Deviation (a measure of spread)

- 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.

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

(# 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 (a 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.**

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.

(# Facebook friends someone has)

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

Sum of Squared Deviations, *aka* Sum of Squares (SS)

- 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 =

Sum of Squared Deviations, *aka* Sum of Squares (SS)

- 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**

Variance (s^2)

- Because it is a *sum*, the SS (sum of squared deviations) gets larger as sample size gets larger

  → can't use SS to compare the variability of samples of different sizes

- More useful to work with a measure of "*average*" spread, known as the sample **variance**
  - To calculate the variance, divide SS by N-1

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

FB friends-squared

Standard Deviation (s)

- The variance (s^2) gives us a measure of spread that's in *units squared*.
  - E.g., we'd have to say that the average variability (spread) in our data was 3224.6 *FB-friends squared*…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*)**.

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

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

$$= \sqrt{3224.6}$$

$$= 56.79 \text{ FB friends}$$

Standard Deviation (s)

- The variance (s^2) gives us a measure of spread that's in *units squared*.
  - E.g., we'd have to say that the average variability (spread) in our data was 3224.6 *FB-friends squared*…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*)**.

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

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

$$= \sqrt{3224.6}$$

The standard deviation, $s = 56.79$ Facebook 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

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**

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/population of scores are *clustered closely* around the mean or *widely scattered* from the mean
  - **Small** *s* indicates: people's scores are **tightly packed** near the mean score
  - **Large** *s* indicates: people's scores are **more spread** from the mean score

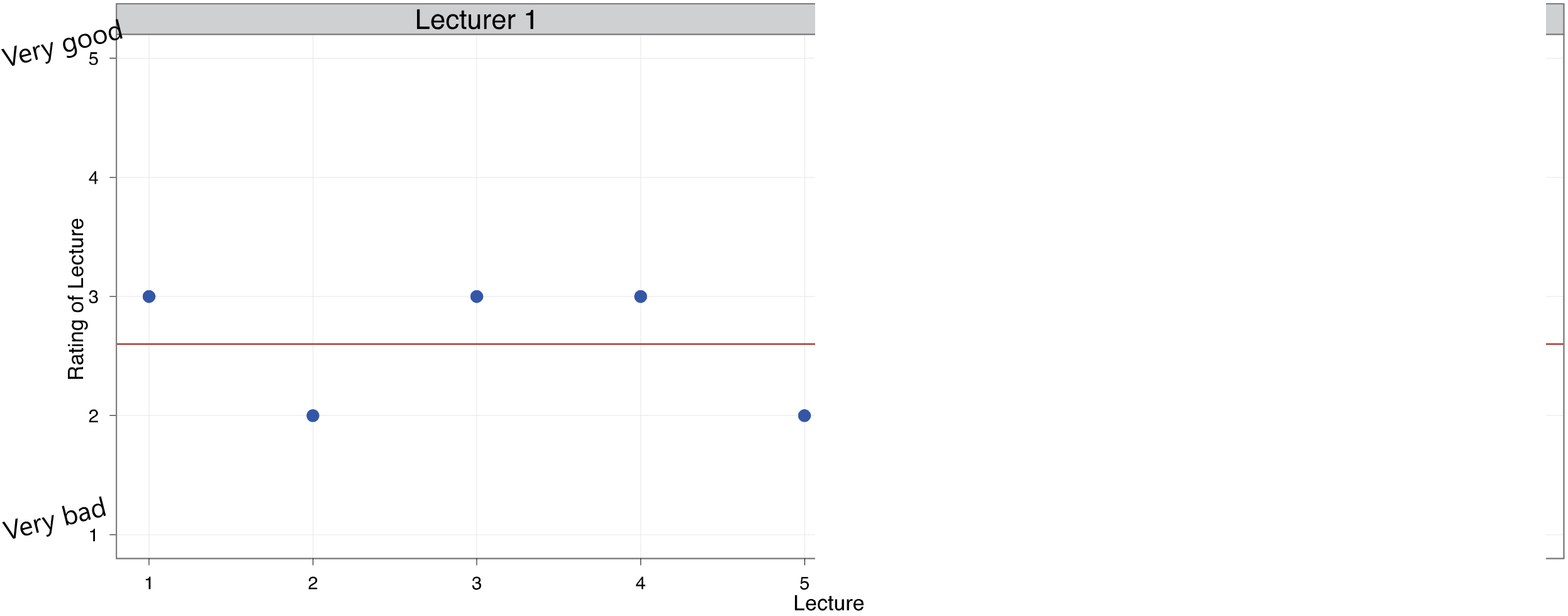
Samples w/equal *mean*, and unequal *standard deviations*

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	

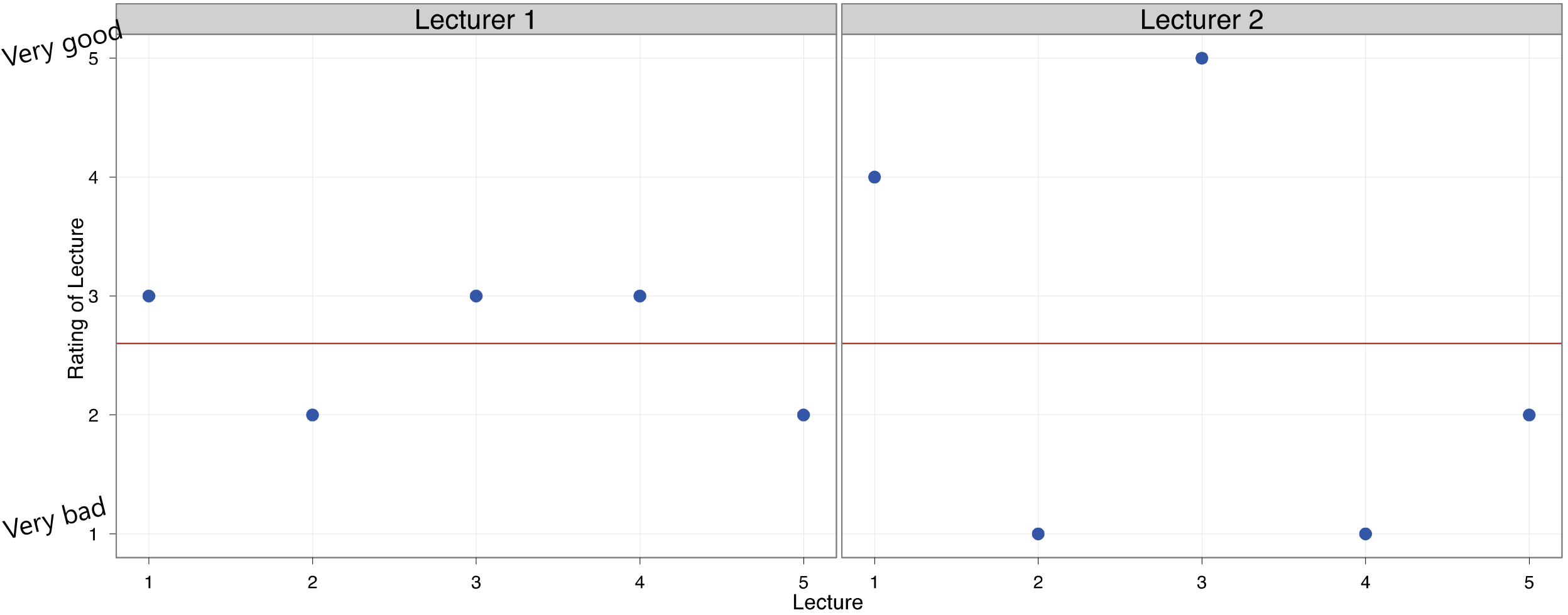
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of individual scores from the mean

Equal Mean, Unequal Standard Deviations



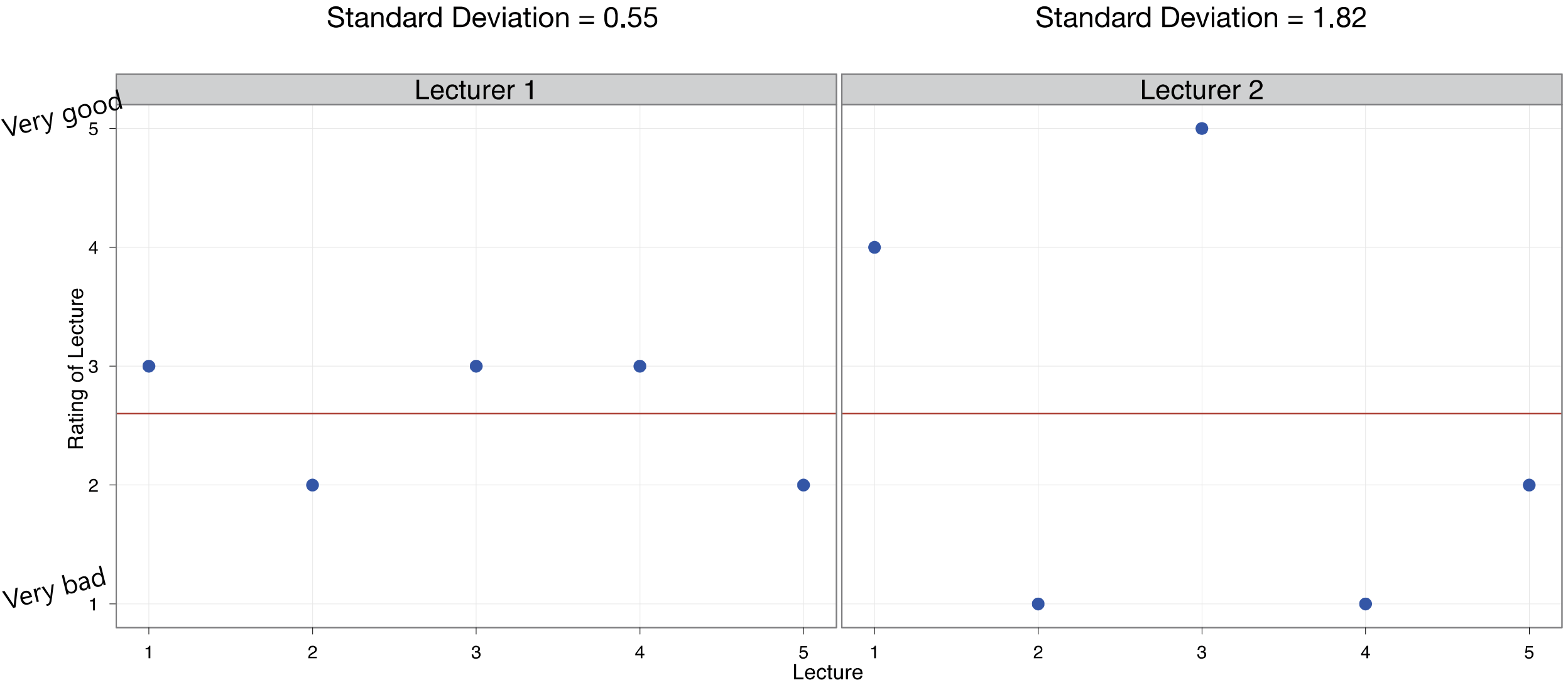
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

Equal Mean, Unequal Standard Deviations



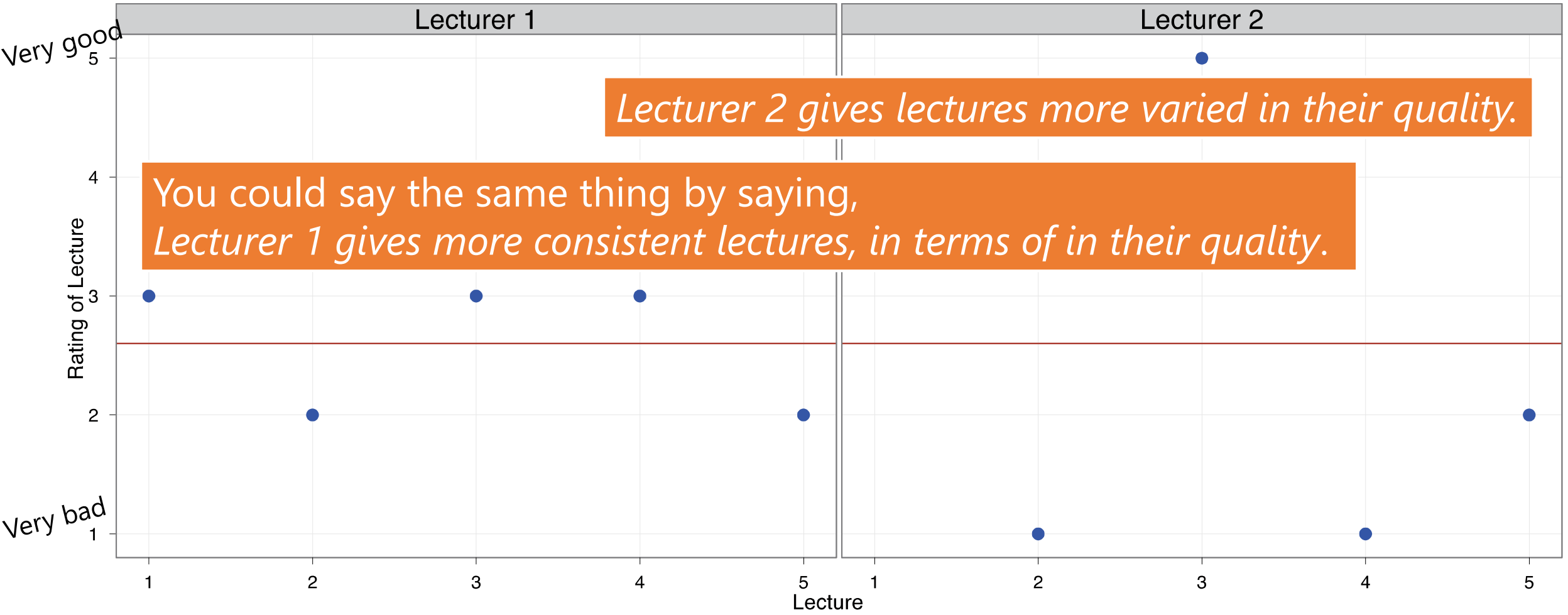
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

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



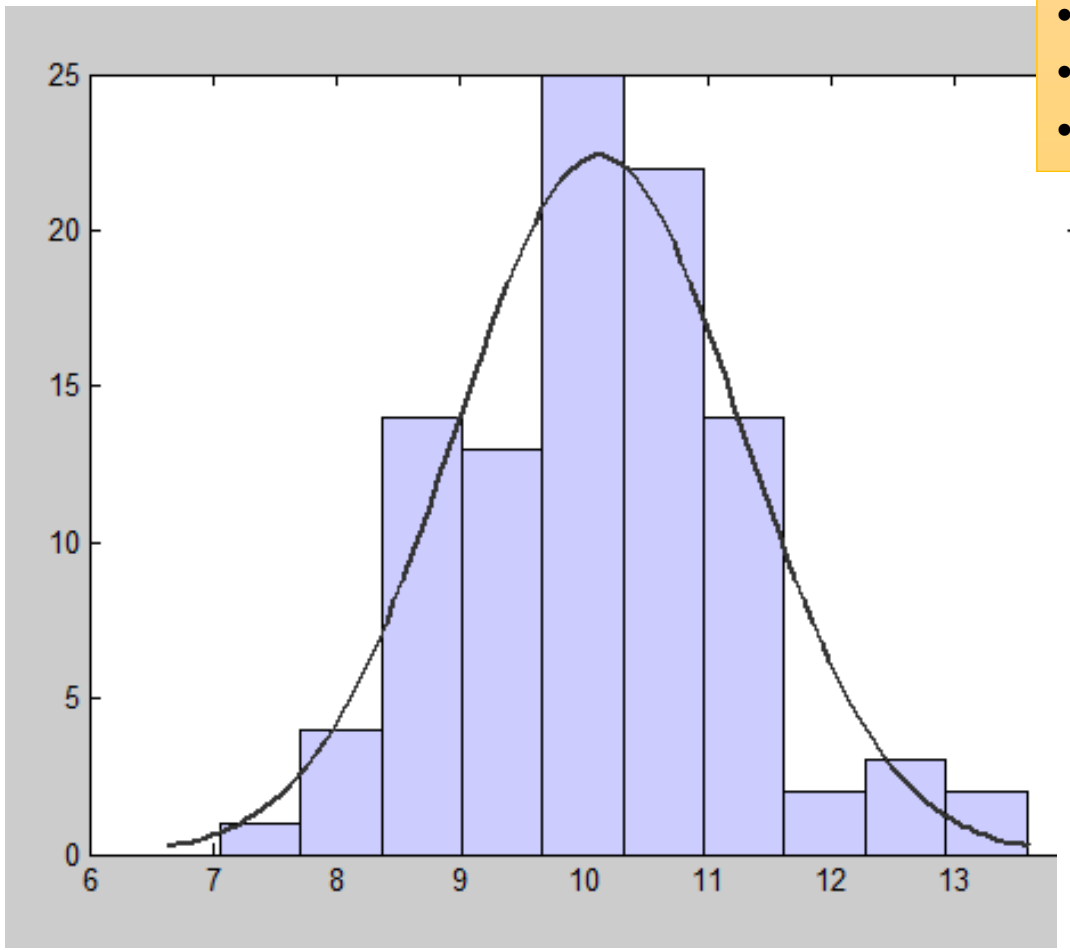
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of our scores from the mean

Frequency distributions are back!

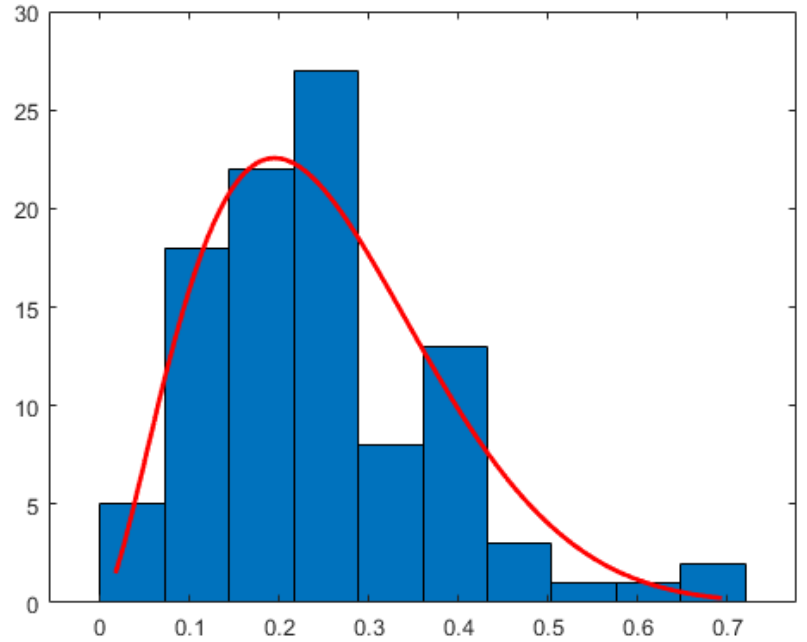
Remember, by "score," I just mean Ps' responses, what researcher measured.
EX:

- # of Twitter 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)

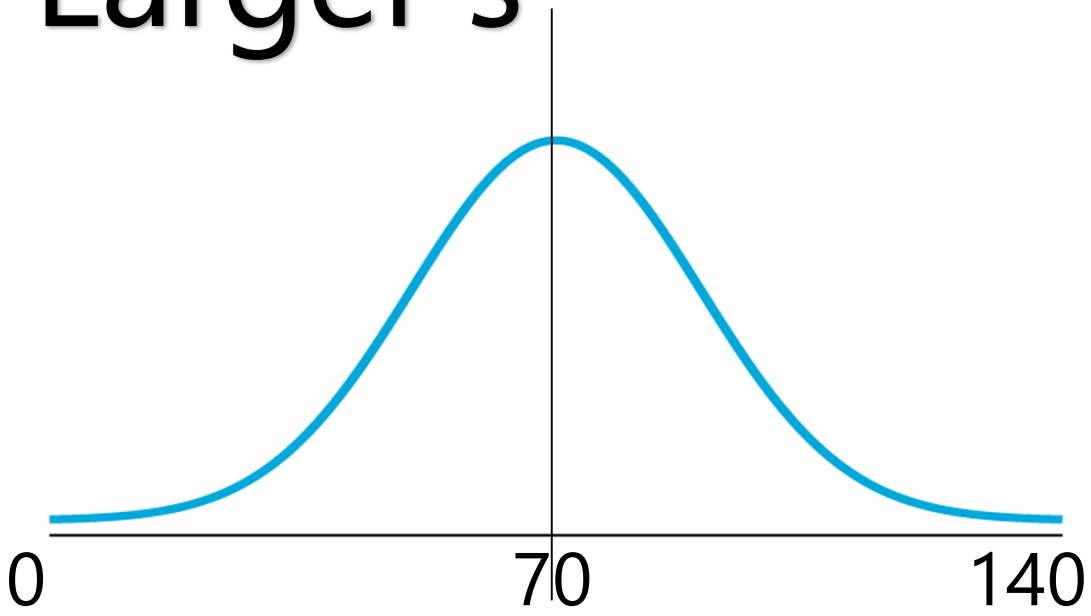


score



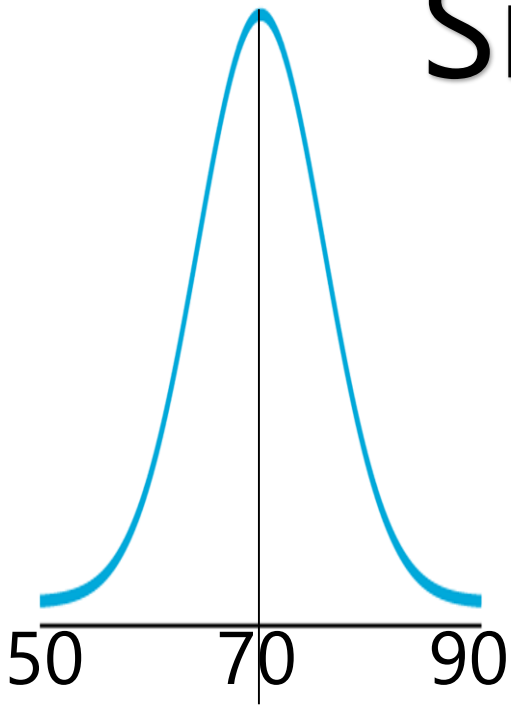
Equal Mean, Unequal Standard Deviations

Larger *s*



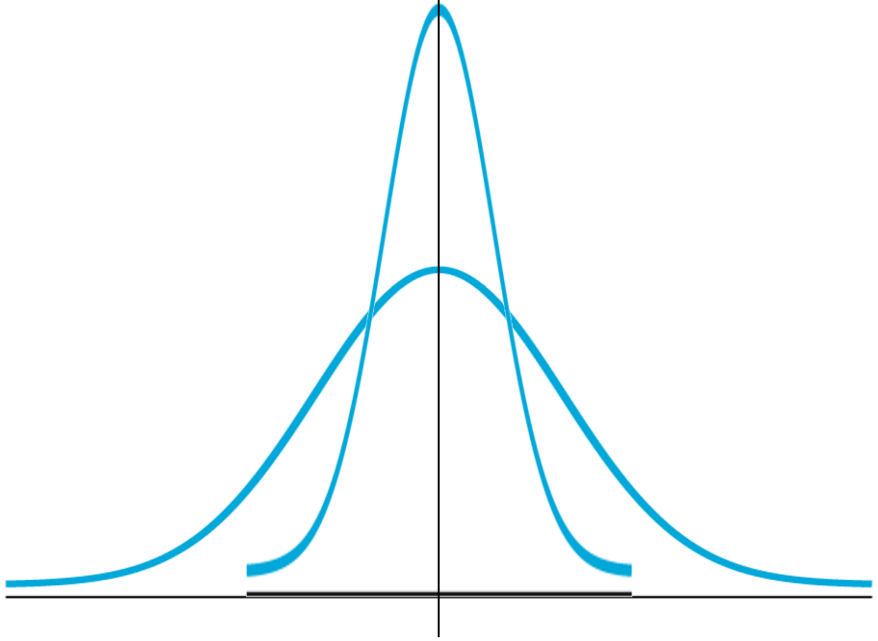
Mean score = 70

Smaller *s*

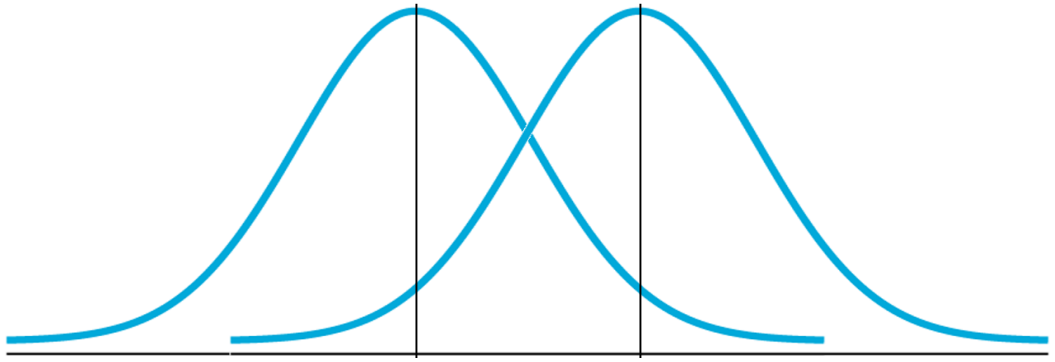


Mean score = 70

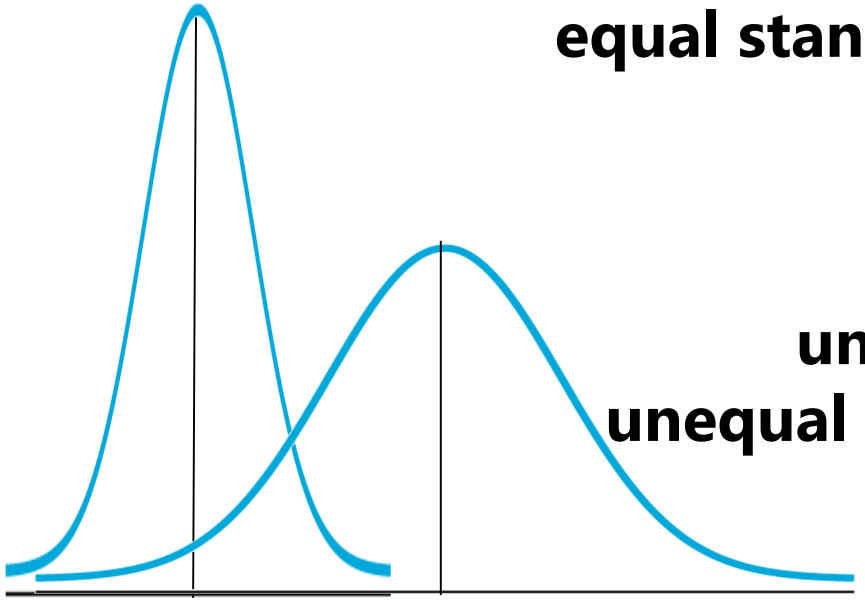
REMINDER: *s* provides a measure of the *standard* (*typical*, *average*) distance of individual scores from the mean



equal means,
unequal standard deviations



unequal means,
equal standard deviations



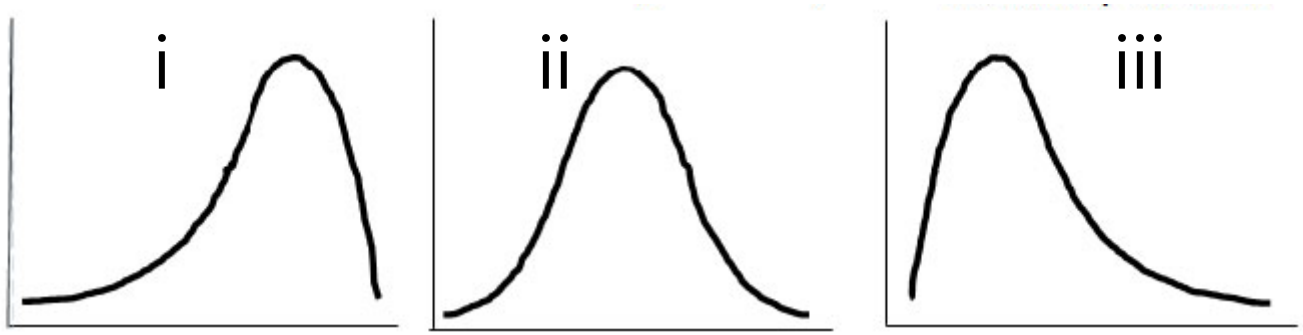
unequal means,
unequal standard deviations

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

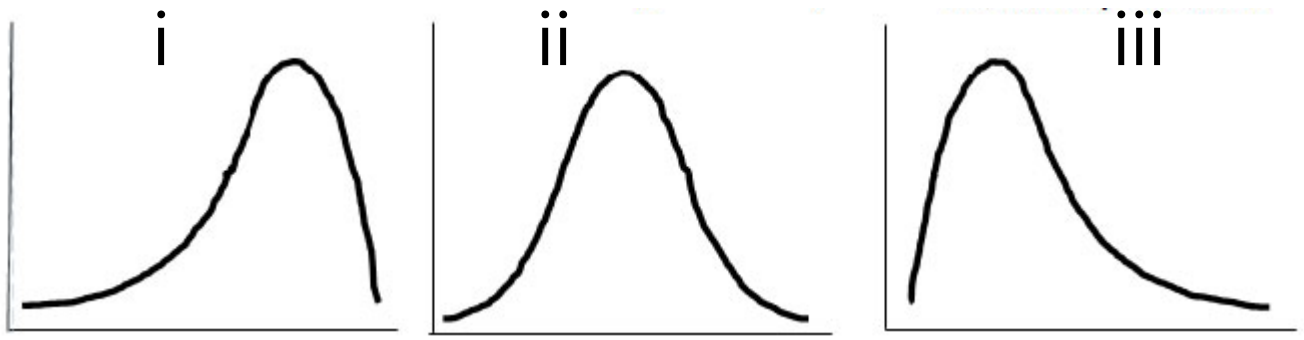
VARIABLE	A	B	C
N	150	150	150
Mean	6.28	13.94	3.81
Median	6.0	6.5	6.0
Standard deviation	2.75	6.51	4.5
Range	10	20	20

For each Q below, select variable ***A, B, or C,*** whichever best answers each question.

1. Which variable above has scores that are clustered most closely around the sample mean?
2. Which variable above has scores that are spread farthest from the sample mean?
3. Which variable above has the most symmetrical (aka, *normal*) distribution?
4. Which variable above is most likely to have a distribution with a positive skew?
5. Which variable above is most likely to have a distribution with a negative skew?
6. Match the variables A, B, and C to the distribution below that fits best.



1. Which variable has scores that are clustered closely around the sample mean?
   - A
2. Which variable has scores that are spread farthest from the sample mean?
   - B
3. Which variable has the most symmetrical (aka, *normal*) distribution?
   - A
4. Which variable is most likely to have a distribution with a positive skew?
   - B
5. Which variable is most likely to have a distribution with a negative skew?
   - C
6. Match the variables A, B, and C to the distribution below that fits best.
   - Variable A matches with distribution **ii** the best
   - Variable B matches with distribution **iii** the best
   - Variable C matches with distribution **i** the best



VARIABLE	A	B	C
N	150	150	150
Mean	6.28	13.94	3.81
Median	6.0	6.5	6.0
Standard deviation	2.75	6.51	4.5
Range	10	20	20

Time to Draw! Practice distinguishing between *skew* and *outliers*

1. Draw a distribution with a positive skew, and no outliers
2. Draw a distribution with a negative skew, and an outlier with a very low score
3. Draw a distribution with a negative skew, and an outlier with a very high score
4. Draw a distribution with no skew, and some very high, and some very low, outliers