

Name: Caleb McWhorter — Solutions
MATH 108
Spring 2024
HW 16: Due 04/10

*“I should let you know, I read a book on
Jiu-Jitsu, and I am prepared to throw it at
you.”*

— Sheldon Cooper, Young Sheldon

Problem 1. (10pts) For a general model, what is the difference between interpolation and extrapolation? For a linear regression, what is R^2 and what does it tell you?

Solution. Interpolation is using a model for values between the most extreme values used to create the model. Extrapolation is using this model for values beyond the most extreme values used to create the model.

For a linear regression, R^2 is the coefficient of determination. This is the square of the (Pearson) regression coefficient, R . The coefficient of determination gives the ‘percentage linearity.’ That is, it tells one the amount of variation in the data explained by the linear model. We only have $R^2 = 1$ if the data is perfectly linear.

Problem 2. (10pts) A researcher is trying to determine if one can predict college success from a student's ACT scores. Specifically, whether one can predict a student's first semester college GPA using their ACT score. The researcher gathers data and creates a linear regression to fit the data. The researcher finds $G = 0.061A + 2.03$, where A is the student's ACT scores and G is the student's GPA. The researcher finds an R^2 value of 0.0726.

- (a) Identify b_0 and b_1 for this model.
- (b) Predict a student's first semester GPA that receives an ACT score of 20.
- (c) If a student that received an ACT score of 20 had a first semester GPA of 3.010, compute the residual for this student.
- (d) Does there appear to be a (linear) relationship between ACT score and GPA? Explain using the coefficient of determination.

Solution.

- (a) We know that a simple linear regression will take the form $\hat{y} = b_0 + b_1x$. But given the model $G = 0.061A + 2.03$, we see that $b_0 = 2.03$ and $b_1 = 0.061$.

- (b) We have...

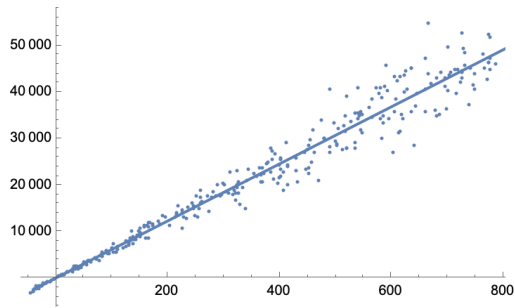
$$G(20) = 0.061(20) + 2.03 = 1.22 + 2.03 = 3.250$$

- (c) The residual would be...

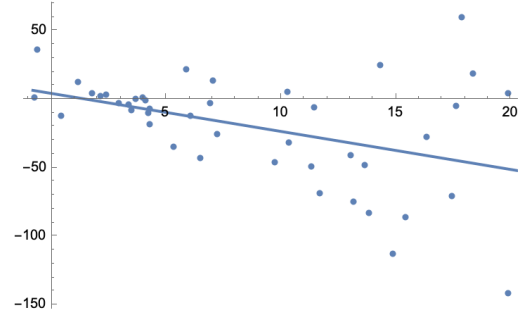
$$e_i = \text{Actual} - \text{Predicted} = 3.010 - 3.250 = -0.240$$

- (d) The R^2 value is 0.0726. That is, the data is only 7.26% linear, i.e. only 7.26% of the variation in the data is explained by the linear model. This is an extremely low R^2 value. Therefore, this linear regression is a bad model. Equivalently, one's ACT score is likely a poor (linear) predictor of one's first semester college GPA.

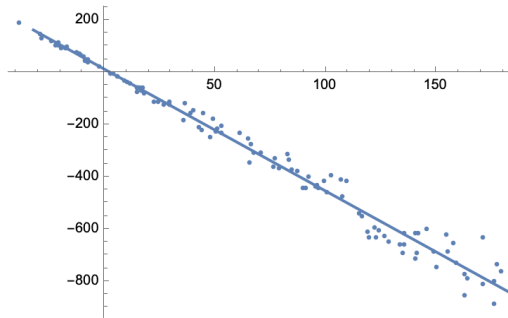
Problem 3. (10pts) Match each regression coefficient to its corresponding graph.



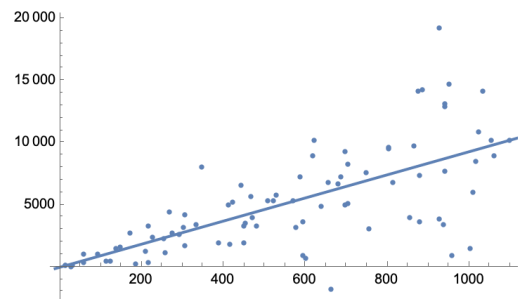
(a)



(b)



(c)



(d)

(i) a : $R = 0.981608$

(ii) d : $R = 0.693245$

(iii) b : $R = -0.425885$

(iv) c : $R = -0.991438$