Seminar 5 Data Collection, Preparation and Analysis Using SPSS

By Dr. Muhammad Ramzan
mramzaninfo@gmail.com,
03004487844

Edited by Ahsan Khan Eco
ahsankhaneco@yahoo.com
03008046243

Data Collection-Methods

- Data collection method is impacted by the method of research you choose. It is usually done through:
  - Data collection by the individual researcher

  - Collection through hired researchers

  - Collection through firms

Data Collection-Formats

- Format of data is influenced by the method of research, as it could be
- Printed questionnaires
- Interview sheets (in-person or telephonic)
- Focus group notes
- Observation notes
- Email or web responses
- Content analysis notes, pictures, documentaries
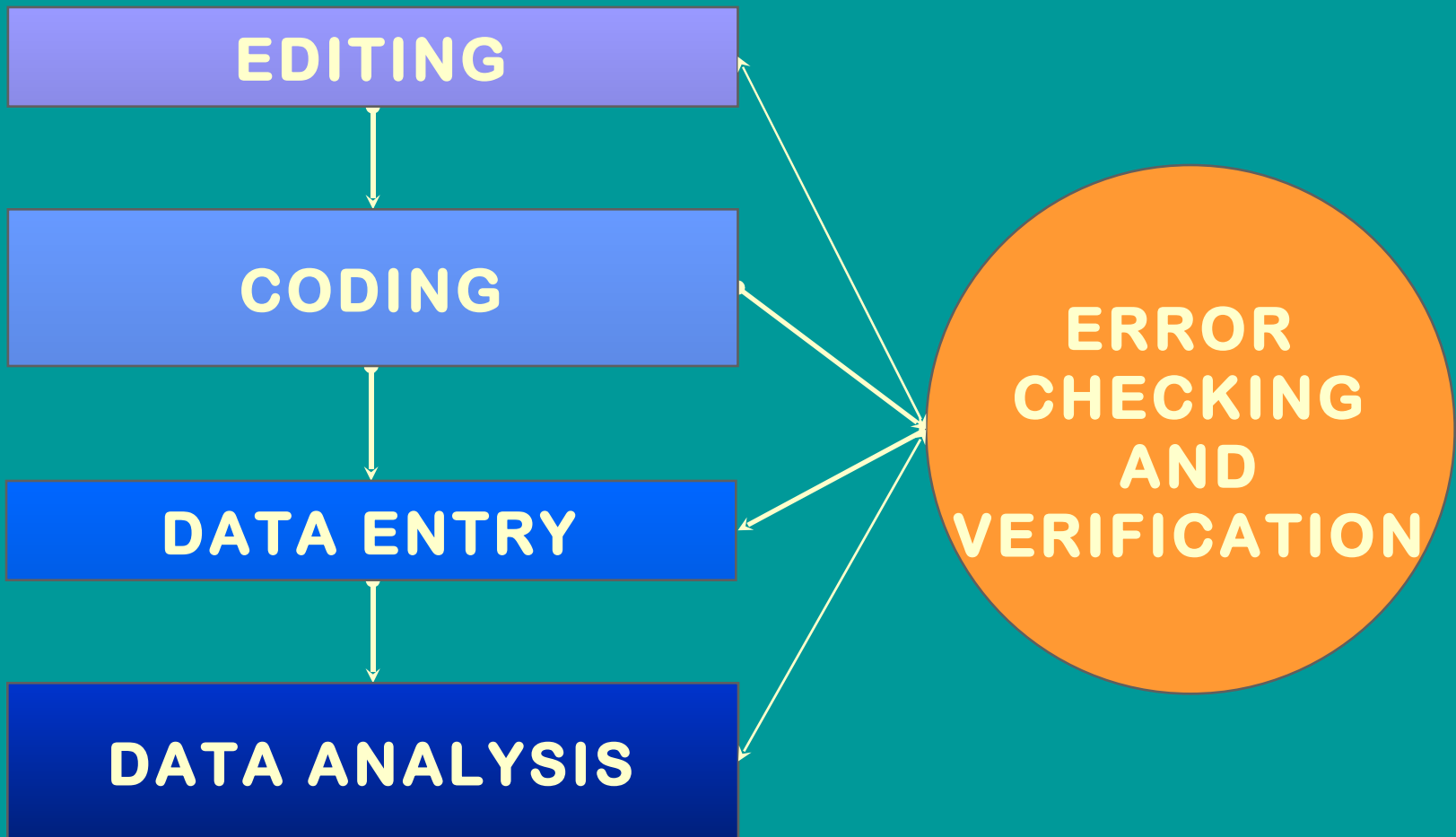- Printed, e-records, scanned data
- Literature review

Data Preparation

- Preparation of data file

- It is important to convert raw data into a usable data for analysis

- The analysis and results will surely depend on the quality of data

- There are possibilities of errors in handling instruments, raw data, transcribing, data entry, assigning codes, values, value labels

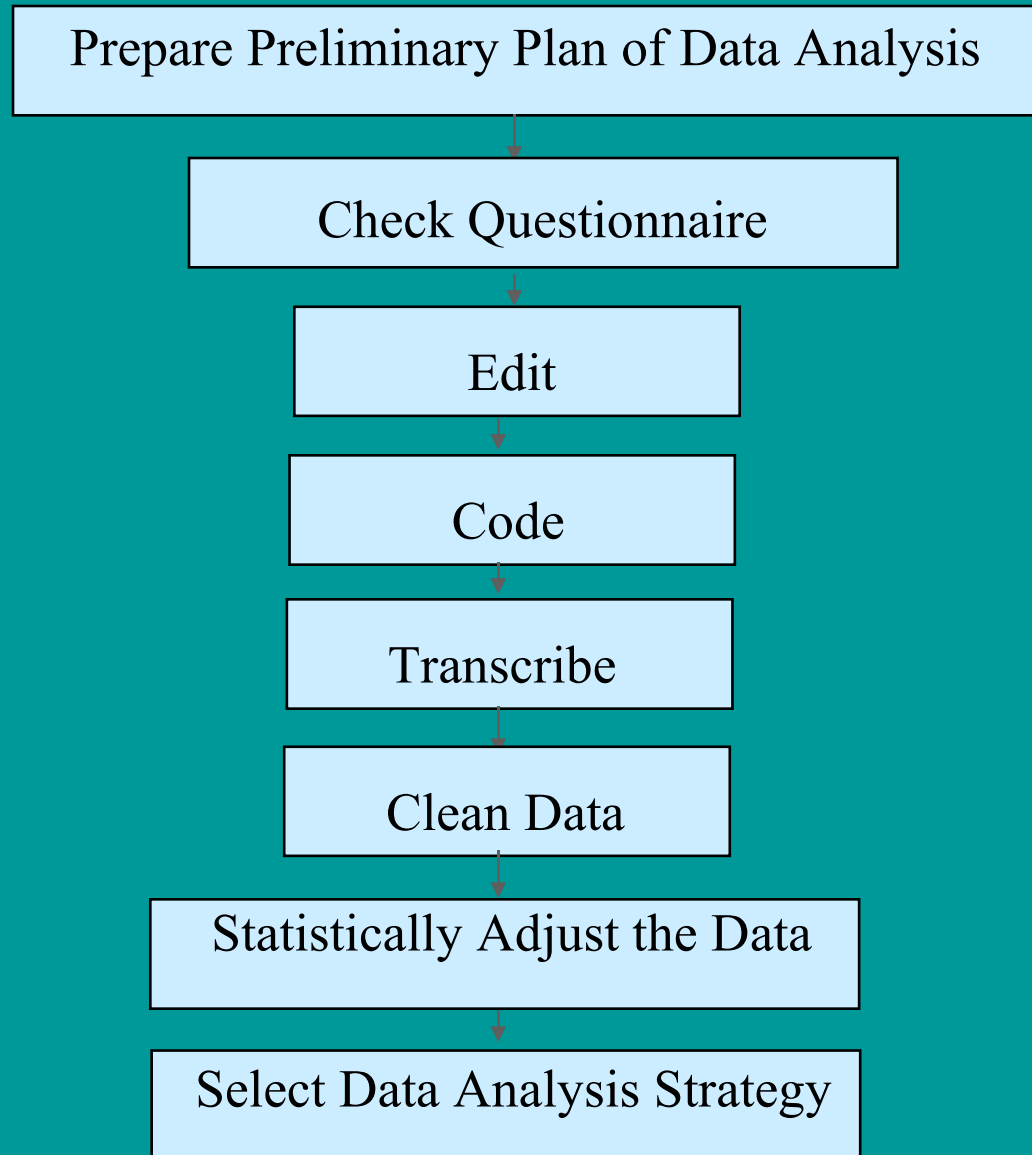- Data need to be cleaned to fulfill the analysis conations

Data Analysis

- Analysis of data is influenced by a number of factors. They are but not limited to:
- The purpose of research
- The type research questions and hypothesis
- The method of research and format of data
- Use of software for management, manipulation and analysis of data
- Researchers skills and capabilities
- Techniques used for data
- The quality of the data

Data Preparation Process

Prepare Preliminary Plan of Data Analysis

↓

Check Questionnaire

↓

Edit

↓

Code

↓

Transcribe

↓

Clean Data

↓

Statistically Adjust the Data

↓

Select Data Analysis Strategy

Questionnaire Checking

A questionnaire returned from the field may be unacceptable for several reasons.

- Parts of the questionnaire may be incomplete.
- The pattern of responses may indicate that the respondent did not understand or follow the instructions.
- The responses show little variance.
- One or more pages are missing.
- The questionnaire is received after the pre-established cutoff date.
- The questionnaire is answered by someone who does not qualify for participation.

Questionnaire Checking

We need to find valid questionnaires for data analysis

Each questionnaire/response need allotment of a case number for future reference

Questionnaire/response need filing in an order for retrieval and verification

Editing of Responses

Treatment of Unsatisfactory Results

- **Returning to the Field** – The questionnaires with unsatisfactory responses may be returned to the field, where the interviewers re-contact the respondents.

- **Assigning Missing Values** – If returning the questionnaires to the field is not feasible, the editor may assign missing values to unsatisfactory responses.

- **Discarding Unsatisfactory Respondents** – In this approach, the respondents with unsatisfactory responses are simply discarded.

Reasons for Editing

CONSISTENCY

COMPLETENESS

QUESTIONS ANSWERED OUT OF ORDER

Editing

- The process of checking and adjusting the data
  - for omissions
  - for legibility
  - for consistency
- And readying them for coding and storage
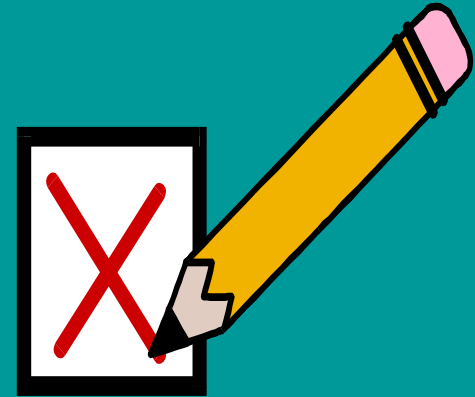
Codes

- The rules for interpreting, classifying, and recording data in the coding process
- The actual numerical or other character symbols

Coding

Coding means assigning a code, usually a number, to each possible response to each question. The code includes an indication of the column position (field) and data record it will occupy.

Coding Questions

- **Fixed field codes**, which mean that the number of records for each respondent is the same and the same data appear in the same column(s) for all respondents, are highly desirable.

- If possible, standard codes should be used for missing data. Coding of structured questions is relatively simple, since the response options are predetermined.

- In questions that permit a large number of responses, each possible response option should be assigned a separate column.

Coding

Guidelines for coding unstructured questions:

- Category codes should be mutually exclusive and collectively exhaustive.

- Only a few (10% or less) of the responses should fall into the "other" category.

- Category codes should be assigned for critical issues even if no one has mentioned them.

- Data should be coded to retain as much detail as possible.

Codebook

A **codebook** contains coding instructions and the necessary information about variables in the data set. A codebook generally contains the following information:

- column number

- record number

- variable number

- variable name

- question number

- instructions for coding

Coding Questionnaires

- The respondent code and the record number appear on each record in the data.

- The first record contains the additional codes: project code, interviewer code, date and time codes, and validation code.

- It is a good practice to insert blanks between parts

- Here are examples of coding

1a. How many years have you been playing tennis on a regular basis?
    Number of years: _____

b. What is your level of play?

Novice		-1	Advanced		-4
Lower Intermediate		-2	Expert		-5
Upper Intermediate		-3	Teaching Pro		-6

c. In the last 12 months, has your level of play improved, remained the same or decreased?

Improved		-1	Decreased		-3
Remained the same		-2			

2a. Do you belong to a club with tennis facilities? Yes . . . . [____] -1

No . . . . [____] -2

b. How many people in your household - including yourself - play tennis?
Number who play tennis _____

3a. Why do you play tennis? (Please "X" all that apply.)

To have fun . . . . [____] -1
To stay fit. . . . . [____] -2
To be with friends. . . . . . [____] -3
To improve my game . . . [____] -4
To compete. . . . . [____] -5
To win. . . . . [____] -6

b. In the past 12 months, have you purchased any tennis instructional
books or video tapes? Yes . . . . [____] -1

No . . . . [____] -2

4. Please rate each of the following with regard to <u>this</u> flight, if applicable.

	Excellent 4	Good 3	Fair 2	Poor 1
Courtesy and Treatment from the:				
Skycap at airport	☐	☐	☐	☐
Airport Ticket Counter Agent	☐	☐	☐	☐
Boarding Point (Gate) Agent	☐	☐	☐	☐
Flight Attendants	☐	☐	☐	☐
Your Meal or Snack.	☐	☐	☐	☐
Beverage Service	☐	☐	☐	☐
Seat Comfort.	☐	☐	☐	☐
Carry-On Stowage Space.	☐	☐	☐	☐
Cabin Cleanliness	☐	☐	☐	☐
Video/Stereo Entertainment	☐	☐	☐	☐
On-Time Departure	☐	☐	☐	☐

"I believe that people judge your success by the kind of car you drive."

Strongly agree	5	Strongly agree	+ 1
Mildly agree	4	Mildly agree	+2
Neither agree nor disagree	3	Neither agree nor disagree	0
Mildly agree	2	Mildly agree	- 1
Strongly disagree	1	Strongly disagree	- 2

Data Transcription

- Transcribe raw data into testable form

- Determine variables

- Convert raw data into meaningful for further processing and answering the research questions and testing hypothesis

- Assign values, weights, value labels

- Scanning, data entry

Data Entry

- The process of transforming data from the research project to computers

- Transferring data files from excel to SPSS

- Optical scanning systems
  - Marked-sensed questionnaires

    In SPSS open the data view

    And enter the data

    Practical session

Data Cleaning: Consistency Checks

Consistency checks identify data that are out of range, logically inconsistent, or have extreme values.

- Computer packages like SPSS, SAS, EXCEL and MINITAB can be programmed to identify out-of-range values for each variable and print out the respondent code, variable code, variable name, record number, column number, and out-of-range value.

- Extreme values should be closely examined.

Data Cleaning Through SPSS

Click analyze in main menu of SPSS data, then click on descriptive analysis, then frequencies

Select variable that you want to check

Click on statistics and tick minimum and maximum values

Click on continue

Summary of results will provide each of variable you selected and then breakdown of responses

Check if there are inconsistencies

Go to data file and remove if there is any

You can clean your data using SPSS descriptive analysis features

Data Cleaning: Treatment of Missing Responses

- **Substitute a Neutral Value** – A neutral value, typically the mean response to the variable, is substituted for the missing responses.

- **Substitute an Imputed Response** – The respondents' pattern of responses to other questions are used to impute or calculate a suitable response to the missing questions.

- In **casewise deletion**, cases, or respondents, with any missing responses are discarded from the analysis.

- In **pairwise deletion**, instead of discarding all cases with any missing values, the researcher uses only the cases or respondents with complete responses for each calculation.

Statistically Adjusting the Data: Weighting

- In **weighting**, each case or respondent in the database is assigned a weight to reflect its importance relative to other cases or respondents.

- Weighting is most widely used to make the sample data more representative of a target population on specific characteristics.

- Yet another use of weighting is to adjust the sample so that greater importance is attached to respondents with certain characteristics

- Example

Variable Re-specification

- **Variable respecification** involves the transformation of data to create new variables or modify existing variables.

- E.G., the researcher may create new variables that are composites of several other variables.

- Dummy variables are used for respecifying categorical variables. The general rule is that to respecify a categorical variable with K categories, K-1 dummy variables are needed.

Variable Re-specification

Product Usage Category	Original Variable Code	Dummy Variable Code		
		X_1	X_2	X_3
Nonusers	1	1	0	0
Light users	2	0	1	0
Medium users	3	0	0	1
Heavy users	4	0	0	0

Note that $X_1 = 1$ for nonusers and 0 for all others. Likewise, $X_2 = 1$ for light users and 0 for all others, and $X_3 = 1$ for medium users and 0 for all others. In analyzing the data, X_1, X_2, and X_3 are used to represent all user/nonuser groups.

Data Transformation

- Data conversion
- Changing the original form of the data to a new format
- More appropriate data analysis
- New variables

New Variables

- Collapsing 5-point scale into 3-point scale
- Collective, average data of respondents and variables
- Reversal of negative statements

- Example

Collapsing a Five-Point Scale

- Strongly Agree

- Agree

- Neither Agree nor Disagree

- Disagree

- Strongly Disagree

- Strongly Agree/Agree

- Neither Agree nor Disagree

- Disagree/Strongly Disagree

Descriptive Analysis

- The transformation of raw data into a form that will make them easy to understand and interpret; rearranging, ordering, and manipulating data to generate descriptive information

Tabulation

- Tabulation - Orderly arrangement of data in a table or other summary format

- Frequency table

- Percentages

Frequency Table

- The arrangement of statistical data in a row-and-column format that exhibits the count of responses or observations for each category assigned to a variable

Central Tendency

Type of Scale	Measure of Central Tendency	Measure of Dispersion
Nominal	Mode	None
Ordinal	Median	Percentile
Interval or ratio	Mean	Standard deviation

Cross-Tabulation

- A technique for organizing data by groups, categories, or classes, thus facilitating comparisons; a joint frequency distribution of observations on two or more sets of variables

- Contingency table- The results of a cross-tabulation of two variables, such as survey questions

Cross-Tabulation

- Analyze data by groups or categories
- Compare differences
- Contingency table
- Percentage cross-tabulations

Type of Measurement		Type of descriptive analysis
↓		↓
Ordinal	→	Rank order Median

You are good students-NOW PRACTICE

By Dr. Muhammad Ramzan
mramzaninfo@gmail.com,
03004487844

Edited by Ahsan Khan Eco
ahsankhaneco@yahoo.com
03008046243