2023GFA_ANA_500_02 Foundations of Data Analytics

Announcements
Read Me First
Syllabus
Course Guide
Assignments

Modules
Discussions
Learning Journals
Exams

gretl

Library Resources
Writing Center
STEM Center
ODEI
Purchase Course Materials

My Grades
Help

Review Test Submission: Module 04 Week 8 gretl Assignment

User	Kohei Nishitani
Course	2023GFA_ANA_500_02 Foundations of Data Analytics
Test	Module 04 Week 8 gretl Assignment
Started	10/19/23 7:52 PM
Submitted	10/20/23 2:19 AM
Due Date	10/20/23 11:59 PM
Status	Completed
Attempt Score	300 out of 450 points
Time Elapsed	6 hours, 26 minutes
Instructions	This is the gretl assignment for Module 04, Week 8. In this assignment you will continue to explore ordinary least squares regression, particularly multiple variable or multivariable regression. I am uploading a complete Word doc below. As was the case last week, this document contains everything you need to complete the gretl assignment as well as a discussion about some of the concepts covered. The intention is to help you develop an intuitive understanding for what is going on with this type of regression. As usual, select the choice that best answers a question and round numeric answers to two decimal places. **As always, if you have question please ask!** ANA500 Week 8 gretl and PP assignments.docx

Question 1 0 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

The price variable is:

Question 2 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

The values in the price variable are in:

Question 3 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

The values in the sqft variable are in:

Question 4 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

The values in the age variable are in:

Question 5 0 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Based on the answers above, no adjustment or transformation should be required to interpret the results of analyses using these variables.

Question 6 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Restrict your data to traditional-style houses. Consider descriptive and summary statistics for your restricted dataset. Use the restricted dataset to answer the following questions.

How many observations are there?

Question 7 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Restrict your data to traditional-style houses. Consider descriptive and summary statistics for your restricted dataset. Use the restricted dataset to answer the following questions.

The correlation between traditional-style house prices and size is statistically significant.

Question 8 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Restrict your data to traditional-style houses. Consider descriptive and summary statistics for your restricted dataset. Use the restricted dataset to answer the following questions.

The value of the correlation coefficient is _____.

Question 9 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Create a scatter plot of house price versus house size for traditional style homes. Does the relationship between price and size appear to be linear?

Question 10 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Is the data skewed?

Question 11 0 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Based on your answer about any apparent skew after taking the natural log of the price variable, do you believe you may have to **further transform** your data to meet the assumptions required to build a regression model?

Question 12 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Now that you have transformed the price variable is the data still skewed?

Question 13 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Based on your answer about any apparent skew after taking the natural log of the price variable, do you believe you may have to further transform your data to meet the assumptions required to build a regression model?

Question 14 0 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Which of the following assumptions could be violated? (Select one)

Question 15 10 out of 10 points

Start by considering your dataset. Make a record of your answer to each of these questions, anyway you want to make this record, to use in answering later questions on this exam. Generate descriptive statistics for the variables price, sqft, and age. Use these statistics to answer the following questions.

Save your reduced dataset to a new data file, e.g. batonRouge-trad.gdt.

Create a scatter plot of the natural log of house price versus house size for traditional style homes that are owner occupied. Does the relationship between price and size appear to be linear now?

Question 16 10 out of 10 points

Generate a simple linear model for traditional style houses with price as a function of house size. That is,

(Equation provided in attached Word doc. Be sure to save the value for the sum of squares error (SSE) for this linear model.)

Interpret the estimates to answer the following questions.

Is house size statistically significant?

Question 17 10 out of 10 points

Generate a simple linear model for traditional style houses with price as a function of house size. That is,

(Equation provided in attached Word doc. Be sure to save the value for the sum of squares error (SSE) for this linear model.)

Interpret the estimates to answer the following questions.

How do these house prices vary with changes in size (change per square foot)?

Question 18 10 out of 10 points

Generate a simple linear model for traditional style houses with price as a function of house size. That is,

(Equation provided in attached Word doc. Be sure to save the value for the sum of squares error (SSE) for this linear model.)

Interpret the estimates to answer the following questions.

The intercept for the simple linear model is, practically speaking, realistic.

Question 19 0 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

What is the intercept value?

Question 20 0 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

The intercept for the quadratic model is, practically speaking, realistic.

Question 21 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

What is the coefficient of ?

Question 22 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

What is the marginal effect for a home with 2000 square feet of living area?

Question 23 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

What is the expected price of the 2000 square foot home?

Question 24 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

What is the elasticity of price with respect to living area for a traditional-style home with 2000 square feet of living area?

Question 25 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

Generate a scatter plot with both the linear and quadratic trend lines on it. Which seems to fit the data better?

Question 26 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

Generate a plot of the residuals from both the linear and quadratic models. Does homoscedasticity appear to be a problem?

Question 27 10 out of 10 points

Generate a quadratic model for this situation, that is , and use this model to answer the following questions.

(Insert or think of the equation provided in the attached Word doc. Be sure to save the value of the sum of squares error (SSE) for this quadratic model.)

Would this indicate that heteroscedascity or heteroskedascity is present in the data?

Question 28 10 out of 10 points

Generate a log-linear model for this situation, that is , and use this model to answer the following questions. (Be sure to save the sum of squares error (SSE) for this log-linear model.)

The house size in square feet is statistically significant.

Question 29 10 out of 10 points

Generate a log-linear model for this situation, that is , and use this model to answer the following questions. (Be sure to save the sum of squares error (SSE) for this log-linear model.)

The intercept of the log-linear model is statistically significant.

Question 30 0 out of 10 points

Generate a log-linear model for this situation, that is , and use this model to answer the following questions. (Be sure to save the sum of squares error (SSE) for this log-linear model.)

The intercept for the log-linear model is, practically speaking, realistic..

Question 31 0 out of 10 points

Generate a log-linear model for this situation, that is , and use this model to answer the following questions. (Be sure to save the sum of squares error (SSE) for this log-linear model.)

Visually, the _____ model appears to be the best fit for the data.

Question 32 10 out of 10 points

Generate a log-linear model for this situation, that is , and use this model to answer the following questions. (Be sure to save the sum of squares error (SSE) for this log-linear model.)

Compare the sum of squares error (SSE) for each model and select the model listed below that actually results in the least error.

Question 33 0 out of 10 points

Ultimately, the log-linear model results in higher house prices for very large houses.

Question 34 10 out of 10 points

Based on the results of the various tests for normality, _____ satisfy/satisfies the assumption of normality. (Hint: these tests are based on **the hypothesis that the data are normal to begin with**, i.e. If the P-value is < 0.05 we must reject the null hypothesis. In other words, when evaluating your results, keep in mind what it means to have a given hypothesis and the P-values you get from your results!)

Question 35 10 out of 10 points

Visually inspecting plots of residuals indicates that _____ satisfy/satisfies the assumption of normality.

Question 36 10 out of 10 points

Consider the plots of residuals generated in the part of your assignments. From visually inspecting the plot do the residuals appear to be relatively evenly distributed about zero?

Question 37 0 out of 10 points

Consider the differences in value for owner-occupied houses versus vacant/rental houses. You will need to subset the full dataset by the variable owner to do this. That is, you will have one where you restrict the data to owner=1, the other where owner=0. Generate limited log-linear models including the variables price, square feet (sqft) and age; one restricted to owner-occupied houses, the other for vacant or rental houses. Use your results to answer the following questions.

The mean of the price for owner-occupied houses is _____.

Question 38 0 out of 10 points

Consider the differences in value for owner-occupied houses versus vacant/rental houses. You will need to subset the full dataset by the variable owner to do this. That is, you will have one where you restrict the data to owner=1, the other where owner=0. Generate limited log-linear models including the variables price, square feet (sqft) and age; one restricted to owner-occupied houses, the other for vacant or rental houses. Use your results to answer the following questions.

The mean of the price for a vacant or rental house is _____.

Question 39 10 out of 10 points

Compare the frequency plots after transforming the price variable using a natural log transformation. Do the frequency plots indicate that by taking the natural log of price we have improved the normality of the distribution?

Question 40 0 out of 10 points

Using the original simple linear model developed earlier for traditional-style houses, test the null hypothesis that the expected price of a 2000 square foot house is equal to or less than $120,000. Use a level of significance equal to 0.05. Use your results to answer the following questions.

The upper limit of the 95% confidence interval is _____.

Question 41 0 out of 10 points

Using the original simple linear model developed earlier for traditional-style houses, test the null hypothesis that the expected price of a 2000 square foot house is equal to or less than $120,000. Use a level of significance equal to 0.05. Use your results to answer the following questions.

The lower limit of the 95% confidence interval is _____.

Question 42 10 out of 10 points

Using the original simple linear model developed earlier for traditional-style houses, test the null hypothesis that the expected price of a 2000 square foot house is equal to or less than $120,000. Use a level of significance equal to 0.05. Use your results to answer the following questions.

The P-value for sqft is _____.

Question 43 10 out of 10 points

Using the quadratic model developed earlier for traditional-style houses that are 2000 square feet in size, test the null hypothesis that the marginal effect of an additional square foot of living area is $75 against the alternate hypothesis that the effect is less than $75. Use a level of significance of 0.01. Based on the results of your hypothesis test you fail to reject the null hypothesis and conclude that for a 2000 square foot house, the marginal effect of adding a square foot of living area is less than $75.

Question 44 0 out of 10 points

Using the quadratic model developed earlier for traditional-style houses that are 4000 square feet in size, test the null hypothesis that the marginal effect of an additional square foot of living area is $75 against the alternate hypothesis that the effect is less than $75. Use a level of significance of 0.01. Based on the results of your hypothesis test you fail to reject the null hypothesis and conclude that for a 4000 square foot house, the marginal effect of adding a square foot of living area is less than $75.

Question 45 0 out of 10 points

Another plot of residuals was generated in the part of your assignments. The model the residuals were obtained from added the variable age to the variables used to generate the OLS model. Does it appear that adding another variable helped satisfy the assumption of regression?

Friday, October 20, 2023 2:19:27 AM EDT

← OK