Started on	Tuesday, 28 September 2021, 2:33 PM
State	Finished
Completed on	Tuesday, 28 September 2021, 3:56 PM
Time taken	1 hour 23 mins
Grade	**40.90** out of 77.00 (**53**%)

Information

You have 80 minutes to complete the test. You can skip questions, and you can "flag" questions to remember to return to them. If you wish to return to a question, please use the question navigation in the upper-right of the screen, rather than the "back" button on your browser.

Be sure to show your work and justify your responses, citing graphs, tables, numbers, and/or models in the R output, for full credit. Please answer questions as concisely and precisely as you are able. You will not be rewarded for unnecessary explanation. Make sure to state all conclusions in context. Throughout the test, you may assume that the conditions of the model are met when making conclusions (at least, until explicitly asked about those conditions).

This test is open-resource. That means that you may use: the book, your notes, the videos, your homework and my solutions, and activities and their solutions. You may also Google things to find different explanations of concepts beyond those given in the textbook.

You may NOT ask another human being for help, accept help from another human being, or post a question on the internet for others to solve. You may ask me for clarification questions, but you should not speak (verbally or virtually) to *anyone else* about the test. Violating these rules is a serious breach of academic integrity and will result in earning a 0% on the Test. In addition, the Dean for Curriculum and Academic Engagement will be notified; depending on the student's history of academic misconduct, this could lead to expulsion from the College.

A reminder of the Wooster Ethic:

I hereby join this community with a commitment to the Wooster Ethic upholding academic and personal integrity and a culture of honesty and trust in all my academic endeavors, social interactions, and official business of the College. I will submit only my own original work, and respect others and their property. I will not support by my actions or inactions the dishonest acts of others.

Question 1

Complete

Not graded

By typing my full name here, I agree to abide by the Wooster Ethic and the academic honesty policies described above.

Sarah Wright

Question 2

Partially correct

Mark 0.90 out of 3.00

Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol. An ANOVA F-test for model utility is performed and returns a p-value of 0.003. Assuming all the conditions for the model are met, which of the following is true? (select all that apply)

Select one or more:

- a. The probability that there IS a linear relationship is 0.003.
- b. If there is NO linear relationship between HDL and total cholesterol, the probability of getting results as or more extreme than ours is about 0.003.
- c. If there IS a linear relationship between HDL and total cholesterol, the probability of getting results as or more extreme than ours is about 0.003. ✖
- d. The relationship between HDL and total cholesterol is strong.
- e. The evidence for an association between HDL and total cholesterol is strong. ✔
- f. The relationship between HDL and total cholesterol is important (that is, important in the "real world").
- g. The probability that there is NO linear relationship is 0.003.

Your answer is partially correct.

You have correctly selected 1.
The correct answers are: If there is NO linear relationship between HDL and total cholesterol, the probability of getting results as or more extreme than ours is about 0.003., The evidence for an association between HDL and total cholesterol is strong.

Question 3

Partially correct

Mark 0.50 out of 2.00

Conditions for a linear regression model include (select all that apply):

Select one or more:

- a. the y-variable must be normally distributed ✖
- b. the x-variables must be normally distributed ✖
- c. the relationship between y and x must be normal ✖
- d. the error terms must be normally distributed ✔
- e. none of the above

Your answer is partially correct.

You have selected too many options.
The correct answer is: the error terms must be normally distributed

Question 4

Correct

Mark 2.00 out of 2.00

A simple linear regression model for predicting the number of calories in breakfast cereals from their sugar content results in the fitted line:

$$\hat{Calories} = 87.428 + 2.481(Sugar)$$

A cereal has a sugar content of 2 grams. Use the fitted line to predict its calorie count, and report the prediction to 2 decimals. (Do not include units in your answer.)

Answer: 92.39 ✔

The correct answer is: 92.39

Question 5

Partially correct

Mark 1.00 out of 2.00

R-squared is a measure of: (select all that apply)

Select one or more:

- a. how much variability in Y is explained by the model ✔
- b. if the conditions of a linear model are met
- c. if a linear model is appropriate for this data set
- d. how well the model fits the data

Your answer is partially correct.

You have correctly selected 1.
The correct answers are: how well the model fits the data, how much variability in Y is explained by the model

Question 6

Incorrect

Mark 0.00 out of 2.00

Influential points should always be dropped from the data set.

Select one:

- True ✖
- False

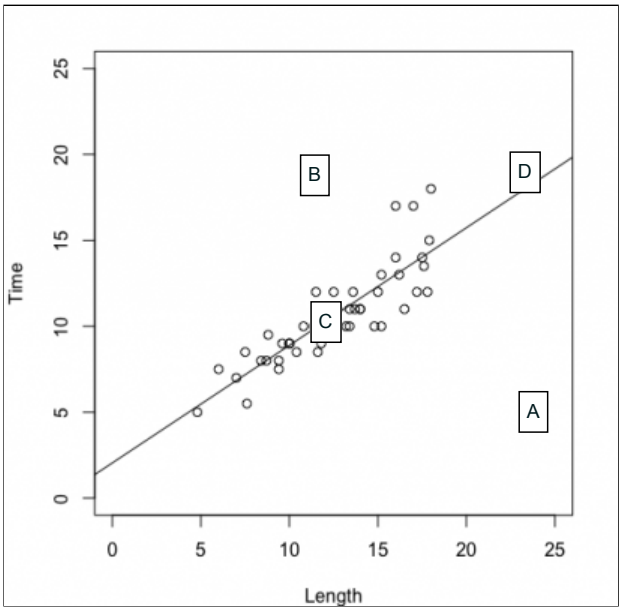
The correct answer is 'False'.

Question 7

Correct

Mark 2.00 out of 2.00

Below is a scatterplot. You should add two points to the graph. Specifically, add: 1) A point that is an outlier AND is influential, which you will label "A"; 2) A point that is an outlier but is not influential, which you will label "B". (Moodle will ask you to place "C" and "D" as well, but it doesn't matter where you put them.)



Your answer is correct.

Comment:

Question 8

Partially correct

Mark 1.50 out of 2.00

A group of breakfast cereals are used to build a linear model that predicts number of calories from sugar content. We wish to test if there is a linear relationship between these two variables. What would be valid hypotheses to test this? (select all that apply)

Select one or more:

- [x] a. Ho: correlation between sugar and calories = 0; Ha: correlation between sugar and calories does not equal 0 ✔
- [ ] b. Ho: slope of linear model does not equal 0; Ha: slope of linear model = 0
- [ ] c. Ho: intercept of linear model = 0; Ha: intercept of linear model does not equal 0
- [ ] d. Ho: X=0; Ha: X does not equal 0
- [x] e. Ho: slope of linear model = 0; Ha: slope of linear model does not equal 0 ✔
- [x] f. Ho: linear model = 0; Ha: linear model does not equal 0 ✖

Your answer is partially correct.

You have selected too many options.
The correct answers are: Ho: slope of linear model = 0; Ha: slope of linear model does not equal 0, Ho: correlation between sugar and calories = 0; Ha: correlation between sugar and calories does not equal 0

Information

The rest of questions on this test relate to the data set and R code/output described in the Output packet you were handed. This packet totals 6 pages.

Read the description of the data set and variables now. The questions below ask about "Var1", "Var3", etc., but all your answers should be *in context* for your data set. (That is, you should use variable names in your answers, not "Var1", "Var3", etc.)

Question 9

Incorrect

Mark 0.00 out of 3.00

Based on the correlation matrix on the first page of output, which variable is the best single predictor of *Var1*?

Select one:

- a. Var3
- b. Var4
- c. Var6
- d. Var5
- e. No way to tell based on the information given ✖

Your answer is incorrect.

The correct answer is: Var6

Question 10

Correct

Mark 2.00 out of 2.00

Use the regression output from **lm1** to conduct a test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$. What is the p-value?

- a. No way to tell based on the information given
- b. 7.22*10^-13
- c. 363.2
- d. approximately 0 ✔
- e. 11.924
- f. -7.698
- g. 0.4237

Your answer is correct.

The correct answer is:
approximately 0

Question 11

Complete

Mark 3.00 out of 3.00

Use the regression output from **lm1** to conduct a test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$. Based on the p-value from the previous question, make a conclusion in context.

The test of slope=0 has a p-value of approx. 0, so yes, there is a significant association between the points total and minutes played per game for 193 NBA basketball Association basketball players from 2018-2019.

Comment:

Question 12

Complete

Mark 0.00 out of 3.00

Using **lm1**, a 90% prediction interval is given in the output. Interpret this interval in context.

We are 90% confident that as the minutes per game increases by 1%, we expect the points total to increase between 1331.904 and 1477.806.

Comment:

Question 13

Correct

Mark 3.00 out of 3.00

We wish to predict *Var1* using *Var6*. Previous analysis indicates that $Var6^2$ might be useful in the model, so we fit this model, called **lm2**. Based on this output, was including $Var6^2$ a good idea?

- a. No, we definitely should not include Var6^2 in the model
- b. Maybe? I'll explain below. ✔
- c. Yes, we should definitely have Var6^2 in the model

Your answer is correct.

The correct answers are:
Yes, we should definitely have Var6^2 in the model,

No, we definitely should not include Var6^2 in the model,

Maybe? I'll explain below.

Question 14

Complete

Not graded

If you answered "Maybe" above, explain your answer here.

It depends on if you want to add more complexity to the model. They would both be equally effective based on the adj r^2

Question 15

Complete

Mark 0.00 out of 3.00

We wish to predict *Var1* using *Var6*, *Var4*, and *Var3*; this is called **lm3**. Interpret the coefficient on *Var3* in context.

Based on the coefficient for lm3, we can conclude that Game Starts is not an effective predictor

Comment:

Information

Consider the set of plots for **lm3**; notice that they are labeled "PLOT A" and "PLOT B". Based *just on the information in each plot*, select one of the choices for each of the conditions of a linear model listed below to indicate whether the plot shows the condition is "reasonable", "problematic" or you "can't judge" the condition *from the plot shown*.

Question 16

Correct

Mark 1.00 out of 1.00

Using PLOT A, what are your thoughts about **linearity**?

Select one:

- a. Reasonable ✔
- b. Problematic
- c. Can't judge

Your answer is correct.

The correct answer is: Reasonable

Question 17

Incorrect

Mark 0.00 out of 1.00

Using PLOT A, what are your thoughts about **independence**?

Select one:

- a. Reasonable ✖
- b. Problematic
- c. Can't judge

Your answer is incorrect.

The correct answer is: Can't judge

Question 18

Correct

Mark 1.00 out of 1.00

Using PLOT A, what are your thoughts about **normality**?

Select one:

- a. Reasonable
- b. Problematic
- c. Can't judge ✔

Your answer is correct.

The correct answer is: Can't judge

Question 19

Incorrect

Mark 0.00 out of 1.00

Using PLOT A, what are your thoughts about **constant variance**?

Select one:

- a. Reasonable
- b. Problematic
- c. Can't judge ✖

Your answer is incorrect.

The correct answer is: Reasonable

Question 20

Incorrect

Mark 0.00 out of 1.00

Using PLOT A, are there any influential points?

- a. Yes ✖
- b. No
- c. Can't judge

Your answer is incorrect.

The correct answer is:
Can't judge

Question 21

Complete

Not graded

If you would like to add explanation or comments about any of your answers above, please do so here.

Thank you for asking. ^_^

Question 22

Correct

Mark 1.00 out of 1.00

Using PLOT B, what are your thoughts about **linearity**?

Select one:

- a. Reasonable
- b. Problematic
- c. Can't judge ✔

Your answer is correct.

The correct answer is: Can't judge

Question 23

Correct

Mark 1.00 out of 1.00

Using PLOT B, what are your thoughts about **independence**?

Select one:

- a. Reasonable
- b. Problematic
- c. Can't judge ✔

Your answer is correct.

The correct answer is: Can't judge

Question 24

Incorrect

Mark 0.00 out of 1.00

Using PLOT B, what are your thoughts about **normality**?

Select one:

- a. Reasonable ✖
- b. Problematic
- c. Can't judge

Your answer is incorrect.

The correct answer is: Can't judge

Question 25

Incorrect

Mark 0.00 out of 1.00

Using PLOT B, what are your thoughts about **constant variance**?

Select one:

- a. Reasonable ✖
- b. Problematic
- c. Can't judge

Your answer is incorrect.

The correct answer is: Can't judge

Question 26

Incorrect

Mark 0.00 out of 1.00

Using PLOT B, are there any influential points?

- a. Yes ✖
- b. No
- c. Can't judge

Your answer is incorrect.

The correct answer is:
No

Question 27

Complete

Not graded

If you would like to add explanation or comments about any of your answers above, please do so here.

Thank you, again, for asking.

Question 28

Correct

Mark 2.00 out of 2.00

What proportion of the variability in *Var1* is explained by the regression model lm3?

a. 0.5452

b. 322.6

c. No way to tell based on the information given

d. 0.5523 ✔

e. 77.72

Your answer is correct.

The correct answer is:
0.5523

Question 29

Correct

Mark 2.00 out of 2.00

Based only on the output from lm3, would you remove any of the variables from the model?

Select one:

a. I would not remove any variables ✔

b. I would remove *Var6*

c. I would remove *Var3*

d. I would remove *Var4*

Your answer is correct.

The correct answer is: I would not remove any variables

Question 30

Complete

Mark 0.00 out of 1.00

Briefly justify your answer to the previous question.

These variables are not highly correlated.

Comment:

Question 31

Complete

Mark 0.00 out of 4.00

Using the output below model **lm3**, report and interpret a confidence interval for the coefficient on *Var4*.

We are 97.5% confident that the for the 193 NBA basketball players the Total Rebounds is between 0.45677 and 0.9498 pts.

Comment:

Question 32

Correct

Mark 2.00 out of 2.00

Given the other variables in model **lm3**, is *Var4* a useful predictor of *Var1*?

a. No

b. Yes ✔

c. No way to tell based on the information given

d. Somewhat

Your answer is correct.

The correct answers are:
Yes,

Somewhat

Question 33

Complete

Mark 2.00 out of 2.00

Justify your answer to the previous question. Be specific!

Because the p value is almost zero, we can assume the Total Rebounds of a player are a good predictor of the total points scored during the season.

Comment:

Question 34

Correct

Mark 2.00 out of 2.00

Are you concerned about multicollinearity in the lm3 model?

a. No way to tell based on the information given

b. Yes

c. No ✔

d. Somewhat

Your answer is correct.

The correct answer is:
No

Question 35

Complete

Mark 2.00 out of 2.00

Justify your answer to the previous question.

This is because the VIF for all the variables presented in lm3 are smaller than 5.

Comment:

Question 36

Correct

Mark 3.00 out of 3.00

We wish to predict *Var1* using *Var6, Var4, Var3*, and *Var2*; this is called **lm4**. Using this model, the two groups in *Var2* would have different

Select one:

- a. slope
- b. intercepts
- c. both (a) and (b) ✔
- d. neither (a) nor (b)

Your answer is correct.

The correct answer is: both (a) and (b)

Question 37

Complete

Mark 0.00 out of 3.00

Consider model **lm4**. Interpret the value 71.7264 (it's one of the coefficient estimates) in context.

For every additional Player under 30 in the NBA the predicted Total Points goes up by 71.7264 pts.

Comment:

Previous activity

◀ Test A practice solutions

Jump to...

Next activity

Test-A-solutions ▶

Stay in touch

- https://technology.wooster.edu
- (330) 287-4357
- helpdesk@wooster.edu

Data retention summary

Get the mobile app