

Tony's Research

Introduction to Research Statistics 1:

December 2006



Tony (Research Student)



Rose (Statistician)

By

Professor *Courtney A. Pindling, PhD*

Introduction

The purpose of this textbook is to provide a supplementary text that forms the basis of Tony's research. Tony's Research is a multimedia presentation of an introductory statistical course. The main purposes and strategies of this learning media is to guide students working on their masters or doctoral researches in statistics to understand fundamental research statistics, develop skills in analyzing data statistically using appropriate statistical tools, and help researchers draw conclusions and make decisions based on statistical analyses. The foundation of Tony's research is to teach by examples and demonstration of applications of statistical methodology, software, and interpretation of results. The learning resources consist of four major educational deliveries: 1. an online textbook, 2. stream lectures and notes, 3. tutorials on the use of statistical software, and 4. interactive workshops. Tony's Research follows Tony has he learns and implements statistics for his graduate research with the aid of a mentor, Rose. In Tony's Research, the student asks all the questions and the mentor provides the solutions and answers. May all your courses be so easy!

Online Textbook

The online textbook provides comprehensive summaries of the major topics of an introductory statistical text. It is part 1 of a 2-part course plan. It is intended as a supplemental resource to a more detailed statistical textbook. The objectives of the

textbook are to introduce the various statistical topics and to show how the SPSS (or Stats4U) program is used to perform various basic statistical analyses. The resource is divided into three parts: 1. data inputs and outputs with SPSS (tutorials), 2. descriptive statistics (describing data), and 3. inference statistics (making inference about data).

The text is organized into 15 chapters. Each chapter is further organized as follows: 1. A brief introduction of topic, 2. Definitions of new terms throughout text (see also glossary), 3. Illustrated examples presented in conversational language, and 4. Formula summaries.

The section on central tendency deals with statistics used to describe a typical data value for a data set. It employs such statistics as: mean, median, mode, and frequency polygon (shows mode, sometimes mean and median).

The section on variability deals with statistics that describe the variability of a sample distribution with such measures as: range, standard deviation, and variance.

The standard score and normal curve allow one to make statements about how far a data point is from its mean and estimate the probability of other points relative to the mean.

While descriptive statistics describes data, inference statistics makes predictions or inference about data. The t -test is a tool useful for comparing the mean of a data set to some constant value or comparing the means of two samples or distributions. We also use the t -test to compare both independent samples and correlated samples. The t -test assumes that both data sets are fairly normally distributed, but it can be used for non-normal data sets within certain limits.

The Pearson Chi-square test is used to measure the degree of associations between two or more categorical data sets (based on observed versus expected frequencies); the data are often nominal or ordinal numbers. This is a distribution independent analysis; however, the samples must be random samples.

The null hypothesis states that what are being compared are the same – means or distributions. We reject the null hypothesis if: a. the significance of the test (p -value) is less than 0.05 (95% confidence or 90%: $\alpha = 0.10$), b. the test statistics is greater than the table lookup value, and c. the confidence interval for the test does *not* contain zero (for differences) or the hypothesized population parameter value. Table 1 shows a summary of the statistics used in this workbook.

Table 1 *Summary of Statistical Topics*

Statistics	Description	Remarks
Descriptive Statistics		
Central Tendency	Typical Value	Mean, median, mode, and frequency
Variability	Spread of Distribution	Range, variance, and standard deviation (std dev.)
Standard Score	Number of standard deviations from the mean	68%: mean \pm 1 std dev. 95%: mean \pm 2 std dev. 99%: mean \pm 3 std dev.
Normal Distribution	Bell-shape, symmetric distribution	Assumed by most statistics
Correlation	Measures relationships between 2 variables	Strength and direction of relationship, r
Inference Statistics		
One-sample t -test	Compares mean of a variable to a constant value	Different if p -value $<$ 0.05
Two-sample t -test (independent)	Compares two sample means	Different if test statistics $>$ table lookup value
Correlated t -test	Compares means of two related samples	Different if confidence interval does not contain 0
Pearson Chi-square	Measures associations between categorical data	Nominal numbers

Streamed Lectures

Short streamed lectures are presented that covers the essential of each topic. Some topics are broken up into several mini-lectures so as to shorten Internet download time. Lectures are streamed to facilitate dialup users. Once a lecture is loaded, it can be fast-forward and replay as often as desired to reinforce learning. At least one complete example of a topic's statistical analysis is presented in a lecture to demonstrate implementation. Each lecture is accompanied by a *pdf* file that students can save and use as notes for the streamed lectures.

Statistical Software Tutorials

A streamed demonstration of how to use the statistical software to perform the statistical analysis covered by each topic is presented for two statistical packages: 1. SPSS version 13 or above and 2. Stat4U: a free statistical software with identical features and functions as SPSS. It is assumed that students, using Tony's research, have access to MS word and MS Excel and is able to use the basic features of these applications or have access to SPSS and other word processing application.

Interactive Workshop

After reading the text and previewing the lectures and tutorials, it is expected that students will complete a topic's workshop assignment. Each workshop assignment consists of ten questions that try to assess concept mastery, applications of statistics skills and demonstration of the usage of either SPSS or Stats4U statistical software. At least

eight of the questions have similar solutions illustrated in both the text and lecture or tutorial. Further illustrations are provided through examples of MS Excel demonstrations of numerical solutions to problems students are asked to solve manually (not using statistical software) or by using computational formulas. These are provided as guide to help students check their computational work and skills.

The Table of Content has a separate links to each chapter's learning or instructional resources in *pdf* form at: <http://www.pindling.org/Math/Statistics1>.

Table of Content

I. Descriptive Statistics	II. Inference Statistics
1. Scales of Measurement	6. Introduction to Hypothesis Testing
2. Central Tendency Frequency Distribution Stem and Leaf	7. Comparing One-Sample Means
3. Variability Outliers - Box Plot	8a. One-Sample Correlation ($\rho = 0$)
4. Standard Scores Normal Curve	8b. One-Sample Correlation ($\rho = a$)
5. Correlation	9. Two-Sample Means (Independent and Homogeneous Variances)
	10. Two-Sample Means (Independent and Non-Homogeneous Variances)
	11. Two-Sample Means (Dependent or Correlated)
	12. Linear Regression
	13. Chi-Square Goodness of Fit Test
	14. Chi-Square Test for Independence