

ANSWER KEY

DO WORK on Quiz before Submitting Answers

STAT1051: QUIZ 9 Take Home

Name:

1. Six students were interviewed in 1957 in Moscow, U.S.S.R. about their statistics and English grades. The following information was obtained:

Table 1 English and Statistics Scores

Student	1	2	3	4	5	6
Statistics Score (x)	70	92	80	74	65	83
English Score (y)	74	84	63	87	78	90
$\widehat{E}[y]$	76.221	85.556	80.464	77.918	74.099	81.737
$\hat{\epsilon}$	-2.221	-1.556	-17.464	9.081	3.901	8.263

$$\sum_{i=1}^6 x_i = 464 \quad \sum_{i=1}^6 x_i^2 = 36354 \quad \sum_{i=1}^6 x_i y_i = 36926$$

$$\sum_{i=1}^6 y_i = 476 \quad \sum_{i=1}^6 y_i^2 = 38254 \quad SS_{xx} = 471.33.$$

- (a) (5 pts) A plot of the data is shown in the figure to the right. Does there appear to be a conspicuous linear trend in the data?

Answer 1 *No*

- (b) (5 pts) Assert that a linear relation exists between Statistics and English grades. Write the linear model that describes the assertion:

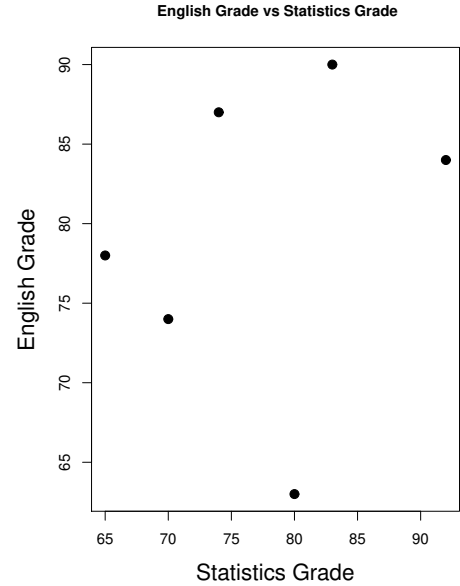
Answer 2 $English_i = \beta_0 + \beta_1 Statistics_i + \epsilon_i$
 where $\epsilon_i \sim N(0, \sigma^2)$

- (c) Identify the
 i. (5 pts) Explanatory variable

Answer 3 *Statistics*

- ii. (5 pts) Response variable

Answer 4 *English*



- (d) (5 pts) What criterion is used for deriving the least squares estimates of the parameters for a linear model?

Answer 5 *Minimize SSE*

- (e) (5 pts) Compute $SS_{xy} =$

Answer 6 $\sum_{i=1}^6 x_i y_i - \frac{(\sum_{i=1}^6 x_i)(\sum_{i=1}^6 y_i)}{n} = 36926 - \frac{(464)(476)}{6} = 115.33$

- (f) Assume $SS_{xy} = 200$.

- i. (5 pts) Compute the least squares estimate of $\beta_0 =$

Answer 7 $\bar{y} - \hat{\beta}_1 \bar{x} = \frac{476}{6} - (.4243281) \frac{464}{6} = 46.51862$

ii. (5 pts) Compute the least squares estimate of $\beta_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{200}{471.33} =$

Answer 8 $\frac{200}{471.33} = .4243281$

(g) **Answer 9** (5 pts) Enter the remaining values for $\hat{\epsilon}$ in Table 1 based on the assumption that $SS_{xy} = 200$.

i. **-2.221**

ii. **-1.556**

iii. **9.081**

(h) Compute

i. **Answer 10** (5 pts) $SSE = \sum \hat{\epsilon}_i^2 = (-2.221)^2 + (-1.556)^2 + (-17.464)^2 + (9.081)^2 + (3.901)^2 + (8.263)^2 = 478.3206$

ii. **Answer 11** (5 pts) $s = 10.935$

(i) The lower limit of the 95% confidence interval for the estimate of the predicted sample average is given by equation (1). Calculate the 95% confidence interval for $E[\widehat{E}[y_p]]$ when $x_p = 83$

$$a = \widehat{E}[y_p] - st_{n-2, \frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

$$a = 81.73 - 30.36118 \sqrt{\frac{1}{6} + \frac{(83 - 77.333)^2}{471.33}} \quad (1)$$

i. **Answer 12** (5 pts) $\alpha = .05$ $n=6$

ii. **Answer 13** (5 pts) $t_{n-2, \frac{\alpha}{2}} = 2.7764$

iii. **Answer 14** (5 pts) $\bar{x} = \frac{464}{6} = 77.333$

iv. **Answer 15** (5 pts) Lower limit $a = 81.73 - 30.36118 \sqrt{\frac{1}{6} + \frac{(83-77.333)^2}{471.33}} = 67.01$

v. **Answer 16** (5 pts) 95% CI is: **(67.01, 96.44)** *Parentheses and comma are required for correct notation*

vi. (5 pts) Interpret this confidence interval.

Answer 17 *At the end of semester, I will obtain the English grades of everyone in the class and calculate the average English grade. We are 95% confident that the average English grade for the class will be in this confidence interval.*

- (j) What are the two principal reasons for using linear models.
- i. **Question 1** (5 pts) **To interpolate**
 - ii. **Question 2** (5 pts) **To show an inherent relationship between x and y.**
- (k) (5 pts) To what extent do you think that this fitted model can be applied to GWU students; in other words, does the theory make sense?

Answer 18 *No, the theory does not make sense Explain your answer. The population of Russian students in 1957 and the population of GWU students are not compatible. For one reason, the native language of the University of Moscow students is Russian, whereas the native language of GWU students is English. Therefore, it is not possible to use the one population to explain the behaviour of the other population by means of this model. Hence, the model is not good.*

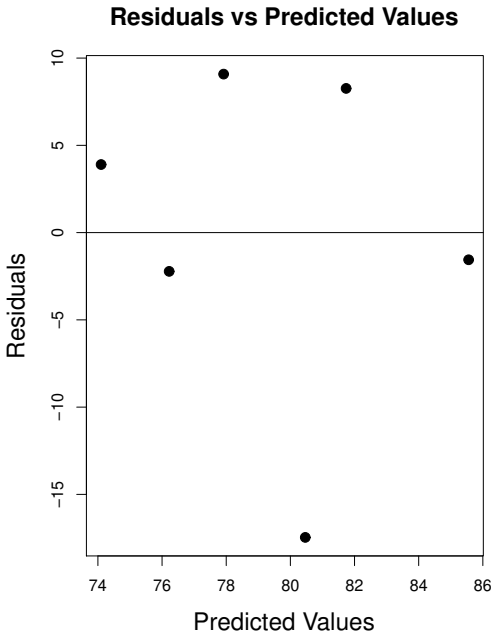


Figure 1

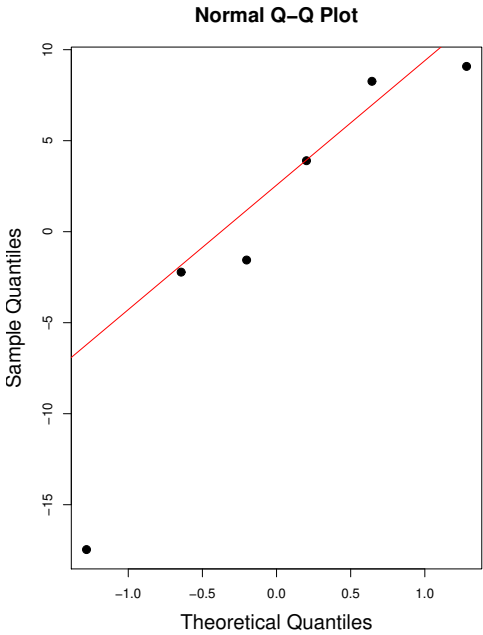


Figure 2

2. **BONUS** (5 pts) Does the plot of residuals vs predicted values show a random pattern?

Answer 19 *Yes, the pattern looks random.*

3. **BONUS** (5 pts) Do the points in the QQ plot follow a straight line?

Answer 20 *No, there is one obvious outlier.*

4. **BONUS** (5 pts) Does the plot of the data shown above and Figures 1 and 2 substantiate or discredit the model?

Answer 21 *The plot of the data itself discredits the model. Explain The theory does not make sense, the plot of the data does not show a conspicuous linear trend, the QQ plot is not linear. Any one of these will invalidate the model. The model is defective.*