Audience Sizing

White Paper

Author: Neeti Singh & Raghu Raman M
February 2009
www.mu-sigma.com

Audience Sizing

An audience, or target group is the primary group of people at whom a product/service/solution is aimed. Discovering the appropriate target audience to market a product or service, is one of the most important stages in marketing. A target audience can segmented based on age group, gender, marital status, etc. (For example, teenagers, females, single people, etc.)

Need for sizing an audience

Determining the right audience, and most importantly the size of the audience, forms a critical part of any business. Audiences become central to an organization's business for a number of reasons. Some of these are:

- To analyze market opportunities and measure performance
- Determine the cost to reach an audience
- To find out the potential size of the market
- To enable accurate measurement of the effectiveness of marketing
- To estimate the scale of the challenge to better size investments, etc.
- Determine the coverage of an audience in a database.

Methodology Overview

One of the methods to size an audience is to build an econometric model that estimates the size of the target audiences across countries or across groups of countries. This model estimates the size of the audiences in each country and sub-region as a mathematical function of its demographic, economic, and/or other requisite characteristics. The purpose of modeling is to synthesize various audience estimates from various sources, overcoming variations such as definitional or methodological differences, incorporating "sometimes conflicting views" using a methodology that is mathematically rigorous and repeatable.

Methodology

Econometric models are typically developed using linear regression. Log linear regression comes into use when economic variables are a part of regression. This is because growth among economic variables is usually exponential. Regression provides us with not just the relationship between variables or a group of variables, but it also explains the individual impact of each input variable on the overall quality of the relationship.

It is quite likely that implementing one model alone might not work as it may not be the right fit for the data available. In such cases, models with different variable combinations should be tried. Model that fits best with respect to the training data should be used for further analysis and interpretations.

A model should be such that it addresses clearly why and for whom the audience estimates are being developed. The audience should be identified beforehand. The factors/variables this audience should be comparable needs to be decided upon before building the model.

Data:

Projects based on audience sizing involve dealing with a variety of datasets. The first and foremost step is fitting an econometric model. This involves determining the right input variables that could possibly drive the output variable. For this purpose, data on appropriate variables are obtained from reliable sources. However, there may be a possibility that though the source is dependable, data available may be fragmented; intending one may encounter the problem of missing data.

Missing data treatment:

Listed below are a few ways to estimate missing values:

I. Use other reliable data sources:

The aim should be to fill in as much data as possible using reliable sources for the same variable; typically looking for missing data points for a variable from multiple sources. However, the different data sources may have the numbers available in different units or may have different base years (in case of economic data). In case of annual data, the values available could be of a previous year. In this case, if on observation the change is not noticeable in the data points in comparison to the previous year's, the previous year's data from the other source in place of the missing data point can be reflected.

II. Data treatment using growth rates:

Another method would be to reflect growth rates to fill in the missing values. It is possible that even after looking at multiple data sources a few data points may still be missing. In such instances, proxies or running regression could be looked at to get the desired complete set of data variables.

III. Use of Proxies:

Taking proxies based on geographic, economic or demographic proximity is a convenient method to fill in the missing values. Proxies may be used in two ways:

- One could be taking direct proxies of a country. This is similar to the country that has missing data points for some input variable. However, if there are a number of input variables of similar type, say, economic, and the proxy of another country is directly reflected, then in this case there will be a number of data points having the same value. For example, there are 4 different economic variables as input variables and Country A has data points for three of these as missing and Country B has all the data points missing. If Country C is similar to Country A and Country B and has all the data points, then, these data points from Country C may be directly reflected in place of the missing data points for both Country A and Country B. In this scenario, it is advisable to use suitable scaling methods.
- The second method could be to reflect some scaled ratio based on a proxy data point. For example, if country A had some missing data points for an economic variable and is similar to country B in terms of economic proximity, then take the ratio of an available data point in another economic variable in country A to the same variable in country B and reflect the same ratio for the missing data point in country A.

Note: Country has been used to explain the concept of proxy. However, this method can be used on any segment, market, etc. for which audience sizes have to be determined.

IV. Regression:

Regression is another effective and convenient way for filling in missing data points. One may model the input variable containing missing data points as a function of some suitable variables to obtain the estimates. One should note that the variables should be such that they drive the output variable and give a significant R square value. There are cases however, when models with small R square values will be considered. This is a subjective decision which can be perfected by experience.

Modeling:

Once the dataset is complete, the econometric model is run to obtain the estimates of the audience size for the desired segment/ market/ country. The model that best fits the training data is selected. To establish the credibility of the model, a confidence range within which the model holds well is required.

Estimating Confidence Intervals:

The following formula is a simple method to obtain the confidence range with α % confidence:

Upper limit = estimated audience size + Zα * sqrt (Standard Error)

Lower limit = estimated audience size - Zα * sqrt (Standard Error)
where, Zα is the value at significance level α

The numbers can then be published or used in conjunction with the confidence intervals.

Summary:

Audience forms an imperative part of any business. The various audiences can be sized using regression models to analyze market opportunities effectively. Though it may seem difficult to size an audience, defining the right econometric model and the data that goes into the model solves most of the problem. Once this is done the estimates can be used to reach the target audience cost effectively and determine marketing effectiveness.

About Mu Sigma:

Mu Sigma helps clients institutionalize analytics in their organizations using global delivery. We are headquartered in Chicago, USA with a delivery center in Bangalore, India. Mu Sigma's scientific community, which consists of practitioners from leading educational institutions in the United States and India, enable us to deploy cutting edge analytics for our clients. Our best-in-class processes leverage expertise in statistics and econometrics in the areas of marketing, risk and supply chain. The techniques our professionals use range from conventional statistical and operations research techniques to advanced artificial intelligence techniques.