INSTRUCTOR:
Xiaoping Shi

Irving K. Barber Faculty of Science
UBC, Okanagan Campus

This exercise set has an additional 7 questions, for a total of 20 points. These questions, in addition to those on other practice sets, are representative of what might be seen on Quiz 2.

READ THE QUESTIONS CAREFULLY

Notes and electronic devices are allowed, but they MUST be kept in airplane mode. You may use the back of a page if you run out of room on the front.

SURNAME, GIVEN NAME (print) _____

STUDENT NUMBER. _____

Signature: _____

Possible multiple choice–pick the MOST accurate answer

1. (2 marks) Suppose we fit a multiple linear regression $y = a + b_1x_1 + b_2x_2 + \varepsilon$, where X_1 is a numeric predictor and X_2 is a binary predictor. We can interpret this model as:

   (a) a single line with slop b_1.

   (b) two lines with the same slop b_1 but one with intercept a and the other with intercept $a + b_2$.

   (c) two lines with the same intercept a but one with slop b_1 and the other with slop $b_1 + b_2$.

   (d) two lines with the different slops, one with slop b_1 and the other with $b_1 + b_2$, and with different intercepts, one with a and the other with $a + b_2$.

   (e) None of the above.

   Sol. b.

2. (2 marks) Regarding the multiple linear regression $y = a + b_1x_1 + b_2x_2 + b_3x_1x_2 + \varepsilon$, which of the following statements are true?

   (a) $E(\varepsilon) = 0$.

   (b) $E(y) = a$ when $x_1 = 0$.

   (c) The change in $E(y)$ is b_1 when there is a 1 unit increase in x_1.

   (d) The change in $E(y)$ is b_2 when there is a 1 unit increase in x_2.

   (e) The change in $E(y)$ is $b_1 + b_3$ when there is a 1 unit increase in x_1.

   Sol. a.

3. (2 marks) Suppose we fit a model of the form $y = a + b_1x + b_2x^2 + \varepsilon$, where ε is assumed to be normal distributed. Which of the following R codes are correct?

   (a) lm(y∼x+x^2).

   (b) lm(y∼x+I(x^2)).

   (c) glm(y∼x+I(x^2), family = binomial).

   (d) glm(y∼x+I(x^2), family = poisson).

   Sol. b

4. (2 marks) Suppose we fit a model of the form $\log(E(y)) = a + b_1x + b_2x^2$, where y is a count variable. Which of the following R codes are correct?

   (a) lm(y∼x+x^2).

   (b) lm(y∼x+I(x^2)).

   (c) glm(y∼x+I(x^2), family = binomial).

   (d) glm(y∼x+I(x^2), family = poisson).

   Sol. d.

5. (2 marks) Suppose we fit a model of the form $logit(P(Y = 1)) = a + b_1x + b_2x^2$, where y takes 1 or zero. Which of the following R codes are correct?

(a) lm(y∼x+x^2).

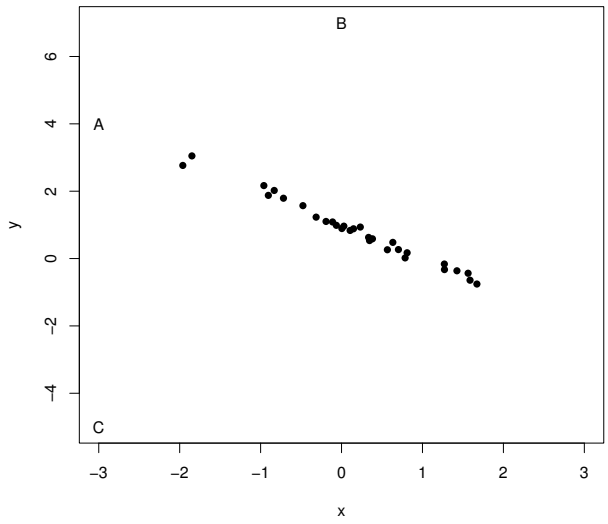
(b) lm(y∼x+I(x^2)).

(c) glm(y∼x+I(x^2), family = binomial).

(d) glm(y∼x+I(x^2), family = poisson).

Sol. c.

6. (6 marks) Match the points indicated by letters in the following scatterplot to the appropriate terms.

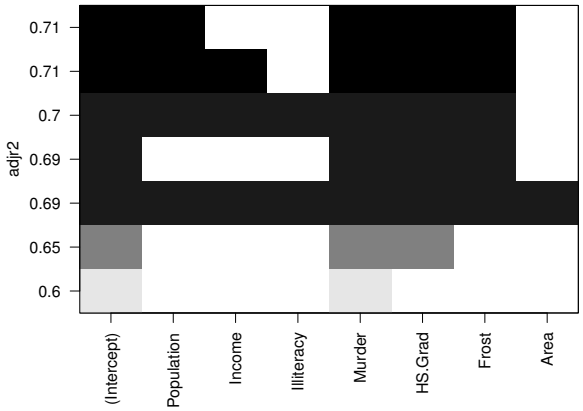


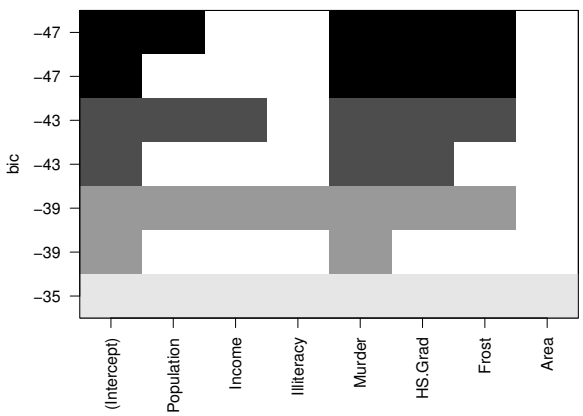
2 (a) Outliers SOL. B,C

2 (b) Influential points SOL. B,C

2 (c) High leverage points SOL. A,C

7. (4 marks) Consider the following partial R ouptut of a multiple linear regression model.





2 (a) What's the best model based on adjusted R^2? Sol. Population+Murder+HS.Grad+Frost

2 (b) What's the best model based on BIC? Sol. Population+Murder+HS.Grad+Frost