Take Test: Module 04 Week 8 gretl Assignment

Test Information

Description	As was the case last week, I have setup this week's assignments so that you should complete the gretl assignment first and then the paper and pencil assignment. This week's assignments are intended to help you continue to prepare for the Final Examination scheduled to begin in just over 24 hours! The Final examination will be comprehensive, i.e. it will cover virtually everything we have covered in ANA 500. I sent out by email and am also uploading a complete Word doc desogmed tp go step-by-step through all the course material to help with your review. As was the case last week, this document contains everything you need to complete the gretl assignment as well as an expanded discussion about some of the concepts covered. The intention is to help you not only review this material but to continue to develop an intuitive understanding for what is going on with this type of regression. Through discussions with other Data Analytics faculty I've been told that one of the things students need more work on is building and interpreting models. So, you will see that this week's assignments focuses on that. 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
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
Multiple Attempts	This test allows 2 attempts. This is attempt number 2.
Force Completion	This test can be saved and resumed later.
	Your answers are saved automatically.

Question Completion Status:

QUESTION 1 — 10 points — Saved

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:

- ○ Numeric, discrete
- ◉ Numeric, continuous
- ○ Logical
- ○ Categorical

QUESTION 2 — 10 points — Saved

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:

- ◉ Dollars (USD)
- ○ 10's of dollars (USD)
- ○ 100's of dollars (USD)
- ○ 1,000's of dollars (USD)

QUESTION 3 — 10 points — Saved

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:

- ◉ 1's
- ○ 10's
- ○ 100's
- ○ 1,000's

QUESTION 4 — 10 points — Saved

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:

- ◉ 1's
- ○ 10's
- ○ 100's
- ○ 1,000's

QUESTION 5 — 10 points — Saved

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.

- ◉ True
- ○ False

QUESTION 6 — 10 points — Saved

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?

582

QUESTION 7 — 10 points — Saved

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.

- ◉ True
- ○ False

QUESTION 8 — 10 points — Saved

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 _____.

0.80

QUESTION 9 — 10 points — Saved

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?

- (●) Yes
- ( ) No

QUESTION 10 — 10 points — Saved

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?

- ( ) Left skew
- ( ) Apparent uniform distribution
- (●) Right skew
- ( ) No apparent skew

QUESTION 11 — 10 points — Saved

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?

- ( ) Yes
- (●) No

QUESTION 12 — 10 points — Saved

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?

- (●) No apparent skew
- ( ) Right skew
- ( ) Left skew
- ( ) Apparent uniform distribution

QUESTION 13 — 10 points — Saved

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?

- ( ) Yes
- (●) No

QUESTION 14 — 10 points — Saved

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)

- ( ) Linearity
- ( ) Independence
- ( ) Normality
- (●) Homoscedasticity (or Homoskedasticity)

QUESTION 15 — 10 points — Saved

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?

- (●) Yes
- ( ) No

QUESTION 16 — 10 points — Saved

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?

- (●) Yes
- ( ) No

QUESTION 17 — 10 points — Saved

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)?

73.77

QUESTION 18 — 10 points — Saved

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.

- ( ) Yes
- (●) No

QUESTION 19 — 10 points — Saved

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?

6.871e+04

QUESTION 20 — 10 points — Saved

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.

- (●) Yes
- ( ) No

QUESTION 21 — 10 points — Saved

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 ?

1.206e-02

QUESTION 22 — 10 points — Saved

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?

48.25

QUESTION 23 — 10 points — Saved

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?

116963

QUESTION 24 — 10 points — Saved

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?

0.83

QUESTION 25 — 10 points — Saved

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?

- (●) Quadratic fit
- ( ) Linear fit
- ( ) Neither the linear or quadratic fit appear to be a "better" fit than the other
- ( ) Both the linear and the quadratic fits appear to be equally good

QUESTION 26 — 10 points — Saved

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?

- (●) Yes
- ( ) No

QUESTION 27 — 10 points — Saved

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?

- (●) Yes
- ( ) No

QUESTION 28 — 10 points — Saved

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.

- (●) Yes
- ( ) No

QUESTION 29 — 10 points — Saved

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.

- (●) Yes
- ( ) No

QUESTION 30 — 10 points — Saved

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..

- ( ) Yes
- (●) No

QUESTION 31 — 10 points — Saved

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.

- (●) Log-linear
- ( ) Linear
- ( ) Quadratic
- ( ) All models appear to be equally good fits to the data

QUESTION 32 — **10 points** — Saved

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.

- ( ) All result in the same SSE
- ( ) Linear
- (●) Quadratic
- ( ) Log-linear

QUESTION 33 — **10 points** — Saved

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

- ( ) True
- (●) False

QUESTION 34 — **10 points** — Saved

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!)

- (●) None of the models
- ( ) All of the models
- ( ) the simple linear model
- ( ) the log-linear model
- ( ) the quadratic model

QUESTION 35 — **10 points** — Saved

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

- ( ) the log-linear model
- ( ) the quadratic model
- ( ) All of the models
- (●) None of the models
- ( ) the simple linear model

QUESTION 36 — **10 points** — Saved

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?

- ( ) Yes
- (●) No

QUESTION 37 — **10 points** — Saved

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 _____.

1.7978e+05

QUESTION 38 — **10 points** — Saved

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 _____.

1.3103e+05

QUESTION 39 — **10 points** — Saved

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?

- (●) Yes
- ( ) No

QUESTION 40 — **10 points** — Saved

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 _____.

1.1490e+05

QUESTION 41 — **10 points** — Saved

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 _____.

1.2337e+05

QUESTION 42 — **10 points** — Saved

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 _____.

1.562955e-130

QUESTION 43 — **10 points** — Saved

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.

- (●) True
- ( ) False

QUESTION 44 — 10 points — Saved

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.

- ( ) True
- (●) False

QUESTION 45 — 10 points — Saved

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?

- (●) Yes
- ( ) No

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers | Save and Submit