

# A Commentary on Stata for Business

## Dr Con Menictas Applied Statistician in Business and Marketing Sciences

Modern business analytics has come a long way in recent years with the focus on statistical modelling providing a competitive advantage. In the past, few models were used for prediction in business, such as OLS, logistic regression and simple time series models. In today's business world and especially with the advances we have seen in computer technology, model complexity is more often called for than in the past.

Increases in market competition mean that the business analyst needs to act quickly and be able to manipulate data and build complex models fast, in order to gain market advantage. Even though there is a current trend of finding new and complex modelling software in the Open Source arena, the business analyst is usually time pressed and relied upon for accurate and optimal results quickly. In serious organisations, where there is little time to be searching for the correct code or the optimal graphics to convey a story to management, Stata outshines all other statistical software, because of its ease of use.

Stata is the leading package that allows the business analyst to do three very important things. *First*, it easily handles over 120,000 variables and over 2 billion rows of data. *Second*, Stata has the gentlest programming learning curve of any statistical package out there today and *third*, it has the most comprehensive suite of the latest statistical models to be found in any statistical package. Stata is the only platform that can manage data and analysis at the same time at this level, and offers coding speed like no other. Stata's interactive coding is vetted and approved before it is adopted. Strict convention is adhered to for example:

```
model depvar [indepvars] [if] [in] [weight] [, options]
```

Source	SS	df	MS	Number of obs	=	74
Model	195877890	4	48969472.6	F(4, 69)	=	7.69
Residual	439187506	69	6365036.32	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.3084
				Adj R-squared	=	0.2683
				Root MSE	=	2522.9

  

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpg	-264.2765	64.70349	-4.08	0.000	-393.3565 -135.1966
headroom	-574.618	467.3848	-1.23	0.223	-1507.025 357.7891
trunk	154.6971	103.7686	1.49	0.141	-52.31561 361.7097
foreign	1825.733	710.6067	2.57	0.012	408.1117 3243.355

Figure 1: Regression Model Output of the `auto` Stata Dataset as an Example

The business analyst is often required to provide visual stories for management because of the need to illustrate the results of an analysis. Stata graphics offer publication standard graphs in addition to markdown and dynamic documents, using very simple and intuitive coding for the purposes easily communicating and recording practically any analysis. Here is an example of a Stata graph (Figure 2) exhibiting very straightforward code.

```
histogram age, title(Histogram of Age) subtitle(Experienced Doctors)
```

### Survey Design and Analysis Services

PO Box 5044  
LYNEHAM ACT 2602  
AUSTRALIA

02 6247 0177 (AU); 09 889 2231 (NZ)  
sales@surveydesign.com.au  
surveydesign.com.au





Figure 2: Example of a Stata Graph

An example of Stata’s use in a business setting was for a complex choice model to estimate the optimal menu bundles for a fast food retailer that has global presence with outlets in all Australian states. A large dataset was provided in excess of 80 million rows and 17,000 variables on which to conduct an analysis prior to going to market research. The project required the construction of a choice model and a detailed multi-tiered segmentation.

A choice experiment was designed in a way that enabled the data that was collected to be used for a variety of choice models, in order to extract the maximum insights for both the product offering and customer type. We needed to estimate classical conditional logit choice models, mixed logit choice models, hierarchical choice models and MNL finite mixture models, in addition to latent class and segmentation models, for the purposes of satisfying client requirements.

Multinomial logistic regression
Multinomial probit regression
Conditional logistic regression
Alternative-specific conditional logit
Alternative-specific multinomial probit
Mixed logit model
Nested logit regression
Setup for nested logit regression
Display nested logit tree structure
Stereotype logistic regression
Multinomial logistic regression finite mixture model
Bayesian multinomial logistic regression
Bayesian multinomial probit regression
Bayesian conditional logistic regression

Figure 3: A Small Extract of Stata 's Choice Models

STATA was able to provide everything in one package, without the need to be using multiple packages throughout the project, with the fastest turnaround times every experienced by the project team for such a complex and demanding project. Given the intuitive drop-down menus (Figure 3) and easy to understand code, the project team had more time to concentrate on the problem at hand, rather than fumbling through code fragments trying to get things to ‘work’.

STATA is relatively new in the space of business analysis, but it is the most exciting piece of software that quite honestly has to be experienced to be believed. You will not be disappointed!

#### Survey Design and Analysis Services

PO Box 5044  
LYNEHAM ACT 2602  
AUSTRALIA

02 6247 0177 (AU); 09 889 2231 (NZ)  
sales@surveydesign.com.au  
surveydesign.com.au

