

Modeling Count Data

Cambridge University Press: 2014

Joseph M Hilbe

Arizona State Univ. & Jet Propulsion Laboratory, Caltech

USEFUL INFORMATION AND ACKNOWLEDGEMENTS

Ten data sets are used in *Modeling Count Data*. The R data sets and their respective description files are to be found in the CRAN package, COUNT. To load one of the data sets into memory, after installing COUNT

```
> library(COUNT)
> data(medpar)      # for example
```

Stata, R, SAS, CVS, and Excel data files are located on my personal website and on the Cambridge University Press site for the book:

http://works.bepress.com/joseph_hilbe/

www.cambridge.org/9781107611252

An *Errata and Comments* PDF document is located at both my website and Cambridge, although I will be able to keep my site up to date. The R COUNT package shares data files, functions and scripts with *Modeling Count Data* (2014) and *Negative Binomial Regression*, 2nd edition (2011). Both are published by Cambridge University Press. Stata ado and do software programs used and associated with the book are found on my personal website, and will be kept up to date. R files not provided in the COUNT package may also be found on my site, and will be enhanced on a timely basis. Likewise SAS files.

Data files for *Modeling Count Data* are displayed in the table below. Files in common with those in *Negative Binomial Regression*, 2nd edition, are also indicated.

FILENAME	SECTION 1 st	DESCRIBED	PAGE	ALSO IN NBR2
smoking	1.2.1		5	
rwm1984	2.4		48	
rwm5yr	2.4		52	NBR2
fasttrakg	2.6		64	NBR2
medpar	3.4.1		93	NBR2
titanic	3.4.3		103	NBR2
nuts	5.4.3		156	
azprocedure	8.0		211	NBR2 (named <i>azpro</i>)
azcabgptca	9.1		219	NBR2 (named <i>azdrg112</i>)
fishing	9.3		233	
titanicgrp		grouped format of titanic		

A 16 page PDF and 959 line text document containing all of the R code in the book will be available on the CRAN COUNT package, and will also be on both my personal website and Cambridge University Press site for the book. Stata and SAS program code will be available in text form so that it can be copied and pasted into the software's editor for execution or amendment. You are encouraged to adapt the code for your own needs.

Prof. **James Hardin** (Univ. of South Carolina, Dept of Biostatistics) and myself have co-authored a number of count model functions or commands in Stata over the years. A number of these are associated with our book, *Generalized Linear Models and Extensions*, which is now in its 3rd edition (2012). In addition, more have been published during the past 12 months, or are soon to be published. These include commands -- which are like R functions and SAS procedures -- for the count models listed below (as of June 2014). We also developed a number of binomial models, which were also used in our GLM book, and recently newer ones that will be used in the second edition of my book, *Logistic Regression Models* (Chapman & Hall/CRC), that is due out in 2015.

ZI = zero inflated version ZT = zero-truncated version

=====

Greene's NB-P (NBP)	+ ZINBP	+ ZTNBP
Poisson inverse Gaussian (PIG)	+ ZIPIG	+ ZTPIG
Generalized Poisson (GP)	+ ZIGP	
3-parameter Waring NB (NBW)	+ ZINBW	(with Tammy Harris)
3-parameter Famoye NB (NBF)	+ ZINBF	(with Tammy Harris)
Zero-inflated heterogeneous NB (called generalized NB in Stata)		

Left, right, and interval truncated count model (treg) for

Poisson	Negative binomial
Generalized Poisson	Poisson Inverse Gaussian
Greene's NBP	3-parameter NBF

A left, right and interval censored count model with the same options is currently being developed. I published censored Poisson and negative binomial models some 20 years ago, and use the *cpoisson* command in this book.

I had previously published a number of hurdle models in the past, but with a combination of logit, probit, cloglog, loglog or any other binary formatted model, a hurdle model can be constructed using the zero-truncated or truncated count models listed above; e.g., for instance, a PIG-logit hurdle.

Count models software authored by others using Stata are included in this book as well. Aside from the above, and those that are currently part of the official Stata package, I should mention:

fmm – a finite mixture model authored by **Partha Deb** of Hunter College and the City University of New York.

qcount – a quantile count model authored by **Alfonso Miranda** of Keele University, UK.

ztpnm - a zero-truncated Poisson-lognormal mixture model by **Helmut Farbmacher** or the Univ. of Munich.

I wish to thank each of these authors for the work they did in writing these programs, and for being able to incorporate the software for examples used in *Modeling Count Data*. Deb and Miranda's programs were also used and acknowledged in *Negative Binomial Regression*, 2nd ed.

I wish to thank James Hardin for his work and contributions to developing Stata software for modeling count data, and to Andrew Robinson (Univ of Melbourne) for his work on the structure of the R *nbinomial* function, which we wrote for use in our book:

Hilbe & Robinson (2013), *Methods of Statistical Model Estimation*, Chapman & Hall/CRC.

I recommend that *nbinomial* be used in place of R's default *glm.nb* function when desiring to model negative binomial data, which I discuss in the book. It has many more options and output.

The above book has its own CRAN package, *msme*, that includes a number of R functions, including *nbinomial*. It is now automatically loaded when the COUNT package is loaded. I should advise you that when data are not appropriate for a particular model, the software algorithm used to model the data may not converge. For instance, if data are not negative binomial distributed, R's *nbinomial* and *glm.nb* may not converge, or if they do, may result in incorrect coefficients, or more likely incorrect standard errors. For R users, I recommend *nbinomial* in preference to *glm.nb* for modeling negative binomial data.

I encourage you to advise me of any typos or errors that you spot in the book. No matter how many editors and how many statistical reviewers, including myself, go over the manuscript and proofs, errors creep in nonetheless. These will be maintained in the *Errata and Comments* document I will post. It will be updated immediately after learning of any error, and posted to my website. I will also acknowledge those who do provide such information. Suggestions and comments are also welcome. Email me at

hilbe@asu.edu or j.m.hilbe@gmail.com

Joseph M Hilbe
Florence, AZ (USA)
3 July, 2014