Warm-Up 4/26/18

A random group of high school students was surveyed. Each student was asked whether it should be mandatory for all high school students to participate in a sport. The results are partially summarized below.

1. In the freshmen group, about what percentage of students agree?

$\frac{53}{72} = 0.736$ or 74%

	Agree	Disagree	No Opinion	Total
Freshman	53	12	7	72
Sophomore	65	37	2	104
Junior	18	42	12	72
Senior	56	67	4	127
Total	192	158	25	375

2. What percent of students agree?

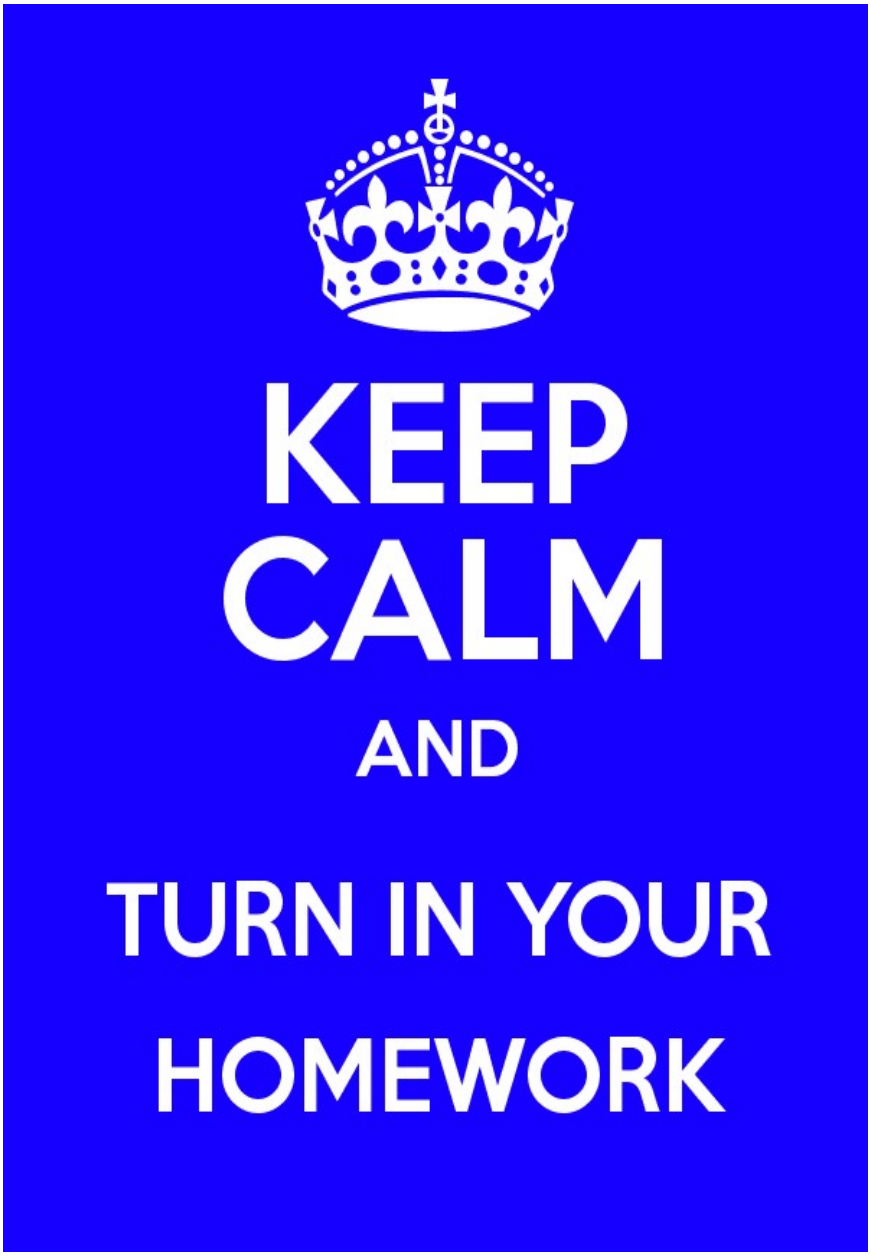
$192/375 = 51\%$

3. What percent of juniors have no opinion?

$12/72 = .1667$ or 17%

4. What percent of seniors disagree?

$67/127 = 52\%$



Day 2: Dot Plots & Histograms # 2 & 4; Day 3: Box Plots #s 1-3; Day 4: Comparing Data # 1-5:

Problem Solving #1-3

Day 6: Conditional Probabilities # 1-3

Unit 6 4/26/18

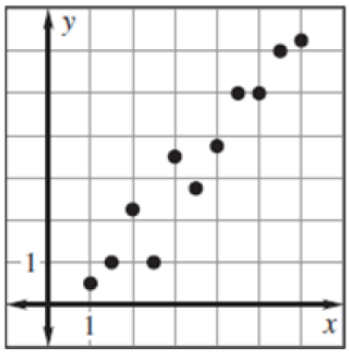
Essential Questions:

1. How can we create and interpret a scatterplot?
2. How can we interpret the correlation coefficient?

Notes 4/26/18

Day 7 – Scatterplots

A **scatterplot** is a graph of data pairs (x, y). Scatterplots are typically used to describe relationships, called **correlations,** between two variables (bi-variate). The **correlation coefficient** describes how well a line fits the data. A **trend line** can be drawn to help determine correlation.

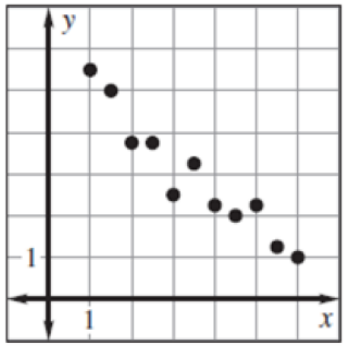


Positive Correlation

As x values increase, y values increase

Correlation Coefficient is close to 1

Positive Slope

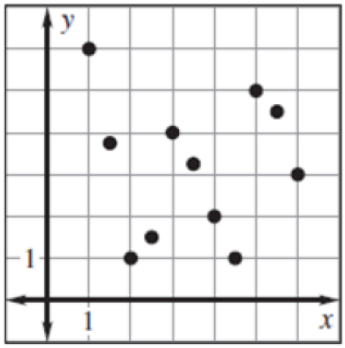


Negative Correlation

As x values increase, y values decrease

Correlation Coefficient is close to -1

Negative Slope



No Correlation

No relationship between x and y

Correlation Coefficient is close to 0

No line

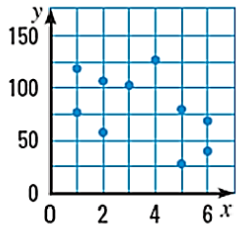
r

Correlation Coefficients			
0.70 to 1.00	Strong Positive	- 0.70 to -1.00	Strong Negative
0.30 to 0.69	Moderate Positive	- 0.30 to - 0.69	Moderate Negative
0.00 to 0.29	None to Weak Positive	0.00 to - 0.29	None to Weak Negative

-1 -.7 -.5 -.2 0 .2 .5 .7 1

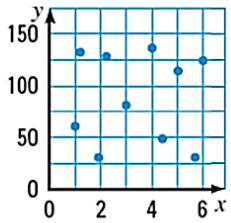
Example: Determine if the following graphs have positive, negative, or no correlations. Then tell if the correlation coefficient is strong, moderate, or weak positive or negative.

a.



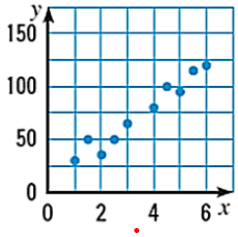
None

b.



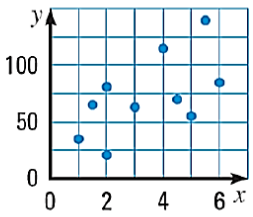
None

c.



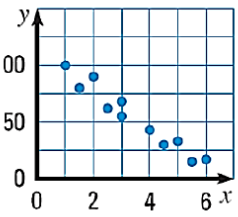
positive strong

d.



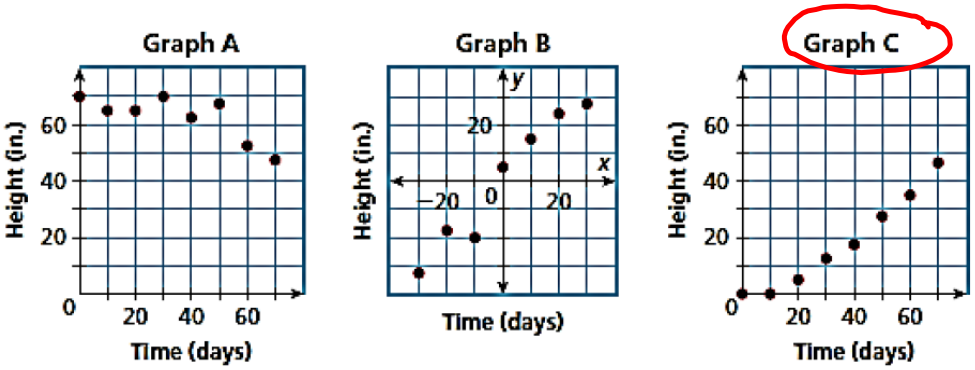
positive weak

e.



negative strong

Example: Describe the scatterplot that best describes the scenario below and explain why:
The relationship between the number of days since a sunflower seed was planted and the height of the plant.



Example: Describe the correlation you would expect to see between each pair of data sets. Explain your choice:

a. The number of hours you work vs the amount of money in your bank account:

Positive Correlation

b. The number of hours workers receive safety training vs the number of accidents on the job:

Negative moderate

c. The number of students at Hillgrove vs the number of dogs in Atlanta:

No Correlation

d. The number of heaters sold versus the months in order from April to September:

Negative Correlation

e. The number of rice dishes eaten vs the number of cars on I-75 throughout the day:

No correlation

f. The number of calories burned/lost vs the amount of hours you worked out:

positive Correlation

Correlation vs Causation

Correlation: implies a mutual relationship between two or more things. It is very IMPORTANT to understand that just because two variables are strongly correlated does NOT imply a cause and effect relationship. A strong relationship between two variables could be a coincidence or caused by additional factors. Typically, correlations use the words noticed and showed.

Correlations only show relationships…they cannot be used to make conclusions!!

Causation: implies a relationship in which one action or event is the direct consequence of another (cause and effect).

Correlation	Causation
• Smoking is correlated with alcoholism *(but it doesn't cause it)*. • The more ice cream consumed on a beach, the increased number of people who go in the water *(eating ice cream doesn't cause you to go in the water more)*.	• The more you smoke, the chances of developing lung cancer increase. *(Does smoking cause lung cancer?)* • The less calories you eat, the more weight you lose *(Does eating less cause you to lose weight?)*

Example: Determine if the following relationships show a correlation or causation:

A. A recent study showed that college students were more likely to vote than their peers who were not in school.

Correlation

B. Dr. Shaw noticed that there was more trash in the hallways after 2nd period than 1st period.

Correlation

C. You hit your little sister and she cries.

Causation

D. The number of miles driven and the amount of gas used on your trip to Disneyworld.

Causation

E. The age of a child and his/her shoe size.

Correlation

F. The amount of cars a salesman sells and the amount of commission he makes during the month of July.

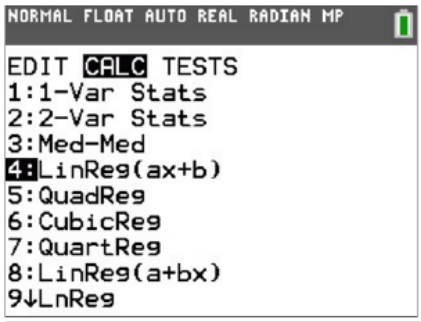
Causation

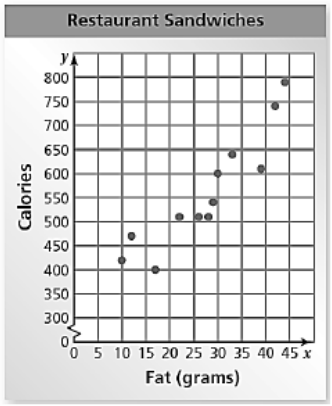
Steps for Calculating the Correlation Coefficient & Creating a Model

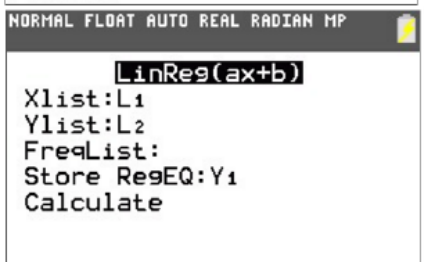
1. Once your data is entered into a list, Press [STAT] → [CALC] and choose your regression.
   - 4: LinReg – Linear Regression $y = mx + b$ $(a = m)$
   - 5: QuadReg – Quadratic Regression $y = ax^2 + bx + c$
   - 0: ExpReg – Exponential Regression $y = ab^x$

2. If you want your graphing calculator to automatically input the equation into y = , do the following:
   On the 2nd screen, hit ENTER until STORE REGEQ is highlighted.
   Hit Vars → Y-VARS → 1: Function → 1: Y_1

3. Hit ENTER until CALCULATE is highlighted. You should see your variables (a, b, and possibly c) unless with r^2 and r.
   R: correlation coefficient – this tells you how much correlation exists between your data
   R^2: this tells you how well the equation fits your data. The closer to 1, the better the fit.







TI 30 MULTIVIEW LINE OF BEST FIT STEPS

1. DATA (type in data)
2. 2nd DATA
3. 2 VAR L1 L2 CALC (enter)
   TI-36 Pro 2 VAR L1 L2 Frequency of 1 Calc
4. a =
   b =
   r =
   ★ You can use the x variable button to find a, b, and r.
5. The equation of the line is y = **a**x + **b.**
6. Correlation Coefficient is r.
7. To predict use **a(predict #) + b**. *Estimated method*

(12,9) (3,8) (10,13) (8,5)
X Y X Y X Y X Y

L1 : X values

L2 : Y-values

D: a = 0.024 - slope

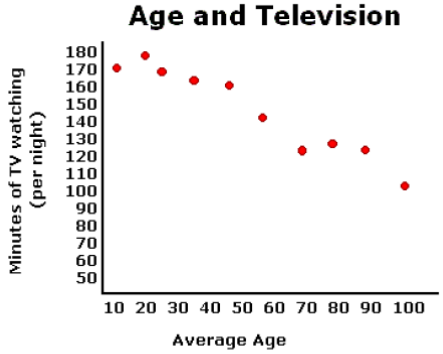
E: b = 8.695 - y-int

F: r = 0.0735 no correlation

$y = 0.02x + 8.7$

Practice Predicting with Scatter Plots

1. What can be concluded from the scatterplot below?



A. The older a person gets, the more television they watch.

B. As a person gets older, their taste in television changes.

C. The older a person gets, the less television they watch.

D. There is no relationship between age and television watching.

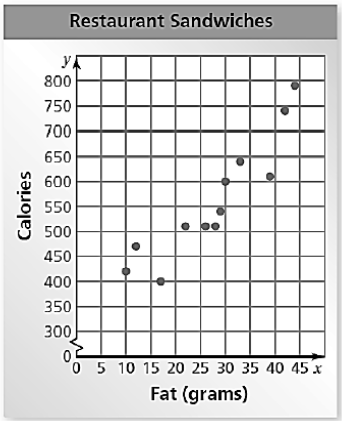
2. The scatterplot shows the number of fat (grams) in a restaurant and the number of calories in a sandwich.

a. How many grams of fat would you predict to be in a sandwich that contains 650 calories?

About 33 grams of fat

b. How many calories would you predict to be in a sandwich with 20 grams of fat?

About 500 calories.

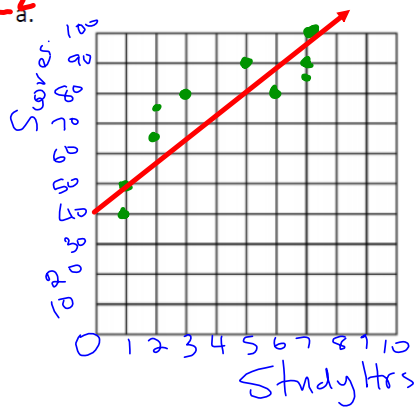


3. Make a scatterplot for each data set. Then find the correlation coefficient using your calculator.

L1 X hrs L2

a.

Study Hours	Regents Score
3	80
5	90
2	75
6	80
7	90
1	50
2	65
7	85
1	40
7	100

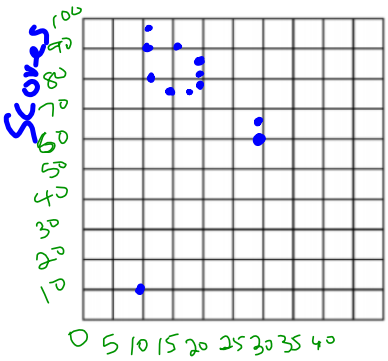


r = 0.85

L1 L2

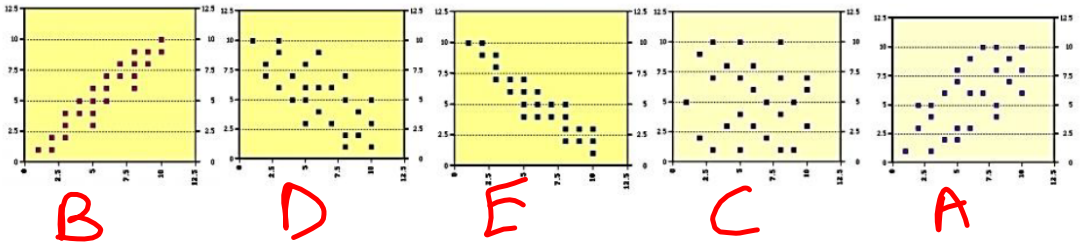
TV Hrs/week	Test Score
30	60
12	80
30	65
20	85
10	10
20	78
15	75
12	95
15	75
11	90
16	90
20	80
19	75

b.



r = -0.02

4. Match the graph with its correlation coefficient.



B D E C A

Choices

A. r = 0.45

B. r = 0.94

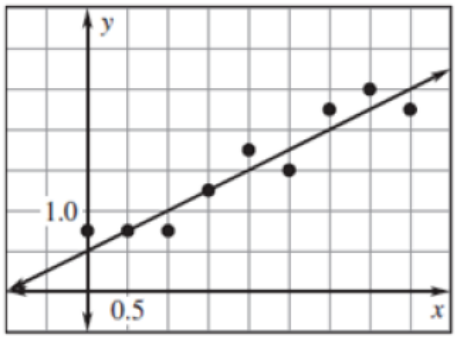
C. r = 0.07

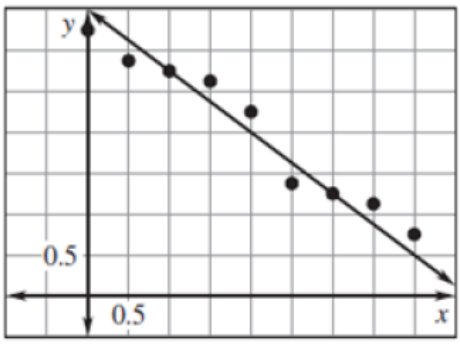
D. r = -0.39

E. r = -0.89

Day 8 – Linear Regression

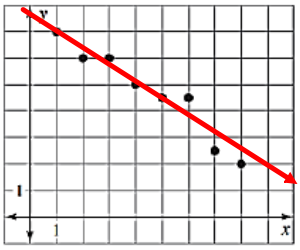
Yesterday, we drew trend lines to help us see if a scatter plot had any types of correlation. A **trend line** is a line that closely models the data. A **line of best fit** is the line that comes closest to all of the points in the data set. The line of best fit provides the predicted values for a set of data.

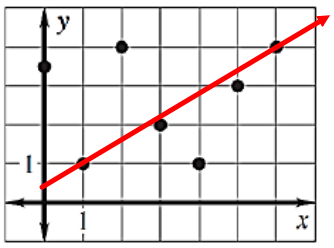


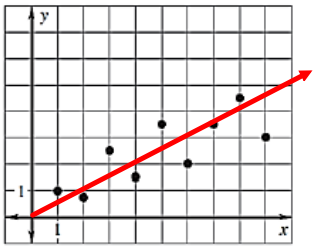


If a line is a good line of best fit, it will have data points above and below the line.

Example: Draw a line of best fit for each graph:







Example: The table shows test averages of eight students. The equation that best models the data is $y = 0.77x + 18.12$ and the correlation coefficient is 0.87. Discuss correlation and causation for the data set.

U.S. History Test Average	90	70	75	100	90	85	80	90
Science Test Average	80	75	72	95	92	82	80	92

$r = 0.87$
positive strong

No causation

Example: Eight adults were surveyed about their education and earnings. The table shows the survey results. The equation that models the data is $y = 0.59x + 30.28$ and the correlation coefficient is 0.86. Discuss correlation and causation for the data set.

Years of Education	12	16	20	14	18	16	16	18
Earnings Last Year (thousand $)	40	65	75	44	70	50	54	86

$r = 0.86$
strong positive

No causation

Calculating a Line of Best Fit

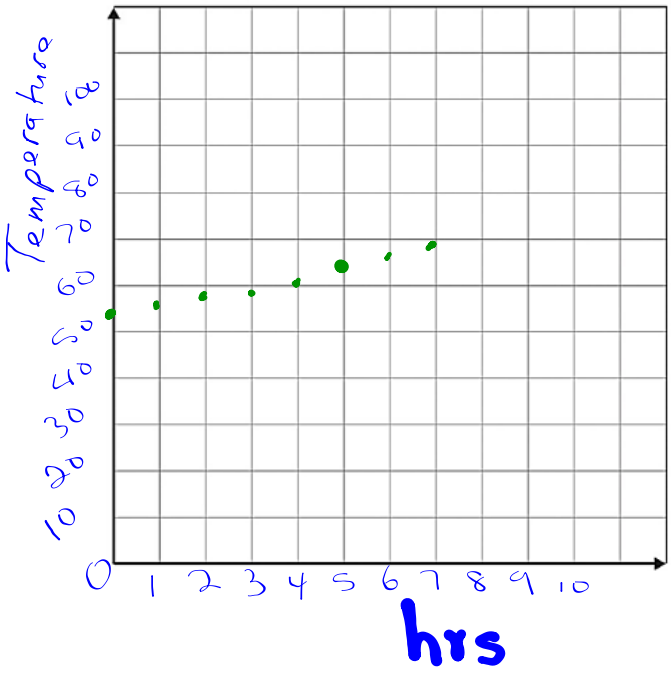
Scenario 1: A weather team records the weather each hour after sunrise one morning in May. The hours after sunrise and the temperature in degrees Fahrenheit are in the table below. Create a graph to represent the data and calculate a linear equation to represent the table.

L1 L2

Hours after sunrise	Temperature in °F
0	52
1	53
2	56
3	57
4	60
5	63
6	64
7	67

y-int

a: 2.19 — slope
b: 51.3 — y-int
r: 0.993
y = 2.2x + 51.3



Calculate by Hand

Step 1: Pick two points and calculate the slope (must go through trend line):

(3, 57) (7, 67)

x_1 y_1 x_2 y_2

$m = \frac{10}{4} = 2.5$

Step 2: Estimate/determine the y-intercept:

y-int: (0, 52)

$y - y_1 = m(x - x_1)$

$y - 57 = 2.5(x - 3)$

Step 3: Enter into y = mx + b

$y - 57 = 2.5x - 7.5$

$+57 \quad +57$

$y = 2.5x + 49.5$

Calculate using Regression

Step 1: Enter data into a list (Stat → Edit)

Step 2: Calculate a regression (Stat → Calc → 4: Lin Reg)

a: 2.19

b: 51.3

r: 0.9939

3. Enter into y = mx + b

$y = 2.2x + 51.3$

a. Interpret what the slope of each equation means in terms of the problem context.

Hand: For every hour after sunrise, the temperature increases by 2.5 degrees.

Calculator: For every hour after sunrise, the temperature increases by 2.2 degrees.

b. Interpret what the y-intercept of each equation means in terms of the problem context.

Hand: The y-intercept means that the temperature at sunrise is 49.5 degrees.

Calculator: The y-intercept means that the temperature at sunrise is 51.3 degrees.