Quiz 6
Stat 217

Name ANSWER KEY

FORMULAS: $DFE = n - 2$ $MS = \frac{SS}{DF}$ $R^2 = \frac{SSM}{SST}$ $t = \frac{b_1}{SE_{b_1}}$

SHOW YOUR WORK!!!

1. The principal at Super High School is interested in using the number of hours a high schooler studies each week to predict the student's academic performance, rated on a scale from 0 to 10. The MINITAB output follows:

Predictor	Coef	StdErr	T	P
Constant	-0.1872	0.8834	-0.212	0.83350
Hours	0.6100	0.1879	*****	0.00268

S = *** R-Sq = ****% R-Sq(adj) = 21.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	******	101.39	10.541	0.00268
Residual Error	33	317.39	****		
Total	34	418.78			

(a) (1/2 pt) Write out the Simple Linear Regression model.

$y = \beta_0 + \beta_1 x + \varepsilon$, where $\varepsilon \sim N(0, \square)$.

(b) (1/2 pt) Write out the least-squares regression line for this problem.

Perf = -.1872 + .61 Hours

(c) (1/2 pt) How many students were used in this SLR?

DFE = n – 2 = 33, so n = 35 students.

(d) (1/2 pt) The Simple Linear Regression model assumes the error terms have a constant variance, σ^2 . Give an unbiased estimate of σ^2 .

$$\sigma^2 \approx MSE = \frac{SSE}{DFE} = \frac{317.39}{33} = 9.617$$

(e) (1/2 pt) Calculate R^2 .

$$R^2 = \frac{SSM}{SST} = \frac{101.39}{418.78} = .2421$$

(f) (1 pt) Interpret the value of R^2 .

24% of the variability of academic performance is explained by the linear relationship with hours studied.

(g) (1/2 pt) What is the correlation between number of hours studied and academic performance?

$$r = \sqrt{R^2} = \sqrt{.2421} = .4920$$

(h) (1 ½ pts) Is there sufficient evidence to suggest that the slope, β_1 , is non-zero? Give the following regression output which help you answer this question:

- the value of the test statistic
- the distribution of the test statistic
- the p-value

The test statistic is $t = \frac{b_1}{SE_{b_1}} = \frac{.61}{.1879} = 3.246$, t~t(33), the p-value is .00268, and so we REJECT $H_0: \square_1 = 0$.

Thus, the evidence suggests that the slope is non-zero.

2. (1/2 pt) Does the evidence suggest that there is a linear relationship between number of hours studied and academic performance? Explain how you decided Yes or No.

Since the slope is non-zero, then we conclude that the there is a linear relationship between academic performance and number of hours studied.

3. (1 pt) Interpret the slope of the least squares regression line, b_1 **in terms of the problem**.

$b_1 = .61/1$. Thus, for each additional hour studied, academic performance increases on average by .61.

4. (1 pt) William, a student at Super High, studies five hours a week. Based on the above regression output, what is Williams academic performance?

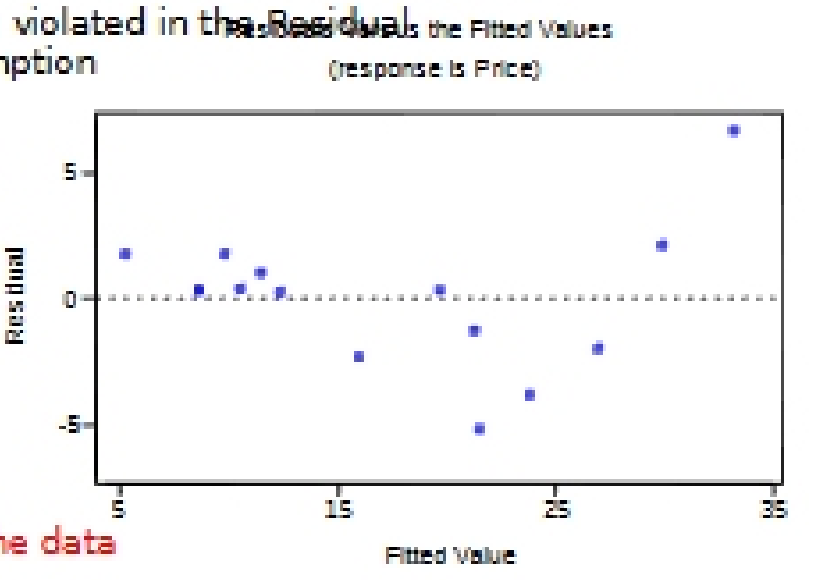
$\hat{y} = -.1872 + .61(5) = 2.86$. Thus, we predict that William's academic performance is 2.86.

5. (1 pt) A new student is starting at Super High this year. She studies 7 hours per week. The principal is not sure whether to use a confidence interval (CI) or a prediction interval (PI) to estimate this student's academic performance at the end of the year. Do you recommend that the principal use a CI or a PI? Explain why your answer is correct.

Use a Prediction Interval since we are predicting for a new student who studies 7 hours a week, not for a mean for all students who study 7 hours a week.

6. (1/2 pt) The constant variance assumption does not appear to be violated in the Residual versus Fitted Value plot shown on the right. Is there another assumption that appears to be violated for these data? Why or why not?

Linearity appears to be violated due to the parabolic shape in the plot.



7. (1/2 pt) For the least squares regression line,

A. SSE is the smallest when compared to all other lines fit to the data