Regression Analysis

A survey such as the Bus Passenger Satisfaction Survey (BPSS) allows great scope for understanding what drives customer satisfaction, and with such a rich resource available, we are able to conduct extra analysis of key drivers that will add value and depth to the data we collect.

The survey allows us to develop a market model showing the importance of the various factors in terms of their impact upon overall customer satisfaction. We frequently use **regression analysis** to model this kind of data. Regression analysis is a statistical technique that allows one to assess the relationship between a dependent variable (in this case customer satisfaction) and several independent variables (the individual service features, e.g. length of time waited for the bus). The method is a multivariate analysis technique that will derive the relative importance of each service feature.

It is often preferable to use modelling techniques to derive customer priorities (derived importance) rather than asking the respondent direct (stated importance). The advantage of modelling the data to discover the importance of various factors is that the respondent is only asked to rate performance on relevant factors, which helps to reduce questionnaire length and respondent fatigue and boredom. Also, respondents are then totally unaware that we are deriving importance ratings, and so cannot consciously (or subconsciously) 'inflate' the importance of particular factors. For instance, when using stated importance, we often find that 'price' and similar cost-related factors may appear to have fairly high impact upon overall satisfaction. However, when we use modelled data, 'price' and similar factors often slip down the importance scale.

We can then look at the relationship between the satisfaction with a service aspect and its importance, to identify the key priorities for improvement. This information can be presented in a number of user-friendly ways.

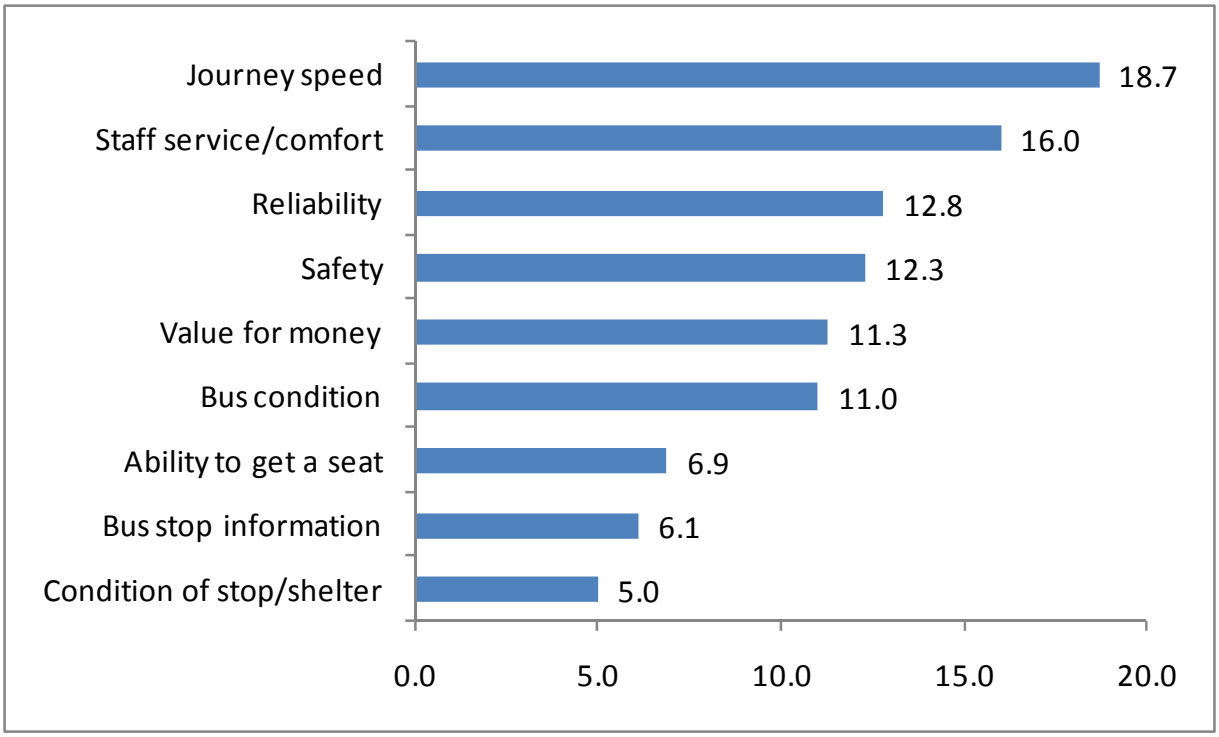
Key Influencers of Overall Satisfaction

Using data modelling techniques, we have derived the importance of the various service areas to the passenger. The analysis involves looking at the satisfaction scores for each of the KPIs and seeing how they correspond with the passengers' satisfaction with the overall service.

We have used a technique called ridge regression analysis to assess this relationship. Thus we can calculate derived importance scores for each service feature; the higher the percentage the more important the area in driving satisfaction with the overall service.

The graph below shows how important each aspect is in determining the overall satisfaction with the service. (Figures from Bus Passenger Satisfaction Survey Quarter 4, 2010)

Key Influencers of Overall Satisfaction



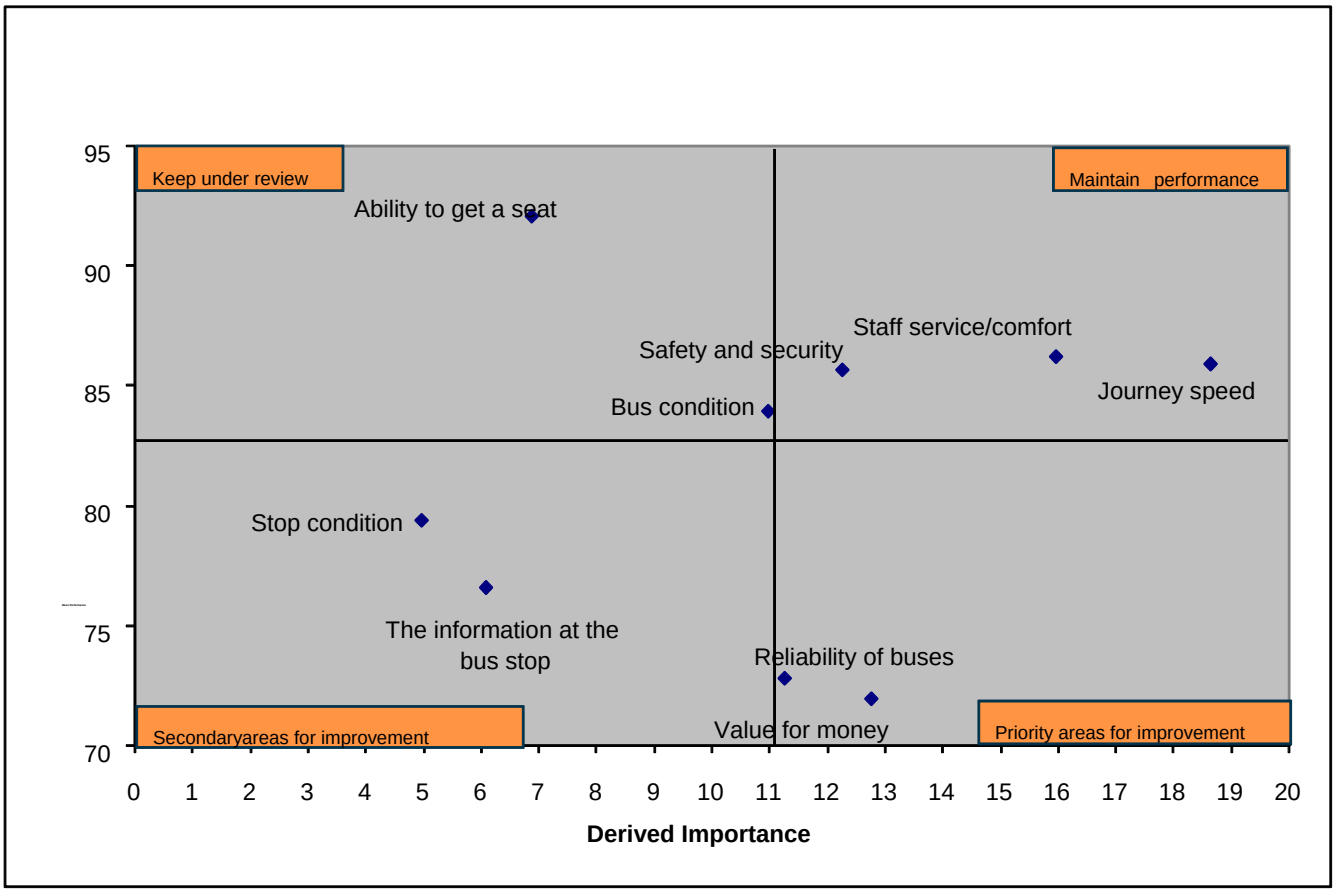
We also calculate a figure to show well the model fits (known as R^2) and it should be noted that only about 39% of any change in overall satisfaction is accounted for by changes in satisfaction with the different service aspects. This is slightly lower than would be expected of a good fit model (50%). However,

the analysis can still be used to establish the *relative* importance of each service feature in driving overall satisfaction.

Quadrant Maps

In the chart below, the derived importance and satisfaction score for each attribute have been mapped onto a Quadrant chart. The further to the right an attribute appears, the higher its importance score. The higher an attribute appears, the higher its satisfaction rating. We can then plot axes that represent the average satisfaction score and the average importance score. Thus, in the bottom right hand box we have attributes whose importance is high, but whose satisfaction levels are low: these are the areas of prime importance for the bus operator to address. In the bottom left hand box are attributes whose importance is lower, but whose satisfaction scores are still below average: these are secondary areas of importance for the bus operator to address. The top right box shows the attributes where the operator is performing well, but must continue to maintain standards, and the top left box shows attributes which are below average importance which are also performing well: these are the areas where attention is least needed.

Priorities for improvement



Ridge Regression

In market research we often use regression analysis to work out how much influence different service attributes have on overall satisfaction. However if the service attributes we measure are in some way correlated (for example as one improves another one is also likely to improve) then this ordinary type of regression can give the wrong answer.

Ridge regression is a special type of regression which is used to deal with this problem (often called collinearity).

In the example below we used ordinary regression and ridge regression to work out the relative influence of the 6 service attributes on customer satisfaction. These are beta coefficients rescaled such that they add to 100% for ease of comparison. The results are quite different with attributes 1 and 2 having much less influence in the ordinary regression.

Attribute number	Ordinary Regression	Ridge Regression
1	8%	13%
2	9%	14%
3	12%	16%
4	20%	18%
5	24%	20%
6	27%	19%

The reason for the difference can be seen below in the table of correlation coefficients. It is apparent that attributes 1 & 2 have slightly lower correlations with satisfaction. However it is also clear that attributes 1 & 2 are very closely related to attributes 3, 4 & 5. Whenever 2 or more attributes are correlated ordinary regression tends to give most of the influence to just one of the attributes. Ridge regression is very good at sharing out the communal influence between the attributes. Because Ordinary regression cannot deal with correlated attributes, it has basically given too much influence to attributes 4, 5 & 6. Ridge regression has used the inter-correlations more carefully to present a less biased picture.

Correlation Coefficients

	Overall Satisfaction	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5	Attribute 6
Attribute 1	.632	1.00	.66	.56	.40	.51	.27
Attribute 2	.646	.66	1.00	.59	.40	.55	.26
Attribute 3	.730	.56	.59	1.00	.57	.59	.28
Attribute 4	.747	.39	.40	.57	1.00	.58	.30
Attribute 5	.827	.51	.55	.59	.58	1.00	.36
Attribute 6	.706	.27	.26	.28	.30	.36	1.00