

Statistical Reasoning *key*

Scatterplots and Correlation

skip #8

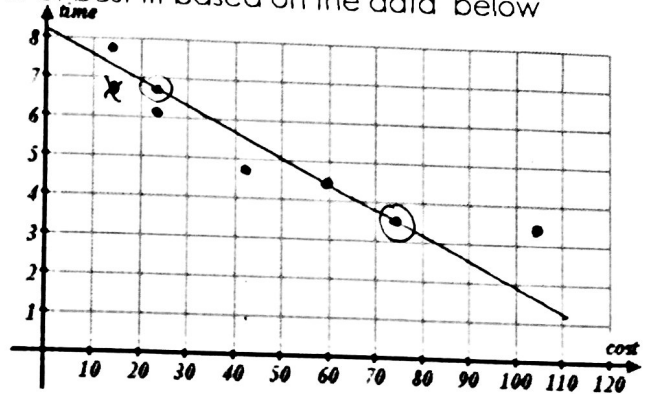
Name: _____

Date: _____

Class: _____

1. Create a scatter plot and approximate a trend line of best fit based on the data below

Model	Cost of Car	0-60 mph acceleration
Scion xB	\$16 K	7.8 sec
Mitsubishi Eclipse	\$24 K	6.1 sec
Chev. Corvette	\$106 K	3.4 sec
Nissan GT-R	\$76 K	3.5 sec
SSC Ultimate Aero	\$42 K	4.8 sec
Lotus Elise	\$60 K	4.4 sec
Honda Civic Si	\$22 K	6.7 sec



Choose 2 points from your trend line and identify the linear equation. $y = a + bx$

$(22, 6.7)$ $m = \frac{6.7 - 3.5}{22 - 76} = \frac{3.2}{-54} = -0.06$
 $(76, 3.5)$

$6.7 = -0.06(22) + b$
 $b = 8.02$

$y = -0.06x + 8.02$

Using your trend line, predict the 0-60 time for a car that costs \$120 K?

$y = -0.06(120) + 8.02 = 0.82 \text{ sec}$

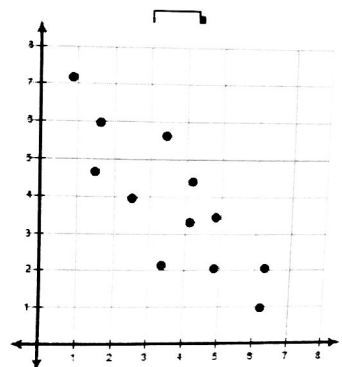
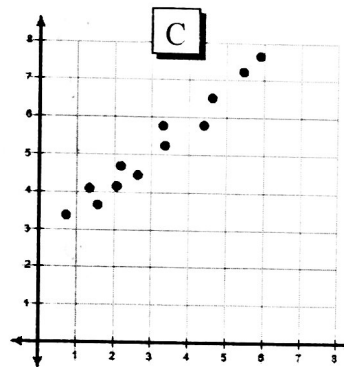
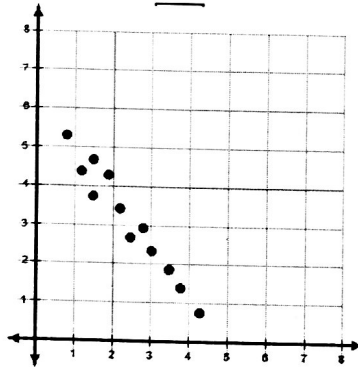
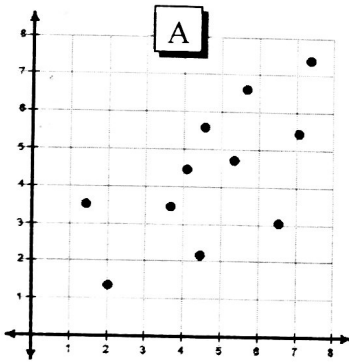
2. Consider the following scatter plots:

C strong positive correlation

A weak positive correlation

B strong negative correlation

D weak negative correlation



3. Consider the following situations. Determine whether you think they have a **positive** or **negative** correlation.

N a. Usually as a car increases in age, its value decreases.

P b. Usually the more hours that a person works the larger their paycheck.

P c. Usually the younger a child is, the smaller their height.

N d. Usually the longer you use a smart phone, the amount of battery life decreases.

4. Consider the following situations and answer the following **True** or **False** Questions.

A researcher noticed a relatively strong positive correlation between a student's score on the SAT and their GPA at the high school they attend.

F a. If one student has one of the lower SAT scored at one high school then they probably have one of the higher GPA's at their school.

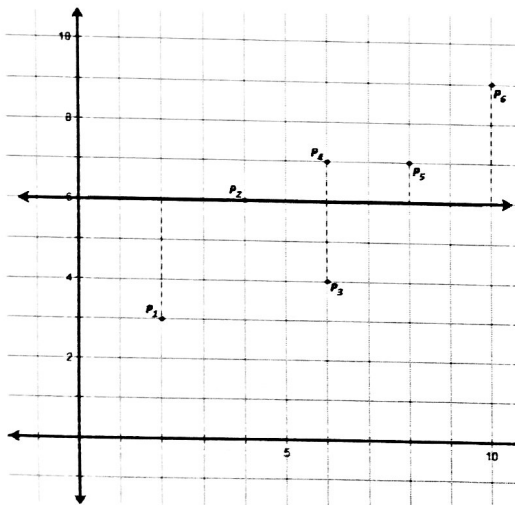
F b. If one student has the highest SAT score at one high school then they must have the highest GPA at their high school too.

T c. If one student has one of the higher SAT scores at one high school then they probably have one of the higher GPA's at their school.

5. Most trend lines that are considered to be a "good fit" will be balanced such that the total **RESIDUAL** above and below the trend line is equal. **RESIDUAL** can be defined as the difference between the actual value (y) and expected value (\hat{y}). A more succinct definition, **RESIDUAL** can be described as the vertical distance each data point is away from the trend line (with signed difference for above and below the trend line).

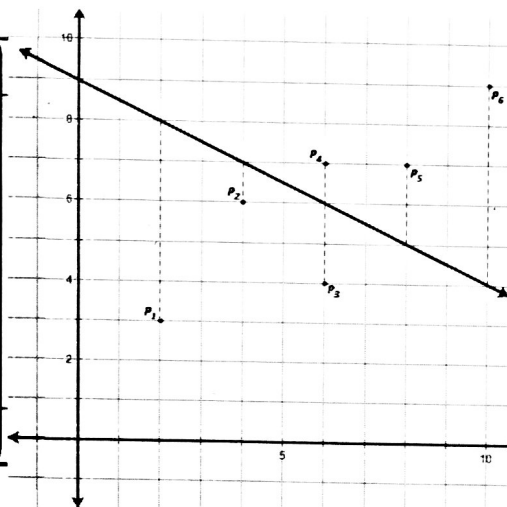
Find the **RESIDUALS** for each of the TREND LINES below (the SCATTER PLOT is the **same** in each graph).

TREND LINE 1



Data Point	Residual
P ₁	-3
P ₂	0
P ₃	-2
P ₄	1
P ₅	1
P ₆	3
Sum of Residuals	0

TREND LINE 2



Data Point	Residual
P ₁	-5
P ₂	-1
P ₃	-2
P ₄	1
P ₅	2
P ₆	5
Sum of Residuals	0

6. To better analyze which trend line is best, it is common to consider comparing the sum of the squares of the residuals. Which trend line do you think is the best based on this new information? Is it the one you expected?

7.

TREND LINE 1

Data Point	Residual	Residual Squared
P ₁	-3	9
P ₂	0	0
P ₃	-2	4
P ₄	1	1
P ₅	1	1
P ₆	3	9
Sum	0	24

TREND LINE 2

Data Point	Residual	Residual Squared
P ₁	-5	25
P ₂	-1	1
P ₃	-2	4
P ₄	1	1
P ₅	2	4
P ₆	5	25
Sum	0	60

8. Work in a group of four students. Each group member enters one of the data sets below into a graphing calculator, makes a scatterplot, and performs a linear regression residual analysis. Which of the 4 graphs has the highest and lowest residual value?

x	y
-2	-4
-1	-2
0	0
1	2
2	4

x	y
-2	2
-1	-3

x	y
-2	9
0	0
1	7
5	-2
7	4

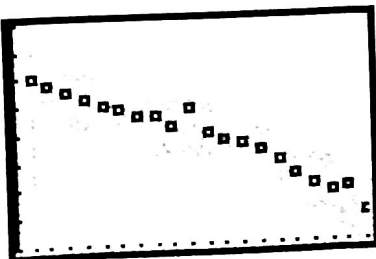
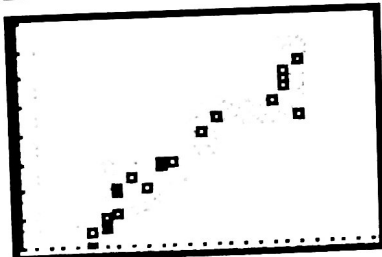
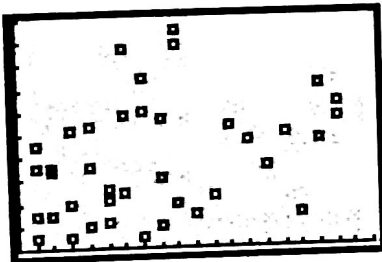
x	y
0	8
1	5

Regression Line Equation

$y =$ _____

Data Point	Residual	Residual Squared
P ₁		
P ₂		
P ₃		
P ₄		
P ₅		
Sum		

9. Consider each scatterplot below. Draw a line to match each r-value to a scatterplot.



$r = .972$

$r = .333$

$r = -.976$