STAT-UB 103 Homework 9

Ishan Pranav

April 30, 2023

1 A study

a. Without looking at the data, we can expect a negative relationship between a person's intelligence test score and the age at which they spoke their first word.

b. Yes, as expected, the plot in Figure 1 indicates a negative relationship between age and score.

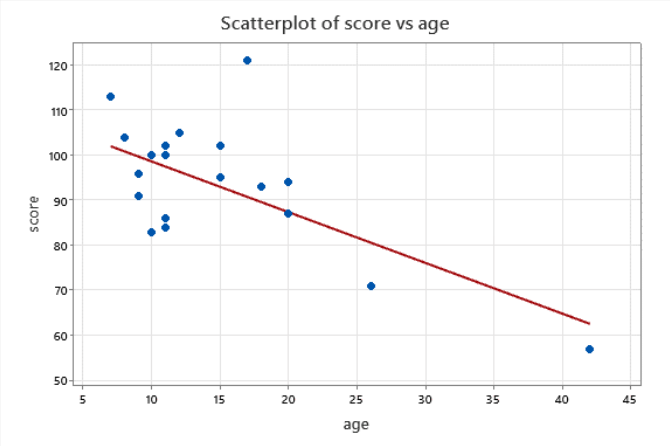

Figure 1: Scatterplot for intelligence test score (response) and age at which first words were spoken (predictor).

c. See Figure 2.

d.

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 \neq 0.$$

$$P(t = -3.63 \mid H_0) \approx 0.002.$$

There is evidence at the one-half-percent significance level that score is related to age.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	109.87	5.07	21.68	0.000	
age	-1.127	0.310	-3.63	0.002	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
11.0229	41.00%	37.89%	27.15%

Fits and Diagnostics for Unusual Observations

Obs	score	Fit	Resid	Std Resid	
18	57.00	62.54	-5.54	-0.85	X
19	121.00	90.72	30.28	2.82	R

R Large residual
X Unusual X

Figure 2: Linear regression analysis for intelligence test score (response) and age at which first words were spoken (predictor).

e.

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 < 0.$$

$$P(t = -3.63 \,|\, H_0) \approx \frac{0.002}{2} \approx 0.001.$$

Given that the null hypothesis is true, the probability of obtaining a Student t-statistic as extreme as (or more extreme than) -3.63 is approximately 0.1 percent. Since 0.001 is smaller than the significance level ($\alpha = 0.005$), there is sufficient evidence to reject the null hypothesis at the half-percent significance level. There is evidence that score is related to age.

f.

$$r^2 \approx 0.4100$$

Approximately 41 percent of the variation in a person's intelligence test score can be explained by variations in the age at which they spoke their first words using a least-squares regression line. This suggests that the linear relationship is moderate.

g. The point (42, 57) is a leverage point because its leverage is 0.65161, which is greater than $\frac{2(k+1)}{n}$. The Cook's distance for this point is 0.678112, which is less than 1, meaning that it is not necessarily a bad leverage point. However, it is still significantly higher than the others: The next-highest is 0.081498. Visually, the point is a good leverage point, but it is still worth evaluating the point.

h.

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 \neq 0.$$

$$P(t = -1.51 \mid H_0) \approx 0.149.$$

After removing the point, the P-value is no longer statistically significant, meaning that without the point (42, 57), we cannot conclude beyond a reasonable doubt that there is a linear relationship between age and score.

$$r^2 \approx 0.1122.$$

After removing the point, the r^2 coefficient of determination has decreased by almost 30 percentage points. This means that the point (42, 57) was extremely influential on the least-squares regression model.

i. Yes, it may be justifiable to remove the point (42, 57) because of its high leverage and influence on the model. Without the point, the linear relationship between age and score is inconclusive.