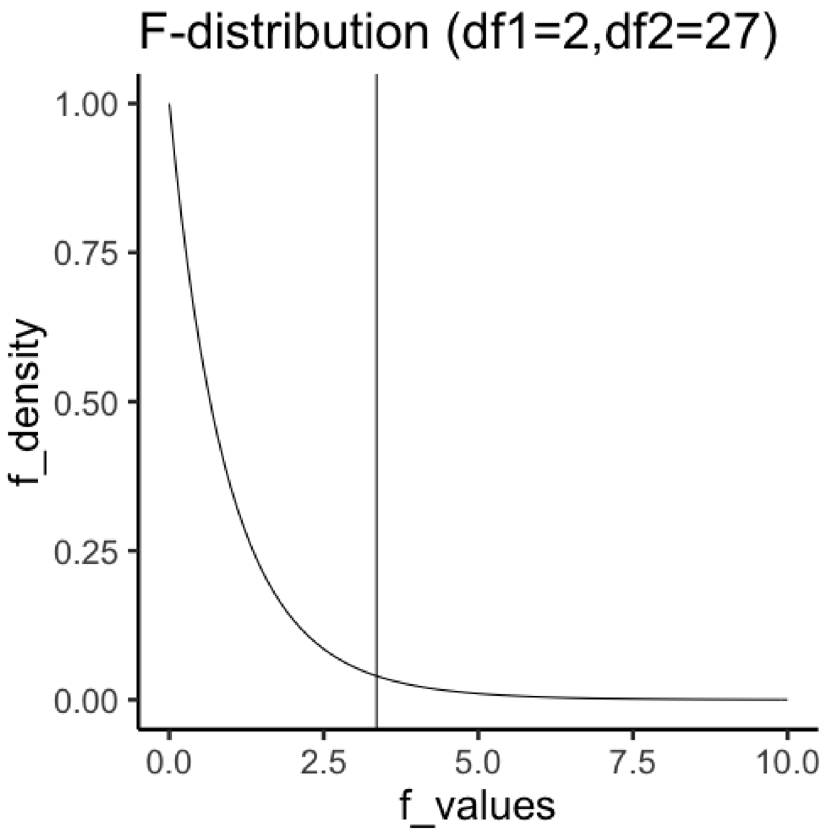
Source	df	SS	MSE	F	p
Effect	$k-1$	SS_{Effect}	$MS_{Effect} = \frac{SS_{Effect}}{k-1}$	$\frac{MS_{Effect}}{MS_{Error}}$	Calculated from F-distribution
Error	$n-k$	SS_{Error}	$MS_{Error} = \frac{SS_{Error}}{n-k}$		

k = number of groups; n = total sample-size

$$SS_{Effect} = \sum_{i=1}^{k} n_i (X_i - \bar{X})^2$$

$$SS_{Error} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{X}_i)^2$$


F-distribution (df1=2,df2=27)
f_density
f_values
0.00
0.25
0.50
0.75
0.0
2.5
5.0
7.5
10.0


F-distribution (df1=2,df2=27)
f_density
f_values
0.00
0.25
0.50
0.75
1.00
0.0
2.5
5.0
7.5
10.0

One-way ANOVA

- **Test Statistic**

$$F = \frac{\text{Between-group variability}}{\text{Within-group variability}} = \frac{\text{MS}_{\text{between}}}{\text{MS}_{\text{within}}}$$

- **Global test**
  - H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
  - Our F-test tells us if we can reject this

Example:
ANOVA: H_0: $\mu_1 = \mu_2 = \mu_3$
↓ (Reject H_0)

Example

```{r}
A <- c(20,11,2)
B <- c(6,2,7)
C <- c(2,11,2)
IV <- as.factor(rep(c("A","B","C"),each=3))
DV <- c(A,B,C)
df <- data.frame(IV,DV)
```

```{r}
df
```

IV <fctr>	DV <dbl>
A	20
A	11
A	2
B	6
B	2
B	7
C	2
C	11
C	2

9 rows

```{r}
aov(DV~IV,df)
```

```
Call:
   aov(formula = DV ~ IV, data = df)

Terms:
                 IV Residuals
Sum of Squares   72       230
Deg. of Freedom   2         6

Residual standard error: 6.191392
Estimated effects may be unbalanced
```

```{r}
aov_results <- aov(DV~IV,df)
summary(aov_results)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
IV           2     72   36.00   0.939  0.442
Residuals    6    230   38.33
```

F(2,6) = 0.939, p = 0.442

One-way ANOVA

- **Use Case:** Comparing means across three or more independent groups

- **As a Linear Model:**
  - $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + \varepsilon$
    - where X's are dummy coded for k-1 groups

  - NHST
    - Traditional Form: $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ vs H_1: not all μ_i equal
    - lm() equivalent: $H_0: b_1 = b_2 = \ldots = 0$ vs H_1: not all $b_i = 0$

Follow-up comparisons

- ANOVA only tests the ***omnibus*** question…
  - Are there any differences anywhere?
- Need to conduct additional tests to compare specific means…
  - Questions like:
- Numerous recommendations for the "right" way to do this
  - Simplest: follow-up t-tests
    - T-test on every group pair!
      - Increases Type-1 error rate

Increasing Type-1 error rate

- **One test:** One decision (binary) with error α of messing up
- **Three tests:** three chances to mess up

  Test 1: Group A vs B → reject?
  Test 2: Group A vs C → reject?
  Test 3: Group B vs C → reject?

  - Each has 5% error.
  - And cross all three?
    - ___ % chance you reject **at least one** incorrectly.

- **A fix**
  - Adjust α for each test to keep overall error at 5%.

Post-hoc corrections

- **Bonferroni:**
  - New $\alpha = 0.05 / 3 = 0.017$
  - Use 0.017 for ALL three tests.

- **Holm**
  - Sequential adjustment for p-values
    - Smallest p-value: compare to $\alpha' = 0.05/3 = 0.017$
    - Second smallest: compare to $\alpha'' = 0.05/2 = 0.025$
    - Largest: compare to $\alpha''' = 0.05/1 = 0.05$
  - Different threshold for each test

One-way ANOVA

- **Test Statistic**

$$F = \frac{\text{Between-group variability}}{\text{Within-group variability}} = \frac{\text{MS}_{\text{between}}}{\text{MS}_{\text{within}}}$$

- **Global test**
  - H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
  - Our F-test tells us if we can reject this
  - **If We Reject H_0 ($p < \alpha$):**
    - We can then investigate specific patterns through:
      - a) Planned Contrasts
      - b) post-hoc tests (Bonferroni, etc.)

Example:
Initial ANOVA: H_0: $\mu_1 = \mu_2 = \mu_3$
↓ (Reject H_0)

Specific Questions:
- Is $\mu_1 > \mu_2$? (Pairwise)
- Is $\mu_1 > (\mu_2 + \mu_3)/2$? (Contrast)

Type-1 error rate (planned contrasts)

- **We know**
  - **One test:** One decision (binary) with error α of messing up
  - **Three tests:** three chances to mess up

    Test 1: Group A vs B → reject?
    Test 2: Group A vs C → reject?
    Test 3: Group B vs C → reject?
    - Each has 5% error.
  - N= ${}^{K}C_2$ pairs exists for any test : k(k-1)/2

- **Planned contrasts**
  - Ahead of time, pick the few (n') you care about
  - So you don't have to correct across all pairs
    - Bonferroni: α' = 0.05/ n'

One-way ANOVA types

- Between-subjects (independent groups)
- Within-subjects (repeated measures)

One-way ANOVA (between subjects)

You know these:

- k group means – $\mu_1, \mu_2, \ldots, \mu_k$
- k samples
  - k sample mean – $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$
  - Sample size – $n_1, n_2, \ldots, n_k$
  - Within group variation

Question: **Are the group means different?**

- Not: "Which ones differ?"
- Not: "By how much?"
- Just: "Is there a difference *somewhere*?"

Variance due to manipulation / random error

Test-statistic

- Earlier
  $t = (\bar{X} - \mu) / (s/\sqrt{N})$

Now, comparing k means

$$F = \frac{MS_{between}}{MS_{within}}$$

- **"How much do conditions vary compared to each person's inconsistency?"**

One-way ANOVA (within subjects)

One-way ANOVA (within subjects)

Condition differences /
(Noise – subject variance)

You know these:

- k ***conditions*** – $\mu_1, \mu_2, \ldots, \mu_k$
- n subjects – each measured k times
  - Condition means – $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$
  - Subject means – $\bar{X}_{s1}, \bar{X}_{s2}, \ldots, \bar{X}_{sk}$
  - Residual variation

Question: **Are the conditions means different?**

- Same question as between-subjects
- But now: *same* people across conditions

Test-statistic

B/w subjects ANOVA

$$F = \frac{MS_{between}}{MS_{within}}$$

Error: all within-group variance

Now,

$$F = \frac{MS_{effect}}{MS_{residual}}$$

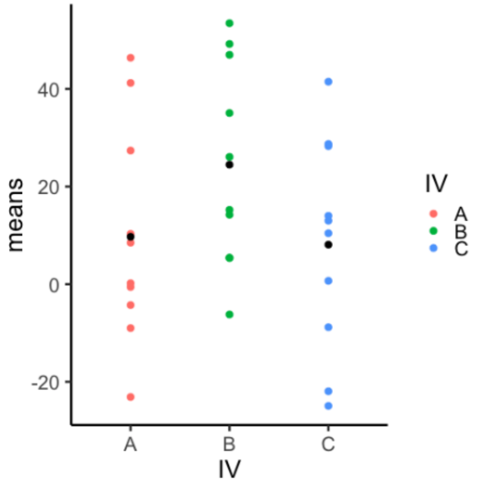
Residual = what's left after removing subject differences from error

Key idea:

- each person serves as their own control
  - we can partition out stable individual differences.
    - Smaller error term → more power.

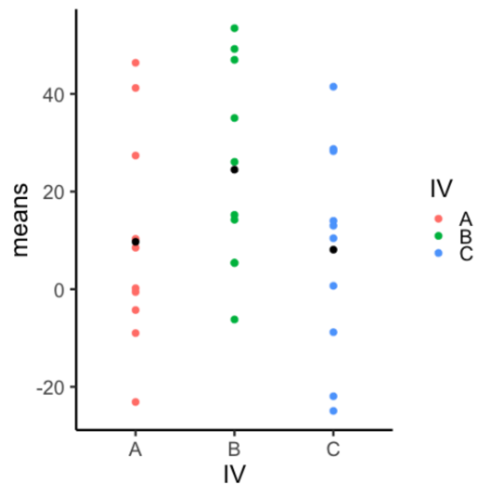
"How much do conditions vary compared to each person's inconsistency?"

B/W Subjects



- MS Effect = 815.97
- MS Error = 472.21
- F = 1.73

B/W Subjects

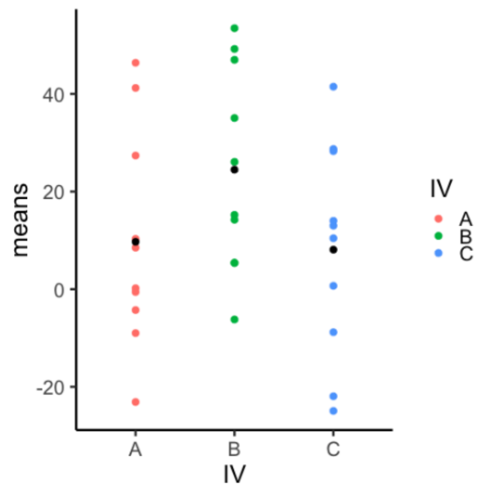


- MS Effect = 815.97
- MS Error = 472.21
- F = 1.73

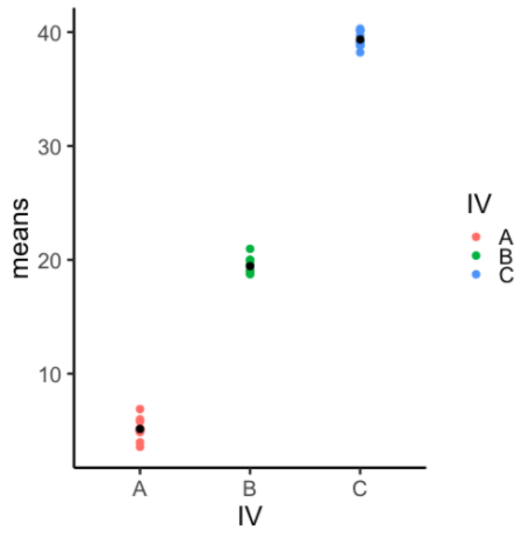
Hope:

- Big gaps between groups, small spread within
    - Tight clusters, far apart

B/W Subjects



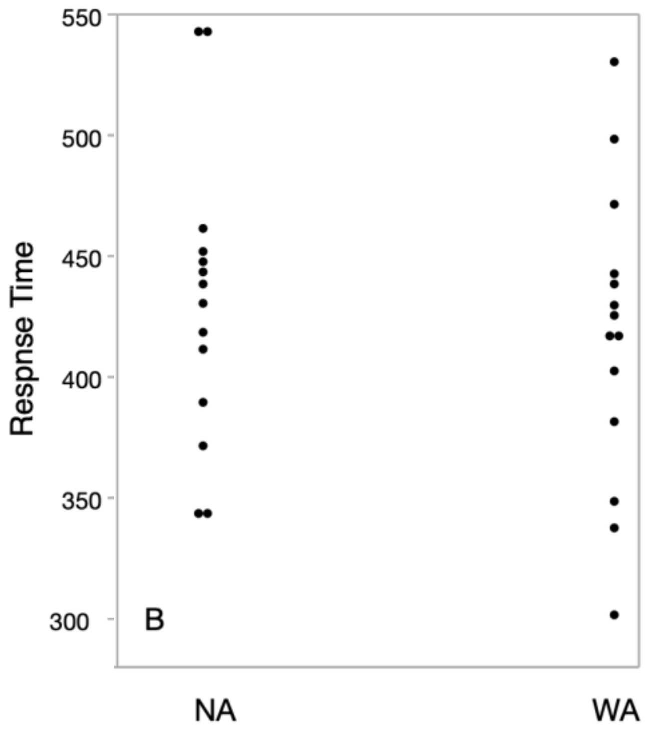
- MS Effect = 815.97
- MS Error = 472.21
- F = 1.73



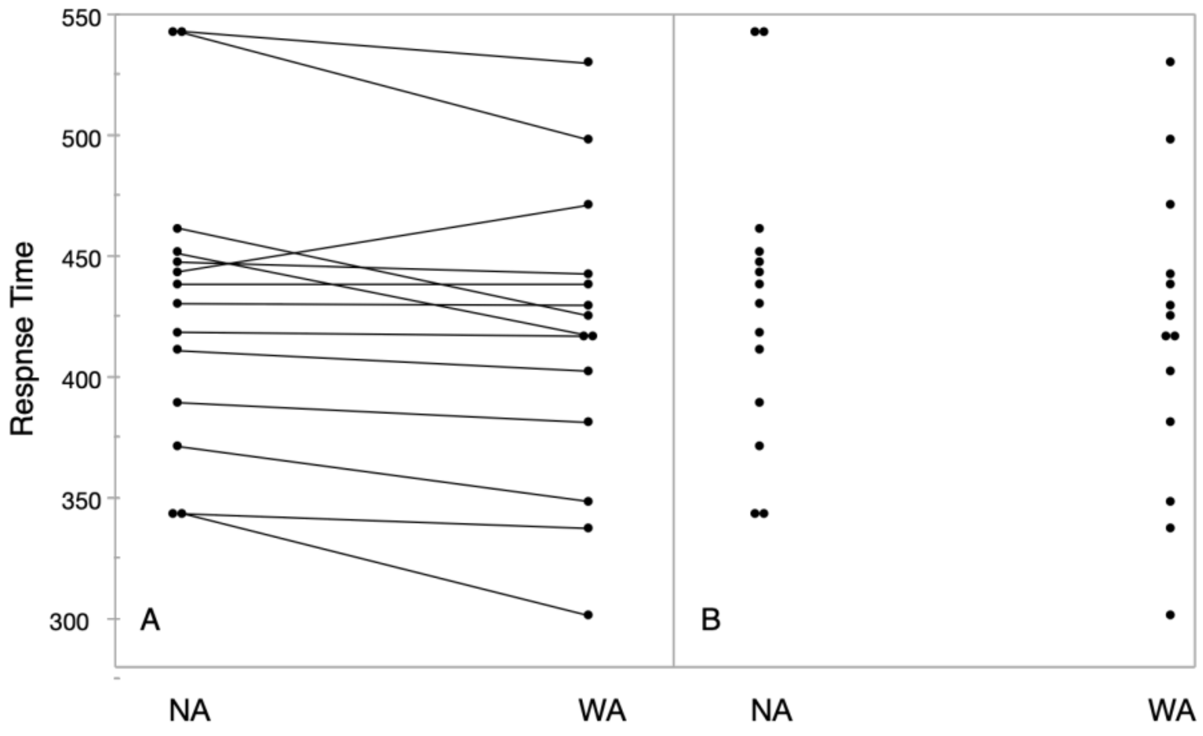
- MS Effect = 2954.08
- MS Error = 0.63
- F = 4720.71

Hope:
- Big gaps between groups, small spread within
    - Tight clusters, far apart

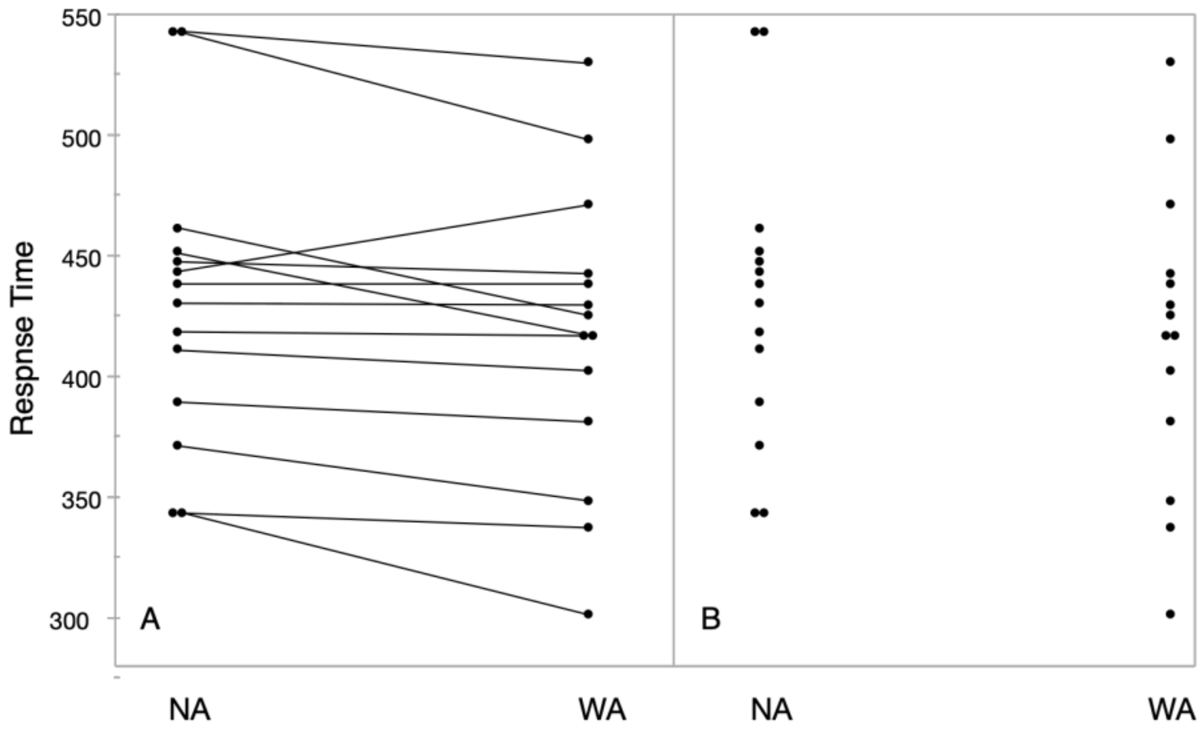
Within Subjects



Within Subjects



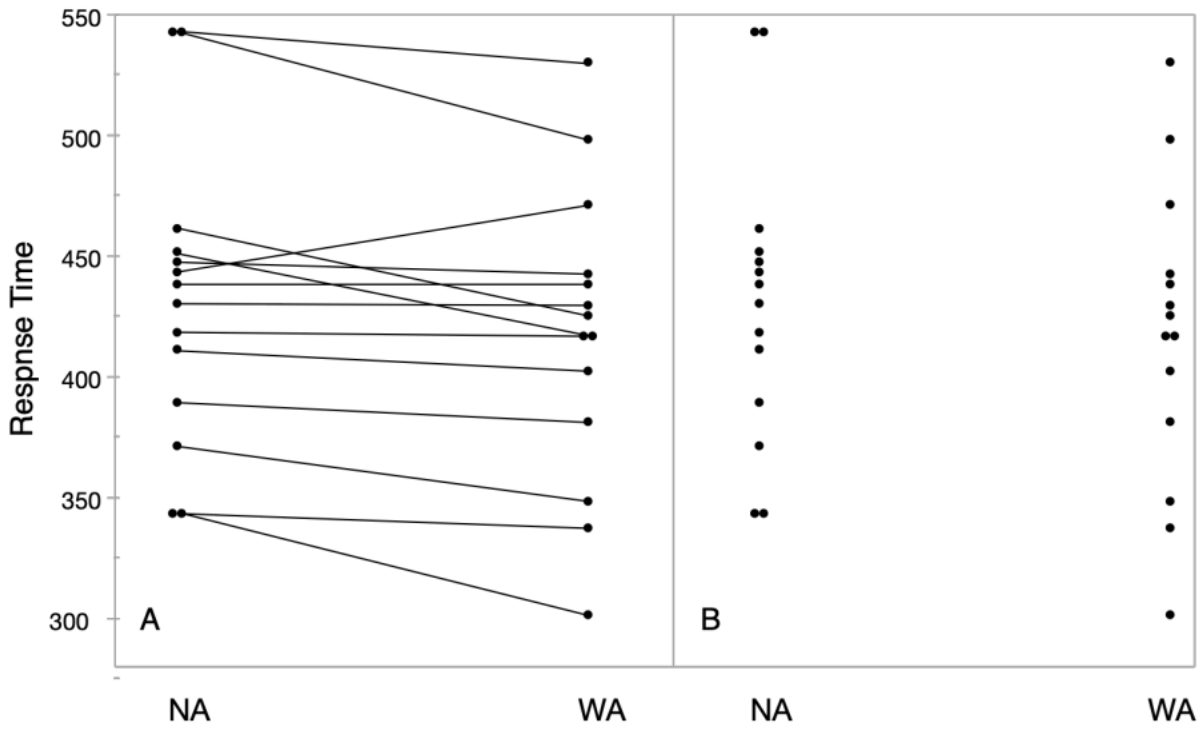
Within Subjects



Hope:
- Everyone shifts in the same direction, by similar amounts

Within Subjects

$$F = \frac{MS_{effect}}{MS_{residual}} = \frac{\text{Condition differences}}{\text{Noise} - \text{subject variance}}$$

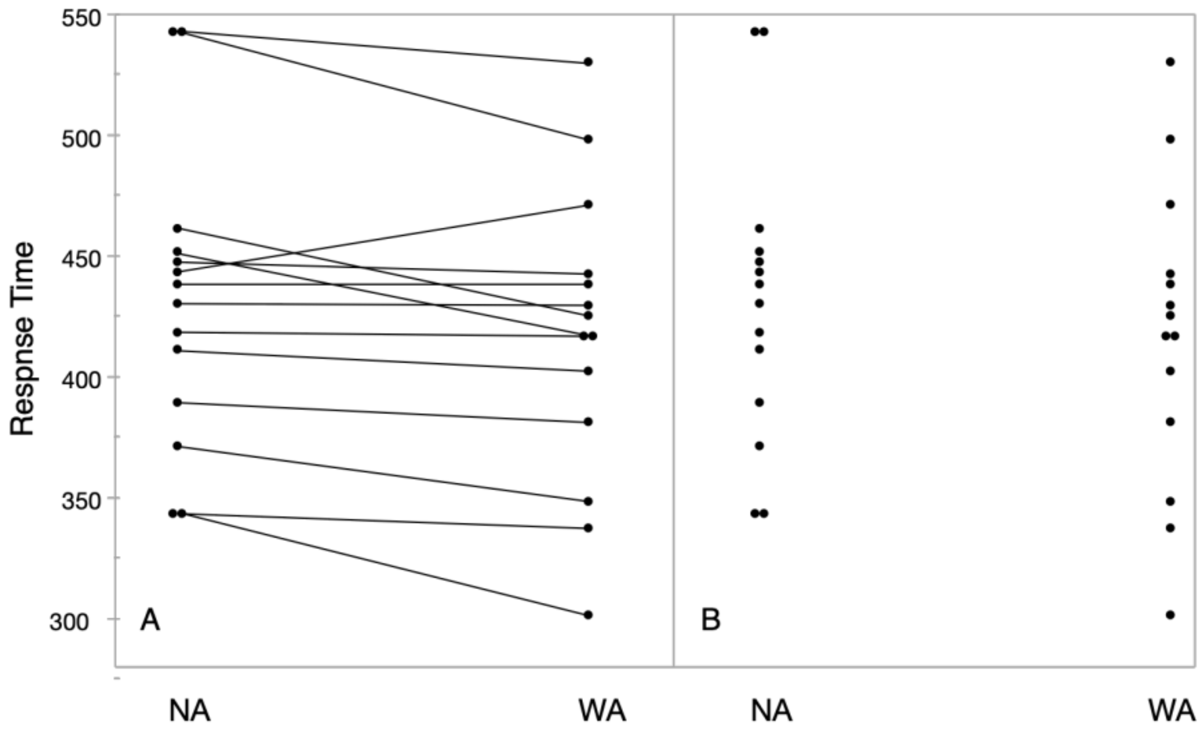


Hope:

- Everyone shifts in the same direction, by similar amounts

Within Subjects

$$F_{within} = \frac{MS_{effect}}{MS_{residual}} = \frac{\text{Condition differences}}{\text{Noise – subject variance}}$$

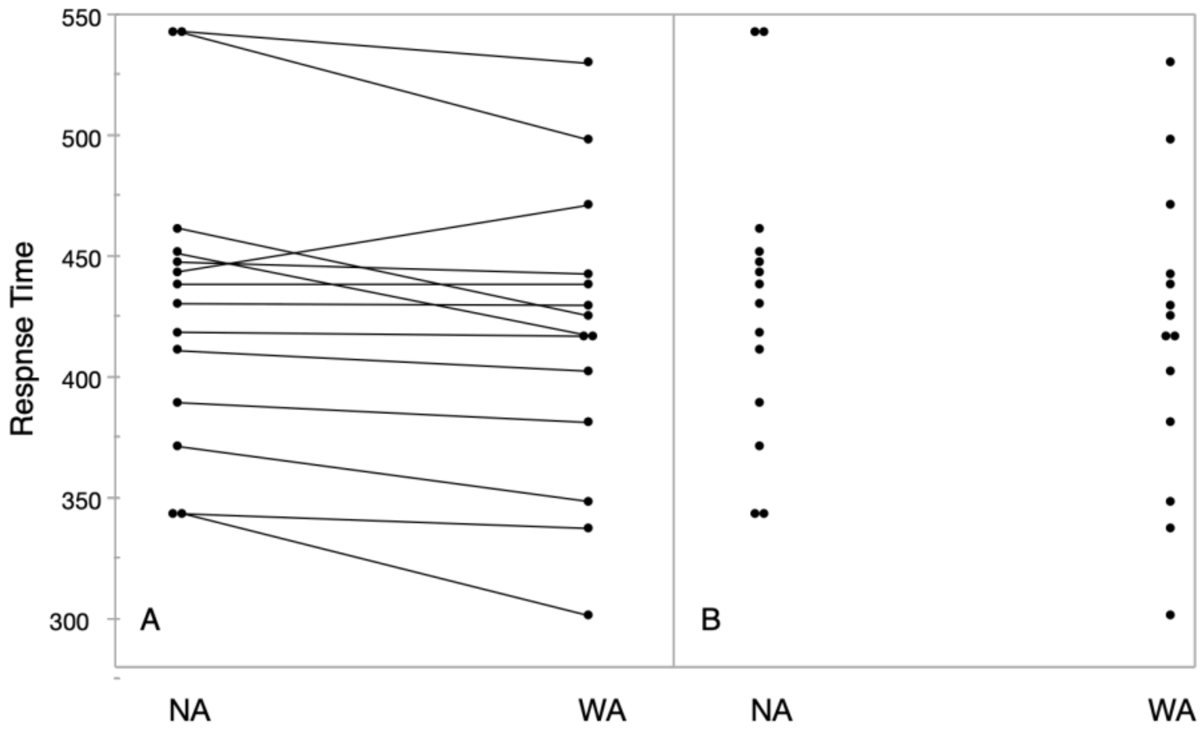


Hope:

- Everyone shifts in the same direction, by similar amounts

Within Subjects

$$F = \frac{MS_{effect}}{MS_{residual}} = \frac{\text{Condition differences}}{\text{Noise – subject variance}}$$



Discuss:

What makes F big?

What makes F small?

Hope:

- Everyone shifts in the same direction, by similar amounts

Between-Subjects Design

SS TOTAL

SS Effect (SS between-groups) | SS Error (SS within-groups)

$$\text{SS TOTAL} = \underbrace{\text{SS Effect}}_{\text{(SS between groups)}} + \underbrace{\text{SS Error}}_{\text{(SS within groups)}}$$

Repeated-Measures Design

SS TOTAL

SS Effect (SS between-groups) | SS Error (SS within-groups)

SS Subjects | SS Error

$$\text{SS TOTAL} = \underbrace{\text{SS Effect}}_{\text{(SS between groups)}} + \boxed{\underbrace{\text{SS Error}}_{\text{(SS within groups)}}}$$

This get's split up into two parts

split up

$$\text{SS TOTAL} = \underbrace{\text{SS Effect}}_{\text{(SS between groups)}} + \boxed{\overbrace{(\underbrace{\text{SS Subjects}}_{\text{(SS Subjects)}} + \underbrace{\text{SS Error}}_{\text{(SS Left-over Error)}})}^{\text{(SS within groups)}}}$$

Within-subjects ANOVA

Source	df	SS	MSE	F	p
Subjects		$SS_{Subjects}$			
Effect	$k-1$	SS_{Effect}	$MS_{Effect} = \frac{SS_{Effect}}{k-1}$	$\frac{MS_{Effect}}{MS_{Error}}$	Calculated from F-distribution
Error	$(n-1)$*(k-1)	SS_{Error}	$MS_{Error} = \frac{SS_{Error}}{(n-1)(k-1)}$		

k = number of groups; n = number of subjects

$$SS_{Total} = SS_{Effect} + SS_{Subjects} + SS_{Error}$$

$$SS_{Effect} = \sum_{i=1}^{k} n_i (X_i - \bar{X})^2$$

$$SS_{Subjects} = k \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Notes: $\bar{X}$= Grand Mean, X_i = condition mean (SS effect), or subject mean (SS Subjects)

Within-subject designs

- Between-subjects
  - we hope that individual differences cancel out between (e.g. treatment & control) conditions
- Within-subjects
  - we *know* they cancel out — same people, every condition

Within-subject designs

- Between-subjects
  - we hope that individual differences cancel out between (e.g. treatment & control) conditions

- Within-subjects
  - we *know* they cancel out — same people, every condition

- **Benefits**
  - More power
  - Fewer participants
    - Cost-effective
  - Better controls

Within-subject designs

- Between-subjects
  - we hope that individual differences cancel out between (e.g. treatment & control) conditions
- Within-subjects
  - we *know* they cancel out — same people, every condition
- **Benefits**
  - More power
  - Fewer participants
    - Cost-effective
  - Better controls

The tradeoff

Carryover effects. Practice effects. Fatigue.

Within-subject designs

- Between-subjects
  - we hope that individual differences cancel out between (e.g. treatment & control) conditions
- Within-subjects
  - we *know* they cancel out — same people, every condition
- **Benefits**
  - More power
  - Fewer participants
    - Cost-effective
  - Better controls

The tradeoff

Carryover effects. Practice effects. Fatigue.

The fix

Counterbalance

Difference to Detect 90% Confidence & 80% Power	Within-Subjects Sample Size	Between-Subjects Sample Size
20%	50	150
10%	115	614
5%	246	2468
4%	312	3860
3%	421	6866
2%	640	15452
1%	1297	61822

From: https://measuringu.com/between-within/

One-way ANOVA

- **Use Case:** Comparing means across three or more independent groups

- **As a Linear Model:**
  - $Y = b_0 + b_1X_1 + b_2X_2 + \ldots +$ **subject effects** $+ \varepsilon$
    - where X's are dummy coded for k-1 conditions
    - subject effects absorb individual differences

  - NHST
    - Traditional Form: $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ vs H_1: not all μ_i equal
    - lm() equivalent: $H_0: b_1 = b_2 = \ldots = 0$ vs H_1: not all $b_i = 0$

Implementation in R (Example)

````
```{r}
A <- c(1,2,1,2,4)
B <- c(9,7,6,4,4)
C <- c(2,9,2,3,4)
DV <- c(A,B,C)
subjects <- as.factor(c(1,2,3,4,5))
IV <- rep(c("A","B","C"), each=5)
df <- data.frame(subjects,IV, DV)
```
````

````
```{r}
aov(DV~IV + Error(subjects/IV), df)
```
````

Implementation in R (Example)

```{r}
A <- c(1,2,1,2,4)
B <- c(9,7,6,4,4)
C <- c(2,9,2,3,4)
DV <- c(A,B,C)
subjects <- as.factor(c(1,2,3,4,5))
IV <- rep(c("A","B","C"), each=5)
df <- data.frame(subjects,IV, DV)
```

	subjects	IV	DV
1	1	A	1
2	2	A	2
3	3	A	1
4	4	A	2
5	5	A	4
6	1	B	9
7	2	B	7
8	3	B	6
9	4	B	4
10	5	B	4
11	1	C	2
12	2	C	9
13	3	C	2
14	4	C	3
15	5	C	4

Implementation in R (Example)

````
```{r}
A <- c(1,2,1,2,4)
B <- c(9,7,6,4,4)
C <- c(2,9,2,3,4)
DV <- c(A,B,C)
subjects <- as.factor(c(1,2,3,4,5))
IV <- rep(c("A","B","C"), each=5)
df <- data.frame(subjects,IV, DV)
```
````

````
```{r}
aov(DV~IV + Error(subjects/IV), df)
summary(aov(DV~IV+Error(subjects/IV),df))
```
````

Implementation in R (Example)

````
```{r}
A <- c(1,2,1,2,4)
B <- c(9,7,6,4,4)
C <- c(2,9,2,3,4)
DV <- c(A,B,C)
subjects <- as.factor(c(1,2,3,4,5))
IV <- rep(c("A","B","C"), each=5)
df <- data.frame(subjects,IV, DV)
```
````

````
```{r}
aov(DV~IV + Error(subjects/IV), df)
summary(aov(DV~IV+Error(subjects/IV),df))
```
````

```
Error: subjects
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  4     18     4.5

Error: subjects:IV
          Df Sum Sq Mean Sq F value Pr(>F)
IV         2     40      20       4 0.0625 .
Residuals  8     40       5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

New assumption to check for (within subjects)

- Earlier,
  - Between: Homogeneity of variance — equal spread in each group.
  - ..
- Now,
  - Within: **Sphericity** — equal variance of *differences* (between all pairs of conditions)
    - Stricter. Easier to violate.
    - Relevant when there are more than 2 conditions
      - var(A−B) = var(A−C) = var(B−C)?
      - Visual check (very rough) : Lines crossing, diverging unpredictably → sphericity violated
    - Mauchly's test — returns p-values if assumption violated

Post-hoc corrections (within subjects)

- Applicable when there are multiple measurements
  - Not relevant for a simple before/after

- Same ideas as before otherwise
  - One-way repeated measures anova
    - Start with Ombinus test
    - Then use ***paired*** t-tests
      - Bonferroni
      - Holm,
      - e.t.c

Questions we’ve asked so far

- **t-test (independent):** Is there a difference between two groups?
- **t-test (paired):** Is there a difference between two conditions? (same people)
- **One-way ANOVA (between):** Is there a difference somewhere among k groups?
- **One-way ANOVA (within / repeated measures):** Is there a difference somewhere among k conditions? (same people)

*All are variants of **“Is there an effect that exists?”***

But there are often other questions

- Does the effect of A depend on B?

But there are often other questions

- Does the effect of A depend on B?
- Examples
  - Does the effect of distraction on memory depend on age?
  - Does the bystander effect depend on group size?
  - Does stereotype threat depend on task difficulty?
  - Does parenting style affect outcomes differently for boys vs girls?
  - Does CBT work better for anxiety vs depression?
  - …

Does the effect of A depend on B?

- Equivalent to:
  - **"Is there an interaction between A and B?"**
    - Are A and B acting alone?
      Or are they working together?
  - **"Does B moderate the relationship between A and the outcome/DV?"**

- Depends on there being more than 1 (predictor / X) variable.
  Answered by:
  - Two-way anova or higher (factorial ANOVA)
  - ANCOVA / regression with interaction
  - Regression

Methods to assess interaction

- Y is continuous

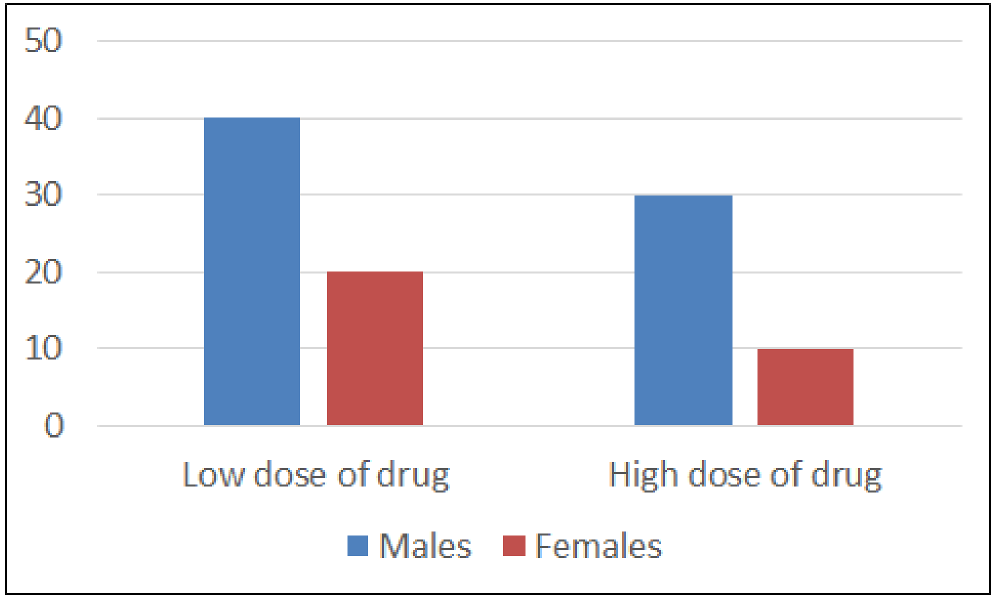
A	B	Method
Categorical	Categorical	Factorial ANOVA
Categorical	Continuous	ANCOVA / regression with interaction
Continuous	Categorical	Moderated regression
Continuous	Continuous	Moderated regression

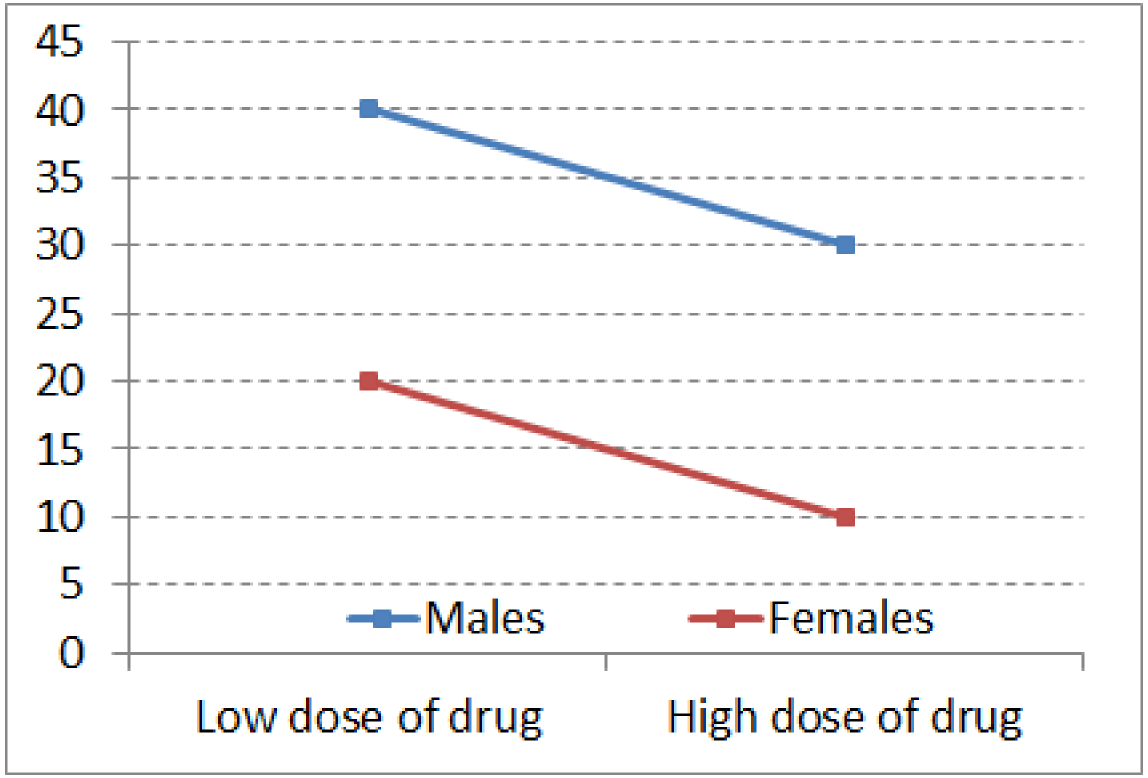
- Y is binary: logistic regression with interaction term
- Y is ordinal: ordinal regression with interaction term

Visualizing interactions

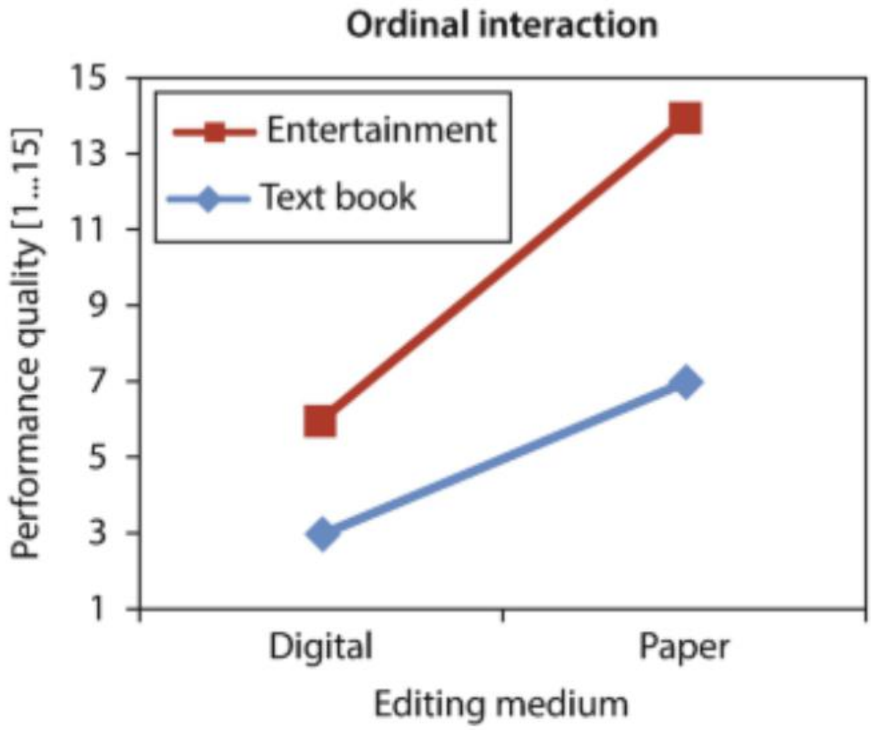
- **Golden Rule of Plots:** Look at the lines.
  - **Parallel Lines = No Interaction**
    - The effect of Variable A is the same, regardless of Variable B.
    - They are additive (independent).
  - **Non-Parallel Lines = Interaction**
    - The lines diverge, converge, or cross.
    - The effect of Variable A *changes* based on the level of Variable B.
    - This is the "It Depends" effect.

Visualization: no-interaction

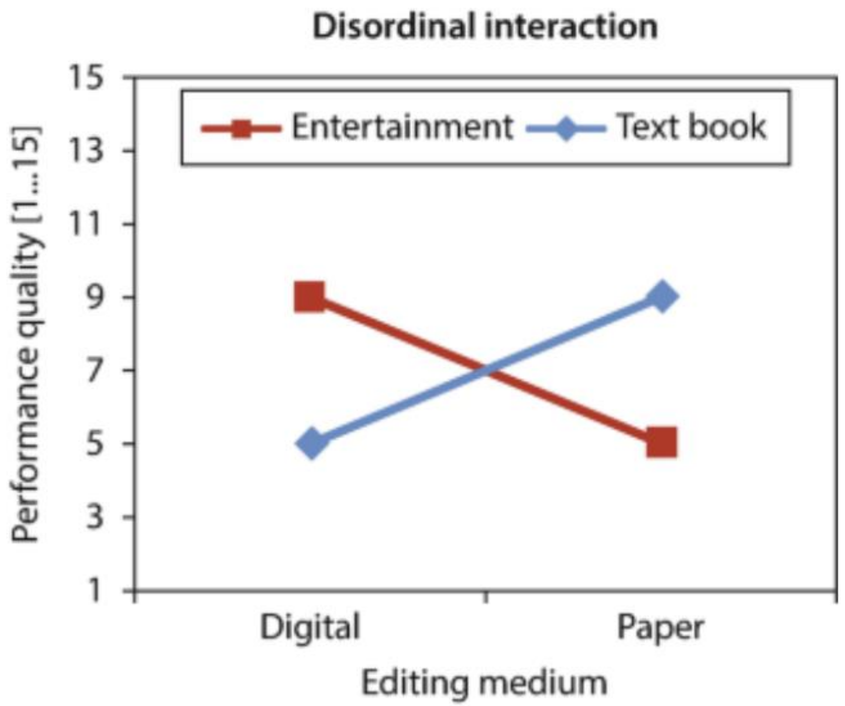




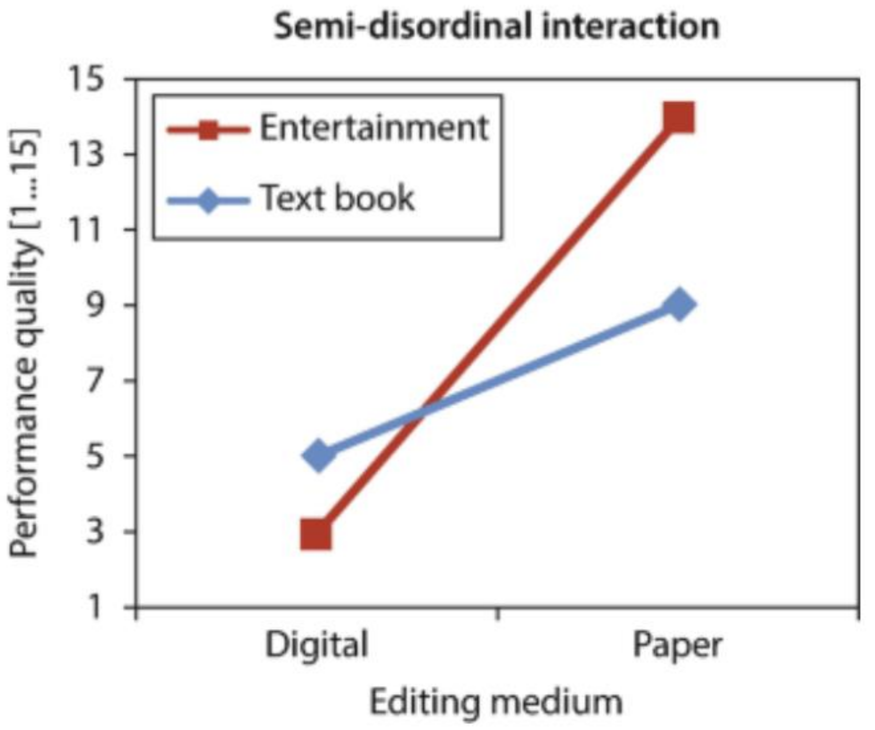
Visualization: interaction



Visualization: interaction



Visualization: interaction



Two-way ANOVA

Two-way ANOVA

- Two factors. One outcome.
- Questions:
  - Main effect of A?
  - Main effect of B?
  - Interaction: Does A depend on B?

Two-way ANOVA

Between group variation /
variation within groups

You know these:

- Two factors – A and B
  - A has i levels
  - B has j levels
- that result in i*j cells
  - Cell means
    - $\bar{X}_{11}, \bar{X}_{12}, \ldots, \bar{X}_{ij}$
  - Cell sizes
    - $n_{11}, n_{12}, \ldots, n_{ij}$
  - Within group variation

Question: **Are there main effects? Is there an interaction?**

Test-statistic

- Earlier

$z = (\bar{X} - \mu) / (\sigma/\sqrt{N})$

$z = (\bar{X} - \mu) / SE$

$t = (\bar{X} - \mu) / (s/\sqrt{N})$

Now, comparing k means

$$F = \frac{MS_{between}}{MS_{within}}$$

(comparing k means)

"Does Factor A matter? Factor B? *Do they interact?*"

Two-way ANOVA

- **Use Case:** Comparing means across groups while considering two different categorical factors and their interaction

- **As a Linear Model:**
  - $Y = b_0 + b_1X_1 + b_2X_2 + b_3(X_1 \times X_2) + \varepsilon$
    - where X's are dummy coded

  - NHST
    - Traditional Form:
      - H_0^A: No main effect of factor A
      - H_0^B: No main effect of factor B
      - $H_0^{A \chi B}$: No interaction between A and B
    - lm() equivalent:
      - H_0^A: $b_1 = 0$
      - H_0^B: $b_2 = 0$
      - $H_0^{A \chi B}$: $b_3 = 0$

Main Effects vs Interaction

- **Main effect of A:** Is there an average difference across levels of A? (Collapsing across B)
- **Main effect of B:** Is there an average difference across levels of B? (Collapsing across A)
- **Interaction:** Do the effects combine in a non-additive way?

When to use what?

Design	**Method**
1 factor, 3+ groups	One-way ANOVA
1 factor, repeated	RM-ANOVA
2+ factors	Factorial ANOVA
Factor + covariate	ANCOVA
Interaction question	Two-way or higher