Calibrating Cross-National Panel Surveys
Laura Wronski
Mingnan Liu

2016 3MC Conference
July 27, 2016
SurveyMonkey Internationally
Q1 2016

20 million accounts
60%

1 million surveys
54%

70 million responses
54%
Nonprobability Panel

Audience + panel providers

Thank you for taking this survey. Powered by SurveyMonkey, the world’s #1 online survey platform.

Feeling Charitable? Help Protect Animals.

Take a survey to raise $0.50 for a charity that cares for animals—or your favorite participating charity. And you could win $100.

Take Surveys Now or create your own survey.

Copyright © 1999 - 2014 SurveyMonkey. About SurveyMonkey • Privacy Policy • Contact Us
Goals

• Can we demonstrate that our panel is high quality?

• Reliability particularly important for SurveyMonkey customers who are purchasing panel specifically to run regular (monthly, quarterly, etc.) surveys

• Validity particularly important for SurveyMonkey customers who are trying to produce nationally representative estimates
Study setup

- From December 2015 through February 2016, we fielded the same survey:

  - ... with 25 questions...
  - ... for one week each month...
  - ... in Australia, Brazil, China, India, the UK, and the US...
  - ... receiving more than 1,000 responses per country per wave.
<table>
<thead>
<tr>
<th>Study setup</th>
<th>Preview link</th>
</tr>
</thead>
</table>

### Data quality
- Straight lining
- Open end validity
- Trap question

### Reliability
- Personal characteristics
- Factual questions
  - Credit card
  - Number of surveys
- Opinion questions
- Demographic questions

* Paradata: time to complete survey

### Validity
- Personal characteristics
- Factual questions
  - Credit card
- Demographics
Results should be consistent from one wave to the next.

Why? Even if members cycle in and out of the panel, we always want to be sure we maintain a panel that is consistent.

Test: Chi-square test across three waves for each variable, look for p-values < .05 to determine significant differences.

Examine both unweighted and weighted data:
- Some customers will weight their data
- Some will not
Reliability

• Personal characteristics
  • Eye color
  • Birth month
  • etc

• Credit/debit card ownership and use

• Morality
  • Alcohol
  • homosexuality

• Data quality
  • Picture verification
  • Open end validity

• Demographics
  • Age
  • Sex
  • Education
  • Race
  • Income
  • Device type

• Internet use/experience
  • Time spent online
  • Number of surveys taken
  • Survey quality rating

• Time to complete the survey
• Good reliability across all countries, but numbers vary across countries
  • BR, CH, IN had panel solely from an external panel provider
  • AU, UK, US had combined external and SM panel sources
• Inconsistent questions weren’t likely to be one category more than another
• Most improve with weighting
  • Exception is Brazil

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Brazil</th>
<th>China</th>
<th>India</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted</td>
<td>85%</td>
<td>92%</td>
<td>73%</td>
<td>81%</td>
<td>92%</td>
<td>85%</td>
</tr>
<tr>
<td>Weighted</td>
<td>100%</td>
<td>77%</td>
<td>85%</td>
<td>88%</td>
<td>96%</td>
<td>88%</td>
</tr>
</tbody>
</table>

% variables that passed the chi-square test
Validity

- When weighted, results should be comparable to gold standard benchmarks
  - Weights come from each country’s census
  - Benchmarks come from
    - Pew 2013 Global Attitudes survey
    - World Bank 2015 Global Financial Inclusion Database
- Why? Even if an internet panel has some inherent coverage bias, we want our panel to be as representative of each country as possible

The Pew Research Center’s 2013 Global Attitudes survey asked 40,117 respondents in 40 countries what they thought about eight topics often discussed as moral issues: extramarital affairs, gambling, homosexuality, abortion, premarital sex, alcohol consumption, divorce, and the use of contraceptives. For each issue, respondents were asked whether this is morally acceptable, morally unacceptable, or not a moral issue. The chart below displays the median responses for each question across the 40 countries.

<table>
<thead>
<tr>
<th>Median percentages of each issue based on 40 countries</th>
<th>UNACCEPTABLE</th>
<th>ACCEPTABLE</th>
<th>NOT A MORAL ISSUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extramarital Affairs</td>
<td>78%</td>
<td>7%</td>
<td>10%</td>
</tr>
<tr>
<td>Gambling</td>
<td>62%</td>
<td>11%</td>
<td>18%</td>
</tr>
<tr>
<td>Homosexuality</td>
<td>69%</td>
<td>20%</td>
<td>13%</td>
</tr>
<tr>
<td>Abortion</td>
<td>56%</td>
<td>15%</td>
<td>12%</td>
</tr>
<tr>
<td>Premarital Sex</td>
<td>46%</td>
<td>24%</td>
<td>16%</td>
</tr>
<tr>
<td>Alcohol Use</td>
<td>42%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>Divorce</td>
<td>24%</td>
<td>30%</td>
<td>22%</td>
</tr>
<tr>
<td>Contraception Use</td>
<td>14%</td>
<td>84%</td>
<td>21%</td>
</tr>
</tbody>
</table>
Validity

- Personal characteristics
  - Eye color
  - Birth month
  - etc

- Credit/debit card ownership and use

- Morality
  - Alcohol
  - homosexuality

- Demographics
  - Age
  - Sex
  - Education
  - Race
  - Income

Pew

World Bank

Census

Validity

Pew

World Bank

Census

8.33333%
## Validity

### % difference from benchmarks

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Brazil</th>
<th>China</th>
<th>India</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has a debit card</td>
<td>4.4%</td>
<td>45.3%</td>
<td>54.3%</td>
<td>275.6%</td>
<td>1.5%</td>
<td>7.4%</td>
</tr>
<tr>
<td>Homosexuality unacceptable</td>
<td>10.0%</td>
<td>41.0%</td>
<td>82.1%</td>
<td>22.4%</td>
<td>14.9%</td>
<td>24.5%</td>
</tr>
<tr>
<td>February</td>
<td>0.4%</td>
<td>3.9%</td>
<td>29.4%</td>
<td>26.4%</td>
<td>8.5%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>

- AU/UK/US better validity than BR/CH/IN
  - Pew benchmarks particularly off in BR/CH/IN

- Differences could be due to
  - Sampling/representativeness
  - Mode
  - Real changes over time (1 year lag)
Problems encountered

Reliability

• Different holidays/internet use patterns in each country
  • Could affect demographics
  • Could affect internet use questions
    • More surveys in December for end of year satisfaction or party planning?

Validity

• Personal characteristics
  • Validity benchmarks for things like eye color, height, etc. vary by country
    • Green eyes nearly nonexistent in China, India, much more common in Brazil
    • Often difficult to find hard statistics for comparison

• Have to treat the Pew/WB data as the gold standard, without a real rationale for that
  • When our numbers differ, don’t know which is true value
Discussion

• Reliability
  • Overall, SM has good reliability across all countries
  • Using the same panel source doesn’t provide equivalent reliability from country to country
  • Weighting data helps minimize the variation

• Validity
  • Hard to know whether differences are true differences or due to
    • Sampling differences
    • Mode differences

• What this means for others doing cross-national surveys
  • Can’t rely on panel to provide an identical sample
  • Treat each country individually
Thanks for listening!
Email us at:
lauraw@surveymonkey.com
mingnanl@surveymonkey.com