2012 Minnesota Internet Survey

Data Collection Methodology Report

Social Science Research Institute
University of North Dakota
Knowledge to Bring People
and Resources Together

Cordell A. Fontaine, Director
Cordell.Fontaine@email.und.edu
Office: 701.777.3593

Table of Contents

2012 Minnesota Internet Survey

Methodology Summary

Introduction

The 2012 Minnesota Internet Survey, commissioned by the Center for Rural Policy and Development conducted by SSRI, interviewed a total of 1,662 adults in Minnesota. Table 1 presents the final dual-frame sample statistics. Statistical results were weighted to correct known demographic discrepancies.

Table 1. Final Dual-frame Sample Statistics

Target Area	Popu. 18+	Landline	Total Cellular	Total	CPO Only	Margin of Error (95%)
Twin Cities Counties	2,148,174	584	242	826	(116)	3.58%
Rest of Minnesota	1,860,424	641	195	836	(114)	3.58%
Statewide						2.53%
Total	**4,008,598**	**1,225**	**437**	**1,652**		

Sample Design

A combination of landline and cellular random digit dial (RDD) samples was used to represent adults in the target areas who have access to either a landline or cellular telephone. Both samples were provided by Marketing Systems Group (MSG), Genesys Sampling Systems[1] according to SSRI specifications. Landline telephone numbers were generated using GENESYS, a stand-alone, in-house RDD windows based program through MSG. Cellular RDD sample replicates were purchased from MSG based upon cellular prefixes in the respective survey areas based on switch locations[2].

SSRI starts with a database of all listed telephone numbers, updated on a four- to six-week rolling basis, 25 percent of the listings at a time. All active blocks—contiguous groups of 100 phone numbers for which more than one residential number is listed—are added to this database. Blocks and exchanges that include only listed business numbers are excluded.

Numbers for the landline sample were drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained three or more residential directory listings. The cellular sample was not list-assisted, but was drawn through a systematic sampling from dedicated wireless 100-blocks and shared service 100-blocks with no directory-listed landline numbers.

Contact Procedures

The telephone interviews were conducted from May 16 through June 23, 2012. As many as eight attempts were made to contact every sampled telephone number. Sample was released for interviewing in replicates, which are representative subsamples of the larger sample. Using replicates to control the release of sample ensures that complete call procedures are followed for the entire sample. Calls were staggered over times of day and days of the week to maximize the

[1] Marketing Systems Group, 565 Virginia Drive, Fort Washington, PA, 19034, 1-800-336-7674 www.genesys-sampling.com

[2] Please see Appendix A: Construction of Cellular RDD Sampling Frames based on Switch Locations.

chance of making contact with potential respondents. Each phone number received at least one daytime call.

Response Rates

The disposition of all sampled telephone numbers ever dialed from the original telephone number samples are presented for the two county-based regions - Twin Cities (Table 1) and Rest of State (Table 2). The response rate estimates the fraction of all eligible respondents in the sample that were ultimately interviewed.

Table 1. Twin Cities Sample Dispositions

	Landline	Cellular
Total	9,500	3,000
Released	9,500	3,300
Unreleased	0	0
Usable	**4,893**	1,848
Unusable	4,608	1,152
Qualified	4,853	1,808

DETAIL		
Disconnected	4,569	1,080
Fax	29	5
Govt./Business	10	67
Unusable	4,608	1,152
No Answer/Busy	493	16
Usability Unknown	493	16
Complete	584	242
Break-off	6	8
Usable/Eligible	590	250
Refused	2,147	817
Language Barrier	51	8
Voice Mail	1,336	523
Call Back	236	194
Usable/Eligible Unknown	3,770	1,542
Under 18		125
Response Rate	**12.0%**	**13.4%**

Table 2. Rest of Minnesota Sample Dispositions

	Landline	Cellular
Total	8,000	3,000
Released	8,000	3,000
Unreleased	0	0
Usable	4,720	1,920
Unusable	3,280	1,080
Qualified	4,910	1,409
DETAIL		
Disconnected	3,249	1,056
Fax	21	10
Govt./Business	10	14
Unusable	3,280	1,080
No Answer/Busy	493	44
Usability Unknown	493	44
Complete	641	198
Break-off	6	20
Usable/Eligible	647	218
Refused	2,147	614
Language Barrier	51	19
Voice Mail	1,336	380
Call Back	236	134
Usable/Eligible Unknown	3,770	1,147
Under 18		213
Response Rate	**13.1%**	**13.8%**

WEIGHTING METHODOLOGY FOR MINNESOTA INTERNET SURVEY

Design Overview

This study has secured a total of 1,662 interviews with adults 18 years of age or older residing in Minnesota. In order to provide a probability-based sample representative of all households, a dual-frame RDD sampling methodology was used whereby both landline and cellular telephone numbers were used to reach eligible adults. Moreover, the entire state was divided into two county-based regions – Twin Cities and the balance of the state – for each of which about 800 interviews were secured. The following table provides a summary of the sampling design specifications and respondent counts for this survey.

Table 1. Sampling design summary and respondent counts by region and frame type

Region	Frame Type	Universe	Sample	Respondents
Twin Cities	Landline	2,763,800	9,500	584
	Cellular	4,015,000	3,000	242
Rest of State	Landline	2,766,400	8,000	641
	Cellular	2,906,000	3,000	195
Total		**12,451,200**	**23,500**	**1,662**

Weighting

Virtually all survey data are weighted before they can be used to produce reliable estimates of population parameters. While reflecting the selection probabilities of sampled units, weighting also attempts to compensate for practical limitations of a sample survey, such as differential nonresponse and undercoverage. The weighting process for this survey consisted of two major steps. In the first step, *base weights* were calculated to reflect selection probabilities and the selection of one adult per household. In the second step, base weights were adjusted so that the resulting final weights aggregate to reported totals for the target population in each region.

For the second step, final weights were calculated using the method of *raking*. Specifically, design weights were simultaneously adjusted along the following raking dimensions using the *WgtAdjust* procedure of SUDAAN. It should be noted that survey data for a number of demographic questions, such as race and age included missing values. All such missing values were first imputed using a *hot-deck* procedure before construction of the survey weights. As such, respondent counts reflected in the following tables correspond to the post-imputation step. Also, since the two regions were sampled disproportionally, an additional adjustment step was necessary to ensure that results of this survey would properly represent the entire state.

Table 2. First raking dimension for weight adjustments by gender and age

Age	Universe			Respondents		
	Male	Female	Total	Male	Female	Total
18-24	239,175	237,791	**476,966**	76	71	**147**
25-34	376,984	377,785	**754,769**	109	129	**238**
35-44	320,127	310,155	**630,282**	133	130	**263**
45-54	419,938	382,286	**802,224**	181	198	**379**

55-64	273,274	307,937	**581,211**	146	147	**293**
65+	286,533	393,203	**679,736**	145	197	**342**
Total	**1,916,031**	**2,009,157**	**3,925,188**	**790**	**872**	**1,662**

Table 3. Second raking dimension for weight adjustments by gender and race/ethnicity

Race/Ethnicity	Universe			Respondents		
	Male	Female	Total	Male	Female	Total
Hispanic	78,461	58,034	**136,495**	26	9	**35**
Non-Hispanic White	1,688,216	1,776,941	**3,465,157**	718	798	**1,516**
Non-Hispanic Other	149,354	174,182	**323,536**	46	65	**111**
Total	**1,916,031**	**2,009,157**	**3,925,188**	**790**	**872**	**1,662**

Table 4. Third raking dimension for weight adjustments by gender and education

Education	Universe			Respondents		
	Male	Female	Total	Male	Female	Total
Less than high school	168,239	149,571	**317,810**	27	18	**45**
High school diploma	575,409	517,916	**1,093,325**	165	205	**370**
Some college, no degree	291,885	399,593	**691,478**	147	146	**293**
Associates or equivalent	282,369	297,925	**580,294**	114	109	**223**
Bachelors or equivalent	423,937	501,425	**925,362**	221	292	**513**
Masters degree or higher	174,192	142,727	**316,919**	116	102	**218**
Total	**1,916,031**	**2,009,157**	**3,925,188**	**790**	**872**	**1,662**

Table 5. Fourth raking dimension for weight adjustments by gender and region

Region	Universe			Respondents		
	Male	Female	Total	Male	Female	Total
Twin Cities	1,068,860	1,109,608	**2,178,468**	389	437	**826**
Rest of the State	847,171	899,549	**1,746,720**	401	435	**836**
Total	**1,916,031**	**2,009,157**	**3,925,188**	**790**	**872**	**1,662**

Table 6. Fifth raking dimension for weight adjustments by phone status and region

Region	Universe			Respondents		
	Twin Cities	Rest of State	Total	Twin Cities	Rest of State	Total
Cell-only	603,435	452,400	**1,055,835**	116	114	**230**
Others	1,575,033	1,294,320	**2,869,353**	710	722	**1,432**
Total	**2,178,468**	**1,746,720**	**3,925,188**	**826**	**836**	**1,662**

Variance Estimation for Weighted Data

Survey estimates can only be interpreted properly in light of their associated sampling errors. Since weighting often increases variances of estimates, use of standard variance calculation formulae with weighted data can result in misleading statistical inferences. With weighted data, two general approaches for variance estimation can be distinguished: *Taylor Series Linearization* and *Replication*. There are several statistical software packages that can be used to produce design-proper estimates of variances using linearization or replication methodologies, including: SAS, SUDAAN, SPSS, and Stata.

Also, an approximation method for variance estimation can be used to avoid the need for special software packages. Researchers who do not have access to such tools for design-proper estimation of standard errors can approximate the resulting variance inflation due to weighting and incorporate that in subsequent calculations of confidence intervals and tests of significance. With W_i representing the final weight of the i^{th} respondent, the inflation due to weighting, which is commonly referred to as *Design Effect*, can be approximated by:

$$\delta = 1 + \frac{\sum_{i=1}^{n} \frac{\left(W_i - \bar{W}\right)^2}{n-1}}{\bar{W}^2}$$

For calculation of a confidence interval for an estimated percentage, $\hat{p}$, one can obtain the conventional variance of the given percentage $S^2(\hat{p})$, multiply it by the approximated design effect, δ, and use the resulting quantity as adjusted variance. That is, the adjusted variance $\hat{S}^2(\hat{p})$ would be given by:

$$\hat{S}^2(\hat{p}) = \delta\, S^2(\hat{p}) = \delta \frac{\hat{p}(1-\hat{p})}{n-1}\left(\frac{N-n}{N}\right)$$

Subsequently, the (100-α) percent confidence interval for P would be given by:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\delta \frac{\hat{p}(1-\hat{p})}{n-1}\left(\frac{N-n}{N}\right)}$$

Appendix A

Construction of Cellular RDD Sampling Frames Based on Switch Location

Construction of Cellular RDD Sampling Frames Based on Switch Locations

Constructing cellular sampling frames for small geographic domains is subject to both operational and definitional challenges. Many of such challenges are due to the simple fact that, unlike landline telephone numbers, cellular numbers are assigned to mobile devices that may be located across the nation – if not the globe. In spite of this indeterminable mobility, however, most US cellular telephone numbers are assigned to exchanges that are native to specific locations as is the case with landline telephone numbers. Cognizant of these dynamics, MSG has developed a methodology for constructing cellular sampling frames for small areas based on the location each 1000-series block of cellular numbers is associated with. While not a one-to-one correspondence, with this methodology it is possible to identify the broader area (e.g., county) where the subscriber of a specific cellular number has a higher likelihood of residing.

Briefly, the North American Numbering Plan Administration (NANPA) is the governing body that regulates the assignment of all area codes, exchanges, and 1000-series blocks of telephone numbers in the US. The NANPA assignment protocols, which tend to be location-centric, apply uniformly to all types of numbers including those used for landline, cellular, and paging services. While area codes conform to state boundaries, for 1000-series blocks Switch Centers3 serve as the basic unit of geography for the telecom industry. Moreover, newly activated cellular numbers are assigned within a finite set of 1000-series blocks allocated to these switch centers.

Given that each switch center has a unique latitude and longitude, cellular switch centers and the set of 1000-series blocks they serve can be identified and included in the sampling frame for specific geographic locations. Unlike landlines for which their associated centers blanket the entire country, cellular switch centers tend to cluster around larger population centers. As such, in metro areas with high volume of telephone calls there can be many cellular switch centers whereas in rural areas such centers may cover several counties. In fact, less than half of the counties in the US have dedicated cellular switch centers. Consequently, the proposed methodology tends to have better coverage properties in populated areas. In order to better understand this situation, in what follows a brief description of the US cellular network topology is provided.

When a call is initiated by a cellular device the resulting signal is detected by the nearest Cell Site, which typically includes a tower or other elevated structure for mounting antennas and associated equipments for signal transmission. Most cell sites are connected to switch centers on a wired network, while others may rely on microwave technology for transmitting information through radio waves. Once a call has been detected and transmitted – either over the wired network or radio waves – the corresponding switch center determines the destination point for the given call and routes it out on the US telephony network. If the destination is a wired residence or business, the call is routed to the local Central Office to be connected to its final destination point. When the destination point is another cellular telephone, however, the closest

[3] Switch or wire centers describe the organization of the local telephone exchange system, with each center serving a unique set of exchanges and their associated telephone numbers.

cell site to the cellular device is identified in order to route the call to the corresponding switch center.