Modelling buyer behaviour/3

How survey results can mislead

By Chuck Chakrapani

From the past to the future

So far in this series of articles, we have discussed what the consumer has done. Yet what the consumer did in the past is of very little interest unless it leads to an understanding of what s/he will do in the future.

We should be able to go from the manifest purchase frequency of a consumer to his or her *underlying rate of purchase*. It is this latent variable - the rate, which is not directly measurable - that provides the tool to predict the future behaviour of a consumer.

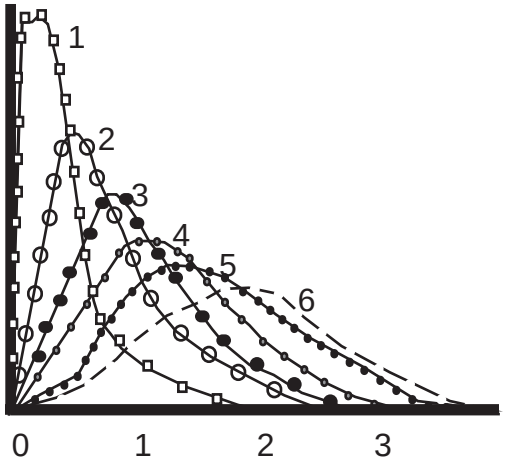
Different frequency classes and their distributions

As we noted earlier in this series, different purchase frequencies give rise to different distributions. We can plot some of these curves for different underlying rates of purchase (see figure below). Each curve in the figure is a gamma distribution (as described in the previous article). As we noted in that article, the mean of the distribution is .785. To illustrate the point, let us assume that we obtained 2 weeks of data and the purchase frequency over the two weeks was 1.57 and we averaged this for each week to obtain .785 per week. Each curve on the figure below represents the rate distribution that corresponds to each frequency class obtained in the survey over a two week period. The horizontal (x) axis represents the underlying rate of purchase. To keep the graph simple, underlying rates over 3 are not plotted.

You will note from the graphs that the purchase frequency curves of consumers with different latent rates of purchase overlap a great deal. This shows that consumers with higher short term frequencies can have lower underlying rates while consumers with lower short term frequencies can have higher underlying rates.

Rate distributions

Each curve represents the frequency distribution of rates for given survey frequencies. The horizontal axis shows the underlying rate of purchase. The vertical axis shows the proportion of customers



Past and future behaviour

To see how the model really can help us understand buyer behaviour patterns, let us look at the actual data.

If we had not used the model, we could easily make the mistake of assuming that if a consumer makes 2 purchases in 2 weeks, then his/her weekly rate is 1 (2+2=1). but in fact, according to the Gamma-Poisson model, the mean rate should be calculated according to the following formula (The proof is not provided here):

Mean rate

= (m+a) divided by (T + a/c)

where, m = Survey frequency of purchase; a = Mean survey frequency + length of survey reporting period; c= Square of survey mean + (Variance - mean) For details see Imprints 5/1993.

The table below shows the difference in estimates of underlying purchase frequency obtained by using simple averages and using the model.

Simple averages can be misleading. They assume that consumers act consistently from one time period to another. Many researchers calculate simple averages to understand buyer behaviour. Consumer behaviour, however, is somewhat more complex than what can be inferred by calculating simple averages.

For instance if a person eats out twice a month during the survey period, it is assumed that over a period of 3 months he/she would eat out 6 times. This may not necessarily be so as the 'model mean' column indicates.

One should note that the reason why a consumer may not repeat his or her past behaviour as measured in a survey has nothing to do with sampling error. Even when the sampling error is not a factor, simple averages will not predict future behaviour.

Simple & Model based means

Obs.Freq.	*% HHs*	*Simple mean*	*Model mean*
0	60.0	0.0	.17
1	21.0	0.5	.43
2	9.5	1.0	.69
3	4.6	1.5	.96
4	2.3	2.0	1.22
5	1.2	2.5	1.48
6	0.6	3.0	1.74
7	0.3	3.5	2.01
8	0.2	4.0	2.27
9+	0.1	4.5	2.53+

• Simple mean is obtained by dividing the number of purchases by the number of weeks.
• Model mean is obtained by the formula (see text)
• % HHs - also obtained through the model

Spurious relationships

In any given period, most consumers exhibit behaviours that are not 'typical' as represented by the average frequency.

If we calculate the average number of cups of coffee drunk by a consumer during a given week, the frequency is most likely to be above or below the underlying rate of consumption. For some consumers it could be much above the average and for others it could be much below the average. If we assume the average to be the 'typical' behaviour, during any given period most consumers will behave atypically.

This is because when we study buyer behaviour, we not only sample buyers but also the time period. Just as any given consumer's behaviour may not be assumed to represent the behaviour of all consumers, the buying frequency of a consumer at any given time may not be assumed to represent his/her buying frequency at other times as well.

Because a consumer was observed to have a high purchase frequency during the survey period, we cannot simply assume that that frequency can be projected to the future.

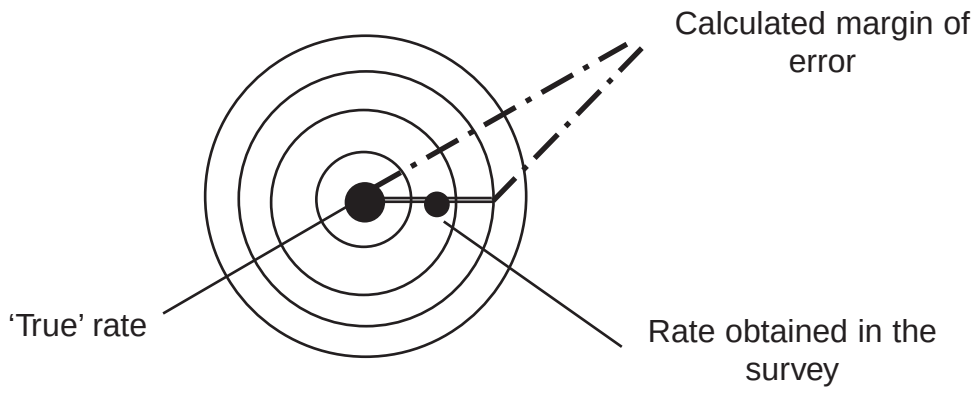
This can be shown to be the case not only from a theoretical point of view, but also from actual panel data gathered on the same consumers over a period of time.

To understand the relevance of this observation, consider two consumers who have identical underlying rates of purchase. Our model shows that although they both have the same underlying rate of purchase, during any given time period either one may exhibit a low or high purchase frequency. To simplify our discussion, let us assume that, in a given population, all consumers have the same underlying rate of purchase.

Even in this case some consumers will exhibit a high frequency of purchase while others will exhibit a low frequency, no matter when the survey is carried out.

If the analyst were to classify those with a low frequency as "light users' and those with a high frequency as 'heavy users' and study their demographic and psychographic characteristics, s/he would be effectively assigning different people to these groups arbitrarily. Such analyses can lead to potentially misleading conclusions thanks to the confounding effects of extraneous

There can be two types of sampling errors in measuring latent variables such as the 'true' rate of buying



Consider the 'bull's eye' to be the 'true' buying rate, a parameter to be estimated from survey research data. If the researcher uses routine interpretation of the data he or she might assume that the survey results will not hit the bull's eye because of the sampling error. Further, the researcher may set the margin of error based on this assumption. But because there is another type of sampling involved (sampling of one time period out of al possible time periods), the true rate could be much farther away and the assumption that the margin of error would include the 'true' rate at a given level of significance may be false. In other words, the obtained rate may not be where it is assumed to be by the researcher but probably in a circle that is further away from he 'true' rate.

variables.

Extending the time frame

What happens if we extend the time frame to a longer time period?
In this case, we are increasing the time period samples. Just as increasing the sample of consumers brings us closer to the population mean of any attribute, so increasing the sample of time periods brings us closer to the underlying rate for any given consumer.
This phenomenon - known as the regression towards the mean - is described below.

Regression towards the mean

In our arbitrary example, we saw that, in a survey, different consumers may have different rates of purchase during the survey period although they all may have the same underlying rate of purchase.
If, instead of sampling one week we sample 13 weeks, we may see that those who had a low purchasing frequency during the first time period compensate for it later by buying more and those who had a high purchasing frequency during the first time period compensate for it by buying less during subsequent time periods.
As a consequence, all consumers with the same underlying rates converge to the mean which is the underlying rate of purchase. Simply put, when sampled time periods increase, the simple means will converge towards the underlying rate. This phenomenon is known as regression towards the mean.

Survey analysis vs. buyer behaviour models

The implicit model used by marketing researchers is as follows:

Frequency of buying
= underlying rate + sampling error

The sampling error is assumed to be purely a function of the sample size. The model is correct for the survey period, but it breaks down when extended to cover future behaviour.
A more complete model would be of the form:

Frequency of buying
= underlying rate + specific influences + sampling error

In other words, a person's frequency is constantly being influenced by external factors.
Sampling error is not the only factor to be considered.

For instance, a person's buying rate could be influenced by factors such as being on holidays, being ill, guests arriving and the product being on sale.
Sampling error can be reduced by increasing the sample size. Variations arising from time sampling can be reduced through the use of a larger time frame or may be estimated through mathematical models.

Level of measurement

We argued that when we compare two consumers with different frequencies of purchase on a given week, it does not necessarily follow that they have different underlying rates of pur-

chase. The converse of this - that consumers who have the same frequency of purchase on a survey may have different underlying rates of purchase - directly follows from the above argument.

It should be noted, however, that when we look upon groups of customers with different purchase frequencies and rank them accordingly, we are likely to get closer to the underlying rates on the ordinal scale of measurement.

In other words, there are likely to be more consumers buying more among those with a high underlying rate of purchase than among consumers with a low underlying rate of purchase. Therefore, grouping consumers as per survey data is likely to preserve the rank order, but not the metric properties of the scale.

Minimizing errors of analysis

Understanding the underlying patterns through an analysis of panel data over a period of time or approximating them through buyer behaviour models will reduce the erroneous analysis of data. This can be particularly critical when we relate a consumer's usage to his or her demographic, media and psychographic characteristics.

Dr. Chuck Chakrapani is a marketing research consultant.