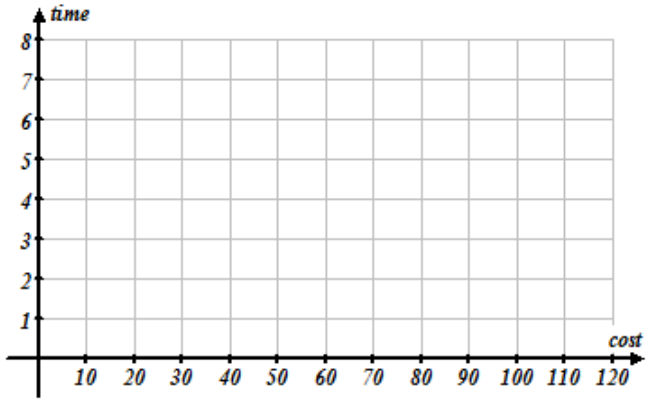
Sec 4.6 – Using Mathematical Models
Linear Regression Models

Name: _____

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

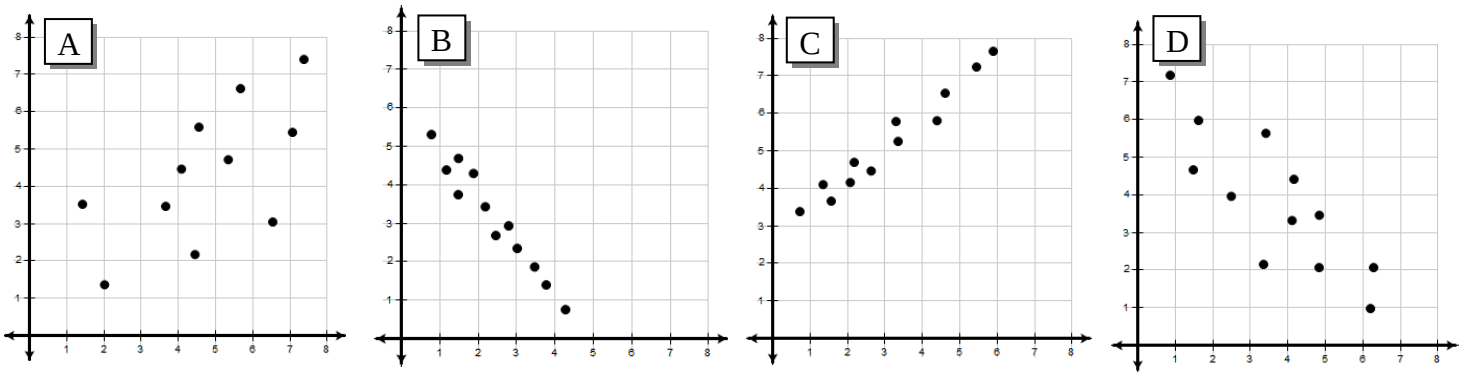
Model	Cost of Car	0-60 mph acceleration
Scion xB	$16 K	7.8 sec
Mitsubishi Eclipse	$24 K	6.1 sec
Chev. Corvette	$106 K	3.4 sec
Nissan GT-R	$76 K	3.5 sec
SSC Ultimate Aero	$42 K	4.8 sec
Lotus Elise	$60 K	4.4 sec
Honda Civic Si	$22 K	6.7 sec



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

2. Consider the following scatter plots:

_____strong positive correlation _____weak positive correlation _____strong negative correlation _____weak negative correlation



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

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

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

_____c. Usually the younger a child is, the smaller their height.

_____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.

_____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.

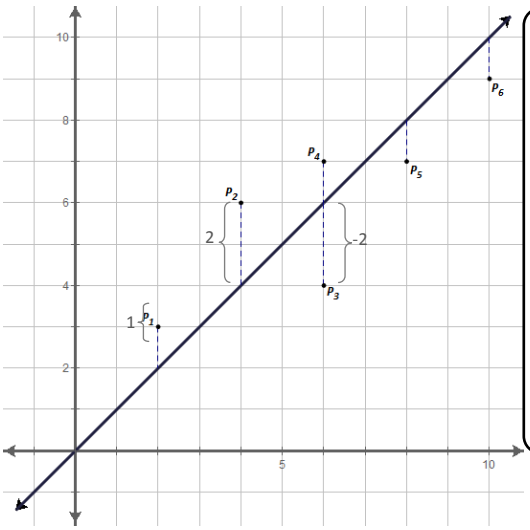
_____b. If one student has the highest SAT score at one high school then they must have the highest GPA at their high school too.

_____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).

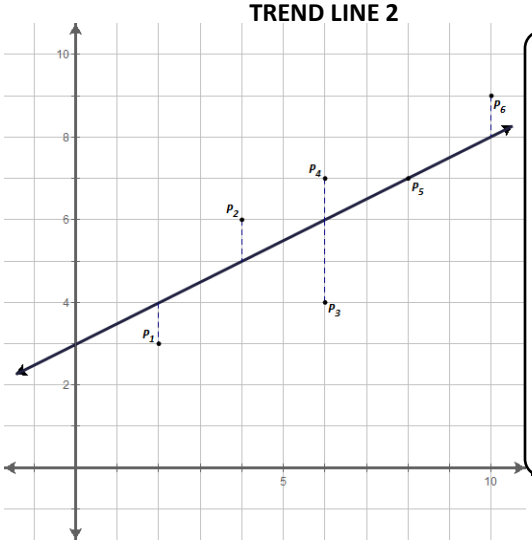
Find the **RESIDUAL**s for each of the TREND LINES below (the SCATTER PLOT is the **same** in each graph).

TREND LINE 1



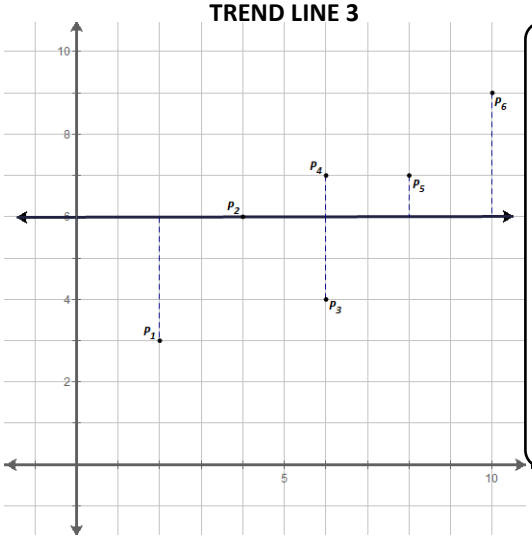
Data Point	*Residual*
P_1	1
P_2	2
P_3	−2
P_4	
P_5	
P_6	
Sum of Residuals	

TREND LINE 2



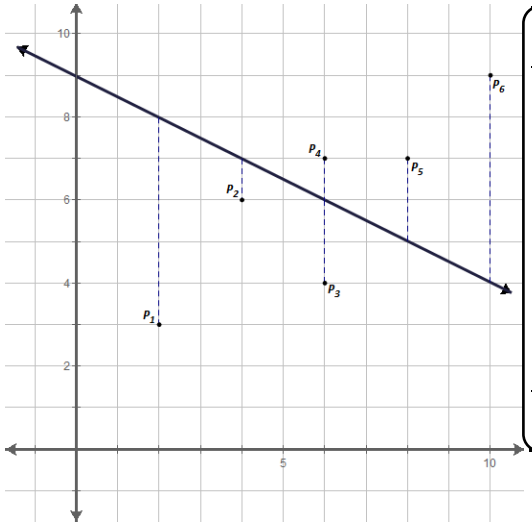
Data Point	*Residual*
P_1	
P_2	
P_3	
P_4	
P_5	
P_6	
Sum of Residuals	

TREND LINE 3



Data Point	*Residual*
P_1	
P_2	
P_3	
P_4	
P_5	
P_6	
Sum of Residuals	

TREND LINE 4



Data Point	*Residual*
P_1	
P_2	
P_3	
P_4	
P_5	
P_6	
Sum of Residuals	

6. What do all 4 trend lies above have in common? *(optional: what is the approximate residual of your trend line from earlier)*

7. 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?

TREND LINE 1

Data Point	*Residual*	*Residual Squared*
P_1	1	1
P_2	2	4
P_3	−2	4
P_4		
P_5		
P_6		
Sum		

TREND LINE 2

Data Point	*Residual*	*Residual Squared*
P_1		
P_2		
P_3		
P_4		
P_5		
P_6		
Sum		

TREND LINE 3

Data Point	*Residual*	*Residual Squared*
P_1		
P_2		
P_3		
P_4		
P_5		
P_6		
Sum		

TREND LINE 4

Data Point	*Residual*	*Residual Squared*
P_1		
P_2		
P_3		
P_4		
P_5		
P_6		
Sum		

8. The line that minimizes the squares is called the LEAST SQUARES REGRESSION LINE. Most scientific calculators are capable of determining the equation of this trend line. Consider again the data about the cars. **The following are the directions for the TI-83/84:**

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

1) First, it will be helpful to turn on additional diagnostic information in your calculator.

CATALOG — SCROLL DOWN TO DianosticOn

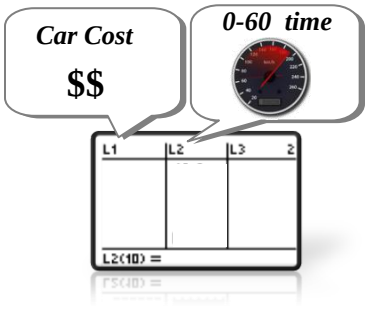
2nd 0 ⌄ ⌄ ..... ⌄ ⌄ ENTER ENTER

2) Under the Stat menu, press STAT 5 ENTER.
(This just resets the list menus)

3) Next, press STAT ENTER

4) If there is OLD data already in the lists that needs to be cleared press the up arrow, ⌃, to highlight L1 and then press CLEAR ENTER to clear out the old data. Do the same for L2 if it has OLD data that needs to be cleared.

5) Next, enter the Cost of the Car in L1 and the 0-60 mph time in L2.



6) Return to the home screen by pressing 2nd MODE and then to calculate the linear regression press STAT ⌄ 4 ENTER.

7) This represents the an equation of a line that minimizes the total residuals squared.

Fill in the blanks to complete the LEAST SQUARES REGRESSION LINE equation.

$$y = \underset{a}{\_\_\_\_\_} \cdot x + \underset{b}{\_\_\_\_\_}$$

Use this equation to reattempt your prediction of how fast a car can go from 0-60mph that costs $120 K

$$y = \underset{a}{\_\_\_\_\_} \cdot (120) + \underset{b}{\_\_\_\_\_} =$$

9. When a prediction is made between two given data points the prediction is called an **INTERPOLATION**. When a prediction is made outside the range of given data points the prediction is an **EXTRAPOLATION**. Which type of prediction was used when you predicted the 0 – 60 mph time of a car that cost $120 K?

10. A calculation called the **correlation coefficient (r)** is used to measure the extent to which the data for the two variables show a linear relationship. The closer the value is to 1 or –1 the stronger the linear relationship. Describe the relationship of the car data.

r: