

Started on	Sunday, May 31, 2020, 5:09 PM
State	Finished
Completed on	Sunday, May 31, 2020, 5:25 PM
Time taken	15 mins 56 secs
Grade	76.50 out of 100.00

Question **1**

Correct

1.00 points out of 1.00

I certify that all work contained within is my own. I neither received any assistance from anyone nor gave assistance to anyone.

Select one:

☒ True

☐ False

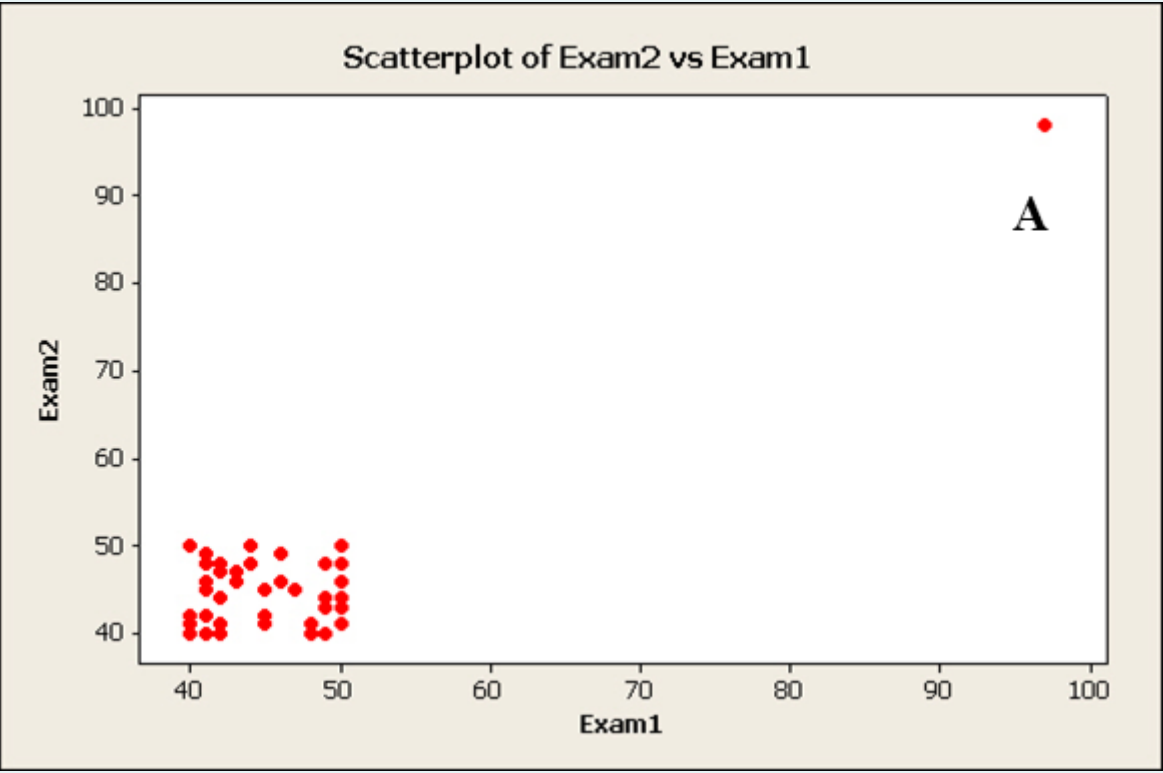
The correct answer is 'True'.

Question **2**

Incorrect

0.00 points out of 7.00

There is an outlier in this data (marked with the letter "A"). If we removed this outlier the following would happen to the correlation coefficient.



Select one:

- ☐ a. Get closer to zero.
- ☒ b. Get closer to one. ✖
- ☐ c. Stay the same.
- ☐ d. Need more information.

The correct answer is: Get closer to zero.

Question **3**

Correct

7.00 points out of 7.00

If a categorical variable has 5 levels, how many dummy variables are needed?

Select one:

- ☐ a. 2
- ☐ b. 3
- ☐ c. 5
- ☐ d. None. Do not need to create dummy variables.
- ☒ e. 4 ✔
- ☐ f. 1

Your answer is correct.

The correct answer is: 4

Question 4

Correct

7.00 points out of 7.00

A marketing researcher develops two models to try and explain the usage of products at a convenience store. His first model has an $R^2 = 0.722$ and an adjusted $R^2 = 0.703$. His second model is the same as the first model with one additional variable. His second model has an $R^2 = 0.727$ and an adjusted $R^2 = 0.691$. What can we conclude?

Select one:

- ☐ a. You cannot compare adjusted R-squared values between models.
- ☐ b. The second model is probably better because it has a higher R-squared value.
- ☒ c. The additional variable in the second model probably hinders the overall model more than it helps it. ✓
- ☐ d. Something is calculated incorrectly because the adjusted R-squared is always larger than R-squared.

The correct answer is: The additional variable in the second model probably hinders the overall model more than it helps it.

Question 5

Incorrect

0.00 points out of 7.00

Which of the following is **not** a characteristic of the F-distribution?

Select one:

- ☐ a. Skewed Right
- ☒ b. Doesn't take on negative values. ✗
- ☐ c. Symmetric
- ☐ d. Has two sets of degrees of freedom.

The correct answer is: Symmetric

Question 6

Incorrect

0.00 points out of 7.00

An analyst is building a multiple regression model to model income. In the model, one of the analyst's variables is a categorical variable for age group that has four categories - teen, young adult, middle-aged, elderly. The analyst is interested in individually comparing the teenage group to each of the other groups individually. How should they code their dummy variables?

Select one:

- ☐ a. Use response coding with 3 dummy variables for young adult, middle-aged, and elderly.
- ☐ b. Use effects coding with 3 dummy variables for young adult, middle-aged, and elderly.
- ☒ c. Use response coding with 4 dummy variables (one for each category). ✗
- ☐ d. Multiple regression cannot determine relationships between categorical variables and continuous variables.

The correct answer is: Use response coding with 3 dummy variables for young adult, middle-aged, and elderly.

A data analyst for a major car manufacturer is trying to develop a pricing model for the company's cars. The analyst came up with the following model from a sample of 54 cars.

- x_1 = Maximum speed of vehicle (mph)
- x_2 = Estimated highway miles per gallon
- x_3 = Maximum number of occupants
- x_4 = Weight of vehicle (pounds)

$\hat{y} = 620.8 + 127.1x_1 - 6.87x_2 + 985.4x_3 + 5.5x_4$
SSE = 101578, SSR = 236165

Use this information to answer the next 4 questions.

Question **7**

Complete

7.50 points out of 10.00

Interpret the coefficient for x_3 in terms of the problem.

As the number of occupants that can fit in a car increases by one, the average value / price of the car is expected to increase by \$985.40.

Comment:
All else constant.

Question **8**

Complete

10.00 points out of 10.00

Calculate the R^2 and adjusted R^2 value.

$R^2 = 0.6992$

$R^2 \text{ Adjusted} = 0.6755$

Comment:

Question **9**

Complete

10.00 points out of 10.00

What is the null and alternative hypotheses for the F-test for overall model significance for this model?

$H_0: B_1 = B_2 = B_3 = B_4 = 0$

H_A : At least one coefficient is nonzero

Comment:

Question **10**

Complete

10.00 points out of 10.00

Calculate the F-statistic for overall model significance in this model. Do not compute the entire hypothesis test, just the F-statistic for the hypothesis test.

28.4807

Comment:

Question **11**

Correct

7.00 points out of 7.00

Why do we calculate adjusted R^2 values in multiple linear regression instead of just R^2 values?

Select one:

- ☐ a. Mathematically R-squared values always increase for extra variables added to a model.
- ☐ b. R-squared values only describe a simple linear regression, while adjusted R-squared values describe a multiple linear regression.
- ☐ c. The adjusted R-squared only increases if the addition of another variable outweighs the loss of degrees of freedom.
- ☒ d. Both a and c, but not b. ✓

The correct answer is: Both a and c, but not b.

Question **12**

Correct

7.00 points out of 7.00

A data analyst is trying to interpret the following results of a simple linear regression with an $R^2 = 0.84$.

Parameter	Estimate	Std. Error	D.F.	T-statistic	P-value
Intercept	22.57	3.53	48	6.394	< 0.0001
Slope	-9.82	1.32	48	-7.439	< 0.0001

What is the value of the correlation coefficient r between the response and explanatory variable?

Select one:

- ☐ a. 0.917
- ☐ b. 0.84
- ☒ c. -0.917 ✓
- ☐ d. Not enough information

The correct answer is: -0.917

Question **13**

Correct

10.00 points out of 10.00

You have the following regression model:

$y\text{-hat} = 2 + 3x_1 - 6x_2$

For a given value of $x_2=10$, what is the predicted change in y for a single unit increase in x_1 ?

Answer: ✓

The correct answer is: 3