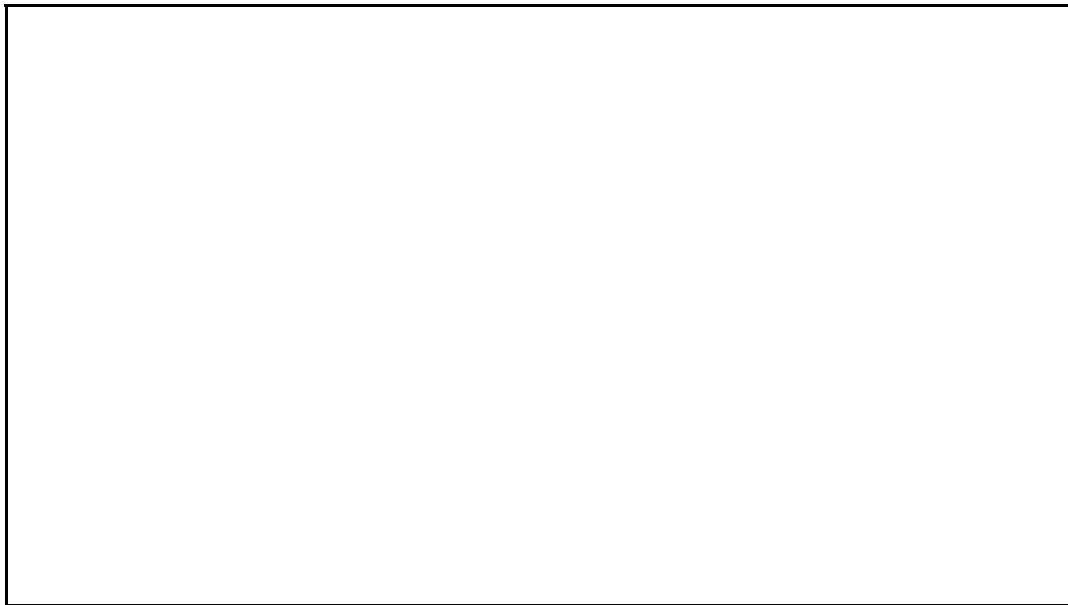


[Courseware](#) [Course Info](#) [Discussion](#) [Syllabus](#) [Download R and RStudio](#) [R Tutorials](#) [Readings](#) [Contact Us](#)
[Progress](#) [Office Hours](#)

CENTER AND SPREAD



SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

When we numerically summarize a variable from our dataset in statistics, we want to give our reader two main things. We want to give them a measure of the variable's center, and we want to give them a measure of the variable's spread around that center.

Now we've got a couple of 1/17/2014 4:30 PM

choose

from when it comes to the measure of center.

We've got the mean, the median, and the mode.

For our measure of spread, on the other hand,

we have the candidates of Standard Deviation or the Interquartile Range (or IQR).

Now, we've all heard of the mean: it's the average.

It's used all over the place; we see it all the time

in headlines or in major statements.

Here are just a few of those headlines that include a measure of mean.

Average Workweek In America Is Actually Much Longer Than 40 Hours A Week

Average Student Debt Climbs To \$29,000

House Prices Almost Five Times the Average, Local Salary

Sea Levels Rising In Antarctica Have Been

Nearly A Third Faster Than The Global Average

But what about these other measures of center: the median and the mode?

How are these alike or how are these

different from the mean?

More importantly, when should we use each of them?

Under what conditions?

Are these headlines accurate when it comes down to it?

Should they all be using the average?

Well, in the end, it all comes down to shape.

We know that the shape of a distribution will push and pull

the measurement of center around.

When the distribution is skewed to the right (or a positive skew),

the mean gets pulled to the right and brings along with it the median.

The mode stays right where it is, at the peak of the data,

at the most common occurring value.

When the distribution is skewed to the left, the mean gets pulled to the left,

and again pulls the median with it, while again the mode stays right

where it was.

Isn't that interesting.

If we want a measure of center, and the distribution is symmetrical

(it's not skewed in either direction), then the mean makes sense to use.

It's a common measure that everyone can understand.

In fact, all three measures of center will be actually

really close to each other.

But if the distribution is skewed, then the mean turns out to be a bad idea.

With a positive skew, the mean over estimates the measure of center.

The same thing happens with a skew in the other direction;

the mean under estimates the central value of the distribution.

So which measure of center should we use?

The way the measures act informs this decision.

A symmetrical variable distribution can use the mean;

it's a good measure that people understand.

If the distribution of the variable is skewed, then the median is the measure of choice.

It's NOT the most common variable (or the mode);

it's the measure of center that informs the center of the distribution

without being too affected by the skew as the mean is.

This comes in handy when we want to look at two subgroups and their data

on the same variable, and one group shows a skewed distribution

while the other is symmetrical.

Remember where the measure of centers are for EACH type of distribution.

If we want to compare the two subgroups, using the mean for the comparison, that would mean that one subgroup is misrepresented.

But, if we use the median instead, then the comparison is a little bit better.

The symmetrical subgroup is still the same value,

and the measure of center for the skewed data is a better estimate.

In our case, for the cost-of-out of state tuition for Public

institutions is symmetrical.

Its mean, median, and mode are pretty close to one another

(within about \$1,000).

But the distribution of Private institutions

is positively skewed - notice how the mean is pulled to the right,

followed by the median and then the mode.

In fact, the distribution is SO skewed that

the mean is over \$5,000

higher than the median.

This same rationale holds for the measure of spread.

When the distribution is symmetrical, the Standard Deviation

is a great measure of spread.

It gives us a measure of variability around the mean.

But, there's a hiccup - the Standard Deviation

is actually describing a particular space in the distribution

as long as it's symmetrical.

It's right here about, just as the curve changes.

But what happens when the distribution is skewed?

Where that curve changes is different on one side of the mean vs. the other.

And remember, with a skewed distribution, the mean is actually

is a poor choice to use.

And the mean is used in the calculation of Standard Deviation.

So we turn to the Interquartile Range (or IQR).

This measure of spread is the difference

between the first

and the third quartiles, which are informed
by the median - the preferred

measure of center for a skewed
distribution.

So, to answer our question, because a
visual inspection of the distribution

for both groups shows both symmetrical
and skewed data,

we use the median for both - the stable
measure.

Private universities have a higher
out-of-state tuition

than Public schools.

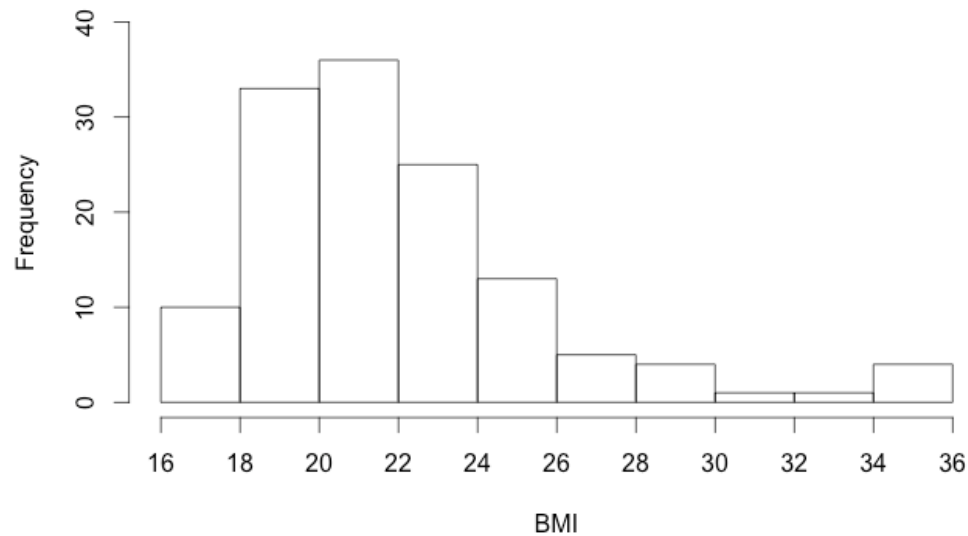
The median cost for a Public school is
\$16,500 while for a Private school

that cost is around \$23,000.

We use the median for both - the stable
measure.

Comprehension Check

Below are data from a recent national survey on the BMI (body mass index) of a sample of Americans.



Identify the missing values in this portion of the frequency table that was used to generate the BMI histogram.

x	f
[16, 18)	(a)
[18, 20)	33
[20, 22)	36
[22, 24)	(b)
[24, 26)	13
[26, 28)	5

(6/6 points)

The value of (a) is:

10

The value of (b) is:

25

25

The shape of this distribution can best be described as:

- ☐ bimodal
- ☐ left skewed
- ☒ right skewed
- ☐ uniform



A BMI of 30 is considered obese. Which of the following is the best description of the BMI of this sample?

- ☐ This sample demonstrates that a large percentage of Americans are obese.
- ☒ The majority of individuals are not obese, but there are some extreme values in the tail of the distribution. ✓
- ☐ The prevalence of obesity cannot be determined from this graph because BMI is an unreliable measure.
- ☐ There are no obese individuals in this sample.

When you are comparing two sets of data, and one is strongly skewed and the other is symmetric, which measure of center should you choose for the comparison?

- ☐ Mean
- ☒ Median ✓
- ☐ Mode
- ☐ The distributions cannot be compared

Which of the following is an accurate interpretation of the median of a distribution?

- ☐ About 50% of the population have the median value.
- ☐ The median is the numerical average of the dataset.
- ☐ The middle 50% of scores in a distribution constitute the median.
- ☒ Half the distribution has a lower score than the median, and half has a higher score than the median. ✓

[Check](#)[Show Answer](#)

Twelve coworkers log their hours worked overtime in the past month:

{10, 2, 6, 12, 14, 15, 15, 24, 15, 25, 3, 12}

[Help](#)

(12/12 points)

Please attempt to answer the following questions by hand if you can:

What is the mean of the sample? (*Round to 2 decimal places.*)

What is the standard deviation of the sample? (*Round to 2 decimal places.*)

What is the 5 number summary for the sample?

min =

Q1 =

Q2 =

Q3 =

Help

max =


The mean number of overtime hours was *(Round to 1 decimal place.)*

The median number of overtime hours was

The range was

The IQR was

What is the shape of this distribution of scores?

☒ approximately normal 

☐ left skewed

☐ right skewed

☐ uniform

Check

Show Answer



Help



EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

© 2014 edX, some rights reserved.

[Terms of Service and Honor Code](#)

[Privacy Policy \(Revised 4/16/2014\)](#)

About & Company Info

[About](#)

[News](#)

[Contact](#)

[FAQ](#)

[edX Blog](#)


[Donate to edX](#)

[Jobs at edX](#)

Follow Us

 [Twitter](#)

 [Facebook](#)

 [Meetup](#)

 [LinkedIn](#)

 [Google+](#)