WORKSHOP 1: INTRODUCTION TO SPSS 1

SPSS Basics:

MINJEONG PARK, GRADUATE ACADEMIC ASSISTANT
MOHAMMAD ZUBAIR, GRADUATE ACADEMIC ASSISTANT

UBC Library Research Commons
RESEARCH COMMONS: AN INTERDISCIPLINARY RESEARCH-DRIVEN LEARNING ENVIRONMENT

- Literature review
- Systematic Review Search Methodology
- Citation Management
- Thesis Formatting
- Nvivo Software Support
- SPSS Software Support
- R Group
- Multi-Disciplinary Graduate Student Writing Group
SPSS SERVICES BY RESEARCH COMMONS

• Workshops

• One-on-one Consultation

Request form to book SPSS consultation:

Background

• Have you used SPSS before?

• Have you worked with other data analysis software?

• Have you conducted any quantitative data analysis before?
Outline

Part 1: Introduction to Statistical Analysis
• Statistical analysis and statistical models
• Various statistical models
• Common criteria for selecting statistical models
• Various software for statistical analysis
• Commonly used software in social sciences

Part 2: Data Preparation in SPSS
• SPSS Interface
• Data preparation with SPSS using example data set
  – Data import/entry
  – Descriptive statistics (Frequencies)
  – Descriptive statistics (Graphs)
Part 1
Introduction to statistical analysis

• Statistical analysis and statistical models
• Various statistical models
• Common criteria for selecting statistical models
• Various software for statistical analysis
• Commonly used software in social sciences
Statistical analysis and statistical model

Broadly speaking,

- **Analysis?**
  - Detailed examination of phenomenon

- **Statistical analysis?**
  - Examination of phenomenon *based on “statistical theories”*

- **Statistical model?**
  - A mathematical structure (simply a tool) that helps you to analyze your data
Various statistical models

Different statistical models are available:

- Correlation
- Independent sample T-test
- Independent one-way ANOVA
- Multiple regression
- Logistic regression
- Paired-sample T-test
- Repeated measures ANOVA
- ...
- Multilevel model (a.k.a., mixed model or hierarchical linear model)
Various statistical models

Different statistical models are designed to handle different

- *Research design*
- *Research question*
- *Type of data*
- *Inference*
- ...

Select statistical model selection can be based on…
Common criteria for selecting statistical models

• **Research design:**
  
  *Cross-sectional* vs. *Longitudinal (repeated measures)*
  
  (collect data in “one time point”) (collect data in “multiple time point”)

• **Inference:**
  
  *Descriptive* vs. *Inferential*
  
  (Inference about sample) (Inference about population)

• **Research question:**
  
  *Association* vs. *Group difference* vs. *Prediction* vs. *Change*

• **Type of data:**
  
  *Continuous data* vs. *Categorical data* vs. *Ordinal data*
  
  (numbers-“amount”) (numbers/strings-”type”) (numbers-”order”)
  
  e.g., Total scale scores e.g., Gender, Country ID e.g., Rank, Size (S, M, L)
Selecting statistical models based on the criteria

Among **inferential statistics**…
(make an inference about population)

<table>
<thead>
<tr>
<th>Cross sectional</th>
<th>Association</th>
<th>Group difference</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous</td>
<td>Continuous</td>
<td>Continuous</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>Categorical</td>
<td>Categorical</td>
</tr>
</tbody>
</table>

- **Inference**
- **Research design**
- **Research question**
- **Data type**

→ *We will discuss more in SPSS 2 workshop!*
Software for statistical analysis

• Once you have…
  - “Data” collected
  - “Research questions” to be answered
  - “Statistical models” selected

• **Statistical analysis software** can help you to conduct statistical analysis
Various software for statistical analysis

Different software programs are available:

- SPSS
- R
- Stata
- SAS
- Matlab
- Minitab
- Mplus
- HLM7
- ...

Software selection can be depending on...

- Convention in disciplines
- Availability
- Characteristics of software
## Software commonly used in social sciences

<table>
<thead>
<tr>
<th></th>
<th>SPSS</th>
<th>Stata</th>
<th>R &amp; R studio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price</strong></td>
<td>Gradpack (Base version) 34.49 USD* (6 month) Free trial available</td>
<td>Stata IC 125 USD* (annual) 30 days money back option</td>
<td>Non-commercial purpose (Free)</td>
</tr>
<tr>
<td><strong>Availability in UBC</strong></td>
<td>Koerner Rm 217A, 218A Woodward Rm B25</td>
<td>Koerner Rm 217A, 218A Woodward Rm B25</td>
<td>Koerner Rm 217A, 218A Woodward Rm B25</td>
</tr>
<tr>
<td><strong>Interface</strong></td>
<td>clicking button + syntax</td>
<td>clicking button + syntax</td>
<td>syntax/command</td>
</tr>
<tr>
<td><strong>Data management</strong></td>
<td>Basic data management options</td>
<td>Basic data management options</td>
<td>Basic data management options + more data management options</td>
</tr>
<tr>
<td><strong>Statistical models</strong></td>
<td>Well-established statistical models</td>
<td>Well-established statistical models</td>
<td>Well-established statistical models + recent statistical models are updated</td>
</tr>
<tr>
<td><strong>Data visualization</strong></td>
<td>Basic plots, graphs, charts</td>
<td>Basic plots, graphs, charts</td>
<td>Basic plots, graphs, charts + more visualization options</td>
</tr>
<tr>
<td><strong>Discussion forum</strong></td>
<td>Stack Overflow Cross Validated Nabble</td>
<td>Statalist</td>
<td>Stack Overflow Cross Validated Nabble</td>
</tr>
</tbody>
</table>
Part 2
Data Preparation in SPSS

• SPSS Interface

• Data preparation with SPSS using example data set
  – Data import/entry
  – Descriptive statistics (Frequencies)
  – Descriptive statistics (Graphs)
SPSS interface

SPSS interface is composed of 3 main windows:

- **Data Editor window**  
  (Data View + Variable View)

- **Output window**

- **Syntax window**
## SPSS interface

- **Data Editor** - Data View: present whole “data”

![SPSS Data View](image)
SPSS interface

- **Data Editor - Variable View:** present information of all “variables”
SPSS interface

- **Data Editor - Variable View**: “Key” information of variables

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Decimal points of data</th>
<th>Description of variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>Gender [0, male]</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>Age</td>
</tr>
<tr>
<td>Marital</td>
<td>0</td>
<td>Marital status [1, married, cohabiting]</td>
</tr>
<tr>
<td>Employment</td>
<td>0</td>
<td>Employment [1, no job]</td>
</tr>
<tr>
<td>QOL</td>
<td>0</td>
<td>Quality of life None</td>
</tr>
<tr>
<td>Distress</td>
<td>0</td>
<td>Distress None</td>
</tr>
</tbody>
</table>

E.g. 0 decimal point = 1
1 decimal point = 1.0
SPSS interface

- **Data Editor - Variable View:** “Key” information of variables

Meaning of values in variable
  E.g. 0 = male, 1 = female

Numbers indicating missing data
**SPSS interface**

- **Data Editor - Variable View**: “Key” information of variables

![Data Editor - Variable View](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Decimals</th>
<th>Label</th>
<th>Values</th>
<th>Missing</th>
<th>Measure</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>999</td>
<td>999</td>
<td>Scale</td>
<td>Input</td>
</tr>
<tr>
<td>Gender</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>[0, male]...</td>
<td>999</td>
<td>Nominal</td>
<td>Input</td>
</tr>
<tr>
<td>Age</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>999</td>
<td>999</td>
<td>Ordinal</td>
<td>Input</td>
</tr>
<tr>
<td>Marital</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>[1, married, co...</td>
<td>999</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>[1, no job]...</td>
<td>999</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>QOL</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>999</td>
<td>999</td>
<td>Scale</td>
<td>Input</td>
</tr>
<tr>
<td>Distress_total</td>
<td>Numeric</td>
<td>8</td>
<td>0</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Input</td>
</tr>
</tbody>
</table>

**Type of data in variable**
- **Scale** = continuous data
- **Nominal** = categorical data
- **Ordinal** = ordinal data

For more information about scale, nominal, and ordinal options,
https://stats.idre.ucla.edu/other/mult-pkg/whatstat/what-is-the-difference-between-categorical-ordinal-and-interval-variables/
SPSS interface

- *Drop-down menu in Data Editor:*

![Image of SPSS Data Editor](image-url)
SPSS interface

- **Output**: present “history of your analysis” and all “outputs”
SPSS interface

- **Syntax**: write “syntax”
Break – 10 min.
Data preparation with SPSS using example data

• Cross-sectional design:
  A set of variables measured from each person in one time point

• A set of variables:
  - Gender (Male = 0, Female = 1)
  - Age (range 10 – 80)
  - Marital status (Married, common law = 1, Widow, divorce, separate = 2, Single, never married = 3)
  - Employment (no job = 1, part time = 2, full time = 3)
  - Quality of life_total (range 0 – 20)
  - Distress_total (range 0 – 20)
  - Self-esteem items (range 0 – 3)

• Missing data are coded as 999

Data preparation in SPSS:
Importing the data
Looking at descriptive statistics
Data import/entry

• Open data in “SPSS format” (.sav):
  File > Open > Data > Select SPSS data file

• Import data in “different formats” (spreadsheet, text etc.):
  File > Open > Data…
Data import/entry

- Enter your data in SPSS *Data Editor – Data View*:

![Data Editor - Data View]

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Gender</td>
<td>Age</td>
<td>Marital</td>
<td>Employment</td>
<td>QOL</td>
<td>Distress_to</td>
<td>Esteem_Q1</td>
<td>Esteem_Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>0</td>
<td>74</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>1</td>
<td>75</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>193</td>
<td>1</td>
<td>79</td>
<td>3</td>
<td>2</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>194</td>
<td>0</td>
<td>43</td>
<td>2</td>
<td>2</td>
<td>17</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>1</td>
<td>79</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>196</td>
<td>0</td>
<td>45</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>1</td>
<td>79</td>
<td>3</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>1</td>
<td>43</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>1</td>
<td>32</td>
<td>2</td>
<td>2</td>
<td>19</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>1</td>
<td>36</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>201</td>
<td>0</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>206</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>207</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Checking information of variables in Variable View

• A set of variables:
  - Gender (Male = 0, Female = 1)
  - Age (range 10 - 80)
  - Marital status (Married, common law = 1, Widow, divorce, separate = 2, Single, never married = 3)
  - Employment (No job = 1, Part time = 2, Full time = 3)
  - Quality of life_total (range 0 - 20)
  - Distress_total (range 0 – 20)
  - Self-esteem items (range 0 – 3)

• Missing data is coded as 999
A set of variables:

- **Gender** (Male = 0, Female = 1)
- **Age** (range 10 - 80)
- **Marital status** (Married, common law = 1, Widow, divorce, separate = 2, Single, never married = 3)
- **Employment** (No job = 1, Part time = 2, Full time = 3)
- **Quality of life_total** (range 0 - 20)
- **Distress_total** (range 0 – 20)
- **Self-esteem items** (range 0 – 3)

**Missing data** is coded as **999**

**E.g., Numbers indicating missing data**

Click it
Descriptive statistics

• Descriptive statistics provide a summary of your data

• Purpose of looking at descriptive statistics:

  (1) Check whether valid data are loaded properly
      E.g., unexpected values (e.g., 999, -2) in “Age” variable (range 10-80)

  (2) Explore data
      E.g., potential group differences, associations between variables

  (3) Sample description
      E.g., % of gender, mean and standard deviation of quality of life score
Descriptive statistics

Descriptive statistics in SPSS:
Descriptive statistics - Frequencies

• Frequencies for “categorical data”
Descriptive statistics - Frequencies

- Frequencies for “categorical data” – Descriptive statistics

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>91</td>
<td>45.5</td>
<td>45.5</td>
<td>45.5</td>
</tr>
<tr>
<td>female</td>
<td>109</td>
<td>54.5</td>
<td>54.5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marital status</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>married, common law</td>
<td>63</td>
<td>31.5</td>
<td>31.5</td>
<td>31.5</td>
</tr>
<tr>
<td>widow, divorce, separate</td>
<td>65</td>
<td>32.5</td>
<td>32.5</td>
<td>64.0</td>
</tr>
<tr>
<td>single, never married</td>
<td>72</td>
<td>36.0</td>
<td>36.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employment</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>no job</td>
<td>58</td>
<td>29.0</td>
<td>29.0</td>
<td>29.0</td>
</tr>
<tr>
<td>part time</td>
<td>75</td>
<td>37.5</td>
<td>37.5</td>
<td>66.5</td>
</tr>
<tr>
<td>full time</td>
<td>67</td>
<td>33.5</td>
<td>33.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
Descriptive statistics - Frequencies

- Frequencies for “categorical data” - Bar plots

**Gender**

- Male
- Female

**Marital status**

- Married
- Widow
- Single

**Employment**

- No job
- Part-time
- Full-time
Descriptive statistics - Frequencies

- Frequencies for “continuous data”
Descriptive statistics - Frequencies

• Frequencies for “continuous data” – Descriptive statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Age</th>
<th>Quality of life</th>
<th>Distress</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>196</td>
<td>200</td>
<td>198</td>
</tr>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Mean</td>
<td>44.20</td>
<td>11.45</td>
<td>10.68</td>
</tr>
<tr>
<td>Minimum</td>
<td>13</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>80</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
Descriptive statistics - Frequencies

- Frequencies for “continuous data” – Histograms

- Age
- Quality of life
- Distress
Descriptive statistics - Graphs
Descriptive statistics - Graphs

- **Pie charts**: Graphs > Legacy Dialogs > Pie…

   → *Useful to see the proportion of “categorical variable”*
Descriptive statistics - Graphs

- **Pie charts**: output
Descriptive statistics - Graphs

- **Scatter/Dot plots**: Graphs > Legacy Dialogs > Scatter/Dot…

→ *Useful to explore associations between variables*
Descriptive statistics - Graphs

• Scatter plots: output
**Summary of SPSS 1 workshop**

In Part 1:
- Statistical analysis as an examination based on “statistical theories”
- Common criteria for selecting statistical models
- Various statistical analysis software (SPSS, Stata, R)

In Part 2:
- SPSS Interface
- Data preparation in SPSS (Data import/entry, Descriptive statistics)
SPSS 2 workshop

SPSS 2 workshop will introduce applying statistical models in SPSS

How to select statistical models?
• Common criteria to select statistical models

How to run statistical models in SPSS?
• Independent sample T-test
• Independent one-way ANOVA
• Correlation
• Simple & Multiple regression
THANK YOU!
QUESTION, COMMENT, IDEAS