Chapter 14 Review

Quiz 14.1C AP Statistics Name: Mr. Morton

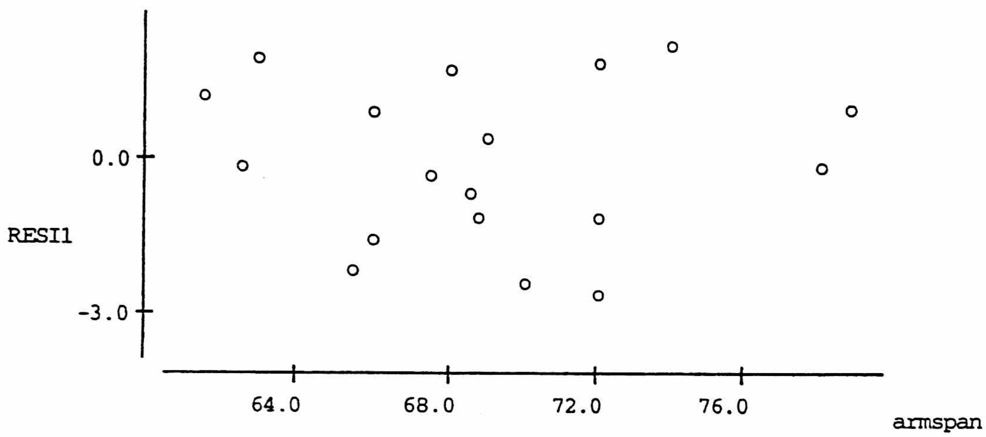
Ideal proportions Once upon a time, a class like yours made measurements of their arm span and height. They entered their results into a Minitab worksheet, requested least squares regression of height on arm span (both in inches) and obtained the following output:

```
Predictor        Coef        Stdev     t-ratio       p
Constant       11.547        5.600        2.06   0.056
Arm span      0.84042      0.08091       10.39   0.000

s = 1.613        R-sq = 87.1%      R-sq(adj) = 86.3%
```

n=18

A residual plot for the data looks like this:



1. Determine the equation of the least squares regression line from the printout.

$\hat{y} = 11.547 + 0.84042x$, $r = 0.93327$, $r^2 = 0.871$

2. In your opinion, is the least squares line an appropriate model for the data? Would you be willing to predict a student's height, knowing that his arm span is 76 inches? Explain. Then do it – use this model to predict the height of a student whose arm span is 76 inches.

- Yes, the residual plot has no clear pattern + $r^2 = 0.871$
- Yes, 76" is within the data set.

$\hat{y} = 11.547 + 0.84042(76)$

$\hat{y} = 75.4189$ inches.

3. Estimate the parameters α, β, and σ.

$\alpha = 11.547$, $\beta = 0.84042$, $\sigma = 1.613$

4. Construct a 95% confidence interval for the true slope of the regression line.

$b \pm t^* SE_b$

$0.84042 \pm 2.12(0.08091)$

$(0.66889, 1.01195)$

Test 14A AP Statistics Name: Mr. Morton

Directions: *Work on these sheets. A table of t distribution critical values is attached.*

Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.* pg. 783

1. Which of the following is NOT one of the basic assumptions that must be satisfied in order to perform inference for regression of y on x?
   (a) For each value of x, the corresponding population of y-values is normally distributed.
   (b) The standard deviation σ of the population of y-values corresponding to a particular value of x is always the same regardless of the specific value of x.
   (c) The sample size (the number of paired observations (x, y) in the sample data) exceeds 30.
   (d) There exists a straight line $y = \alpha + \beta x$ such that for each value of x, the mean μ_y of the corresponding population of y-values lies on that straight line.

2. If the assumptions for regression inference are met, then a normal probability plot of the residuals should be
   (a) Bell shaped
   (b) A group of randomly scattered points
   (c) Roughly linear
   (d) Clearly curved

3. If a test of hypotheses rejects H_0: $\beta = 0$ in favor of the alternative hypothesis H_a: $\beta > 0$, where β is the population regression slope, then the least-squares regression line
   (a) Slopes downward and to the right when plotted on the scatterplot of paired observations (x, y)
   (b) Is useful for predicting y given x (within the limits of x-values covered by the data)
   (c) Can be extrapolated beyond the limits of the x-values covered by the data to predict y at any possible x
   (d) Is not useful for predicting y given x

4. Inference for regression on the population regression slope β is based on which of the following distributions?
   (a) The t distribution with n – 1 degrees of freedom
   (b) The standard normal distribution
   (c) The chi-square distribution with n – 1 degrees of freedom
   (d) The t distribution with n – 2 degrees of freedom

5. Suppose that inference for regression is conducted on the following small data set:

x	12	14	16	18
y	2	3	5	6

   The number of degrees of freedom for our test statistic is
   (a) 4
   (b) 3
   (c) 2
   (d) Inference cannot be conducted on this data set because it is too small.
   (e) The answer cannot be determined from the information given.

26

6. In inference for regression, the statistic s represents
   (a) The estimate of the standard deviation s in the regression model
   (b) The standard deviation of the x-values in the paired observations (x, y)
   (c) The estimate of the y-intercept
   (d) The standard deviation of the y-values in the paired observations (x, y)

Part 2: Free Response

Answer completely, but be concise. Write sequentially and show all steps.

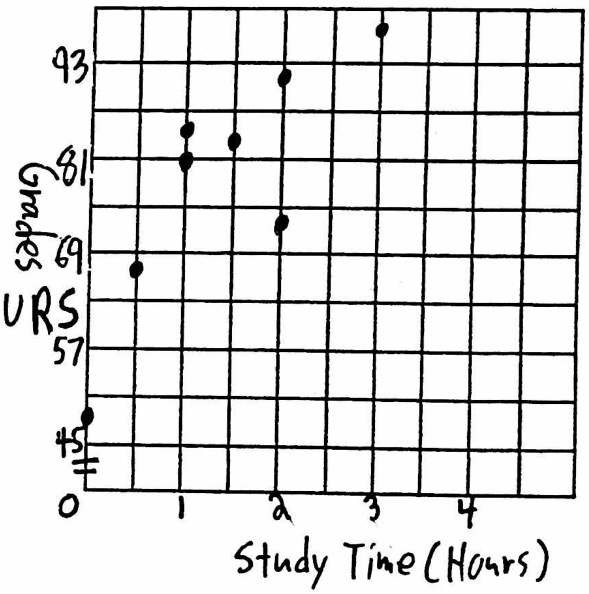
A teacher asked her 8 introductory statistics students to record the total amount of time they spent studying for a particular test. The amounts of study time x (in hours) and the resulting test grades y are given below.

$n=8$, $df=6$

hours	x	2	1	1.5	0.5	1	3	0	2
grade	y	92	81	84	68	85	96	48	74

7. Make a scatterplot of the data.

Grades VS. Study Time



8. Use your calculator to obtain the equation of the least-squares regression line and the correlation.

$$\hat{y} = 60.7059 + 12.9412x$$

$$\widehat{GRADE} = 60.7059 + 12.9412\ HOURS$$

$$r = 0.8098,$$

$$r^2 = 0.6558$$

9. Explain in words what the slope β of the true regression line says about hours studied and grade awarded.

For every one hour spent studying, the test grade increases by 12.9412 points.

10. What is the estimate of β from the data? What is your estimate of the intercept α of the true regression line?

$\beta = 12.9412$, $\alpha = 60.7059$

11. Use your calculator to calculate the residuals. Report the sum of the residuals and the sum of the squares of the residuals. Then use these results to estimate the standard deviation σ in the regression model.

$$\Sigma(y-\hat{y}) = 0$$

$$\Sigma(y-\hat{y})^2 = 560.3529$$

$$s = \sqrt{\frac{560.3529}{8-2}}$$

$$s = 9.6640$$

27

12. The standard error of the slope SE_b is defined as $SE_b = \frac{s}{\sqrt{\sum (x-\bar{x})^2}}$ Calculate SE_b.

$s = 9.6640$

$SE_b = \frac{9.6640}{\sqrt{6.375}}$ ~OR~ $SE_b = \frac{b}{t} = \frac{12.9412}{3.3811}$

$SE_b = 3.8275$ $SE_b = 3.8275$

13. Suppose we want to find out if the number of hours studied helps predict grade awarded on this statistics test. Formulate null and alternative hypotheses about the slope of the true regression line. State a two-sided alternative.

$H_0: \beta = 0$, $H_a: \beta \neq 0$

14. Determine the test statistic, the degrees of freedom, and the P-value of t against the alternative.

$t = 3.3811$

$df = 6$

p-value $= 0.0148$

15. Would you reject the null hypothesis at the 1% significance level? Explain briefly.

$0.0148 > 0.01$, fail to reject.

16. Write your conclusions in plain language.

Since $0.0148 > 0.01$, we fail to reject H_0 at a 1% significance level. Therefore, there is NOT enough evidence to say that there is a useful linear relationship between study time and grade

17. Compute a 95% confidence interval for the slope ß of the true regression line.

$b \pm t^* SE_b = 12.9412 \pm 2.447(3.8275)$

$\Rightarrow (3.5756, 22.307)$

I pledge that I have neither given nor received aid on this test. ☺

28

Test 14B A P Statistics Name: Mr. Morton

Directions: *Work on these sheets. A table of t distribution critical values is attached.*

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

The following information is used in questions 1–5.
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 fit 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 Coefficient     P-value
Constant     -176.644     61.16                   0.0050
Sales        0.092498     0.0075                  ≤0.0001
```

1. Using the above data, approximately what is the intercept of the least-squares regression line?
(a) 0.0925
(b) 0.0075
(c) –176.64
(d) 61.16
(e) None of the above. The answer is____________________.

2. 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) 0.0925 ± 0.012
(c) –0.0925 ± 0.0075
(d) –0.0925 ± 0.012
(e) None of the above. The answer is ____________________.

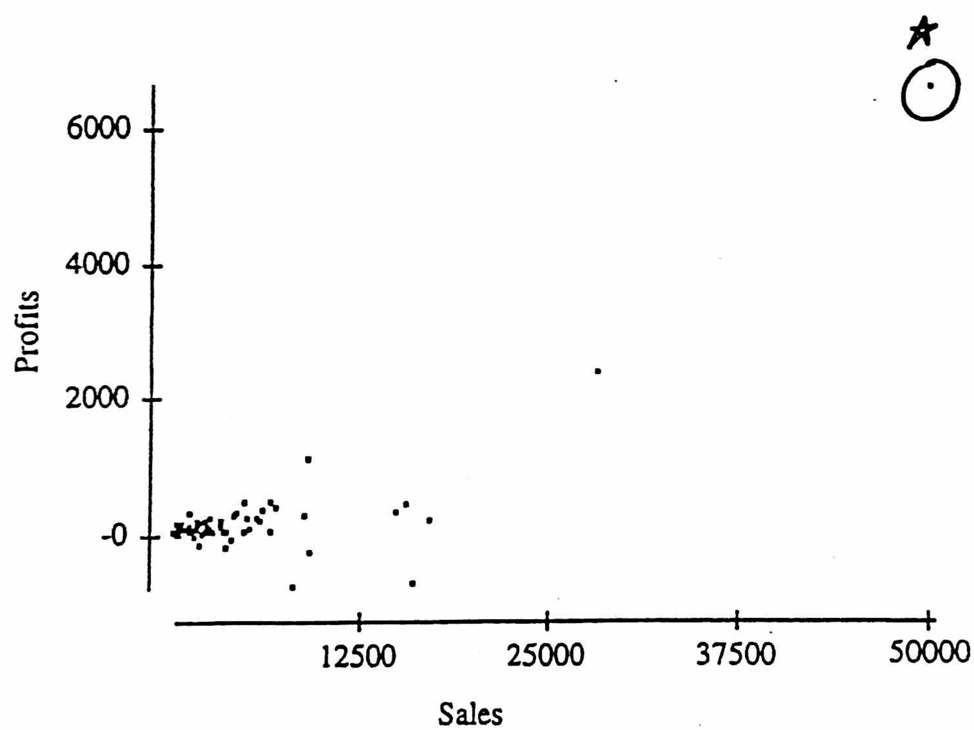
$0.0925 \pm 1.664(0.0075)$

3. 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
(b) 0.082
(c) 0.092
(d) 12.73
(e) None of the above. The answer is 12.3331.

$t = \frac{b}{SE_b}$

4. 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) 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. The answer is______________________________.

5. Use the above data for this question. A scatterplot of sales versus profits is given below:



Which of the following statements is supported by the plot?
(a) There is no striking evidence in the plot that the assumptions for regression are violated and there is a clear straight-line trend.
(b) There are very influential observations in the plot suggesting that our above results must be interpreted with extreme caution.
(c) The plot contains dramatic evidence that the standard deviation of the response about the true regression line is not even approximately the same everywhere.
(d) The plot contains many fewer points than were used to fit the least-squares regression line in the previous problems. Obviously there is a major error present.

Part 2: Free Response

Answer completely, but be concise. Write sequentially and show all steps.

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, PAPERS = # papers in hundreds, and the intervals listed at the bottom are computed for 1,600 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%

Fit       Stdev. Fit           95% C.I.                    95% P.I.
229.13          2.34      (223.93 , 234.34)          (213.68 , 244.58)
```

6. Formulate null and alternative hypotheses about the slope of the true regression line. Adopt the two-sided alternative.

$H_0: \beta = 0$

$H_a: \beta \neq 0$

7. What is the least-squares regression equation?

$\hat{y} = 35.80 + 12.0835x$, $\widehat{COST} = 35.8 + 12.1\ PAPERS$

8. What is the standard error about the line (also known as the standard deviation *s* in the regression model)?

$s = 6.526$

9. What is the slope of the least-squares regression line?

$b = 12.0835$

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

$\alpha = 35.8$, $\beta = 12.0835$, $\sigma = 6.526$

11. What is the value of the test statistic for testing the hypotheses?

$$t = \frac{b}{SE_b} = \frac{12.0835}{0.9738} = 12.4086$$

12. How many degrees of freedom does t have? $df = 10$

13. What is the P-value for the test?

p-value ≈ 0

14. Is the number of papers graded useful for predicting the amount paid? Use a significance level of 0.01. Explain briefly. Since $0 < 0.01$, we reject H_0.

15. What is the estimated cost of grading 1,600 papers? $X = 16$

$\approx \$229.14$

16. Find the 95% confidence interval for the average amount paid to all graders who grade 1,600 papers.

statistic $\pm t^* SE_b$

$229.14 \pm 2.228(0.9738)$

$(226.97, 231.31)$

ON CHART:
$(223.93, 234.34)$

I pledge that I have neither given nor received aid on this test.

32