Determining Sample Size; How to Calculate Survey Sample Size

Hamed Taherdoost

Determining Sample Size; How to Calculate Survey Sample Size

HAMED TAHERDOOST [a, b]
[a] Research and Development Department, Hamta Business Solution Sdn Bhd, Malaysia
[b] Research and Development Department, Ahoora Ltd | Management Consultation Group, Malaysia
hamed.taherdoost@gmail.com http://www.ahooraltd.com http://www.hamta.org

Abstract: - The sample size is a significant feature of any empirical study in which the goal is to make inferences about a population from a sample. In order to generalize from a random sample and avoid sampling errors or biases, a random sample needs to be of adequate size. This study presents a summary of how to calculate the survey sample size in social research and information system research.

Key-Words: - Sample Size, Survey Design, Questionnaire Development, Calculate Sample Size, and Research Methodology

1 Survey Sample Size

In order to generalize from a random sample and avoid sampling errors or biases, a random sample needs to be of adequate size. What is adequate depends on several issues which often confuse people doing surveys for the first time. This is because what is important here is not the proportion of the research population that gets sampled, but the absolute size of the sample selected relative to the complexity of the population, the aims of the researcher and the kinds of statistical manipulation that will be used in data analysis (Taherdoost, 2016). While the larger the sample the lesser the likelihood that findings will be biased does hold, diminishing returns can quickly set in when samples get over a specific size which need to be balanced against the researcher's resources (Gill et al., 2010).
To put it bluntly, larger sample sizes reduce sampling error but at a decreasing rate. Several statistical formulas are available for determining sample size.
There are numerous approaches, incorporating a number of different formulas, for calculating the sample size for categorical data.

$$n = \frac{p\,(100-p)z^2}{E^2}$$

n is the required sample size
P is the percentage occurrence of a state or condition
E is the percentage maximum error required
Z is the value corresponding to level of confidence required

There are two key factors to this formula (Bartlett et al., 2001). First, there are considerations relating to the estimation of the levels of precision and risk that the researcher is willing to accept:
E is the margin of error (the level of precision) or the risk the researcher is willing to accept (for example, the plus or minus figure reported in newspaper poll results). In the social research a 5% margin of error is acceptable. So, for example, if in a survey on job satisfaction 40% of respondents indicated they were dissatisfied would lie between 35% and 45%. The smaller the value of E the greater the sample size required as technically speaking sample error is inversely proportional to the square root of n, however, a large sample cannot guarantee precision (Bryman and Bell, 2003).
Z concern the level of confidence that the results revealed by the survey findings are accurate. What this means is the degree to which we can be sure the characteristics of the population have been accurately estimated by the sample survey. Z is the statistical value corresponding to level of confidence required. The key idea behind this is that if a population were to be sampled repeatedly the average value of a variable or question obtained would be equal to the true population value. In management research the typical levels of confidence used are 95 percent (0.05: a Z value equal to 1.96) or 99 percent (0.01: Z=2.57). A 95 percent level of confidence implies that 95 out of 100 samples will have the true population value within the margin of error (E) specified.
The second key component of a sample size formula concerns the estimation of the variance or heterogeneity of the population (P). Management

researchers are commonly concerned with determining sample size for issues involving the estimation of population percentages or proportions (Zikmund, 2002). In the formula, the variance of a proportion or the percentage occurrence of how a particular question, for example, will be answered is P(100-P). Where, P= the percentage of a sample having a characteristic , for example, the 40 % of the respondents who were dissatisfied with pay, and (100-P) is the percentage (60%) who lack the characteristic or belief. The key issue is how to estimate the value of P before conducting the survey? Bartlett et al. (2001) suggest that researchers should use 50% as an estimate of P, as this will result in the maximization of variance and produce the maximum sample size (Bartlett et al., 2001).
The formula for determining sample size, of the population has virtually no effect on how well the sample is likely to describe the population and as Fowler (2002) argues, it is most unusual for it (the population fraction) to be an important consideration when deciding on sample size (Fowler, 2002).
Table 1 and 2 present sample size that would be necessary for given combinations of precision, confidence levels, and a population percentage or variability of 50% (the figure which many researchers suggest to maximize variance).

Table 1: Sample size based on Desired Accuracy with Confidence Level of 95%
Source: (Gill et al., 2010)

`	**Variance of the population P=50%**		
	Confidence level=95% Margin of error		
Population Size	5	3	1
50	44	48	50
75	63	70	74
100	79	91	99
150	108	132	148
200	132	168	196
250	151	203	244
300	168	234	291
400	196	291	384
500	217	340	475
600	234	384	565
700	248	423	652
800	260	457	738
1000	278	516	906
1500	306	624	1297
2000	322	696	1655
3000	341	787	2286
5000	357	879	3288
10000	370	964	4899
`	**Variance of the population P=50%**		
	Confidence level=95% Margin of error		
25000	378	1023	6939
50000	381	1045	8057
100000	383	1056	8762
250000	384	1063	9249
500000	384	1065	9423
1000000	384	1066	9513

Table 2: Sample size based on Desired Accuracy with Confidence Level of 99%
Source: (Gill et al., 2010)

`	**Variance of the population P=50%**		
	Confidence level=99% Margin of error		
Population Size	5	3	1
50	46	49	50
75	67	72	75
100	87	95	99
150	122	139	149
200	154	180	198
250	181	220	246
300	206	258	295
400	249	328	391
500	285	393	485
600	314	452	579
700	340	507	672
800	362	557	763
1000	398	647	943
1500	459	825	1375
2000	497	957	1784
3000	541	1138	2539
5000	583	1342	3838
10000	620	1550	6228
25000	643	1709	9944
50000	652	1770	12413
100000	656	1802	14172
250000	659	1821	15489
500000	660	1828	15984
1000000	660	1831	16244

2 Conclusion

As mentioned, for sample size calculation, Table 1 and 2 or the provided formula can be used. The sample sizes reflect the number of obtained responses, and not necessarily the number of questionnaires distributed (this number is often increased to compensate for non-response). However, in most social and management surveys, the response rates for postal and e-mailed surveys are very rarely 100% (Taherdoost, 2016). Probably the most common and time effective way to ensure minimum samples are met is to increase the sample size by up to 50% in the first distribution of the survey (Bartlett et al., 2001).

References:

[1] BARTLETT, J. E., KOTRLIK, J. W. & HIGGINS, C. C. 2001. Organizational Research: Determining Appropriate Sample Size in Survey Research. Learning and Performance Journal, 19, 43-50.

[2] BRYMAN, A. & BELL, E. 2003. Business Research Methods, Oxford, Oxford University Press.

[3] FOWLER, F. J. 2002. Survey research methods, Newbury Park, CA, SAGE.

[4] GILL, J., JOHNSON, P. & CLARK, M. 2010. Research Methods for Managers, SAGE Publications.

[5] TAHERDOOST, H. 2016. Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. International Journal of Advance Research in Management, 5(2), 18-27.

[6] TAHERDOOST, H. 2016. How to Design and Create an Effective Survey/Questionnaire; A Step by Step Guide. International Journal of Advance Research in Management, 5(4), 37-41.

[7] ZIKMUND. 2002. Business Research Methods, Dryden, Thomson Learning.