

INTRODUCTION TO STATISTICAL METHODS

SPRING 2021 | EXAM 2

Total Scores: 20 Points + 3 Bonus Points

1. Suppose you have collected a mini-dataset of 8 observations with two variables: father's years of schooling and son's years of schooling.

- (a) [2pts] State a relevant research question based on these two variables.
- (b) [2pts] Write down the formula for a bivariate linear regression model related to your research question. (Note: You should define the variable notations first. e.g. $Y=?$, $X=?$)
- (c) [4pts] You have generated the following information using R. With these statistics, calculate your OLS estimators for the slope and intercept parameters.

```
> describe(school)
      vars n  mean   sd median
father   1  8 12.50  3.59     12
son      2  8 14.38  3.29     14
> cov(school$father,school$son)
[1] 10.21429
```

* NOTE: In the output, "vars" stands for "variables" and "sd" stands for "standard deviation."

- (d) [2pts] Based on information in the R output above, what is the correlation between father's years of schooling and son's years of schooling?
- (e) [2pts] What proportion of the variation in Y can be explained by X?

2. Consider the following prediction equation for a multiple regression:

$$\hat{Y} = 2 + 2X_1 - 3X_2$$

- (a) [2pts] If we fix X_2 at zero, that is, set $X_2 = 0$, the prediction equation above will become a simple bivariate regression prediction equation that involves only Y and X_1 :
 $\hat{Y} = 2 + 2X_1 - 3X_2 = 2 + 2X_1 - 3 \cdot 0 = 2 + 2X_1$

Now, suppose that we fix X_2 at 1, that is, $X_2 = 1$. Follow the similar steps above and write down what would the prediction equation become.

(b) [2pts] If we draw two lines to represent the two equations in (a), will the slopes of these two lines be the same or different? What does this imply?

(d) [2pts] Consider a different prediction equation:

$$\hat{Y} = 2 + 2X_1 - 3X_2 + 2X_1X_2$$

What would be the prediction equation if we set $X_2 = 0$ and $X_2 = 1$, respectively?

(e) [2pts] Are the slopes of the two lines in (d) the same? What does it imply?

3. (Bonus 3pts) Consider the prediction equation:

$$\hat{Y} = a + bX,$$

where a and b are the intercept and slope coefficients respectively.

Suppose we construct two new variables Y^* and X^* :

$$Y^* = 10 \cdot Y + 2$$

$$X^* = 0.1 \cdot X$$

Suppose our sample data remain the same, but we now regress Y^* on X^* and obtain the following prediction equation:

$$\hat{Y}^* = a^* + b^*X^*$$

.

How will the new intercept (a^*) and slope (b^*) relate to the original a and b ? (i.e., I'm asking you to express a^* and b^* as functions of a and b .)