Topic 12

Data Analysis and Interpretation
What is data analysis?

- A process where data collected are analyzed to answer research question about characteristics & relationship, patterns / influences in some social phenomena (Blaikie, 2003).

- For quantitative data – analyzes data using appropriate statistical techniques.
  - Data analysis packages
    - SPSS for Windows, Excel

- The result of statistical analysis are descriptions, relationships, comparison and predictions.

- Objectives:
  1. getting a feel for the data
  2. testing the goodness of data
  3. testing the hypotheses
Getting a Feel for the Data

• Get mean, variance and standard deviation for each variable.
• See if all items, responses range over the scale, and not restricted to one end of the scale alone.
• Obtain Pearson Correlations for all variables.
• Tabulate your data.
• Descriptive statistics for your sample’s key characteristics (e.g. demographic details).
• See Histograms, Frequency Polygons, etc.
For each variable measured, obtain:

*Reliability
  • Split half
  • Internal consistency

*Validity
  • Convergent
  • Discriminant
  • Factorial
Testing Hypotheses

Using appropriate statistical analysis, test hypotheses, examples:

- **T-test** to test the significance of differences of the means of two groups.
- Analysis of variance (ANOVA) to test significance of differences among the means of more than two different groups, using the F-test.
- Using **regression analysis** to establish the variance explained in the DV through independent variables.
What is the choice of method of data analysis?

- Depends on the objective of study.
- Depending on whether you are studying the whole population / sample.
  - Depending on type of sampling techniques used
    - probability or non-probability sampling.
- Data analysis can be divided into 4 types:
  - Univariate descriptive
  - Bivariate descriptive
  - Explanatory
  - Inferential
Univariate Descriptive

Purpose:

- To describe / summarize individual variables or a set of variables in a given data set.
- It provides important information about variable.
- 3 major characteristics of a single variable that we tend to look at are:
  - The frequency distribution & percentage
  - The central tendency
  - The dispersion
Bivariate Descriptive

Purpose:
- To establish similarities or differences between the variables of study.
- Describe patterns or connections or strength of relationships between the variables of study.

Statistical procedure commonly used:
- Chi-square Test
- Measures of association – Pearson’s Correlation ($r$)
Explanatory

Purpose:
- To establish variables/factors that are responsible to produce the state of the phenomenon/trends.
  - Establish influence. Association alone does not allow one to infer that one variable has an influence on another.

Statistical procedure commonly used:
- Simple linear/multiple regression
- Partial Correlation
- Mediation/moderation analysis
- Mean analysis
• Used to generalize the result obtained from a random (probability) sample back to the population from which the sample was drawn.

• Use the same type of analysis as in univariate, descriptive & exploratory analyses.

• Conditions to use inferential analysis:
  – Sampling technique – probability sampling
  – Response rate – not less than 85%
• Key concept in inferential analysis:
  – Bivariate/Explanatory Analysis – significance p-value [or t-value (t-test) or F-value for ANOVA, multiple regression].
  – Univariate Descriptive Analysis – confidence level and confidence interval.

  ❖ **Confidence level** – level/degree of confidence (level of probability) a researcher wish to set – 90%, 95%, or 99%. If 95% = 0.05 significance level. Indicate how confident we are about our population estimate. Known as the p-value or significant value.

  ❖ **Confidence interval** – range of value around the sample value within which we expect the population value to lie (upper & lower limits of confidence interval).

  \[ CI \ (interval/ratio) = M \pm (z \times se_M) \]
  \[ se_M = \frac{SD}{\sqrt{n}} \]

  \[ CI \ (Categorical \ variables) = p \pm z \times \sqrt{\frac{p*(1-p)}{n}} \]

  Refer to this website for CI calculator:
  http://dimensionresearch.com/resources/calculators/conf_prop.html

  \[ z \text{ score of the p-value} = t-value \]
  - 95% {0.05} = 1.960
  - 99% {0.01} = 2.576

  (See: Kumar 2005)
SPSS

Windows
**SPSS Data Editor:**
Define Variables and enter data

**Variable view:** to define variables

**Data view:** to enter data
2 SPSS Viewer:
Display results of data analyses

Navigator to help easy browsing
Results of data analyses
SPSS Syntax Editor:
Write, display, retrieve, run and save syntax/commands

Syntax or commands
Steps in Data Analysis:

1. Define Variables
2. Enter Data
3. Run Frequency
4. Data Editing
5. Reliability Test
6. Transformations
7. EDA

Data Analyses:
- Frequency
- t-test
- ANOVA
- Chi-square
- Correlation
- Regression
- Factor analysis
- Repeated measures ANOVA
Define and Enter Data
In the Data Editor, you can:

1. **Define variables** in **Variable View** window
2. **Enter data** in **Data View** window
Define:

- Name
- Label
- Values

Required
Optional; recommended
2 Enter Data

In Data View window

One column refers to one variable

One row refers to one case or observation
Data Editing and Transformations
Data Editing:

- Change data value
- Cut, copy and paste data value
- Add or delete case
- Add or delete variable
- Change ordering of variables
Data Transformations:

Two commonly used data transformations:

1. **Compute**
   - Create new variable from existing variables.
   - Example: Calculate mean cumulative attitude scores.
     - Attitude comprises 7 items (A1 to A7).
     - Procedure: 
       - Click menu Transform.
       - Compute Attitude = Mean (A1 to A7).

![Compute Variable window](image)
2. Recode

- Create categories from continuous variable
- Example:
  Categorize mean cumulative Attitude scores into the following three categories:
  - Low (1) 1 – 2.33
  - Moderate (2) 2.34 – 3.66
  - High (3) 3.67 – 5

  RECODE Attitude (1 thru 2.339=1) (2.34 thru 3.669=2) (3.67 thru 5=3) INTO Attitude_CAT

- Procedure:
  - Click menu Transform | Recode into Different Variable
Statistical Procedures:

- Reliability test
- Exploratory data analysis
- Descriptive Statistics
  - Frequencies, crosstabs
- Compare group means
  - t-test and ANOVA
- Relationship between variables
  - Chi-square, Spearman rho, Pearson PM correlation, and regression
Reliability Test:
Exploratory Data Analysis:
Frequencies:
ANOVA:
Correlation Analyses:
SPSS Coaches
1 **Statistic Coach**

Guide in selecting the most appropriate statistical analysis
2 Result Coach
Help to interpret statistical results

Job satisfaction

Job satisfaction Stem-and-Leaf

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Stem &amp; Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>15</td>
<td>7 8 9 0 1</td>
</tr>
</tbody>
</table>

Tests of Normality

<table>
<thead>
<tr>
<th>Division</th>
<th>Kolmogorov-Smirnov Statistic</th>
<th>df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer Products</td>
<td>.156</td>
<td>173</td>
<td>.000</td>
</tr>
<tr>
<td>Business Products</td>
<td>.192</td>
<td>211</td>
<td>.000</td>
</tr>
</tbody>
</table>

The Kolmogorov-Smirnov statistic tests the hypothesis that the data are normally distributed.
First step in data analysis

Information collected – raw data

1st step – ensure data are clean – free from inconsistencies and incompleteness

Prepare for data entry.
- Coding data – prepare a code book.
  * How you code depends on the type of scale of measurement of the variable
- Create SPSS file
  * Enter coding in SPSS file
- Perform data entry
  * Cleaning of data
    * Look for missing values
- Perform data transformation
- Perform data computation

Conduct a reliability analysis pilot test & final test
### Sample Code Book

#### DEMOGRAPHIC PROFILE

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Type</th>
<th>Label</th>
<th>Values</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Numeric</td>
<td>Respondent ID</td>
<td>None</td>
<td>Nominal</td>
</tr>
<tr>
<td>group</td>
<td>String</td>
<td>Group name</td>
<td>None</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
| gender        | Numeric | Gender              | 1. Male  
2. Female         | Nominal |
| age           | Numeric | Student Age         | None                    | Scale   |
| marital       | Numeric | Marital Status      | 1. Single  
2. Married  
3. Divorced  
4. Widowed         | Nominal |
| race          | Numeric | Race                | 1. Malay  
2. Chinese  
3. Indian  
4. Others          | Nominal |
| orace         | String  | Other races         | None                    | Nominal |
| relig         | Numeric | Religion            | 1. Islam               | Nominal |