Introduction to IBM SPSS Statistics 20

by
Stian Lydersen

NTNU

Revised 13 aug 2012

http://folk.ntnu.no/slyderse/medstat/SPSS/Introduksjon_SPSS.pdf

http://folk.ntnu.no/slyderse/medstat/SPSS/Introduction_SPSS.pdf

Purpose:

Introduction to:

- Establish a data file. Enter data. Edit the file.
- Presentation of data – descriptive statistics. Tables and graphs.
- Basic analyses.

Some statistics programs:

- General:
  - R[0)], SAS, Stata[2)],
  - SPSS[1)] , MINITAB[1)], Statistica
- Special:
  - SamplePower[1)], Amos[1)] , LISREL, M-plus, StatXact, LogXact
- Graphics:
  - SigmaPlot[1)]
- Spreadsheet:
  - Excel[1)]

[0)] Free, [1)] NTNU-licence, [2)] licence at DMF

SPSS

- General statistics package
- Easy to use
- Easy data entry, file structure, and editing of file
- Much output
- Limited documentation of methods
- Limited on advanced or special methods
- Difficult syntax structur
- “dominating” in medical research in Norway.

Litterature

- Kinnear & Gray: ” IBM SPSS Statistics 19 Made Simple”, 2011
- Bowers, David: ”Medical Statistics from Scratch. An Introduction for Health Professionals” 2nd ed, Wiley, 2008. ISBN 978-0-470-51301-9.

Documentation fromSPSS Inc.

- Statistics Base / Regression / Advanced Models / Categories / Conjoint / Exact Tests / Missing Values / Bootstrapping / and several others
  - http://download.spss.no/SPSS_Statistics_19_Doc_en.zip
  - Some are also availble in books
- Help -> Topics
- Help -> Algorithms: Technical description / definiition of methods

File types

- Data file (*.sav)
- Viewer file (*.spv)
- Syntax file (*.sps)
- … and some other less used file types

7

Moving files between SPSS versions

- Data file:
  - Only minor format changes. Usually no problems
- Viewer file:
  - Can be difficult or impossible
  - Save important results as syntax file and/or pdf file
- Syntax file:
  - Usually no problems

8

Structure of data file: Cases, variables and values

- Case:
  - Example: Person
- Variable:
  - Example: id number, age, height, sex(gender)
- Values:
  - Example: 205, 45, 178.2, “female”

9

Repeated measurements - 2 alternative formats:

- 1. Each patient as case (“wide format”)
- 2. Each time point for each patient as case (“long format”)
- Svitsje mellom format 1 og 2 vha Data editor -> data -> Restructure

10

Windows

- Data Editor:
  - Data View
  - Variable View
- Viewer
- Syntax Editor
- … and some other less used windows

11

Variable view - 1

- Name:
  - Up to 64 characters (letters, numbers, @, #, _,$, …)
  - Start with letter
  - No space, no *, ?, !, …
  - The letters æ, ø, å are not recommended
  - Not ”and”, ”or”, ”not”, ...
- Type: Numeric, date, string, etc
- Width (on data file), especially relevant for ”string”

12

Variable view - 2

- Decimals (on screen / output)
- Label: Up to 120 characters, all characters allowed
- Values: F.ex 1 = male, 2 = female
- Missing: F.ex 98 = was not asked, 99=not answered
- Column (on screen / output)
- Align
- Measure: Scale, Ordinal, Nominal
- Role: Input, Target, Both, None, Partition, Split

13

Missing:

- "System missing" – no value is entered. Easiest and almost always OK! (Not for variable type "string")
- User definet missing:
  - Can distinguish between causes of missing
  - Can distinguish between missing and forgotten to enter value
  - Can give problems when saving in other formtats than SPSS

14

Measure, examples:

- Scale:
  - Height in cm
- Ordinal:
  - How is your health? 1 = "poor" , 2 = "not very good", 3 = "good", 4 = "very good"
- Nominal:
  - Marital status: 1 = "not married", 2 = "married", 3 = "cohabiting" , 4 = "divorced", 5 = "widow(er)"

15

Time between dates

- Transform -> Date and Time Wizard -> Calculate with dates and times
  - Default: Gives time truncated (down) to nearest integer (f.ex number of whole years)
  - Recommended: Keep as decimal number

16

Computation of new variables

- Example: Body mass index
- BMI=(veight in kg)/(height i meters) ^{2}
- Transform -> Compute variable

17

Entering data

- Keyboard entry in data editor
- Opening an SPSS file
- From another file type, f.ex EXCEL
  - File -> Open -> Data -> Files of Type (Usually recommended)
  - vha Copy - Paste i Windows (Useful with small files or small data areas. Can loose some information)

18

Descriptive statistics - tables

- One scale variable (or categorical variable):
  - Descriptive statistics -> Descriptives
- One categorical variable:
  - Descriptive statistics -> Frequencies
- Two categorical variables:
  - Descriptive statistics -> Crosstabs

19

Descriptive statistics – tables (cont.)

- One scale variable and one categorical variable:
  - Compare means -> Means
- Two scale variables:
  - Categorize one of the variables
  - Alternatively: Simple linear regression

20

Finding and correcting errors

- Descriptive statistics
  - Frequencies, Descriptives, Crosstabs
- Correcting errors:
  - Find the error(s):
    - Edit - Find (In Data Editor - Data View), or
    - Data -> Sort cases
  - Correct or delete erroneous values

21

Copying tables / graphs to Word, Excel or Power Point:

- In SPSS:
  - Finish editing the object (table 7 graph)
  - Edit -> Copy
- In Word:
  - Place the pointer at the right place
  - Edit -> Paste special -> (choose a suitable format)
- In Excel (table)
  - Place the pointer at the right place
  - Edit -> Paste

22

Hypotesis testing

- Establish nullhypothesis and alternative hypotesis. Example:
  - H_0: Expected response is equal in the goups
  - H_1: Expected response differs
- The p-value (”signifikanssannsynlighet”, sig.) isthe probability of observing the observed values or more extreme, given H_0 is true.
- Reject H_0 if the p-verdi is less than the signifikance level (usually 0.05 or 0.01)

23

Choice of statistical test method.

- Some tests assume a certain model.
  - Example: Student’s t-test assumes data to be (approximately) normally distributed
- Non-parametric tests are more flexible
  - Example: Compare two medians: Nonparamteric tests > 2 indenpendent groups (Mann-Whitney)

24

Exact versus asymptotic

- Exact:
  - 0.215 * 0.529 = 0.113735
- Approximately:
  - 0.215* 0.529 ≈ 0.2 * 0.5 = 0.1
- Asymptotic means approximately, with better approximation with larger n

25

Calculation of p-values in SPSS:

- Asymptotic and exact methods are available in
  - Crosstabs
  - Nonparametrics (Wilcoxon-Mann-Whitney, Kruskall-Wallis and others)
- Asymptotic is default
- Exact is recommended for small data sets
- Exact is too time-comsuming or impossible for large data sets.

26

Some useful graphs

- One categorical variable:
  - Bar chart ("stolpediagram")
  - Pie chart ("kakediagram"?)
- Two categorical variables:
  - Clustered bar chart ("klynget stolpediagram")

27

Some useful graphs (cont.)

- One scale variable:
  - Histogram
  - Compare data with the normal distribution: Q-Q plot is easier to read and interpret than "normal curve overlay" in histogram
- Two scale variables:
  - Scatterplot

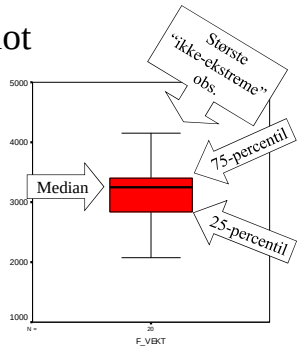
28

Some useful graphs (cont.)

- One scale variable and one categorical variable (compare the scale variable between two or more groups):
  - Dot plot or scatter plot (when "few" observations)
  - Box plot (with "many" observations)

29

Box-plot



30