|  |  |  |
| --- | --- | --- |
| **Function** | **Stata** | **Python** |
| **Read Data** |  |  |
| Read csv |  |  |
| Read stata file | Use “dir/filename.dta”, clear | df = pd.read\_stata(‘dir/filename.dta’) |
| **Generate variables** |  |  |
| Generate y = exp(x) | gen y = exp(x) | \*\* import math as math  df[y] = np.exp(df.x) |
| Generate random variable with normal distribution | gen x=rnormal(mean,SD) | \*\* import numpy as np  x = random.normal(loc=mean, scale=SD, size=# of obs) |
| **Drop** |  |  |
| Drop observations if x > # | Drop if x > # | df.drop(df[df['x'] > #].index, inplace = True) |
| **Collapse data** |  |  |
| Generate frequency table for variable x | collapse (count) x | collapsed\_data = df.groupby('x')['x'].count() |
| **OLS** |  |  |
| Simple regression of y on x | reg y x | \*\* import statsmodels.formula.api as smf  mod = smf.ols(formula = "y ~ x", data=df).fit()  print(mod.summary()) |
| Regression of y on a set of dummy variables for a | reg y i.a | \*\* import statsmodels.formula.api as smf  mod = smf.ols(formula = "y ~ C(a)", data=df).fit()  print(mod.summary()) |
| Fixed effects regression | xtset panel\_var  xtreg y x, fe | \*\* import statsmodels.formula.api as smf  mod = smf.ols(formula = "y ~ x+ C(a)", data=df).fit()  print(mod.summary()) |
| **2sls** |  |  |
| 2sls |  |  |
| 2sls with FE |  | \*\* from statsmodels.sandbox.regression.gmm import IV2SLS  \*\* from patsy import dmatrices, dmatrix  y, x = dmatrices('y~ x + C(a)', df)  z = dmatrix('x + C(a)', df)  print(IV2SLS(y, x, instrument = z).fit().summary()) |
| **Graphing** |  |  |
| Create a scatterplot of x and y | Scatter y x, xlab(x ) ylab(y) | \*\* import matplotlib.pyplot as plt  plt.scatter(df.x, df.y)  plt.xlabel("x-label")  plt.ylabel("y-label")  plt.show() |
|  |  |  |
|  |  |  |