EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- ANOVA
  - Family-wise Alpha
- Hypotheses Formation
- ANOVA Visualized
- Sources of Variance
  - Within Group Variance
  - Between Group Variance
- Ratio of Between and Within Group Variance

Two or More

□ Thus far we have compared one group to another group (whether a known group, another sample group, or a pre/post group)

What if we want to compare more than two groups?

ANOVA

ANOVA

AN = Analysis
O = of
VA = Variance

Historical Moment: Ronald Fisher

- **Ronald Aylmer Fisher** (Feb. 17, 1890 – July 29, 1962)
- British statistician and geneticist
- Combined Mendelian Generics (think peas) and natural selection using math.
- Lots of other contributions:
  - Maximum likelihood estimation
  - Population genetics
  - ANOVA

Historical Moment: Ronald Fisher

- **Ronald Aylmer Fisher** (Feb. 17, 1890 – July 29, 1962)

- But… Like Francis Galton, he was really into this idea of heredity and planned reproduction.

- He was a loud and proud eugenicist…

- He had no doubt the Nazi's "sincerely wished to benefit the German racial stock, especially by the elimination of manifest defectives" and he would have given "his support to such a movement"

Why Not Multiple t-tests?

- If we have four groups (A, B, C, D), we would have to do:
  - A-B
  - A-C
  - A-D
  - B-C
  - B-D
  - C-D

- We would actually have to do SIX t-tests to check all the different combinations
- When we do this, our .05 significance level actually compounds and gets a lot bigger than .05.
  - Called "family-wise" error rate, and it dramatically increases the chances of making a Type I error.
- Instead we conduct an ANOVA

Hypotheses for ANOVA

What do you think the null and alternative hypotheses for ANOVA are?

Hypotheses for ANOVA

- As usual, the NULL states something like,
  - "The means for all groups are equal."
    - "There is no difference between the groups."

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

- The alternative hypothesis states that AT LEAST ONE group mean is different from the other.
  - "At least one group mean is not equal to the others."

$$H_0: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

ANOVA Studying Example

- I want to know whether there is any difference in using different materials to learn Statistical Literacy as assessed by an exam.
  - Three Groups:
    - Group 1: Using book only to study
    - Group 2: Using lecture notes only to study
    - Group 3: Using lecture notes and the book to study

What am I going to compare to test if at least one group differs from the others?

Comparing Variance

- What am I going to compare to assess differences?
  - "Analysis of the *Variance*"
- Eventually we will compare the means, which is what we are used to, but first... what we are really testing is the VARIANCE between and within the groups...
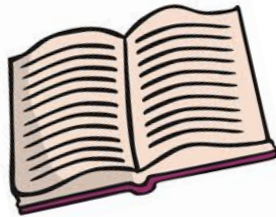
What is variance, again?
Why would it matter?

ANOVA Visualized
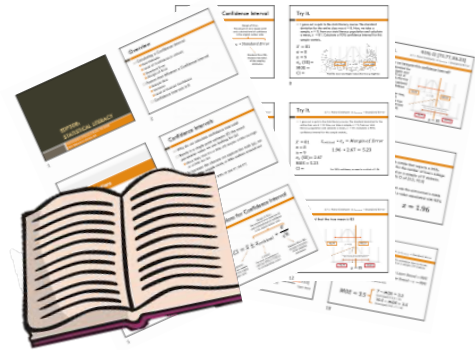
ANOVA Studying Example

Population 1	Population 2	Population 3

Sample 1
(Book Only)
$n_1 = 10$
$\bar{x}_1 = 7.60$

Sample 2
(Lecture Notes)
$n_2 = 10$
$\bar{x}_2 = 17.60$

Sample 3
(Lecture Notes and Book)
$n_3 = 10$
$\bar{x}_3 = 19.30$

Just by looking at the results, do you think the results will be significant?

Which group means do you think will be significantly different from each other?

Variance WITHIN One Group

- Variance, s^2 ,is the amount of spread (difference) WITHIN the observations for ONE group
  - What we've been using all along
  - Group 1(Book Only) has it's own variability
    - Ex. One person might have scored high (15), but the average score was only 7.60

Variance WITHIN Group 1

$$\bar{x}_1 = 7.60$$
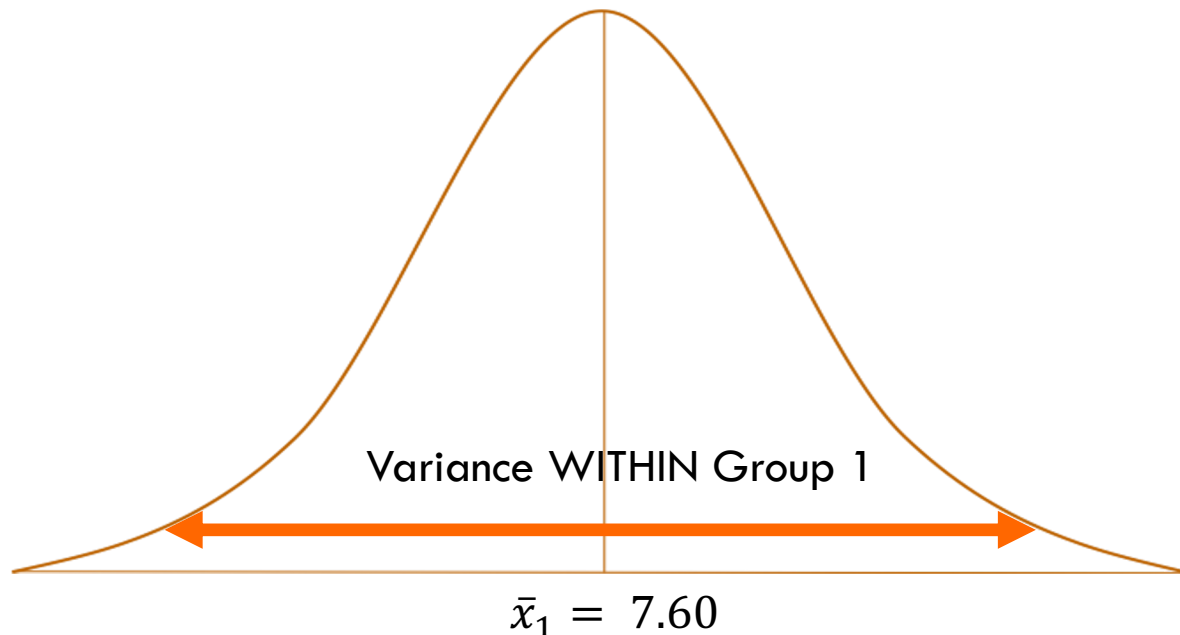
Variance WITHIN Groups

- Each group has its own WITHIN group variance
  - Regardless which group you're in, there will be variation from person to person based on some preexisting variables
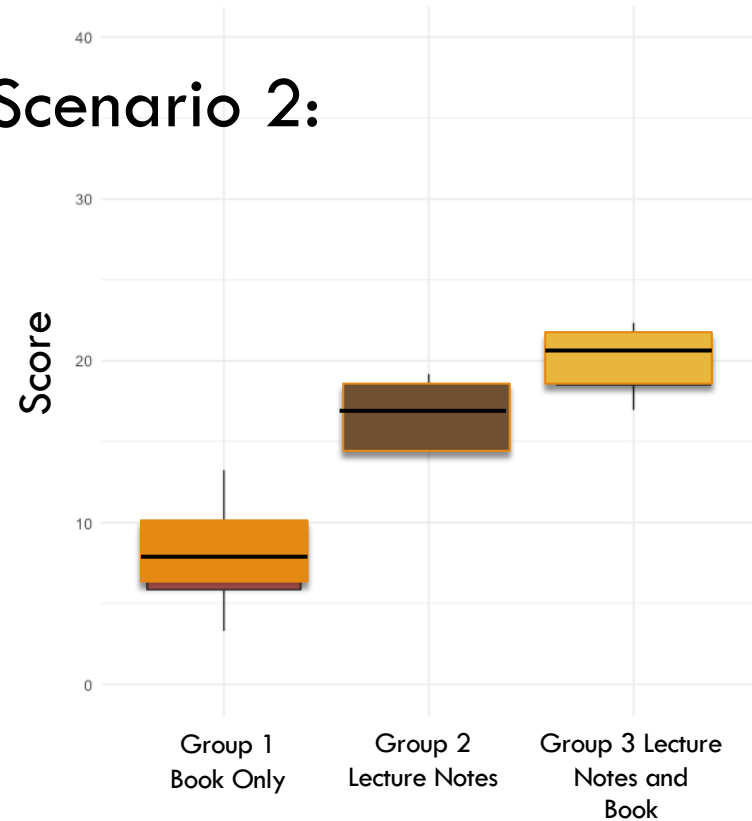   - Ex. Some people like math more than other

$$\bar{x}_1 = 7.60 \qquad \bar{x}_2 = 17.60 \qquad \bar{x}_3 = 19.30$$

Group 1
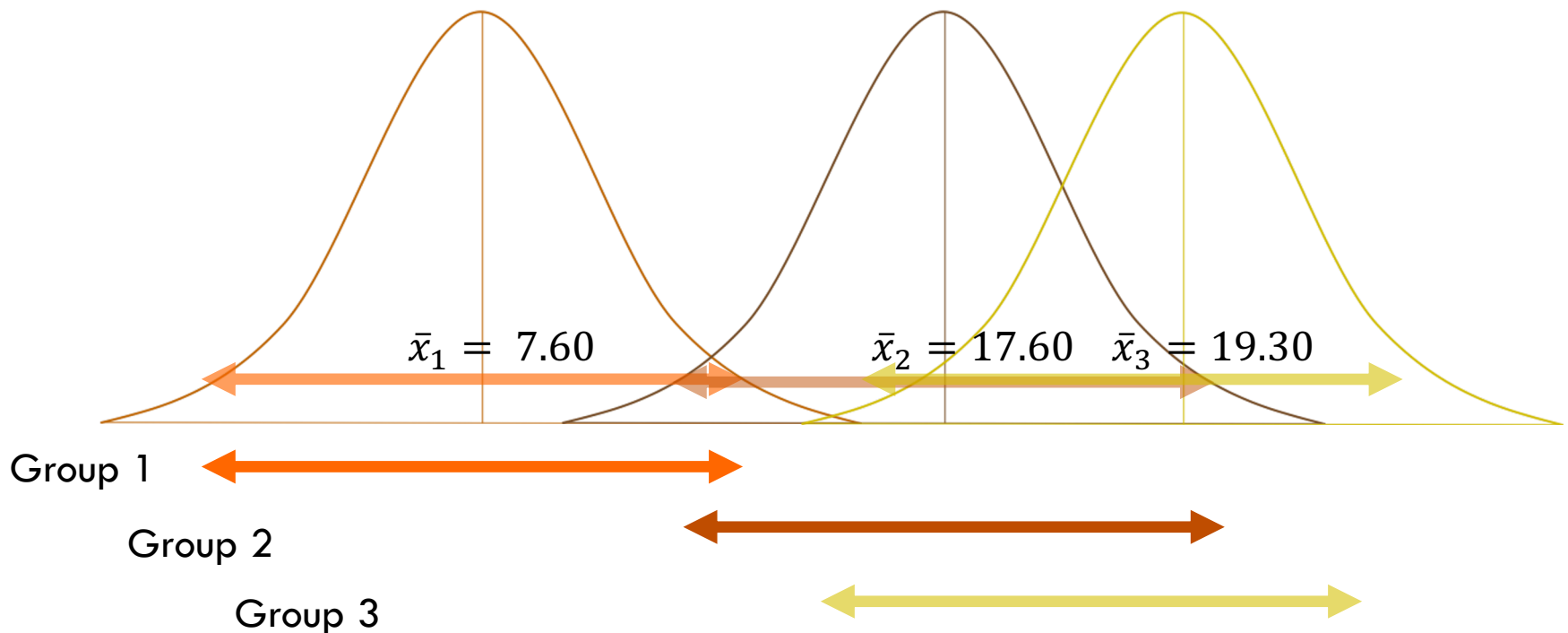
Group 2

Group 3

Variability

Which of the two scenarios below has more within group variability?
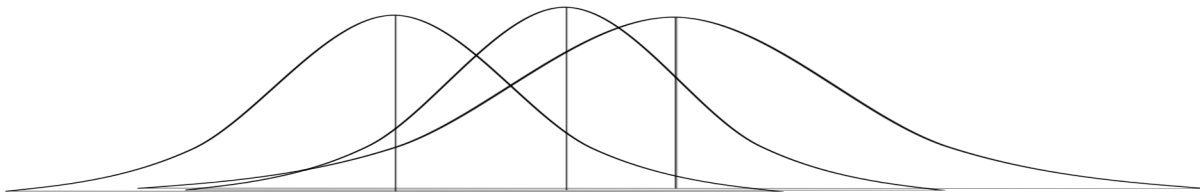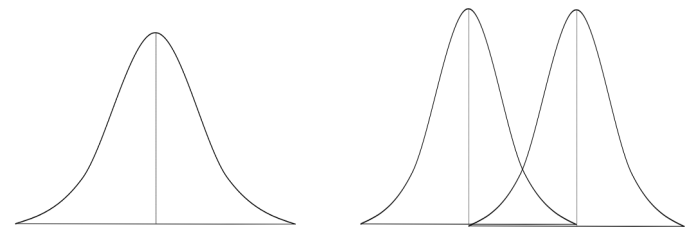
Scenario 1:
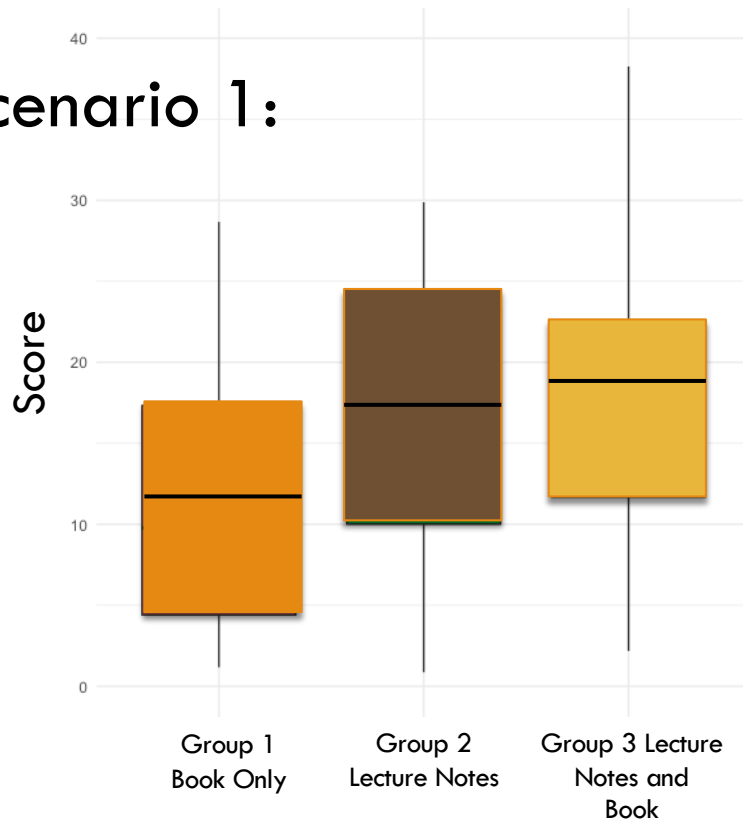
Scenario 2:

Think About It...

- If you were trying to figure out if there was any difference DUE TO GROUPING (the independent variable, treatment, intervention, etc.), in which scenario would that be easier?
  - Scenario 1 where WITHIN each group there is a lot of variability and a lot of overlap between the groups
  - Scenario 2 where WITHIN each group there is little variability and the groups are quite separate from each other

Scenario 1

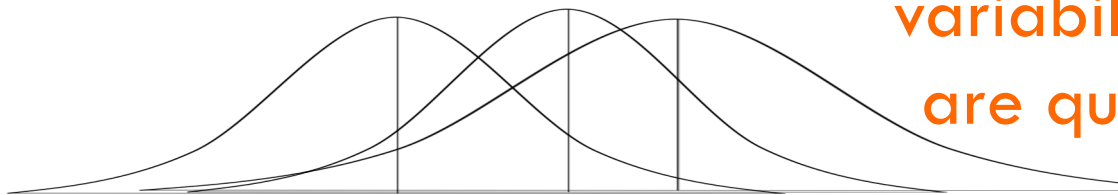Would it be easier to tell which group you were in in Scenario 1 or Scenario 2?
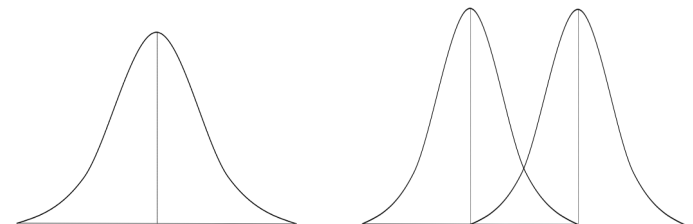
Scenario 2

Think About It…

- If you were trying to figure out if there was any difference DUE TO GROUPING (the independent variable, treatment, intervention, etc.), in which scenario would that be easier?

Scenario 2 where WITHIN each group there is little variability and the groups are quite separate from each other
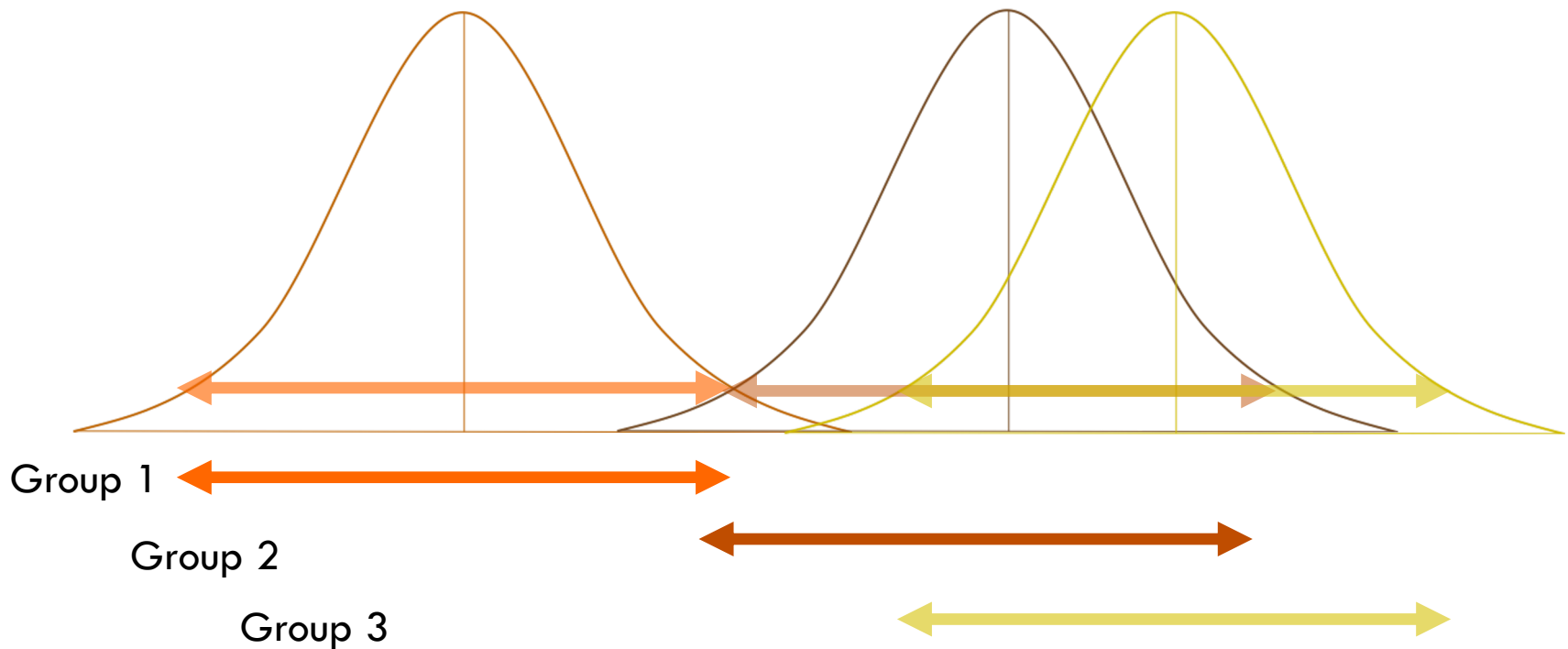
Scenario 1

Would it be easier to tell which group you were in in Scenario 1 or Scenario 2?

Scenario 2

Variance WITHIN Groups

- WITHIN group variance is the variability that is **NOT** due to the independent variable (not due to being in a particular group)
  - Just regular ole differences between people
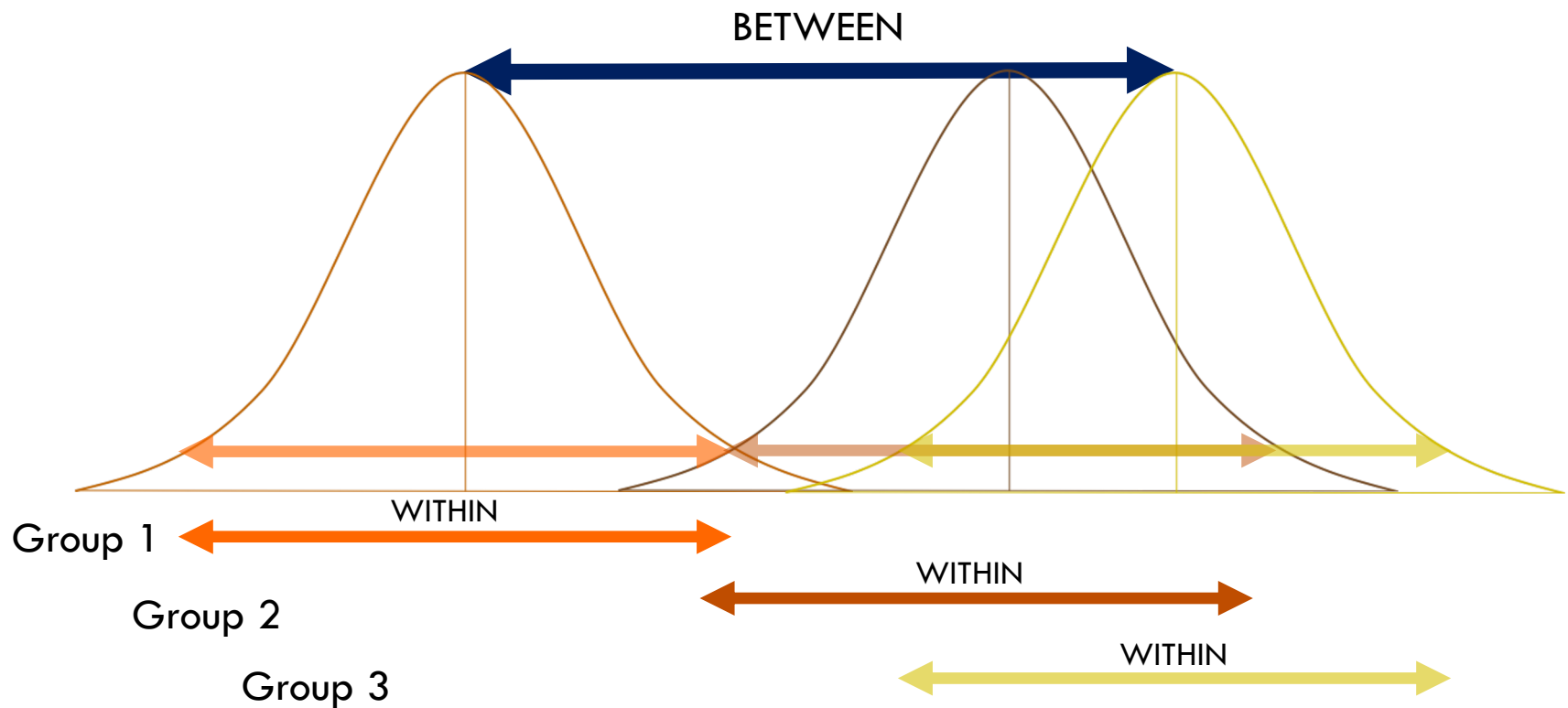 - Just people being different form each other

Group 1

Group 2

Group 3

More Sources of Variance

If WITHIN group variance is just the preexisting differences among people…

**What other type of variance are we interested in looking at?**

Variance BETWEEN Groups

- We can also compare the variance BETWEEN the group means

BETWEEN

WITHIN

Group 1

WITHIN

Group 2

WITHIN

Group 3

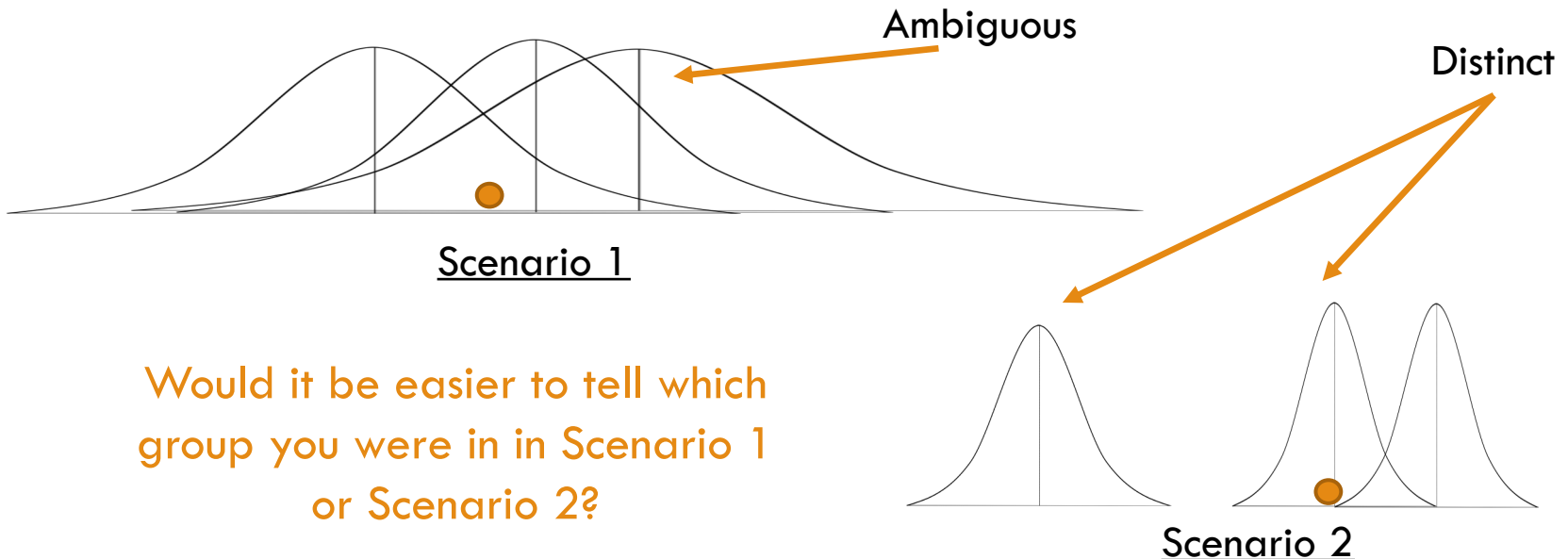# Variance BETWEEN Groups

- BETWEEN group variance is the good stuff!
  - How much did the groups vary from each other based on which group they were?
    - The grouping is our independent variable we are interested in.

BETWEEN

WITHIN

Group 1

WITHIN

Group 2
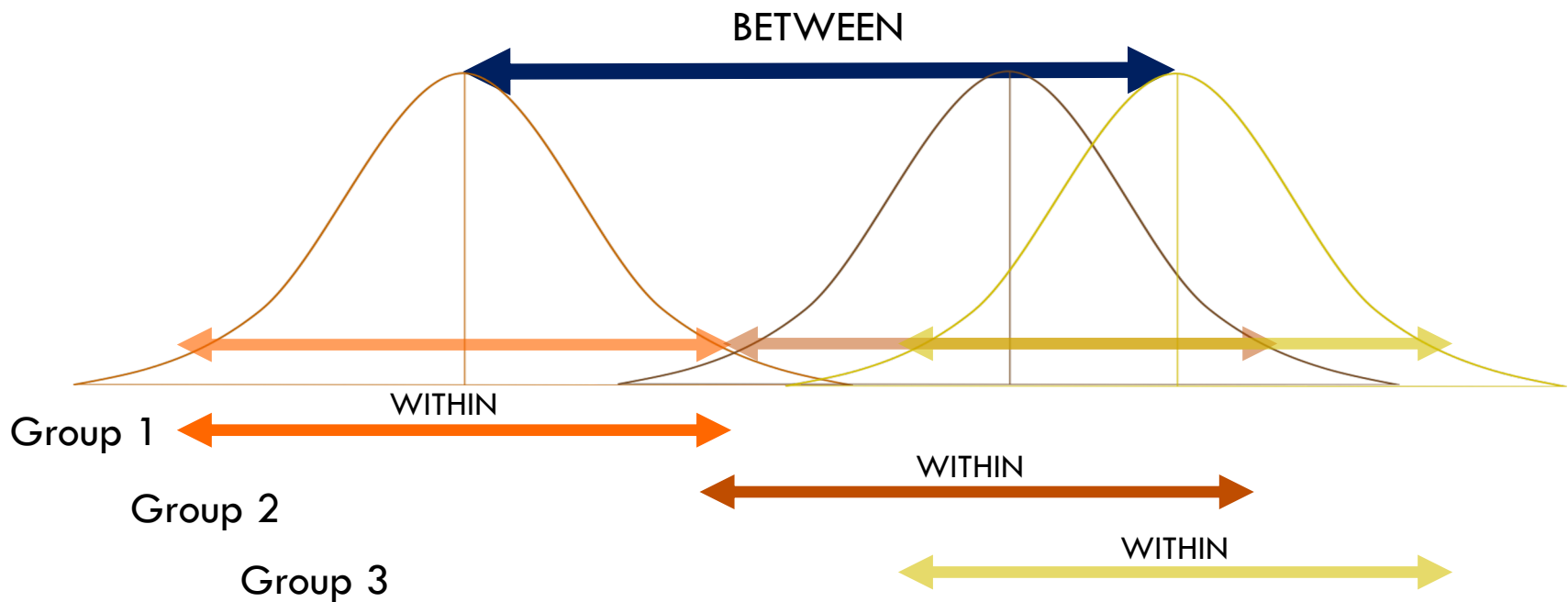
WITHIN

Group 3

Think About It… Again.

- If you were trying to figure out if there was any difference DUE TO GROUPING (the independent variable, treatment, intervention, etc.), in which scenario would that be easier?
  - Scenario 1 where WITHIN each group there is a lot of variability and a lot of overlap between the groups
  - Scenario 2 where WITHIN each group there is little variability and the groups are quite separate from each other

Ambiguous

Distinct

Scenario 1

Would it be easier to tell which group you were in in Scenario 1 or Scenario 2?

Scenario 2

Best Variance Case?

What is the optimal variance situation? Why?

1) High WITHIN and High BETWEEN

2) High WITHIN and Low BETWEEN

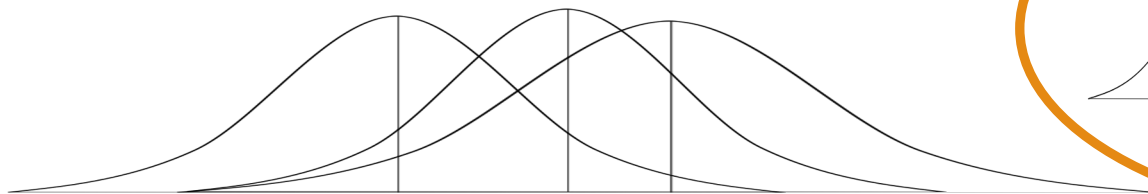3) Low WITHIN and Low BETWEEN

4) Low WITHIN and High BETWEEN

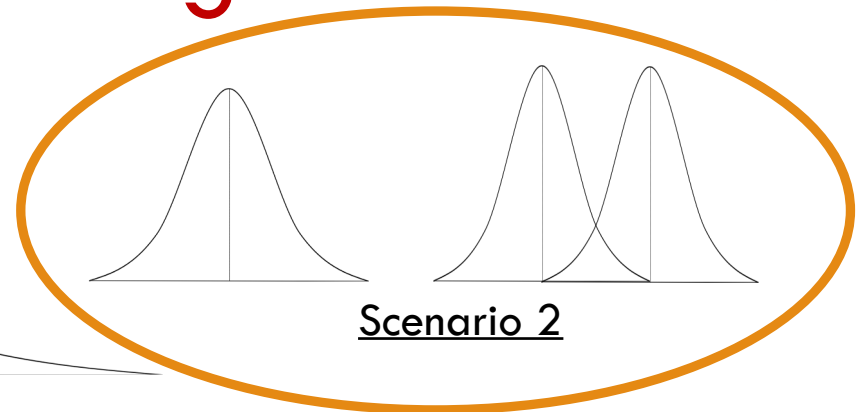# Best Variance Case?

What is the optimal variance situation? Why?

1) High WITHIN and High BETWEEN

2) High WITHIN and Low BETWEEN

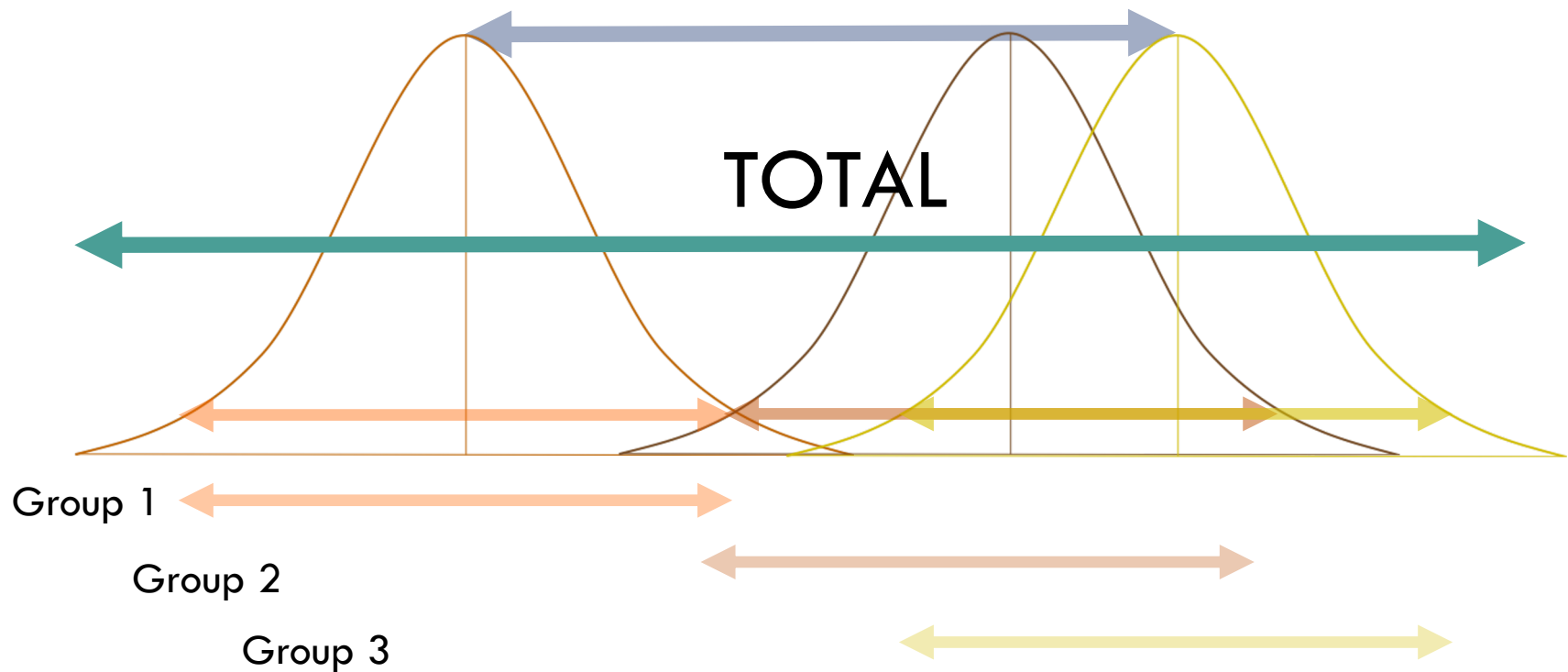3) Low WITHIN and Low BETWEEN

4) Low WITHIN and High BETWEEN

Scenario 1

Scenario 2

TOTAL Variance Among Groups

□ Lastly, we have TOTAL variance among all the groups, the teal color

TOTAL

Group 1

Group 2

Group 3

ANOVA and Ratios

Round Up the Variances

So what are our three types of variance in ANOVA?

Which one is related to Error?

Which one is related to the Independent Variable?

Sources of Variance

□ This is how the three types of variance work with each other

WITHIN		BETWEEN		TOTAL
Error Variance (Individual Differences)	+	Group Variance (Independent Variable Differences)	=	All Types of Variance (All Source of Difference)

Which of these sources of variance do I want to compare to determine if group assignment (the independent variable) has any effect?

Low Within, High Between Ratio

- What does a ratio mean again?
  - "relation between two amounts showing the number of times one value contains or is contained within the other"

$$Surplus \; or \; Debt = \frac{Your \; monthly \; income}{Your \; monthly \; bills}$$

If the ratio between your monthly income and your monthly bills = 1, what does that mean?

If the ratio between your monthly income and your monthly bills > 1, example (2.0) what does that mean?

If the ratio between your monthly income and your monthly bills < 1, example (.50) what does that mean?

Low Within, High Between Ratio

$$Surplus\ or\ Debt = \frac{Your\ monthly\ income}{Your\ monthly\ bills}$$

If the ratio between your monthly income and your monthly bills = 1, what does that mean?
You can exactly pay off your bill are not left with any extra money at the end of the month.

Monthly Income = Monthly Bills
Ratio = 1

If the ratio between your monthly income and your monthly bills > 1, example (2.0) what does that mean?
You can pay your bills and have some left over, surplus!

Monthly Income > Monthly Bills
Ratio > 1

If the ratio between your monthly income and your monthly bills < 1, example (.50) what does that mean?
You can't pay your bills and are going into debt...

Monthly Income < Monthly Bills
Ratio < 1

Low Within, High Between

Which situation would we hope to see if we were testing out new group interventions?

$$F = \frac{Between\ Group\ Variability}{Within\ Group\ Variability}$$

Situation 1

Between = Within

Ratio = 1

Situation 2

Between > Within

Ratio > 1

Situation 3

Between < Within

Ratio < 1

Low Within, High Between

Which situation would we hope to see if we were testing out new group interventions?

$$F = \frac{Between\ Group\ Variability}{Within\ Group\ Variability}$$

Situation 1

Between = Within

Ratio (F) = 1

Situation 2

Between > Within

Ratio (F) > 1

Situation 3

Between < Within

Ratio (F) < 1

Low Within, High Between

Which situation would we hope to see if we were testing out new group interventions?

$$F = \frac{Explained\ Variability}{Unexplained\ Variability}$$

Situation 1

Between = Within

Ratio (F) = 1

Situation 2

Between > Within

Ratio (F) > 1

Situation 3

Between < Within

Ratio (F) < 1

Next Up…

We will look at some ANOVA output and interpret it…