

# Session 7.2

## Introduction to Quantitative Research Methods

- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Questionnaire Design
- 7 Weighting the sample

- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Questionnaire Design
- 7 Weighting the sample

## 1: Things to consider

- Why is the data being collected?
- What is the research question?
- Consider data requirements
- Consider what factors are of interest
- Define the population of interest
- Arrive at a judgement of what is feasible
- Select the sample & calculate the sample size
- Define the data collection procedure
- Is there bias in the process?

- What is the objective of the study?
- Identify the null and alternative hypothesis:
  - The null hypothesis,  $H_0$ , represents a theory that has been put forward but has not been proved.
  - The alternative hypothesis,  $H_A$ , relates to the statement to be accepted if/when the null hypothesis is rejected.

## Defining the population of interest

- Think about the population of interest
- Can the population be defined?
- Is there a sampling frame?
- How can subjects be identified?
- If physical objects, e.g. core samples, vehicles, road surfaces, roads etc. are they readily available?

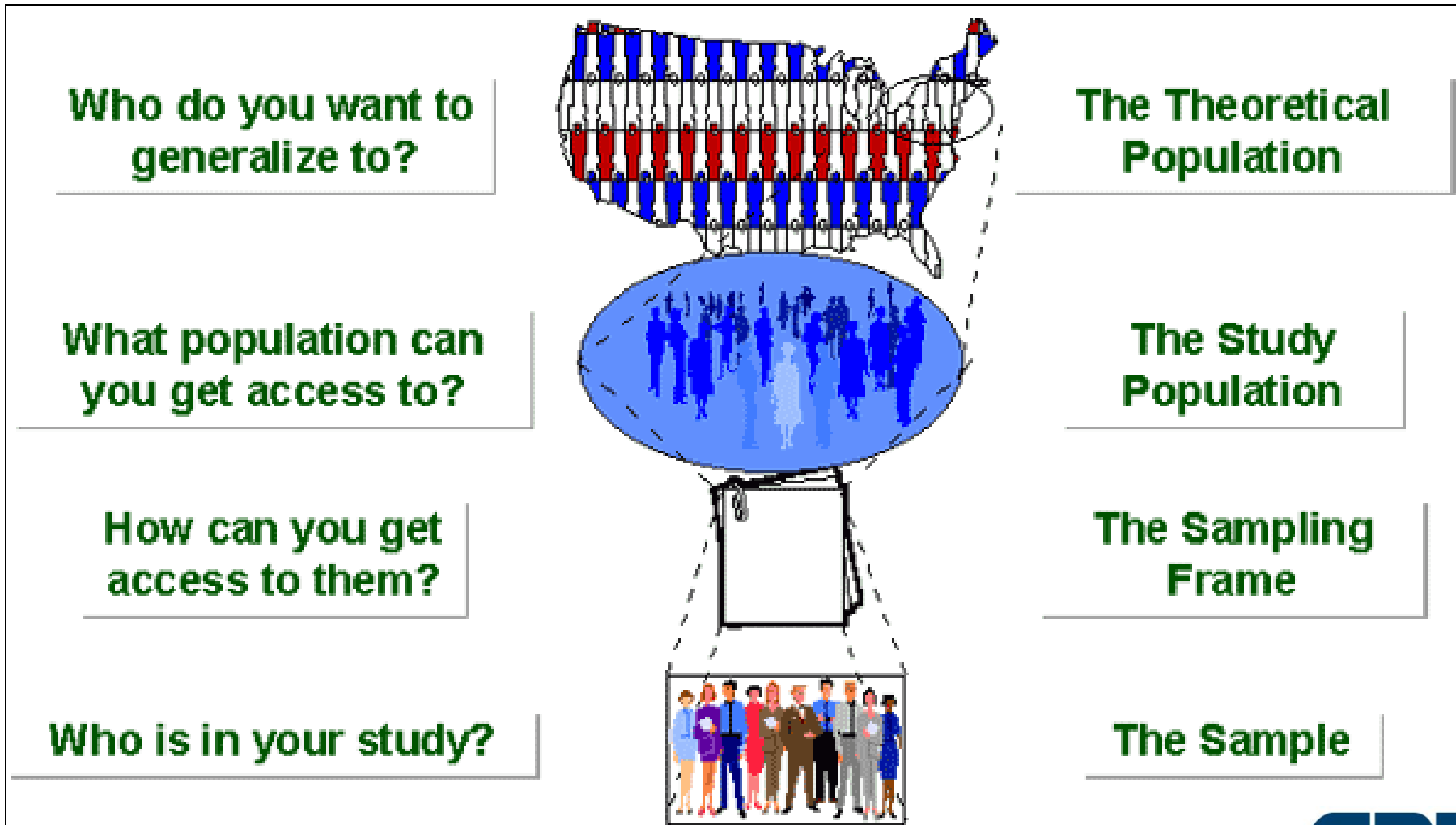
- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Data storage & validation
- 7 Weighting the sample

## 2: What is a sample?

- Sampling is counting in the real world!
- It is not a science, it is not precise
- Sampling gives an insight into what is really happening
- We assume that:
  - Some data is better than no data – but it can be improved on (sometimes at great expense)

- Is there a sampling frame, i.e. a population identifiable and contactable?
- Is the sample unit representative of the population of interest?
- Is there information about the population which could be used in selecting the sample?
- What sampling process will be appropriate?

- **Population (of interest):** The set of units that you hope to investigate
  - e.g. people in Arusha
- **Accessible population:** The population that is accessible to you
  - e.g. people in the phone book
- **Sampling frame:** Procedure for sampling accessible population (this may be a list)
  - e.g. the phone book
- **Sample:** The group of units who you select to be in your sample
  - e.g. a simple random sample from the phone book



## Error and bias

- Access to whole population?
- Complete and accurate sampling frame?
- Sampling bias
- Sample does not respond or drop out of study

**a** Simple random

**b** Systematic

**c** Stratified

**d** Cluster

**e** Quota

**f** Multi-stage

### Summary

- Each unit has an equal chance of being selected
- Suitable for small sample with complete sampling frame

### Advantages and disadvantages

#### Advantages:

- Simple

#### Disadvantages:

- Not statistically efficient
- Poor representation of subgroups possible

### Procedure

1. Select population and sampling frame
2. Give each unit of population a unique number
3. Calculate/guess your sample size ( $n$ )
4. Retrieve  $n$  unique numbers from your random number table

### Summary

- Randomly order the population
- Select every  $n$ th unit

### Advantages and disadvantages

#### Advantages:

- Easy and efficient

#### Disadvantages:

- Vulnerable to periodicity (i.e. values repeat periodically)

### Procedure

1. Select population and sampling frame
2. Give each unit of population a unique random number and sort on this number
3. Write down your sample size ( $n$ )
4. Calculate your sampling fraction  $f = n/N$
5. Start at a random number in the list
6. Pick every  $f$ th value

### Summary

- Organise your distinct categories
- Randomly sample within each category
- May want to over sample small categories to ensure good estimates

### Advantages and disadvantages

#### Advantages:

- More economical

#### Disadvantages:

- May not equally represent all responses
- Lose variability information

### Procedure

1. Select population and sampling frame
2. Choose categories of your population to sample
3. Assess whether you need to over sample from small categories
4. Calculate your sampling fraction for each category
5. Sample  $n_1 * f_1$
6. Repeat for each category

### Summary

- Population dispersed over a large area? Reduce mileage by cluster analysis
- Split geographical region into areas and select a number  $p$  of them.
- Each unit of population within these  $p$  regions should be measured estimates

### Procedure

1. Select population and sampling frame
2. Split your population into small regions
3. Write down your sample size  $p$
4. Sample  $p$  regions using simple random sampling
5. Measure every unit within each sampled region

### Advantages and disadvantages

#### Advantages:

- More economical

#### Disadvantages:

- May not equally represent all responses

### Summary

- Generally used for street surveys
- Sample selection is made by an interviewer
- A quota is specified from a subset or set of subsets of the population

### Advantages and disadvantages

#### Advantages:

- Quick and cheap to organise

#### Disadvantages:

- Not necessarily random
- Possibly biased samples
  - more approachable people

### Procedure

1. Select population and sampling frame
2. Choose your subsets of population
3. Specify a quota for each subset
4. Interview or measure units until the quota is filled for a subset and continue until all other subset quotas are complete

### Summary

- Combine the sampling strategies above
- Similar to cluster sampling, but sample within your chosen sample
- For example, sample  $n$  administration regions, sample  $m$  streets in those wards and sample every  $p$ th house in each street

### Procedure

1. Select population and sampling frame
2. Split your population into small regions
3. Sample regions
4. Sample areas within regions
5. Sample subsets of areas within regions choosing the most appropriate sampling strategy each time

### Advantages and disadvantages

#### Advantages:

- Convenience, economy and efficiency

#### Disadvantages:

- Lower accuracy due to higher sampling error

- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Questionnaire Design
- 7 Weighting the sample

- If possible obtain estimates of the variability of the data
- Consider a pilot exercise
- Consider size of effect that is important
- Consider Type I & Type II errors
- Consider constraints, i.e. budget, time etc.

$H_0$  = Null Hypothesis of no effect

$H_1$  = Alternative hypothesis that there is an effect

Reality	Decision	
	Accept $H_0$	Reject $H_0$
$H_0$ is true	OK	Type I error
$H_1$ is true	Type II error	OK

Type I error is often set to 5%.

Type II error is often set to 10%.

A Type II error is frequently due to sample sizes being too small.

- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Questionnaire Design
- 7 Weighting the sample

There are two types of data source:

- **Primary data:**
  - Collected for the purpose of the experiment
  - Possibly only a small number of observations
- **Secondary data:**
  - Collected for a different purpose
  - May not directly answer the research questions
  - Meta-analysis – possibly a large dataset

## Analysing the data

- Define hypothesis to be tested, preferably as a null-hypothesis, i.e. the test will be to determine the probability of the null-hypothesis being likely
- Identify the likely statistical tests to be used
- Discuss with statisticians if unsure and/or if software not available/clear

- 1 Things to consider
- 2 Sampling
- 3 Sample size
- 4 Data Sources
- 5 Data collection
- 6 Questionnaire Design
- 7 Weighting the sample

- The data collection system should be well defined
- It should be representative of the population of interest
- Pilot a questionnaire, especially if it includes new questions
- Standardise the data collection and apply throughout study
- Always train data collectors

## Data collection by questionnaires or surveys

**a** Face to Face interview

**b** By phone

**c** By post

**d** On-line/email

### Face to face interview

- The interviewer and respondent communicate directly either where the interviewer selects passers-by in the street or in a respondent's home

### Advantages and disadvantages

#### Advantages:

- Complex structure
- Able to give explanations
- Completed consistently
- Higher response rate
- All questions likely to be completed

#### Disadvantages:

- Expensive/requires data entry
- Respondent may not answer truthfully

### Telephone interview

- A cheaper alternative to face-to-face interviews

### Advantages and disadvantages

#### Advantages:

- Complex structure
- Able to give explanations

#### Disadvantages:

- Bias – time of call, ex-directory households
- Requires data entry

### Postal

- The questionnaire is sent out for self-completion and returned by post

### Electronic

- The questionnaire is on-line or sent via email

### Advantages and disadvantages

#### Advantages

- Cheaper than interviews
- More likely to answer truthfully as anonymous (e.g. a questionnaire on drugs/alcohol)

#### Disadvantages

- Low response rate
- May not be representative of the population
- Requires data entry

### Advantages and disadvantages

#### Advantages

- Cheap & fast
- Data accuracy – automatic validation checks
- Flexible design
- May be more likely to return than if by post

#### Disadvantages

- Bias – access to internet, less control over sample population

## Bias

- A statistic is **biased** if it is calculated in such a way that it is systematically different from the population parameter of interest.
- Types of bias:
  - **Selection bias**, where individuals or groups are more likely to take part in a research project than others, resulting in biased samples.
  - **Spectrum bias** arises from evaluating diagnostic tests on biased samples, leading to an overestimate of the sensitivity and specificity of the test.
  - **Omitted-variable bias** appears in estimates of parameters in a regression analysis when the assumed specification is incorrect, in that it omits an independent variable that should be in the model.
  - **Funding bias** may lead to selection of outcomes, test samples, or test procedures that favour a study's financial sponsor.
  - **Reporting bias** involves a skew in the availability of data, such that observations of a certain kind may be more likely to be reported and consequently used in research.
  - **Data-snooping bias** comes from the misuse of data mining techniques.
  - **Analytical bias** arises due to the way that the results are evaluated.
  - **Exclusion bias** arises due to the systematic exclusion of certain individuals from the study.

## Addressing bias

- Carefully design surveys which aim for a random sample of the population
- Recognise that bias is a risk not a certainty
- Sampling is inevitable – there is too much to count properly
- Undertake a pilot and adjust approach if necessary

1 Things to consider

2 Sampling

3 Sample size

4 Data Sources

5 Data collection

6 Questionnaire Design

7 Weighting the sample

## 6. Questionnaire design

- Use simple, short questions on the recent past
- Keep the length of questionnaire short
- Ask general questions first followed by more specific questions
- Open questions are difficult (and costly) to analyse; consider closed questions with multiple choice answers
- Ensure all possibilities are included in multiple choice questions
- Do not ask leading questions or have a biased set of multiple choice answers
- Ensure any ranges are non-overlapping
- **Ensure that the questions are disaggregated by gender and age**

- Consider what data you need to collect to address the research question/hypothesis
- Conduct checklist interviews
- Conduct a pilot survey with a small sample
- Survey enumerators should be trained on how to complete the questionnaire with respondents and provided with comprehensive guidance notes
- Each questionnaire should be numbered and contain the following information:
  - Date of survey
  - Location of survey: settlement name, road number and district/province
  - Name of interviewer and interviewee (if they will provide it)

- Key informants can provide information on the socio-economic characteristics of the community
  - Population and number of households
  - Distance to roads and services
  - Local economy
  - Condition of local roads
  - Provision of transport services
- Key informants include but are not limited to
  - School teachers
  - Health workers
  - Religious leaders
  - Local government officials or village elders

## Key informant questionnaire

Rural Transport Survey: Village Level Questionnaire

Interview start time..... Finish time .....

Questionnaire No..... Date.....

Village Name..... District.....

Interviewer.....

No. Km from junction with secondary or main road.....

### 1. Village Demography

1.1 Number of households..... Village population.....

1.2 No. of men<sup>a</sup>..... No. of women<sup>a</sup>..... No. of children<sup>b</sup>.....

<sup>a</sup> 18 years old or over

<sup>b</sup> Under 18 years old

### 2. Village Topography

1.1 Describe the village terrain: (tick one of the following)

Flat..... Rolling..... Hilly..... Mountainous.....

1.2 Describe the vegetation cover around the village: (tick one of the following)

Dambo (marsh)..... Open..... Forested.....

Grassland ..... Semi-arid .....

**3. What is the distance to the nearest educational facilities?** (If educational facility in village, write 0)

Type	Primary	Secondary	Technical	Kindergarten	Other
Km					

**4. What is the distance to the nearest health facilities?** (If health facility in village, write 0)

Type	Health post	Clinic	Hospital	Other
Km				

4.1 a) Are there any vehicles that come to the village? Yes / No

b) What type are these vehicles?.....

c) How frequently do they visit the village.....(week / month / year )

- The household questionnaire should be administered to the head of the household
- The questionnaire is designed to obtain the following information:
  - Household size and composition
  - Household livelihood and income
  - Agricultural marketing activities: crops and food, livestock and dairy
  - Employment
  - Transport requirements: trip frequency, duration, distance, fare, age/gender, mode
  - Trip satisfaction
  - Ownership and use of means of transport
- Tabulate questions to facilitate quick and straightforward responses

5 What are your transport requirements outside the village:

Destination	Roughly how many journeys ( <i>round trips</i> )? State whether per day, week, month or year.	Time taken for <i>one way trip</i>	Distance for one way trip (Km).	Who makes these trips? Tick whether men (M), women (W), boys (B) and/or girls (G)				Mode of transport (See Key)
				M	W	B	G	
Travel to market: to sell								
To buy								
Travel to grinding mill								
Education: Primary								
Secondary								
Transport of harvest								
Health – travel to: Health post								
Local clinic								
Hospital								
Social visits (weddings, funerals, visits to friends and relatives)								
Religion								
Travel to employment								
Travel to farms								
Use of post office/ public telephone								
Collection of farm inputs								
Other .....								

Key: 1 walk, 2 Bicycle, 3 Motorcycle, 4 Car, 5 Bus, 6 Truck, 7 Boat, 8 Ox-drawn cart, 9 Tractor, 10 Other, 11 Delivered

# AFCAP Transport operator questionnaire



- Transport service operators – operating minibuses, pickups and intermediate means of transport (IMTs) such as motorcycle taxis
- The questionnaire is designed to obtain the following information:
  - Vehicle type, maximum capacity, actual number of passengers carried, goods load
  - Vehicle operations: lease and ownership of vehicles
  - Vehicle operating costs and maintenance
  - Vehicle utilisation
  - Vehicle age and condition
  - Use of feeder roads and problems experienced



## Transport operator questionnaire

1 Is the driver a member of an association? Yes / No

### Vehicle operating costs and maintenance

2 Does the driver have to pay to park the vehicle? Yes / No

3 If so, how much does he have to pay? ..... frs (CFA)

4 What fuel does the vehicle use (Circle appropriate)?      Petrol / Diesel

5 What is the cost *per litre* of fuel? ..... frs (CFA)

6 How much is spent on fuel? ..... frs (CFA) per day / week /  
month

7 How much is spent on maintenance? ..... frs (CFA) per day / week /  
month / year

8 How regularly is maintenance carried out?

i) Daily .....      ii) Weekly .....      iii) Monthly .....

iv) When the vehicle develops a mechanical problem .....

v) Other (Please state) .....

### Vehicle utilisation

9 How long ago was the vehicle last out of service? ..... days / weeks /



- A travel diary can be used to obtain more accurate travel data of every household member among a larger sample of households in the community
- ‘Live’ trip making can be captured by a large sample of households across the community for the same period of time
- The travel diary collects information on:
  - Details about the respondent: relationship to head, gender, age, occupation
  - Origin/destinations, trip purpose, trip duration, mode of transport
  - Household access to vehicles/means of transport
  - Income earning trips for one week
- The respondent should describe the trips to the enumerator when the travel diary is collected for validation

Travel Diary	HH Survey No:			HH member Name:		Day of the Week:	
	No:			No:		Date:	
TIME OF DAY	Before 6 am	6 am - 9 am	9 am - 12am	12 am - 3 pm	3 pm - 6 pm	6 pm - 9 pm	9 pm - 12 pm
<b>TRAVEL WHERE</b>							
<i>Within locality</i>							
<i>Within area</i>							
<i>Outside area</i>							
<i>Long distance</i>							
<b>TRAVEL PURPOSE (list all in order of importance)</b>							
<b>INCOME-EARNING (None=0, or list income source)</b>							
<b>TRAVEL MODE</b>							
<b>TRAVEL TIME (minutes)</b>							
<b>TRAVEL WITH WHO</b>							
Family Member/s (list numbers)							
Non-Family (how many people)							
Alone							
<b>FURTHER REMARKS</b>							



- Sensitisation
- Obtaining informed consent
- Participant confidentiality and anonymity
- Protection of personal data
- Manage expectations
- Providing incentives for participation
- Avoid 'survey fatigue'

1 Things to consider

2 Sampling

3 Sample size

4 Data Sources

5 Data collection

6 Questionnaire Design

7 Weights

## 7: Weights - 1

**IF**

- You applied a non proportional sampling fraction to some of your categories

**OR**

- The data you have collected are not distributed similarly to the population

**AND**

- You want to generalise to the whole population

**THEN**

- You might need to apply weights before analysing



## Weights - 2

- If population and sample are not similarly distributed then you may wish to apply some weights
- Weights help to redistribute the sample so that the population distribution is appropriately represented
- Weight =  $\frac{\text{proportion of population in category}}{\text{proportion of sample in category}}$



Now read  
Session 7.2  
notes!