Test 12A AP Statistics Name:

Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

Use the following for questions 1 – 3.

A statistically-minded English teacher wonders if she can predict the lengths of essays that her students submit on the basis of how large the computer files for the essays are. She selects a random sample of 13 student papers (all produced by the same word-processing software) and compares file size (in kilobytes or KB) to the word count for each. A computer regression analysis of her data is given below. Assume all conditions for inference for slope have been met.

```
Predictor      Coef  SE Coef      T      P
Constant   -1450.3    773.4  -1.88  0.088
File Size    76.50    28.60   2.67  0.022

S = 145.851   R-Sq = 39.4%   R-Sq(adj) = 33.9%
```

1. Which of the following is the estimate (from this sample) of the standard deviation of the sampling distribution of *b*, the slope of the sample regression line for this relationship?

(a) 773.4 (b) 145.851 (c) $\frac{773.4}{\sqrt{13}}$ (d) 28.60 (e) $\frac{28.60}{\sqrt{13}}$

2. Which of the following is the best interpretation of the quantity S = 145.851?

(a) The average distance between the file size of each of these essays and the mean file size for all the essays in this sample is about 146 KB.
(b) The sum of the squared deviations between each observed file size and the file size predicted by the regression equation is 146 KB.
(c) The average distance between the words counts of each of these essays and the mean word count for all the essays in this sample is about 146 KB.
(d) The sum of the squared deviations between each observed word count and the word count predicted by the regression equation is 146 KB.
(e) Predictions of word count from file size based on this regression model will be off by an average of about 146 KB.

3. If the teacher uses these data to test the hypotheses $H_0 : \beta = 0$ *versus* $H_a : \beta \neq 0$ at the α = 0.05 level, which of the following is an appropriate conclusion?

(a) Since the *P*-value of 0.022 is less than α, reject H_o. We have convincing evidence that there is a linear relationship between file size and word count.
(b) Since the *P*-value of 0.022 is less than α, fail to reject H_o. We do not have convincing evidence that there is a linear relationship between file size and word count.
(c) Since the *P*-value of 0.088 is greater than α, reject H_o. We have convincing evidence that there is a linear relationship between file size and word count.
(d) Since the *P*-value of 0.088 is greater than α, fail to reject H_o. We do not have convincing evidence that there is a linear relationship between file size and word count.
(e) Since the *P*-value of 0.088 is greater than α, accept H_o. We have convincing evidence that the regression line relating word count to file size has a slope of 0.

4. Which of the following conditions must be satisfied in order to perform inference for regression of y on x?

I. The population of values of the independent variable (x) must be normally distributed.

II. The standard deviation of the population of y-values for a given value of x is the same for every x-value.

III. There is a linear relationship between x and the mean value of y for each value of x.

(a) I only
(b) II only
(c) I and III
(d) II and III
(e) All three must be satisfied.

Use the following for questions 5 – 6:

A random sample of 40 companies on the Forbes 500 list was selected and the relationship between sales (in hundreds of thousands of dollars) and profits (in hundreds of thousands of dollars) was investigated using regression. A least-squares regression line was fitted to the data using statistical software, with sales as the explanatory variable and profits as the response variable. Here is the output from the software:

```
Dependent variable is Profits
R-squared = 66.2%
s = 466.2 with 40 - 2 = 38 degrees of freedom

Variable    Coefficient  s.e. of Coefficient        P-value
Constant      -176.644          61.16          0.0050
Sales   0.092498          0.0106            ≤0.0001
```

5. Which of the following is an appropriate interpretation of the number 0.092498?

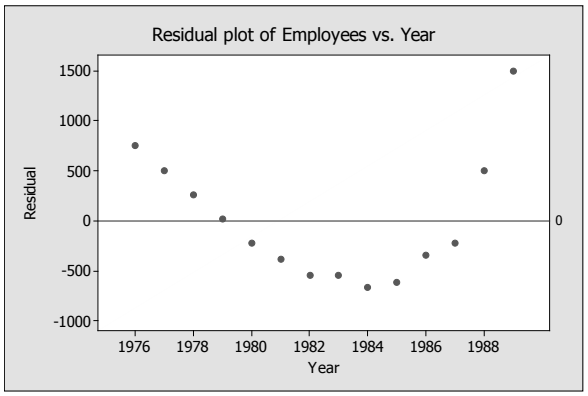
(a) An increase in profits of $100,000 increases predicted sales by $9,249.80.
(b) An increase in sales of $100,000 increases predicted profits by $9,249.80.
(c) Sales of $100,000 correspond to predicted profits of $9,249.80.
(d) The ratio of predicted profit to observed profit is, on average, 0.092498.
(e) The ratio of predicted profits to observed sales is, on average, 0.092498.

6. Which of the following expressions best represents the margin of error of a 90% confidence interval for the slope of the population regression line?

(a) $1.686\sqrt{0.0106}$
(b) $1.686\left(\frac{0.0106}{\sqrt{38}}\right)$
(c) $1.686(0.0106)$
(d) $1.645\sqrt{0.0106}$
(e) $1.645\left(\frac{0.0106}{\sqrt{38}}\right)$

Use the following for questions 7 – 9:

A least squares regression analysis of the number of employees at Microsoft *versus* Year (from 1976 through 1989) produced the following residual plot.



7. Based on this residual plot, which of the following statements are true?

I. The relationship between Number of employees and Year is non-linear.
II. This regression equation would overestimate the number of employees at Microsoft in 1982.
III The number of employees at Microsoft decreased from 1976 through 1982.

(a) I only
(b) II only
(c) III only
(d) I and II are true.
(e) None of these statements is true.

8. The scatterplot of Log (Number of employees) *versus* Year shows a strong, positive, linear pattern. Which of the following conclusions can be drawn from this information?

(a) The scatterplot of Log (Number of employees) *versus* Log (Year) will also be strong, positive, and linear.
(b) The relationship between Number of employees and Year can be modeled well by an exponential function.
(c) The relationship between Number of employees and Year can be modeled well by a power function.
(d) The residual plot for the regression of Log (Number of employees) on Year will show a curved pattern similar to the one shown above.
(e) The number of employees at Microsoft in 1994 can be predicted accurately by the linear regression of Log (Number of employees) on Year from these data.

9. Below is the computer regression analysis of the relationship between Log (Number of employees) at Microsoft and Year.

```
Predictor      Coef   SE Coef       T      P
Constant    -449.71     18.91  -23.78  0.000
Year       0.228005  0.009540   23.90  0.000

S = 0.143896   R-Sq = 97.9%   R-Sq(adj) = 97.8%
```

Which of the following is the correct regression equation from this analysis?

(a) $\hat{y} = -449.7 + 0.228x$

(b) $\widehat{\log y} = -449.7 + 0.228x$

(c) $\widehat{\log y} = -449.7 + 0.228(\log x)$

(d) $\widehat{\log y} = 0.228 - 449.7x$

(e) $\widehat{\log y} = 0.228 - 449.7(\log x)$

10. A fisheries research report gives the following regression equation for the relationship between the length (*L)* in cm. and weight (*W*) in grams, of the gracile lizardfish, a small marine fish that lives in the Indian Ocean:

$$\widehat{\ln W} = -5.36 + 3.216(\ln L).$$

Which of the following is closest to the predicted weight of a gracile lizardfish that was 12 cm long, based on this model?

(a) 2.63 grams
(b) 3.50 grams
(c) 13.89 grams
(d) 33.23 grams
(e) 426.58 grams

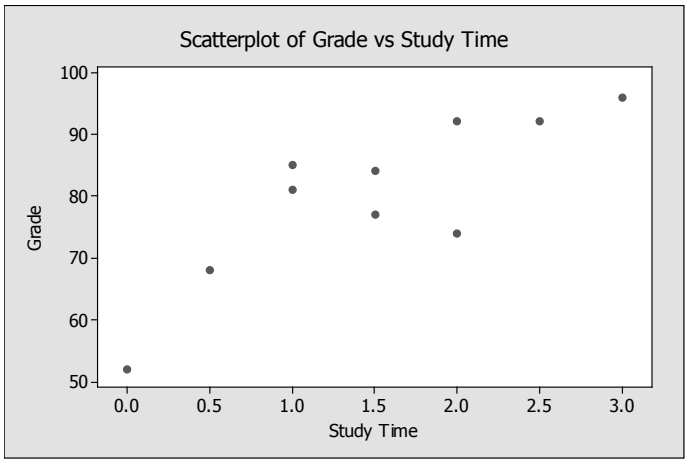
Part 2: Free Response

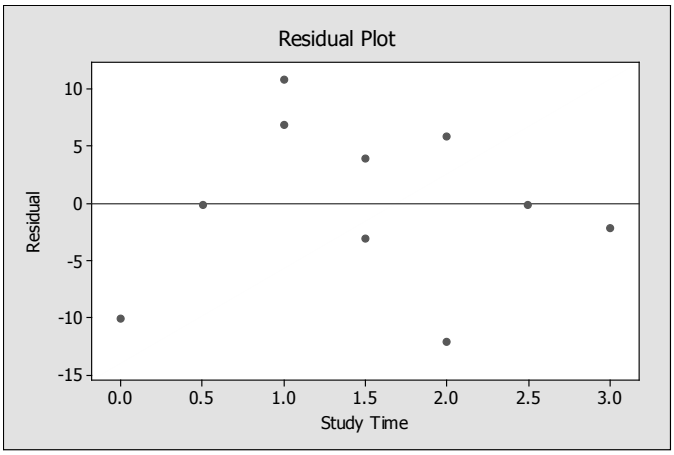
Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

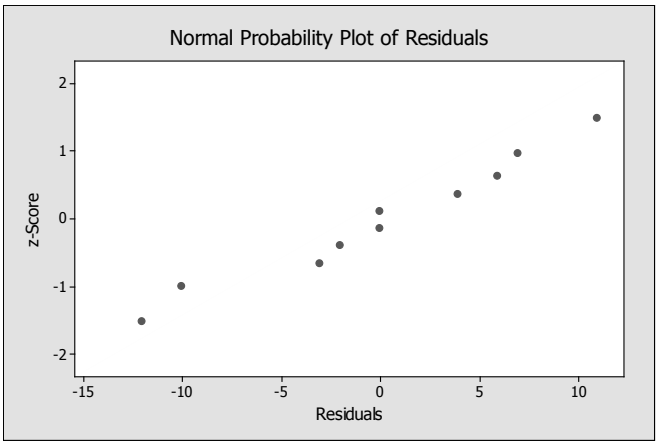
11. A college teacher asked a random sample of 10 of the 250 students in her introductory statistics class to record the total amount of time (in hours) they spent studying for a particular test and then combined these times with the students' scores on the test. She then performed a regression analysis on the data. Below is numerical and graphical output from her computer software.

Predictor	Coef	SE Coef	T	P
Constant	62.100	4.893	12.69	0.000
Study Time	12.000	2.825	4.25	0.003

S = 7.73709 R-Sq = 69.3% R-Sq(adj) = 65.4%







(a) Use the computer output to discuss whether the conditions for regression inference have been met. If you do not have enough information to check a condition, describe what further information would be required.

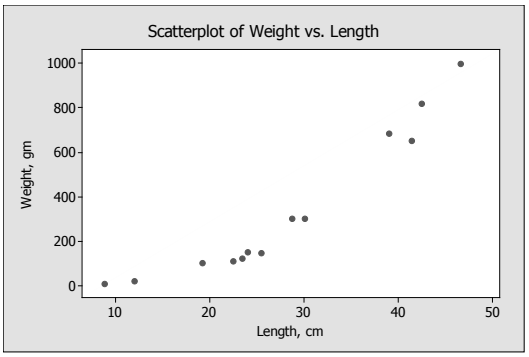
For the remaining questions using these data, assume that the conditions for inference have been satisfied.

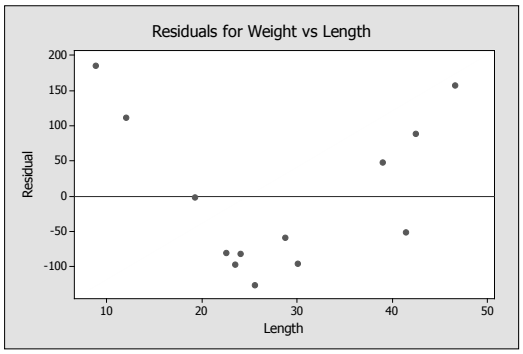
(b) Do these data provide convincing evidence that there is a linear relationship between time spent studying and test score? Perform the appropriate significance test to support your conclusion.

(c) Construct and interpret a 95% confidence interval for the slope of the population regression line for predicting test score from study time.

(d) Can the teacher tell her students that her data establishes that studying more will improve their grades? Why or why not?

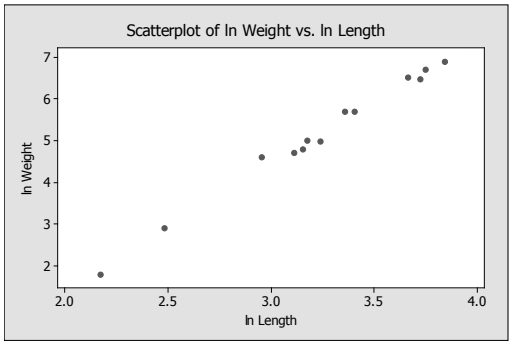
12. A fisheries biologist wants to predict the weight (in grams) of perch (a type of fish) caught in a certain lake from their length (in cm). He catches, measures, and weights 13 perch whose lengths were between 8 and 48 cm. Below is a scatterplot of his data, along with the residual plot from a linear regression analysis.





(a) Is a linear model appropriate for these data? Justify your answer.

(b) Below is a scatterplot of the natural logarithm of weight *vs.* the natural logarithm of length. This relationship is clearly more linear that the one above. Does this suggest that the relationship between length and weight can be modeled by an exponential function or by a power function? Explain.



(c) Computer output from the regression of ln (Weight) *vs.* ln (Length) is given below. Use it to predict the weight of a fish that is 75 cm long.

```
Predictor      Coef  SE Coef       T      P
Constant   -4.7017   0.2880  -16.33  0.000
ln Length  3.04223  0.08812   34.52  0.000

S = 0.149747   R-Sq = 99.1%   R-Sq(adj) = 99.0%
```

(d) Comment on the reliability of the prediction made in part (c).