

A Commentary on Stata for Marketing Analytics

Dr Con Menictas
Applied Statistician in Business and Marketing Science

Marketing and marketing analytics are very complex disciplines. They span anthropology, psychology, sociology, economics, consumer behaviour, just to state a few. To optimally predict what customers and markets do, marketing analytics regularly call for advanced data manipulation and statistical and econometric models to be estimated quickly to predict human behaviour.

The current trend for a lot of analysts and modellers is largely divided into two camps. In the one camp, we have the *establishment* analyst, that feels at home using large and expensive corporate statistical environments. In the other camp, the recently *liberated* and in many cases *citizen* analyst/keyboard jockey, that prefers to use Open Source software to build their statistical solutions. At times, analysts delve into both camps which do require a lot of programming in both data preparation and management, and the building of statistical models.

However, in *real world* applied marketing environments, the requirement is for lightning fast turnaround times of model building and implementation. This is particularly the case when the business and the market it operates in are mature and saturated with competitors, where the main improvement in revenues comes from small percentage incremental gain. Speed here is a critical requirement and the number of variables and observations can be very large. There is no time to be searching for code or for making last minute changes to both software and hardware to accommodate big data.

Out of all the currently available statistical packages, no other package offers data management and statistical capability with speed as Stata does. Stata allows the analyst and model builder to focus more on solving the problem rather than wading through code and at times endlessly searching for code snippets on the internet.

This is where Stata shines regarding data management where it can easily accommodate in excess of 120,000 variables and more than 2 billion observations. In this space, SQL data manipulation is no match, be it via commercial or Open Source systems. Stata also comes in a 64MP version, meaning it can make use of up to 64 processors, thus turning around big data models in only a small fraction of the time that it would take competing software. Stata's interactive coding is vetted and approved before it is adopted. Strict convention is adhered to for example:

```
model price mpg headroom
```

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
_cons	10842.69	2342.933	4.63	0.000	6168.665 15516.71

Graphing in Stata is also extremely intuitive and fast and incorporates ggplot and ggplot2 aspects where required (Figure 2), including graphs for visual impairment. For example, a scatter plot with two variables

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

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



scatter price mileage

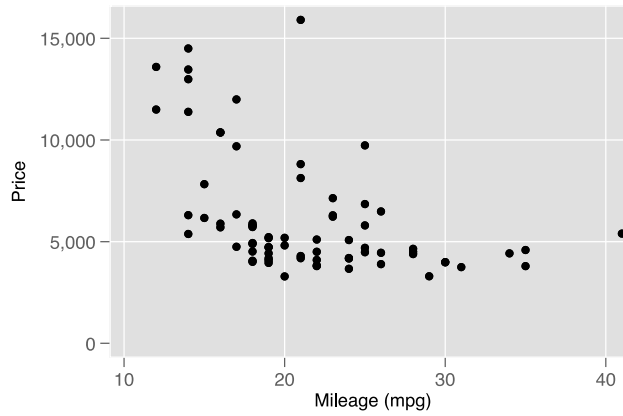
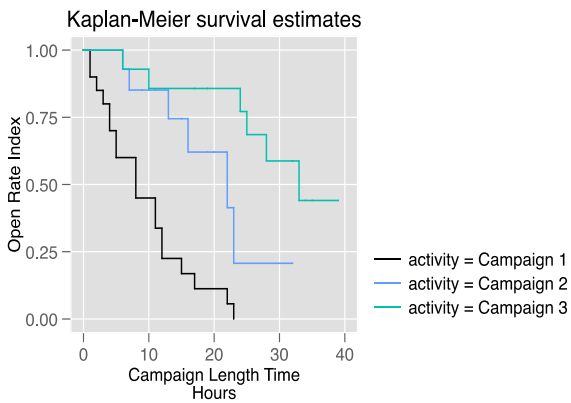


Figure 2: Example of a Stata Scatterplot

An example of Stata’s use in a marketing setting was to test the effectiveness of 3 versions of an email campaign that was sent to the client’s membership base over a weekend starting Saturday afternoon. A survival analysis was used to determine the open rate success rate for each campaign.

There were 10 mail-out experimental cells by 30 versions across ~30,000 customers and each test was modelled to see the best surviving campaign on 3 email version open rates. The analysis was conducted on-the-fly using Stata’s survival analysis capabilities over a weekend with results ready for the marketing team on Monday. Figure 3 illustrates Campaign 3 as being the most successful of the three campaigns tested and also shows the vastness of the survival models offered in Stata, to make the modelling tasks easy and quick!



Cox proportional hazard model
Test proportional-hazards assumption
Graphically assess proportional-hazards assumption
Kaplan-Meier versus predicted survival
Competing-risks regression
Parametric survival models
Interval-censored parametric survival models
Panel data parametric regression
Multilevel parametric regression
Parametric regression finite mixture model
Bayesian parametric survival models
Bayesian multilevel parametric survival models
Plot survivor, hazard, cumulative hazard, or cumulative incidence function

Figure 3: Stata’s Survival Models and Example

If you are serious about preparing data for statistical modelling and need to be confident and on time in a fast-paced marketing environment, look no further than Stata. It will free you up from coding nightmares and allow you to focus on improving the business and simply getting better at what you do!

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

