

Activity 3.2A -
INTRO TO REGRESSION

AP Statistics

Name: **KEY**

The table and scatterplot below show the relationship between student enrollment (in thousands) and total number of property crimes (burglary and theft) in 2006 for eight colleges and universities in a certain U.S. state.

Enrollment (in 1000s) (x)	No. of Property Crimes (y)	Predicted Values (units) \hat{y} crime	Residual Values (y - \hat{y})
16	201	Do in Step I	Do as part of question 4
2	6		
9	42		
10	141		
14	138		
26	601		
21	230		
19	294		

INSTRUCTIONS:

A) Identify your variables:

EXPLANATORY is **STUDENT ENROLLMENT**

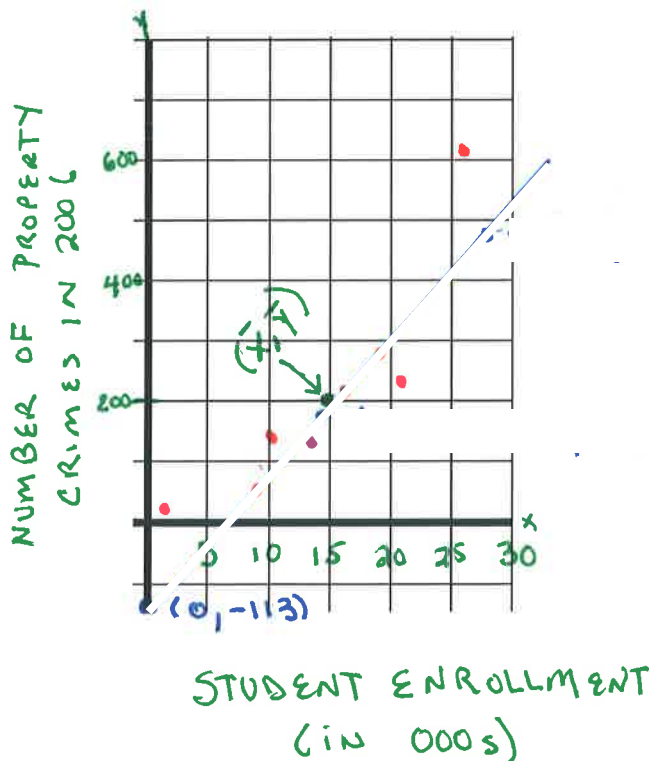
RESPONSE is **PROPERTY CRIME**

B) Create a scatter plot and describe the association in context

C) Find the mean and standard deviation for both variables. Use [STAT] 2-var stats

ENROLLMENT $\bar{x} = 14.63$ $S_x = 7.60$
CRIMES $\bar{y} = 206.63$ $S_y = 185.21$

D) Plot (Xmean, Ymean) on the scatter plot with a star. $(\bar{x}, \bar{y}) = (14.63, 206.63)$



← The scatter plot shows a positive, **linear**, moderately strong association between student enrollment and number of property crimes in 2006.

Activity 3.2A -
INTRO TO REGRESSION

AP Statistics

Name:

KEY

The table and scatterplot below show the relationship between student enrollment (in thousands) and total number of property crimes (burglary and theft) in 2006 for eight colleges and universities in a certain U.S. state.

Enrollment (in 1000s) (x)	No. of Property Crimes (y)	Predicted Values (units) \hat{y} crime	Residual Values (y - \hat{y})
16	201	236	Do as part of question 4
2	6	-68	
9	42	84	
10	141	106	
14	138	193	
26	601	455	
21	230	346	
19	294	302	

INSTRUCTIONS:

A) Identify your variables:

EXPLANATORY is STUDENT ENROLLMENT

RESPONSE is PROPERTY CRIME

B) Create a scatter plot and describe the association in context

C) Find the mean and standard deviation for both variables. Use [STAT] 2-var stats

ENROLLMENT $\bar{x} = 14.63$ $S_x = 7.60$

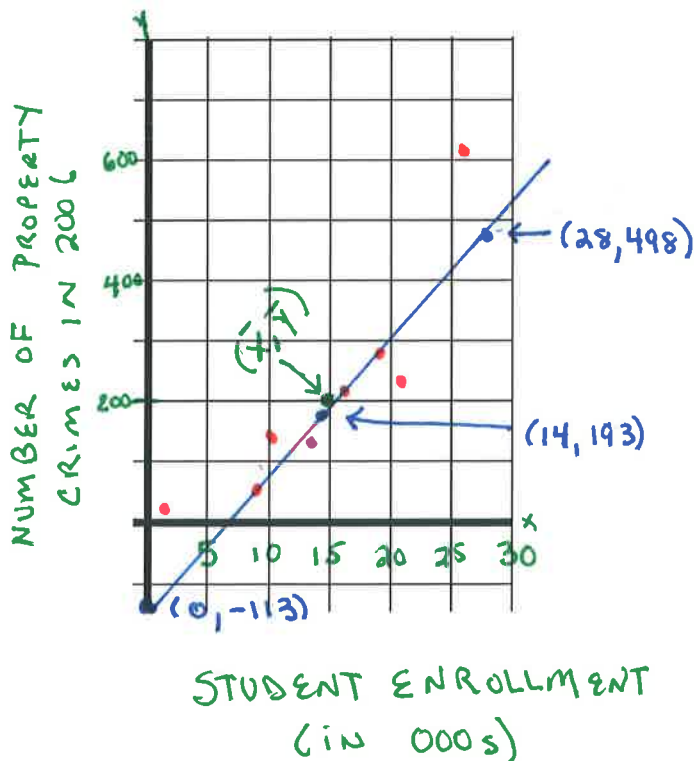
CRIMES $\bar{y} = 206.63$ $S_y = 185.21$

D) Plot (Xmean, Ymean) on the scatter plot with a star. $(\bar{x}, \bar{y}) = (14.63, 206.63)$

E) Find the correlation coefficient $r = .895$. Does this make sense? Explain.

YES. IT CONFIRMS WHAT WE SAW IN THE SCATTER PLOT

Now go to the next page and answer questions...



The scatter plot shows a positive, linear, moderately strong association between student enrollment and number of property crimes in 2006.

NOTICE (\bar{x}, \bar{y}) IS ON THE LSRL.

$$\hat{y} = a + bx$$

where \hat{y} read "y hat"
 $a = y$ intercept
 $b = \text{slope}$

INSTRUCTIONS:

F) Find the LSRL

The LSRL equation: $\text{CRIME} = -112.58 + 21.8 (\text{ENROLLMENT})$

where

$\hat{y} =$ PREDICTED NUMBER OF PROPERTY CRIMES

$x =$ STUDENT ENROLLMENT IN THOUSANDS

G) Put the equation in [y=]

H) Pick 2 points from the [table] and plot the LSRL on the scatter plot.

I) Fill in the predicted values in the table on the front page. The [table] has the predicted values. Look up each x-value for Enrollment in the [table] and the y-value is the predicted crimes.

1. Interpret the slope of the least-squares line in the context of the problem.

$$b = 21.8$$

FOR EACH 1,000-STUDENT INCREASE IN ENROLLMENT, THE PREDICTED NUMBER OF PROPERTY CRIMES INCREASES BY 21.83.

2. How many crimes would you predict on a campus with enrollment of 14 thousand students? Show your work.

$$x = 14$$

$$\text{Crime} = -112.58 + 21.83(14) = 193.04 \text{ crimes}$$

3. Find the residual for the campus with 14 thousand students and 138 property crimes. Show your work. Interpret the value of the residual in the context of the problem.

$$\text{Residual} = y - \hat{y} = 138 - 193.04 = -55.04$$

$$x = 14 \\ y = 138$$

$$\hat{y} = \text{crime} = 193.04$$

The actual number of property crimes at this college is 55.04 crimes fewer than the number of crimes predicted by this linear model.

Activity 3.2A -
INTRO TO REGRESSION

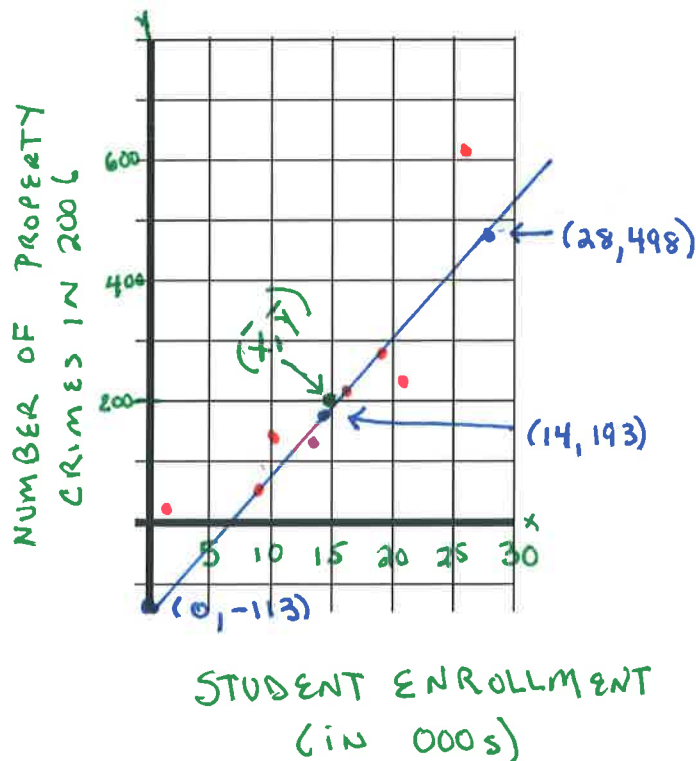
AP Statistics

Name: **KEY**

The table and scatterplot below show the relationship between student enrollment (in thousands) and total number of property crimes (burglary and theft) in 2006 for eight colleges and universities in a certain U.S. state.

Enrollment (in 1000s) (x)	No. of Property Crimes (y)	Predicted Values (units) \hat{y} <i>crime</i>	Residual Values ($y - \hat{y}$)
16	201	236	-35
2	6	-68	74
9	42	84	-42
10	141	106	35
14	138	193	-55
26	601	455	146
21	230	346	-116
19	294	302	-8

Question 4 - ADD RESIDUALS

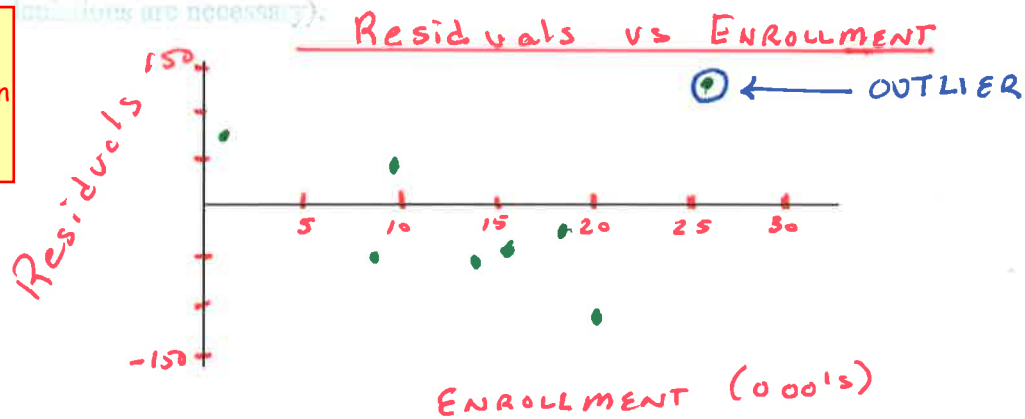


The scatter plot shows a positive, linear, moderately strong association between student enrollment and number of property crimes in 2006.

NOTICE (\bar{x}, \bar{y}) IS ON THE LSRL.

4. Use the scatterplot to make a rough sketch of the residual plot for these data.

Q4:
Calculate the residuals ($y - \hat{y}$) in the table on the front page.



5. Would the slope of the regression line change if the point (26, 601) were removed from the data set? In what direction?

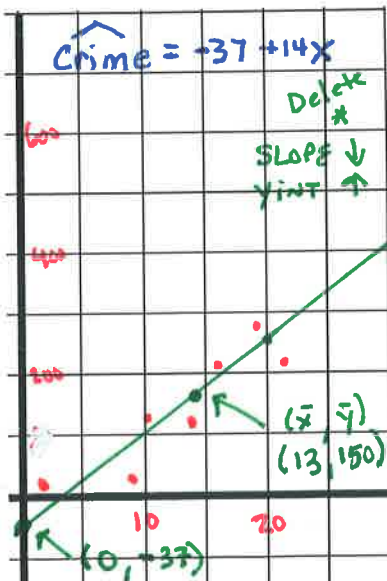
YES. FROM THE RESIDUAL PLOT YOU CAN SEE (26, 601) HAS A HIGH POSITIVE RESIDUAL SO IT TENDS TO PULL THE LINE TOWARDS IT. REMOVING IT WOULD REDUCE THE SLOPE.

NOW REMOVE THE POINT AND SEE THE NEW LSAL (see below)
 $r = .923$ increased and r^2 increased too.

6. The value of r^2 for these data is 0.801. Interpret this value in the context of the problem.

80.1% OF THE VARIATION IN PROPERTY CRIMES CAN BE ACCOUNTED FOR BY THE REGRESSION OF PROPERTY CRIME ON ENROLLMENT.

7. Is the given least-squares regression line a good model for these data? Support your answer with appropriate evidence from your answers above.

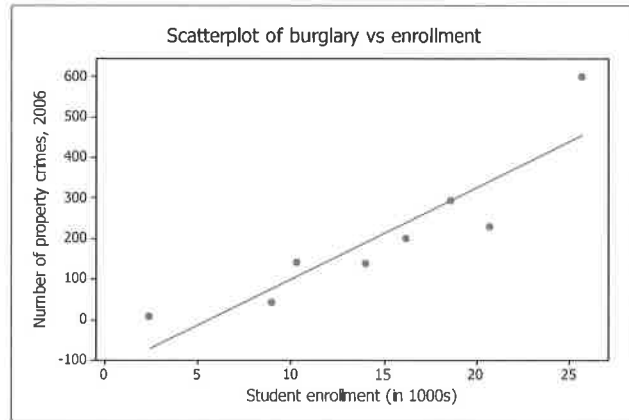


① 80.1% OF THE VARIATION IS ACCOUNTED BY THE MODEL

② THE APPROPRIATENESS OF THE LINEAR MODEL IS BASED ON REVIEWING THE RESIDUAL PLOT BECAUSE THE RESIDUAL PLOT SHOWS NO PATTERN.

The table and scatterplot below show the relationship between student enrollment (in thousands) and total number of property crimes (burglary and theft) in 2006 for eight colleges and universities in a certain U.S. state.

Enrollment (in 1000s) (x)	No. of Property Crimes (y)
16	201
2	6
9	42
10	141
14	138
26	601
21	230
19	294



The equation of the least-squares regression line is $\hat{y} = -112.58 + 21.83x$, where \hat{y} = predicted number of property crimes and x = student enrollment in thousands.

- Interpret the slope of the least-squares line in the context of the problem.
- How many crimes would you predict on a campus with enrollment of 14 thousand students? Show your work.
- Find the residual for the campus with 14 thousand students and 138 property crimes. Show your work. Interpret the value of the residual in the context of the problem.

Quiz 3.2A

1. For each 1000-student increase in enrollment, the predicted number of property crimes increases by 21.83. 2. $\hat{y} = -112.58 + 21.83(14) = 193.04$ crimes.

3. $y - \hat{y} = 138 - 193.04 = -55.04$. The actual number of property crimes at this college is 55.04 fewer than the number of crimes predicted by this linear model. 4. See graph below. 5. That point has a high, positive residual, so it tends to “pull” the line toward it. Removing it would reduce the slope. 6. 80.1% of the variation in property crimes can be accounted for by the regression of property crimes on enrollment. 7. Answers will vary and will depend on the appearance residual plot sketched in #4. Some students will suggest that there is enough of a “U” shape in the residual plot to suggest that the linear model is not appropriate. Others will say the lack of pattern in residuals suggests that a linear fit is appropriate.

