Skew The Script

All solutions and teacher notes in blue

AP Statistics Handout Key: Lesson 1.4

Topics: measures of center, measures of spread, using technology to find summary stats

Lesson 1.4 Guided Notes

The table below shows the salaries of 12 employees at a company.

Measures of Center

Salaries (thousands of $)
39
34
34
35
34
32
43
34
185
35
29
67

1) Write the formula for **the mean** (in words and symbols). Then, find the mean salary (show your work).

$$\bar{x} = \frac{\text{Sum of data values}}{\text{number of data values}} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{39+34+34+35+34+32+43+34+185+35+29+67}{12} = \frac{601}{12} = \mathbf{50.1}$$

Mean: $50.1 thousand dollars ($50,100)

2) Find the median salary (show your work). For convenience, the salaries (in thousands of $) are shown in order below:

~~29~~, ~~32~~, ~~34~~, ~~34~~, ~~34~~, 34, 35, ~~35~~, ~~39~~, ~~43~~, ~~67~~, ~~185~~

$\frac{34+35}{2} = 34.5$ → **Median: $34.5 thousand ($34,500)**

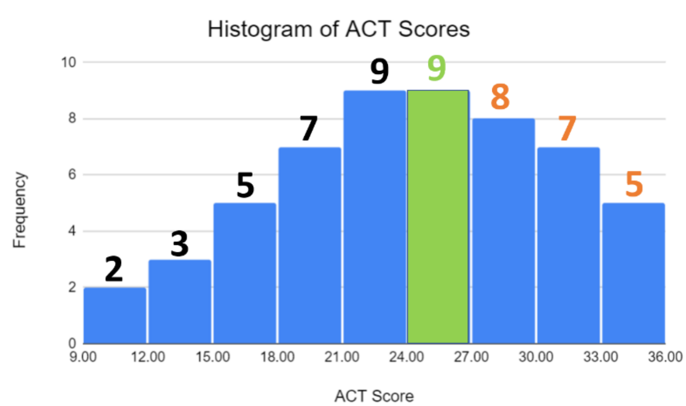
Question Preview (for later discussion): *The boss is trying to hire you to work at this company. She says, "Our typical salary is $50,100." Is this misleading? Why or why not?*

Approximating a median in a histogram: The graph below describes 55 ACT scores.

3) Approximate the median ACT score. Show your work.

55/2 = 27.5 → The median is the 28th data point. We reach the 28th data value in the green highlighted bin (24 – 27).

The median is an ACT score between 24-27



Measures of Spread

4) Write the formula for the **range**. Then, find the range of the salaries (show your work).

Range = Max – Min
Range = 185 – 29 = **$156 thousand**

5) The formula for the **standard deviation** (s_x) is displayed to the right:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Let's break down the formula piece-by-piece:

a) Why do you think we find the differences between each data point and the mean ($x_i - \bar{x}$)? Hint: the standard deviation is a measure of *spread.*

These differences provide a way to measure the distance between data values and a measure of center (the mean). This shows us how far spread out each data value is from the "center."

b) Why do you think we square the differences $(x_i - \bar{x})^2$ before we sum them?

We square the differences to make them all positive. This way, the negatives don't cancel the positives when we add them (which would ruin our attempt to get the "total" spread).

c) We find the sum ($\sum$) and divide by n – 1. This is similar to what we did before for another statistic. Which statistic was that? Why do you think we are doing the (almost) same thing here?

This is similar to the mean, where we found the sum of the data and divided by n. By doing something similar, we get something like the "average" / "typical" squared distance from the mean.

d) At the end, we take the square root ($\sqrt{}$). Why do you think we do this?

To cancel the squaring we did earlier and put our measure back in original units (dollars rather than squared dollars)

6) Write down and interpret the standard deviation of the salaries.

$s_x = 43.61$ → The salaries in the dataset are typically vary $43,610 away from the mean.

7) Write down the formula for the **interquartile range (IQR)**. Then, find the IQR (show your work). For convenience, the salaries (in thousands of $) are shown in order below:

IQR = Q3 – Q1

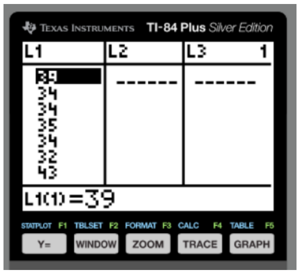
~~29~~, ~~32~~, [34, 34,] ~~34~~, ~~34~~, ~~35~~, ~~35~~, [39, 43,] ~~67~~, ~~185~~

Quartile 1: 34 Quartile 3: $\frac{39+43}{2} = 41$

IQR= 41-34 = 7 → **IQR: $7 thousand**

Question to ponder (you'll answer it later in the lesson): *Which measure of spread (range, standard dev., or IQR) best represents the "typical" distance between salaries? Why?*

Technology: Summary Statistics

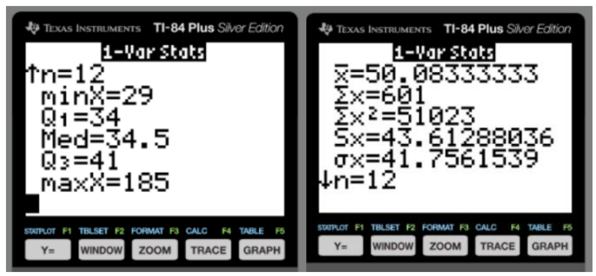


1. Put data into List 1 (STAT → EDIT)

2. Find 1-Var Stats (STAT → CALC → 2)

3. Select Data List (→ Calculate)

4. Scroll through the summary stats (use Sx for stdev.)



Lesson 1.4 Discussion

Recommended discussion norms:
skewthescript.org/discussion-norms

Salaries (thousands of $)
39
34
34
35
34
32
43
34
185
35
29
67

1. The boss is trying to hire you to work at this company. She says, “Our typical salary is $50,100.” Is this misleading? Why or why not?

Misleading: the mean is dragged up by high outliers. Specifically, the salary of $185,000 raises the overall sum of the data values significantly, resulting in a mean value that is higher than the vast majority of salaries at the company. It’s dubious to say that earning $50k is “typical” when it’s above 10 out of the 12 employee salaries.

Measures of Center
Mean: 50.1 ($50,100)
Median: 34.5 ($34,500)

2. Which measure of spread (range, standard dev., or IQR) best represents the “typical” distance between salaries? Why?

Most of the data values are between 29-39, with only a few employees that have unusually high salaries compared to the rest. It looks like the range and standard deviation were greatly affected by the high outlier salaries and had high estimates of the spread in the data (greater than this $10k range between $29k-$39k). The IQR shows more of the typical spread between the majority of data values.

Measures of Spread
Range: $156,000
S_x: $43,610
IQR: $7,000

"Resistance is futile"

The median is **resistant to** (not seriously affected by) skew and outliers. The mean **isn't resistant** to skew and outliers. The mean follows skew/outliers.

The interquartile range (IQR) is resistant to skew and outliers. The range and **standard deviation** are not resistant to skew and outliers.

Why are the median and IQR resistant to outliers? *Let's explore with the salary data:*

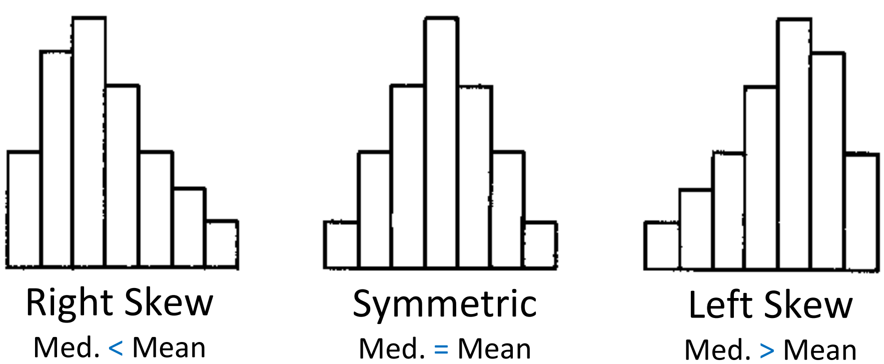
29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185

For the mean: The outlier salary – $185,000 – drags up the mean because its high value is given **a lot of weight** in the calculation

For the median: The **position** matters more than the **value**. Because $185,000 is the highest data point, it's crossed off right away. The outlier is **not** given a large weight in the calculation.

For the IQR: Like the median, the position matters more than the value for the IQR.

Applet: Use the (very cool) simulation linked here to explore these properties of the mean and median.
Link: http://digitalfirst.bfwpub.com/stats_applet/stats_applet_6_meanmed.html
Source: Digital First project from Bedford, Freeman, & Worth publishers



Lesson 1.4 Practice

Teachers: We recommend providing additional practice exercises from your AP Stats textbook or from prior AP Stats exams. The following textbook sections and AP exam questions are aligned to the content covered in this lesson.

- *The Practice of Statistics (AP Edition)*, 4th-6th editions: section 1.3
- *Stats: Modeling the World (AP Edition)*, 4th & 5th editions: chapter 3, 3rd edition: chapter 4
- *Statistics: Learning from Data (AP Edition)*, 2nd edition: section 3.1 - 3.3
- *Advanced High School Statistics*, section 2.2
- AP Exam Free Response Questions (FRQs): 2014 Q4 (part a), 2019 Q6 (part b), 2018 Q5

Handout Key by statistics student Greyson Zuniga