

Introduction to SPSS Handout

Sunthud Pornprasertmanit
Faculty of Psychology, Chulalongkorn University

Basic Command

Data viewer, Output, Syntax File, Table Editor, Chart Editor

.sav, .sps, .spo

Data view, variable view

Paste Command

Using asterisk in syntax file

End file with period

Paste Variables

Edit → Options

General, variable lists

Viewer, display command in log, font

Output Label

Interactive

Pivot Tables

Data

Fonts, Value Labels

Sort Cases

Merge Files

Aggregate

Split Files

Select Cases

Weight Cases

Compute

Random Number Seed

Count

Recode

Categorized Variables

Rank cases

Automatic Recodes = Condensed rank data

Replace Missing Values

Descriptive Statistics

Frequencies

Descriptive

How to clean data

Data entry error

Missing data

Outlier

Author Note

This article was written in December 2006 for teaching in Computer Application in Psychology Class, Faculty of Psychology, Chulalongkorn University
Correspondence to Sunthud Pornprasertmanit. Email: psunthud@gmail.com

Missing data

How missing data come from?

Impact of missing data

Reduction of sample size

Statistical bias

Understanding the process of missing data (e.g. are missing data random?) and cure their impact (delete case, imputation)

Four-step process in identify missing data

1) Is it ignorable missing data?

a. Ignorable missing data

i. Parameter

ii. Missing data from design that known

iii. Censored data

b. Not ignorable

i. Known and can control

ii. Known but cannot control

iii. Unknown

2) How much missing data?

a. If not much, skip to step 3

b. If much (more than 10%), may delete cases or variables

3) Diagnose the randomness of data

Missing at random (MAR) or Missing at completely random (MCAR);

Test by Little's MCAR test or analyse all variables to explore missing data process

4) Imputation method

a. MAR – EM approach

b. MCAR

i. Only valid data; Listwise, pairwise

ii. Replacement value; Hot-cold deck imputation, case substitution, Mean substitution, Regression imputation, EM approach

SPSS – Missing value analysis

Outlier

Influence of outlier

Effect on result of analysis (Both improve or prevent relationship)

Don't present population

Question that frequently asked: Do the outliers represent population?

Four classes of outliers

1) procedural error

2) extraordinary event: has explanation

3) extraordinary observation: has no explanation

4) unique in their combination (bivariate outlier or multivariate outlier)

Detecting outlier

Univariate: z-score, boxplot

Bivariate: Scatterplot (Confident eclipse), boxplot

Multivariate: Mahalanobis D^2 (Euclidian distance/df) dispersed by t-distribution

If found outlier, should find the process of outlier and decide whether retention or deletion

If deleted, less generalizability.

Maybe analysis both with outlier and without outlier

Testing assumption of statistical analysis

Why are there assumptions of statistical analysis?

For accuracy of statistical analysis

Explore more information from data

Concept of robustness

Assessing individual variable and variate

Four statistical assumptions

1) Normality: univariate normality, multivariate normality

How they affect if violated: made t or F test inaccuracy, make unequal error in prediction

How to detect

(1) graphical display: histogram, normal p-p plot

(2) statistical detection: skewness, kurtosis, Komogorov-

Smirnov test

Large sample size reduce detrimental effects of nonnormality

Remedy: choose another statistical technique, data transformation

2) Homoscedasticity; Homogeneity of variance, Homogeneity of variance/covariance matrices

How they are occur: variable type, skewed distribution

How they affect if violated: unequal accuracy of prediction, made statistical testing too liberal or too conservative

How to test

(1) graphical display: scatterdiagram, boxplot

(2) statistical detection: Levene test, Hartley F_{\max} test, Box's F

test, Box's M test

Remedy: Data transformation, Weighted least square approximation

3) Linearity

How they affect if violated: reduce true correlation b/w variables

How to detect: scatterdiagram, examine residual, nonlinear

relationship

Remedy: Data transformation

4) Absence of correlated errors: No patten of error or unexplained systematic relationship exists in the dependent variable

How they affect if violated: specification error, inaccuracy in prediction, no maximum variance extracted

Example: Time-series data, group that not included in model

How to detect: examine residual

Remedy: correcting the specification error

Data transformation

From theory or data derived

Change in interpretation of results

Maybe analysis both transformed and not transformed

Should not transform dependent variable b/c difficult in interpretation

Dummy variable

Changing from nonmetric variable to metric variable

How to build dummy variable

- 1) indicator coding
- 2) effect coding
- 3) orthogonal coding