

R and Stata Parallels

For Categorical Data Analysis

Andy Grogan-Kaylor

2023-12-03

Table of contents

This is an evolving and growing document. Comments, questions, corrections and clarifications are all welcome.

Concept	Stata	R
Get Data	<code>use "....dta"</code>	<code>load("....RData")</code>
Descriptives	<code>summarize</code>	<code>summary()</code>
Cross-Tabulation	<code>tabulate x y</code>	<code>table(x, y)</code> ¹
Cross-Tabulation With Row and Column Percentages	<code>tabulate x y, row col</code>	<code>prop.table(table(x, y), margins = ...)</code> <code>gmodels::CrossTable(x, y)</code> ²
ChiSquare Test	<code>tabulate x y, row col</code> <code>chisq</code>	<code>chisq.test(table(x, y))</code> ³
Logistic Regression	<code>logit y x</code>	<code>glm(y ~ x, data = ..., family = "binomial")</code>

¹For the sake of parsimony, in my R `table`, `prop.table`, and `gmodels::CrossTable` syntax, I am using single columns of data, e.g. `x` and `y`, but R could as easily use the `dataset$variable` syntax e.g. `table(dataset$variable)`

²`gmodels::CrossTable` offers nicer formatting that may be easier to read, especially for those accustomed to Stata or SAS.

³Stata shows the Cross-Tabulation Table together with the χ^2 test while R only shows the results of the χ^2 test.

Concept	Stata	R
Probit Regression	<code>probit y x</code>	<code>glm(y ~ x, data = ..., family = binomial(link = "probit"))</code>
Ordered Logistic Regression	<code>ologit y x</code>	<code>polr(y ~ x, data = ..., Hess = TRUE)⁴</code>
Multinomial Logistic Regression	<code>mlogit y x</code>	<code>multinom(y ~ x, data = ...)⁵</code>
Poisson Regression	<code>poisson y x</code>	<code>glm(y ~ x, family="poisson", data= ...)</code>
Negative Binomial Regression	<code>nbreg y x</code>	<code>glm.nb(y ~ x, data = ...)⁶</code>

Created by Andy Grogan-Kaylor: <https://agrogan1.github.io/>; agrogan@umich.edu

⁴Requires `library(MASS)`

⁵Requires `library(nnet)`

⁶Requires `library(MASS)`