



# Session 6.6: Data Structure, Exploration and Description

1 Data Types and Levels of Measurement

2 Data Structure and Sources

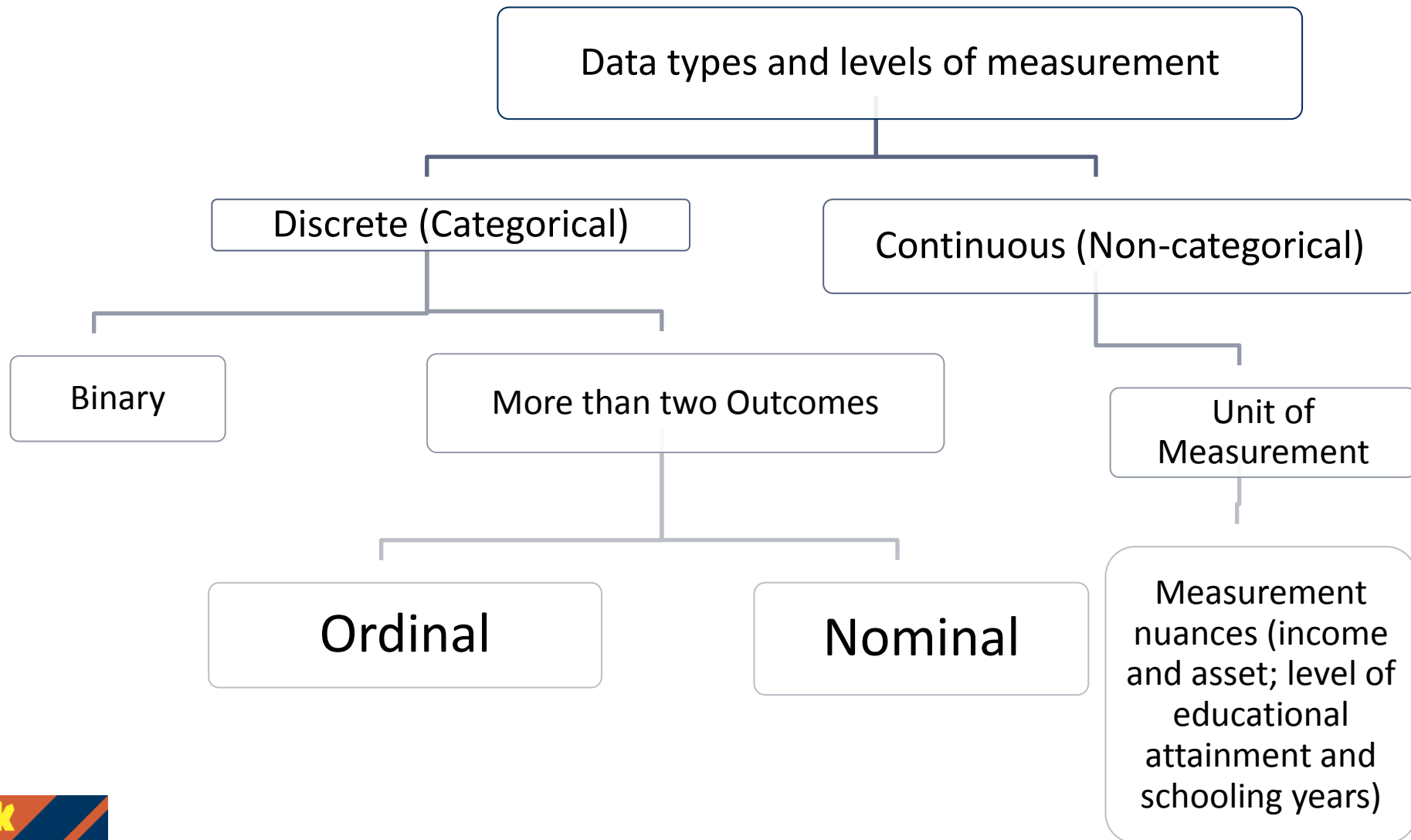
3 Data Exploration and Management

1 Data Types and Levels of Measurement

2 Data Structure and Sources

3 Data Exploration and Management (Practical Session)

## Data Types and Levels of Measurement



1 Data Types and Levels of Measurement

2 Data Structure and Sources

3 Data Exploration and Management (Practical Session)

- Data Structures
  - Cross Sectional (n – varies with t– constant)
    - Determinants of student’s CGPA in 2013
  - Time Series (t- varies with n- constant)
    - Determinants of Ghana’s GDP post independence
  - Panel (both n and t varies)
    - Cross country analysis of transport revenue over the period 2005 to 2010

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

- Positive Attributes
  - Great learning curves
  - Control over choice and measurement of variables
- Challenges
  - Need to validate descriptive statistics with national level information
    - Exploratory studies are exempted
  - Did your sampling deal with usual concerns
    - Representativeness based on both size and reflection of population characteristics?
    - Weights
    - Inflation

- Secondary Data
  - Ghana specific – Household
    - Ghana Living Standard Survey (GLSS) [five rounds, sixth round on-going]
    - World Bank Development Indicators
      - <http://data.worldbank.org/data-catalog/world-development-indicators>

- 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

## 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

- 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>							



**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 =  $\frac{\text{proportion of population in category}}{\text{proportion of sample in category}}$

1 Data Types and Levels of Measurement

2 Data Structure and Sources

3 Data Exploration and Management (Practical Session)

- Any observation that can take on more than one value.
- *Dependent variable* – The supposed effect in an experimental study. The effect is dependent on another variable.
- *Independent variable* – The supposed cause in an experimental study. All other variables that may affect the dependent variable are controlled.
- *Moderating Variable* – A variable that contributes to or influences the relationship between a dependent and independent variables.

- *Intervening Variable* - These are variables that occur between the dependent and independent variables. The relationship between income and longevity could be mediated (other name of intervening) by medical care.
- *Confounding variable* – A variable that obscures the effect of another variable.
- *Latent Variable* – An unobservable underlying variable.
- *Manifest Variable* – An observed variable assumed to indicate the presence of a latent variable.



Now read  
Session 6.6  
Notes!