

What is the annotated output for a Tobit regression in Stata?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the annotated output for a Tobit regression in Stata?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160607>

The annotated output for a Tobit regression in Stata is a comprehensive summary of the results obtained from running a Tobit regression analysis on a dataset. It includes detailed information such as the model specifications, coefficient estimates, significance levels, goodness-of-fit measures, and diagnostic tests. Additionally, it provides graphical representations of the data and the regression model, as well as a discussion of the interpretation and implications of the results. This annotated output serves as a valuable tool for understanding and communicating the findings of a Tobit regression analysis in Stata.

Tobit Regression | Stata Annotated Output

This page shows an example of tobit regression analysis with footnotes

explaining the output. The data in this example were gathered on undergraduates applying to graduate school and includes undergraduate GPAs, the reputation of the school of the undergraduate (a topnotch indicator), the students' GRE score, and whether or not the student was admitted to graduate school.

The range of possible GRE scores is 200 to 800. This means that our outcome variable is both left censored and right-censored. In other words, if two students score an 800, they are equal according to our scale but might not truly be equal in aptitude.

(In other words, we have a ceiling effect.) The same is true of two students scoring 200 (a floor effect). Tobit regression generates a model that predicts the outcome variable to be within the specified range.

If we are interested in predicting a student's GRE score using their undergraduate GPA and the reputation of their undergraduate institution, we should first consider GRE as an outcome variable.

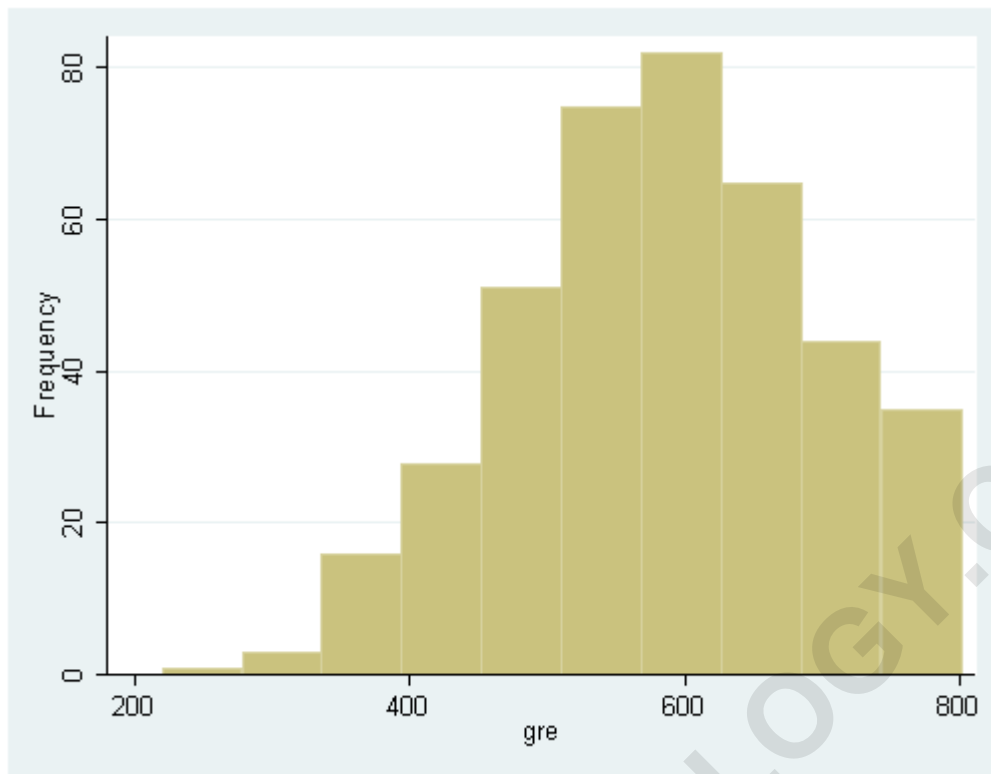
use <https://stats.idre.ucla.edu/stat/stata/dae/logit.dta>,
clear

summarize(gre)

Variable | Obs Mean Std. Dev. Min Max

-----+-----
gre | 400 587.7 115.5165 220 800

histogram gre, bin(10) freq



To generate a tobit model in Stata, list the outcome variable followed by the predictors and then specify the lower limit and/or upper limit of the outcome variable. The lower limit is specified in parentheses after ll and the upper limit is specified in parentheses after ul. A tobit model can be used to predict an outcome that is censored from above, from below, or both.

tobit gre gpa topnotch, ll(200) ul(800)

Refining starting values:

Grid node 0: log likelihood = -2332.8456

Fitting full model:

Iteration 0: log likelihood = -2332.8456

Iteration 1: log likelihood = -2331.4413

Iteration 2: log likelihood = -2331.4314

Iteration 3: log likelihood = -2331.4314

Tobit regression Number of obs = 400

Uncensored = 375

Limits: Lower = 200 Left-censored = 0

Upper = 800 Right-censored = 25

LR chi2(2) = 70.93

Prob > chi2 = 0.0000

Log likelihood = -2331.4314 Pseudo R2 = 0.0150

gre | Coefficient Std. err. t P>|t|

-----+-----
gpa | 111.3085 15.19669 7.32 0.000 81.43266 141.1843

```

topnotch | 46.65774 15.75359 2.96 0.003 15.68709
77.6284
_cons | 205.8515 51.24085 4.02 0.000 105.115 306.5881
-----+-----
var(e.gre)| 12429.62 923.9586 10739.66 14385.5
-----

```

Tobit Regression Output

Tobit regression Number of obs = 400

LR chi2(2) = 70.93

Prob > chi2 = 0.0000

Log likelihood = -2331.4314 Pseudo R2 = 0.0150

```

-----
gref| Coef. Std. Err. z P>|t| [95% Conf. Int.]
-----+-----

```

```

gpa | 111.3085 15.19665 7.32 0.000 81.43273 141.1842

```

```

topnotch | 46.65774 15.75356 2.96 0.003 15.68716
77.62833

```

```

_cons | 205.8515 51.24073 4.02 0.000 105.1152 306.5879
-----+-----

```

```

var(e.gre)| 12429.62 923.9586 10739.66 14385.5
-----

```

a. Log likelihood - This is the log likelihood of the fitted model. It

is used in the Likelihood Ratio Chi-Square test of whether all predictors' regression coefficients in the model are simultaneously zero.

b. Number of obs - This is the number of observations in the dataset

for which all of the response and predictor variables are non-missing.

c. LR chi2(2) - This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression

coefficient is not equal to zero. The number in the parentheses indicates the

degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square

statistic and is defined by the number of predictors in the model (2).

d. Prob > chi2 - This is the probability of getting a LR test

statistic as extreme as, or more so, than the observed

statistic under the null hypothesis; the null hypothesis is that all of the regression coefficients are simultaneously equal to zero. In other words, this is the probability of obtaining this chi-square statistic (70.93) or one more extreme if there is in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to accept a type I error, which is typically set at 0.05 or 0.01. The small p-value from the LR test, <0.0001 , would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero. The parameter of the chi-square distribution used to test the null hypothesis is defined by the degrees of freedom in the prior line, $\text{chi2}(2)$

e. Pseudo R2 - This is McFadden's pseudo R-squared. Tobit regression does not have an equivalent to the R-

squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics. Because this statistic does not mean what R-square means in OLS regression (the proportion of variance of the response variable explained by the predictors), we suggest interpreting this statistic with great caution. For more information on pseudo R-squareds, see [What are Pseudo R-Squareds?](#).

f. gre - This is the response variable predicted by the model.

We are using a tobit model because this response variable is censored: the GRE scores are scaled from 200 to 800 and cannot fall outside of this range.

g. Coef. - These are the regression coefficients. Tobit regression coefficients are interpreted in the similar manner to OLS regression coefficients; however, the linear effect

is on the uncensored latent variable, not the observed outcome. The expected GRE score changes by Coef. for each unit increase in the corresponding predictor.

gpa - If a subject were to increase his gpa by one point, his expected GRE score would increase by 111.3085 points while holding all other variables in the model constant. Thus, the higher a student's gpa, the higher the predicted GRE score.

topnotch - If a subject attended a topnotch institution for her undergraduate education, her expected GRE score would be 46.65774 points higher than a subject with the same grade point average who attended a non-topnotch institution. Thus, subjects from topnotch undergraduate institutions have higher predicted GRE scores than subjects from non-topnotch undergraduate institutions if grade point

averages are held constant.

_cons - If all of the predictor variables in the model are evaluated at zero, the predicted GRE score would be **_cons = 205.8515**. For subjects from non-topnotch undergraduate institutions (topnotch evaluated at zero) with zero gpa, the predicted GRE score would be 205.8515. This may seem very low, considering the mean GRE score is 587.7, but note that evaluating gpa at zero is out of the range of plausible values for gpa.

h. Std. Err. - These are the standard errors of the individual regression coefficients. They are used in both the calculation of the t test statistic, superscript **i**, and the confidence interval of the regression coefficient, superscript **k**.

i. t - The test statistic **t** is the ratio of the Coef. to the Std. Err. of the respective predictor. The

t value is used to test against a two-sided alternative hypothesis that the Coef. is not equal to zero.

j. $P > |t|$ - This is the probability the t test statistic (or a more extreme test statistic) would be observed under the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors are in the model. For a given alpha level, $P > |t|$ determines whether or not the null hypothesis can be rejected. If $P > |t|$ is less than alpha, then the null hypothesis can be rejected and the parameter estimate is considered statistically significant at that alpha level.

gpa - The t test statistic for the predictor gpa is $(111.3085/15.19665) = 7.32$ with an associated p-value of <0.001 . If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for

gpa has been found to be statistically different from zero given topnotch is in the model.

topnotch -The t test statistic for the predictor topnotch is $(46.65774/15.75356) = 2.96$ with an associated p-value of 0.003. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for topnotch has been found to be statistically different from zero given gpa is in the model.

_cons - The t test statistic for the intercept, _cons, is $(205.8515/51.24073) = 4.02$ with an associated p-value of < 0.001 . If we set our alpha level at 0.05, we would reject the null hypothesis and conclude that _cons has been found to be statistically different from zero given gpa and topnotch are in the model and evaluated at zero.

k. - This is the Confidence Interval (CI) for an

individual coefficient given that the other predictors are in the model. For a given predictor with a level of 95% confidence, we'd say that we are 95% confident that the "true" coefficient lies between the lower and upper limit of the interval. The CI is equivalent to the t test statistic: if the CI includes zero, we'd fail to reject the null hypothesis that a particular regression coefficient is zero given the other predictors are in the model with alpha level of zero. An advantage of a CI is that it is illustrative; it provides a range where the "true" parameter may lie.

l. var(e.gre) - This is the estimated variance of the regression.

In earlier versions of Stata, sigma was given in the output. Sigma is the square root of the variance that is given in the current output.