Lecture 7: Understanding human behaviour through survey design

DATA5207: Data analysis in the social sciences

Dr Shaun Ratcliff

The United States Studies Centre, University of Sydney

Why surveys?

Why surveys?

They are the source of much of our data.

They let us measure behaviours or attitudes that are otherwise hard or impossible to obtain.

Examples:

▶ Political preferences, attitudes, behaviours.

Why surveys?

They are the source of much of our data.

They let us measure behaviours or attitudes that are otherwise hard or impossible to obtain.

Examples:

- ▶ Political preferences, attitudes, behaviours.
- ▶ Health and well-being.

Why surveys?

They are the source of much of our data.

They let us measure behaviours or attitudes that are otherwise hard or impossible to obtain.

Examples:

- ▶ Political preferences, attitudes, behaviours.
- ▶ Health and well-being.
- ▶ Marketing (brand awareness, testing product attributes, willingness to pay).

Why surveys?

They are the source of much of our data.

They let us measure behaviours or attitudes that are otherwise hard or impossible to obtain.

Examples:

- ▶ Political preferences, attitudes, behaviours.
- ▶ Health and well-being.
- ▶ Marketing (brand awareness, testing product attributes, willingness to pay).
- ▶ Workplace engagement and satisfaction.

Survey modes

- In person.

Survey modes

- In person.
- Mail.

Survey modes

- In person.
- Mail.
- Phone (live, robo).

Survey modes

- In person.
- Mail.
- Phone (live, robo).
- Online.

Where do respondents come from?

Where do respondents come from?

What is the population of interest?

Political poll: electorates, states, country, men, women, young people, non-white voters?

Health and well-being: current patients/clients, all potential patients/clients?

Sampling

- Random sampling of respondents from population of interest guarantees survey results generalise validly to the population.

Sampling

- ▶ Random sampling of respondents from population of interest guarantees survey results generalise validly to the population.

- ▶ The normative ideal of public opinion research.

Sampling

- ▶ Random sampling of respondents from population of interest guarantees survey results generalise validly to the population.

- ▶ The normative ideal of public opinion research.

- ▶ But, potential problems: Can the population of interest be listed/enumerated and sampled from? Or is the population marginalised, hard-to-reach? Do all sampled persons respond? Is non-response ignorable?

Sampling

- ▶ Random sampling of respondents from population of interest guarantees survey results generalise validly to the population.

- ▶ The normative ideal of public opinion research.

- ▶ But, potential problems: Can the population of interest be listed/enumerated and sampled from? Or is the population marginalised, hard-to-reach? Do all sampled persons respond? Is non-response ignorable?

- ▶ Non-response: callbacks and follow-ups, incentives, offer mode of choice to respondent.

Types of response formats

- Categorical.

Types of response formats

- Categorical.
- Ordinal.

Types of response formats

- Categorical.
- Ordinal.
- Likert (a specific form of ordinal).

Types of response formats

- Categorical.
- Ordinal.
- Likert (a specific form of ordinal).
- Continuous.

What makes a good question?

Surveys

Typically have highly structured form:

▶ Questions and response formats identical for all respondents.

Surveys

Typically have highly structured form:

▶ Questions and response formats identical for all respondents.
▶ Typically respondents answer individually ("microdata" or "unit-record" data).

Surveys

Typically have highly structured form:

- Questions and response formats identical for all respondents.
- Typically respondents answer individually ("microdata" or "unit-record" data).
- In most academic settings, respondents answer anonymously.

Surveys

Typically have highly structured form:

- ▶ Questions and response formats identical for all respondents.
- ▶ Typically respondents answer individually ("microdata" or "unit-record" data).
- ▶ In most academic settings, respondents answer anonymously.
- ▶ Designed to help ensure valid and reliable measurement of phenomena of interest.

Surveys

Typically have highly structured form:

- ▶ Questions and response formats identical for all respondents.
- ▶ Typically respondents answer individually ("microdata" or "unit-record" data).
- ▶ In most academic settings, respondents answer anonymously.
- ▶ Designed to help ensure valid and reliable measurement of phenomena of interest.
- ▶ Standardised (consistent) question wording and responses.

Surveys

Typically have highly structured form:

- ▶ Questions and response formats identical for all respondents.
- ▶ Typically respondents answer individually ("microdata" or "unit-record" data).
- ▶ In most academic settings, respondents answer anonymously.
- ▶ Designed to help ensure valid and reliable measurement of phenomena of interest.
- ▶ Standardised (consistent) question wording and responses.
- ▶ Attempt to remove inter-subjective variation.

Surveys

Typically have highly structured form:

- ▶ Questions and response formats identical for all respondents.
- ▶ Typically respondents answer individually ("microdata" or "unit-record" data).
- ▶ In most academic settings, respondents answer anonymously.
- ▶ Designed to help ensure valid and reliable measurement of phenomena of interest.
- ▶ Standardised (consistent) question wording and responses.
- ▶ Attempt to remove inter-subjective variation.
- ▶ Obviously limited when employed cross-culturally, across languages.

Sources of respondent misreporting

► Social desirability bias: tendency to present a positive version of self.

Sources of respondent misreporting

▶ Social desirability bias: tendency to present a positive version of self.

▶ Complier effects: giving answers that R thinks will "please" or "satisfy" the interviewer/researchers.

Sources of respondent misreporting

▶ Social desirability bias: tendency to present a positive version of self.

▶ Complier effects: giving answers that R thinks will "please" or "satisfy" the interviewer/researchers.

▶ Acquiesence: tendency to "agree" rather than "disagree".

Sources of respondent misreporting

- Social desirability bias: tendency to present a positive version of self.

- Complier effects: giving answers that R thinks will "please" or "satisfy" the interviewer/researchers.

- Acquiesence: tendency to "agree" rather than "disagree".

- Recall: question asks about events long ago.

Sources of respondent misreporting

▶ Social desirability bias: tendency to present a positive version of self.

▶ Complier effects: giving answers that R thinks will "please" or "satisfy" the interviewer/researchers.

▶ Acquiesence: tendency to "agree" rather than "disagree".

▶ Recall: question asks about events long ago.

▶ Guessing: R is making up an opinion.

Sources of respondent misreporting

Social desirability is one of several demand characteristics. Weber and Cook have described some demand characteristics as involving the participant taking on a role in the experiment. These include:

► The good-participant role. Participant attempts to discern researcher aims and confirm them.

Sources of respondent misreporting

Social desirability is one of several demand characteristics. Weber and Cook have described some demand characteristics as involving the participant taking on a role in the experiment. These include:

▶ The good-participant role. Participant attempts to discern researcher aims and confirm them.

▶ The negative-participant role (the screw-you effect). Participant attempts to discern research aims and confound them.

Dealing with respondent misreporting

- **Question wording.**
- Anonymous surveys without human interviewer (mail or online, however, downsides).
- List and endorsement experiments.

Questionnaire design

- ▶ Much harder than you think! Don't be embarrassed to copy others!

Questionnaire design

- ▶ Much harder than you think! Don't be embarrassed to copy others!

- ▶ Balanced questions. Neutral tone. Don't bias response.

Unbalanced questions

The coal mine proposed by the Indian company Adani for the Galilee Basin in Queensland is dirty and will pollute all the rivers in the local area, which is why so many people in the community hate the mine and say that it will not provide jobs for locals and will ruin their health.

Do you believe the Queensland government should or should not approve the development of the mine?

1. Strongly support
2. Support
3. Neither support nor oppose
4. Oppose
5. Strongly oppose

A balanced way to ask this question

Some people say the Adani coal mine proposed for the
Galilee Basin in Queensland is bad for the environment
and will not provide much economic benefit. Others say
it will provide jobs for locals and provide tax revenues
to pay for public services.

Do you believe the Queensland government should or should
not approve the development of the mine?

1. Government should approve the mine
2. Government should not approve the mine
3. Not sure

Questionnaire design

- Much harder than you think! Don't be embarrassed to copy others!
- Balanced questions. Neutral tone. Don't bias response.
- Short and sweet: avoid double-barrelling.

Double-barrelled questions

```
Should the government spend less money on the
military and more on education?
```

```
1. Yes
2. No
3. Don't know
```

Another way to ask this question

Some say military spending by the Australian
government is too high, others that it is too low.

Do you think the government should spend less on the
military, spend more on the military, or keep spending
at current levels?

1. Spend much less on the military
2. Spend less on the military
3. Keep spending at current levels
4. Spend more on the military
5. Spend much more on the military
6. Don't know

Questionnaire design

- Much harder than you think! Don't be embarrassed to copy others!
- Balanced questions. Neutral tone. Don't bias response.
- Short and sweet: avoid double-barrelling.
- Understand your respondents. For public, avoid jargon, overly complex concepts.
- Many survey writers (eg academics) overshoot when writing for mass opinion surveys.

Jargon

You may have heard some people believe that increasing
tariffs will be good for the American economy, while
others say they will be bad for the American economy.

Do you believe that increased tariffs between the United
States and its trading partners will be:

1. Good for the American economy
2. Bad for the American economy
3. Neither good nor bad for the American economy

Another way to ask this question

Some people believe that increasing
tariffs (taxes on imports) will be good for the American
economy, while others say they will be bad for the
American economy.

Do you believe that increased tariffs between the United
States and its trading partners will be:

    1. Good for the American economy
    2. Bad for the American economy
    3. Neither good nor bad for the American economy

Questionnaire design

- ▶ Much harder than you think! Don't be embarrassed to copy others!
- ▶ Balanced questions. Neutral tone. Don't bias response.
- ▶ Short and sweet: avoid double-barrelling.
- ▶ Understand your respondents. For public, avoid jargon, overly complex concepts.
- ▶ Many survey writers (eg academics) overshoot when writing for mass opinion surveys.
- ▶ Response formats for 'pick one' must be intelligible, mutually exclusive and exhaustive.

Other considerations when using survey data

- Sample size.

Other considerations when using survey data

- Sample size.

- Sampling methodology (ie, mode: online, ftf, phone live or robo, mail).

Other considerations when using survey data

- Sample size.

- Sampling methodology (ie, mode: online, ftf, phone live or robo, mail).

- Non-response (and its bias), and solutions to this.

Other considerations when using survey data

- Sample size.

- Sampling methodology (ie, mode: online, ftf, phone live or robo, mail).

- Non-response (and its bias), and solutions to this.

- Causality and survey experiments (more on this next week).

The Census is also a kind of survey (sort of).

In the (virtual) lab.

Post your questions via the chat function. I will answer them once I have walked you through the lab exercise.

Assessable group work

Assessable group work

- Today, Wednesday Friday.
- Due after next day after class.
- Submit via canvas.

In this morning's lab

- I will give you a survey and corresponding data. The World Values Survey.

In this morning's lab

- I will give you a survey and corresponding data. The World Values Survey.

- You will use these data to identify predictors for quality of life.

In this morning's lab

▶ I will give you a survey and corresponding data. The World Values Survey.

▶ You will use these data to identify predictors for quality of life.

▶ We will create the dependent variable in the next class. This will be a measure of quality of life created with a combination of four questions from the survey using factor analysis.

In this morning's lab

- ▶ I will give you a survey and corresponding data. The World Values Survey.

- ▶ You will use these data to identify predictors for quality of life.

- ▶ We will create the dependent variable in the next class. This will be a measure of quality of life created with a combination of four questions from the survey using factor analysis.

- ▶ This morning, you will develop theories for what might predict higher/lower quality of life. Think about what drives life quality. Look at questionnaire and codebook. Identify possible predictors. Write up the descriptive analysis. Means, sd, or histograms, tables.

In this morning's lab

- ▶ I will give you a survey and corresponding data. The World Values Survey.

- ▶ You will use these data to identify predictors for quality of life.

- ▶ We will create the dependent variable in the next class. This will be a measure of quality of life created with a combination of four questions from the survey using factor analysis.

- ▶ This morning, you will develop theories for what might predict higher/lower quality of life. Think about what drives life quality. Look at questionnaire and codebook. Identify possible predictors. Write up the descriptive analysis. Means, sd, or histograms, tables.

- ▶ This afternoon we will create your dependent variable and you will test your ideas.

Second quiz

The second quiz

▶ Open book and 20 minutes long. You will be expected to provide short written answers. It will be run through canvas.

The second quiz

- ▶ Open book and 20 minutes long. You will be expected to provide short written answers. It will be run through canvas.

- ▶ Although open book, all questions must be in your own words. Any evidence of plagiarism will result in a zero grade.

The second quiz

- ▶ Open book and 20 minutes long. You will be expected to provide short written answers. It will be run through canvas.

- ▶ Although open book, all questions must be in your own words. Any evidence of plagiarism will result in a zero grade.

- ▶ I have not tried to set this test at too hard a level. We are not trying to trick or fail you. It is designed to tell you (and us) where you are at.

The second quiz

Some key things you might want to know:

- ▶ The test will consist of three questions randomly selected from a larger pool.

The second quiz

Some key things you might want to know:

▶ The test will consist of three questions randomly selected from a larger pool.

▶ It will focus on concepts from Week 1, and interpreting the output from regression models (you should know how to do this for linear and logistic regression).

The second quiz

Some key things you might want to know:

- ▶ The test will consist of three questions randomly selected from a larger pool.

- ▶ It will focus on concepts from Week 1, and interpreting the output from regression models (you should know how to do this for linear and logistic regression).

- ▶ There will be two conceptual questions, and you will be required to predict a probability from one logistic regression model.

The second quiz

Some key things you might want to know:

- ▶ The test will consist of three questions randomly selected from a larger pool.

- ▶ It will focus on concepts from Week 1, and interpreting the output from regression models (you should know how to do this for linear and logistic regression).

- ▶ There will be two conceptual questions, and you will be required to predict a probability from one logistic regression model.

- ▶ It is timed, you will have 20 minutes to complete it once you start. It will be accessed from the assessment page on canvas.

The second quiz

Some key things you might want to know:

- ▶ The test will consist of three questions randomly selected from a larger pool.

- ▶ It will focus on concepts from Week 1, and interpreting the output from regression models (you should know how to do this for linear and logistic regression).

- ▶ There will be two conceptual questions, and you will be required to predict a probability from one logistic regression model.

- ▶ It is timed, you will have 20 minutes to complete it once you start. It will be accessed from the assessment page on canvas.

- ▶ I will launch the test at 5pm today (Monday 1 February). It will close 5pm tomorrow (Tuesday, AEST). You can take the test any time within this 24 hour period.

Research project (Sunday, 14 February).

Questions about assessments?