Surveys and Sampling

How to avoid dumb answers to dumb questions

I. Goals
A. Gain a lot of information
B. Do it in a cost effective fashion
C. Get a representative sample
D. Take as little time from your respondents as necessary

II. Question construction

III. Administering the Survey

IV. Sampling

V. Dumb Surveys

VI. Conclusions
II. Question Construction

A. General Principles
   1) avoid ambiguity
      Example: (question from OIT survey)
      Are you hooked on the internet?

   2) do not use “leading” questions
      Example: “Do you think pornographers should be allowed to sell their smut on the internet?”

   3) give responders appropriate options:
      Example: Which of the following is true for you:
      1) I use the internet everyday.
      2) I use the internet 3 times a week.
      3) I use the internet 3 or more times a week.
4) ask only ask what you need to know
Example: What is your email address?

5) consider the order of questions
Example:
1) There is a lot of pornography on the internet (agree or disagree)
2) I use the internet everyday (agree or disagree)

6) Pretest the survey
II. Question Construction

B. Types of Questions

1) Open ended questions:
   When you are “online,” what types of activities do you do?
   - get a lot of information
   but
   - hard to answer
   - hard to score

2) Fixed-alternative questions:
   Do you believe the internet is (circle one)
   - good  or  bad  for society.
   - easy to fill out
   - easy to score
   But - limits the information you can obtain.
   Consider adding qualifiers (“but”)
   Consider adding “other”

3) Closed-ended questions:
   How much of your free time do you spend “surfing” the internet?
   1. A lot
   2) Some
   3) A little
   4) None
4) Likert scales
Likert (1932) technique for measuring attitudes.
Example:
The internet has transformed society:
<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Uncertain</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

5) Semantic Differential
Osgood’s (1952) research on relations between concepts.
- Bipolar adjectives used to rate specific concepts
- Ratings typically done on a 7-point scale

Semantic Differential Example:
Please rate the Internet on the following dimensions (put an X on the appropriate line):
Fun                      Dull
Good                     Bad
Informative              Stupid
III. Administering the Survey

A. Face-to-Face Interviews
advantages:
  clarify questions
  see the “set” of the responders
  probe for additional information

A. Face-to-Face Interviews (cont.)
Disadvantages:
  potential bias
  expensive
  time consuming
  sampling bias?

B. Telephone Interviews
   Many of the same advantages of Face-to-Face
   More cost effective
But:
  limit the time (15 min)
  sampling bias?
C. Mail-in Questionnaires
advantages:
- mail anywhere
- cost effective
disadvantage
- low return rate
  - 75% - good
  - above 50% acceptable
  - below 50% questionable

C. Mail-in Questionnaires (cont)
How to improve return rates:
- use first class mail
- postage paid envelopes
- professional quality questionnaire
- easy to answer

D. Internet Surveys (research)
Online surveys:
- e.g., Survey Monkey (http://www.surveymonkey.com)
Online research:
- e.g., Project Implicit (http://projectimplicit.net/index.html)
Problems? (potential bias, “negativistic subjects”)


IV. Sampling

A. Goal: Obtain a representative data set to enable generalization to the population.

Famous example: 1936 Presidential Election
Literary Digest:
\[ n = 2,376,523 \]
sampled from phone directories and automobile registration
Predicted Alfred Landon (Republican) would win

B. Two Types of Sampling:

Probability Sampling
- relies on chance to determine inclusion in the sample
Nonprobability Sampling
- sample selected in other ways
Generally: Probability sampling is preferred

1936 Presidential Election (cont.)
Gallup, with a much smaller sample,
Correctly predicted Roosevelt would win.

Today: Most opinion polls contain less than 2000 individuals.
1) Probability Sampling

a) Simple Random Sampling:
   - each member of a population has an equal chance of being selected
   - good generalizability
   - difficult to accomplish

b) Systematic Sampling
   Select from the entire population by pulling every nth individual from a list.

   Example: Sampling registered voters from a list ordered by SS#

   - must have a complete list
   - must have list ordered in an unbiased fashion

c) Stratified Random Sampling
   Used to ensure that certain groups are systematically represented.

   Example: You may want to be sure that some African Americans included in your sample. Construct your sample to include 10% African Americans.
d) Cluster Sampling:
Randomly select a grouping (cluster) of individuals, and then evaluate everyone in that cluster.
Example: Randomly select a classroom time and place. Give the Internet Survey to everyone in that class.

e) Multistage Sampling:
Variant of Cluster Sampling
Randomly select a cluster, and then randomly select individuals from the cluster.
Example: Randomly select a University, then randomly select students from the University.

2. Nonprobability Sampling
a) Convenience Sampling
Sampling individuals who are readily available
Example: people passing by in a shopping mall
Not generalizable at all!
b) Quota Sampling
   fixed numbers of specific types of people
   Example: 10 Psych. majors, 10 Biology majors
c) Snowball Sampling
   Allow sampled people to help find other people to measure.
   Example: prostitutes and risky sexual behavior

C. Sample Size

1) Large samples increase generalizability
   (law of large numbers)
2) Large samples increase power
   (the ability to detect true differences)

3. How to calculate needed sample size:

   \[
   \text{Sample size} = \left( \frac{Z_{\text{confidence level}} \times \text{standard dev}}{\text{desired precision}} \right)^2
   \]

   However: standard deviation is typical unknown
Sample Size Example

\[ Z (95\% CI) = 1.96 \]
Assume population standard deviation of 20
Assume desired precision of 2

Sample size = \( \frac{1.96 \times 20}{2} \approx 20 \)
\[ 20^2 = 400 \]

Sample size and surveys: (Yamane, 1967)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>400</td>
<td>5</td>
</tr>
<tr>
<td>1,111</td>
<td>3</td>
</tr>
<tr>
<td>2,500</td>
<td>2</td>
</tr>
<tr>
<td>10,000</td>
<td>1</td>
</tr>
</tbody>
</table>

V. Dumb Surveys

Some Examples:
VI. Conclusions

Surveys should:
1) be short
2) ask appropriate questions
3) use representative samples
4) use large samples