Quiz 2

BSTA 512, Winter 2024

Feb 21, 2024

Name: Answer Key

Instructions

There are **10 total pages** in the exam and **14 questions** (11 multiple choice and 3 free response). Please make sure you have all of the pages!

1. I have written a "30 minute" quiz. However, you have 50 minutes from 2:00 - 2:50pm.
2. The quiz is open book and open notes. You may use books other than the class textbook, you may use anything on our course webpage, and you may use reference websites (like Wikipedia, Googling expected value of specific distribution, etc.).
3. No cheating will be tolerated. If one person is caught cheating, I will need to reconsider how to administer Quiz 2. Cheating includes:
    - Using ChatGPT
    - Using question and answer threads typically seen on sites like StackExchange, WikiHow, Quora, Reddit, StackOverflow, Chegg, etc.
    - Asking other students in the room or looking at other students' quiz work.
4. Each multiple choice question is worth 3 points. The free response questions are labelled with their point value.
5. You may use headphones during the quiz.

Grading

For Nicky to fill out:

Question	Points	Potential Points
Questions 1-9		27
Questions 10-14		17
Total		44

Questions

1. In the following simple linear regression model, with only one continuous predictor, what is the F-test measuring?

$$Y = \beta_0 + \beta_1 X + \epsilon$$

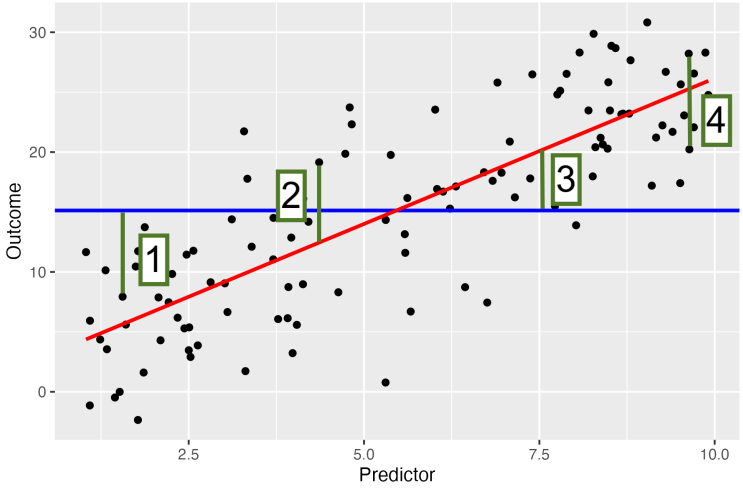
   a. The proportion of variation not explained by the model to total variation of the outcome (Y)

   b. The proportion of variation explained by the model to variation not explained by the model

   c. The proportion of variation explained by the model to total variation of the outcome (Y)

   d. The proportion of total variation of the outcome (Y) to variation not explained by the model

2. In the following picture, which number corresponds to the deviation that is part of the total sum of squares (SSY or SST)? Note that the red and blue lines correspond to the fitted line and the mean outcome, respectively.



   a. 1

   b. 2

   c. 3

   d. 4

3. When using an F-test to compare a full model (under the alternative hypothesis) to the reduced model (under the null hypothesis), what are we testing?

   a. We are testing if the unexplained variation in the full model is significantly greater than the unexplained variation in the reduced model.

   b. We are testing if the total variation in the full model is significantly less than the total variation in the reduced model.

   c. We are testing if the unexplained variation in the full model is significantly less than the unexplained variation in the reduced model.

   d. We are testing if the explained variation in the full model is significantly less than the explained variation in the reduced model.

4. For the model: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$. Which of the following would be the reduced model if I was testing if the addition of X_4 significantly added to the prediction of Y?

   a. $Y = \beta_0 + \epsilon$

   b. $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \epsilon$

   c. $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \beta_3 X_3 + \epsilon$

   d. $Y = \beta_0 + \beta_4 X_4 + \epsilon$

5. For the model: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$. Which of the following would be the reduced model if I was testing if any of the variables X_1, X_2, X_3, X_4 significantly add to the prediction of Y?

   a. $Y = \beta_0 + \epsilon$

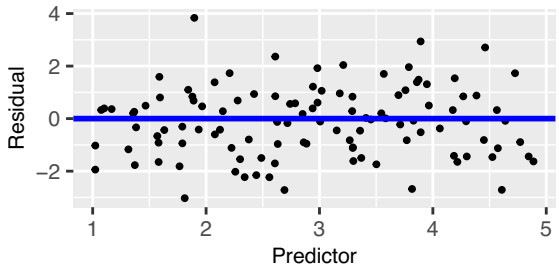
   b. $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \epsilon$

   c. $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_2 + \beta_3 X_3 + \epsilon$

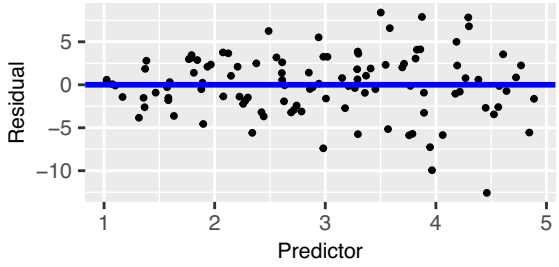
   d. $Y = \beta_0 + \beta_4 X_4 + \epsilon$

6. Which of the following residual plots show approximately constant variance?

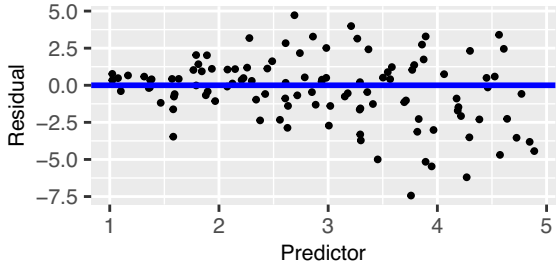
a. Residual plot 1:



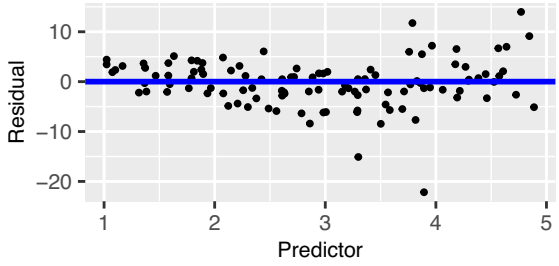
b. Residual plot 2:



c. Residual plot 3:

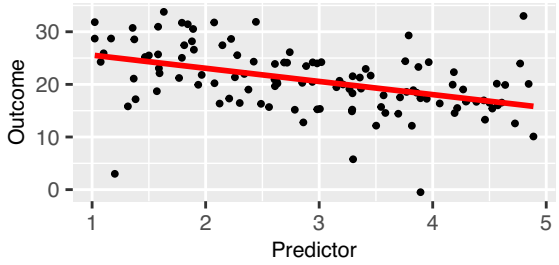


d. Residual plot 4:

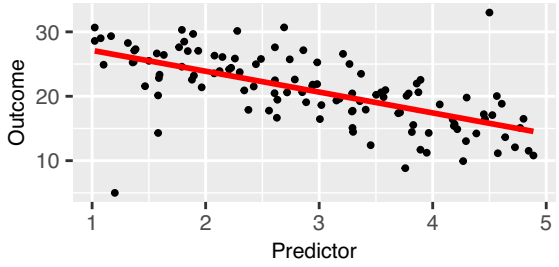


7. Which of the following scatterplots does **not** have potential outliers?

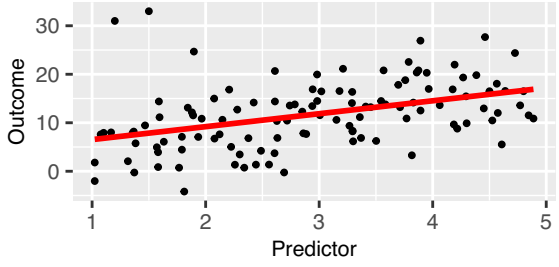
   a. Scatterplot 1:



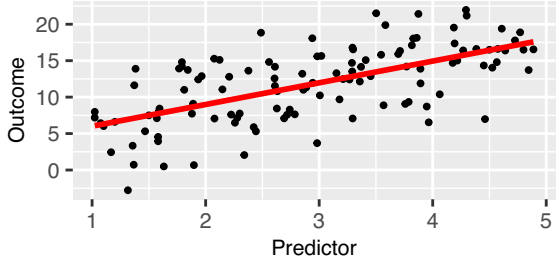
   b. Scatterplot 2:



   c. Scatterplot 3:

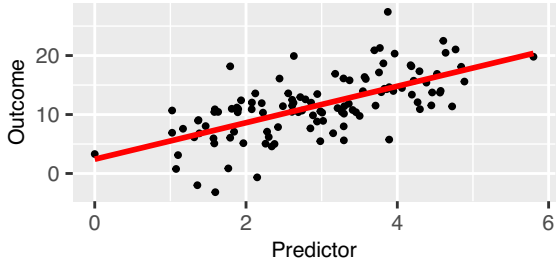


   d. Scatterplot 4:

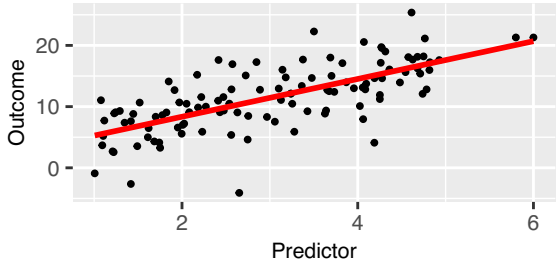


8. Which of the following scatterplots do **not** have high leverage observations?

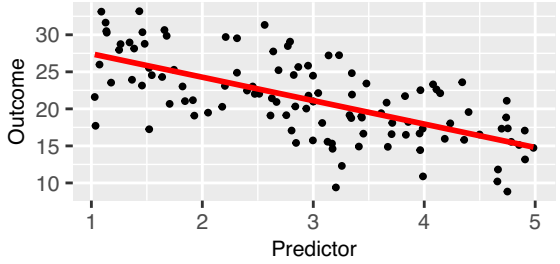
   a. Scatterplot 1:



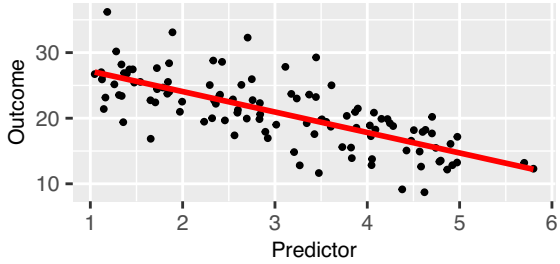
   b. Scatterplot 2:



   c. Scatterplot 3:

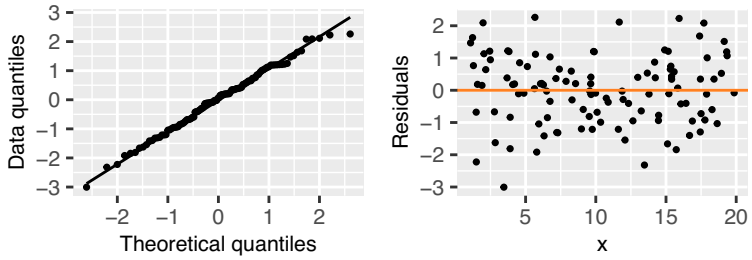


   d. Scatterplot 4:

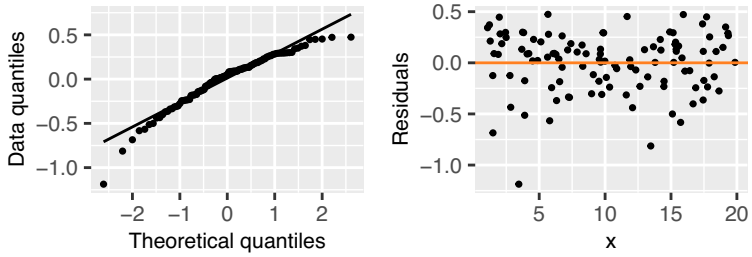


9. Below are QQ plots and residual plots for different transformations of X and Y in SLR. Which of the following models is the most appropriate for our analysis?

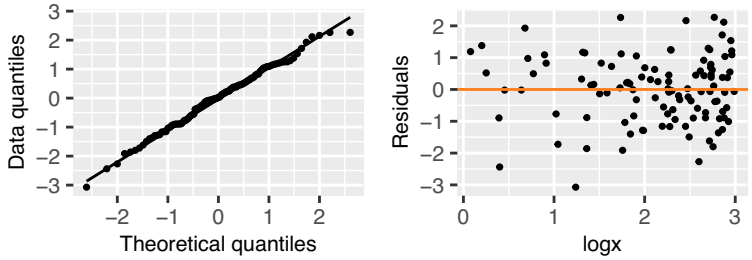
a. No transformation: $Y = \beta_0 + \beta_1 X + \epsilon$



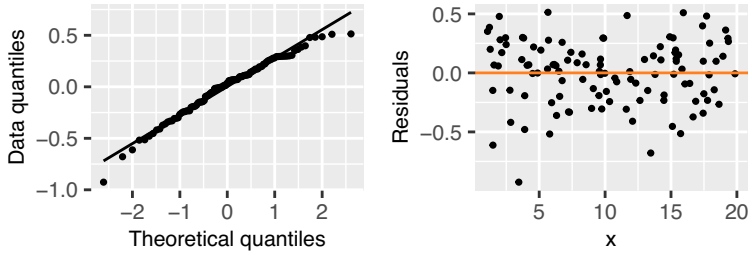
b. Log-transformation of Y: $log(Y) = \beta_0 + \beta_1 X + \epsilon$



c. Log-transformation of X: $Y = \beta_0 + \beta_1 log(X) + \epsilon$



d. Square root of Y: $\sqrt{Y} = \beta_0 + \beta_1 X + \epsilon$



Question 10-14 will use the following study and analysis:

A study is conducted to examine the role of age (measured in days) and birth weight (BWT in ounces) as predictors of infant systolic blood pressure (SBP, mm Hg). Let's say we fit the following model:

$$\text{SBP} = \beta_0 + \beta_1 \text{BWT} + \beta_2 \text{Age} + \epsilon$$

We have the following regression table:

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	53.45	4.53	11.79	0	43.66	63.24
bwt	0.13	0.03	3.66	0	0.05	0.20
age	5.89	0.68	8.66	0	4.42	7.36

10. For the above model, the coefficient of determination (R^2) is 0.8809. Please interpret this value in the context of the study.

    a. 88.1% of the variation in fitted infant systolic blood pressure is explained by the linear model with birth weight and age.

    b. 88.1% of the variation in infant systolic blood pressure is explained by the linear model with birth weight.

    c. 88.1% of the variation in infant systolic blood pressure is explained by the linear model with birth weight and age.

    d. 88.1% of the variation in infant systolic blood pressure is not explained by the linear model with birth weight and age.

11. (3 points)From the above regression table, please write out the fitted regression equation. Please include the numeric values for each estimated coefficient and keep round them to the 2nd decimal place.

$$\widehat{SBP} = 53.45 + 0.13\,BWT + 5.89\,Age$$

OR

$$SBP = 53.45 + 0.13\,BWT + 5.89\,Age + \hat{\varepsilon}$$

0.5 pt: BWT/Age have NO hat

0.5 pt: correct values

0.5 pt: (Y has hat + no ε) OR (Y no hat + $\hat{\varepsilon}$)

1 pt: includes all vars in correct spots

$\widehat{SBP}$ can be:
- $\hat{Y}$
- $\hat{E}(Y|Age, BWT)$
- $E(Y|Age, BWT)$

12. (5 points) Please interpret the coefficient for birth weight in the above regression equation. Please include the 95% confidence interval.

For every 1 ounce (0.5 pts; 1+ units) increase in birth weight (0.5 pts; inc in right variable), there is an expected (0.5 pts; exp/avg/mean) increase (0.5 pts) of 0.13 (0.5 pts; value) mmHg (0.5 pts; units) in systolic blood pressure (0.5 pts) (95% CI: 0.05, 0.20 (included 0.5 pts; 0.25 pts each)), adjusting for age. (0.5 pts)

13. (3 points) Using the above regression table and the fitted regression equation, what is the fitted regression line for systolic blood pressure for 4-day old infant?

$$\widehat{SBP} = 53.45 + 0.13\,BWT + 5.89\,Age$$

$\hookrightarrow Age = 4$

$$= 53.45 + 0.13\,BWT + 5.89(4)$$

$$= 53.45 + 0.13\,BWT + 23.56$$

$$\widehat{SBP} = 77.01 + 0.13\,BWT$$

1 pt: using correct age

0.5: keeping $0.13 \cdot BWT$ in line

0.5: simplifying the math

1 pt: keeping coef w/ right variable

14. Using the above regression table and the fitted regression equation, what is the expected systolic blood pressure for 5-day old infant with had a birth weight of 115 ounces?

a. 91.96 mmHg

b. 731.45 mmHg

c. 38.51 mmHg

d. 97.85 mmHg

$$\widehat{SBP} = 53.45 + 0.13\,BWT + 5.89\,Age$$

$$= 53.45 + 0.13(115) + 5.89(5)$$

$$= 53.45 + 14.95 + 29.45$$

$$= 97.85 \text{ mmHg}$$