

INSTRUCTOR:  
XIAOPING SHI



IRVING K. BARBER FACULTY OF SCIENCE  
UBC, OKANAGAN CAMPUS

This exercise set has an additional 6 questions, for a total of 13 points. These questions, in addition to those on other practice sets, are representative of what might be seen on Quiz 1.

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) Regarding the  $R^2$  value from the linear regression  $y_i = a + bx_i + \varepsilon_i$ ,  $i = 1, 2, \dots, n$ , which of the following statements are true?
  - (a) total amount of variation in the model explained by the response variable
  - (b) proportion of variation in the model explained by the response variable
  - (c)  $0 \leq R^2 \leq 1$ .
  - (d) If  $x$  and  $y$  are uncorrelated, then  $R^2 = 0$ .

Sol. c, d.

2. (2 marks) Regarding the  $p$ -value, which of the following statements are true?
  - (a) The probability of observing a test statistic as or more extreme than that which we observed.
  - (b) The probability of observing a test statistic as or more extreme than that which we observed, assuming the alternative hypothesis is true.
  - (c) If  $p$ -value is extremely small, then we reject the null hypothesis.
  - (d) If  $p$ -value is extremely small, then we reject the alternative hypothesis.

Sol. c.

The next two questions will refer back to this paragraph. When we fit a model, say  $\hat{f}$ , to data arising from some unknown true model  $f$ , it is useful to keep in mind the following. If  $Y$  is numeric response,  $\hat{f}$  is the predicted response from our model,  $X$  is the set of predictors, and  $\varepsilon$  the error term, it is possible to show that the mean squared error for a model is

$$E[(Y - \hat{f})^2] = \text{Var}(\hat{f}) + [\text{Bias}(\hat{f})]^2 + \text{Var}(\varepsilon)$$

3. (2 marks) Regarding what was discussed above, which of the following statements regarding  $\varepsilon$  are FALSE?
  - (a)  $\varepsilon$  is the error that exists in the "true" model:  $Y = f(X) + \varepsilon$ .
  - (b)  $\varepsilon$  is considered irreducible error in the MSE.
  - (c)  $\varepsilon$  and  $X$  are dependent.
  - (d) The mean value of  $\varepsilon$  is zero.

Sol. c.

4. (2 marks) Again, regarding the above paragraph,, the bias of the fitted model is...
  - (a)  $f(X) - E[\hat{f}(X)]$
  - (b)  $\{f(X) - E[\hat{f}(X)]\}^2$
  - (c)  $\varepsilon$
  - (d)  $\text{Var}(\varepsilon)$

Sol. a.

5. (2 marks) Which of the following statements generally hold true about testing and training sets?

- (a) The MSE of the test set equals the MSE of the training set.
- (b) The MSE of the test set is less than the MSE of the training set.
- (c) The MSE of the test set is larger than the MSE of the training set.
- (d) The MSE of the test set does not generally have any relationship with the MSE of the training set.

Sol. c.

6. Consider the following partial R output of a simple linear regression model.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.34602	0.40176	***	0.414
x	0.09667	***	0.330	0.750

Residual standard error: 1.232 on 8 degrees of freedom

Multiple R-squared: 0.01347, Adjusted R-squared: -0.1099

F-statistic: 0.1092 on 1 and 8 DF, p-value: 0.7495

1 (a) Write the equation of predicted response.

2 (b) b) Fill in the spaces marked with “\*\*\*”.

Sol. a)  $\hat{y} = 0.34602 + 0.09667x$ . b) 0.34602/0.40176, 0.09667/0.330.