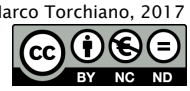
Population Surveys

Empirical Methods in Software Engineering (01OPJIU)

http://softeng.polito.it/EMSE/

Version 2.2

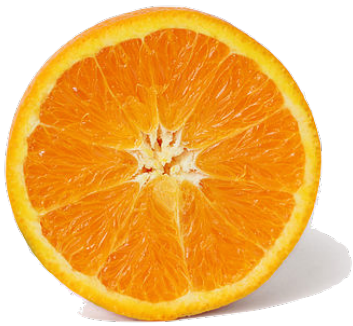


Licensing Note

Agenda

- Error and inference
  - Measurement
  - Representation
- Sampling
- Questionnaire design
  - Cognitive issues
  - Guideline for writing questions
  - Privacy

Observational Studies



Survey
(Cross-Sectional)

Case study

Case-Control

Survey

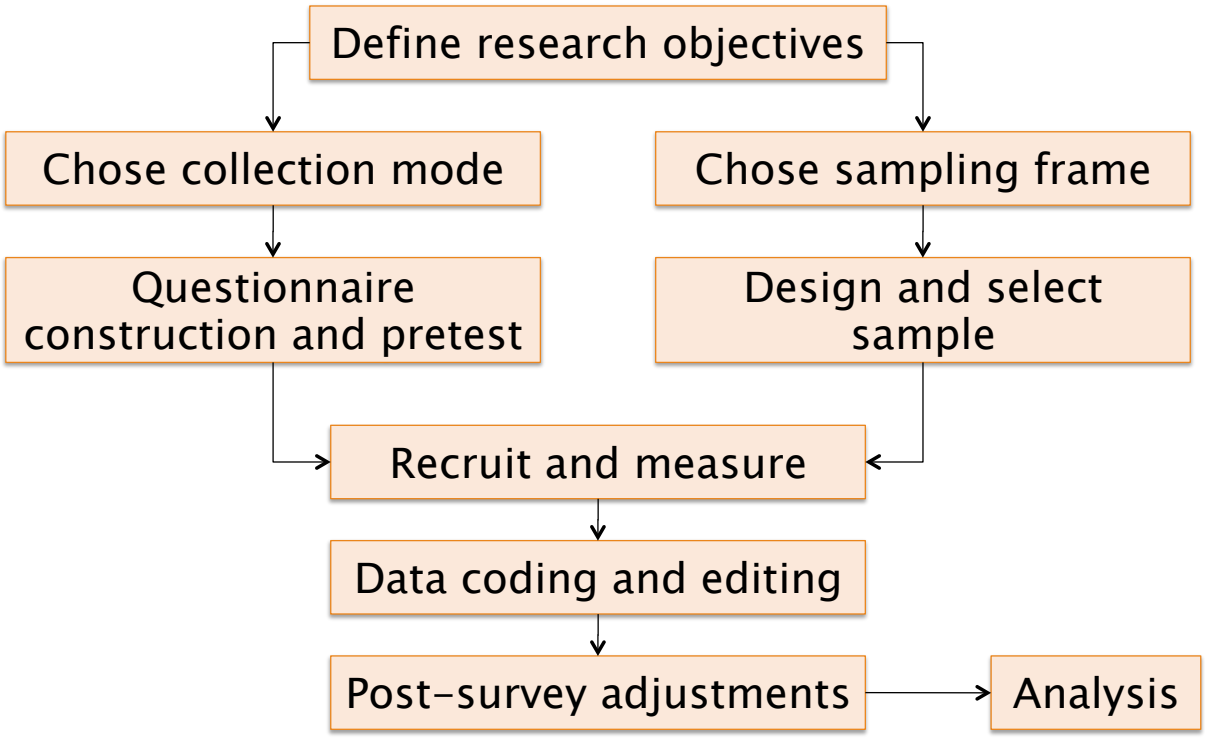
Systematic observational method

to gather information from
(a sample of) entities

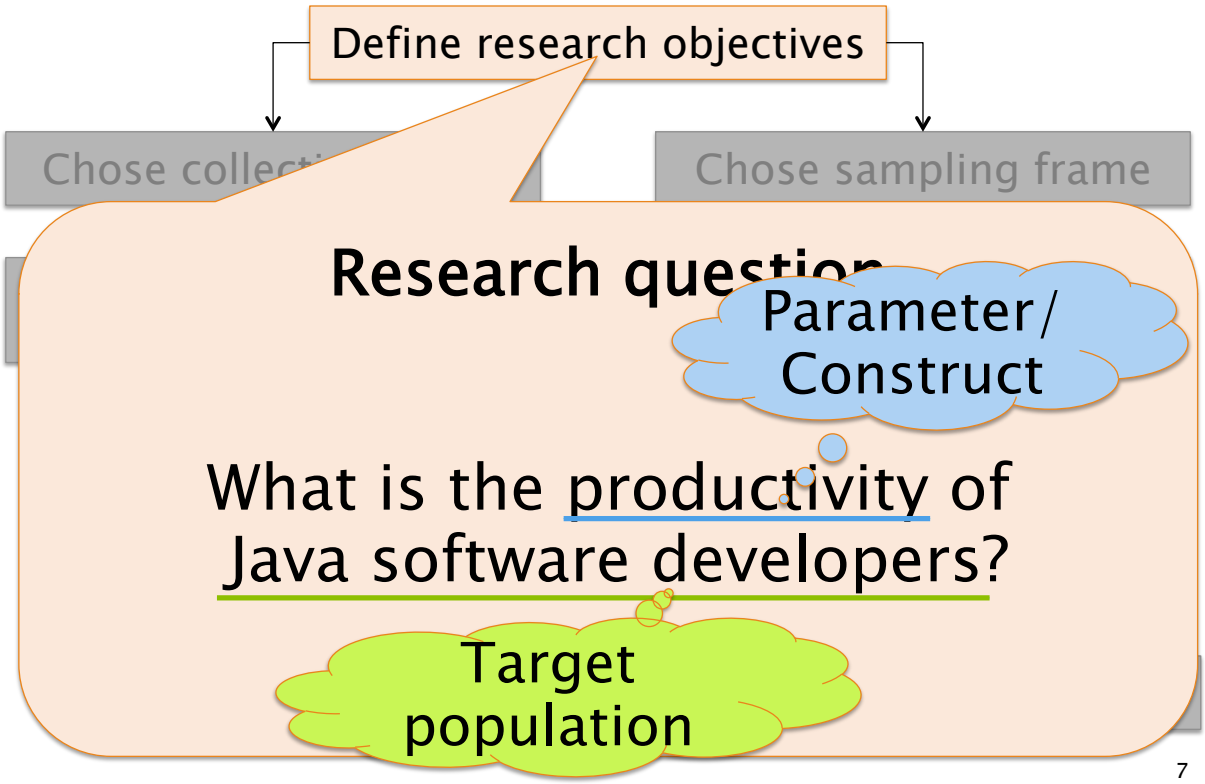
for the purpose of constructing
quantitative descriptors of a
population

Descriptive statistics + Analytic statistics

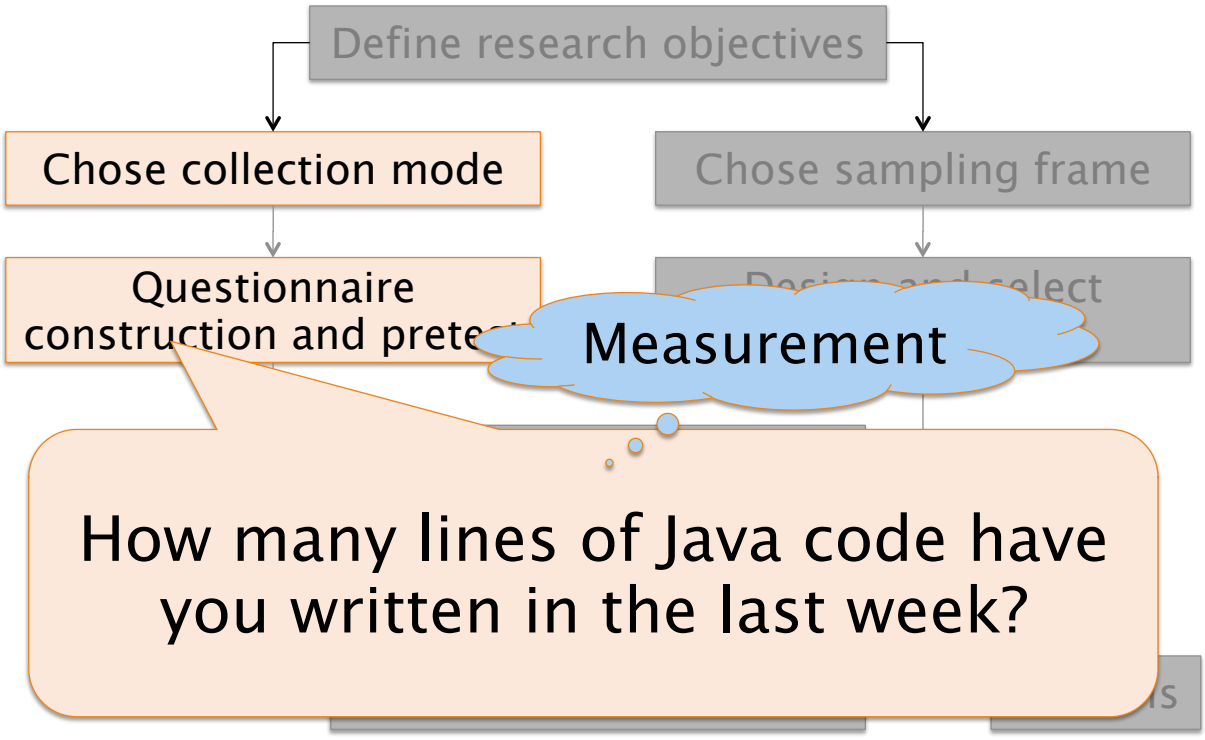
Survey process



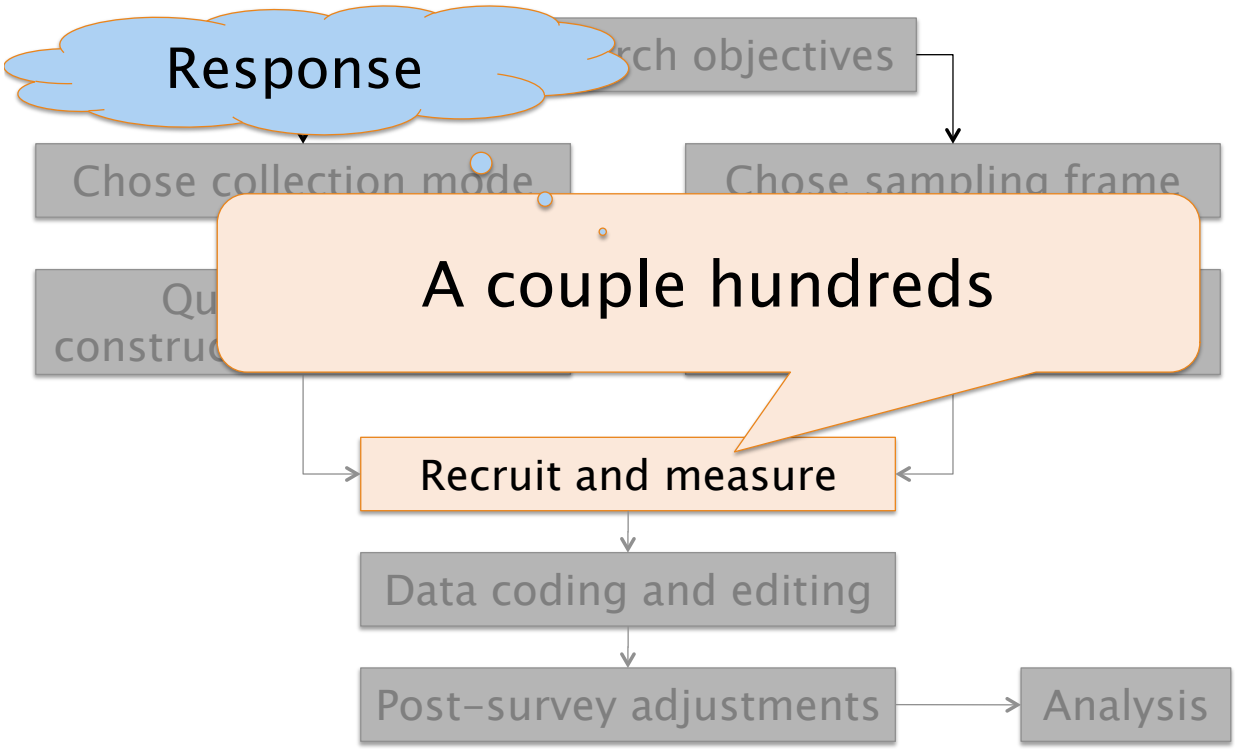
Survey process



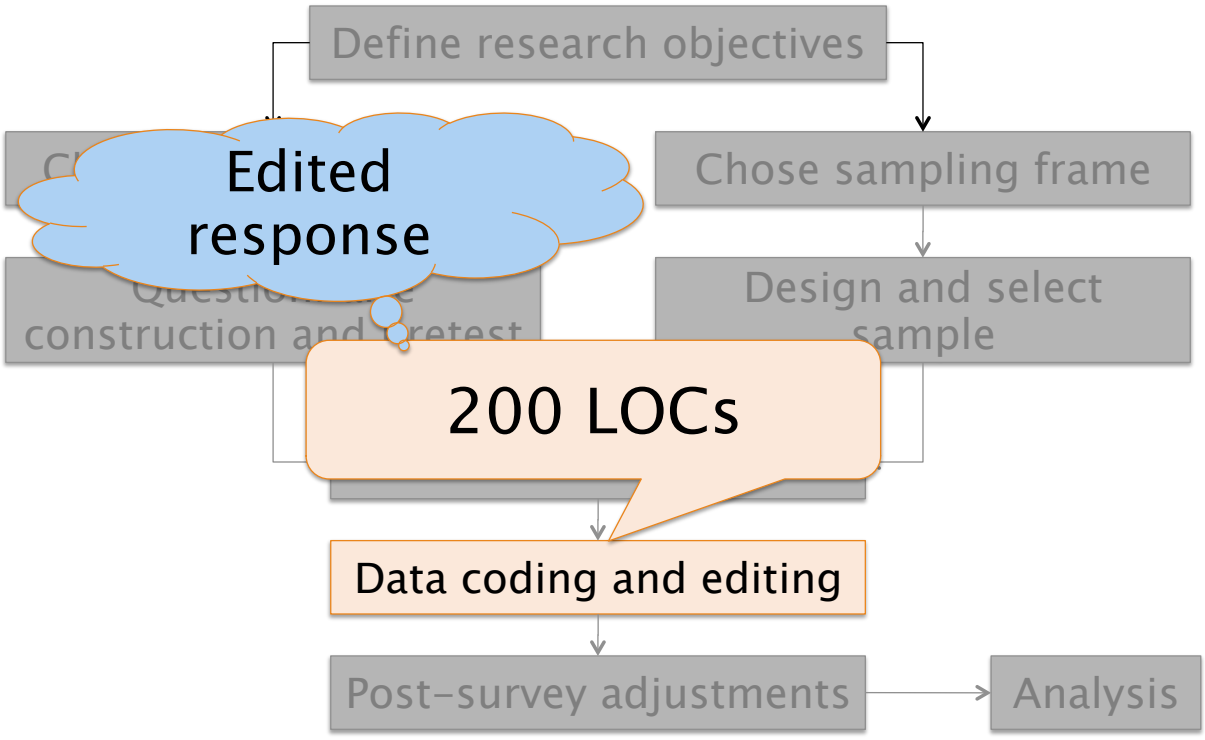
Survey process



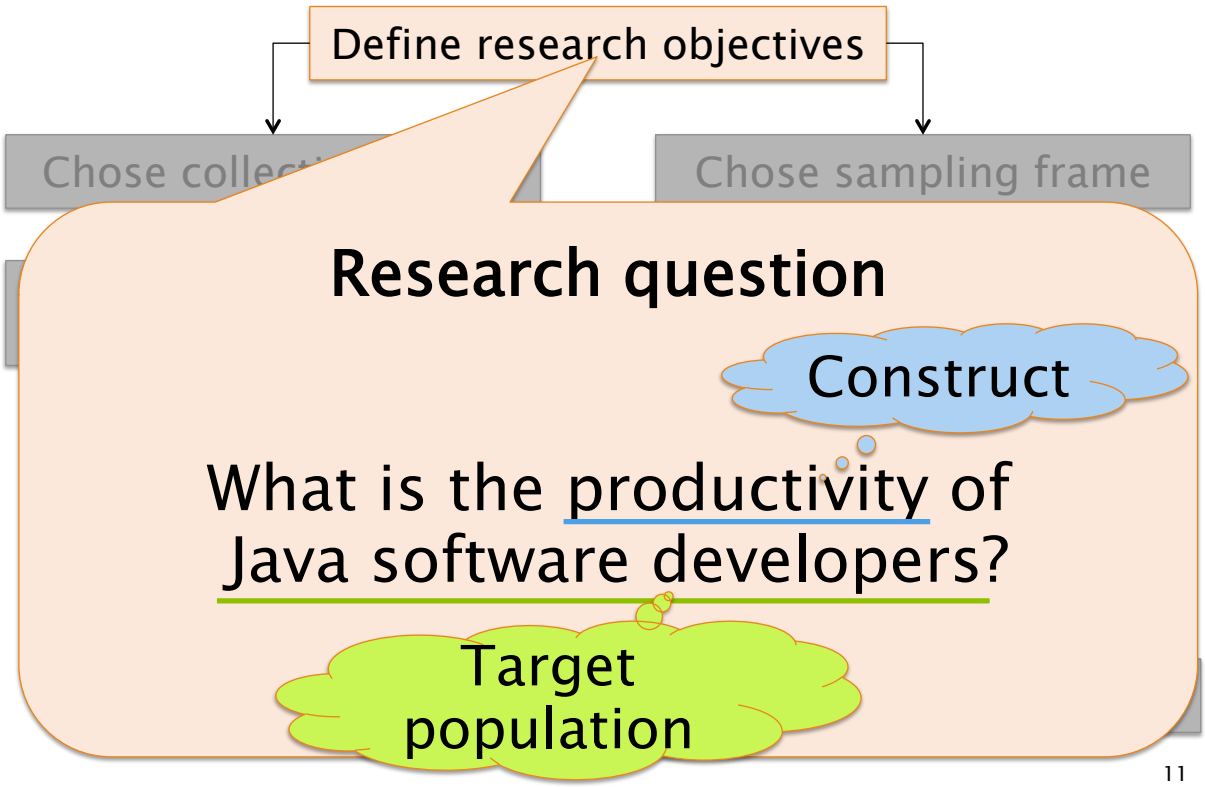
Survey process



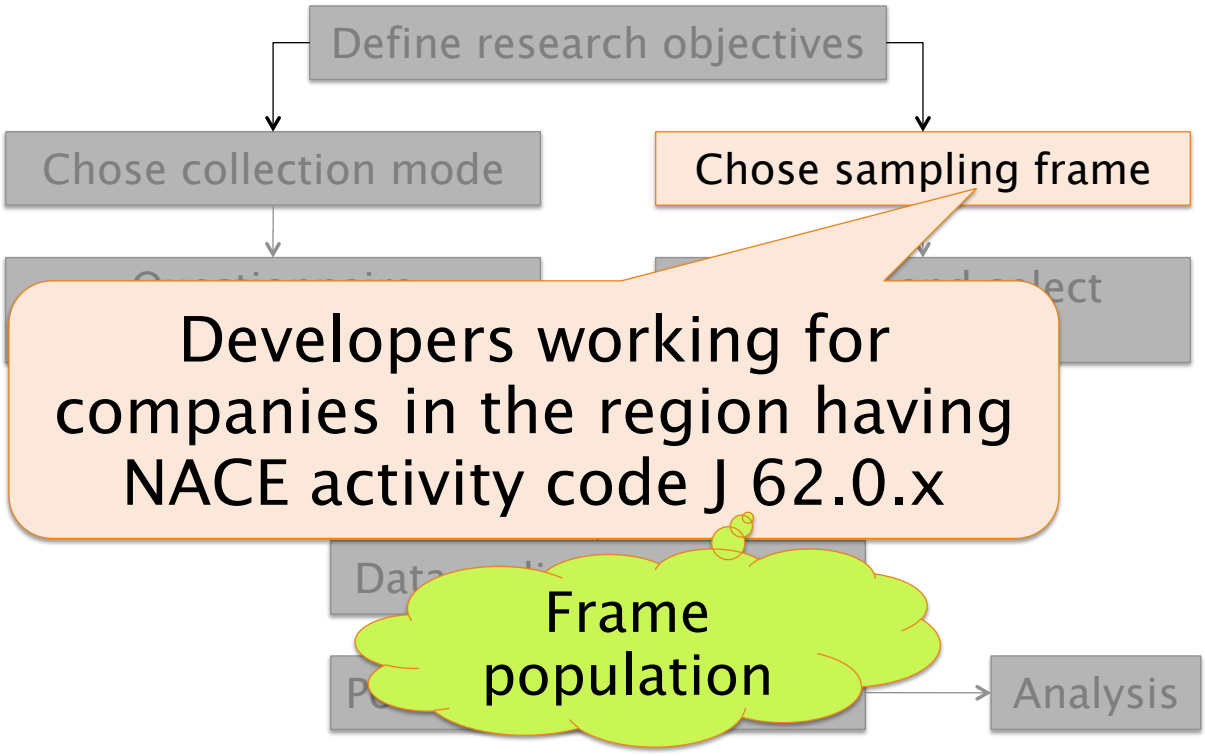
Survey process



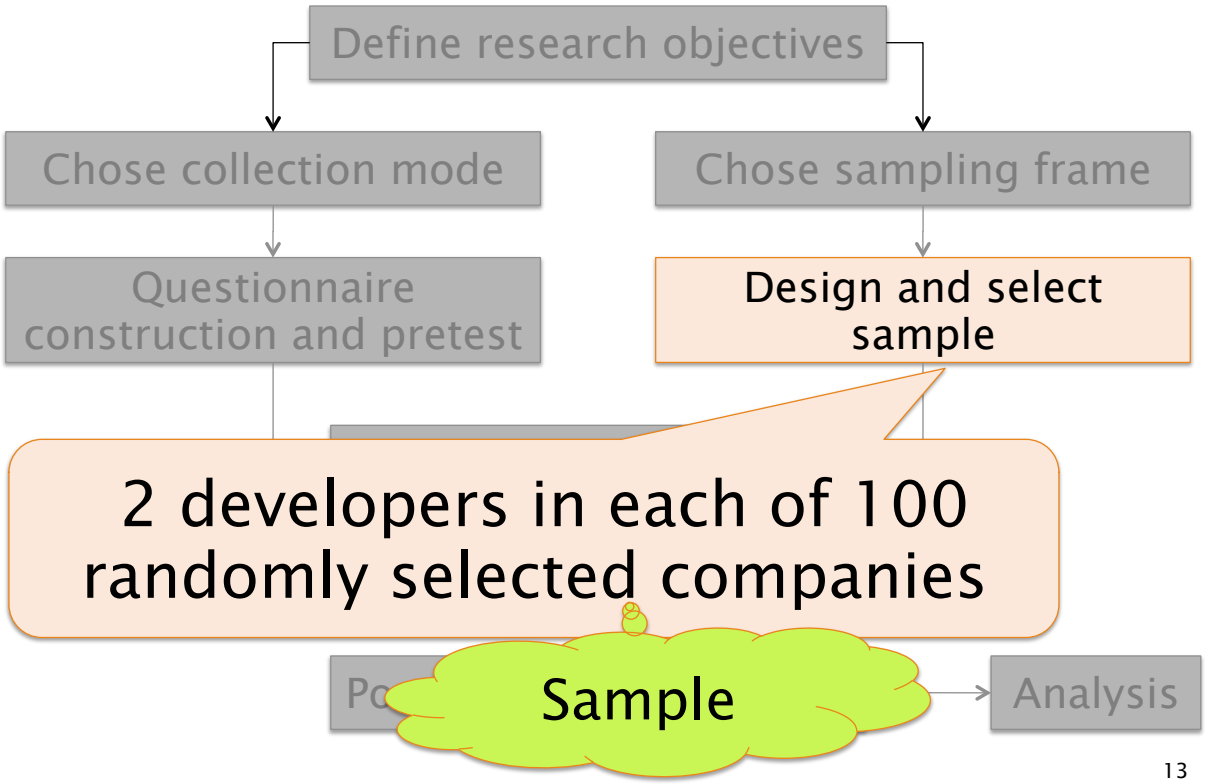
Survey process



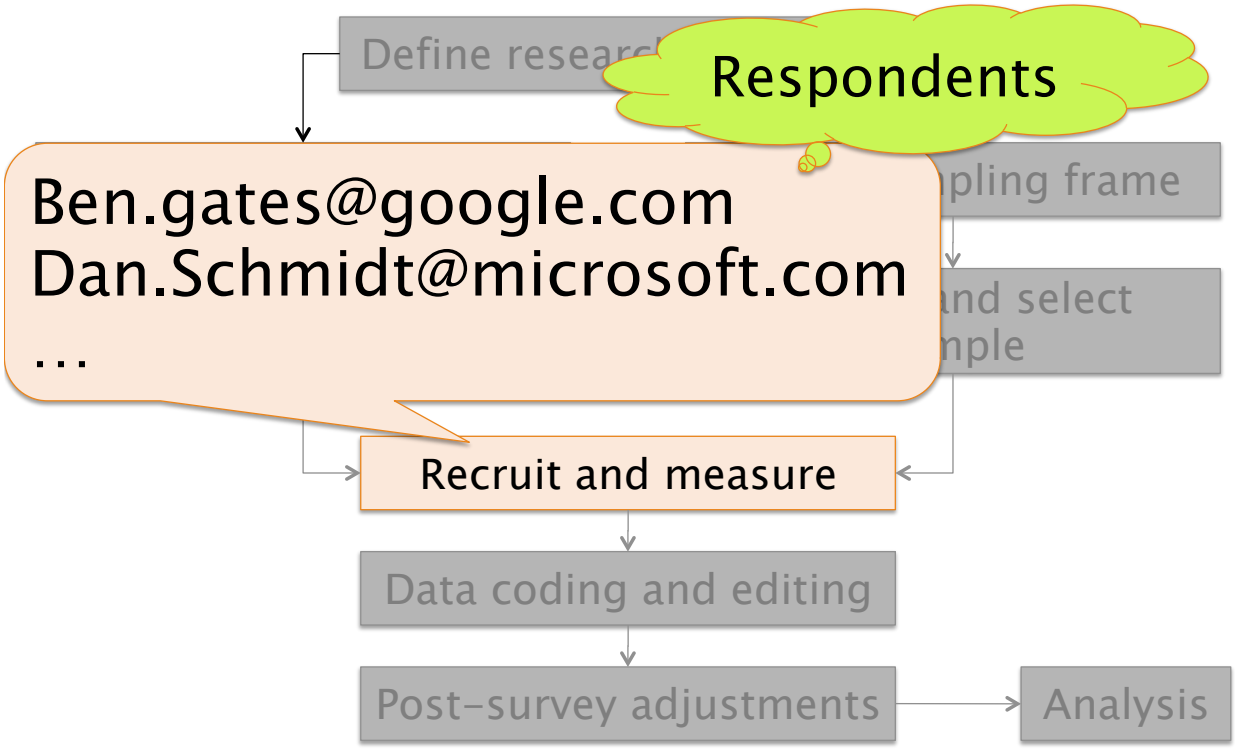
Survey process



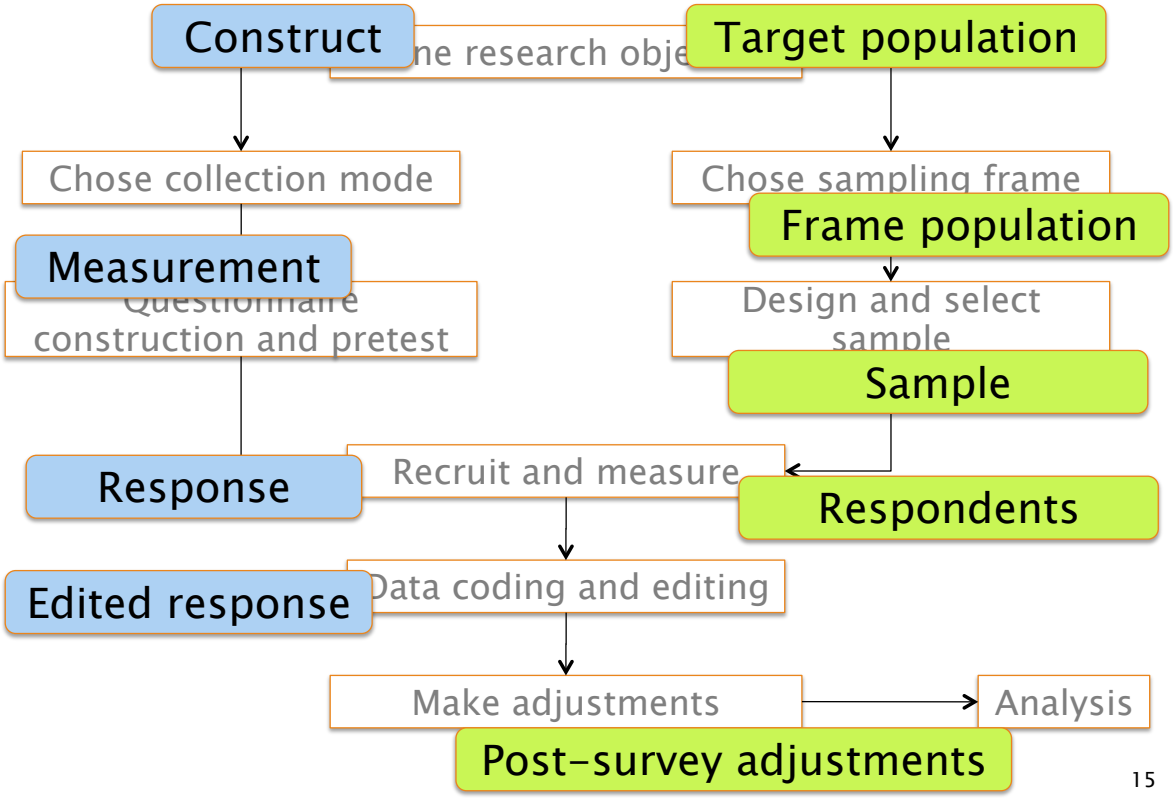
Survey process



Survey process

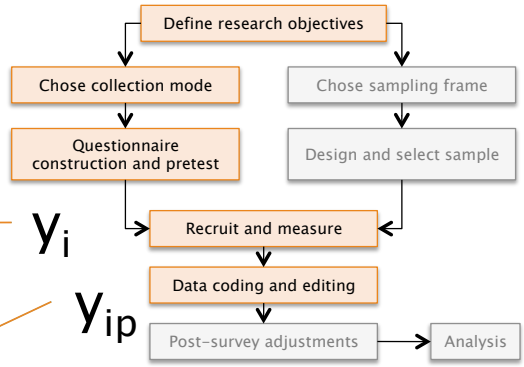


Measurement vs Representation



Measurement perspective

- Construct — μ_i
- Measurement — Y_i
- Response — y_i
- Edited response — y_{ip}



Construct

- Element of information sought by researchers
- Examples
  - How many new jobs created
  - How many incidents of crime with victims
  - Which developments tools used
- Formulation
  - Easy to understand
  - Imprecise
  - Abstract

Construct

- Abstraction
  - Directly observable
    - E.g. Staff for a project
    - A few defined ways to measure
  - Non directly observable
    - Intention to adopt a technology
    - No single well-defined measure

Measurement

- How to gather information about constructs
  - Objective measures
    - Electronic
    - Physical
  - Answers to questions
    - Visual
    - Oral

Validity

- Gap between constructs and measurement
  - Ideally the measure is the result of just one among several possible trials
  - In practice the measurement may introduce an error
    - Each trial introduces a different error

$$Y_{it} = \mu_i + \epsilon_{it}$$

- Validity=correlation between Y and μ

Response

- The actual data collected through the survey
  - A question may require
    - Search own memory
    - Access records
    - Ask other persons
  - Closed questions already contain possible answers
  - Sometimes a response is not provided

Measurement error

- Gap between the ideal measurement outcome and response obtained

$$y_i - Y_i$$

  - Response bias
    - Systematic misreporting
  - Reliability
    - Variability over several trials

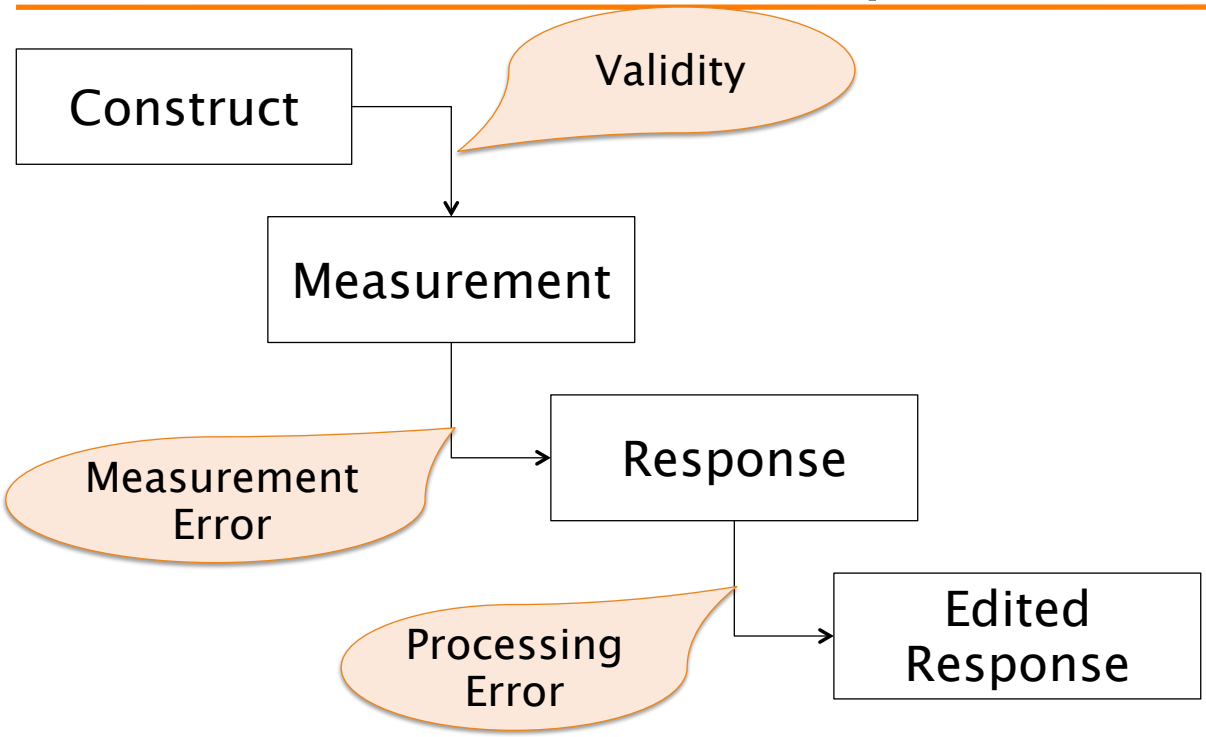
Edited response

- Review process before using data
  - Range checks
  - Consistency checks
  - Illegible answers detection
  - Skipped questions
  - Outlier detection

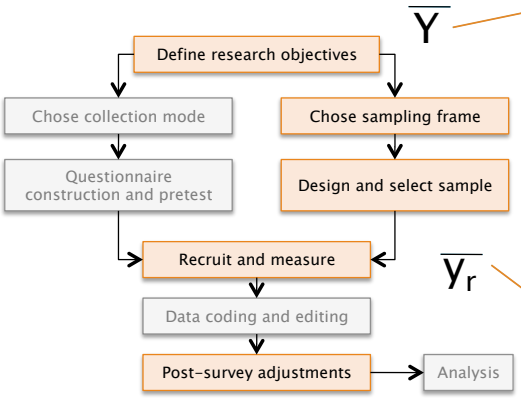
Processing error

- Gap between variables used in analysis and those provided by the respondent
  - Erroneous outlier identification
  - Coding error

Errors – measurement pov

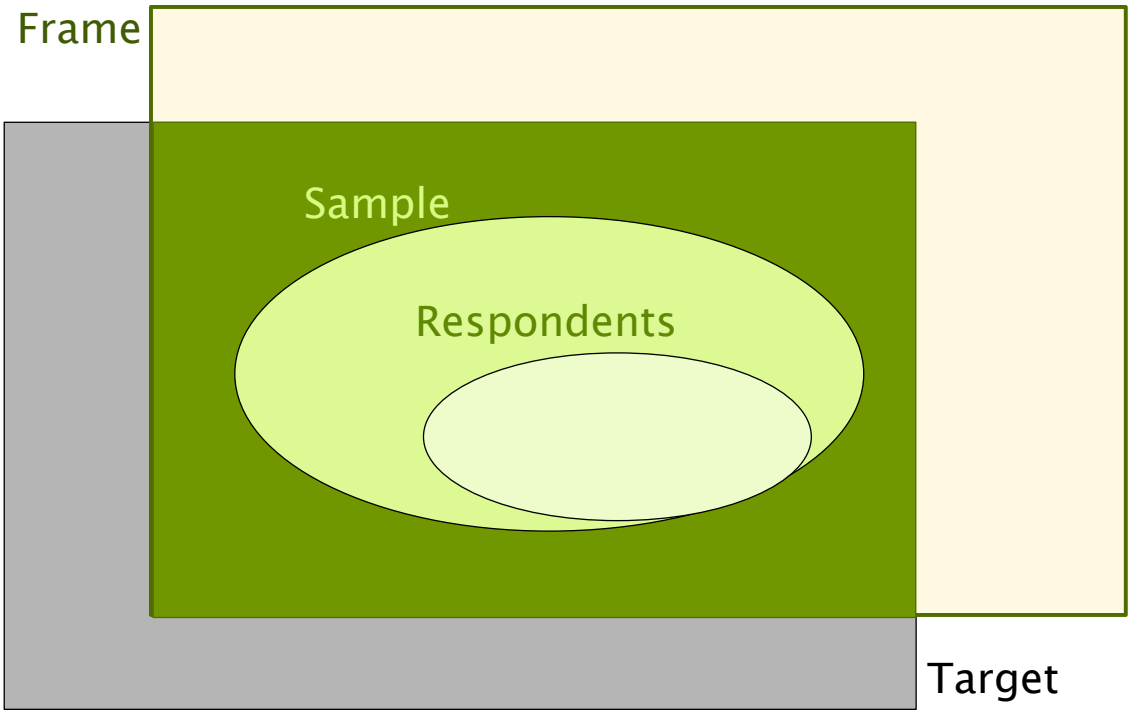


Representation perspective



- $\overline{Y}$ Target population
- $\overline{Y}_C$ Frame population
- $\overline{y}_a$ Sample
- $\overline{y}_r$ Respondents
- Post-survey adjustments

Representation perspective



Target population

- The set of units to be studies
  - Abstract population definition
  - E.g. software projects
    - Time?
    - In sw companies only?
    - Italian companies only?
    - Completed or just started projects?

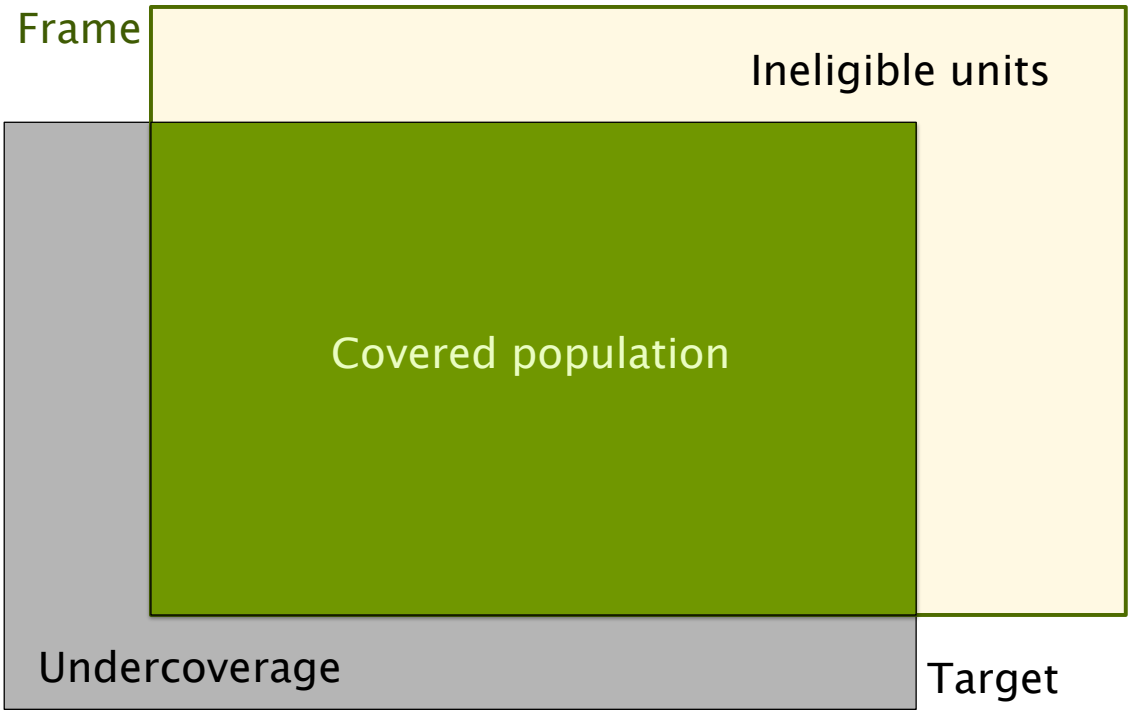
Frame population

- In theory
  - The subset of target population that has a chance to be selected
- In practice
  - a set of units imperfectly linked to the target population members
    - E.g. telephone numbers

Framing instrument

- The (conceptual) instrument used to identify the units of study
  - Household phone numbers to get persons
  - Company records to get employees
  - Customer IDs to get customers

Coverage Error



Reality check

- During 2012 USA Presidential Elections Campaign because of an effect of Federal regulations polling cellphones was more expensive
- As a result, many public polls leave cellphone users out of their samples.
- Due to the growing popularity of cellphones as the only point of contact for young voters and minorities, pollers left key constituencies for Obama out of the polls and skewed the numbers for Romney in some samples.

"*That's why some polls looked so difficult for the president, because they were under-polling the electorate for the president*"

J. Messina (Campaign Manager for Obama)

http://www.politico.com/news/stories/1112/84103.html?hp=l1

Coverage bias

- Two factors
  - Difference between covered and not covered population
    - $\overline{Y}$: mean of target
    - $\overline{Y}_C$: mean of covered $\overline{Y}_U$: mean of uncovered
  - Proportion of non covered population
    - C: # covered units U: # uncovered units

$$\overline{Y}_C - \overline{Y} = \frac{U}{C}(\overline{Y}_C - \overline{Y}_U)$$

Sample

- Units selected from the frame population
  - Time and cost opportunity
- Deliberate non-observation
  - May introduce deviation between
    - Sample statistic
    - Full frame statistic

Sampling error

- Sampling bias
  - Systematic exclusion of some members
    - Or significantly reduced chance of selection
- Sampling variance
  - Ideal set of samples all drawn from the same frame

$$V_s = \frac{\sum_{s=1}^{S} \left(\overline{y}_s - \overline{Y}_C\right)^2}{S}$$

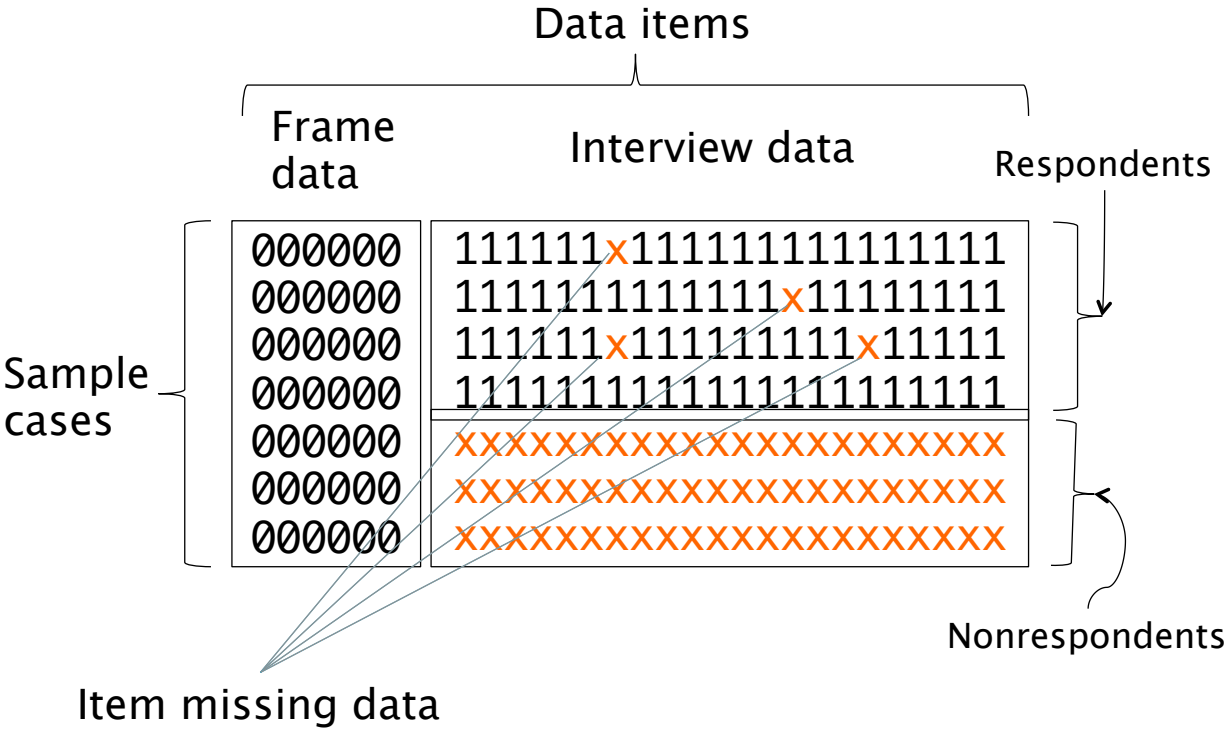
Sampling error reduction

- Probability sampling
  - All units have non zero selection probability
- Stratification
  - Representation of key sub-populations is controlled
- Element samples
  - As opposed to cluster samples
- Sample size

Respondents

- The subset of sample for which a measurement could be collected
  - Item missing data: incomplete measures
- Full participation (i.e. 100% response rate) realistically possible only for inanimate units

Respondents



Non-response error

- Non-response bias
  - Non-response rate: m_S / n_S
  - Difference between respondents and non-respondents

$$\bar{y}_r - \bar{y}_s = \frac{m_s}{n_s}(\bar{y}_r - \bar{y}_m)$$

Post-survey adjustment

- Weighting
  - Compensate under-representation due to
    - Non response patterns
    - Mismatch between frame and target population
- Imputation
  - Item missing data are replaced by estimates

Fitness for use

- Credibility
  - Neutrality
  - Method
    - Documentation
    - Disclose possible weaknesses
- Relevance
  - Survey construct vs. user's concern
- Timeliness
  - How actual are data?

SAMPLING

Terminology

- Elements = fundamental units of population
- Target population
  - Well defined type of units
    - E.g. household, housing unit
- Survey population
  - Restriction of target population practically accessible
    - E.g. military projects

Terminology

- Sampling frame
  - Materials and methods to identify the elements of the target populations
- When available frames miss target
  - Redefine target population
  - Accept coverage errors

Coverage issues

- Undercoverage
- Ineligible units
- Clustering
- Duplication

- Example
  - Sampling software projects though companies using ATECO/NACE codes to identify software companies

Undercoverage

- Introduces errors from non-observation
- Companies
  - NACE codes
    - J 62.0.1 – Computer programming activities
    - J 62.0.2 – Computer consultancy activities
  - Large companies that “also” develop software are not filed under those codes
- Projects
  - What exactly qualifies as software development project?
- Possible use of multiple frames

Ineligible units

- Identification of foreign units
  - In advance
    - Easily solved by protocol
  - At data collection
    - Screening
    - Oversampling if prevalence is known
- Too many foreign units make the frame not cost effective.
  - E.g. sampling all companies in area asking for software projects

Clustering

- Multiple elements of the target population are represented by the same frame element
  - One company runs many projects
- Selection of all target elements
  - Difficult to collect information successfully
  - There may be communication between elements
  - Sample size depends on cluster size

Clustering

- Typically only a sample of cluster elements is selected
  - Unequal probability of selection
  - Elements in large clusters have lower probability of being selected
    - Projects in small companies are more likely being selected than projects in large ones
- Weighting can be used to compensate

Duplication

- A single target element is associated with multiple frame elements
  - Developers from different companies work at the same project
- Target sub-population with multiple frame units have higher chances of being selected
  - Purging duplicates
  - Discarding duplicates

SAMPLE SIZE – AN EXAMPLE

Java vs. C#

- We aim at finding out which is the preferred language between
  - Java
  - C#
- Let's assume 2 out of 3 favor Java
- We want to
  - Be 99% sure of collecting enough answers
  - Conclude with a 99% confidence
  - That Java is most preferred language
- We need to ask 185 developers at least!

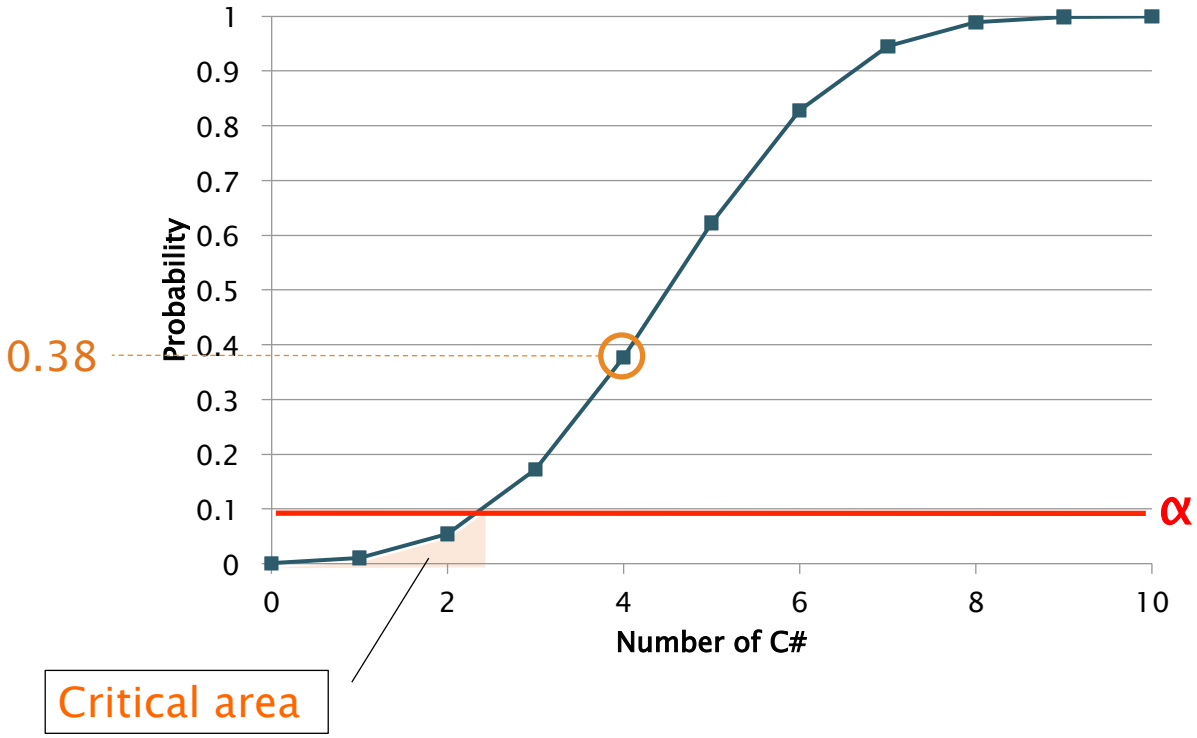
Example – Java vs. C#

- Experiment result: 4 C# preferences from 10 respondents

Java Java Java C# C# Java C# Java C# Java

- What is the probability of having 4 or less C# preferences out of 10 replies?
  - Binomial distribution
    - Cumulative function

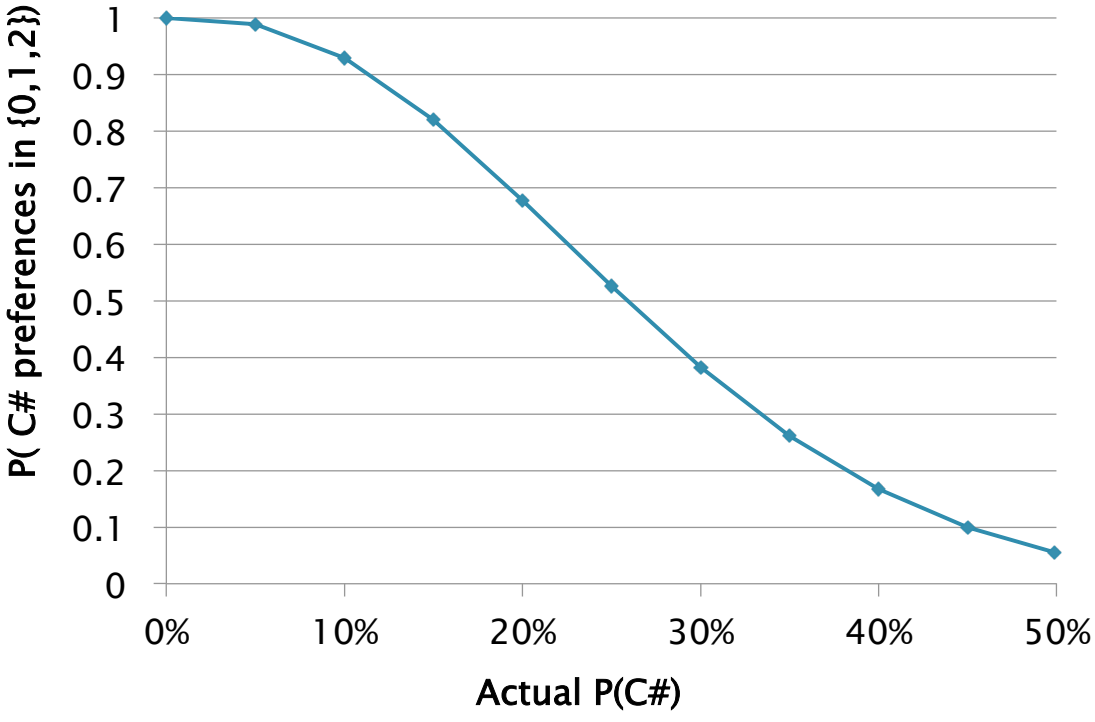
Example – Testing



Power

- We conclude C# is less preferred if we are in the critical area
  - 0, 1, or 2 preferences
- What is the real capability of discovering a real (dis)preference?
  - Power
- Assuming C# is actually less preferred than Java what is the probability of discovering it?
  - P(C#) < 0.5

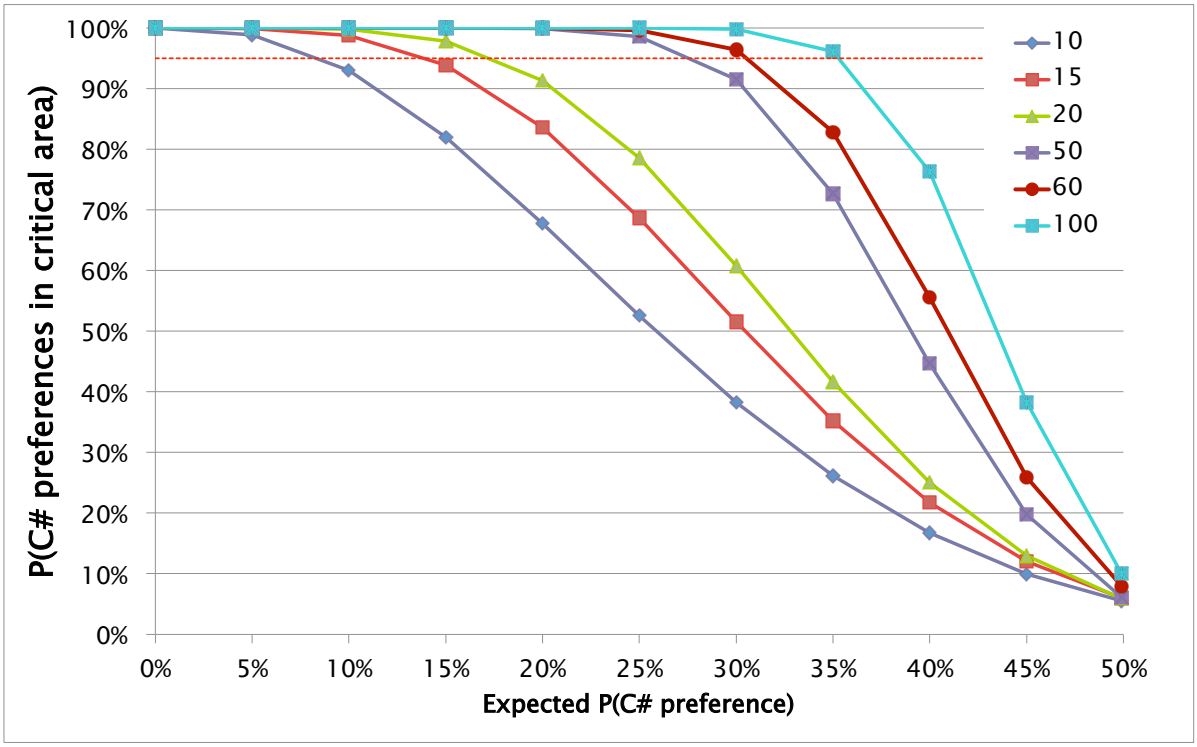
Example – Power



Example – Power

- Let's suppose we suspect that C# is preferred by 30% of the times
  - P(C#) = 30%
  - P(C# preferences in {0, 1, 2}) ~ 38%
- How many developers should we ask to have at least 95% of chances to discover such a bias?
  - Power > 0.95

Power vs. Sample size



Power vs. Sample size

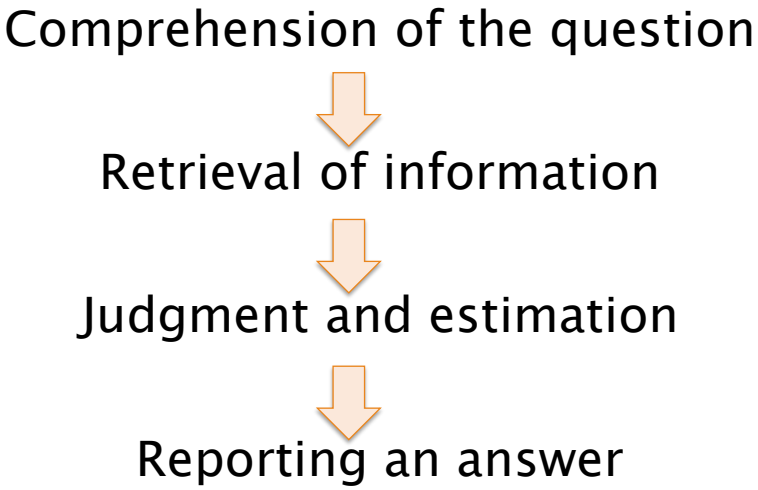
Expected P(C# preferences)	Required sample size
30%	55
40%	223
45%	893

Required power: 95%
Significance level: 10%

QUESTIONNAIRE DESIGN

Cognitive Process

- Answering questions



Comprehension

- Focus on
  - Question
  - Instructions
  - Set of permissible answers
- Interpretation
  - Parse the question
  - Assign meaning to key substantives
  - Infer the purpose behind the question
  - Determine boundaries and overlaps among permissible answers

Retrieval

- Retrieve information relevant to answering the question from long-term memory
  - Start with retrieval cue
    - Number and richness
  - Difficulty for far events

Judgment and estimation

- Combining and supplementing the results of retrieval
  - Retrieved hints are not precise
  - Must be put into relation to other cues to improve precision
    - E.g. relate one event time to another event time

Reporting

- Selection and communication of an answer
  - Map the answer onto question's options
  - Question type:
    - Open
    - Closed

Problems in answering

- Failure to encode the information sought
- Misinterpretation of the question
- Memory problems
- Flawed judgment or estimation strategies
- Problems in formatting the answer
- Misreporting
  - More or less deliberate
- Failure to follow instructions

Encoding problems

- In general living through and event does not imply absorbing much information about it
  - For routine event accounts may be largely based on what usually happens
- If people didn't encode information in the first place no question will elicit accurate responses
  - People cannot provide information they do not have

Misinterpretation

- Even most common terms may be interpreted in several different ways
- Respondents typically are reluctant to ask for clarification
  - They do not admit not understanding
    - A large share (40%) venture an opinion on fictitious issue
  - Unfamiliar or technical terms may result in a wide range of interpretations

Misinterpretation

- Grammatical ambiguity
  - Did you eat the cookies on the couch?
  - Spoken: I scream for an ice cream
- Excessive complexity
- Faulty presupposition
  - How many classes are in your program?

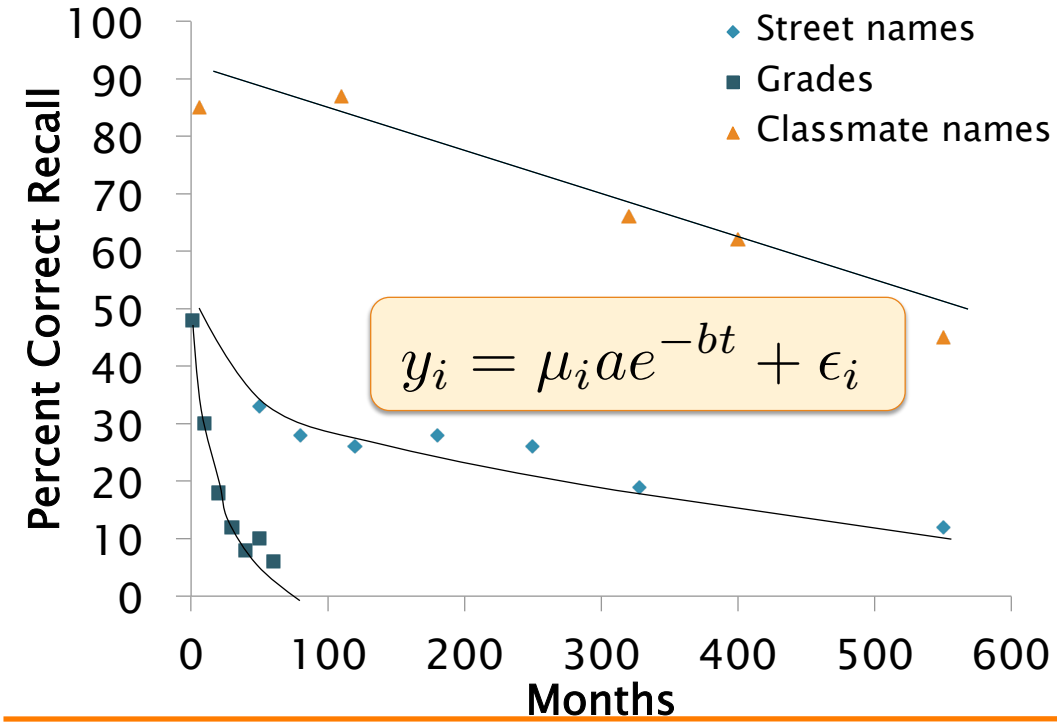
Misinterpretation

- Vague concepts / quantifier
  - How many large programs did you write?
  - Disambiguating may lead to complexity
- Unfamiliar terms
  - W.r.t. respondent background
- False inference
  - A general question after a specific one may be interpreted regarding all but the specific topic

Memory problems

- Mismatch between terms used in question and those used for encoding
- Distortion in the representation of the events over time
  - Post-event information from rehearsal
- Retrieval failure
- Reconstruction errors
  - Filling missing pieces by projecting usual events or current situations
  - Telescoping

Recall accuracy



Estimation problems

- Behavioral frequency questions
- Recall-and-count
  - Remember specific incidents and total them up
    - Adjust upward to allow for forgotten ones
- Rate-based estimation
  - Recall the rate at which incidents typically occur
- Impression-based estimation

Judgment problems

- Attitude questions
- Start with a clearly defined view
- Bottom-up
  - Start from specific beliefs
    - recall-and-count
- Top-down
  - Derive a position from more general values
- Exact wording affects answers from respondents without clear view

Formatting issues

- Open-ended questions with numerical answer
  - Respondent tend to report round values
    - Percentages cluster around 50 and 100
- Closed questions with ordinal scale
  - Positivity bias
  - Detailed labeling improves reliability
  - Two-step branching improves reliability
- Closed questions with nominal scale
  - Primacy effect
  - Recency effect

Misreporting

- Sensitive questions
  - Intrusive question may prompt non response
  - Embarrassing question may prompt
    - Under-reporting of embarrassing behavior
    - Over-reporting socially desirable behavior
- Computer supported self-administered questionnaires reduce misreporting

Navigational errors

- Typical for paper-based self-administered questionnaires

GUIDELINES FOR WRITING QUESTIONS

Introduction

- Survey intro page
  - Stimulate participation
  - Provide context
  - Define terminology
  - Avoid bias

Bad Intro

Focus

- Avoid normative questions (
  - What should we do?
- Focus on descriptive questions
  - "What is happening in your environment?
- Ask explanatory questions
  - Why is this happening?

General principles

- Simplicity
- Clarity
- Effortlessness
- Responsiveness

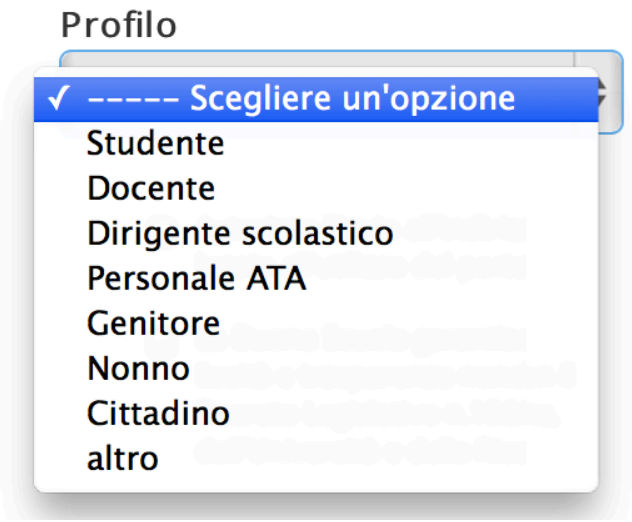
Question types

- Demographics
- Behavior
  - Non-sensitive
  - Sensitive
- Attitude

Demographics

- Measure all possibly significant (co)factors
- Adopt standard taxonomies and terminology

Counter-example



Profile
----Chose an option
Student
Teacher
Education manager
Staff
Parent
Grandparent
Citizen
other

Non sensitive behavior

- Wording of questions
  - in closed questions include all reasonable possibilities
    - Also improves precision
  - Make questions as specific as possible
    - E.g. time frame
  - Use words that virtually any respondent will understand
    - Avoid technical terms
    - Replace vague quantifies with explicit frequencies categories

Non sensitive behavior

- Lengthen questions including memory cues to improve recall
  - List of examples
- When forgetting is likely, use aided recall
  - Ask separate question for subcategories
- When long recall periods must be used, use a life event calendar to improve reporting
- To reduce telescoping errors, ask respondents to use records

Non sensitive behavior

- If cost is a factor consider whether proxies might be able to provide accurate information

Sensitive behavior

- Use open rather than closed questions
- Use long rather than short questions
- Use familiar words
- Load the question to reduce misreporting
- Ask about long periods
- Embed sensitive questions among others
- Use self-administered questionnaires
- Consider use of a diary
- Ask how much sensitive questions were
- Collect validation data

Loading

- Wording a question in such a way that invites a socially undesirable answer
  - Everybody does it
  - Assume the behavior
  - Authorities recommend it
  - Reasons for doing it

Loading examples

- Everybody downloads some amount of illegal music, how many songs you download per week?
- We expect you listen to illegal music as most people, how many illegal songs you download per week?
- Some studies suggest that listening to illegal music actually favor the later purchase of the legal version. How many illegal songs you download per week?
- The quality of songs in an album is diverse, so it makes sense to listen illegally before buying. How many illegal songs you download per week?

Attitude

- Specify the attitude object clearly
- Avoid double-barreled questions
- Measure the strength of the attribute
  - Possibly using separate items
- Use bipolar items
  - Except when they may miss key information
- Alternatives have impact on answers
  - Middle options should be included
    - Unless compelling reasons exists (e.g. elections)

Double barreled question

- Often derive from multifaceted constructs
  E.g.
  - How satisfied are you with your pay and job conditions?
  - Is this tool interesting and useful?
- Should be split in two distinct items addressing different measures

Attitude

- To measure change ask the same question each time
- Ask general questions before specific ones on a topic
- When asking questions about multiple items start with the least popular ones

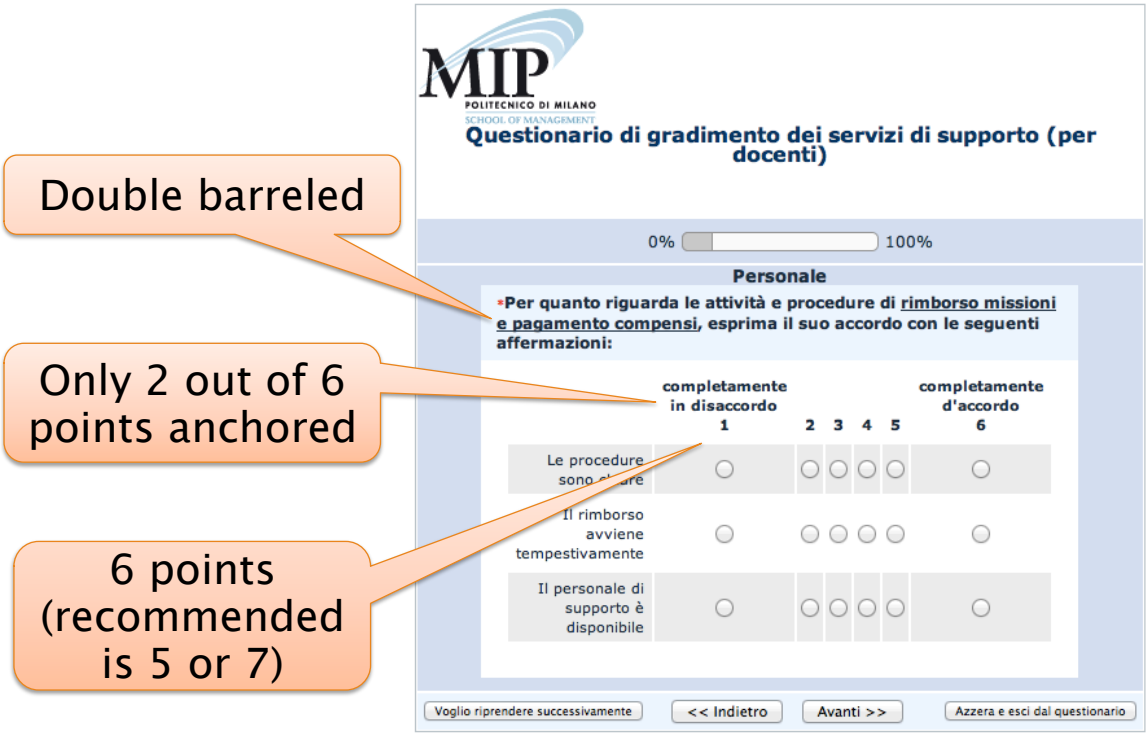
Attitude

- Likert scale
  - A statement which the respondent is asked to evaluate the level of agreement or disagreement is measured
  - Present a symmetry of categories about a midpoint with clearly defined linguistic qualifiers
  - Typical
    - Strongly disagree
    - Disagree
    - Neither agree nor disagree
    - Agree
    - Strongly agree

Attitude

- Use closed questions to measure attitude
- Use five or seven points scale and label each point
- Start with the least popular end of the scale
- Use analogue devices to collect more detailed scale information (>7 points)
- Use rankings only if all alternative are visible to the respondent otherwise use pairwise comparisons
- Get ratings to every item of interest
  - Do not use check-all-that-apply items

Counter-Example



Self administered

- Consistent use of visual elements
  - To define desired path through the questionnaire
- Alert reader when change conventions
- Directions should be placed
  - Where they are to be used
  - Where they can be seen
- Information to be used together must be placed in the same location
- Ask one question at a time!

PRIVACY

Privacy

- In Italy there is quite a strict law:
  - http://www.garanteprivacy.it/garante/document?ID=1219452
  - Section 7 provides a list of the rights of the subject,
  - Section 13 details the information to be provided to the subjects

Privacy

- Information to be provided
  - Purposes and modalities of the processing for which the data are intended
  - Nature of providing the requested data
  - Consequences of denial to reply
  - Entities or categories of entity of data communication and dissemination
  - Rights
  - Responsible for the data

Privacy information – Example

- Purposes and modalities of the processing for which the data are intended

- The data you provide will be handled for statistical and scientific purposes, aimed at investigating the details of software development. The handling will be carried on by electronic means.
  - Nature of providing the requested data
- The participation in the investigation is voluntary.
  - Consequences of denial to reply
- Denying to answer will have no consequence.
  - Entities or categories of entity of data communication and dissemination
- Personal data collected during the investigation will be shared only among the researchers involved in the project.

Privacy information – Example

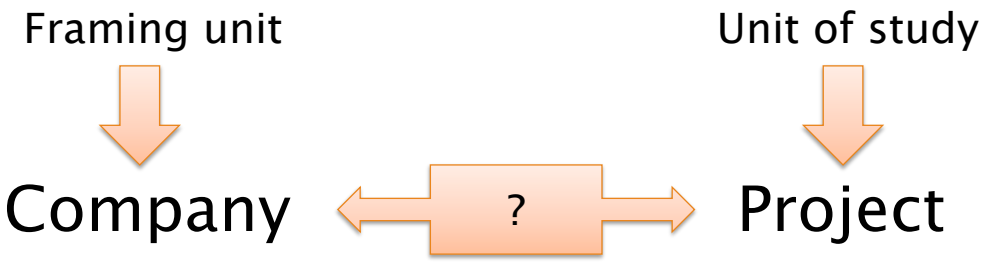
- Rights
  - At any time you will be able to exert your rights with the responsible for the data handling, according to section 7 of D.lgs. 196/2003, which we copy integrally:
  - …
- Responsible for the data
  - The responsible for data treatment is …

SOFTWARE ENGINEERING SURVEYS

Examples from “my” surveys

- Development with Off-The-Shelf components (2004)
- Success factors in software development projects (2006)
- Software migration projects (2007)
- MDD adoption (2011)

Framing and sampling



Which are the relevant companies?

How to get a software project given a company?

Framing

NACE

A – Agriculture

…

D Manufacturing

D 5x.xx : Electrical and electronic devices

…

J 72.0.1 : Consulting and installation of computer based systems

J 72.0.2 : Software implementation

…

Apple, Motorola, Nokia

Official classification

- ICT sector (OECD)
  - 3x.xx ICT Industry
    - Apple, Motorola, Nokia, etc.
  - 5x.xx ICT Trade
    - Mediaworld, Computerland etc.
  - 64.2x ICT Telecom
    - AT&T, BT, Telenor etc.
  - 72.xx ICT Software,
    - Vendors: Oracle etc.
    - Consultants: Accenture, Cap Gemini etc.

Framing by Activity

- ICT Contents sector
- Software development outside the ICT sector
  - Finance
  - Engineering
  - Automotive

Multiple framing instruments

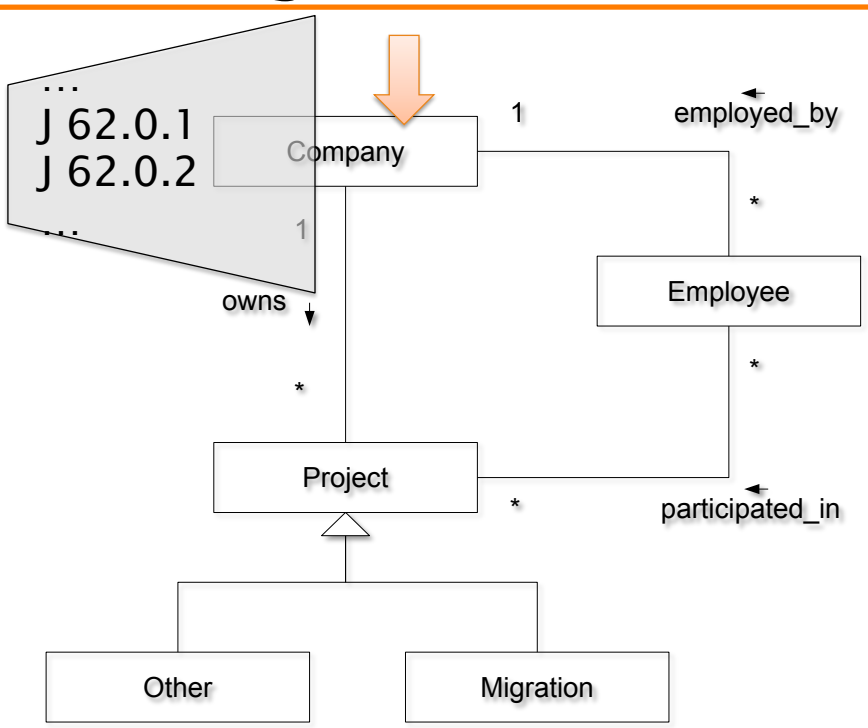
- Convenience subset of companies among research partners
- Largest ICT companies
- Largest non-ICT companies
  - IT departments
- Members of ICT-related industrial organizations
- Public lists of companies,,
  - From public sources, e.g. census bureaus or yellow pages

Company size

- Median number of employees = 1
- Mean employees in ICT ~ 7
- Companies with more than 100 employees = 3%

- Is a uniform coverage of company sizes advisable?

Indirect Framing



Contacting

- External contact
  - switchboard operator
- Gatekeeper a contact person
  - IT manager
- Respondent
  - The person providing the information
  - Should be provided by the gatekeeper
  - Often it is not the unit of study!

Self exclusion

- Are we interested in the target type of project only?
  - Prevalence of target project type
- If we provide too much detail about the target project type people not involved (interested) opt out of the study without providing any information

From Respondent to Project

- Respondent reports about a project
- Which project to select?
  - Respondent's decision
    - Bias towards successful cases
    - Good recall
  - Latest
    - Completed or not?

Bibliography

- Groves, Fowler, Couper, Lepkowski, Singer and Torangeau, 2009. "*Survey Methodology – 2nd edition*" John Wiley and Sons
- Conradi R., Li J., Slyngstad O. P. N., Kampenes V. B., Bunse C., Morisio M., Torchiano M. "*Reflections on conducting an international survey of CBSE in ICT industry*" IEEE 4th International Symposium on Empirical Software Engineering November, 2005

Off-The-Shelf Development

- Li J., Conradi R., Slyngstad O. P. N., Torchiano M., Morisio M., Bunse C. "*A State-of-the-Practice Survey on Risk Management in Development with Off-The-Shelf Software Components*", IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, Vol. 34(2), pp. 271-286, ISSN: 0098-5589, 2008
- Li J., Conradi R., Bunse C., Torchiano M., Slyngstad O. P. N., Morisio M. "*Development with Off-The-Shelf Components: 10 Facts*" IEEE SOFTWARE, Vol. 26(2), pp. 80-87, ISSN: 0740-7459, 2009
- Conradi R., Li J., Slyngstad O. P. N., Kampenes V. B., Bunse C., Morisio M., Torchiano M. "*Reflections on conducting an international survey of CBSE in ICT industry*" IEEE 4th International Symposium on Empirical Software Engineering November, 2005

Success factors

- Egorova E., Torchiano M., Morisio M. "*Actual vs. Perceived Effect of Software Engineering Practices in the Italian Industry*" THE JOURNAL OF SYSTEMS AND SOFTWARE, Vol. 83(10), pp. 1907-1916, ISSN: 0164-1212, 2010

Migration

- Torchiano M., Di Penta M., Ricca F., De Lucia A., Lanubile F. "*Migration of Information Systems in the Italian Industry: a State of the Practice Survey*" INFORMATION AND SOFTWARE TECHNOLOGY, Vol. 53 (1), pp. 71-86, ISSN: 0950-5849, 2011

Exercise

- Each group read the assigned paper and identify
  - Research question
  - Target population
  - Frame population
  - Sampling approach

Target population

- Avoid restricting the target population based on factors such as availability
- The framing instrument is tightly linked to the research questions
- Select the appropriate variables to characterize the units of analysis