Lesson 004
Measures of Variability

September 18, 2023

	Student 1	Student 2	Student 3
Course #1	80	70	60
Course #2	80	75	60
Course #3	80	85	100
Course #4	80	90	100
Mean / Median	**80 / 80**	**80 / 80**	**80 / 80**

Variability

When some data are more *spread out* than others, we say that they have higher **variability**.

There is less concentration around the measures of location.

Measures of Variability

- The simplest measure of variability is the **range**, given by

$$\text{range} = x_{\max} - x_{\min}.$$

Measures of Variability

- The simplest measure of variability is the **range**, given by

$$\text{range} = x_{\max} - x_{\min}.$$

- The **sample variance** is given by squared deviations from the mean.

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

Measures of Variability

- The simplest measure of variability is the **range**, given by

$$\text{range} = x_{\max} - x_{\min}.$$

- The **sample variance** is given by squared deviations from the mean.

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

  - Note, the formula you will often see will be *slightly* different. Ignore this for now!

Measures of Variability

- The simplest measure of variability is the **range**, given by

$$\text{range} = x_{\max} - x_{\min}.$$

- The **sample variance** is given by squared deviations from the mean.

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

  - Note, the formula you will often see will be *slightly* different. Ignore this for now!
  - The square root of the sample variance is called the **standard deviation**, $s = \sqrt{s^2}$.

	Student 1	Student 2	Student 3
Course #1	80	70	60
Course #2	80	75	60
Course #3	80	85	100
Course #4	80	90	100
Mean / Median	**80 / 80**	**80 / 80**	**80 / 80**
Range / Variance	**0 / 0**	**20 / 62.5**	**40 / 400**

Suppose that the following 5 values are observed: {7, 8, 5, 5, 7}. What is the range of the data?

0

Range = $8 - 7 = 1$

0%

Range = $7 - 7 = 0$

0%

Range = $7 - 5 = 2$

0%

Range = $8 - 5 = 3$

0%

Suppose that the following 5 values are observed: {7, 8, 8, 5, 7}. What is the variance of the data?

Variance = $\frac{(7-7)^2+(8-7)^2+(8-7)^2+(5-7)^2+(7-7)^2}{5} = 1.2$

0%

Variance = $\frac{(7-7)+(8-7)+(8-7)+(5-7)+(7-7)}{5} = 0$

0%

Variance = $\frac{7^2+8^2+8^2+5^2+7^2}{5} = 50.2$

0%

Variance = $\frac{7^2+8^2+8^2+5^2+7^2}{5} - 7 = 43.2$

0%

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.
- We have that $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.
- We have that $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.
- Adding constants to all of the data will not change the variance.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.
- We have that $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.
- Adding constants to all of the data will not change the variance.
- Multiplying all of the data by a constant, c, multiplies the variance by c^2

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.
- We have that $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.
- Adding constants to all of the data will not change the variance.
- Multiplying all of the data by a constant, c, multiplies the variance by c^2
  - The standard deviation will be multiplied by $|c|$.

Properties of the Variance and Standard Deviation

- We have both $s^2 \geq 0$ and $s \geq 0$, with equality only in constant data.
- The standard deviation makes most sense to discuss in conjunction with the mean.
- We define $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, so that $s^2 = \frac{S_{xx}}{n}$.
- We have that $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.
- Adding constants to all of the data will not change the variance.
- Multiplying all of the data by a constant, c, multiplies the variance by c^2
  - The standard deviation will be multiplied by $|c|$.
  - This can be useful for unit conversions.

The variance of temperature, measured in Celsius, was observed to be 25. These temperatures are converted to Kelvin, by adding 273.15. What is the variance and standard deviation of the new data?

Variance is $25 + 273.15 = 298.15$ and standard deviation is $\sqrt{298.15} = 17.267$.

0%

Variance is $25 \times 273.15 = 6828.75$ and standard deviation is $\sqrt{6828.75} = 82.636$.

0%

Variance is 25 and standard deviation is 5.

0%

There is not enough information given to answer either.

0%

The variance of temperature, measured in Celsius, was observed to be 25. These temperatures are converted to Fahrenheit, by first multiplying by 1.8 then adding 32. What is the variance and standard deviation of the new data?

Variance is $1.8 \times 25 + 32 = 77$ and standard deviation is $\sqrt{77} = 8.775$.

0%

Variance is $1.8^2 \times 25 = 81$ and standard deviation is $\sqrt{81} = 9$.

0%

Variance is 25 and standard deviation is 5.

0%

Variance is $1.8 \times 25 = 45$ and standard deviation is $\sqrt{45} = 6.708$.

0%

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?
  - This quantity is called the p-**th percentile**.

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?
  - This quantity is called the p-**th percentile**.
  - The median is the 50-th percentile.

Generalizing from the Median

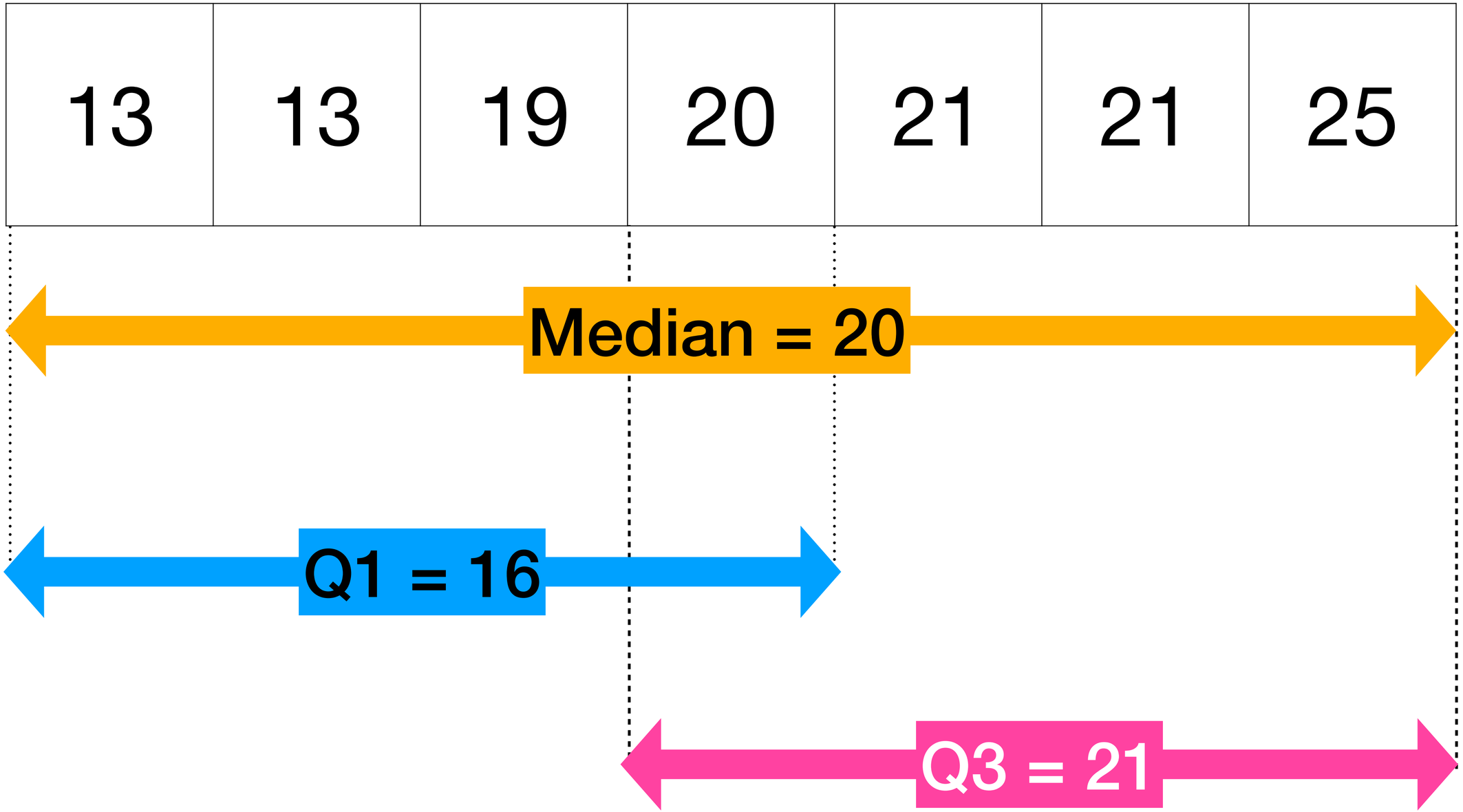
- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?
  - This quantity is called the p-**th percentile**.
  - The median is the 50-th percentile.
- We call the 25th percentile Q1, and the 75th percentile Q3.

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?
  - This quantity is called the p-**th percentile**.
  - The median is the 50-th percentile.
- We call the 25th percentile Q1, and the 75th percentile Q3.
  - This stands for **q**uartile 1 and 3.

Generalizing from the Median

- Recall that the median divides the data so that 50% is above it and 50% is below it.
- What if we swapped from 50% to p% (below, and $(100 - p)$% above)?
  - This quantity is called the p-**th percentile**.
  - The median is the 50-th percentile.
- We call the 25th percentile Q1, and the 75th percentile Q3.
  - This stands for **q**uartile 1 and 3.
  - These can be computed as the median of the lower and upper half of the data.


13
13
19
20
21
21
25
Median = 20
Q1 = 16
Q3 = 21

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.
  - IQR is a measure of spread, related to the median.

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.
  - IQR is a measure of spread, related to the median.
  - Useful for detecting outliers.

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.
  - IQR is a measure of spread, related to the median.
  - Useful for detecting outliers.
  - Data which are $1.5 \times$ IQR away from the nearest quartile are mild outliers; more than 3 times are extreme outliers.

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.
  - IQR is a measure of spread, related to the median.
  - Useful for detecting outliers.
  - Data which are $1.5 \times \text{IQR}$ away from the nearest quartile are mild outliers; more than 3 times are extreme outliers.
- If we list min, $Q1$, median, $Q3$, max for data, this is the **five number summary**.

Interquartile Range and Five Number Summary

- The interquartile range, or IQR, is calculated as $\text{IQR} = Q3 - Q1$.
  - IQR is a measure of spread, related to the median.
  - Useful for detecting outliers.
  - Data which are $1.5 \times$ IQR away from the nearest quartile are mild outliers; more than 3 times are extreme outliers.
- If we list min, $Q1$, median, $Q3$, max for data, this is the **five number summary**.
  - We can display the five number summary using a **box plot**

13	13	19	20	21	21	25

Min	Q1	Median	Q3	Max	IQR
13	16	20	21	25	5

Consider the data given by {1, 1, 2, 2, 3, 5}. What is Q1, Q3, and the IQR?

$\mathrm{Q1} = 1; \mathrm{Q3} = 3; \mathrm{IQR} = 4$

0%

$\mathrm{Q1} = 1.5; \mathrm{Q3} = 2.5; \mathrm{IQR} = 1$

0%

$\mathrm{Q1} = 1; \mathrm{Q3} = 3; \mathrm{IQR} = 2$

0%

$\mathrm{Q1} = 1.5; \mathrm{Q3} = 2.5; \mathrm{IQR} = 4$

0%