Basic Statistical Analysis and Interpretation

http://orcid.org/0000-0003-0006-8841
Dave E. Marcial
www.davemarcial.net
demarcial@su.edu.ph

Concepts and Variables
Structure of Statistical Analysis
Statistical Packages/Software

C. Statistical Packages

e-Statistical Tool (open source)

- ADaMSoft – a generalized statistical software with Data mining algorithms and methods for data management.
- ADMB – a software suite for non-linear statistical modeling based on C++ which uses automatic differentiation.
- Bayesian Filtering Library
- Chronux – for neurobiological time series data
- DAP – A free replacement for SAS
- ELKI a software framework for development of data mining algorithms in Java.
- Fityk – nonlinear regression software (GUI and command line)
- gretl – gnu regression, econometrics and time-series Library
- JAGS – Just another Gibbs sampler (JAGS) is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) developed by Martyn Plummer. It is similar to WinBUGS.
- JHepWork – Java-based statistical analysis framework for scientists and engineers. It includes an advanced IDE and Jython shell.
- JMulTi
- Octave – programming language (very similar to Matlab) with statistical features
- Mondrian (software) - data analysis tool using interactive statistical graphics with a link to R.
- OpenBUGS

e-Statistical Tool (open source)

- Orange, a machine learning and bioinformatics software
- Ploticus – software for generating a variety of graphs from raw data
- PSPP – A free software alternative to IBM SPSS Statistics
- R – A free implementation of the S language.
  - R Commander – GUI interface for R
  - Rattle GUI – GUI interface for R
- RapidMiner, a machine learning toolbox
- Revolution Analytics – Production-grade software for the enterprise big data analytics
- Salstat - Menu driven statistics software
- Scilab – uses GPL compatible CeCILL license
- SciPy (a Python library for scientific computing) contains the *stats* sub-package which is partly based on the venerable |*STAT* (a.k.a. *PipeStat*, formerly *UNIX*|*STAT*) software
  - scikit-learn extends SciPy with a host of machine learning models (classification, clustering, regression, etc.)
- Shogun, an open source Large Scale Machine Learning toolbox that provides several SVM (Support Vector Machine) implementations (like libSVM, SVMlight) under a common framework and interfaces to Octave, Matlab, Python, R
- Simfit – Simulation, curve fitting, statistics, and plotting
- SOCR
- SOFA Statistics – a desktop GUI program focused on ease of use, learn as you go, and beautiful output.
- Statistical Lab – R-based and focusing on educational purposes
- Weka is also a suite of machine learning software written at the University of Waikato.
- Xlisp-stat

e-Statistical Tool (free)

BV4.1

Screenshots

Developer(s)	Federal Statistical Office of Germany
Stable release	1.1 / February 2006
Operating system	Windows NT 4.0 / Windows 98 and higher
Available in	English, German
Type	Statistics / Statistical software / Time series analysis / Seasonal adjustment / Decomposing of time series
License	Freeware
Website	BV4.1

e-Statistical Tool (proprietary)

- Analytica - visual analytics and statistics package
- Angoss - strong products KnowledgeSEEKER and KnowledgeSTUDIO are a cost effective alternative to SAS and SPSS
- ASReml – for restricted maximum likelihood analyses
- BMDP – general statistics package
- CalEst – general statistics and probability package with didactic tutorials
- Data Applied – for building statistical models
- EViews – for econometric analysis
- FAME – a system for managing time series statistics and time series databases
- GAUSS – programming language for statistics
- Genedata Analyst– software solution for the integration and interpretation of experimental data in the life science R&D
- GenStat – general statistics package
- GLIM – early package for fitting generalized linear models
- GraphPad InStat – Very simple with lots of guidance and explanations
- GraphPad Prism – Biostatistics and nonlinear regression with clear explanations
- IBM SPSS Statistics – comprehensive statistics package
- IBM SPSS Modeler – comprehensive data mining and text analytics workbench
- Izenda - Self-service data analysis over HTML5

e-Statistical Tool (proprietary)

- Q research software – quantitative data analysis software for market research
- Quantum – part of the SPSS MR product line, mostly for data validation and tabulation in Marketing and Opinion Research
- RATS – comprehensive econometric analysis package
- SAS – comprehensive statistical package
- SHAZAM – comprehensive econometrics and statistics package
- Simul - econometric tool for multidimensional (multi-sectoral, multi-regional) modelling
- SigmaStat – for group analysis
- SOCR – online tools for teaching statistics and probability theory
- Speakeasy – numerical computational environment and programming language with many statistical and econometric analysis features
- Stata – comprehensive statistics package
- Statgraphics – general statistics package
- STATISTICA – comprehensive statistics package
- StatXact – package for exact nonparametric and parametric statistics
- Systat – general statistics package
- S-PLUS – general statistics package
- Unistat – general statistics package that can also work as Excel add-in
- The Unscrambler (free-to-try commercial Multivariate analysis software for Windows)
- XploRe

SPSS Exercises

- Basic Descriptive Statistics
  - Compute Frequencies of Nominal Data
  - Cross Tabulation: Sex and Party
  - Compute Descriptive Statistics for Score Data
- Compute the Product-Moment Correlation between Age and Income
- Compute the Spearman Rank-Order Correlation between Age and Income
- Find the Linear Regression Equation for Predicting Income from Age
- Reliability
  - Compute the reliability of a test-Retest result
  - evaluate the internal consistency reliability

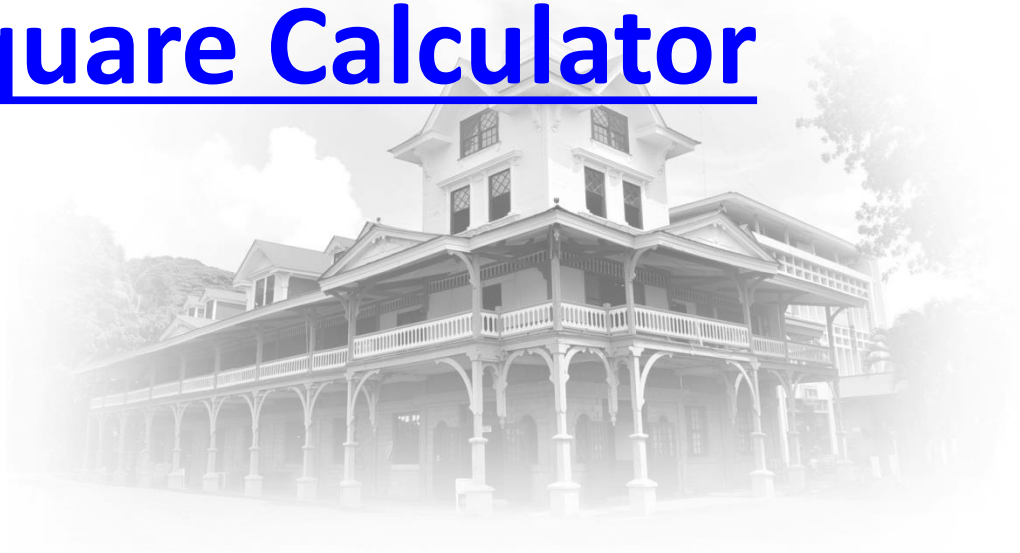
SPSS Demonstration

- Compute an independent samples *t*-test comparing men and women on their income level
- Computer the correlated Samples *t*-Test comparing two conditions (ten and Fifteen)
- Compute the One-Way ANOVA between typing speed and room temperature
- Test the hypothesis that the data from 50 flips of a coin indicates that the coin is honest
- Evaluate whether there is a relationship between sex and political affiliation

Demonstration

(**Measuring Dependency of Two Variables from Categorized Data)**

Online Chi-Square Calculator



Performance vs. Technology Efficacy

- Sample Data:

Count

		Technology_Efficacy					Total
		1	2	3	4	5	
Performance	1	72	9	3	0	0	84
	2	7	64	17	3	0	91
	3	0	20	51	19	0	90
	4	1	5	8	29	2	45
	5	0	0	0	4	8	12
Total		80	98	79	55	10	322

Download SPSS

Free version

More Exercises

- Refer to the data sets I gave.

Descriptive:

- What is socio-demographic profile of the respondents?
- What is the technologic profile of the respondents?
- What is the level of competency of ICT tools as perceived by the respondents?
- What is the level of competency of ICT concept as perceived by the respondents?
- What is the level of competency of ICT operations as perceived by the respondents?
- What is the training motivation profile of the respondents?
- What is the **Social Technographic Profile of the respondents?**

More Exercises

- Refer to the data sets I gave.

Correlation:

- What is the relationship between socio-demographic profile of the respondents and
  - level of competency of ICT concept
  - level of competency of ICT operations
  - training motivation profile
  - Social Technographic Profile?

More Exercises

- Refer to the data sets I gave.

Correlation:

- What is the significant difference between technologic profile of the respondents and
  - level of competency of ICT concept
  - level of competency of ICT operations
  - training motivation profile
  - Social Technographic Profile?

References

- De Leon, R.O. Introduction to Statistics. Slides Presentation. Silliman University
- http://www.mheducation.co.uk/openup/chapters/9780335227242.pdf
- Calderon, J. F. and Gonzales, E. C. (1993). Methods of Research and Thesis Writing
- http://wps.prenhall.com/hss_salkind_exploring_5/4/1035/265001.cw/index.html
- http://experientia.com/services/understanding/ethonographic-research/
- **The Role and Importance of Research.** http://wps.prenhall.com/hss_salkind_exploring_5/4/1035/265001.cw/index.html
- The Foundations of Research. http://www.socialresearchmethods.net/kb/intres.php