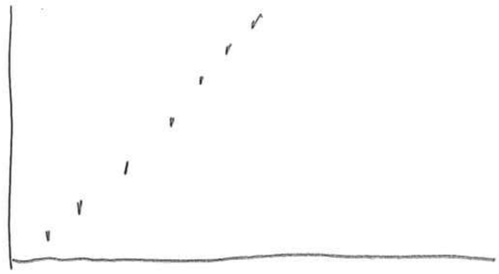
TEST
PRACTICE WKSHT.
~~Quiz~~ 3.2A AP Statistics Name: TEACHER

The AP Statistics exam was first administered in May 1997 to the largest first-year group in any discipline in the AP program. Since that time, the number of students taking the exam has grown at an impressive rate. Here are the actual data. Begin by entering them into your calculator lists.

Year		**Number of students**
1997	1	7,667
1998	2	15,486
1999	3	25,240
2000	4	34,118
2001	5	40,259
2002	6	49,824
2003	7	58,230
2004	8	65,878
2005	9	76,786

1. Use your calculator to construct a scatterplot of these data using 1997 as Year 1, 1998 as Year 2, etc. Describe what you see.



Data pretty much falls in line
Linear Pattern

2. Find the equation of the least-squares line on your calculator. Record the equation below. Be sure to define any variables used.

$\hat{y} = -946.17 + 8488.97x$ or #students $= -946.17 + 8488.97$(year since 1996)

$r = .999$ $r^2 = .998$

3. Interpret the slope of the least-squares line in context.

$\frac{y}{x}$ $\frac{\text{student}}{\text{time}}$ students taking AP increase at a rate of about 8490 per year

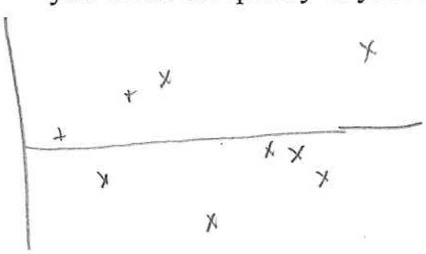
4. How many students would you predict took the AP Statistics exam in 2006? Show your method.

$x = 10$

$\hat{y} = -946.17 + 8488.97(10)$ $\hat{y} \approx 83943.5$ or 83,943 people

$y - \hat{y}$

5. Construct a residual plot. Sketch it in the space below. Comment on what the residual plot tells you about the quality of your linear model.



no pattern (fanning)
even above + below zero
Linear model good fit

6. Interpret the value of r^2 from your calculator in the context of this problem.

$r^2 = .998$ or 99.8% of the variation in the number of students taking the AP Stat exam is accounted for by the Linear relationship as describeb by the LSRL

7. Does there appear to be any outliers or influential Data? If so, explain No. Points are all on the line (almost perfectly)

Are jet skis dangerous? Propelled by a stream of pressurized water, jet skis and other so-ca bikes carry from one to three people, retail for an average price of $5700, and have become most popular types of recreational vehicle sold today. But critics say that they're noisy, dar damaging to the environment. An article in the August 1997 issue of the *Journal of the Am Medical Association* reported on a survey that tracked emergency room visits at randomly s hospitals nationwide. The study recorded data on the number of jet skis in use and the nun accidents for the years 1987–1996. Computer output and a residual plot from a linear regre analysis of the data are shown below.

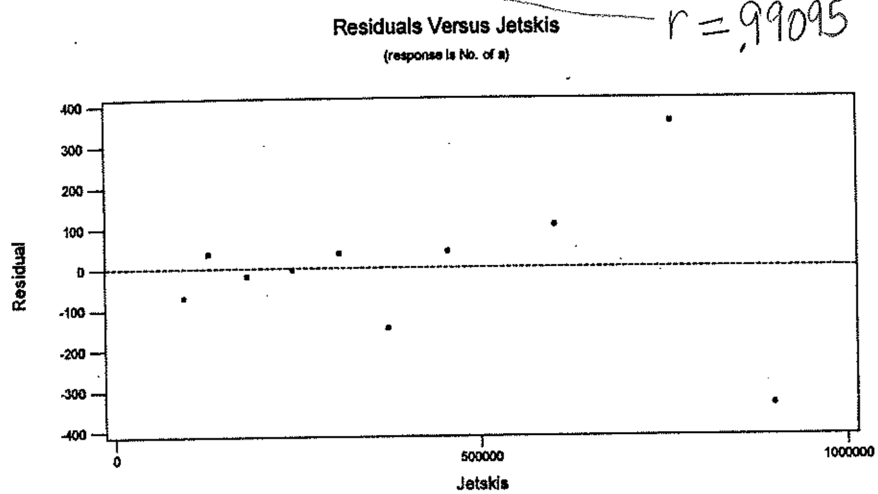
Predictor	Coef	SE Coef	T	P
Constant	-0.8	109.9	-0.01	0.994
Jetskis	0.0048308	0.0002292	21.08	0.000

S = 188.3   R-Sq = 98.2%   R-Sq(adj) = 98.0%

$r = .99095$

$a = \bar{y} - b(\bar{x})$

$b = r\left(\frac{s_y}{s_x}\right)$



1. What is the equation of the least-squares line? Be sure to define any variables you use.

$\hat{y} = -.8 + .0048308x$ or $\widehat{\#ACCD} = -.8 + .0048308(\#\text{jet ski})$

2. Interpret the value of r^2 in the context of this problem.

98.2% of variation in the # accidents can be explained by the linear relationship w/ # jet skis described by the LSRL

3. Is a line an appropriate model for these data? Justify your answer.

A line is appropriate. The correlation $r = .99095$ close to 1 and the correlation of determination is 98.2% → close to 100% so LSRL is good fit. The residual plot shows random points which indicates Linearity.

~~4. Interpret the value of s in the context of this problem.~~

★ Concerned about the two high residuals → indicating the line is more reliable predictor for lower x values than higher x values.

At a conference for high school teachers, some data were collected on the age and current i each teacher's primary vehicle. The data were entered into a TI-84. Here is some output fr two-variable statistics on the data.

```
2-Var Stats
 x̄=5
 Σx=150
 Σx²=966
 Sx=2.729152957
 σx=2.683281573
↓n=30
```

```
2-Var Stats
↑ȳ=60939.83333
 Σy=1828195
 Σy²=1.74171E11
 Sy=46520.87632
 σy=45738.95716
↓Σxy=12371116
```

$a = \bar{y} - b(\bar{x})$

$b = r\left(\frac{S_y}{S_x}\right)$

The correlation between age and mileage for these data is 0.877.

1. Calculate the equation of the least-squares line for predicting a high school teacher's ve mileage from its age. Show your work.

$\hat{y} = -13806.467 + 14949.26x$

$\hat{y}$ = vehicle mileage x = vehicle age

$b = .877\left(\frac{46520.87632}{2.729152957}\right)$

$b = 14949.26$

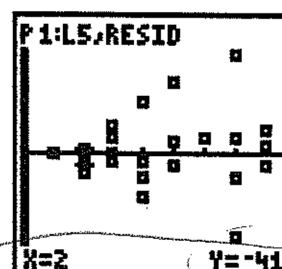
$a = 60939.833 - (14949.26)(5)$

$= -13806.467$

2. Interpret the slope of the regression line in context.

Each year the mileage increases by 14949.26 miles.

Here is a residual plot from this linear regression analysis.



Fans

-4164.764

Highlighted Point.

3. Describe what the residual plot tells you about how well the linear model fits the data.

It Fits the model pretty well. Points split equally above + below 0 + Random

$y - \hat{y}$.

4. What is the actual vehicle mileage for the teacher whose point is highlighted in the plot?

$\hat{y} = -13806.467 + 14949.26(2) = 16092.053 = \hat{y}$

$resid = y - \hat{y} \Rightarrow -4164.764 = y - 16092.053 \Rightarrow y = \boxed{11927.289 \text{ Actual}}$

5. State and interpret the value of r^2 in the context of this problem.

$r^2 = (.877)^2 = .7691 \Rightarrow$ 76.91% of the variation in mileage is explained by the LSRL for the relationship between mileage + age of car

6. The residual plot appears to Fan out as X increases.

What does this mean for our LSRL and data analysis / prediction.

Less Reliability in the prediction values as x increases