AP Statistics Problem Set #5

1. **The SAT Essay: Is Longer better?** Following the debut of the new SAT Writing test in march 2005, Dr. Les Perelman from MIT stirred controversy by reporting, "It appeared to me that regardless of what a student wrote, the longer the essay, the higher the score." He went on to say, "I have never felt a quantifiable predictor in 25 years of grading that was anywhere as strong as this one. If you just graded them based on length without ever reading them, you'd be right over 90 percent of the time." The following are the data summaries: The mean number of words written is 275.7419 words with a standard deviation of 139.49 words. The mean score based on the words written is 4.0323 with a standard deviation of 1.6428. The correlation coefficient between the number of words written and the score is 0.8805.

a) Write the equation of the regression line, in context.
b) Explain what the intercept of the regression line indicates.
c) Interpret the slope of the regression line.
d) Predict the score for someone who writes 420 words.
e) Based upon these statistics, how effective do you think the prediction of the score on the written essay is based on the number of words written in the essay? Explain.
f) As a student, would you rather have a positive or negative residual in this context? Explain.
g) What number of words written would a student have written if they received a score of a 5?

2. **ARMSPAN** Leonardo da Vinci, the renowned painter, speculated that an ideal human would have an armspan (distance from the outstretched fingertip of the left hand to the outstretched fingertip of the right hand) that was equal to his height. The following computer regression printout shows the results of a least-squares regression of armspan on height, both in inches, for a sample of 18 high school students.

Predictor	Coef.	SE Coef	T	P
Constant	11.5474	5.6	2.06	0.0558
Height	0.84024	0.08091	10.4	0.000

R-sq = 87.1% R-sq(adj.) = 86.3%

a) What is the least squares regression line?
b) Predict the height of a person with an armspan of 72 inches.

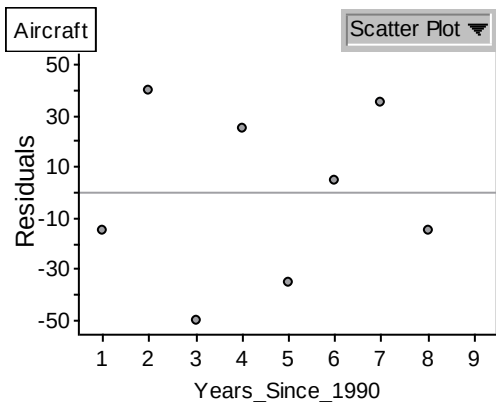
3. **A GROWING CHILD** Sarah's parents are concerned that she seems short for her age. Their doctor has the following record of Sarah's height:

Age (in months):	36	48	51	54	57	60
Height (cm):	86	90	91	93	94	95

(a) Make a scatterplot of these data. Describe the pattern.
(b) Using your calculator, find the equation of the least-squares regression line of height on age.
(c) Find the residuals for each height and make a residual plot of the 6 values.
(d) Based on the scatterplot, the residual plot, and r-square, is the LSRL in part (b) a good linear model?
(e) What is Sarah's rate of growth, in centimeters per month? Normally growing girls gain about 6 cm in height between the ages 4 (48 months) and 5 (60 months). What rate of growth is this in centimeters per month? Is Sarah growing more slowly than normal? Explain.

5. **NUMBER OF AIRCRAFTS** Lydia and Bob were searching the Internet to find information on air travel in the United States. They found data on the number of commercial aircraft flying in the United States during the years 1990-1998. The dates were recorded as years since 1990. Thus, the year 1990 was recorded as year 0. They fit a least squares regression line to the data. The graph of the residuals and part of the computer output for their regression are given below.

$\hat{y} = 2939.93 + 233.517$
$r = 0.88$



(a) Is a line an appropriate model to use for these data? What information tells you this?

(b) What is the value of the slope of the least squares regression line? Interpret the slope in the context of this situation.

(c) What is the value of the intercept of the least squares regression line? Interpret the intercept in the context of this situation.

(d) What is the predicted number of commercial aircraft flying in 1992?

(e) What was the actual number of commercial aircraft flying in 1992?