R and Stata Parallels

For Categorical Data Analysis

Andy Grogan-Kaylor

2023-09-05

Table of contents

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

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

[1] For 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.

[3] Stata 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	`probit y x`	`glm(y ~ x, data = ..., family = binomial(link = "probit"))`
Ordered Logistic Regression	`ologit y x`	`polr(y ~ x, data = ..., Hess = TRUE)`[4]
Multinomial Logistic Regression	`mlogit y x`	`multinom(y ~ x, data = ...)`[5]
Poisson Regression	`poisson y x`	`glm(y ~ x, family="poisson", data= ...)`
Negative Binomial Regression	`nbreg y x`	`glm.nb(y ~ x, data = ...)`[6]

Created by Andy Grogan-Kaylor: http://www.umich.edu/~agrogan; agrogan@umich.edu

[4] Requires `library(MASS)`
[5] Requires `library(nnet)`
[6] Requires `library(MASS)`