Mgmt 469

Transitioning from KStat to Stata

As many of you may recall from DECS 434, KStat is a set of macros added to Microsoft Excel that allows you to perform various statistical analyses. While KStat is easy to use, it is not powerful enough to perform sophisticated analyses or deal with large datasets, both of which are common features of real-world research. Moreover, you cannot readily program in Excel, which makes it difficult for other to reproduce your work. It is for these reasons that we require you to learn Stata for the purposes of this course. Since many of you are well versed with KStat, this document should help in making your transition to using Stata easier.

Since KStat is entirely menu-driven, the following table presents the equivalent Stata command for each option in the KStat "statistics" menu. Stata allows you to use an abbreviated version of the main command – these are presented inside the square brackets [ ] in the following table. Note that all commands in Stata are typed in lower case. In the examples below, if there is no variable(s) following a command, Stata performs that analysis on all the variables in the data.

KStat option	Description	Stata command	Example
Univariate Statistics	Calculates and displays a variety of univariate summary statistics	- summarize [su] - summarize, detail [su, de]	- su - su price1 price2 - su price1 if price1 > 0 - su price1, de
Correlations	Finds the correlation coefficients between all pairs of variables in your data	- correlate [corr] - pwcorr - pwcorr, sig	- corr - corr (price1 price2) - pwcorr (price1 price2), sig
Regression	Perform a linear regression of a dependent variable on one or more independent variables	- regress Y X [reg]	- reg sales1 price1 price2 price3 - reg sales1 price1 price2 if store==5
Analysis of Variance	Fits Analysis-of-variance models	-anova Y X1 X2	- anova sales1 price1 price2 - anova sales1 store
Prediction	Make predictions using most recently performed regression	-predict varname (always follows a regression)	- predict predsales

Residuals	Compute residuals using predictions	-predict varname, resid (always follows a regression)	-predict residsales, resid
Model Analysis	Returns a variety of statistics from different tests, each of which is covered below		
Model Analysis	Breusch-Pagen test: A test for heteroskedasticity in the errors	- hettest* - estat hettest (need to follow a regression)	- reg sales1 price1 promo1 week; estat hettest;
Model Analysis	Durbin-Watson Statistic: a measure of the degree of autocorrelation in the errors	- dwstat* - estat dwatson (need to follow a time-series regression)	- tsset week; regress sales1 price1 promo1; estat dwatson;
Model Analysis	Variance Inflation Factor: a measure used to test for multicollinearity	- vif* - estat vif (need to follow a regression)	- reg sales2 price1 price2 price3; estat vif;
Model Analysis	Studentized Residual: the number of std deviations that residual is from zero	- predict varname, rstudent (needs to follow a regression)	- reg sales1 price1; predict studres, rstudent;
Model Analysis	Cook's D: a measure of the influence a data point has on the regression results	- predict varname, cooksd (needs to follow a regression)	- reg sales1 price1; predict cdist, cooksd;

* - These commands continue to work but are out-of-date as of Stata 9. The next line contains the current version of the command

The Graphics options presented by KStat can also be replicated using Stata (Stata has a lot more options when it comes to plotting graphs too). For plotting scatterplots, histograms, pie charts etc. use the "Easy graphs" option in the "Graphics" tab, as shown below in Figure 1. Similarly, if you are interested in looking at residual plots from your regression, use the "Regression diagnostic plots" option in the "Graphics" tab, as shown in Figure 2.
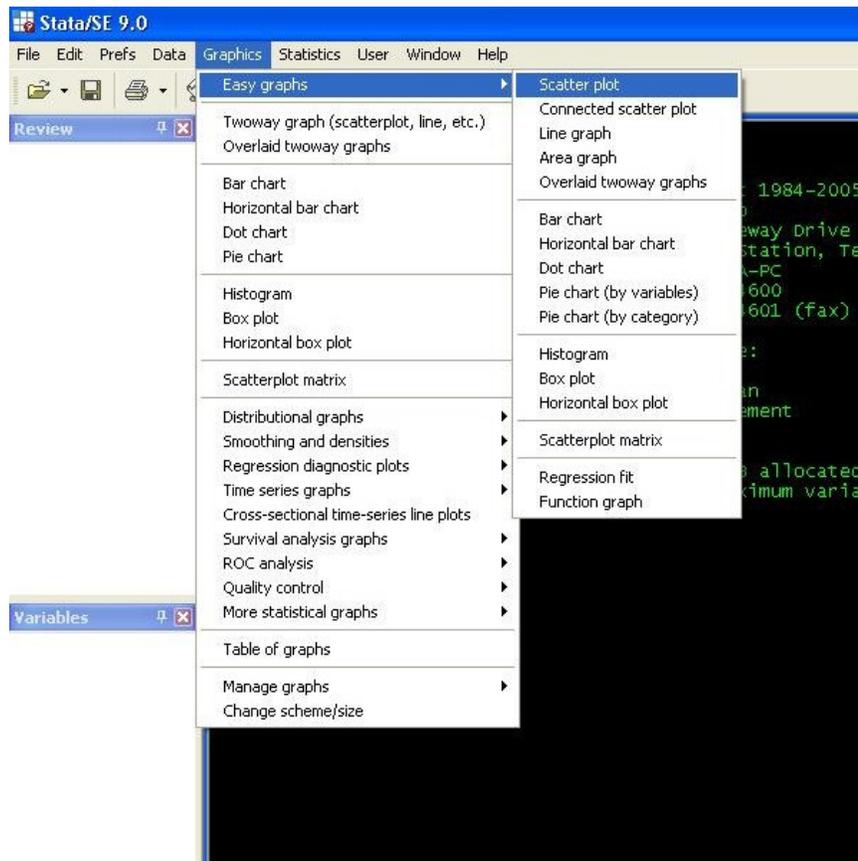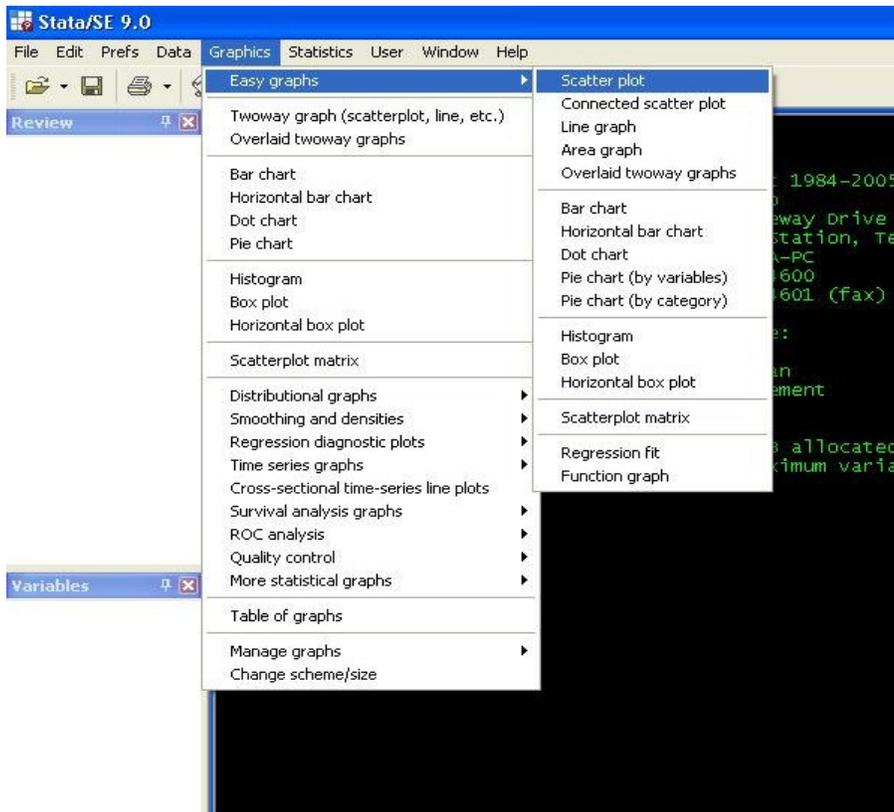
Figure 1. Scatterplots in Stata

Figure 2. Residual plots in Stata