# Reminders

- Lab #1 due Tuesday by 12:30 pm – turn in to Moodle.
  - Please see **Missed Work Policy** in syllabus for how to formally request a 24-hour extension on lab assignments.

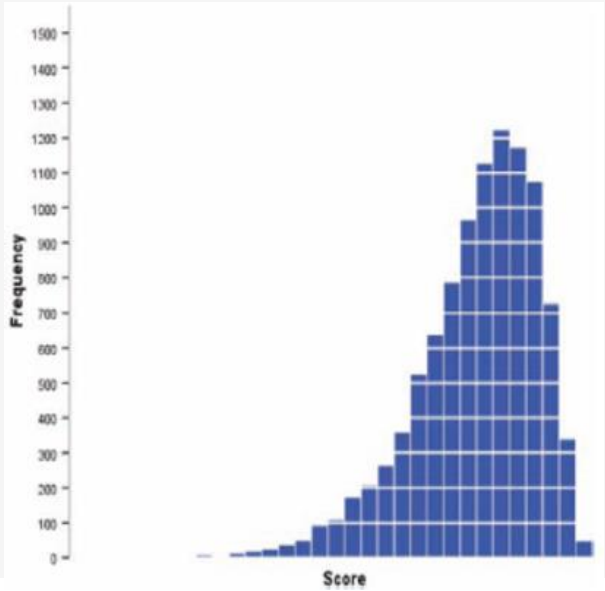
- Remember, office hours can be booked here: https://gvalenti.youcanbook.me/
- Or email me for an appointment (at least 24 hours ahead of time)

- Consider visiting our peer-tutors Mon - Fri: Hanna McNamara (our "embedded peer-tutor"), Alanna Gaines (who you met on Tuesday), Ada Weems or Mary Blake Zeron.

# Mode, Median, & Mean Critical Thinking Qs

1. Which is the best measure of central tendency for *qualitative* data?
2. Which is the only measure of central tendency (mode, median, or mean) that may not exist for a quantitative variable?
3. Which measure of central tendency is the only one that incorporates every single score into its calculation?

4. If the skew is negative (see diagram), which will be *higher*: the median or the mean?
5. Is the median or the mean more sensitive to extreme scores (i.e., outliers)?



# Mode, Median, and Mean Critical Thinking Qs

1. Which is the best measure of central tendency for *qualitative* data?

mode

**Ex: if measure *favorite color*, can't calculate a <u>median</u> because you cannot put the responses in order, and can't calculate <u>mean</u> because responses are not values**

2. Which is the only measure of central tendency (mode, median, or mean) that may not exist for a quantitative variable?

mode

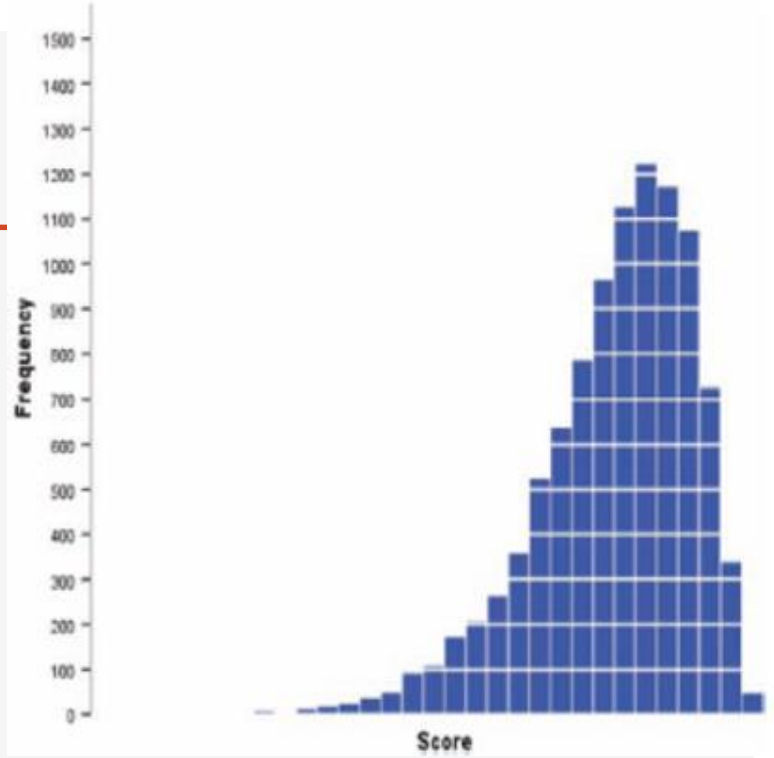
**Each score may occur an equal number of times → there is no mode.**

3. Which measure is the only one that incorporates every single score into its calculation?

mean

All of the x's are summed in the numerator of this equation.

# Mode, Median, and Mean Critical Thinking Qs & Answers

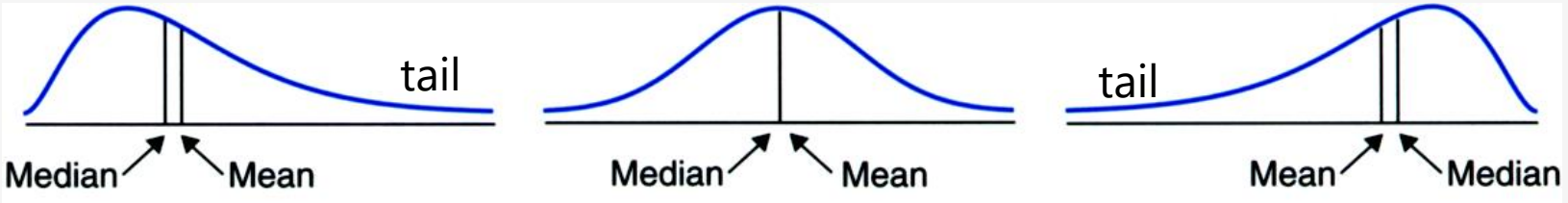



4. If the skew is negative (see diagram), which will be *higher:* the median or the mean? **median**

***SOME ADDITIONAL FACTS:***

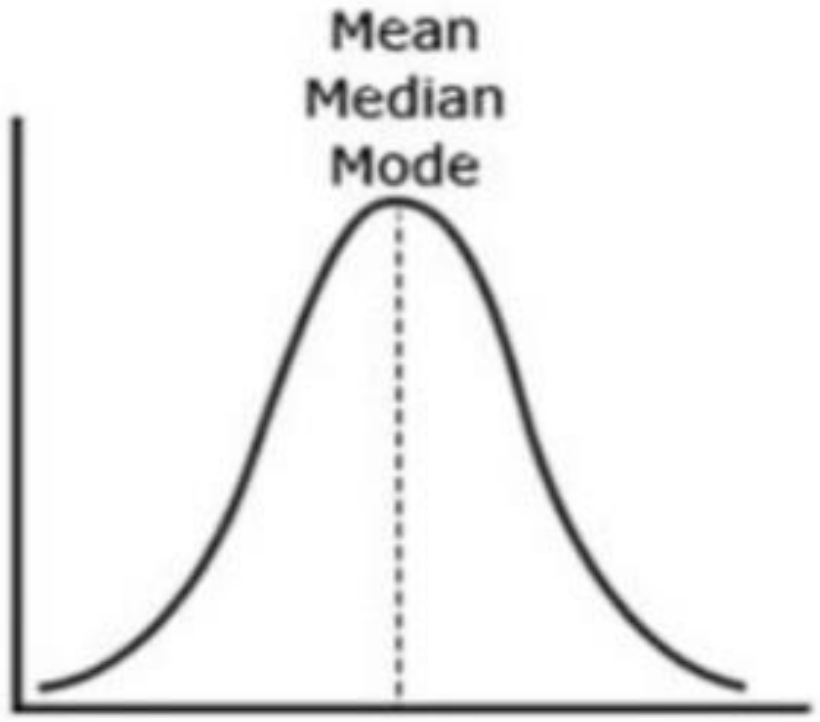
- ***In perfectly symmetrical distributions, the median and mean are identical***
- ***Means get pulled in direction of tail.***
  - ***Medians sometimes get pulled in that direction, too, but not as strongly.***

5. Is the median or the mean more sensitive to extreme scores (outliers)? **mean**




(a) Positive Skew (b) Symmetric (c) Negative Skew

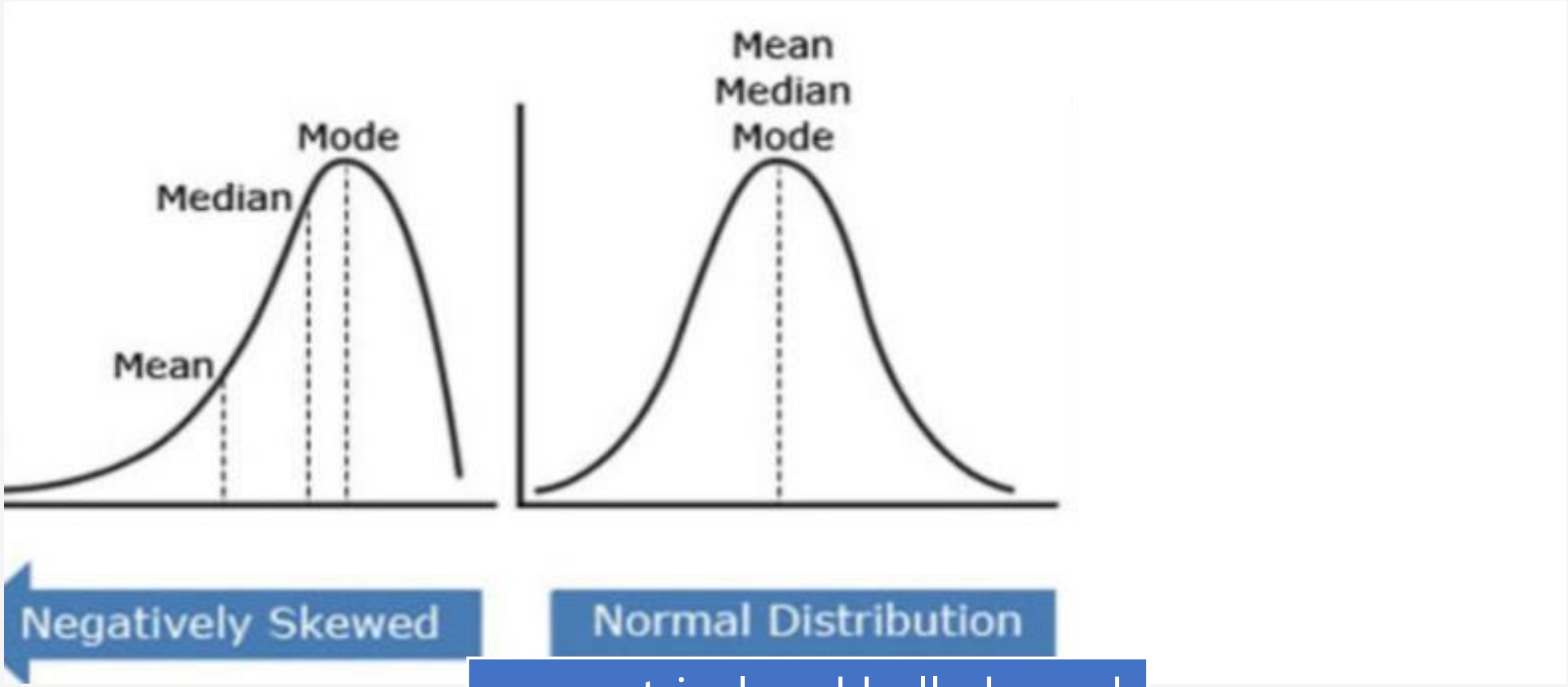
# Another image of the relationship between skew and the three measures of central tendency



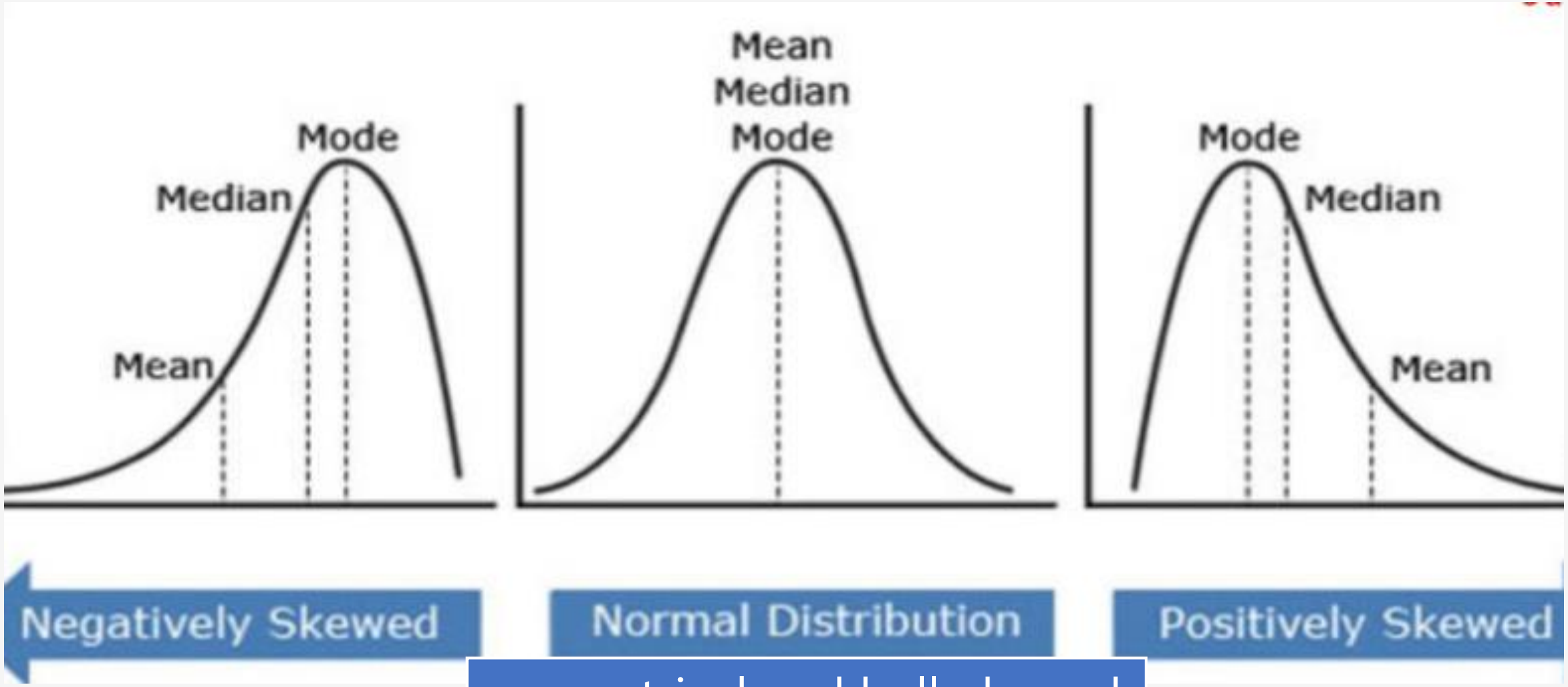

Normal Distribution

symmetrical and bell-shaped

# Another image of the relationship between skew and the three measures of central tendency



# Another image of the relationship between skew and the three measures of central tendency




symmetrical and bell-shaped

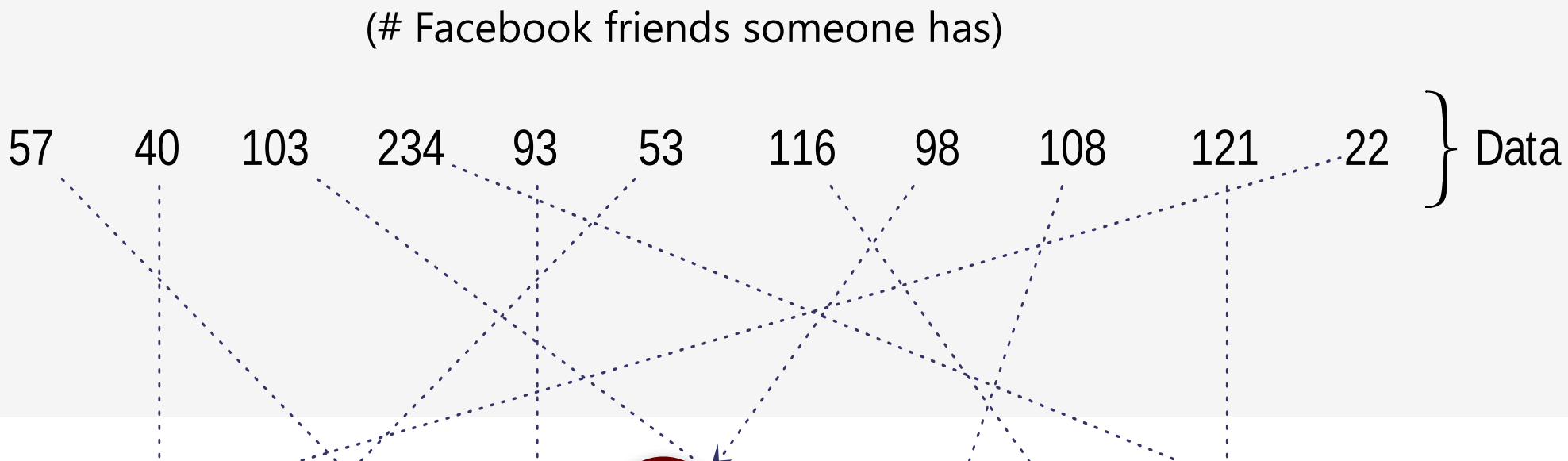
# Outline for Ch. 3

1. Review of frequency distributions
2. Measures of central tendency
3. **Measures of spread (aka, measures of dispersion, aka, measures of variability)**
4. Combining central tendency & spread

Range
Deviation
Sum of Squares (SS)
Variance $s^2$
Standard Deviation $s$

# Range (a measure of spread (*aka dispersion, variability*)

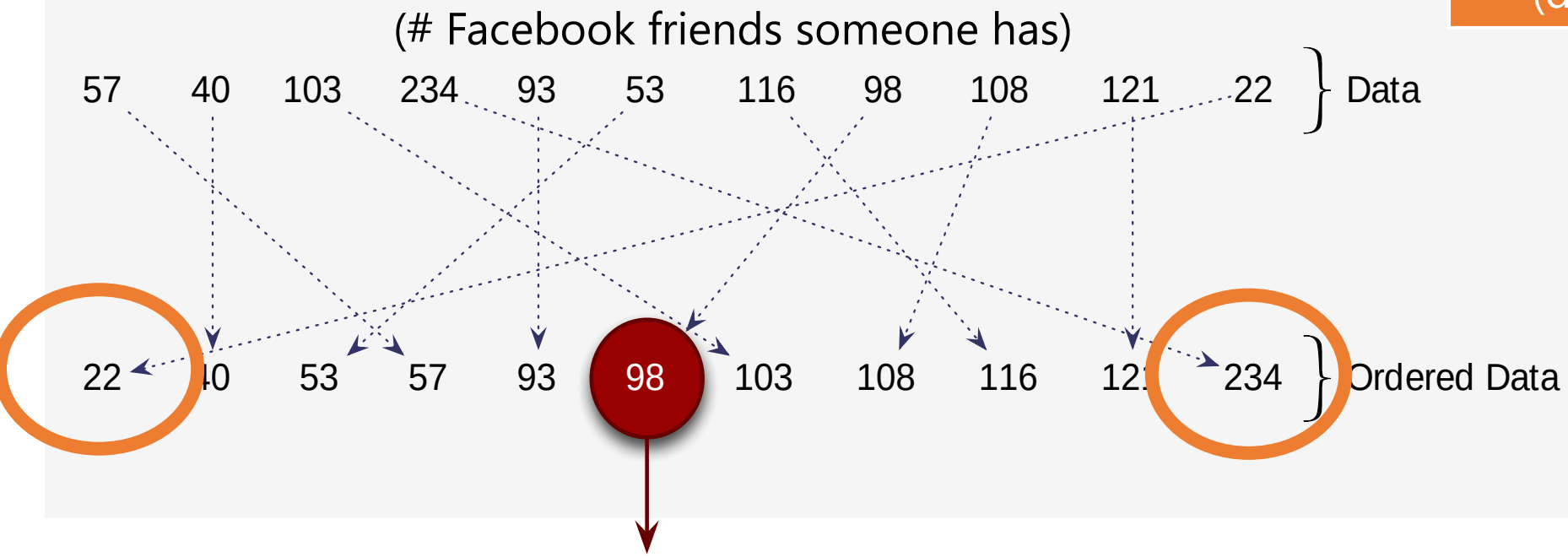
- The distance covered by the scores in a distribution
- To calculate, subtract the smallest score from the largest
- *Practice*: What is the **range** for this data set?



# Range (a measure of spread (*aka dispersion, variability*))

- The distance covered by the scores in a distribution
- To calculate, subtract the smallest score from the largest
- ***Weakness of range*: only uses 2 scores**
- Practice: What is the **range** for this data set?

**Range** = 234 - 22 = **212**
**Range = 212 FB friends**
(don't forget the units!)



# Deviation (another measure of spread)

- how different a given score is from the center of a distribution (i.e., from *the mean*)
  - Unlike with the range, you can calculate a deviation for each individual score in the data set.

$$\text{deviation} = x_i - \bar{x}$$

$x_i$ refers to "the score of the $i^{th}$ person"

$\bar{x}$ refers to "the mean (average) score in the sample"

# Deviation (another measure of spread)

- how different a given score is from the center of a distribution (i.e., from *the mean*)

  - Unlike with the range, you can calculate a deviation for each individual score in the data set.

$$\text{deviation} = x_i - \bar{x}$$

$\bar{X}$ = 95 friends in this sample

(# Facebook friends someone has)

57 40 103 234 93 53 116 98 108 121 22 } Data

| Score ( $x_i$ ) |
| --- |
| 22 |
| 40 |
| 53 |
| 57 |
| Etc. |

# Deviation (another measure of spread)

$$\text{deviation} = x_i - \bar{x}$$

- how different a given score is from the center of a distribution (i.e., from *the mean*)
  - You can calculate a deviation for each individual score in the data set

$\bar{X}$ = 95 friends in this sample

| Score ( $x_i$ ) | Mean ( $\bar{X}$ ) |
|---|---|
| 22 | 95 |
| 40 | 95 |
| 53 | 95 |
| 57 | 95 |
| Etc. | Etc. |

(# Facebook friends someone has)

57 40 103 234 93 53 116 98 108 121 22 } Data

# Deviation (another measure of spread)

$$\text{deviation} = x_i - \bar{x}$$

- how different a given score is from the center of a distribution (i.e., from *the mean*)
  - You can calculate a deviation for each individual score in the data set

$\bar{X}$ = 95 friends in this sample

| Score ( $x_i$ ) | Mean (x-bar) | deviation $(x_i - \bar{x})$ |
|---|---|---|
| 22 | 95 | |
| 40 | 95 | |
| 53 | 95 | |
| 57 | 95 | |
| Etc. | Etc. | |

(# Facebook friends someone has)

57 40 103 234 93 53 116 98 108 121 22 } Data

# Deviation (another measure of spread)

$$\text{deviation} = x_i - \bar{x}$$

- how different a given score is from the center of a distribution (i.e., from *the mean*)
  - You can calculate a deviation for each individual score in the data set
- In practice, researchers rarely use individual deviations.
- *It'd be better to have one value to represent how spread out your scores are. So researchers calculate the* ***standard deviation***.

| Score ( $x_i$ ) | Mean ( $\bar{X}$ ) | deviation $(x_i - \bar{x})$ |
|---|---|---|
| 22 | 95 | -73 |
| 40 | 95 | -55 |
| 53 | 95 | -42 |
| 57 | 95 | -38 |
| Etc. | Etc. | Etc. |

(# Facebook friends someone has)

57 40 103 234 93 53 116 98 108 121 22 } Data

# How do we calculate the standard deviation? Well, it's a process . . .

Using the *deviations*, we next calculate the

**Sum of Squared Deviations**,

*aka* the **Sum of Squares (SS)**

- The SS is technically another measure of spread

- To calculate SS, *square* each deviation, and then *sum* these squared deviations

| Score ( $x_i$ ) | Mean (x-bar) | deviation $(x_i - \bar{x})$ |
|---|---|---|
| 22 | 95 | -73 |
| 40 | 95 | -55 |
| 53 | 95 | -42 |
| 57 | 95 | -38 |
| Etc. | Etc. | Etc. |
| | | SS = |

# How do we calculate the standard deviation? Well, it's a process . . .

- To calculate SS, *square* each deviation, and then *sum* these squared deviations

$$\text{sum of squared deviations (SS)} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

SS = 32,246 FB friends-squared

**VARIANCE TABLE**

| Score ( $x_i$ ) | Mean (x-bar) | deviation $(x_i - \bar{x})$ | deviation$^2$ $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 22 | 95 | -73 | *(-73*-73) 5329* |
| 40 | 95 | -55 | *3025* |
| 53 | 95 | -42 | *1764* |
| 57 | 95 | -38 | *1444* |
| Etc. | Etc. | Etc. | Etc. |
| | | SS = | **32,246** |

# How do we calculate the standard deviation? Well, it's a process . . .

- Using the SS, we next calculate the **variance** ( $s^2$ ) - another measure of spread

- To calculate the variance, divide SS by N-1, where N = the # of scores

$$\text{variance}\left(s^2\right) = \qquad \text{FB friends-squared}$$

In other words, the average variability (spread) in our data is 3224.6 FB-friends squared

This dog is thinking: *What the heck is a "Facebook friend-squared"? That unit makes no sense!*

# Standard Deviation ($s$ ) – the G.O.A.T. measure of spread!!!

- The variance ($s^2$) gives us a measure of spread that's in *units squared*, which is not an intuitive unit of measurement.
- To solve this problem, take the square root of the variance, to get a statistic known as the sample **standard deviation (*s*)**.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1}}$$

$$\text{variance}\left(s^2\right) = \frac{\text{SS}}{N-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1} = \frac{32,246}{10} = 3224.6 \text{ FB friends squared}$$

# Standard Deviation ($s$ ) – the G.O.A.T. measure of spread!!!

- The variance ($s^2$) gives us a measure of spread that's in *units squared*, which is not an intuitive unit of measurement.
- To solve this problem, take the square root of the variance, to get a statistic known as the sample **standard deviation ($s$)**.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1}}$$

$$= \sqrt{3224.6}$$

$$s = 56.79 \text{ Facebook friends}$$

$$\text{variance}\left(s^2\right) = \frac{\text{SS}}{N-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1} = \frac{32,246}{10} = 3224.6 \text{ FB friends squared}$$

The standard deviation for the Facebook friends variable is 56.79. In other words, on average, each participant's number of FB friends differs from the mean number of FB friends by about 57 friends.

# SUMMARY

## Process of calculating standard deviation (i.e., *s* )

**DEVIATION**

**1. Find deviation for each score**

**SS**

**2. Square each deviation & then sum**

**VARIANCE ($S^2$)**

**3. Find "average" of squared deviations**

**(divide by N-1)**

**STANDARD DEVIATION (*S*)**

**4. Take square root**

Please complete the exercise on your handout.

# Outline for Ch. 3

1. Review of frequency distributions
2. Measures of central tendency
3. Measures of spread
4. **Combining central tendency & spread**

Measures of ***Central Tendency***
Mode
Median
Mean

Measures of ***Spread***
Range
Deviation
Sum of Squares (SS)
Variance $s^2$
**Standard Deviation *s***

# More on the Standard Deviation ( *s* )

What is it, exactly?

The standard deviation…

- provides a measure of the *standard* (*typical*, *average*) distance of individual scores from _____

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N - 1}}$$

# More on the Standard Deviation ( *s* )

What is it, exactly?

The standard deviation…

- provides a measure of the *standard* (*typical*, *average*) distance of individual scores from **the sample mean**
- describes whether the scores in a sample of scores are *clustered closely* around the mean or *widely scattered* from the mean
  - **Smaller** *s* indicates: people's scores are **tightly packed** near the mean score
  - **Larger** *s* indicates: people's scores are **more spread** from the mean score

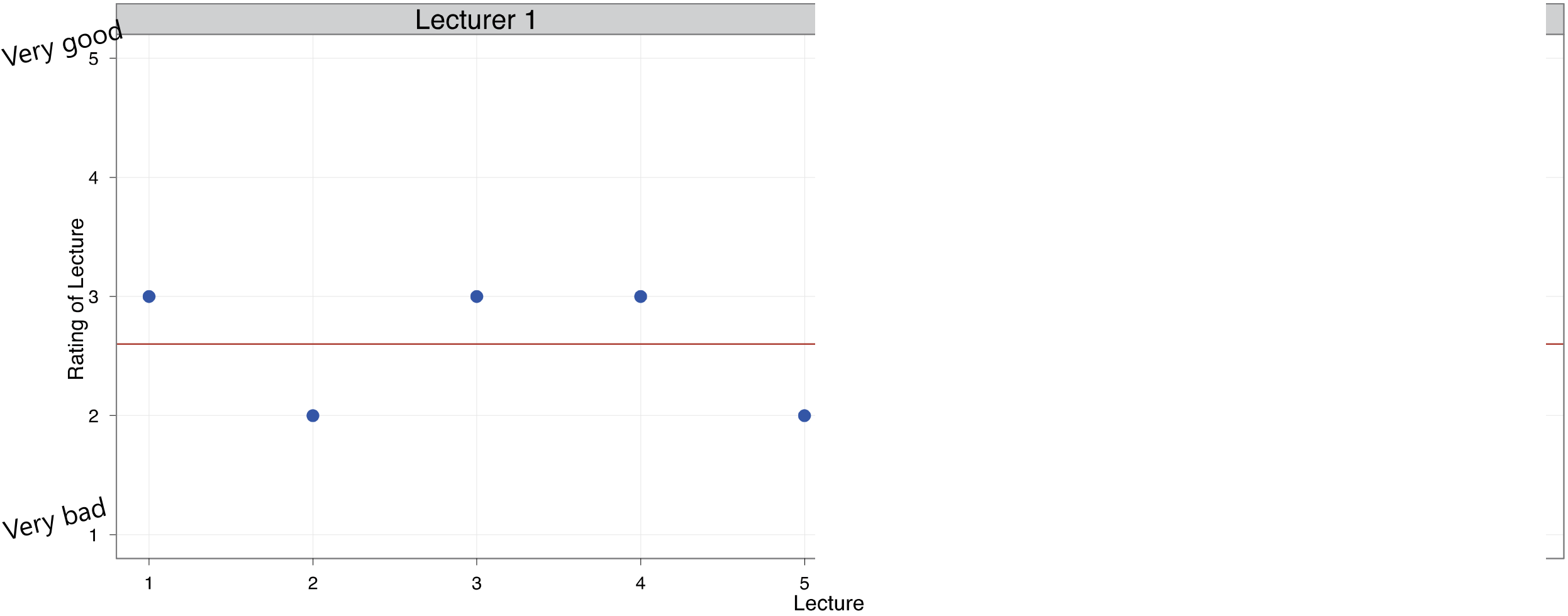
# Samples w/equal *mean*, and unequal *standard deviations*

What if all individual scores in a sample were 50 – what would the ***S*** be? zero

| Sample 1 scores: | | Sample 2 scores: | | Sample 3 scores: | |
|---|---|---|---|---|---|
| 10 | | 30 | | 10 | |
| 30 | Mean = 50 | 45 | Mean = 50 | 10 | Mean = 50 |
| 40 | | 50 | | 15 | |
| 50 | **$S$ = 28.28** | 51 | **$S$ = 10.60** | 40 | **$S$ = 41.93** |
| 60 | | 54 | | 70 | |
| 60 | | 56 | | 95 | |
| 100 | | 64 | | 110 | |

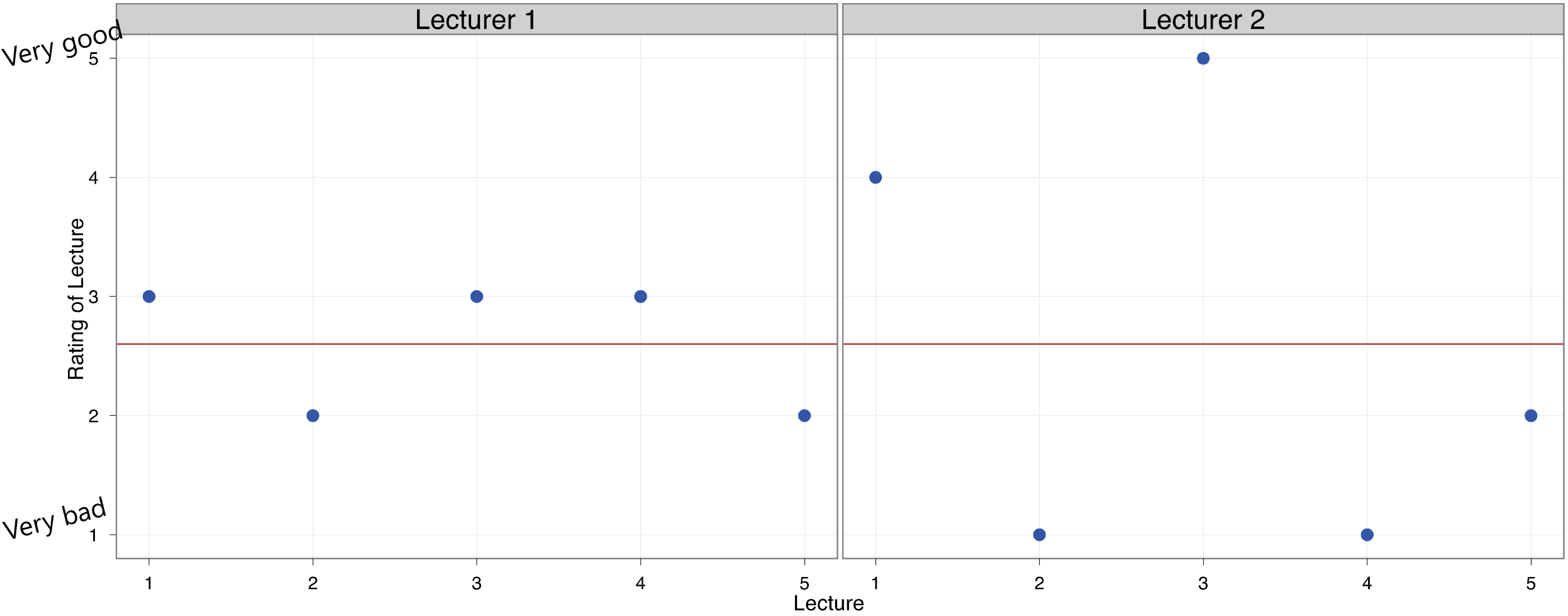
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of individual scores from the mean

# Equal Mean, Unequal Standard Deviations



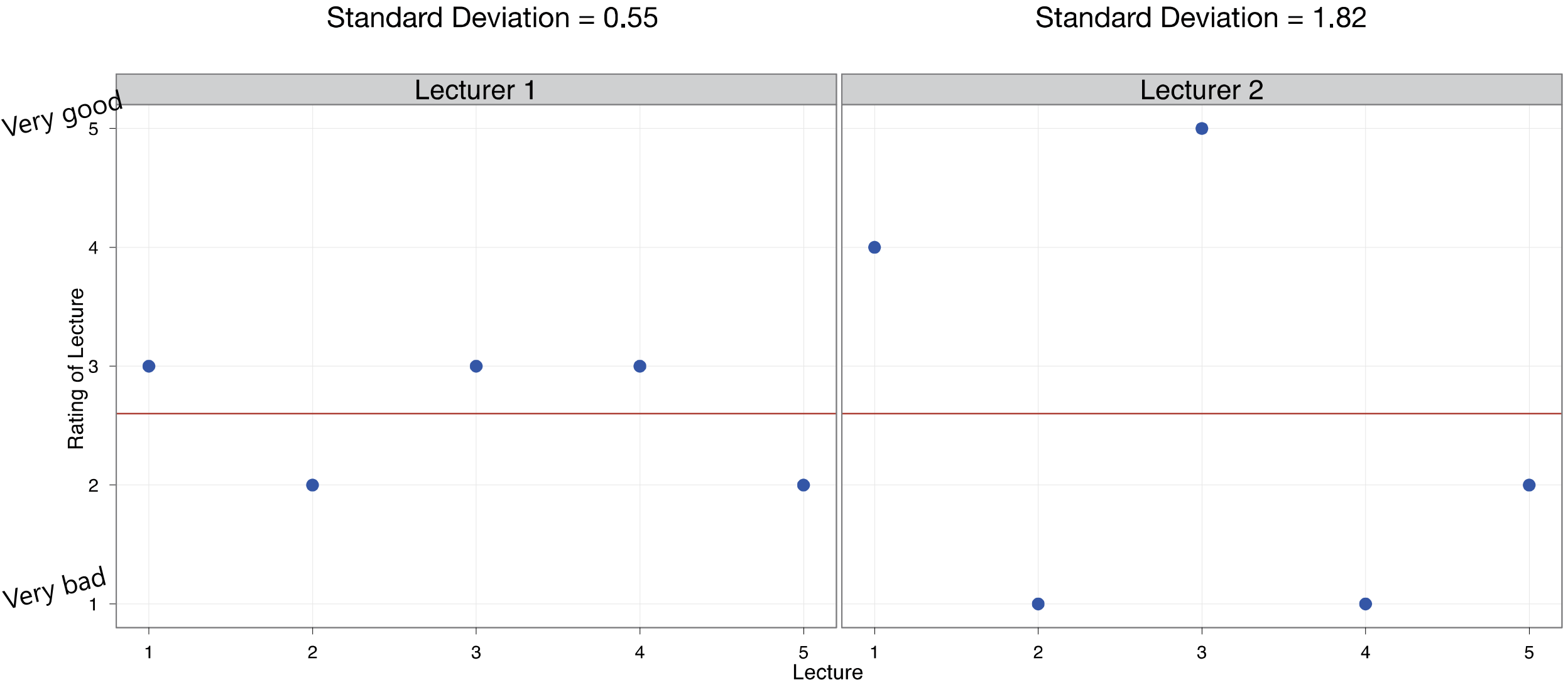

REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

# Equal Mean, Unequal Standard Deviations




REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

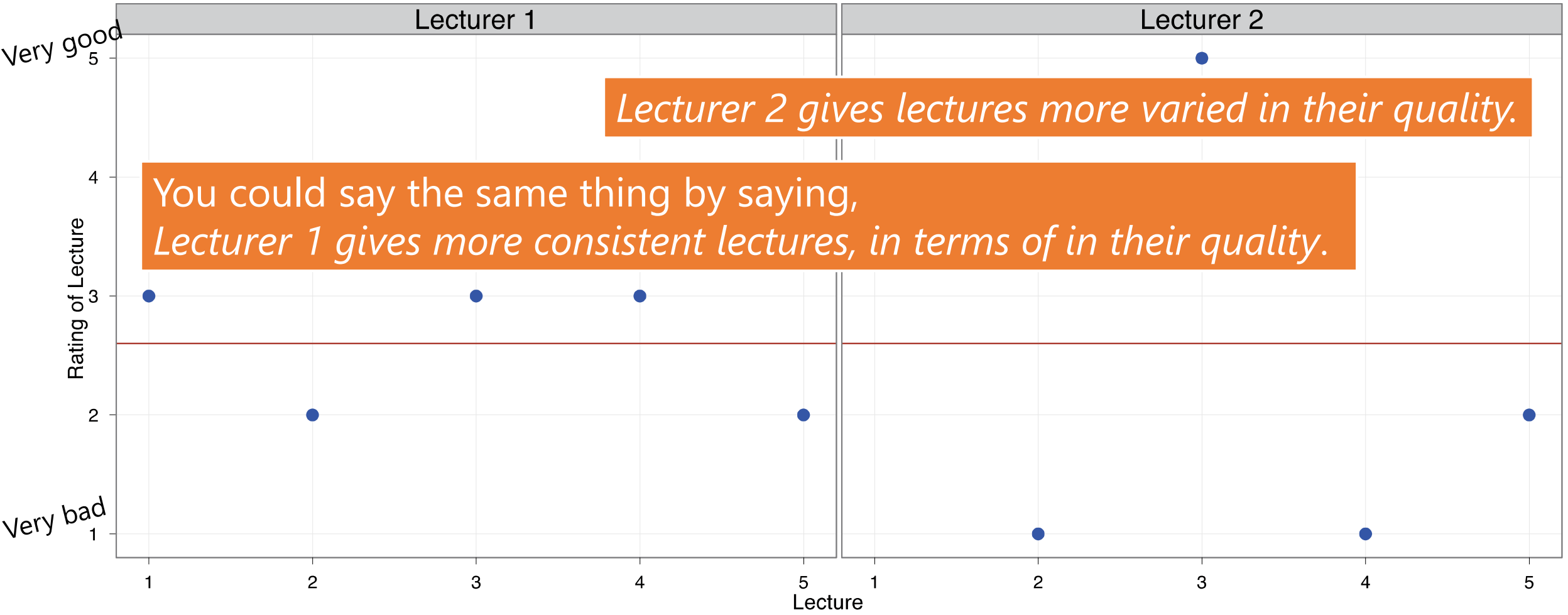
# Equal Mean, Unequal Standard Deviations




REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

# Equal Mean, Unequal Standard Deviations

While the quality of the two lecturers' lectures is the same on average....




*Lecturer 2 gives lectures more varied in their quality.*

You could say the same thing by saying,
*Lecturer 1 gives more consistent lectures, in terms of in their quality.*

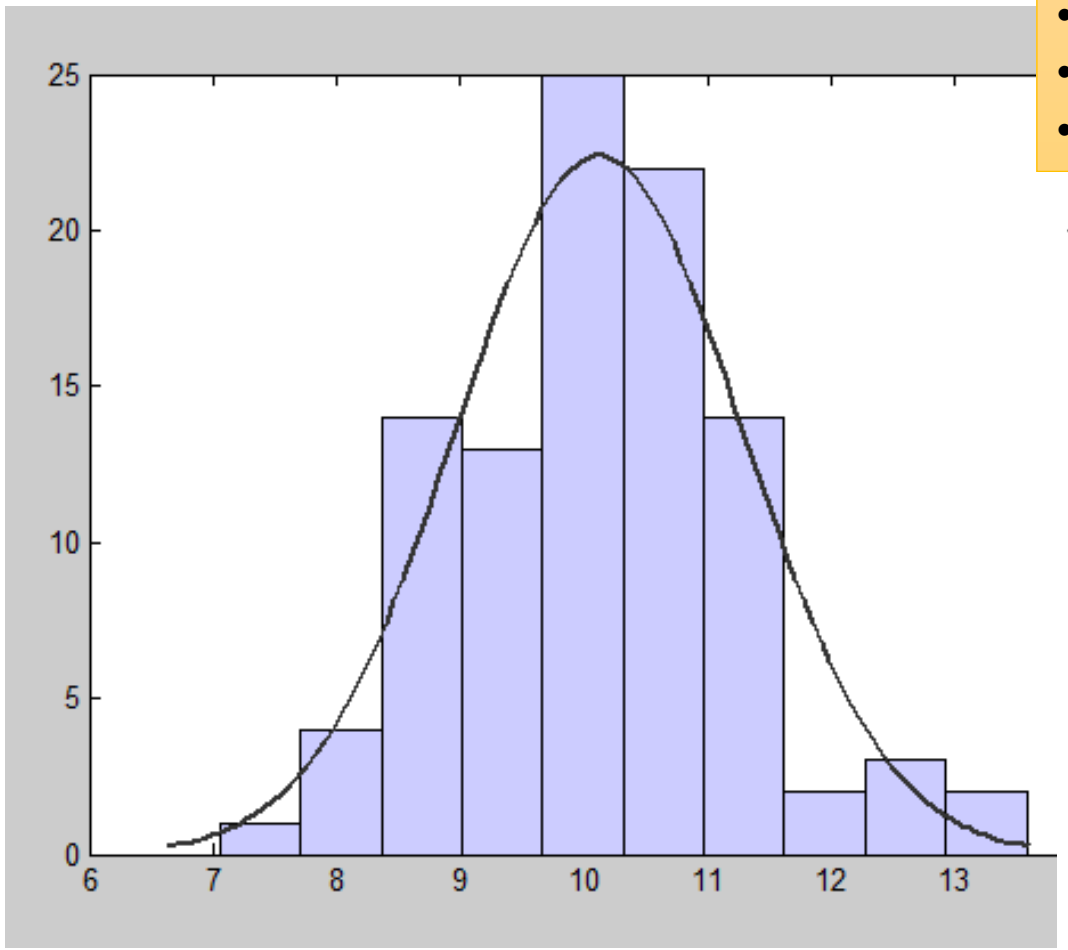
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

# Frequency distributions are back!

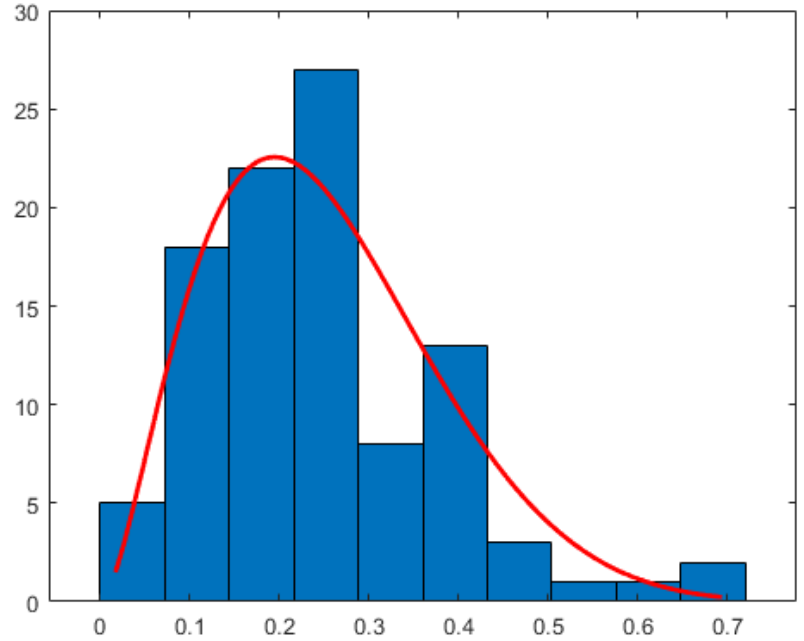
Remember, by "score," I just mean Ps' responses, what researcher measured.
EX:

- # of Instagram followers
- Scale values 1-disagree to 7-agree
- Your score on an exam (0-100%)
- # of hours of sleep you got last night

Frequency
(# of Ps with each score)
or
Percentage
(% of Ps with each score)

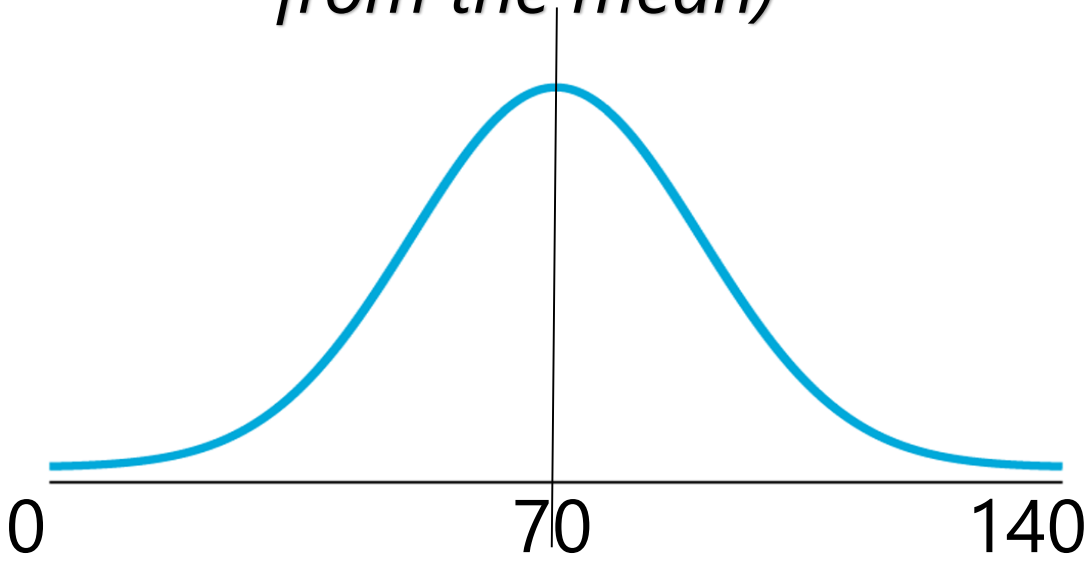



score



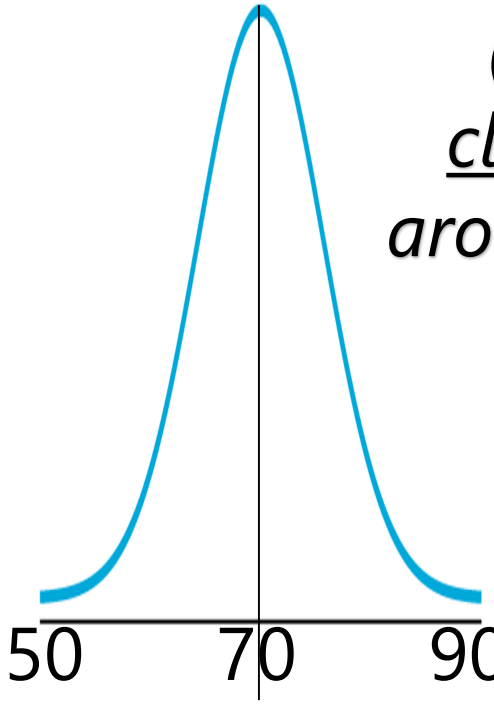
# Equal Mean, Unequal Standard Deviations

**larger**
standard deviation
*(data are more spread out from the mean)*



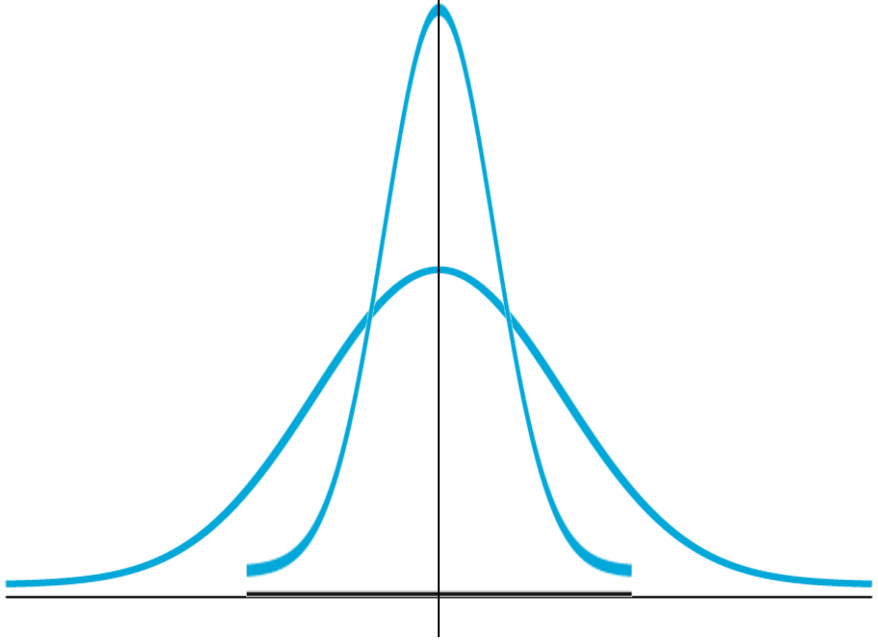

Mean score = 70

**smaller**
standard deviation
*(data are more closely clustered around the mean)*

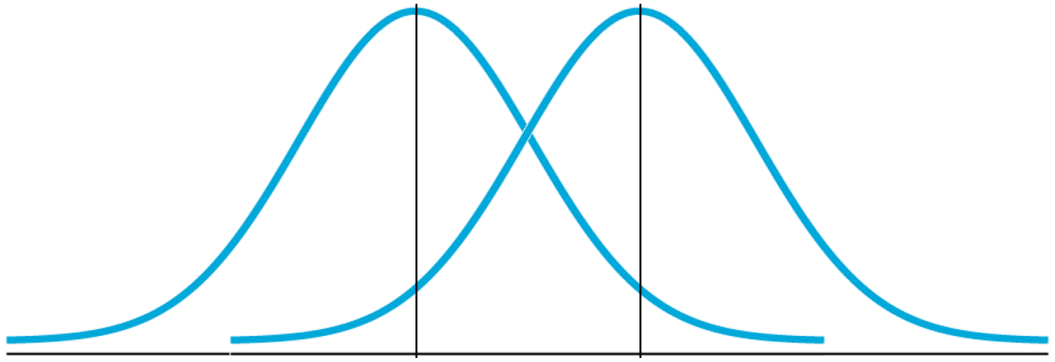



Mean score = 70

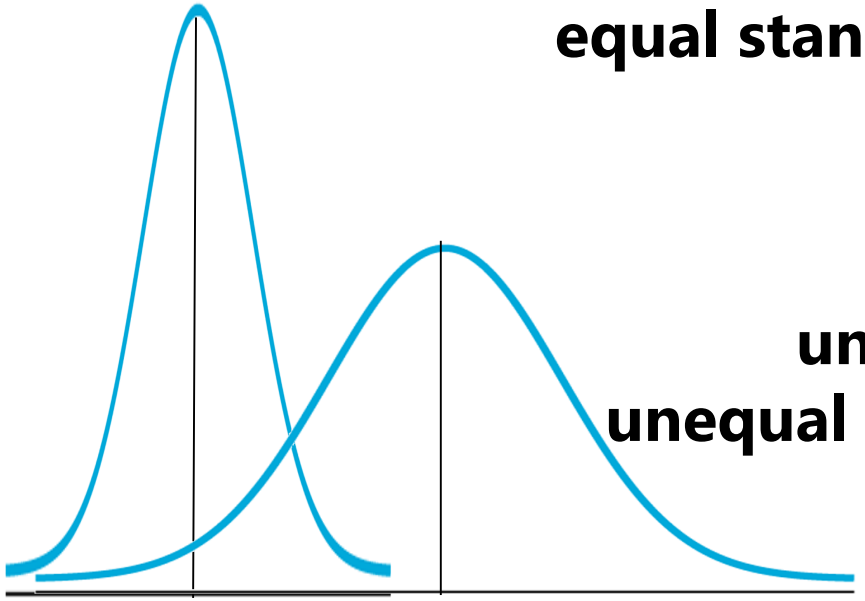
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of individual scores from the mean



**equal means,**
**unequal standard deviations**



**unequal means,**
**equal standard deviations**



**unequal means,**
**unequal standard deviations**

REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of individual scores from the mean