

STATA Cheat Sheet

Chris Hoofnagle

August 2014

1 General

Stata Syntax	command variables, options
Operators	&=and =or others: == != ~= > >= < <=
Syntax for if excluding missing values	summarize variable if sex==1 & age > 30 & age <
xtile, nquantiles()	Generates new variable into equal number of cases
Clonevar	Copies a variable, its label, etc
drop	Deletes a variable

2 Descriptive statistics

sum vars	Summarize Westin_Code, detail
list vars	List Westin_Code testscore in 1/20
codebook vars	
Tabstat vars	Tabstat var, statistics(mean median sd) by(var) columns(statistics)
tab1 vars	Basic one way tab
table vars	Table recage, contents(mean educ count recage)

3 Helpful commands

ssc hot, n(25)	Top 25 most downloaded extensions
fre	Enhancement for tab, ssc install fre
numlabel _all, add	Appends labels to all values
bigtab	Ssc install for “too many values” error in tabulate

4 Comparing variables—nominal, ordinal, and interval

tab dep_var indep_var	Cross tab for two categorical (nominal or ordinal) values, puts dep_var in rows
tab indep_var,sum(dep_var)	Mean comparison for categorical indep_var and interval dep_var
bysort control_var: tabulate dep_var indep_var	Creates many cross tabs, one for each of the control variables
tabulate, generate()	Creates indicator variables from existing; great for dummy variables in mean comparisons
tab control_var indep_var, summarize(dep_var)	Mean comparison for each of the control variables
tab sex privacy_fund chi2	Tab option for chi2
tab age educ taub	Proportional reduction in error measure of the strength of a relationship, both variables ordinal, table must be “square.” PREs answer the “how strong” question
somersd indep_var dep_var	Taub alternative PRE for non-square tables
lambda ideo race	PRE for variables where one or both is nominal. Mode must be different between categories
tab gender Westin chi2 V	Alternative to lambda, not a PRE
chi2 gamma taub V expected	Options for comparisons
tabi	Lets you input values and check chi2, gamma, V
How to report chi2	$2(1, N=1,000) = 3.034; p < .001$

ttest var=test_value	ttest age=0 One sample t test; finds boundaries of mean at 95% confidence level
robvar var, by (group)	Compares variances (use before 2-sample ttest) Must confirm that $p < .05$, otherwise add unequal option to ttest
ttest var, by(group) level(90)	Determines whether differences in means are statistically significant at 90% level, e.g. ttest quiz if WestinCode<3, by(WestinCode) level(90)
corr vars	For interval-level relationships
xi: regress dep i.indep	Estimates effect of categorical variables on independent
Regress dep indep_vars	Dependent is interval; indep is any type. Estimates effect of indep var on dependent. R explains independent's contribution to dep value
Logit	Analyze interval independent variable and binary dependent