Chapter 12 Practice Test – AP Statistics Name Cohen

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

The following information is used in Questions 1 - 3.
A random sample of 80 companies from 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 by 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 squares = 66.2%
s = 466.2 with 80 - 2 = 78 degrees of freedom

Variable      Coefficient        S.E. of Coef       P-value
Constant      -176.644           61.16              0.0050
Sales         0.092498           0.0075             ≤0.0001
```

1. Using the above data, approximately what is a 90% confidence interval for the slope of the least-squares regression line?
(a) 0.0925 ± 0.0075
B (b) 0.0925 ± 0.012
(c) –0.0925 ± 0.0075
(d) –0.0925 ± 0.012
(e) None of the above.

$b \pm t^*(SE_b)$ $d.f = n-2$

$0.0925 \pm 1.671(.0075)$

2. Using the above data, what is the value of the *t* statistic for testing whether the slope of the least-squares regression line is 0?
(a) 0.0075
D (b) 0.082
(c) 0.092
(d) 12.33
(e) None of the above.

$t = \frac{b}{SE_b} = \frac{.0925}{.0075}$

3. Using the above data, is there strong evidence (and if so, why) of a straight-line relationship between sales and profits?
(a) Yes, because the slope of the least-squares line is positive.
B (b) Yes, because the *P*-value for testing if the slope is 0 is quite small.
(c) No, because the value of the square of the correlation is relatively small.
(d) It is impossible to say because we are not given the actual value of the correlation.
(e) None of the above.

4. I measure a response variable Y at each of several times. A scatterplot of log *Y* versus time of measurement looks approximately like a positively sloping straight line. We may conclude that
(a) the correlation between time of measurement and *Y* is negative, since logarithms of positive fractions (such as correlations) are negative.
C (b) the rate of growth of *Y* is positive but slowing down over time.
(c) an exponential curve would approximately describe the relationship between *Y* and time.
(d) a power function would approximately describe the relationship between *Y* and time.
(e) A mistake has been made. It would have been better to plot log *Y* versus the logarithm of time.

The following information is used in Questions 5 to 8.
A marine biologist wants to test the effect of water temperature on the average dive duration for sea otters. Several otters are available for an experiment. The biologist collects the following data:

	Water temp.°(C)	Dive duration (sec)
Otter	x	y
J2	4	63
J1	8	75
B7	8	84
B9	12	91
M3	12	101
D4	16	110
B8	20	115

We want to determine if water temperature is useful in predicting dive duration. Here is output from Minitab for these data:

Predictor	Coef	Stdev	t-ratio	p
Constant	52.789	5.257	10.04	0
H2Otemp	3.3684	0.4216	***	**

s = 5.557 R-sq = 92.7% R-sq(adj) = 91.3%

5. An appropriate null hypothesis for a test would be "the slope of the true regression line is
(a) positive."
(b) 3.3684."
(c) s = 5.557."
(d) not zero."
(e) zero."

E

6. The equation for the least-squares regression line is
(a) $\hat{y}$ = 3.3684 + 52.789 x
(b) $\hat{y}$ = 52.789 + 5.557 x
(c) $\hat{y}$ = 52.78 x + 5.557
(d) $\hat{y}$ = 52.789 + 3.3684 x
(e) $\hat{y}$ = 5.257 + 0.4216 x

D

7. The t statistic for testing H_0 has been left out. From the output, the t-statistic has the value
(a) 7.99. (b) 10.04. (c) 0.124. (d) 0.927. (e) 15.67.

A

$t = \frac{b}{SE_b} = \frac{3.3684}{.4216}$

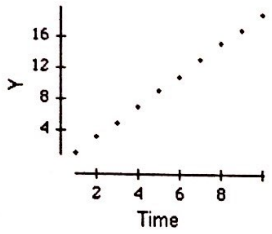
8. The P-value is
(a) less than 0.001.
(b) between 0.001 and 0.01.
(c) between 0.01 and 0.05.
(d) between 0.05 and 0.10.
(e) greater than 0.10.

A

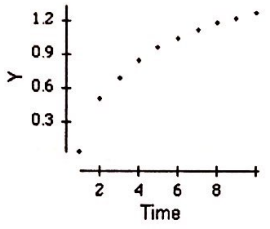
d.f = n-2

tcdf(7.99, 100000, 5)

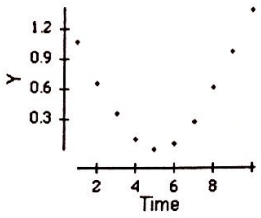
9. Which of the following scatterplots would indicate that Y is growing exponentially over time?



(a)

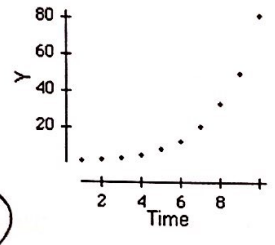


(b)



(c)

D



(d)

(e) none of these

Part 2: Free Response *Answer completely, but be concise. Write sequentially and show all steps.*

10. A mathematics professor wishes to analyze the relationship between the number of papers (in hundreds) graded by his department's student homework graders and the total amount of money paid to the graders. He collects data for 12 randomly chosen graders and uses MINITAB to do regression analysis. Below is a portion of the Minitab output. (Here, COST = amount paid (dollars), PAPERS = number of papers in hundreds, and the intervals listed at the bottom are computed for 1600 papers.)

```
The regression equation is      COST = 35.8 + 12.1 PAPERS

Predictor     Coef          Stdev         t-ratio       P
Constant      35.80         17.06         2.10          0.062
PAPERS        12.0835       0.9738        12.41         0.000

s = 6.526           R-sq = 93.9%        R-sq (adj) = 93.3%
```

(a) Formulate null and alternative hypotheses about the slope of the true regression line. Use a two-sided alternative.

$H_0: \beta = 0$

$H_a: \beta \neq 0$

(b) What is the least-squares regression equation?

$\hat{Cost} = 35.8 + 12.0835(papers)$

(c) What is the standard error about the line (also known as the standard deviation s in the regression model)? Interpret this value in context.

$S = 6.526$ The "typical" error in predicting cost to grading based on # of papers is about \$6.53

(d) Interpret the slope of the least-squares regression line in the context of this problem.

For every 100 papers graded, the amount of money paid to graders increases by about $12.08.

(e) The model for regression inference has three parameters: α, β, and σ. Estimate these parameters from the data.

α is estimated by $a = 35.80$ (y-intercept)

β is estimated by $b = 12.0835$ (slope)

σ is estimated by $s = 6.526$ (standard deviation of residuals)

(f) What is the value of the test statistic for testing the hypotheses?

$t = 12.41$

(g) How many degrees of freedom does *t* have?

$n - 2 = 12 - 2 = 10$

(h) What is the *P*-value for the test?

$P \approx 0.000$

(i) Is the number of papers graded useful for predicting the amount paid? Use a significance level of 0.01. Explain briefly.

Yes. Since the p-value of almost 0 is below any reasonable significance level, we reject the H_0. We have enough evidence of a relationship between number of papers graded and cost of grading.

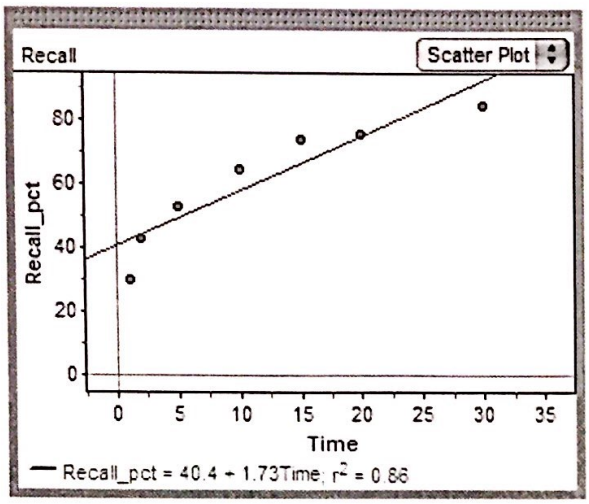
(j) What is the estimated cost of grading 1600 papers?

Since x is in 100s of paper use $x = 16$

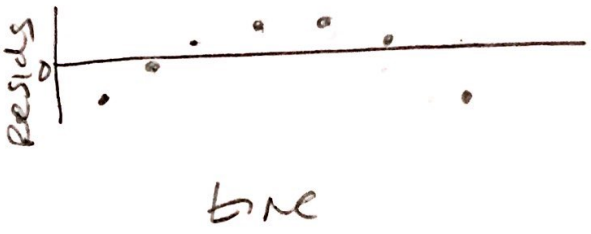
$\hat{Cost} = 35.8 + 12.1(16)$

$= \$229.40$ is the predicted cost of grading 1600 papers.

11. An experiment was conducted to determine the effect of practice time (in seconds) on the percent of unfamiliar words recalled. Here is a Fathom scatterplot of the results with a least-squares regression line superimposed.



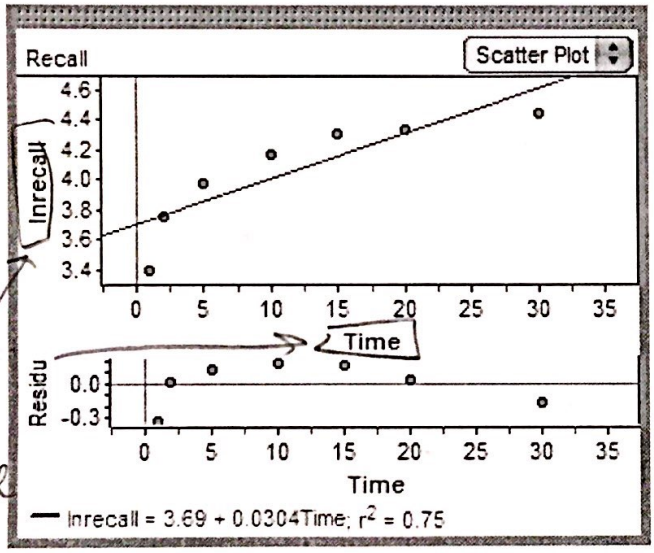
(a) Sketch a residual plot below.



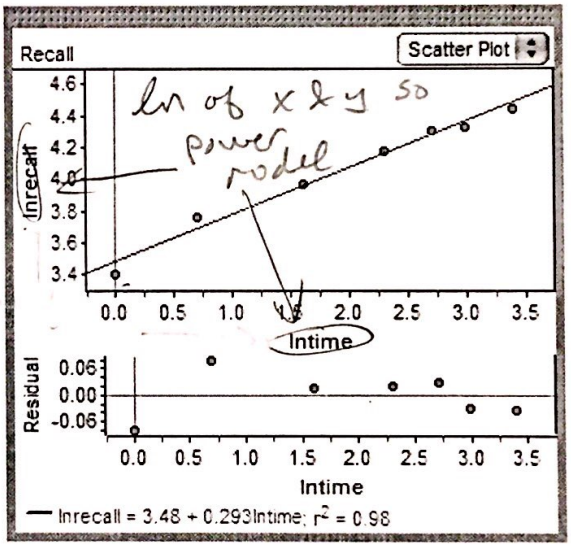
bne

(b) Does a linear model fit the data well? Justify your answer. No, the residual plot shows a clear curved pattern even though the r-value is strong at .93

We used Fathom to transform the original data in hopes of achieving linearity. The screen shots below show the results of two different transformations.



ln of only y → so exponential



(c) Which model would fit the original data better, an exponential model or a power model? Justify your answer.

The power model fits the data better. There is still a slight curve in the residual plot but not a much as the exponential model. In addition the r-value for the power model is .99 which is stronger than the exponential model which is .87.

(d) Use the model you chose in (c) to predict word recall for 25 seconds of practice. Show your method.

$$\ln(\hat{recall}) = 3.48 + 0.293 \ln(25)$$

$$\ln(\hat{recall}) = 4.423$$

$$\hat{recall} = e^{4.423}$$

$$= 83.35\%$$