What is censored regression and how can it be analyzed using Mplus?

Authored by

stats writer

June 30, 2024

Censored regression is a statistical technique used to analyze data that contains censored or truncated values. In other words, some of the data points are known to fall within a certain range, but the exact values are unknown. This type of data is commonly found in fields such as economics, finance, and biostatistics.

Mplus is a statistical software program that can be used to analyze censored regression data. It offers various techniques such as maximum likelihood estimation and Bayesian analysis to handle censored values and produce accurate estimates. Mplus also allows for the inclusion of covariates and the examination of the relationship between the censored variable and other variables of interest. Additionally, Mplus provides graphical representations and summary statistics to aid in the interpretation of the results. Overall, Mplus is a powerful tool for analyzing censored regression data and can provide valuable insights in various research fields.

Censored Regression | Mplus Annotated Output

**This page shows an example of censored regression with footnotes
explaining the output. First an example is shown using Stata, and then an
example is shown using Mplus, to help you relate the output you are likely to be
familiar with (Stata) to output that may be new to you (Mplus). We suggest that
you view this page using two web browsers so you can show the page side by side
showing the Stata output in one browser and the corresponding Mplus output in
the other browser.**

This example is drawn from the Mplus User's Guide (example 3.2) and we suggest that you see the Mplus User's Guide for more details about this example. We thank the kind people at Muthén & Muthén for permission to use examples from their manual.

Example Using Stata

Here is a probit regression example using Stata with two continuous predictors x1 and x2 used to predict a binary outcome variable, u1.

```
infile u1 x1 x3 using https://stats.idre.ucla.edu/wp-content/uploads/2016/02/ex3.2.dat, clear

summarize u1

Variable | Obs Mean Std. Dev. Min Max
-------------+--------------------------------------------------------
u1 | 1000 .9240341 1.113079 0A 6.579389

tobit u1 x1 x3, ll(0)
```

Tobit regression Number of obs = 1000

LR chi2(2) = 697.44

Prob > chi2 = 0.0000

Log likelihood = -1142.8851 Pseudo R2 = 0.2338

--

u1 | Coef. Std. Err. t P>|t|

-------------+--

x1 | 1.074801D .0419657 25.61 0.000 .9924498 1.157152

x3 | .4947541D .0378985 13.05 0.000 .4203842 .569124

_cons | .5154865E .0405066 12.73 0.000 .4359986 .5949743

-------------+--

/sigma | 1.071333F .0316242 1.009276 1.133391

--

Obs. summary: 376 left-censored observations at u1<=0

624 uncensored observations

0 right-censored observations

estat ic

Model | Obs ll(null) ll(model) df AIC BIC

-------------+---

. | 1000 -1491.605 -1142.885B 4 2293.77C 2313.401C

The output is labeled with superscripts to help you relate the later Mplus output to this Stata output. To summarize the output, both predictors in this model, x1 and x2, are significantly related to the outcome variable, u1.

Mplus Example

Here is the same example illustrated in Mplus based on the https://stats.idre.ucla.edu/wp-content/uploads/2016/02/ex3.2.dat data file. Note that by using estimator=wls; (weighted least squares) the results are shown in a probit metric.
Had we specified something like estimator=ml; (maximum likelihood) then the results would be shown in a logit scale.

TITLE:
this is an example of a censored regression for a censored dependent variable with two covariates

DATA:
FILE IS https://stats.idre.ucla.edu/wp-content/uploads/2016/02/ex3.2.dat;
VARIABLE:
NAMES ARE y1 x1 x3;
CENSORED ARE y1 (b);
ANALYSIS:
ESTIMATOR = MLR;
MODEL:
y1 ON x1 x3;

SUMMARY OF ANALYSIS

<some output omitted to save space>
Number of observations 1000

<some output omitted to save space>
SUMMARY OF CENSORED LIMITS
Y1 0.000A

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Loglikelihood

H0 Value -1142.885B

Information Criteria

Number of Free Parameters 4
Akaike (AIC) 2293.770C
Bayesian (BIC) 2313.401C
Sample-Size Adjusted BIC 2300.697
(n* = (n + 2) / 24)

MODEL RESULTS
Estimates S.E. Est./S.E.
Y1 ON
X1 1.075D 0.043 25.101
X3 0.495D 0.037 13.344

Intercepts
Y1 0.515E 0.040 12.810

Residual Variances
Y1 1.148F 0.067 17.235