Surveys Overview

Empirical Software Engineering

IME-USP

October, 17th, 2011

Then, What Is a Survey?

- Today the word "survey" is used most often to describe a method of gathering information from a sample of individuals. This "sample" is usually just a fraction of the population being studied.

- For example:

  - a sample of voters is questioned in advance of an election to determine how the public perceives the candidates and the issues ...

  - a manufacturer does a survey of the potential market before introducing a new product ...

- Give some examples

Then, What Is a Survey?...

- Not only do surveys have a wide variety of purposes, they also can be conducted in many ways
  - including over the telephone, by mail, or in person.
  - Nonetheless, all surveys do have certain characteristics in common.

- Unlike a census, where all members of the population are studied, surveys gather information from only a **portion** of a population of interest
  - the size of the sample depending on the purpose of the study.

Then, What Is a Survey?...

- In a bona fide survey, the sample is not selected haphazardly or only from persons who volunteer to participate.

- It is scientifically chosen so that each person in the population will have a measurable chance of selection.

  - This way, the results can be reliably projected from the sample to the larger population.

- Information is collected by means of standardized procedures so that every individual is asked the same questions in more or less the same way.

  - The survey's intent is not to describe the particular individuals who, by chance, are part of the sample, but to obtain a composite profile of the population.

Then, What Is a Survey?...

- Anonymity:
  - The industry standard for all reputable survey organizations is that individual respondents should never be identified in reporting survey findings.
  - All of the survey's results should be presented in completely anonymous summaries, such as statistical tables and charts.

Questions asked with Surveys

- You can further classify surveys by their content. Some surveys focus on opinions and attitudes (such as a pre-election survey of voters – or poll)

- Others are concerned with factual characteristics or behaviours (such as people's health, housing, consumer spending, or transportation habits).

- Many surveys combine questions of both types. Respondents may be asked if they have:
  - heard or read about an issue ...
  - what they know about it ...
  - their opinion ...
  - how strongly they feel and why...
  - their interest in the issue ...
  - past experience with it ...
  - and certain classification information (such as age, gender, marital status, occupation, and place of residence).
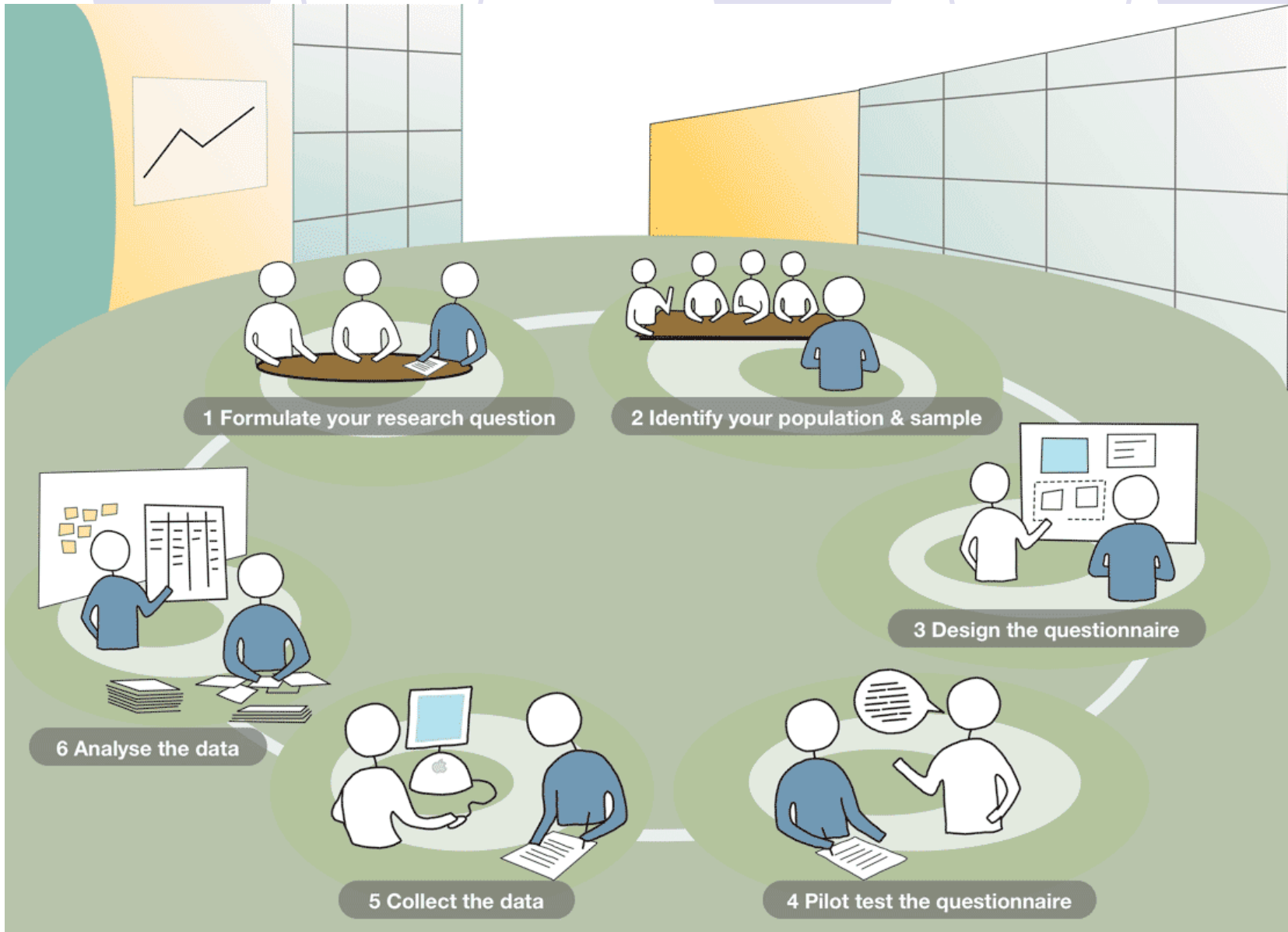
Questions asked with Surveys…

- Questions may be open-ended ("Why do you feel that way?") or closed ("Do you approve or disapprove?").

- The manner in which a question is asked can greatly affect the results of a survey. For example, a recent NBC/Wall Street Journal poll asked two very similar questions with very different results:
  - (1) Do you favour cutting programs such as social security, Medicare, Medicaid, and farm subsidies to reduce the budget deficit?
    - The results: 23% favour; 66% oppose; 11% no opinion.
  - (2) Do you favour cutting government entitlements to reduce the budget deficit?
    - The results: 61% favour; 25% oppose; 14% no opinion.

Questions asked with Surveys…

- The questionnaire may be very brief -- a few questions, taking five minutes or less -- or it can be quite long -- requiring an hour or more of the respondent's time.

- Because changes in attitudes or behaviour cannot be reliably ascertained from a single interview, some surveys employ a "panel design," in which the same respondents are interviewed on two or more occasions.
  - Cross sectional
  - Longitudinal

The Survey Process

1 Formulate your research question

2 Identify your population & sample

3 Design the questionnaire

4 Pilot test the questionnaire

5 Collect the data

6 Analyse the data

The Survey Process

- Many stages in the survey process
  - Goal-setting; survey design; sampling; data collection, capture, cleaning, and analysis; reporting survey results; and decision-making based on these results

- Different people (and often different organisations) are involved at different stages

- Good communication is therefore important
  - E.g. Don't want to be recoding data after you have received it

- Statistics is useful at many points during a survey
  - Especially sampling, weighting, and data analysis

Goal-Setting

- Vital to define research question or survey objectives

Goal-Setting

- Level of formality usually depends on:
  - Whether survey sponsor and researcher are separate people or organisations
  - Whether the survey process must be transparent
    - Usually the case when there are many stakeholders with opposing interests; e.g. media or social/govt research
  - Importance of the survey

Common types of objective

- To **evaluate the rate or frequency of some characteristic** that occurs in a population, for example, we might be interested in the frequency of failing projects (Standish Group, 2003).

- To **assess the severity of some characteristic or condition** that occurs in a population, for example, we might be interested in the average overrun of software projects (Moløkken-Østvold et al., 2004).

- To **identify factors that influence a characteristic or condition**, for example, we might be interested in factors that predispose a process improvement activity towards failure or towards success Dybå (2005).

Kitchenham and Pfleeger, 2008. Personal Opinion Surveys

Common types of objective

- The first two types of survey objective are descriptive: they describe some condition or factor found in a population in terms of its frequency and impact.

- The second type of survey looks at the relationship existing among factors and conditions within a population

Kitchenham and Pfleeger, 2008. Personal Opinion Surveys

Why do we do surveys?

- To describe the populations: What is going on?

- Theoretical reasons: Why is it going on?
  - Develop and test theory
  - Theory should always guide survey development and data collection

Survey Design

- Survey design involves
  - Choice of data collection methodology
    - Usually face-to-face in-home or central location, telephone, mail, e-mail, WWW pop-ups, or some combination of these
  - Questionnaire design
  - Sample design & analysis planning
  - Estimating survey costs
- Aim to achieve survey objectives cost-effectively
  - Usually involves trade-offs between cost, speed and accuracy
  - E.g. in-home face-to-face vs telephone interviewing

Sample Design and Selection

- Necessary unless survey is a census
- Many design options
  - E.g. stratification, clustering sampling
- Has major implications for data analysis (and sometimes for other stages)

Survey sampling

The Need For Sampling

- Large population can be cost prohibitive
- Using a subset reduces the cost
- Exercise caution to select representatives
  - Otherwise, invalidates reliability
- Easier follow-ups with fewer people
  - Follow-ups can be critical

Sampling Methods

Several general methods exist for selecting samples:

- Probability (or random) sampling is where each object has a known, non-zero probability of being selected
  - Can produce unbiased results (if no non-response)
  - Allows for calculation of sampling error (if pairwise selection probabilities known)
  - Most widely accepted sampling method. Strongest acceptance in USA.
- Judgment sampling involves choosing objects that it is believed will give accurate results
  - E.g. three areas (one large city, one town, one rural)

Sampling Methods (continued)

- Quota samples are based on selecting objects until you have a certain number (the quota) of each type
  - Appeals to idea of a "representative" sample
  - Can produce substantial bias (e.g. 1992 UK election polls)
  - Still widely used (especially for telephone surveys with high non-response levels)
- Convenience samples are obtained by choosing the easiest objects available
 - E.g. the first ten people to walk out of a store

Sampling Terminology

- ## Units are the objects to be surveyed
  - Usually people, households, businesses (enterprises or geographic units), customers or activities (e.g. trips, phone calls, nights stayed)
- ## Survey population
  - The collection of units that the survey results should describe or explain
- ## Sample
  - The subset of the population for which survey data is collected (or is intended to be collected)
- ## Sampling frame
  - a method of contacting selected sample units, including the information needed to select them
    - E.g. a list of customers, including the value of their business
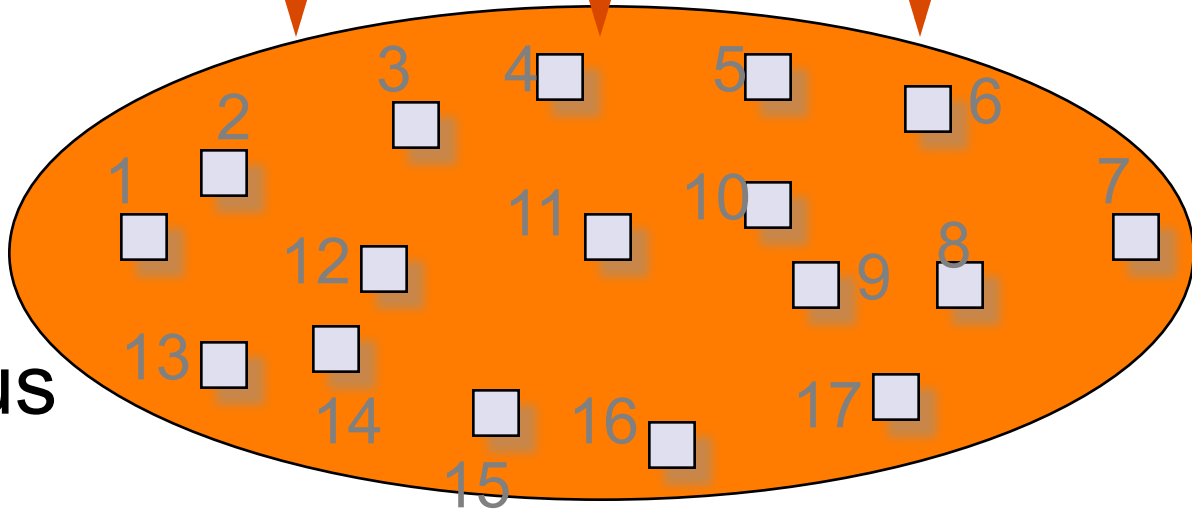- ## Need precise definitions of these for each survey

Benefits of sample surveys as a research method

- Benefits of sample surveys as a research method:

    - Relatively cheap and fast

    - Valuable for finding out how people <u>think</u> or how they <u>behave</u>
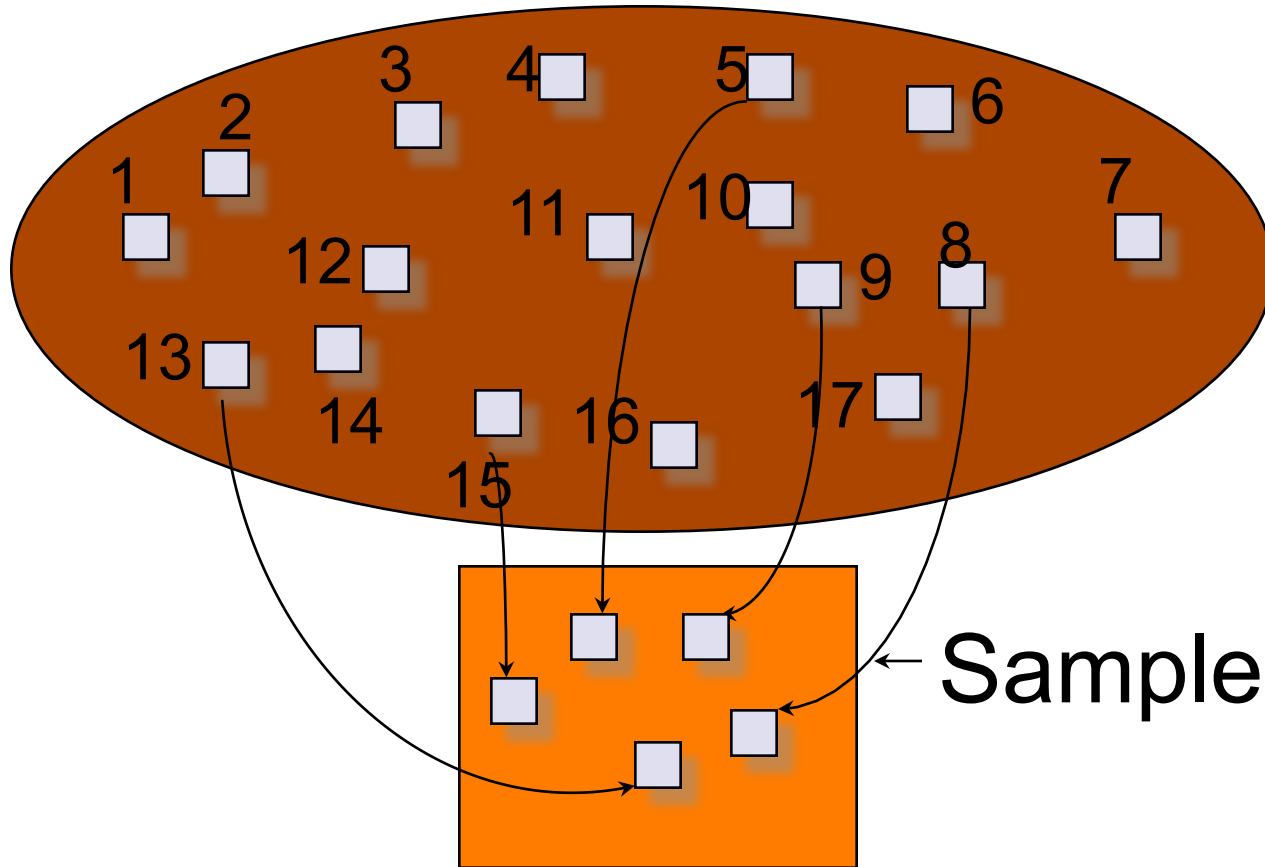
Census
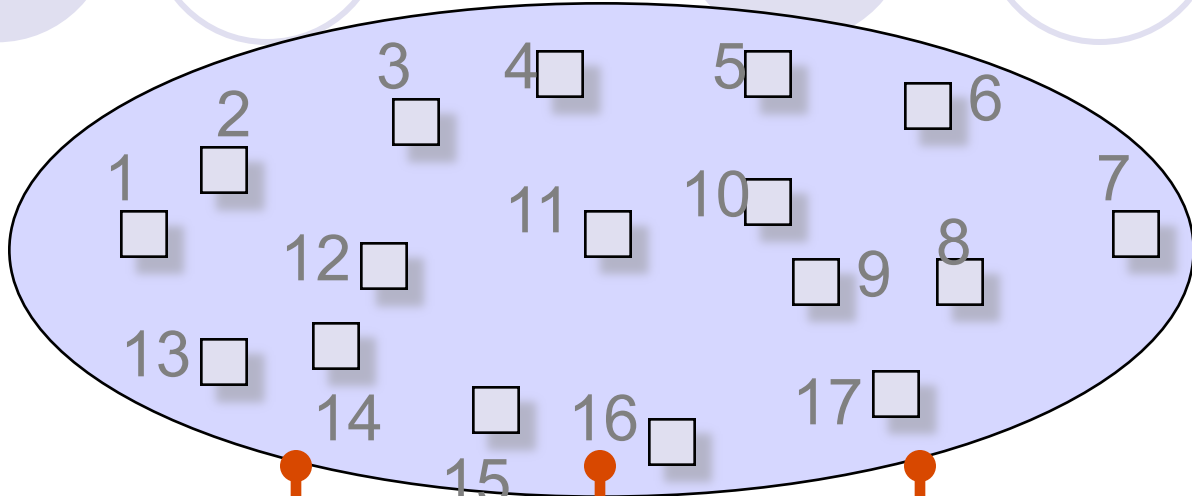
Census

List of Units

1 ✔
2 ✔
3 ✔
4 ✔
5 ✔
6 ✔
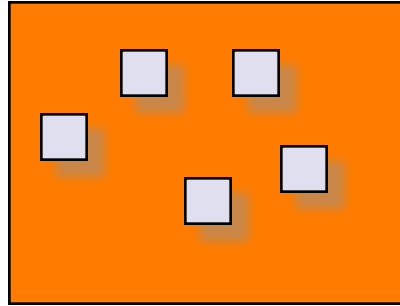7 ✔
8 ✔
9 ✔
10 ✔
11 ✔
12 ✔
13 ✔
14 ✔
15 ✔
16 ✔
17 ✔

Sampling Frame

List of Units

1
2
3
4
5 ✔
6
7 ✔
8 ✔
9 ✔
10
11
12
13 ✔
14
15 ✔
16
17

Sample

Sample Survey

measurements
data

Telephone Sampling of Households

- Undercoverage is a fundamental problem for telephone surveys of households
  - Households without phones are also different in other ways; e.g. they are generally low-income households
- Duplicates also occur
  - i.e. some households have more than one phone number, and thus have more chance of being selected
  - Should ideally correct for this by sub-sampling or weighting the data, or by removing all duplicates from the frame before sample selection

Telephone Sampling Frames

- White Pages
  - Telecom sells random samples of listed numbers
    - Unlisted numbers not included
      - So have lost another 15% of phone numbers
  - May be cheaper to use paper directories instead, but these are out of date (even when just distributed)

Sampling Frames Summary

- Several possible frames exist for household, individual and business surveys
- However many have severe flaws
  - Substantial under coverage
  - Duplicates present in most frames
    - De-duplication of entire frame, subsampling or reweighting needed
    - Note that it is not enough to select a sample and then remove duplicates within the sample
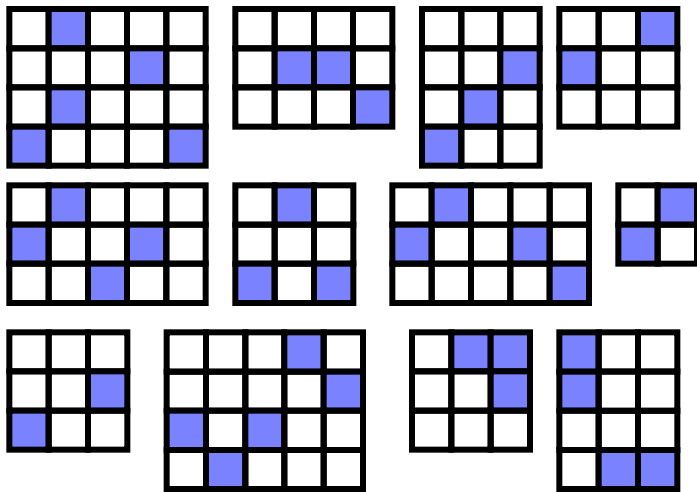- Caution is advisable

Probabilistic Sample Methods

- Known, non-zero chance for inclusion
- Simple random: Equal probability
- Stratified random: Separate into groups
- Systematic: Select every X member
- Cluster-based: Belong to specific group

Difference Between Cluster and Stratified Sampling
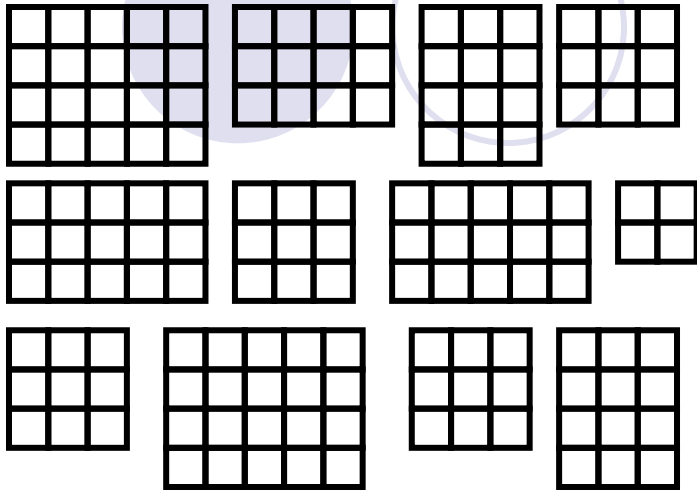
Population of L strata, stratum l contains n_l units

Population of C clusters

Take simple random sample in *every* stratum
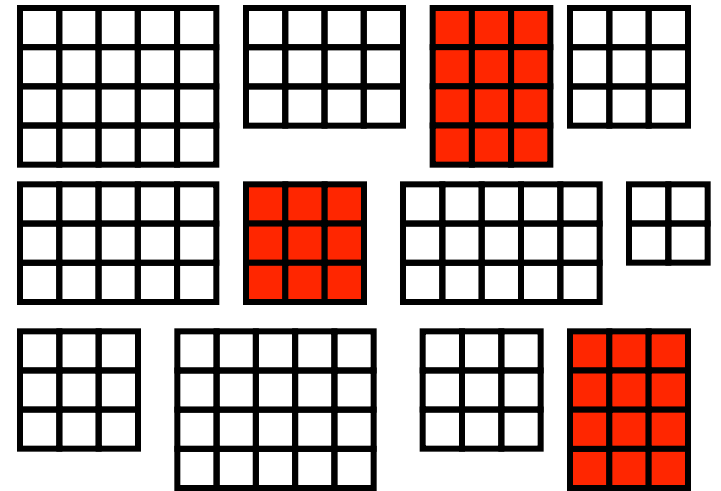
Non-Probabilistic Sampling

- When people are hard to identify or rare
- Convenience: Available and willing
- Snowball: Leverage recommendations
- Quota: Stratified until reach proportions
- Focus groups: Representative of a group

Population and Sampling
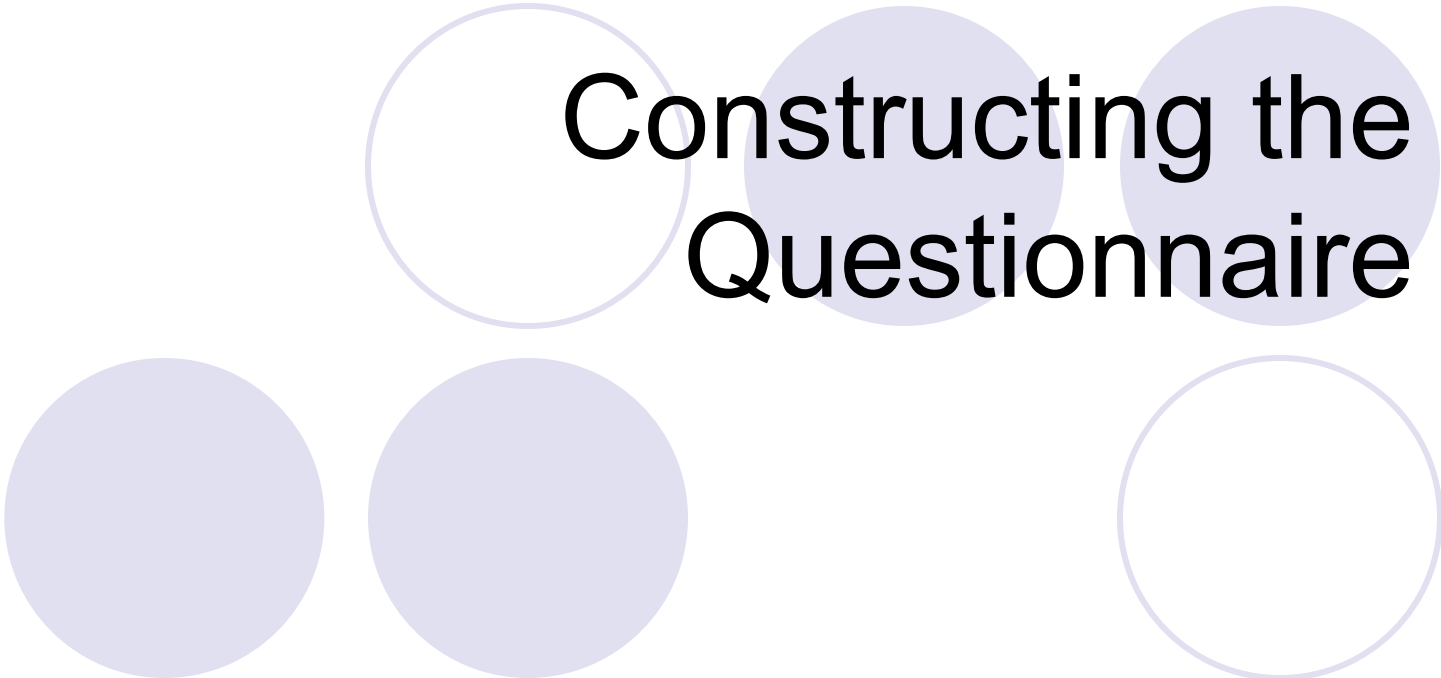
- Non Probabilistic
  - Between 30 and 500
  - 10% of Population
- Probabilistic

Population Size	95 % conf. – 3 % accuracy
50	47
5000	760
100.000	888
900.000	895

Constructing the Questionnaire

Exercise

Task - Design a survey that asks workers at a large organization about their satisfaction with their employer's cafeteria.

Constructing a Survey Instrument

- Search literature for similar surveys
  - Easier to re-validate prior surveys
- Selection of questions crucial
  - Appropriate target and length
  - Be wary of open-ended questions
  - Closed answers should be clear

Questionnaire Evaluation

- Motivation: why should people respond?
  - Help by clarifying purpose of survey
  - Be direct about your intentions
- Researcher bias invalidates survey!
  - May be unintentional, if not trained
  - Must provide balanced questions

Questionnaire Size

- Most dislike answering long surveys
- Long surveys present analysis challenges
- May introduce spurious results
  - Margin of error can overwhelm data
- Inappropriate statistical analysis
  - Yield desired, but insignificant results

Developing High Quality Surveys

- Be clear about what you plan to measure
  - Use theory as a guide.
  - If there is no theory, use a well-formulated definition.
  - If no good definition, conceptualize and describe how the construct relates to existing phenomena (DeVellis, 2003).
- Be as specific as possible about the topic

Developing High Quality Surveys

- Generate multiple items (i.e., a multi-item scale) to assess phenomena of interest
- Rule of Thumb (Campinelli, 2005): Goal = 8-12 items per construct

Multi-item Scales

- Fowler (1995) –
  - Allow more detailed measurements
  - Reduce item-specific effects
- Converse and Presser (1986) –
  - Additional items may compound rather than alleviate wording and context effects
  - Item meanings may not remain consistent over time
- DeVellis (2003)
  - longer scales are less prone to changes in alpha (reliability) for new samples

Generating Multiple Items

- Create items that are of high quality
    - Consider content and face validity (have pool of experts review)
- Check that items conform to highest design principles (have survey methodologist review)

Resources are:

- Fowler, F. J. Jr. (1995). Improving survey questions: Design and evaluation. London, Sage.
- Dillman, D. (1999). Mail and Internet Surveys: The Tailored Design Method, 2nd Edition. New York: John Wiley Company,New York.

Item Generation - Deductive Approach

It requires:

(a) an understanding of the phenomenon to be investigated;

(b) thorough review of the literature to develop the theoretical definition of the construct under examination

From Hinkin (1998)

Item Generation-Deductive Approach

- Advantages: through adequate construct definitions, items should capture the domain of interest, thus to assure content validity in the final scale

- Disadvantages: requires the researchers to possess working knowledge of the phenomena; may not be appropriate for exploratory studies

From Hinkin (1998)

Item Generation - Inductive Approach

- Appropriate when the conceptual basis may not result in easily identifiable dimensions for which items can then be generated

- Frequently researchers develop scales inductively by asking a sample of respondents to provide descriptions of their feelings about their organizations or to describe some aspects of behavior

- Responses classified into a number of categories by content analysis based on key words or themes or using a sorting process

Item Generation - Inductive Approach

- Advantages: effective in exploratory research

- Disadvantages:
  - Without a definition of construct under examination, it is difficult to develop items that will be conceptually consistent.
  - Requires expertise on content analyses
  - Rely on factor analysis which does not guarantee items which load on the same factors share the same theoretical construct

Characteristics of Good Items

- As simple and short as possible
- Language should be familiar to target audience
- Keep items consistent in terms of perspectives (e.g., assess behaviors vs. affective response)
- Item should address one single issue (no double-barreled items)
- Leading questions should be avoided
- Negatively worded questions should be carefully constructed and placed in the survey

What about these items?

- I would never drink and drive for fear of that I might be stopped by the police (*yes* or *no*)
- I am always furious (*yes* or *no*)
- I often lose my temper (*never* to *always*)
- 滿招損，謙受益

Types of data

- **Categorical data**
  - Fixed number of outcomes
  - Eg: Faculty – Sciences, Agriculture, Arts
  - No order in the outcomes
  - Divide sample into groups
- **Continuous data**
  - Measurements (age, height, weight etc)

Types of data

- Ordinal data
  - Order is important
  - Outcomes take on discrete values
  - Eg Income: 0-1000,
        1000-10000,
        10000+
  - Likert scales: 0-disagree, to 5-agree
  - Mostly treated as continuous, some times as categorical

Likert scale

- Way of quantifying people's opinions, feelings experiences etc
- Most common: 1-5 scale or 1-7
- Words assigned to each scale:
  - 1 Completely disagree
  - 2 Disagree
  - 3 Neutral
  - 4 Agree
  - 5 Fully agree

Variables & cases

- Each item on a questionnaire is a variable
- Other variables can be derived from items, eg: BMI=weight/height^2
- Each respondent included in the survey is a case

Latent variables

- Variable that cannot be directly measured:
  - Service quality
  - Economic confidence index
  - Family resilience
  - Emotional intelligence
- It is measured indirectly through measurable variables like items on a questionnaire
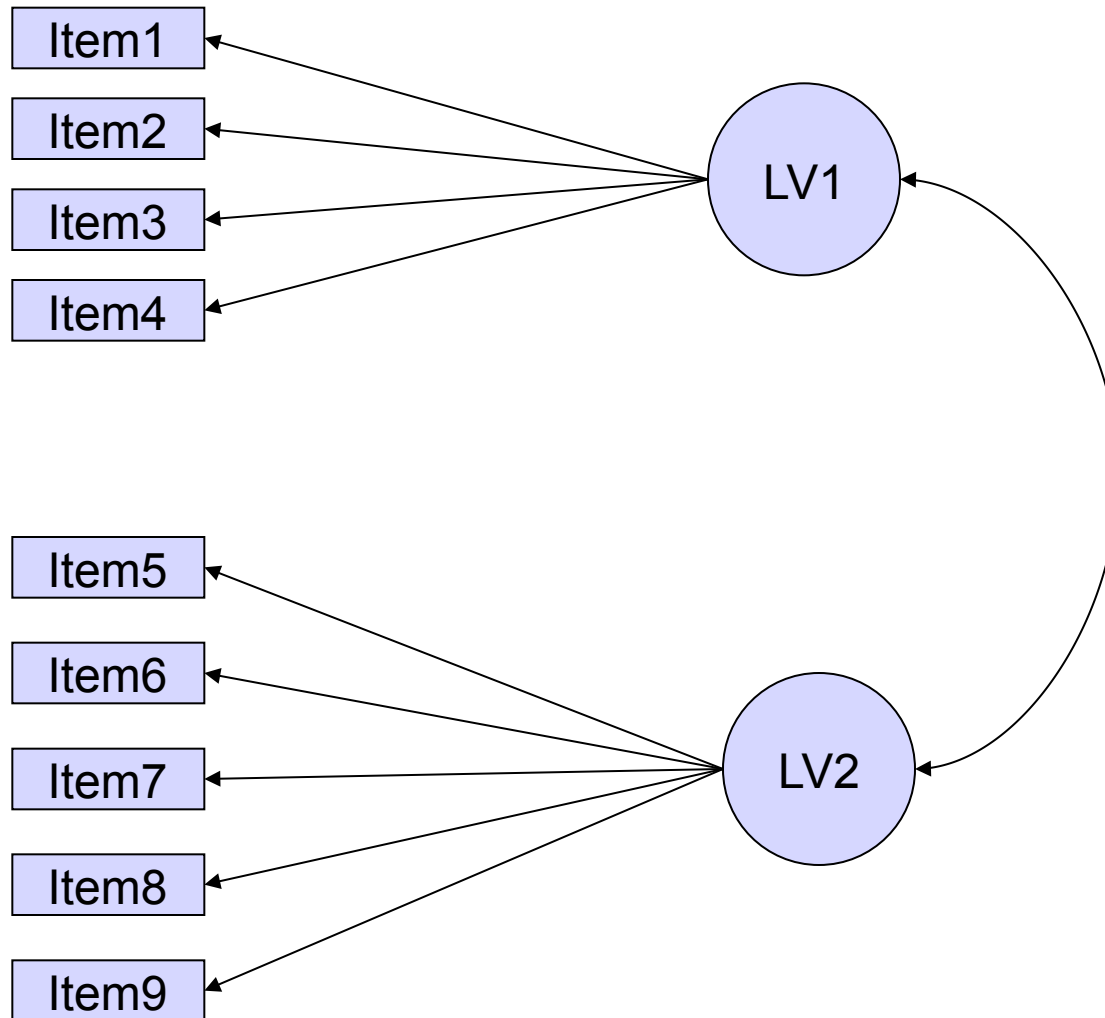
Latent variables
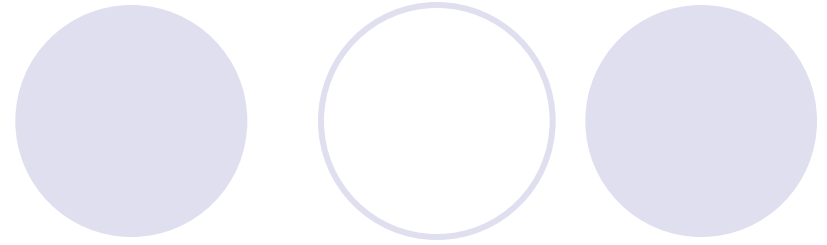
- Other terms used:
  - Constructs
  - Scales & sub scales
  - Dimensions & sub-dimensions
- Score for latent variable:
  - Sum or avg of measured items

Latent variable structure

- More than one latent variable evaluated
- Structure:
  - Indicates which items measure which variable
  - How are the latent variables interrelated

Latent variable structure

What construct does this scale measure? (1)

1. Have a job which leaves you sufficient time for your personal or family life. (.86)
2. Have training opportunities (to improve your skills or learn new skills). (-.82)
3. Have good physical working conditions (good ventilation and lighting, adequate work space, etc.). (-.69)
4. Fully use your skills and abilities on the job. (-.63)
5. Have considerable freedom to adapt your own approach to the job. (.49)
6. Have challenging work to do---work from which you can get a personal sense of accomplishment. (.46)
7. Work with people who cooperate well with one another.(*.20*)
8. Have a good working relationship with your manager.(*.20*)

Adapted from Heine et al. (2002)

Pre-Testing

- Helps identify problems or biases early
- Start out small
  - Focus groups: mediated discussions
  - Pilot studies: small distribution
- Contributes to validity and reliability
  - SE: weak validity and reliability

Developing High quality surveys

- Analyze pilot test responses to assess:
- Uni-dimensional structure of scale (using exploratory factor analysis - EFA)
- How items discriminate – by reviewing means, variances / standard deviations (want moderate variance)
- Inter-item correlations (want high correlations /Cronbach alpha values as scale should assess one construct)
- Item to scale correlations (want high correlation)

Reliability

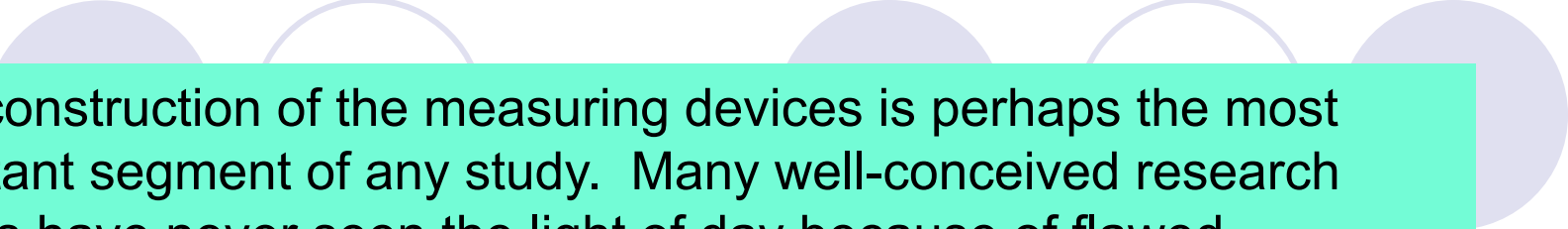- Refers to the extent to which the
scale / survey provides consistent results when surveying similar populations.
- Multiple types of reliability –
  - Internal consistency - correlation among the variables comprising the set (Cronbach's alpha)
    - We can only speak about the reliability of each construct, and not the reliability of a survey as a whole, if the survey measures more than one construct.

Multiple types of validity – (here are just 3)

- Construct validity - the approximate truth of the conclusion that your operationalization accurately reflects its construct.

- Face validity – whether operationalization "on its face"

  ○ it seems like a good translation of the construct.

- Content validity - check of the operationalization against the relevant content domain for the construct.

  ○ This approach assumes that you have a good detailed description of the content domain, something that's not always true.

"The construction of the measuring devices is perhaps the most important segment of any study. Many well-conceived research studies have never seen the light of day because of flawed measures."

Schoenfeldt, 1984

"The point is not that adequate measurement is 'nice'. It is necessary, crucial, etc. Without it we have nothing."

Korman, 1974, p. 194

"Validation is an unending process….Most psychological measures need to be constantly evaluated and reevaluated to see if they are behaving as they should."

Nunnally & Bernstein, 1994, p. 84

How is Reliability Related to Validity?

A measure may be reliable but not valid, but it cannot be valid without being reliable. That is, reliability is a necessary but not sufficient condition for validity.

Survey reporting

Introduction

- Describe the purpose of your study,

- Describe the importance (significance) of the study—why it was worth doing in the first place,

- Explain why you used this particular theory or model and what are its advantages,

- Provide a rationale, very briefly describe the research design and how it accomplished the stated objectives

Method used

- the type of survey used;

- the limitations of the online survey methodology;

- the survey questions;

- the types of response categories—that is, drop-down menus, checklists, open ended;

- the distribution methods: how the survey was distributed, how the participants were recruited, how the nonrespondents were handled;

Method used (II)

- information about incentives if any were offered;
- how the survey was pilot tested;
- how reliability and validity were addressed;
- how the survey was constructed;
- how many people were e-mailed the survey, where you obtained their e-mail addresses from, what were the eligibility criteria;
- if the survey was anonymous;
- how you handled incomplete surveys; and
- how informed consent was handled

Results

- Do not discuss or interpret your results, report background information, or attempt to explain the analysis.

- Never include raw data or calculations in a research paper.

- Do not present the same data more than once.

- Text should complement any figures or tables, not repeat the same information.

- Do not confuse figures with tables. Figures include graphs, photographs, illustrations, diagrams, and so on. Tables have a row-and-column structure—that is, tables contain cells.

Discussion

- Decide if each hypothesis is supported or rejected, or if you cannot make a decision with confidence. Do not simply dismiss a study or a part of a study as "inconclusive."

- Research papers are not accepted if the work is incomplete. Draw what conclusions you can based on the results that you have, and identify the questions that remain unanswered and why.

- You may suggest future directions, such as how the survey might be modified to accomplish another objective.

Discussion (II)

- Explain all your observations as much as possible.
- Decide if the research design adequately addressed the hypothesis and whether or not it was properly controlled.
- Try to offer alternative explanations if reasonable alternatives exist.
- One survey research project will not answer all questions, so keeping the big picture in mind, where do you go next? The best studies open up new avenues of research. What questions remain? What recommendations do you have for future research?

Decision-Making

- Decision-making and action
  - Influenced by survey results (hopefully!)
  - Actions may include further research

Statistics in Survey Research

- In summary:
  - Statistics is generally most useful in the design and analysis stages of surveys
    - Especially sampling, weighting, and data analysis
  - Also relevant at other stages
    - Quality control and quality improvement for survey operations
    - Effect of survey procedures on survey results
    - Interpretation of survey results

Summary: Best practices

- Study the literature & the phenomenon to come up with a broad definition of the construct

- Collect good behavioral incidents (quantity & quality)

- Use inductive and deductive approaches alternately in the development process

Take Away Lessons

- Good survey measures must be grounded on sound theory and conceptual definitions

- Developing good survey measures takes much time, resources, experiences, and commitment, but the payoff can be immense!!

- Avoid convenience measurement at all time!!!

- If there is a good, published measure available, Use it!!! Not to reinvent the vehicle!!!

References

Based on some other presentations:

- Surveys Overview. Department of Statistics - University of Auckland. www.stat.auckland.ac.nz/~balemi/SurveysSampling.ppt. Acesso em setembro/2011.

- José Danado. IDI-NTNU. Survey as a Research Method - Validity and Threats. Presented on June/2011.

These books are highly readable:

- Fowler, F. J. Jr. (1995). Improving survey questions: Design and evaluation. London: Sage.

- DeVellis, R.F. (2003). Scale development: Theory and applications, 2nd edition. Thousand Oaks, CA: Sage

- Dillman, D. (1999). Mail and Internet surveys: The tailored design method, 2nd Edition. New York: John Wiley Company.