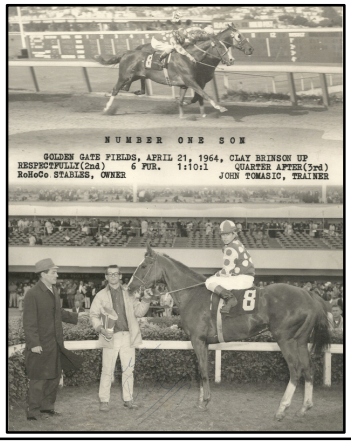
SURVEY DESIGN

POPULATIONS, SAMPLES, AND INSTRUMENTS



BACKGROUND

1. Sampling noninstitutionalized human populations

Many techniques have now been developed that enable us to draw unbiased samples of the noninstitutionalized population

2. The art of asking questions

Enough experience has now accrued in developing survey instruments that yield valid and reliable answers on a wide variety of topics

3. Multivariate data analysis

Data processing and computers, along with developments in statistical analysis now make it possible to feasibly calculate relationships between variables in highly complex models and analyses

TERMS

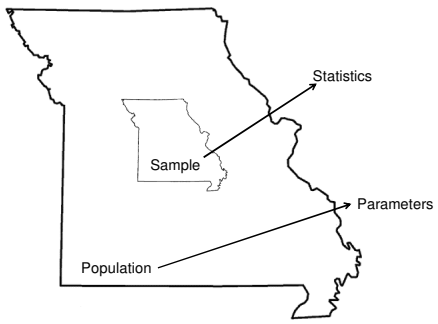
Element or Unit of Analysis
Universe
Population
Survey Population
Sampling Frame
Variable
Parameter
Statistic
Sampling Error

TERMS

Unit of Analysis

Individual persons
Groups
Organizations
Communities
Societies
Things

L%B LRS	H%B HRS	H%B HRS
L%B LRS	H%B HRS	H%B HRS
L%B LRS	L%B LRS	L%B LRS
L%B LRS	H%B HRS	H%B HRS
L%B LRS	L%B LRS	H%B HRS



TYPES OF SAMPLES

Nonprobability samples

- Convenience samples
- Judgemental samples
- Quota samples
- Expert samples
- Purposive samples
- Available data samples

Probability samples

- Simple random samples
- Systematic samples
- Stratified samples
- Cluster samples

SAMPLE SIZE

SAMPLE SIZE ESTIMATION

We use the confidence interval formula to estimate sample size requirements.

$$c.i. = \bar{x} \pm z * \sigma_{\bar{x}}$$

Decision process involves:

- Type of population
- Nature of the resulting variable
- Variability of the resulting variable
- Amount of precision required

SAMPLE SIZE ESTIMATION

	INFINITE	FINITE
MEAN	$n = \frac{z^2 * \sigma^2}{E^2}$	$n = \frac{z^2 * N * \sigma^2}{(z^2 * \sigma^2) + (N * E^2)}$
PROPORTION	$n = \frac{z^2 * \pi * (1-\pi)}{E^2}$	$n = \frac{z^2 * N * \pi * (1-\pi)}{(z^2 * \pi * (1-\pi)) + (N * E^2)}$

1) While you can use any *z* you wish, convention is *z* at **α**=.05 is 1.96 which is rounded to 2
2) Must have estimate of variance → previous studies,
   pilot study,
   R/6 or R/4
   Proportion 50-50 split
3) E = precision - decision based on consequences.
4) N is population size and n is sample size.

!**Always round up**!

SAMPLE SIZE ESTIMATION

Q?: What is the likelihood that a property tax issue be passed by voters?

Solution: take a sample survey poll
Margin of error 3% for an infinite population

$$n = \frac{Z^2 * \pi * (1-\pi)}{E^2} = \frac{4 * .5 * .5}{.0009} = \frac{1}{.0009} = 1111.1 = 1112$$

take sample and find 588 supporters

$$588/1112 = 52.9\%$$

What is the 95% confidence level for these results?

$$c.i. = .529 \pm z * \sqrt{\frac{P_s * (1-P_s)}{n}}$$
$$= .529 \pm 1.96 * \sqrt{\frac{.529 * .471}{1112}}$$
$$= .529 \pm 1.96 * \sqrt{\frac{.249159}{1112}}$$
$$= .529 \pm 1.96 * \sqrt{.00022406}$$
$$= .529 \pm 1.96 * .014968762$$
$$= .529 \pm .029338774$$
$$c.i. = .529 \pm .03$$
$$.499 < \pi < .559$$
$$49.9\% < \pi < 55.9\%$$

SAMPLE SIZE ESTIMATION

Precision (Interval Width)	Approximate Sample Size
±10%	100
±7%	200
±5%	400
±3%	1000
±2%	2400
±1%	9600

ARE WE TESTING WHAT WE THINK WE'RE TESTING?

"I'm *not* being a tattle-tale! — I'm being a reliable source!"

Validity

INTERNAL AND EXTERNAL VALIDITY

Internal validity of a research study is the degree to which the study accurately answers the question it was intended to answer.

External validity refers to the extent to which we can generalize the results of a research.

CAN WE GET THE SAME RESULTS?

Reliability

INDICES & SCALES

COMPOSITE MEASURES

Single indicator problems

Lower levels of measurement

Efficiency

INDEX VS SCALES

Used interchangeably – wrong

Both use an ordinal or lower level of measurement

Both rank order subjects

The distinction is that an index is an accumulation of scores while a scale accumulates and looks at the pattern of scores

INDEX CONSTRUCTION

Items must have face validity

The index must be unidimensional

Be aware of the variance of the items

Must be aware of the bivariate relationship among the items

Must be aware of the multivariate relationship among the items – Chronbach's alpha in spss

INDEX SCORING

An index, like a Likert, is simply the addition of values – adding up

Suppose we want to measure "alienation" as a concept – what are the choices?

Question:

We are all just cogs in the machinery of life.

1 Strongly agree

2 Agree

3 Neutral

4 Disagree

5 Strongly disagree

Would this one question measure alienation?

Could have 10 questions that measure facets of alienation

Thus the range of scores would be from 10 to 50 – which is a much better set of scores

Problem is the middle of the distribution

Person A answers 1111155555 – score =30

Person B answers 5555511111 – score =30

muddy middle!

EXAMPLE
ROSENBERG SELF-ESTEEM SCALE

Description of Measure: A 10-item scale that measures global self-worth by measuring both positive and negative feelings about the self. The scale is believed to be uni-dimensional. All items are answered using a 4-point Likert scale format ranging from strongly agree to strongly disagree.

Typically, such composite measures have been normed with large sample sizes. Always check for reliability results when you use such pre-designed measures. If you have created the measure, then you must check the reliability yourself with something like Chronbach's alpha.

Items 2, 5, 6, 8, 9 are reverse scored. Sum scores for all ten items. Keep scores on a continuous scale. Higher scores indicate higher self-esteem.

QUESTIONS

Instructions: Below is a list of statements dealing with your general feelings about yourself. Please indicate how strongly you agree or disagree with each statement.

1. On the whole, I am satisfied with myself.

Strongly Agree ❹ Agree ❸ Disagree ❷ Strongly Disagree ❶

2. At times I think I am no good at all.

Strongly Agree ❶ Agree ❷ Disagree ❸ Strongly Disagree ❹

3. I feel that I have a number of good qualities.

Strongly Agree ❹ Agree ❸ Disagree ❷ Strongly Disagree ❶

4. I am able to do things as well as most other people.

Strongly Agree ❹ Agree ❸ Disagree ❷ Strongly Disagree ❶

5. I feel I do not have much to be proud of.

Strongly Agree ❶ Agree ❷ Disagree ❸ Strongly Disagree ❹

6. I certainly feel useless at times.

Strongly Agree ❶ Agree ❷ Disagree ❸ Strongly Disagree ❹

7. I feel that I'm a person of worth, at least on an equal plane with others.

Strongly Agree ❹ Agree ❸ Disagree ❷ Strongly Disagree ❶

8. I wish I could have more respect for myself.

Strongly Agree ❶ Agree ❷ Disagree ❸ Strongly Disagree ❹

9. All in all, I am inclined to feel that I am a failure.

Strongly Agree ❶ Agree ❷ Disagree ❸ Strongly Disagree ❹

10. I take a positive attitude toward myself.

Strongly Agree ❹ Agree ❸ Disagree ❷ Strongly Disagree ❶

To the computer

SCALING

Scaling gets around this issue by looking at the pattern of responses

It also looks at the accumulation of scores

Both indices and scales move up from ordinal level of measurement but the scale is a little higher

GUTTMAN SCALE EXAMPLE

The example uses only a Guttman scale

There are many multiple dimension scaling procedures in spss and other stat packages.

Conceptual dimension: Social Distance – The degree of separation maintained by a person in regards to other races and/or ethnic groups.

Indicators: Questions regarding behavior of a person towards persons of other races and/or ethnic groups.

CONCEPT – SOCIAL DISTANCE

- Q1 – Would you invite a person of another race to your parent's home for dinner?
- Q2 – Would you talk with a person of another race?
- Q3 – Would you sit next to a person of another race on a bus?
- Q4 – Would you be willing to have a person of another race for a roommate?
- Q5 – Would you marry a person of another race?

All questions are answered as: 0=NO and 1=YES

Now we ask 10 subjects these questions on a survey.

Data Matrix

Subject	Q1	Q2	Q3	Q4	Q5
A	1	1	1	0	0
B	0	1	1	1	1
C	0	0	0	0	0
D	0	0	1	0	0
E	1	1	1	1	1
F	0	1	1	1	0
G	0	1	0	0	0
H	1	1	1	1	0
I	0	1	1	0	0
J	1	1	0	0	0

Step 1: Sum frequency of the 1's for each row and for each column.

Subject	Q1	Q2	Q3	Q4	Q5	freq 1's
A	1	1	1	0	0	3
B	0	1	1	1	1	4
C	0	0	0	0	0	0
D	0	0	1	0	0	1
E	1	1	1	1	1	5
F	0	1	1	1	0	3
G	0	1	0	0	0	1
H	1	1	1	1	0	4
I	0	1	1	0	0	2
J	1	1	0	0	0	2
freq 1's	4	8	7	4	2	

Step 2: Rearrange matrix in descending order of the frequency of ones. Rows top to bottom and columns left to right.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 3: Build a stair step diagonal.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's
E	1	1	1	1	1	5
B	1	1	1	0	1	4
H	1	1	1	1	0	4
A	1	1	0	1	0	3
F	1	1	1	0	0	3
I	1	1	0	0	0	2
J	1	0	0	1	0	2
D	0	1	0	0	0	1
G	1	0	0	0	0	1
C	0	0	0	0	0	0
freq 1's	8	7	4	4	2	

Step 4: Count errors

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's	errors
E	1	1	1	1	1	5	0
B	1	1	1	0	1	4	2
H	1	1	1	1	0	4	0
A	1	1	0	1	0	3	2
F	1	1	1	0	0	3	0
I	1	1	0	0	0	2	0
J	1	0	0	1	0	2	2
D	0	1	0	0	0	1	2
G	1	0	0	0	0	1	0
C	0	0	0	0	0	0	0
freq 1's	8	7	4	4	2		
errors	1	2	1	3	1		8

Step 5: Estimate moves to reduce the number of errors.

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's	errors
E	1	1	1	1	1	5	0
B	1	1	1	0	1	4	2
H	1	1	1	1	0	4	0
F	1	1	1	0	0	3	0
A	1	1	0	1	0	3	2
I	1	1	0	0	0	2	0
J	1	0	0	1	0	2	2
D	0	1	0	0	0	1	2
G	1	0	0	0	0	1	0
C	0	0	0	0	0	0	0
freq 1's	8	7	4	4	2		
errors	1	2	1	3	1		8

- The normal process would be to iterate between rows and columns until a final solution is achieved (that is, no more error reduction.
- There are two ways this must be done:
  - ✓ Review rows – columns – rows – etc until a solution is reached
  - ✓ Review columns – rows – columns – etc until a solution is reached

In this example, the moves we did is the best we can do.

Step 6: Final Solution

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's	errors
E	1	1	1	1	1	5	0
B	1	1	1	0	1	4	1
H	1	1	1	1	0	4	0
F	1	1	1	0	0	3	0
A	1	1	0	1	0	3	1
I	1	1	0	0	0	2	0
D	0	1	0	0	0	1	1
J	1	0	0	1	0	2	1
G	1	0	0	0	0	1	0
C	0	0	0	0	0	0	0
freq 1's	8	7	4	4	2		
errors	1	0	0	3	0		4

Step 7: Determine scale steps for the final solution

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's	errors	Row modal	Scale Step
E	1	1	1	1	1	5	0	5	5
B	1	1	1	0	1	4	1	4	5
H	1	1	1	1	0	4	0	4	4
F	1	1	1	0	0	3	0	3	3
A	1	1	0	1	0	3	1	3	2
I	1	1	0	0	0	2	0	3	2
D	0	1	0	0	0	1	1	4	2
J	1	0	0	1	0	2	1	3	1
G	1	0	0	0	0	1	0	4	1
C	0	0	0	0	0	0	0	5	0
								38	
freq 1's	8	7	4	4	2				
errors	1	0	0	3	0		4		
Col Modal	8	7	6	6	8	=35			

STATISTICAL DETERMINATION

$$\text{Coef of Reproducibility} = 1 - \left(\frac{\#\text{ errors}}{\#\text{ rows} * \#\text{ columns}}\right) = 1 - \left(\frac{4}{10 * 5}\right) = .92$$

$$\text{Coef of Scalability} = 1 - \left(\frac{\#\text{ errors}}{(\#\text{rows} * \#\text{ col}) - \text{max modal}}\right) = 1 - \left(\frac{4}{(10 * 5) - 38}\right) = .67$$

$$\text{Min Marg Reproducibility} = \left(\frac{\text{max modal}}{\#\text{rows} * \#\text{col}}\right) = \left(\frac{38}{10 * 5}\right) = .76$$

$$\%\text{ Improvement} = \text{CR - MMR} = .92 - .76 = .16 \text{ or } 16\% \text{ improvement over chance}$$

CR must be .9 or better and CS must be .6 or better in order to say one has a scale. In addition, it is desirable to have a percent improvement of at least 20% but this requirement is not as rigid.

Step 7: Determine scale steps for the final solution

Subject	Q2	Q3	Q4	Q1	Q5	freq 1's	errors	Row modal	Scale Step
E	1	1	1	1	1	5	0	5	5
B	1	1	1	0	1	4	1	4	5
H	1	1	1	1	0	4	0	4	4
F	1	1	1	0	0	3	0	3	3
A	1	1	0	1	0	3	1	3	2
I	1	1	0	0	0	2	0	3	2
D	0	1	0	0	0	1	1	4	2
J	1	0	0	1	0	2	1	3	1
G	1	0	0	0	0	1	0	4	1
C	0	0	0	0	0	0	0	5	0
								38	
freq 1's	8	7	4	4	2				
errors	1	0	0	3	0		4		
Col Modal	8	7	6	6	8	=35			

CONCEPT – SOCIAL DISTANCE

- Q1 – Would you invite a person of another race to your parent's home for dinner?
- Q2 – Would you talk with a person of another race?
- Q3 – Would you sit next to a person of another race on a bus?
- Q4 – Would you be willing to have a person of another race for a roommate?
- Q5 – Would you marry a person of another race?

All questions are answered as: 0=NO and 1=YES

Now we ask 10 subjects these questions on a survey.

QUESTIONNAIRE DESIGN

QUESTIONNAIRE DESIGN AND CONSTRUCTION

Phrasing of questions is critical

avoid vague, nebulous questions

- Bad example: "What do you think about abortion?"
- Bad example: "How many children do you have?"

questions must be clear, unambiguous

- Good example: "Which of the following statements best represents your attitude toward elective abortions in the first trimester?"
- Conduct a "readability analysis"

avoid lengthy questions; keep questions short, succinct

- Bad example: Hilary Clinton would be the first female President of the United States. Do you think her gender would be a problem in negotiating with the Arab world, which is a male-dominated culture?
- Good example: Would Barack Obama make a good Supreme Court justice?

QUESTIONNAIRE DESIGN AND CONSTRUCTION--CONTINUED

avoid double-barreled questions (compound questions)

- Bad example: "Do you favor stricter hand gun controls and mandatory minimum sentences for carjackers?"

avoid loaded language (push polling)

- Bad example: "Don't you think that...?" "Isn't it true that...?"
- Bad example: Emotionally charged words: " gang member," "welfare mother," "extremist groups," "spin doctor," etc.

avoid slang, jargon, abbreviations and acronyms

- Bad example: "Should states regulate PETA and the ALF?"
- Bad example: "Do you think hip hop is wack?"

avoid or minimize negative wording

- Bad example: Is the Obama administration right in not establishing a firm deadline for not forcing Russia to not bomb in Syria and Turkey?

OPEN-ENDED VERSUS CLOSE-ENDED QUESTIONS

Schuman, Ludwig, & Krosnick (1986): 60% of respondents selected one of four options in a close-ended format, but only 2.4% mentioned any of the same four responses in an open-ended format.

- **Open-ended**: allows subjects more leeway, flexibility
  "What is your primary ethnic/cultural background?"
- Requires a **content analysis** of responses

- **Close-ended** or "forced-choice" ties respondents' hands somewhat
- Easy to code the data
- Always include an "other_____" category

PHRASING OF QUESTIONS IS CRITICAL

Even slight variations in wording can alter respondents' answers.

- "occupied territory" versus "contested territory"
- "assisted suicide" versus "mercy killing"
- "fetus" versus "unborn child"

Make questions concrete, come down the "ladder of abstraction"

Use negatively worded questions or statements sparingly

- sometimes necessary to include reverse-worded items to identify a "response set"

...MORE ON PHRASING QUESTIONS

Avoid evaluative language

- Phrasing of questions should not imply approval, disapproval
- Follow-ups should not suggest surprise, liking, disliking, etc.

  Bad example: Do you think the Democratic dominated Congress should lift the harsh restrictions on stem cell research?

  Bad example: Despite its poor track record in crisis intervention, do you think the United Nations should intervene in Darfur, Sudan?

Probably stop here?

DILLMAN APPROACH TO SURVEYS

Total Design Method (1978)

What aspects of survey process affect quality or quantity of response

- A detailed and systematic way to design and implement surveys
- Social exchange theory
- High quality mail surveys
- Acceptable response rates

The Tailored Design Method (2000)

Many changes since 1978

- Technology – computers, scanners…
- More complete understanding of social exchange principles in survey research
- Research on survey methods
- Need to tailor the method vs same procedure

Note: Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method 3rd ed. 2009

SOCIAL EXCHANGE THEORY

Theory of human behaviour

- "The theory asserts that actions of individuals are motivated by the return these actions are expected to bring, and in fact usually do bring from others" (Dillman, 2000)
- People will do things that are rewarding

- **Three elements for predicting action**
  - **Rewards** -- What one expects to gain from an activity
  - **Costs** -- What one gives up/spends to obtain rewards
  - **Trust** -- Expectation that in the long run, rewards of doing something will outweigh the costs of doing it

APPLICATION OF SOCIAL EXCHANGE THEORY

In terms of questionnaire design & implementation:

- Increase rewards for responding
- Reduce perceived costs
- Establish trust so the rewards outweigh costs of responding

WAYS OF PROVIDING REWARDS

Show positive regard

- Show respect (e.g. give reason for survey, provide number to call, personally address letters)

Say thank you

- e.g. "we appreciate your help", "thanks in advance"

Ask for help, advice, assistance

- Provides a sense of reward

Support group values

- appeal to widely shared values – the study's social usefulness

WAYS OF PROVIDING REWARDS

Give tangible rewards

- Include a financial incentive (token) vs promise

Make the questionnaire interesting

- Improve layout and design

Give social validation

- Let them know that others have responded

Opportunities to respond are scarce

- Respond quickly so information can be used

WAYS OF REDUCING SOCIAL COSTS

Avoid subordinating language

- Ask for a favour e.g. to solve a problem vs it is necessary for you to complete this survey because….

Avoid embarrassment

- E.g. avoid complex questions and instructions at the start of the questionniare

Avoid inconvenience

- E.g. include return postage paid envelope – real stamp

WAYS OF REDUCING SOCIAL COSTS

Make questionnaire appear short and easy

- Lesson the perceived cost of responding

Minimize requests to obtain personal information

- Offer explanations why info is important & ensure confidentiality

WAYS OF ESTABLISHING TRUST

Provide a token of appreciation <u>in advance</u>

- Small amount – gesture of trust – vs payment for your time
- Including actual stamps on *return* envelope

Sponsorship by legitimate authority

- e.g. university or gov't sponsored vs marketing research)

Make the task appear important

- Personalized cover letters on letterhead paper
- Professional looking questionnaire
- Something useful will be done with results

DILLMAN'S FIVE NEEDED ELEMENTS FOR HIGH RESPONSE RATES

1. Respondent friendly questionnaire
2. Up to 5 (varied) contacts
3. Inclusion of stamped return envelopes
4. Personalized correspondence
5. Token financial incentive

ONLINE VIDEOS HOW TO DO ONLINE SURVEYS

http://info.userzoom.com/online-surveys-design-webinar.html

81 minutes

https://www.youtube.com/watch?v=TFFkNGS0tfA

40 minutes

SECONDARY ANALYSIS

What is Secondary Data Analysis?

- "In the broadest sense, analysis of data collected by someone else" (p. ix; Boslaugh, 2007)
- Analysis of secondary data, where "secondary data can include any data that are examined to answer a research question other than the question(s) for which the data were initially collected" (p. 3; Vartanian, 2010)
- In contrast to primary data analysis in which the same individual/team of researchers designs, collects, and analyzes the data

Where Can I Find Secondary Data?

Searching for secondary datasets:

- Inter-University Consortium for Political and Social Research
  - http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp
- Data.gov
  - http://www.data.gov
- National Center for Education Statistics
  - http://nces.ed.gov
- U.S. Census Bureau
  - http://www.census.gov
- Simple Online Data Archive for Population Studies (SodaPop)
  - http://sodapop.pop.psu.edu/data-collections

Examples of Large Secondary Datasets for Education & Social Sciences Research

- Common Core of Data (CCD)
- Current Population Survey (CPS)
- Early Childhood Longitudinal Study (ECLS): Birth (ECLS-B) and Kindergarten (ECLS-K) Cohort
- General Social Survey (GSS)
- Head Start Family and Child Experiences Survey (FACES)
- Monitoring the Future (MTF)
- National Assessment of Educational Progress (NAEP)
- National Education Longitudinal Study (NELS)
- National Household Education Surveys (NHES)
- National Longitudinal Study of Adolescent Health (Add Health)
- National Longitudinal Survey of Youth (NLSY)
- National Survey of American Families (NSAF)
- National Survey of Child and Adolescent Well-Being (NSCAW)
- National Survey of Families and Households (NSFH)
- NICHD Study of Early Child Care and Youth Development (SECCYD)
- Programme for International Student Assessment (PISA)
- Progress in International Reading Literacy Study (PIRLS)
- Trends in International Mathematics and Science Study (TIMSS)
- U.S. Panel Study of Income Dynamics (PSID): Child Development Supplement (CDS)

Advantages of Secondary Data Analysis

- Study design and data collection already completed
  - Saves time and money
    - Access to international and cross-historical data that would otherwise take several years and millions of dollars to collect
    - Ideal for use in classroom examples, semester projects, masters theses, dissertations, supplemental studies
- Data may be of higher quality
  - Studies funded by the government generally involve larger samples that are more representative of the target population (greater external validity!)
  - Oversampling of low prevalence groups/behaviors allows for increased statistical precision
- Datasets often contain considerable breadth (thousands of variables)

Disadvantages of Secondary Data Analysis

- Study design and data collection already completed
  - Data may not facilitate particular research question
  - Information regarding study design and data collection procedures may be scarce
- Data may *potentially* lack depth (the greater the breadth the harder it is to measure any one construct in depth)
  - Constructs may be operationally defined by a single survey item or a subset of test items which can lead to reliability and validity concerns
  - 'Post hoc' attempts to construct measurement models may be unsuccessful (survey items may not hang together)
- Certain fields or departments (e.g., experimental programs) may place less value on secondary data analysis
- May require knowledge of survey statistics/methods which is not generally provided by basic graduate statistics courses