Quiz 8

Due 5 Dec at 23:59 **Points** 20 **Questions** 20 **Time limit** None
Allowed attempts 2

Attempt history

	Attempt	Time	Score
KEPT	Attempt 2	9 minutes	20 out of 20
LATEST	Attempt 2	9 minutes	20 out of 20
	Attempt 1	5,356 minutes	17 out of 20

Score for this attempt: **20** out of 20
Submitted 5 Dec at 15:33
This attempt took 9 minutes.

Question 1

1 / 1 pts

A regression model in which more than one independent variable is used to predict the dependent variable is called

- a simple linear regression model.
- Correct! a multiple regression model.
- an independent model.
- a chi-square analysis model.

Question 2

1 / 1 pts

We are provided 50 observations to predict house price with 20 candidate predictors. Two different variable selection algorithms provide two multiple linear regression models.

Model 1 includes 18 predictors, shows a R^2 of 0.85 and an adj R^2 of 0.56.

Model 2 includes 3 variables with R^2 of 0.81 and adj R^2 of 0.63.

Which model is considered better?

- Model 1
- Correct! Model 2
- Neither is good

For multiple linear regression model, we evaluate the model performance using adjusted R^2 because it has a penalty term to control the number of predictors in the model. R^2 always increases with increasing number of predictors which can lead to overfitting issue.

Question 3

1 / 1 pts

A multiple regression model has the following equation: $Y = 7 + 2X_1 + 9X_2$

In which of following situations, is Y expected to increase by 2 units?

- Correct! increase x1 by 1 unit while holding x2 constant
- increase x1 by 2 unit while holding x2 constant
- increase x1 and x2 both by 1 unit
- increase x2 by 1 unit while holding x1 constant

Question 4

1 / 1 pts

In backward selection regression with 10 predictor variables, which of the following statements is correct?

Correct!

- (selected) The procedure starts with 10 predictor variables and eliminate one predictor at each iteration
- The procedure starts with 2 predictor variables and add one predictor at each iteration
- The procedure starts with 1 predictor variables and add one predictor at each iteration
- The procedure starts with no predictor variables and add one predictor at each iteration

Question 5

1 / 1 pts

In multiple linear regression,

- there can be any number of response variables but only one predictor.
- there must be only one predictor.
- the R-square value must be larger than 1.

Correct!

- (selected) there can be several predictors, but only one response variable.

Question 6

1 / 1 pts

A health science-kinesiology program to lose weight collected data from ten students. Gender variable was coded as 1=female and 0=male. Time variable represents the number of periods. The regression equation obtained by using a statistical software was: Pounds lost = 15.8 + 0.65time + 6.00gender

The estimated weight loss of a female who stayed in the program for 5 periods is:

- 19.05
- 15.8
- 3.25
- Correct! 25.05

For female, gender takes value of 1, so Pounds lost = 15.8 + 0.65*5+ 6.00*1=25.05

Question 7

1 / 1 pts

A model is built to study the effect of (self-reported) exercise on weight in college students while considering their age. There are 4 values for the exercises: daily, weekly, sometimes, never. If we use dummy variables to represent the exercise levels and set never as the reference, what the equation would be like?

- Correct! weight=a+b_1*daily+$b_2$weekly+b_3*sometimes+b_4*age
- weight=a+b_1*daily+$b_2$weekly+b_3*never+b_4*age
- weight=a+b_1*never+$b_2$weekly+b_3*sometimes+b_4*age
- weight=a+b_1*daily+$b_2$weekly+b_3*sometimes+b_4*never+b_5*age

There should be 4-1=3 dummy variables. 'never' is the reference level so it should not be included in the model. When the level is never, all of dummy variables daily, weekly, and sometimes are taking value of 0.

Question 8

1 / 1 pts

If Forbes includes years of education and age to predict CEO salary, what would collinearity represent?

Correct!

- (●) High correlation between CEO age and years of education.
- ( ) A linear relationship between the predictors and CEO salary.
- ( ) Both predictors correlating with variables not included in the regression model.
- ( ) Independence between years of education and age.

The collinearity in the linear regression is the linearity between independent variables.

Question 9

1 / 1 pts

A regression equation was built to predict student weight with the number of hours spent on the computer (x_1), the number of hours spent watching television (x_2), and the number of hours spent on the telephone (x_3). The estimated regression equation is

$y = 125.5 + 3.26\, x_1 + 4.27\, x_2 + 1.75\, x_3$.

What is the predicted weight if the student spends 4 hours on the computer, 2 hours viewing television, and 3 hours on the telephone?

Correct!

- (•) 152.33
- ( ) 153.22
- ( ) 151.83
- ( ) 154.35
- ( ) 155.86

y = 125.5 + 3.26*4 + 4.27*2+ 1.75*3=152.33

Question 10

1 / 1 pts

To predict students' test scores, which of the following independent variables would NOT best be represented by dummy variable(s)?

- ( ) gender
- ( ) race

Correct!

- (•) number of hours studying for the test
- ( ) marital status

Number of hours studying for the test is a numerical variable so it should not be represenetd by dummy variable(s).

Question 11

1 / 1 pts

Which of these problems arise as collinearity increases?

- Difficulty in detecting significant variables
- Reduced precision of coefficient etimates

Correct!

- (•) All of these answers are correct
- Inflated standard error of the coefficient estimates

Question 12

1 / 1 pts

Forward regression is an iterative procedure that:

Correct!

- (•) adds one independent variable at a time
- deletes one independent variable at a time
- initially contains all predictor candidates in the model
- both A and B are correct

Forward selection is an iterative procedure that starts with no predictor in the model and add one independent variable at a time.

Question 13

1 / 1 pts

A model is built to study the effect of (self-reported) exercise on weight in college students while considering their age. There are 4 values for the exercises: daily, weekly, sometimes, never. We use dummy variables to represent the exercise levels and set daily as the reference.

The equation is estimated to be
weight=105.5+3.56weekly+9.1sometimes+12never+0.72age.

What will be the predicted weight for students who are 30 years old and doing weekly exercises?

- 125.56
- Correct! 130.66
- 132.33
- 135.88

weight=105.5+3.56*1+9.1*0+12*0+0.72*30

Question 14

1 / 1 pts

What factors affect the sale price of oceanside condominium units? To answer this question, the following data were recorded for each of the n = 105 units sold at auction:

y = sale price (**in thousands of dollars**)

x_1 = floor height (1, 2, 3, ..., 8)

x_2 = view type (1 is ocean view; 0 is bay view)

The derived equation is $y = 17.770 - 0.073\ x_1 + 3.137\ x_2$.

Interpret the estimated effect of floor height.

- Expected sale price increases by $73 for each one unit increase in floor height for the same view type.
- Correct! Expected sale price decreases by $73 for each one unit increase in floor height for the same view type.

Expected sale price decreases by $73 for each one floor increase in floor height for different view types

The correlation between selling price and floor height is -0.073.

Question 15

1 / 1 pts

If Forbes includes years of education, age, and their interaction to predict CEO salary, which of the following methods is NOT appropriate to validate the assumptions?

Using the scatter plot of residual vs predicted y plot to check the linearity and constant variances.

Using the QQ plot of residuals to check the normality.

Using the Cook's D to detect outlier.

Correct!

Using the QQ plot of residuals to check the independence

The independence is related to the data collection step not something we can really test for.

Question 16

1 / 1 pts

Forbes has modeled CEO salary by age for an exclusive feature in the magazine. Which of the following statements is **NOT** true about outliers in the model?

Outliers, if included in the model, can influence the coefficient estimates.

Correct!

- Outliers only represent distinguished CEOs in the data set.
- Outliers, if included in the model, can cause the predicted response to be very different compared to model without the outliers
- Outliers are values where CEO salary against age is very different from others in the data set

Question 17

1 / 1 pts

In general, to represent a categorical predictor variable that has *n* possible categories,

(Note: dummy variables are also called indicator variables.)

- We must create (*n* + 2) dummy variables
- We must create (*n* + 1) dummy variables
- We must create *n* dummy variables

Correct!

- We must create (*n* – 1) dummy variables

Question 18

1 / 1 pts

In a multiple regression model involving more than one independent variable, which of the following tests is used to test the significance of the relationship between the dependent variable and **the whole set of independent variables**?

- t-test

Correct!

- F-test

chi-square test

Either a t-test or an F-test can be used -- they are mathematically equivalent in this case.

Question 19

1 / 1 pts

When using a multiple regression model for statistical inference, we assume that errors are following a(n) _____ distribution.

Binomial

Correct!

Normal

Exponential

Poisson

Question 20

1 / 1 pts

Stepwise regression is an iterative procedure combining both of forward and backward selection. Which of the following statement is true?

It adds one independent variable at a timea

It removes one independent variable at a time

Neither A nor B is correct.

Correct!

Both A and B are correct

Stepwise regression is an iterative procedure combining both of forward and backward selection. Therefore, it will add and remove one predictor at each step according to the pre-specified significance level of entry and stay.

Quiz score: **20** out of 20