

# Data Collection: Sampling and Data - from instruments to data

Feb 26 Morning in Pavlodar KZ

Alex Liu, Ph.D.

Director of the RM Institute

Los Angeles, California, USA

Alex@ResearchMethods.org

# Part I

- Sampling
- 9:30 ~ 11:00

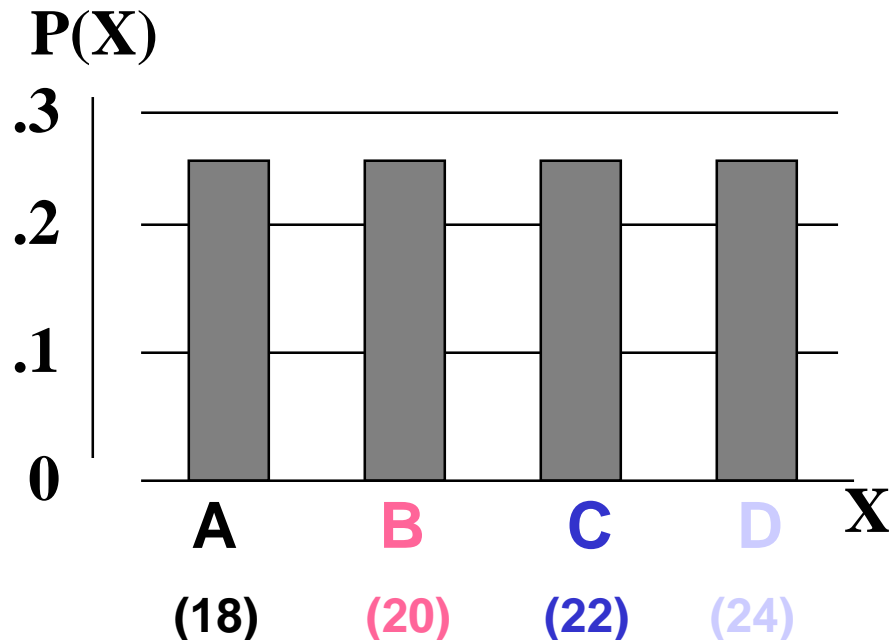
# Logic of Sampling

- Representativeness (sample is similar to the population under study)
- Probability of selection (equal chance of being selected)
- Sampling distribution
- Sampling error
- Confidence levels (intervals)

# Comparing the Population with its Sampling Distribution

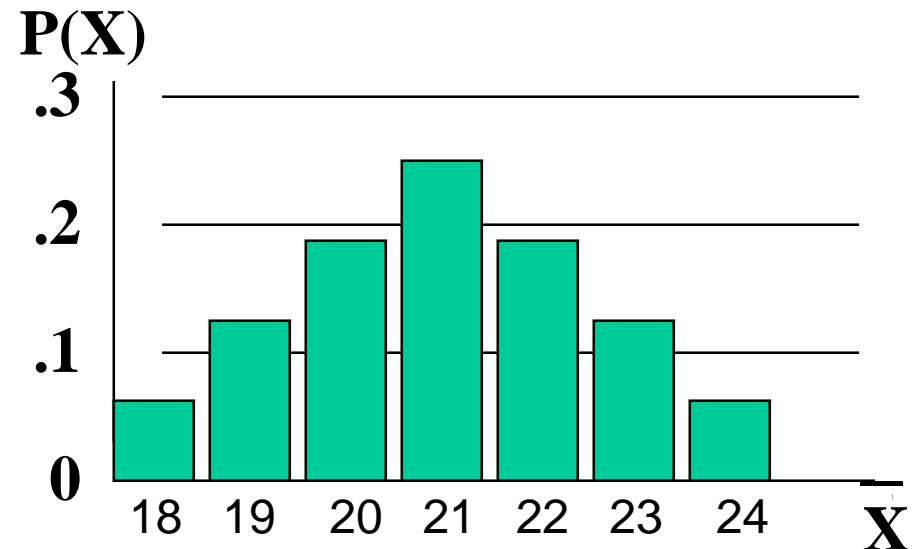
Population

$$\mu = 21, \quad \sigma = 2.236$$



Sample Means Distribution  
 $n = 2$

$$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$$



# Nonprobability sampling

- **Quota sampling - Quota Sampling**
  - Completing quotas - usually in terms of certain demographics such as sex or age band of desired numbers of respondents (sample cases) proportional to their population frequencies. Hence arriving at an approximation of a representative sample of the population **BUT ONLY** in terms of the specified quotas.  
GEM London
- **Snowball sampling**
  - start with a few, then rely on their help to locate others.
  - often used to study homeless, migrants, ...

# Probability Sampling: Definition

- **Probability sampling** is any method of sampling that utilizes some form of *random selection*.
- In order to have a random selection method, you must set up some process or procedure that assures that the different units in your population have equal probabilities of being chosen.
- Simple random sampling, stratified random sampling and cluster sampling are all probability samples, but utilise different procedures.
- by GEM London

# Types of probability sampling

- Simple random sampling
- Stratified
- Cluster

# Simple Random Sampling

- **Simple Random Sampling**
- Basic sampling technique where a group of subjects (a sample) are selected for study from a larger group (a population).
- Each individual is chosen entirely by chance and has an equal chance of being included in the sample.
  
- **Advantages**
  - Ideal for statistical purposes
- **Disadvantages**
  - Requires an accurate list of the whole population
  - The optimum sample is the one which maximises precision per unit cost, and by this criterion, simple random sampling can often be bettered by other methods.
  
- GEM

# Stratified Random Sampling

- **Stratified Random Sampling**

- The sampling frame is divided into non-overlapping groups or strata, e.g. regions. A sample is taken from each stratum, and when this sample is a simple random sample it is referred to as stratified random sampling.

- **Advantages**

- Stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest.
- The size of the sample in each stratum should be taken in proportion to the size of the stratum.
- Choose a larger sample for those strata with greater variability
- Ensures better coverage of the population than simple random sampling.

- **Disadvantages**

- Difficulty in identifying appropriate strata.
- More complex to organise and analyse results.

# Cluster Sampling

- **Cluster Sampling**

- Where units sampled are chosen in clusters, close to each other.
- Examples are households in the same street.
- The population is divided into clusters, and some of these are then chosen at random. Within each cluster, units are then chosen by simple random sampling or some other method. Ideally the clusters chosen should be dissimilar so that the sample is as representative of the population as possible.

- **Advantages**

- saving of travelling time, and consequent reduction in cost
- useful for surveying people in particular areas of interest

- **Disadvantages**

- units close to each other may be very similar and so less likely to represent the whole population
- larger sampling error than simple random sampling

# GEM Sampling Practice

- **Variety of sampling designs employed**
  - **Simple random sampling**
    - Handful of countries use this technique
  - **Stratified random sampling**
    - Over 70% of all telephone interviews conducted were stratified random designs
    - 60% of total samples collected employed stratified random sampling
    - Wide range of stratification units applied
      - e.g. age, gender, region, settlement size, ethnicity, monthly income, occupation, phone area code etc
      - some teams used only one strata, others used many (depends on country demographics and access to whole information)

# GEM Sampling Practice 2

- **Cluster sampling**

- used by nearly all countries that conducted face-to-face interviews

- **Quota sampling**

- Very few countries still using this method

# GEM Sampling Requirements

- **In 2007 the GEM co-ordination team will:**
- **Enforce the minimum requirements set out in the RFP**
  - Each country **MUST** submit a full survey proposal to the GEM co-ordination team prior to commencement of their survey, without exception
  - The RFP provides a detailed example of all requirements at every stage
  - Your proposal must address each and every item
- **e.g. sample frame, population data and distributions, distributions for each stratified unit per strata, coverage, size of clusters and all associated numbers, detailed selection method of respondent, number of call backs etc**
- If the proposal does not meet the RFP requirements, it will be necessary for you must work with your vendor to revise it

Sampling frame is the list or quasi list of elements of which a probability sample is selected.

## Part II

- Data collection NES & APS
- 11:20 ~ 12:30

# GEM Data Collection Practice

- **Forty-two** countries in GEM 2006
- Variety of interview methods used:
  - Telephone
    - Fixed 24
    - Fixed and mobile 4
    - Mobile only 1
  - Face-to-face 10
  - Telephone and Face-to-face 3

# Telephone vs. face-to-face

- Advantages of telephone survey: low cost, fast, interviewers work in comfort & can get easy support
- Advantages of face-to-face survey: may see some characteristics out of what covered by the questionnaire, may help control response quality

# Return rate

- $> 60-70$  is considered as good.
- GEM used “frequently call back” to increase return rate
- Face-to-face interviewing can increase return rate

# Professional Interviewing

- Dress professionally
- Familiarity with the questionnaire
- Follow question wording EXACTLY
- Record responses EXACTLY
- Follow the skip pattern correctly

# GEM Documentation 1

- The after-survey results template (in the RFP) must be completed IN FULL.
- This includes completion of the set of descriptive items in the top half of the “Fieldwork Report”, found in the RFP, as well as all the after-survey interview numbers requested.
- In 2006, many of these numbers were omitted, not recorded or misinterpreted by the vendor, which took a lot of time to resolve.
- Please reiterate to your vendor **at the start of the survey**, which figures they should be recording. If any clarification of the definitions is required, please ask.
  - As much as we appreciated the efforts that many teams made in delivering extra information and break-downs of interview results in great detail, we don’t require this data.
  - Breaking down figures into further categories than listed on the sheet in the RFP takes you time, and interpreting what is often local terminology to add them back up again takes us time. So submitting the sheet from the RFP is adequate.

# GEM Documentation 2

- Acknowledging the shortcomings of any survey is vital to it's future improvement, and why we are publishing a technical report this year, outlining what we have achieved and what we aim to achieve in the future
- GEM London will expand on the importance of the documentation process and propose ways that the results we document can be improved. This of course hinges solely on the quality of the survey itself.
- Handout documentation sheets, summary of sampling characteristics and timeline.

# Some practical issues in sampling

## - some common problems

- Latvia experience

# Some practical issues in data collection

- Latvia experience