Take Test: Problem Set #2

Test Information

Description

This is both the documentation and the online version of Problem Set #2. For additional information about using gretl to do logistic regression you can read through the websites https://analytics4all.org/2016/04/06/logistic-regression-with-gretl/ or https://medium.com/swlh/a-brief-introduction-to-econometrics-with-gretl-792c1e102e97.

The data set you need to complete Problem Set #2 is in the ContractData.gdt file.
ContractData.gdt

This assignment is shown in your textbook, A Second Course in Statistics: Regression Analysis as Example 9.5 in Section 9.6 Logistic Regression. Although the textbook solves this problem using the software package SAS, it is easy to do using gretl.

Notice that the results of the model adequacy tests, i.e. the X^2 or "likelihood Ratio" is shown in the details of the model gretl outputs as the last row, the Likelihood ratio test: Chi-square(2) =18.5377, which matches the results from SAS given in the example in the textbook. Note that there appears to be a typo in the textbook in discussing the coefficient estimates. The SAS output in the textbook matches the gretl output, but the textook has a sign error for $\widehat{\beta}_2$.

A short script you can use, if you want to, to complete the first part of this problem set is PS2aScript.inp.
PS2aScript.inp

Keep in mind that the two models output by this script are actually the same. I set the script up as though there were two models because each has slightly different output. Model1 outputs typical logit model results including a confusion table. Model2 outputs P-values and statistical significance as well as the covariance matrix. However, they are actually the same data and the same model, just different output.

The problem set you need to complete the second part of Problem Set #2 is in the cokeVsPepsi.gdt file.
cokeVsPepsi.gdt

A short script you can use, if you want to, to complete the second part of this problem set is PS2bScript.inp.
PS2bScript.inp

Instructions

You should answer all questions first, either using paper and pencil or another computer program such as gretl. Then, enter your answers in the online assignment. I have setup this assignment so you will have three chances to take it, one to get the questions and one to really enter your answers. Be sure to keep track of your work and answers. If for any reason I have to reset your assignment it will wipe out all the work you did before!

There are a variety of types of questions. You should select the best choice or choices. If you are entering a numeric value you should round your answer to two decimal places unless there are other specific instructions for a specific question. Not all questions are worth the same number of points, i.e. some questions are worth more points than others. If you have any questions – ask! There is additional information in the short description below.

Multiple Attempts: This test allows 2 attempts. This is attempt number 1.

Force Completion: This test can be saved and resumed later.

Your answers are saved automatically.

QUESTION 1

5 points Saved

How many observations are there in the contract/road construction dataset?

31

QUESTION 2

5 points Saved

How many independent variables are in the contract data/road construction dataset?

3

QUESTION 3

5 points Saved

Fill in the exact variable name of the dependent variable in the contract data/road construction dataset. Be careful, this field is case sensitive!

BIDStatusy

QUESTION 4

5 points Saved

The possible values for the dependent variable, BIDStatusy are 0 or 1.

- (●) True
- ( ) False

QUESTION 5

5 points Saved

If BIDStatusy is coded with the number 1 that indicates the bid is a fixed bid. If BIDStatusy is coded with the number 0 that indicates the bid is a competitive bid.

- (●) True
- ( ) False

QUESTION 6

5 points Saved

The reported mean value of BIDStatusy, 0.38710, indicates that (select the best choice to complete this sentence).

- ( ) there are more competitive bids than fixed bids.
- ( ) there are no bids available in this dataset.
- (●) there are more fixed bids than competitive bids.
- ( ) nothing. You cannot tell anything about the bids from the mean of the variable BIDStatusy.

QUESTION 7

5 points Saved

The output of the logit model built for the contract data/road construction dataset indicates that the model is able to accurately predict _____ % of all cases (or observations). Note that the answer is in a percentage from the model output.

83.9

QUESTION 8

5 points Saved

The logit model for the contract data/road construction dataset results in _____ Type II errors.

3

QUESTION 9

5 points Saved

A Type Ii error for the contract data/road construction logit model means that (select the best choice below to complete this sentence).

- ( ) 9 "cases" were incorrectly coded in the first place.
- ( ) you cannot tell what the error was for.
- ( ) 17 "cases" were incorrectly coded in the first place.
- (●) 3 "cases" incorrectly indicated that they were for competitive bids when in reality they were for fixed bids.

QUESTION 10 — 5 points — Saved

For the contract data/road construction logit model as reported in the textbook Example 9.5, the (Hosmer and Lemeshow Goodness-of-Fit) P-value equal to 0.2324 which is greater than the 0.05 level of significance indicates that _____ (select the choice below that best completes this sentence). Note that I have redirected you to the textbook output for this question because we do not typically compute the Hosmer and Lemeshow Goodness-of-Fit test.

- ○ the null hypothesis must be rejected. The model lacks fit.
- ○ the test conducted is not a good test for this model.
- ◉ there is insufficient evidence that the logistic regression model lacks fit. That is, consistent with the practice of failure to reject the null hypothesis when the P-value is greater than the level of significance we find the model (null hypothesis) has sufficient supporting evidence.
- ○ there is insufficient evidence for basically anything. More data is required.

QUESTION 11 — 5 points — Saved

The number of observations in the cokeVsPepsi dataset is _____.

1140

QUESTION 12 — 5 points — Saved

The number of variables in the cokeVsPepsi dataset is _____.

5

QUESTION 13 — 5 points — Saved

For the Coke-Pepsi dataset, the three models produce virtually identical results.

- ○ True
- ◉ False

QUESTION 14 — 5 points — Saved

Although the probit and logit models built for the coke-pepsi dataset are different, both models result in the same error rates. (Hint, look for the confusion tables in the model output!)

- ◉ True
- ○ False

QUESTION 15 — 5 points — Saved

The dependent variable, coke, in the cokeVsPepsi dataset means that _____ (select the choice below that best completes this sentence).

- ○ coke is more often chosen over pepsi.
- ○ people generally prefer pepsi over coke.
- ○ people generally prefer coke over pepsi.
- ◉ coke is the chosen product when the variable = 1, pepsi when the variable = 0.

QUESTION 16 — 5 points — Saved

The number of correctly predicted "cases" by either the probit or logit models is _____. (Enter the number of correct cases this time, NOT the percent.)

765

QUESTION 17 — 5 points — Saved

The number of "cases" or observations that were predicted to have coke chosen when pepsi should have been chosen is _____. (Hint, look at the confusion tables in the model output!)

122

QUESTION 18 — 5 points — Saved

The expectation is that if a store has a coke display then coke is more likely chosen. If a store has a pepsi display then pepsi is more likely chosen. All the models; linear, probit, and logit, show that the coefficients for the variables disp_coke and disp_pepsi have opposite signs. This indicates that _____. (Select the choice below that best completes this sentence.)

- ◉ the effect of coke and pepsi displays have opposite effects on the probability of choosing coke. For example, when a pepsi display is present then coke is less likely to be chosen.
- ○ the effect of displays for either coke or pepsi are negligible.
- ○ coke is more often chosen than pepsi.
- ○ the variables for display do not make sense.

QUESTION 19 — 5 points — Saved

Both probit and logit models can be used to predict "choice". This is due to their ability to predict "discrete" or binary outcomes. (Note, that this does NOT include multinomial models which involve more than 2 choices, e.g. either/or problems.)

- ◉ True
- ○ False

QUESTION 20 — 5 points — Saved

To keep choice probability, "p," within the interval [0, 1], both the probit and logit models are based on a nonlinear "S-curve" or S-shaped relationship between "x" and "p". However, where the probit function is related to the standard normal probability distribution, the logit function is related to the sigmoid function.

- ◉ True
- ○ False

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

Save All Answers | Save and Submit