A discussion on *Sampling* in the Marketing Research

March 28, 2012

muhammad.zubair@foresight.com.pk

foresight|research

There is no set percentage that is accurate for every population !

- What matters is the actual number or size of the sample, not the percentage of the population
- For example, let's take a case of finding out average income of a group of 300 program participants. Taking a 20% sample of this group will give us a sample of 60 people. There is a fairly large chance in this case of underrepresenting the population
- On the other hand, 20% of 30,000 county residents (a sample of 6,000) would be a wastefully large sample, and not significantly more accurate than a sample of around 400 residents chosen randomly

muhammad.zubair@foresight.com.pk

Steps involved in determining sample size

Margin of error

- It is the precision with which the sample predicts the true values in the population, also called sampling error
- For example for ±5% margin of error, 35% TOM awareness in the survey means that in the real-world population, TOM may very well be in between 30% and 40%

Confidence level

- A confidence level of 95% means that, if we were to draw the sample 100 times, 95 of these samples would have the true population value within the range of precision specified earlier, and 5 would be unrepresentative samples
- Higher confidence levels require larger sample sizes.

Degree of variability in the sample

- Variability is the degree to which the attributes are distributed throughout the population. A heterogeneous population, divided more or less 50%-50% on an attribute, will be harder to measure precisely than a homogeneous population, divided say 80%-20%.

muhammad.zubair@foresight.com.pk

Mathematical expression of the problem

$$\Pr\left(|p - P| \geq d\right) = \alpha$$

Where:

- p is the sample estimate of the attribute we want to estimate
- P is the true value of the attribute for the total population
- d is margin of error
- α is significance value (confidence level)

Formula for the sample size is:

$$n = \frac{\frac{t^2\, p\, q}{d^2}}{1+\frac{1}{N}\left(\frac{t^2\, p\, q}{d^2}-1\right)}$$

Where:

- q = (1- p)
- Usually we work at 5% margin of error i.e. d = 0.05, &
- 95% confidence interval, i.e. for α=.05 we get t = 1.96

muhammad.zubair@foresight.com.pk

Sample size table for 5% margin of error with 95% confidence level

Population	Variability in the sample		
	50%	40%	30%
100	80	79	77
125	95	94	91
150	109	107	103
200	132	130	124
250	152	150	142
300	169	166	156
350	184	180	169
400	197	193	179
600	235	229	211
800	260	253	231
1,000	278	270	245
2,000	323	312	278
4,000	351	338	299
6,000	362	348	307
10,000	370	356	313
15,000	375	360	316
20,000	377	363	318
100,000	383	368	322
1,000,000	385	369	323

- This table assumes a 95% confidence level, identifying a risk of 1 in 20 that actual error is larger than the margin of error (greater than 5%)

- A five percent margin of error indicates willingness to accept an estimate within +/- 5 of the given value

muhammad.zubair@foresight.com.pk

For any more questions on sampling contact your client service representative …

muhammad.zubair@foresight.com.pk