ANSWERKEY

Homework 9

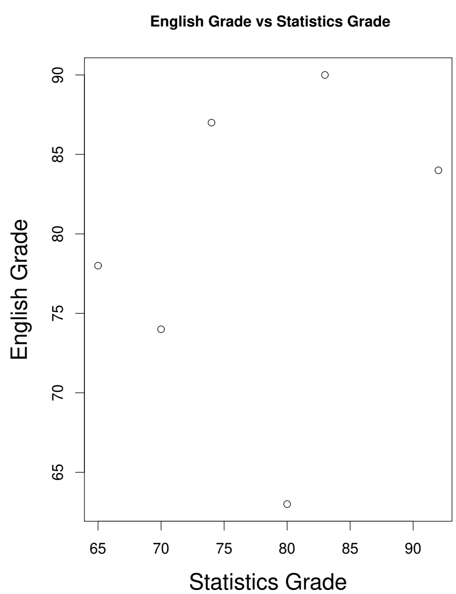
1. Six students were interviewed in 1957 in Moscow, U.S.S.R. about their statistics and English grades. The following information was obtained:

Statistics Grade	70	92	80	74	65	83
English Grade	74	84	63	87	78	90

$\sum_{i=1}^{6} x_i = 464 \quad \sum_{i=1}^{6} x_i^2 = 36354 \quad \sum_{i=1}^{6} x_i y_i = 36926$
$\sum_{i=1}^{6} y_i = 476 \quad \sum_{i=1}^{6} y_i^2 = 38254 \quad SS_{xx} = 471.33.$

(a) Plot the data. Let x be the statistics grade and y be the English grade. Assess the plot as to whether statistics grades are directly proportional to English grades.

Answer 1



(b) Assert that a linear relation exits between the two grades. Write the linear model that describes your assertion. Be sure to state the probability distribution which you are assuming for ϵ.

Answer 2 $y_i = \beta_0 + \beta_i x_i + \epsilon_i$ *where* $\epsilon_i \sim N(0, \sigma^2)$ *NOTE: At most half partial credit is given if* $\epsilon_i \sim N(0, \sigma^2)$ *is omitted.*

(c) Compute the least squares estimates of β_0 and β_1.

Answer 3

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1\bar{x} = \frac{476}{6} - (.2446959)\frac{464}{6} = 79.3333 - .2446959(77.33) = 60.4101839$$
$$\widehat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{115.33}{471.33} = .2446959$$

(d) Suppose everyone in STAT51 will get a statistics grade of 90. Estimate the sample average of the English grades that will be given at the end of the semester to the same class of STAT51 students.

Answer 4 $\widehat{y} = 60.41 + .2446x_0 = 60.41 + .2445(90) = 82.43281$

(e) Compute the 95% confidence interval of the estimate of that predicted sample average.

Answer 5 *Note that* $s^2 = \frac{\sum(\epsilon^2)}{n-2} = \frac{463.1117}{4} = 115.7779 \rightarrow s = 10.76$

$$t_{4,.05} = 2.77$$
$$a = \widehat{E[y]} - st_{n-2,\frac{\alpha}{2}}\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$
$$= 82.43281 - 10.76(2.77)\sqrt{1/6 + \frac{(90 - 77.33)^2}{471.33}} = 61.159$$
$$b = \widehat{E[y]} + st_{n-2,\frac{\alpha}{2}}\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$
$$= 82.43281 + 10.76(2.77)\sqrt{1/6 + \frac{(90 - 77.33)^2}{471.33}} = 103.7062$$

The 95% C.I. is (61.159,103.7062)

(f) Forget about everyone else; predict your English grade based on the fitted least squares line, supposing that you will receive a statistics grade of 95.

Answer 6 $\widehat{y} = 60.41 + .2446x_0 = 60.41 + .2445(95) = 83.6563$

(g) Compute the 95% confidence interval of the estimate of your English grade.

Answer 7

$$
\begin{aligned}
t_{4,.05} &= 2.77 \\
a &= \widehat{E[y]} - st_{n-2,\frac{\alpha}{2}}\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \\
&= 82.43281 - 10.76(2.77)\sqrt{1 + 1/6 + \frac{(90 - 77.33)^2}{471.33}} = 43.255 \\
b &= \widehat{E[y]} + st_{n-2,\frac{\alpha}{2}}\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \\
&= 82.43281 + 10.76(2.77)\sqrt{1 + 1/6 + \frac{(90 - 77.33)^2}{471.33}} = 124.05
\end{aligned}
$$

The 95% C.I. is (43.255,124.05)

(h) Do you trust the model? Explain your answer.

Answer 8 *The population of Russian students in 1957 and the population of GWU students are not compatible. i.e. biased data, hence the model is not good. Note that the upper limits of both confidence intervals exceed 100 which is not possible. This implies that something went wrong. The plot of the data does not show a linear trend, hence the model is defective.*

2. You are planning to sell a used 2004 automobile and want to establish an asking price that is competitive. From a review of newspaper advertisements for used cars, you collect the following data where the asking price is in thousands of dollars and age is in years before the present:

Age of car (x)	1	1	2	3	3	4	7	8	9	10	11	12
Asking Price (y)	9.8	8.9	8.8	7.7	8.4	6.0	3.4	2.0	1.5	1.6	1.4	1.0
$\widehat{E[y]}$	9.23	**9.23**	8.38	7.53	7.53	**6.67**	4.11	**3.26**	2.41	1.56	.70	-.14
$\hat{\epsilon}$	.57	**-.33**	.42	**.17**	**.87**	**-.67**	**-.71**	-1.26	**-.91**	**.04**	**.70**	1.14

$\sum_{i=1}^{12} x_i = 71$ $\sum_{i=1}^{12} x_i^2 = 599$ $\sum_{i=1}^{12} x_i y_i = 205.3$ $\sum_{i=1}^{12} y_i = 60.5$ $\sum_{i=1}^{12} y_i^2 = 441.87.$

(a) Write the linear model:

Answer 9 $y_i = \beta_0 + \beta_i x_i + \epsilon_i$ *where* $\epsilon_i \sim N(0, \sigma^2)$ *NOTE: At most half partial credit is given if* $\epsilon_i \sim N(0, \sigma^2)$ *is omitted.*

(b) SS_{xx}=

Answer 10 *178.9167*

(c) SS_{xy}=

Answer 11 *152.6583*

(d) Given that $\widehat{\beta}_0 = 10.089986$ and $\widehat{\beta}_1 = -0.853237$, enter the remaining values for $\widehat{E[y]}$ in the table.

(e) Enter the remaining values for $\widehat{\epsilon}$ in the table.

(f) Calculate SSE=

Answer 12 $\sum \widehat{\epsilon_i^2} = 6.595417$

Based on your set of data and on the asserted linear model, what would be a good asking price, y_p, for your three year old car i.e.,

(g) Compute $\widehat{E[y_p]}$=

Answer 13 *10.089986-0.853237(3)= 7.530275=$7,530*

(h) Compute the 95% confidence interval of $E[\widehat{E[y_p]}]$.

Answer 14 *i.* $s = \sqrt{\frac{sse}{12-2}} = \sqrt{\frac{6.595417}{10}} = .8121217$

ii. $\alpha = .05$ $\frac{\alpha}{2} = .025$ *n-2=10*

iii. $t_{10,.025} = 2.228139$

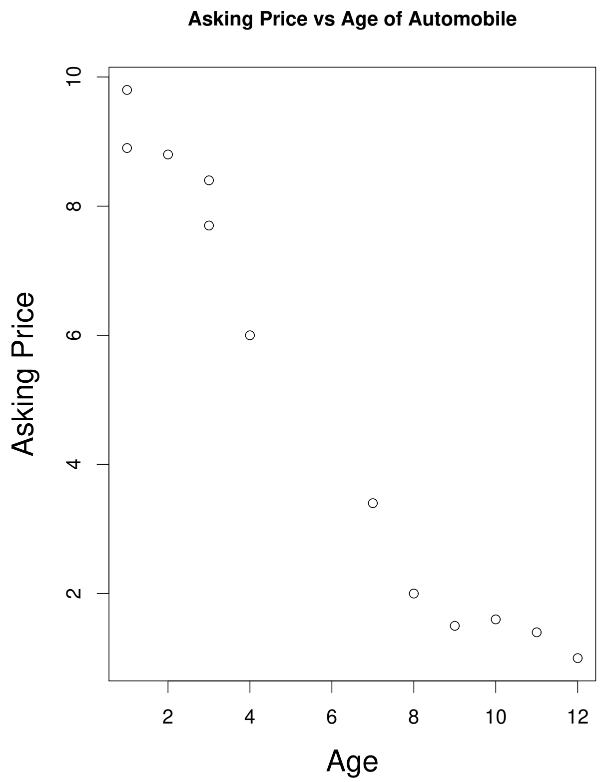
iv. $a = \widehat{E[y_p]} - st_{n-2,\frac{\alpha}{2}}\sqrt{\frac{1}{n} + \frac{(x_p-\bar{x})^2}{SS_{xx}}} = 7.530275 - .8121217(2.228139)\sqrt{1/12 + \frac{(3-5.916667)^2}{178.9167}} = 6.875638$

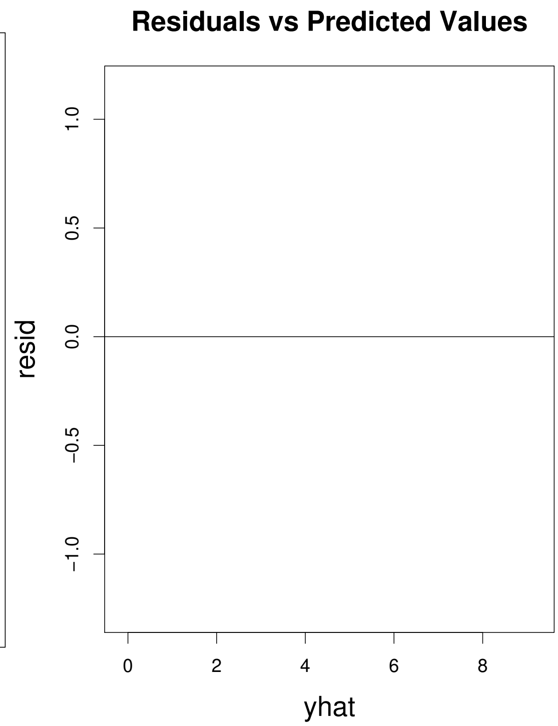
v. $b = \widehat{E[y_p]} + st_{n-2,\frac{\alpha}{2}}\sqrt{\frac{1}{n} + \frac{(x_p-\bar{x})^2}{SS_{xx}}} = 7.530275 + .8121217(2.228139)\sqrt{1/12 + \frac{(3-5.916667)^2}{178.9167}} = 8.184912$

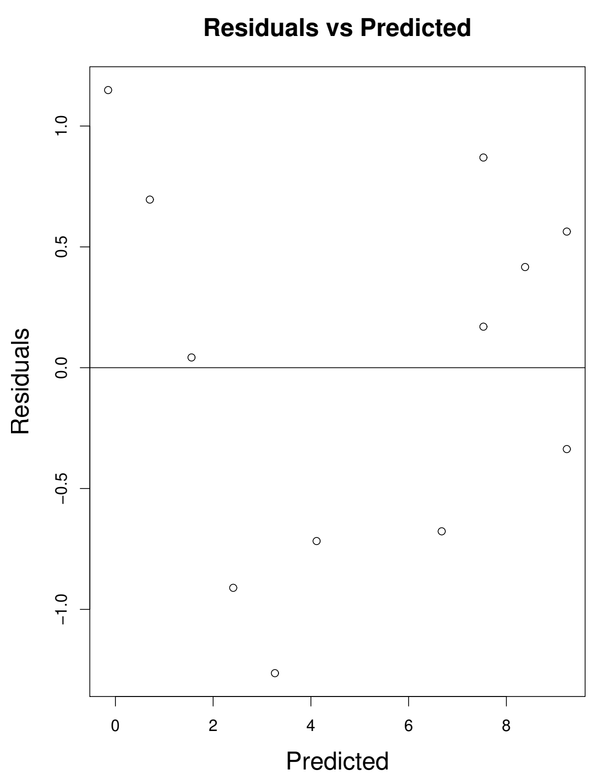
vi. The 95% CI for $E[\widehat{E[y_p]}]$*=($6,875,$8,185)*

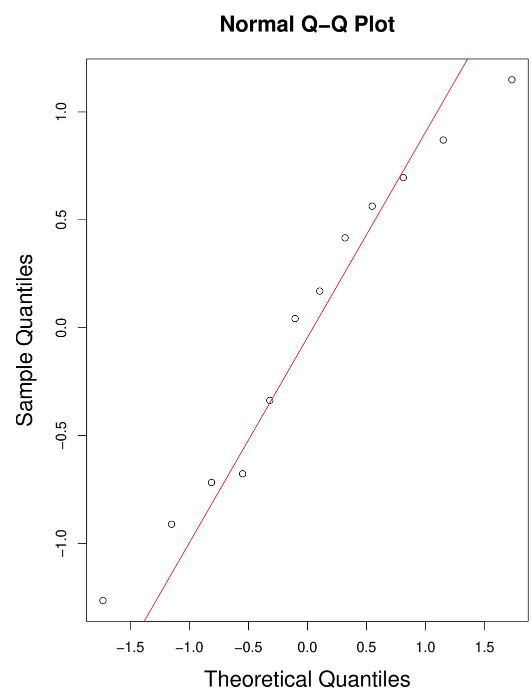
(i) Plot the data.

(j) Plot the residuals versus predicted values.


Asking Price vs Age of Automobile
Asking Price
Age
2
4
6
8
10
12


Residuals vs Predicted Values
resid
yhat
−1.0
−0.5
0.0
0.5
1.0


Residuals vs Predicted
Residuals
Predicted
0
2
4
6
8


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

Answer 15

(k) Do these two plots indicate that the model is a good model? Explain your answer.

Answer 16 *No, the plot of residuals versus predicted values shows a V-shaped pattern. The existence of this pattern invalidates the asserted linear model. Also, note that the plot of the data shows that there is a non-linear trend towards the right of the plot. Both plots indicate that the proposed linear model is not a good model.*

(l) A common transformation which is used to rectify a bad plot of residuals versus predicted values is the square root transformation. The least squares fitted line for a square root transformation is: $\widehat{\sqrt{y_p}} = 3.3159605 - 0.2064453x_p$. Use this model to estimate the asking price, y_p, when the age of the car is $x_p = 3$.

Answer 17 $\widehat{\sqrt{y_p}} = 3.3159605 - 0.2064453(3) = 2.696625$ *hence* $y_p = 2.696625^2 = \$7,271$

Note that this estimate is \$7,530-\$7,271=\$259 less than the previous estimated.