Data with STATA – Summary Sheet

STATA Preparatory Course

Summarize Data

summarize v012 v445, detail
Calculate a variety of univariate summary statistics

ci means v012 v445, level(99)
Compute standard errors and confidence intervals

correlate v012 v445
correlation matrix

pwcorr v012 v438 v445, star(0.05)
pairwise correlation coefficients with sig. levels

mean v012 v445
Estimates of means, including standard errors

proportion v149 v102
Estimates of proportions, including standard errors for categories identified in varlist

Include missing values

Create one indicator variable for each v149 education category using the prefix "educ_"

tabulate v149, mi gen (educ)
One-way table: number of women in each v149 education category

tabulate v149 v102, mi
Two-way table: cross-tabulate number of observations for each combination of education and residence

bysort v149: tabulate v102
For each v149 education category, apply the command tabulate v102

tabstat v012 v438 v445, by(v102) stat(mean sd n)
Create compact table of summary statistics

Replaces data

collapse (mean) v445 (max) v012, by(v102)
Calculate mean BMI and maximum age by residence type

Estimate Models

regress v445 v012 v438, vce(robust)
Estimate an OLS model of BMI using age and height, with robust standard errors

regress v445 v012 i.v149 if v102 == 1, vce(cluster v001)
Estimate the model for urban women only, clustering standard errors by PSU

rreg v445 v012 v438, genwt(reg_wt)
Estimate robust regression that limits the influence of outliers

probit working v012 i.v149, vce(robust)
Estimate a probit model for the generated binary variable working

logit anemia i.v190 v012, or
Estimate a logistic model for anemia and report odds ratios

Continuous Variables: c **Categorical Variables: i** **Interactions: #**

Operator	Description	Command	What it does
i.	Specify indicators	**regress v445 i.v149**	Treat v149 as a categorical variable
ib(#).	Specify base indicator	**regress v445 ib(3).v149**	Use category 3 of v149 as the base category
fvset	Command to change base	**fvset base frequent v149**	Use the most frequent v149 category as the base
!	Omit a category	**regress v445 i.v149 if v149 != 2**	Treat v149 as categorical and omit v149 == 2
#	Specify interactions	**regress v445 v012 c.v012#c.v012**	Include a squared age term
##	Specify interactions	**regress v445 c.v012##c.v012**	Include the age main effect and squared term

Nisha Malhotra, PhD
Economist & Policy|Analyst
Research & Education Consultant
Website: www.nishamalhotra.com
linkedin.com/in/nishamalhotra
youtube.com/@NishaMalhotra

© 2026 Nisha Malhotra | CC BY 4.0 International | Reproduction and adaptation permitted with attribution.