# Let mw Know about HW Questions.

The first test on the Unit of Descriptive Statistics will be next **Tuesday, September 17th.** Starting Friday, you will be given review problems.

Warm Up

Solve an equation from Algebra 2

$$30(2.5)^t = 2000$$

Warm Up from Algebra 2 - Solve for t

$$30(2.5)^t = 2000$$

$$(2.5)^t = \frac{2000}{30}$$

$$\log[2.5^t] = \log\left[\frac{2000}{30}\right]$$

$$t \cdot \log(2.5) = \log\left(\frac{2000}{30}\right)$$

$$t = \frac{\log\left(\frac{2000}{30}\right)}{\log(2.5)} = 4.58$$

$$\log_{2.5}\left(\frac{2000}{30}\right) = t$$

$$t = \frac{\log\left(\frac{2000}{30}\right)}{\log(2.5)}$$

$$= 4.58$$

Next

$$1800(2)^{-t} = 6$$

$$1800\left(\tfrac{1}{2}\right)^{t} = 6$$

$$\left(\tfrac{1}{2}\right)^{t} = \frac{6}{1800}$$

$$t\log\left(\tfrac{1}{2}\right) = \log\left(\frac{6}{1800}\right)$$

$$t = \frac{\log\left(6/1800\right)}{\log\left(\frac{1}{2}\right)} = 8.23$$

$$\frac{1800}{2^t} = \frac{6}{1}$$

$$6 \cdot 2^t = 1800$$

$$2^t = 300$$

HW

QUESTIONS

**15** A sample of 10 measurements has a mean of 15.7 and a sample of 20 measurements has a mean of 14.3. Find the mean of all 30 measurements.

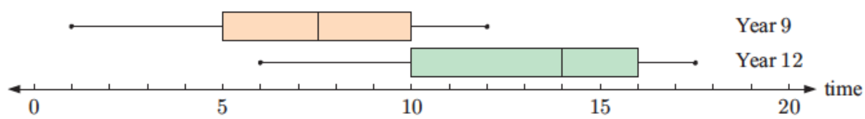
Mean of all 30 $= \dfrac{10(15.7) + 20(14.3)}{30} =$

**5** The table shows the sizes of land blocks on a suburban street.
Use technology to estimate the mean land block size.

| *Land size* ($m^2$) | *Frequency* |
|---|---|
| [500, 600) | 5 |
| [600, 700) | 11 |
| [700, 800) | 23 |
| [800, 900) | 14 |
| [900, 1000) | 9 |

## EXERCISE 6G.2

1 The following side-by-side boxplots compare the times students in years 9 and 12 spend on homework.




a Copy and complete:

| *Statistic* | *Year 9* | *Year 12* |
|---|---|---|
| minimum | | |
| $Q_1$ | | |
| median | | |
| $Q_3$ | | |
| maximum | | |

b For each group, determine the:
   i range
   ii interquartile range.

c Are the following true or false, or is there not enough information to tell?
   i On average, Year 12 students spend about twice as much time on homework as Year 9 students.
   ii Over 25% of Year 9 students spend less time on homework than all Year 12 students.

## EXERCISE 6H

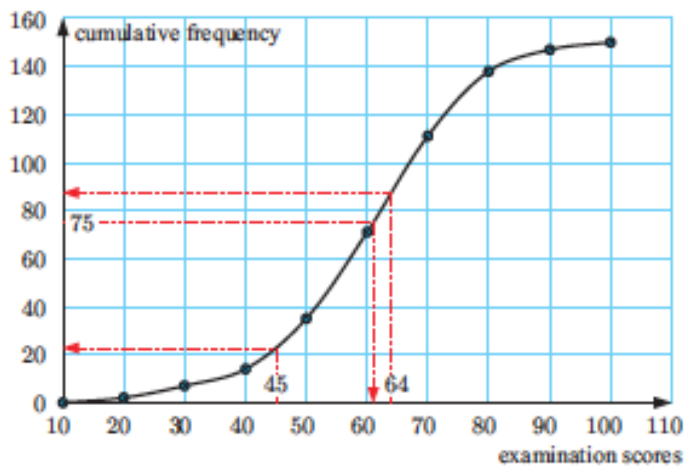
1 The examination scores of a group of students are shown in the table. Draw a cumulative frequency graph for the data and use it to find:

   a the median examination mark
   b how many students scored less than 65 marks
   c how many students scored between 50 and 70 marks
   d how many students failed, given that the pass mark was 45
   e the credit mark, given that the top 16% of students were awarded credits.

| *Score* | *Frequency* |
|---|---|
| $10 \leqslant x < 20$ | 2 |
| $20 \leqslant x < 30$ | 5 |
| $30 \leqslant x < 40$ | 7 |
| $40 \leqslant x < 50$ | 21 |
| $50 \leqslant x < 60$ | 36 |
| $60 \leqslant x < 70$ | 40 |
| $70 \leqslant x < 80$ | 27 |
| $80 \leqslant x < 90$ | 9 |
| $90 \leqslant x < 100$ | 3 |

## EXERCISE 6H

1




**a** ≈ 61 marks **b** ≈ 87 students **c** ≈ 76 students

**d** ≈ 23 students **e** 79 marks

**2** A botanist has measured the heights of 60 seedlings and has presented her findings on the cumulative frequency graph below.

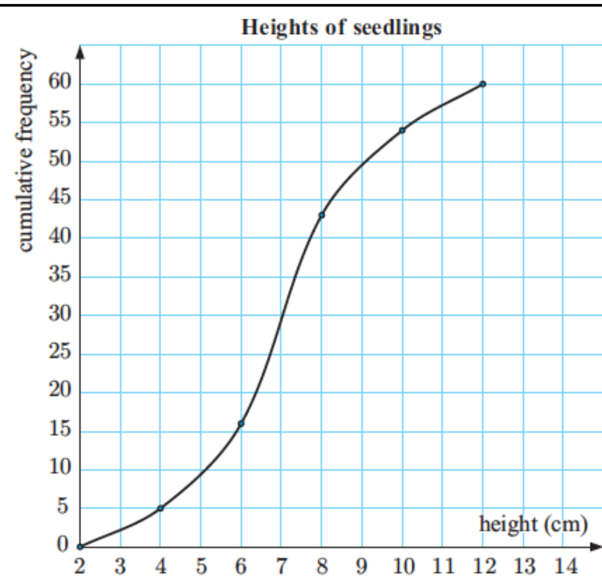
**a** How many seedlings have heights of 5 cm or less?

**b** What percentage of seedlings are taller than 8 cm?

**c** Find the median height.

**d** Find the interquartile range for the heights.

**e** Copy and complete:
"90% of the seedlings are shorter than ......"



Heights of seniors - Table 1

| Name | Height |
|---|---|
| Fred | 6.0 ft. |
| George | 5.8 ft. |
| Harry | 5.9 ft. |
| Melvin | 5.6 ft. |
| Vern | 6.5 ft. |
| Dan | 5.6 ft. |
| Andrew | 5.8 ft. |
| Craig | 5.8 ft. |
| Nate | 5.9 ft. |
| Jeff | 5.8 ft. |

**Mean = 5.87 feet**

**Standard deviation (or σ) = 0.254 feet**

Heights of Seniors - Table 2

| Name | Height |
|---|---|
| Mark | 6.8 ft. |
| Matt | 5.6 ft. |
| Lloyd | 5.2 ft. |
| Jim | 4.6 ft. |
| Cooper | 7.1 ft. |
| Kirk | 5.8 ft. |
| Charlie | 5.7 ft. |
| Cleavon | 5.9 ft. |
| Bob | 5.6 ft. |
| Kenneth | 6.1 ft. |

**Mean = 5.84 feet**

**Standard deviation (or σ) = 0.719 feet**

**Objectives:**

Calculate and interpret the standard deviation of a data set

- by hand
- using GDC
- with a spreadsheet on a laptop using Google Sheets

We'll need 3 or 4 add'l laptops tomorrow.

A

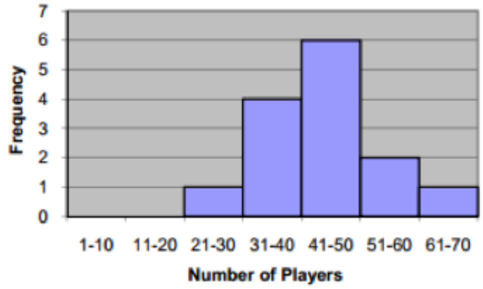
don't write

**The**
**Standard Deviation**
**is a measure of how much variation there is from the center of the data.** actually from the mean.

Basketball Camp

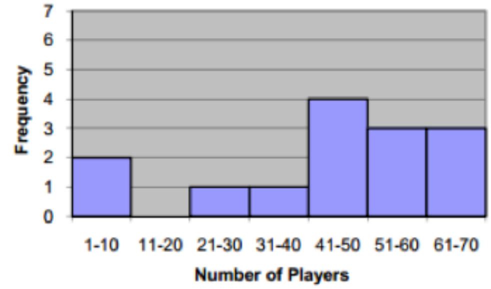
| Data set 1 | Data set 2 |
|---|---|

**Number of Basketball Players Recorded Once Each Day from April 1-14**




Mean = 45.0
Variance = 106.1
Standard Deviation = 10.3

**Number of Basketball Players Recorded Once Each Day from April 15-28**




Mean = 45.0
Variance = 420.3
Standard Deviation = 20.5

## How many Colleges Do You Think You Might Apply To?

How many different colleges is your group of 4 applying to? Find the total number of colleges for your whole group. *Don't double count the same college*

1. Record the data for the class here. ________________________________________

3 13 11 16 9 30 9
11 13

**Finding Quartiles**

2. List these values below in order from least to greatest. Find the **median** of the data by hand (not with a GDC). Draw a line in the data set to mark this and write in the median value. You have now split the data set into two halves

3 9 9 11 11 13 13 16 30

$Q_1 = 9$ $Q_3 = 14.5$ median = 11

$\frac{9+1}{2} = 5$

3. The range is somewhat helpful to indicate how spread out the data is but does not tell us how much all of the data is spread out. The IQR is a little better for that. Now go back to your diagram above find $Q_1$ and $Q_3$ the same way, by drawing lines. Then calculate the IQR = 5.5 colleges
Interpret, in context, the IQR.

$14.5 - 9$

$\frac{10}{2} = 5$

The middle half the values varies by 5.5 colleges

**Finding Standard Deviation**

$\bar{x} = 12.8 \rightarrow 13$

4. An even better way to measure variability in the data is with the standard deviation. How can we find the ***average distance of the values from the mean***?

a. Find the mean of the data (we'll round)
b. Find the distance of each piece of data from the mean.
c. Square the differences to make them positive.
d. Find the "average".
e. Square root to undo the squares.

$$\frac{(3-13)^2 + (13-13)^2 + (11-13)^2 + (16-13)^2 + (9-13)^2 + (30-13)^2 + (9-13)^2 + (11-13)^2 + (13-13)^2}{9}$$

$s^2 = \frac{438}{9} = 48.7$ colleges

$$s = \sqrt{\frac{438}{9}} = 6.98 \text{ colleges}$$

4. Enter the classroom data in your GDC and find the summary statistics. Verify our work above.

5. Finally, *interpret* our standard deviation, in context.

$\bar{x}$ = 13 colleges

S = 6.98 colleges

The number of colleges typically varies by 6.98 colleges from the mean (13 colleges)

B

# The Standard Deviation

**is the square root**

**of the average of the squared deviations from the mean.**

**Deviation just means how far from the normal**

$(x_i - \bar{x})$

each data value

mean

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

| | |
|---|---|
| Range | Standard Deviation |
| IQR | |

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

| | | |
|---|---|---|
| Range = max - min<br>IQR = $Q_3 - Q_1$ | Standard Deviation<br>$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$ | |

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

| | | |
|---|---|---|
| Range = max - min<br>IQR = $Q_3 - Q_1$ | Standard Deviation<br>$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$<br>Use $\sigma_x$ value on GDC | $S$ always $\geq 0$ |

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

| | | |
|---|---|---|
| Range = max - min<br>IQR = $Q_3 - Q_1$ | Standard Deviation<br>$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$<br>Use $\sigma_x$ value on GDC | $S$ always $\geq 0$<br>Large $S$ means greater variation |

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

## Summarize - Measuring Variability

| | | |
|---|---|---|
| Range = max - min<br>IQR = $Q_3 - Q_1$<br>If skewed distrib. safer to use median and IQR | Standard Deviation<br>$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$<br>Use $\sigma_x$ value on GDC | $S$ always $\geq 0$<br>Large $S$ means greater variation |

J

# Entire Population

$\mu$ = population mean

$\sigma$ = population std. deviation

## Sample

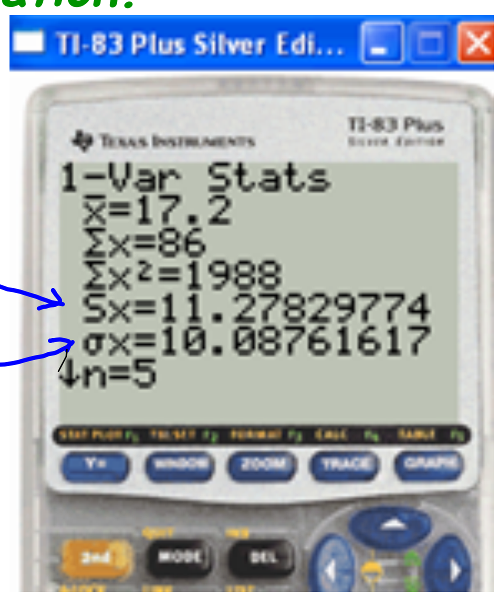
$\bar{x}$ = 14 college

$s_n = \frac{}{n-1}$

$\mu$

$\sigma = \frac{}{n}$

***Your GDC will calculate both versions of the standard deviation.***




Always for IB

The **standard deviation** is the square root of the average of the squared deviations from normal. (normal being the *mean*)

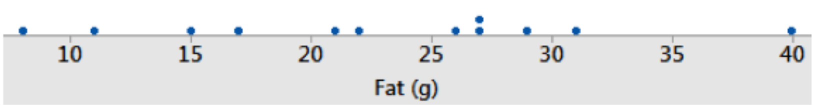
## Summarize - Measuring Variability

| | | |
|---|---|---|
| Range = max - min<br>IQR = $Q_3 - Q_1$<br>If skewed distrib. safer to use median and IQR | Standard Deviation<br>$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$<br>Use $\sigma_x$ value on GDC | $s$ always $\geq 0$<br>Large $s$ means greater variation<br>If roughly symmetric distribution, use $\bar{x}$ and $s$ |

**Have we found the beef?**

Here are data on the amount of fat (in grams) in 12 different McDonald's sandwiches, along with a dotplot. The mean fat content for these sandwiches is $\bar{x} = 22.833$ grams.

27 11 22 21 40 8 17 15 29 31 27 26




1. Find the **range** of the distribution. ____________

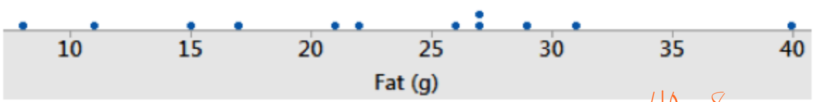
2. Use your GDC to find the **interquartile range**. IQR = ____________

3. Use your GDC to calculate the **standard deviation**. Interpret this value, in context.

**Have we found the beef?**

Here are data on the amount of fat (in grams) in 12 different McDonald's sandwiches, along with a dotplot. The mean fat content for these sandwiches is $\bar{x} = 22.833$ grams.

27 11 22 21 40 8 17 15 29 31 27 26




1. Find the **range** of the distribution. 32 grams  40 − 8

2. Use your GDC to find the **interquartile range**. IQR = 12 grams  28 − 16

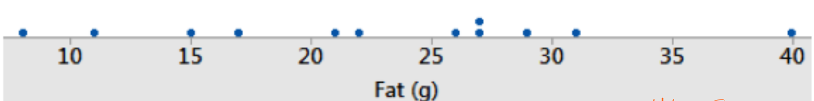
3. Use your GDC to calculate the **standard deviation**. Interpret this value, in context.

Need to know mean as well $\bar{x} = 22.8$

**Have we found the beef?**

Here are data on the amount of fat (in grams) in 12 different McDonald's sandwiches, along with a dotplot. The mean fat content for these sandwiches is $\bar{x} = 22.833$ grams.

27 11 22 21 40 8 17 15 29 31 27 26




1. Find the **range** of the distribution. 32 grams  40 − 8

2. Use your GDC to find the **interquartile range**. IQR = 12 grams  28 − 16

3. Use your GDC to calculate the **standard deviation**. Interpret this value, in context.

Need to know mean as well $\bar{x} = 22.8_g$  $S_x = 8.68$ grams

The fat content typically varies by 8.68 grams from the mean (22.8 g)

4a. The dotplot suggests that the Bacon Clubhouse Burger, with its 40 grams of fat, is a possible outlier. Recalculate the range, interquartile range, and standard deviation for the other 11 sandwiches.

w/o 40

4b Compare these values with the ones you obtained in Questions 1 through 3. Explain why each result makes sense.

4a. The dotplot suggests that the Bacon Clubhouse Burger, with its 40 grams of fat, is a possible outlier. Recalculate the range, interquartile range, and standard deviation for the other 11 sandwiches.

w/o 40 range = 20 grams, IQR = 12 grams Std. Dev = 7.27 grams

4b Compare these values with the ones you obtained in Questions 1 through 3. Explain why each result makes sense.

4a. The dotplot suggests that the Bacon Clubhouse Burger, with its 40 grams of fat, is a possible outlier. Recalculate the range, interquartile range, and standard deviation for the other 11 sandwiches.

w/o 40 range = 20 grams, IQR = 12 grams Std Dev = 7.27 grams

4b Compare these values with the ones you obtained in Questions 1 through 3. Explain why each result makes sense.

range

IQR

Std.Dev

4a. The dotplot suggests that the Bacon Clubhouse Burger, with its 40 grams of fat, is a possible outlier. Recalculate the range, interquartile range, and standard deviation for the other 11 sandwiches.

w/o 40 range = 20 grams, IQR = 12 grams Std Dev = 7.27 grams

4b Compare these values with the ones you obtained in Questions 1 through 3. Explain why each result makes sense.

range decreased because the max went down

IQR stayed the same because the middle half isn't affected much by extreme values

Std.Dev

4a. The dotplot suggests that the Bacon Clubhouse Burger, with its 40 grams of fat, is a possible outlier. Recalculate the range, interquartile range, and standard deviation for the other 11 sandwiches.

w/o 40   range = 20 grams,   IQR = 12 grams   Std Dev = 7.27 grams

4b Compare these values with the ones you obtained in Questions 1 through 3. Explain why each result makes sense.

range   decreased because the max went down

IQR   stayed the same because the middle half isn't affected much by extreme values

Std. Dev   went down because there was no longer a large deviation from the mean created by 40.

$$S_x = \sqrt{\frac{(\ -\ )^2 + (\ -\ )^2 + (40 -\ )^2}{n}}$$

See your LCQ

✓ Remain in class

✓ Each group given a copy of solutions

✓ NO cell phones

**Assignment --- Ch 6 HH packet**

**p.181.... 1**
**p.185.... 4**
**p.196....5, 6**
**p.199...3, 4, 6**

The test on the first Unit of Descriptive Statistics will be next Tuesday, September 17th. Starting Friday, you will be given review problems.

***Oh, yes. It's time for a Statistics Joke***

***Two statisticians were traveling in an airplane from LA to New York. About an hour into the flight, the pilot announced that they had lost an engine, but don't worry, there are three left. However, instead of 5 hours it would take 7 hours to get to New York.***

***A little later, he announced that a second engine failed, and they still had two left, but it would take 10 hours to get to New York.***

***Somewhat later, the pilot again came on the intercom and announced that a third engine had died. Never fear, he announced, because the plane could fly on a single engine. However, it would now take 18 hours to get to New York. At this point, one statistician turned to the other and said,***

***"Gee, I hope we don't lose another engine, or we'll be up here forever!"***