NON-PARAMETRIC STATISTICS

Inference Methods for Qualitative Research

Philibert C. Ndunguru

NON-PARAMETRIC STATISTICS

Inference Methods for Qualitative Research

Philibert C. Ndunguru

NON-PARAMETRIC STATISTICS

Inference Methods for Qualitative Research

Philibert C. Ndunguru, Ph.D.,
Professor of Business Administration and Statistics and Dean of the Faculty of Science and Technology, Mzumbe University

Published by
Mzumbe University

ISBN 9987 – 617 – 82 – 4

BASIC FORMULAS IN NON-PARAMETRIC STATISTICS

Binomial / Sign Test

$$\alpha_c = \sum_{B=B^*}^{n} \binom{n}{B} p^B (1-p)^{n-B} \quad \text{(One-tailed test)}$$

$$\alpha_c = \sum_{B=0}^{n-B^*} \binom{13}{B} \left(\tfrac{1}{2}\right)^B \left(\tfrac{1}{2}\right)^{13-B} + \sum_{B=B^*}^{n} \binom{13}{B} \left(\tfrac{1}{2}\right)^B \left(1-\tfrac{1}{2}\right)^{13-B}$$

(Two-tailed test)

$$Z_c = \frac{B^* - np}{\sqrt{np(1-p)}} \quad \text{(Normal approximation of binomial)}$$

K-sample Median Test

$$\chi^2 = \sum_{i}^{k} \sum_{j}^{2} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$P(M_x = m_x) = \frac{\binom{m}{m_x}\binom{n}{\frac{m+n}{2} - m_x}}{\binom{m+n}{\frac{m+n}{2}}}$$

Kolmogorov-Smirnov Test

$$D_{nc} = \max\left|F_e - F_o\right|$$

$$D_n = \begin{cases} D_{n,\alpha} & for\ n \le 35 \\ \dfrac{1.36}{\sqrt{n}} & otherwise \end{cases}$$

$$Z_c = D\sqrt{n}$$

Wilcoxon Signed-rank Test

$$T = \min\left(R^+, R^-\right)$$

$$E(T) = \frac{n(n+1)}{4}; \quad V(T) = \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Mann-Whitney Test

$$U = \begin{cases} U_1 & if\ n_1 \le n_2 \\ U_2 & otherwise \end{cases}$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1; \quad U_2 = n_2 n_1 + \frac{n_2(n_2+1)}{2} - R_2$$

$$E(U) = \frac{n_1 n_2}{2}; \quad V(U) = \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Kruskal-Wallis Test

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{n_i}\right) - 3(n+1)$$

Friedman Test

$$F_{rT} = \frac{12}{pk(k+1)}\left[\sum R_i^2\right] - 3p(k+1) \text{ or}$$

$$F_{rT} = \frac{12}{pk(k+1)} S_T, \text{ where: } S_T = \sum_{i=1}^{k} R_i^2 - \frac{\left(\sum_{i=1}^{k} R_i\right)^2}{k}$$

Runs Test

$$E(R) = \frac{2n_1 n_2}{n_1 + n_2} + 1; \quad V(R) = \sigma_R^2 = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}$$

Chi-square Test of Equal Proportions

$$\chi^2 = \sum_i^r \sum_j^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \Rightarrow \mathrm{N}(0,1)$$

Note: $P_1 = \dfrac{o_{11}}{n_1}$, $P_2 = \dfrac{o_{21}}{n_2}$ and $P = \dfrac{o_{11} + o_{21}}{n_1 + n_2}$

Chi-square-based Correlations

$$\phi = \sqrt{\frac{\chi^2}{N}}.$$

$$c = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

$$r = \sqrt{\frac{\chi^2}{N(r-1)}}$$

$$\upsilon = \sqrt{\frac{\chi^2}{N\min(r-1,c-1)}} = \sqrt{\frac{\phi^2}{\min(r-1,c-1)}}$$

$$\mathrm{T} = \frac{\chi^2}{N\sqrt{(r-1)(c-1)}} = \frac{\phi^2}{\sqrt{(r-1)(c-1)}}$$

Yule's Q coefficient $Q = \dfrac{o_{11}o_{22} - o_{12}o_{21}}{o_{11}o_{22} + o_{12}o_{21}}$

Fisher's Exact Probability

$$P(a,b) = \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{N}{a+b}}$$

Predictive Efficiency Indices

$$\lambda_p,\ \tau_p,\ \phi_p = \frac{\text{Errors without model} - \text{Errors with model}}{\text{Errors without model}}$$

$$\lambda_p = \frac{\sum_{i=1}^{r} o_{im} - \max(C_j)}{N - \max(C_j)}$$

$$\sigma^2(\lambda_p) = \frac{\left(N - \sum_{i=1}^{r} o_{im}\right)\left(\sum_{i=1}^{r} o_{im} + \max(C_j) - 2o_{jm}\right)}{(N - \max(C_j))^3}$$

$$\tau_p = \frac{\sum_{j=1}^{c} C_j\left(\frac{N - C_j}{N}\right) - \sum_{i=1}^{r}\sum_{j=1}^{c} o_{ij}\left(\frac{R_i - o_{ij}}{R_i}\right)}{\sum_{i=1}^{c} C_j\left(\frac{N - C_j}{N}\right)}$$

$$\phi_p = \frac{\frac{(o_{11}+o_{12})(o_{12}+o_{22})}{N} + \frac{(o_{11}+o_{21})(o_{21}+o_{22})}{N} - (o_{12}+o_{21})}{\frac{(o_{11}+o_{12})(o_{12}+o_{22})}{N} + \frac{(o_{11}+o_{21})(o_{21}+o_{22})}{N}}$$

Pseudo F-Test for PRE-based Coefficients

$$d = \frac{(P_e - p_e)}{\sqrt{\frac{P_e(1-P_e)}{N}}}$$

$$P_e = \frac{errors\ without\ model}{N}$$

$$p_e = \frac{errors\ with\ model}{N}$$

Light-Margolin measure

$$R^2_{psdo} = \frac{SSR}{SST} = \frac{\left(\sum_{i=1}^{r} \frac{1}{R_i} \sum_{j=1}^{c} o_{ij}^2\right) - \frac{1}{N}\sum_{j=1}^{c} C_j^2}{N - \frac{1}{N}\sum_{j=1}^{c} C_j^2}$$

Cross-product Ratio Test for Retrospective Design

$$\theta = \frac{o_{11}o_{22}}{o_{12}o_{21}}$$

$$\sigma^2(\ell n\,\theta) = \frac{1}{\sum\sum o_{ij}} = \frac{1}{o_{11} + o_{12} + o_{21} + o_{22}}$$

$$\ell n\,\theta \pm Z_{1-\frac{\alpha}{2}}\,\sigma(\ell n\,\theta)$$

Kappa coefficient

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{\sum o_{ii} - \sum e_{ii}}{\sum\sum o_{ij} - \sum e_{ii}}$$

Correlation Coefficients for Ordinal Data

$$r_s = 1 - \frac{6\sum d_i^2}{N(N^2 - 1)} \qquad where: d_i = rank\ differences$$

$$Z = \frac{r_s - 0}{1/\sqrt{N-1}} = r_s\sqrt{N-1}$$

$$\tau_a = \frac{C - D}{\frac{1}{2}N(N-1)}$$

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_Y)}}$$

$$\tau_c = \frac{(C - D)}{\frac{1}{2}N^2[(m-1)/m]}$$

$$\gamma = \frac{C - D}{C + D}$$

$$d_y = \frac{C - D}{C + D + T_y} \qquad or \qquad d_x = \frac{C - D}{C + D + T_x}$$

$$C = \sum_{i=1}^{r-1}\sum_{j=1}^{c-1} o_{ij}\left(\sum_{k=i+1}^{r}\sum_{m=j+1}^{c} o_{km}\right)$$

$$D = \sum_{i=1}^{r-1}\sum_{j=2}^{c} o_{ij}\left(\sum_{k=i+1}^{r}\sum_{m=1}^{j-1} o_{km}\right).$$

$\sigma_S = \sqrt{\frac{1}{18}N(N-1)(2N+5)}$ for all $S = C - D$

$$Z = \frac{S-0}{\sigma_S}$$

DEDICATION

This book is dedicated to my aunt Beligna Kosani Ndunguru and my Teachers at Mbuji and Litembo Primary Schools for their role in shaping my formative schooling life.

CONTENTS

PREFACE

The long experience, with internal and external examination of thesis and/or dissertation projects by graduate students, and in the peer-reviewing of research reports by university faculty members, reveal a major enigmatic aspect that: *although a large percentage of such research projects are qualitative, one hardly finds non-parametric techniques in their analysis; and researchers attempting have largely used the famous Chi-square test technique for independence*. For instance, despite having inference stance, very few MBA/MPA/MSc thesis projects, submitted at Mzumbe University since 1988, used non-parametric inference techniques. Why is the use of non-parametric technique so low in qualitative research projects undertaken by university graduates and faculty members?

The low level of application of non-parametric inference techniques, according to the author, seems to be associated with lack of a comprehensive research-based and localized text on non-parametric techniques. Most approved and recommended text-books on statistics dwell solely on parametric inference statistical techniques; and there are few, a single-digit-percentile, that have a chapter or section devoted for elementary non-parametric techniques. This bias creates an impression among researchers and students that parametric techniques are the only inference methods to rely on; a view that is parochial and misleading. *Non-parametric Statistics: Inference Methods for Qualitative Research* is a book that fills that gap fitly. The book is meant to alleviate an acute shortage of a comprehensive research-based and localized material on non-parametric techniques for business, economics, management, and other social sciences. Examples and/or exercises given and solved are based on data that are research-based and this orientates a reader as to the power of non-parametric techniques for research in business and economics.

Besides, *Non-parametric Statistics: Inference Methods for Qualitative Research*, addresses the problem of stereotype that is being associated with over-reliance and over-use of parametric techniques for both descriptive and inference research studies. In such applications, researchers are obliged to spend more time and energy and money looking for large samples; looking for studied populations that are normally distributed; making sure that data are quantified – measured at interval or ratio scale. These efforts are very much appreciated because they foster and enhance interpretation, reliability, and validity of research reports. However, and certainly in some applied work, small samples from non-normal populations could give better results that are reliable and valid, let alone being less costly. It is in this aspect that non-parametric statistical techniques are considered and comprehensively discussed in this book. This book provides a comprehensive and more focused research-based material on non-parametric statistical techniques.

The book is structured as follows: Chapter one introduces the basic concepts of a statistical inference problem. Chapter two focuses on inference techniques that are appropriate for nominal data. Chapters three and four dwell on inference techniques using ordinal data; with chapter three focusing on sample median test techniques and chapter four on rank-sum techniques. Both techniques typically rely on data that are capable of being ranked. Chapter five is exclusively presenting material on testing randomness of a set of sample observations, which is a universal premise for all inference methods. The technique adopted in this book is based on Wald-Wolfowitz one-sample runs test procedure. The final chapter dwells on correlational inference techniques that are appropriate for nominal and/or ordinal data.

The author has streamlined the material to be reader-friendly. The solved examples, exercises, and review questions that are drawn from real and contextualized business, economics, management, and social life situations enable readers to understand the topical concepts covered. The optional exercises on parametric techniques enable readers to reflect on the link between parametric and non-parametric statistical techniques as applied in social science research. Chapters two through six each contain a section on SPSS tutorial, in which the use of SPSS software in solving non-parametric inference problems is demonstrated. In this way, readers get to know and/or exercise their skills in the application of the statistical package.

It is the expectation of the author that the material presented in this book will popularize the use of non-parametric statistical techniques in applied business, management and other social science research projects among university graduates and researchers.

P. C. Ndunguru, *Ph.D.*

Mzumbe University
January, 2009.

1

BASIC CONCEPTS OF INFERENCE IN SOCIAL RESEARCH

The traditional inference methods in research, based on parametric statistical techniques, have strict assumptions for their use. The assumptions include, among others, normal distribution, equal variance, and interval/ratio scale-measured data regarding the phenomenon understudy. These assumptions are generally not met in applied business, economics, and social sciences. Statistical theory provides alternative valid and reliable inference research procedure when data are qualitative or made to be qualitative. This chapter reviews some basic concepts that are related to social science research process.

1.1 *Conceptualization and Operationalization*

A scientific social research study project starts with a conceptualization process. The process entails listing minimum number of *verbs* and *nouns* which describe a studied phenomenon. The studied phenomenon may focus on exploring, describing, explaining comparing, or predicting some traits, characteristics or properties of objects, events or behaviour of individuals. Thus, a research concept is a verb or noun that is used to describe *objects*, *situations*, *events*, *individuals* implied in a research subject being investigated. For instance, management, enterprise performance, democracy, crime, sin, etc., are concepts that may be used to describe a studied phenomenon. However, social science research concepts have multi-valued meanings to different people at different times and/or places.

Conceptualization then is a process of giving **literal** as well as **scientific meanings** of research concepts as well as exploring **how these concepts relate to each other**. A research concept is defined in terms of other concepts already known, literally or scientifically. For instance, a scientific definition of a concept in economics means what economists know and generally agree to be the meaning of the concept in the discipline. The known concepts, which are used to define a new research concept, are referred to as **primitive concepts**. Critical and focused literature review often leads to, having in place, relevant and appropriate definitions of concepts for the purpose of undertaking research.

Concepts must be translated into observable variables, a process known as **operationalization**. Thus, operationalization is about creating indicator-variables that give evidence of presence or absence of observable instances of a concept. The observable instances may be traits, characteristics or properties of objects, situations or events. During data collection, numbers or symbols are assigned to the traits, characteristics or properties of the objects, situations, or events. The assignment of numbers or symbols, which is in essence a measurement process, enables recording responses of respondents on a research variable.

A **variable** is a trait, characteristic or property of object, event, or individual that can be observed or made to be observed. It is an observable property that a set of all study cases in a target or sampled population possess in some degree; some have more while others have less of the property. Variables in a research, like concepts, are defined in terms of other variables or concepts already known by members of a given scientific community. In defining a variable, a researcher decides the exact property and provides a set of standard procedures by which presence or absence of a property can be determined reliably in each study case.

Attitudes, images, decision-making style, life styles, affiliation, expectations, opinions, motivations, achievement, and hostility constitute behavioural units or objects in a social science research. These units have less identified traits, characteristics or properties; in the sense that they cannot be directly observed or measured. As such, the meanings of these behavioural units must be constructed and reconstructed. For instance, to study hostility empirically, one must reconstruct its meaning based on *logical grouping of some primitive concepts*. One such reconstruction of a meaning is a boy *hitting his play-mates frequently* indicates *hostility* among the play-mates. In the process of construction and reconstruction, several indicants or items may be used to measure a construct or concept. For instance, *democracy* as a construct can be measured by: *fairness of election*, *freedom of press*, *freedom to organize*, *rights of minority*, and/or *regularity of elections*.

In a research project, and as a rule of thumb, research concepts are identifiable in the stated broad research objectives and/or questions. Besides, well formulated statement of research problem, regarding objectives and/or questions, contains indicator-variables, on which data are to be collected. Furthermore, it is very pertinent for a researcher to distinguish variables used in a research project. A variable in a research project may be dependent, independent, moderating, intervening or extraneous; and this distinction or classification of variables enable a researcher to make choice of data analysis and/or the interpretation of the results, findings and conclusions. The following tables give summaries of the ideas of conceptualization and operationalization of research concepts and research variable typology.

Table 1.1: Concepts and indicator-variables in a research project

Research concept	Indicator-variable
Social class	Type of occupation or residence in a social setting
Governance	Rule of law; Democratic institutions; Freedom of press
Enterprise-performance	Profit level; Market share; Labour productivity
Citizenry-empowerment	Education level; Social protection; Participation

Table 1.2: Typology of variables

Type	Description
Dependent	Variable whose changes result from outside cause or causes
Independent	Variable that causes or pressurizes others to change
Moderating	Variable that has contingency effect on the originally stated IV-DV relationship
Intervening	Non-measurable factor that theoretically affects the observed phenomenon, but whose effects are inferred from independent and moderating variables
Extraneous	Comprises many factors that are not included in a research design and if not controlled offer plausible alternative explanation to the observed variability or confound interpretation of observed relationship

Thus, conceptualization is very fundamental to any scientific research in the following sense:-

- Research problem is stated and justified in terms of *concepts*.
- Research questions, objectives, significance, and its scope are stated and rationalized in terms of *concepts*.
- Review of theoretical and empirical literature is done on *concepts* that are already conceptualized.
- Theories and empirical studies, relating to a studied problem and helping to underscore research *concepts*, are selected on the basis of conceptualization.

- Research model or framework that elicits variables and their relationships is about *concepts* of the conceptualized study.
- Choice of target population, data collection-design methods precedes conceptualization.
- Operationalization and/or measurement of variables on which data are collected are accomplished in terms of *concepts*.
- Choice of data analysis framework precedes conceptualization.
- Conclusions are framed, rationalized, and reported in terms of *concepts*.
- Throughout the research process the conceptualized *concepts* guide research activities.

1.2 *Measurement and Measurement-scales*

Quality of data is a fundamental issue of great concern for research as it ensures reliable and valid research results, findings, and conclusions. Data that are used in research are obtained through the process of **measurement**, which is the *assignment of numbers or symbols to property, characteristic or a trait of an object, event, or individual according to rules*. The objects, events or individuals possess property, trait, or characteristic that can be a subject of empirical investigation. The assigned number or symbol corresponds to the *degree* or *intensity* of the property possessed by a study case.

How are the numbers or symbols assigned according to the intensity and/or degree of the presence of an empirical property? Assigning numbers or symbols to objects or events that indicate intensity implies that there is a scale: *an instrument that associates numerical values with qualitative attributes*. For instance, one may construct a scale to measure opinion of the acceptability of a new environmental policy to be introduced in a social setting. The numerical values of scale may take a range of 0 to 5 (0 to 10 or 0 to 100); *with 0 indicating not acceptable at all*, and 5 (10 or 100) *indicating very much acceptable*.

Measurement, the construction of a scale, in science involves the following tasks:

- **Classification**, which is about determining a *difference in kind* regarding presence or absence of a property among objects or events.
- **Ordering**, this is about determining a *difference in degree* regarding presence or absence of a property among objects or events.
- **Finding exact differences** of the presence or absence of a property among the objects using a scale.
- **Comparing scores by taking their ratios** so that the statement "one score is twice as high as another" makes sense.

On the basis of measurement, four types of data, namely, *nominal*, *ordinal*, *interval*, and *ratio* scales are generated based on the measurement tasks outlined above. Classification is the most primary level-task of measurement and the data generated are referred to as *nominal-scaled*. This is followed by ordering, in which apart from finding a difference in the property characterizing an object, there is a possibility of ranking. Ranking refers to the possibility of being able to identify a difference in degree of a property presence in an object or event. The resulting data of this second level-task of measurement are referred to as *ordinal-scaled*. Finding exact difference regarding the presence or absence of a property that characterizes an object is at the third level-task of measurement. The task generates interval-scaled data that are capable of being manipulated by using basic arithmetic operations of addition and subtraction. Finally, with ratio-scaled

measurement level, ratios of scores can be compared because the measurement provides scores of data that are capable of being represented by real numbers. This implies that there is a possibility of performing arithmetic operations of multiplication and/or division on ratio-scaled data.

The task of measurement, scale construction, is relatively easy with regard to **primary** rather than **secondary qualities of study objects**[1]. The primary qualities are the real qualities that belong to bodies or matter such as, size, position, density, weight, and height. Primary qualities, as such, can be mathematically measured and manipulated. However, constructing a scale is often difficult when dealing with secondary qualities of study objects. Secondary qualities include colour, tastes, emotions, opinions and attitudes. Others are motivations, integrity, love, faithfulness, kindness, friendly, hostility, beauty, and all other behavioural units or objects in social science research. These secondary qualities reside only in consciousness and do not participate in any way, for most of the time, in the realm of reality. In this respect, mathematical measurement and/or manipulability of secondary qualities are remote.

Each of the four types of data, nominal, ordinal, interval or ratio, dictates a different kind of analysis framework with respect to inference procedures. For instance, parametric inference methods are only appropriate with interval or ratio scale-measured data. Non-parametric inference methods on the other hand are appropriate with nominal and ordinal scale-measured data. Given that each high measurement level incorporates all the properties of the lower measurement levels, non-parametric inference methods are equally used for analysing ratio-scaled data.

Statistical inference methods seek to ascertain or test **descriptive** or **relational** hypothesis. Descriptive hypothesis states or describes existence,

---

[1] Galileo (1564 – 1642) introduced the concepts of primary and secondary qualities of objects, distinguishing between those that belong to physical bodies or matter and those that reside only in consciousness. See Adolf Mihanjo (2005) for more details on these concepts.

size, form, difference, homogeneity, skew-ness or distribution of a variable or set of variables under study with respect to case(s). Relational hypothesis seeks to describe, determine or test a relationship between variables with respect to some case(s). Focused theoretical and/or empirical literature review enables a researcher to comprehend the kind of attributes or relationships between and/or among variables to consider in a study. The relationships between and/or among research variables in a research project being referred to may be:

- Symmetric relationship in which two variables are alternate indicators of a third target variable in a studied relationship. Part or partial correlation analysis is used to determine existence of symmetric relationship.
- Reciprocal relationship in which two variables mutually influence and/or reinforce each other. Raw correlation analysis is used to ascertain reciprocal relationship.
- Asymmetric relationship typically postulates a cause-effect phenomenon; and regression, ANOVA/MANOVA techniques are used to explore asymmetric relationships.

Focused theoretical and/or empirical literature review enables researchers to determine the kind of descriptive variable-attributes and relationships among study variables to be tested. This book explores the use of non-parametric techniques in testing both descriptive as well as relational hypotheses when research data are qualitative or made to be qualitative.

1.3 *Basic Inference Problem in Qualitative Research*

Inference statistical methods are divided into parametric methods and nonparametric methods. Parametric methods deal with the study of sampling distribution of statistics, such as, sample mean $\left(\overline{X}\right)$, sample variance $\left(S^2\right)$, difference of sample means $\left(\overline{X}_1 - \overline{X}_2\right)$, ratio of sample variances $\left(\frac{S_1^2}{S_2^2}\right)$, or sample correlation coefficient (r). These statistics are studied with the aim of testing some hypotheses concerning corresponding population parameters, such as, mean (μ) or variance $\left(\sigma^2\right)$ for a variable phenomenon being studied. As such parametric techniques are applicable with *interval* or *ratio* scales of measurements. The basic assumptions of using parametric tests are:

- Observations are drawn at random from the sampled population, so that resulting sampling errors are uncorrelated.
- The variable phenomenon that is characterizing the sampled population has a normal distribution.
- If there is more than one sampled population, variances of the variable phenomenon that are characterizing the several populations are equal – the sampled populations are homogenous with respect to variability.
- Observations are measured at interval or ratio scale.

There are many applied cases in business, economics, and other social sciences for which the above assumptions of normality, homogeneity, or ratio-scale measurement of sampled population are not met. For instance, more often than not, researchers are confronted with inference problems in which assumptions underlying the use of classical statistical techniques are not met; or study features such as randomness, trend, independence, symmetry, or goodness of fit characterize the inference problem. Besides, there are cases in which measurements characterizing the inference problem are **qualitative**. An atypical inference problem of the sort is presented below.

Let $x_1, x_2, \cdots, x_n$ be a random sample of size n from cumulative density function $F_x(\bullet)$ with corresponding density function $f(x)$. Let also $y_1, y_2, \cdots, y_n$ be another random sample from cumulative density function $F_y(\bullet)$ with density function given by $f(y)$. Assume that observations from $F_x(\bullet)$ are independent of the observations from $F_y(\bullet)$. A pertinent question is: Are attributes of the two samples providing sufficient evidence to indicate that the two populations are the same? Put differently, to what extent are the two populations equal or nearly the same?

In a more formal way, the above problem boils down to testing a null hypothesis that $H_0 : F_x(Z) = F_y(Z)\ \forall Z$ against the alternative hypothesis $H_a : F_x(Z) \neq F_y(Z)$ *for at least one* Z. If $H_0 :$ is true, there are two independent estimators for an attribute or a set of attributes; one based on $F_x(Z)$ and the other based on $F_y(Z)$. Thus, a common estimator can be determined by using $F_x(Z)$ and another by using $F_y(Z)$. Conceptually then, the above problem may be stated in form of a question as: *How do we ascertain that the common estimators based on the unknown cumulative density functions $F_x(Z)$ and $F_y(Z)$ are the same?*

In a parametric test environment, Z represents set of numerically measurable attributes or parameters, such as, population mean (μ) or

population variance (σ^2). However, in non-parametric test environment Z represents qualitative set of attributes, which is typical in qualitative research projects. Qualitative research places emphasis on understanding through looking closely at people's words, actions, and records; and discovering patterns which emerge after close observation, careful documentation and analysis of data. On account of this, qualitative research projects are characterized by high-level theories and a measurement level that is too low to be manipulated mathematically.

Thus, statistical inference problems in qualitative research are mainly addressing **theory-building or grounded-theory**. The focus of inference methods in theory-building or grounded theory is on testing both descriptive and/or relational hypotheses and propositions, when research data are qualitative or made to be qualitative. Theory-building is about determining, identifying or ascertaining existence of a relationship among study variables on the basis of data rather than reviewed existing theories.

The focus of theory-building is about, firstly, determining or **identifying a difference in kind** regarding presence or absence of a property among objects or events. Secondly, it is about **determining a difference in degree** regarding presence or absence of a property among objects or events. These aspects of theory-building are respectively within the realm of classification and ordering dimensions of measurement theory. And, the fundamental idea of measurement theory in social science is the contention that when you can measure what you are talking about and express it in numbers, you know something about it. In conducting non-parametric statistical inference a researcher is providing evidence of what is known by measuring what is being talked about and expressing it in numbers.

1.4 Testing of Statistical Hypothesis

A statistical hypothesis is an assertion or conjecture about a distribution of one or more random variables. As an assertion or conjecture, a hypothesis is a tentative answer to a research question that is being subjected to empirical investigation[2]. If the hypothesis defines or specifies the distribution completely, then it is called *simple hypothesis*; otherwise it is known as *composite hypothesis*. For instance, given a random sample, $x_1, x_2, \cdots, x_n$ the hypothesis that "*the sample comes from a normal distribution*" is a composite hypothesis because it does not specify the distribution in terms of its parameters such as mean and variance. However, if the distribution is completely specified, we have a simple hypothesis, for instance, a hypothesis that "*the sample comes from a normal distribution with mean and variance of* $\mu = 36$ *and* $\sigma^2 = 9$ *respectively*" is a simple hypothesis. Inference problems in qualitative research are, in the main, characterized by testing composite hypotheses that may have either descriptive and/or relational stance.

A hypothesis is also distinguished as being a null or an alternative depending on the background knowledge of the time. A **null hypothesis** is a *plausible* assertion or conjecture about a population attribute in the light of background knowledge of the time. On the other hand, an **alternative hypothesis** is a *speculative*, *bold* or *novel* assertion about the population attribute in the light of existing theoretical structures of the time. An alternative hypothesis; is *speculative* if it is offering a more viable explanation to a population attribute that is yet to be fully and comprehensively known; is considered *bold* if it contradicts or conflicts with the generally accepted theories of the time; and it is *novel* if it predicts some

[2] The term *proposition* is used to describe a hypothesis that is formulated for a conduct of practical affairs rather than being subjected to empirical investigation. For more details on the concept of hypothesis, see Ndunguru, P. C. (2007), *Lectures on Research Methodology for Social Sciences*, Mzumbe University: Research and Publications.

new phenomenon that is neither touched nor ruled out by theoretical structures of the time. Put differently, an alternative hypothesis is speculative, bold and novel if accepting it leads to new knowledge.

Thus, test of a statistical hypothesis is a rule or procedure that enables to decide whether or not to reject a null hypothesis. The focus in statistical hypothesis testing is a null hypothesis so that accepting it is tantamount to rejecting the alternative and vice versa.

In conducting a statistical test two errors are committed, namely, *unjust rejection* or **type I error** and *wrong acceptance* or **type II error**. Type I error is about concluding that there is a relationship when there really is none, and type II error is a failure to detect a relationship that exists. Both errors can not be reduced completely. Probability of committing type I error is called level of significance or size of the test and this probability is denoted by α, and $1-\alpha$ is called level of confidence. The probability of committing type II error is denoted by β, and $1-\beta$ is called the power of the test. Given that the focus of statistical hypothesis is a null hypothesis, type I and II errors and the level of significance and power test are summarized in the table below.

	Ho: True	Ho: False
Reject Ho:	Type I (α)	Correct decision ($1-\beta$)
Accept Ho:	Correct decision ($1-\alpha$)	Type II (β)

Furthermore, there are two types of statistical hypothesis tests namely, non-randomized and randomized tests. Non-randomized tests divide the whole sample space S into two non-overlapping regions, W and $S-W$ and the decision is to reject Ho if the given sample $S \in W$ and accept otherwise. Randomized test is specified by a probability function described by, for instance, $\psi_T(x_1, x_2, \cdots, x_n)$ such that:

$$T = \psi_T(x_1, x_2, \cdots, x_n) = \begin{cases} 1 \ \forall \ x_1, x_2, \cdots, x_n \in W \\ 0 \ otherwise \end{cases}$$

In this case, $\psi_T(x_1, x_2, \cdots, x_n) = \Pr ob\{rejecting \ Ho\} = p,$ and $\psi_T(x_1, x_2, \cdots, x_n)$ is referred to as the critical function. For instance, if $S = (x_1, x_2)$ represents a set of sample values, a critical function $\psi_T(x_1, x_2) = \frac{x_i^2}{x_1^2 + x_2^2} = p$ may be formed such that the null hypothesis is rejected if $\psi_T(x_1, x_2) = \frac{x_1^2}{x_1^2 + x_2^2}$. In this case, the null hypothesis will be rejected with a probability of $\psi_T(x_1, x_2) = \frac{x_1^2}{x_1^2 + x_2^2} = \frac{25}{25 + 64} = 0.28$ when the set is $S = (x_1, x_2) = (5, 8)$

1.5 Power Function of a Test

Power of a test is the probability of rejecting an alternative hypothesis (Ha) when in fact it is true. In practice, one would like to minimize this probability. Note that when probability of type I error is zero, then null hypothesis is always accepted; and if probability of type II error is zero then null hypothesis is always rejected.

To illustrate the concept of a power function let $x_1, x_2, \cdots, x_N$ be a finite population from a normal distribution $N(\mu, \sigma^2)$, whose mean-parameter is unknown while its variance is $\sigma^2 = 25$. It is required to estimate the mean-parameter by conducting a statistical hypothesis testing.

The null hypothesis is Ho: $Ho: \mu \le \mu_0$ and the alternative hypothesis is Ha: $Ha: \mu > \mu_0$. The test criterion is to reject Ho if computed sample mean-statistic is $\overline{X} > \mu + \frac{5}{\sqrt{n}}$.

In general the unknown mean-parameter μ is unrestricted in the sense that the sample space for μ is $S = \{x_1, x_2, \cdots, x_N\} \in R^N \ \forall x_i \in R$. This implies that:

- The set $S = \{\mu = \frac{1}{N}\sum x_i \in (-\infty, \infty)\}$ defines the **unrestricted sample space**;
- The **accept-region for Ho** is defined by a set described by $W_0 = \{(x_1, x_2, \cdots, x_N) : \mu = \frac{1}{N}\sum x_i \in (-\infty, \mu_0)\}$ and;
- The **reject-region for Ho** is defined by a set described by $W_a = S - W_0 = \{(x_1, x_2, \cdots, x_N) : \mu = \frac{1}{N}\sum x_i \in (\mu_0, \infty)\}$.

To compute the probability of rejecting Ho when it is the true, the mean-parameter is fixed and only its value under the alternative hypothesis is allowed to vary. Fixing the parameter allows us to compare the probability values under different parametric values in the alternative hypothesis. In that respect, the size of the test is the maximum probability of rejecting null hypothesis (Ho) when $\mu \in \mu_0$. In formal terms, this probability is equal to:

$$Size \ of \ test = \max_{\mu \in \mu_O} \Pr ob\left(\overline{X} > \mu_0 + \frac{s}{\sqrt{n}}\right) = \alpha .$$

Furthermore, since sample data are used the unrestricted sample space, accept-region as well as the reject-region are evaluated on the basis of sample data, i.e., sample mean in the present case. Henceforth, given a sample, $x_1, x_2, \cdots, x_n$, from such a normal distribution;

- The **unrestricted sample space** is:

$$S = \{\overline{X} = \tfrac{1}{n}\sum x_i \in (-\infty, \infty)\};$$

- The **accept-region for Ho** is:

$$W_0 = \{(x_1, x_2, \cdots, x_n) : \overline{X} = \tfrac{1}{n}\sum x_i \in (-\infty, \mu_0)\} \text{ and;}$$

- The **reject-region for Ho** is:

$$W_a = S - W_0 = \{(x_1, x_2, \cdots, x_n) : \overline{X} = \tfrac{1}{n}\sum x_i \in (\mu_0, \infty)\}.$$

1.6 Non-parametric Techniques

Non-parametric techniques deal with comparing populations whose form of distributions are unknown, and as such there are no parameters to be tested or compared with. In this respect, non-parametric tests deal with composite hypothesis testing, and are therefore considered to be distribution free statistical test techniques. The focus of non-parametric techniques is to test whether or not a given sample characterizes a specified population. In that respect, nonparametric techniques are used even with *nominal* or *ordinal* data and do not make assumptions about values of the population parameters or about the shape of sampled population. However, sample observations for non-parametric testing must be drawn at random so that resulting sampling errors are uncorrelated. Thus, a working definition of non-parametric techniques may be summed up as:

- Statistical techniques whose test statistics do not depend upon the form of the underlying population distribution from which the sample is drawn.
- Statistical techniques which are not concerned with parameters of a population.

- Statistical techniques for which data have insufficient strength to warrant meaningful arithmetic operations.

Non-parametric techniques are broadly divided into: *sign, sample median, rank-sum, randomness and correlation* test techniques. Sign, sample median, rank-sum and randomness test techniques together have exploratory and descriptive stance. These test techniques mainly focus on exploring, describing, or comparing population characteristics or attributes in relation to some hypothesized attributes by testing descriptive hypotheses. The correlation techniques focus on unveiling causal relationships among attributes in a given population using nominal and/or ordinal data. On account of that, correlation test techniques take causal explanatory stance.

The general steps of conducting a non-parametric statistical test are the same as those involved in conducting parametric tests. They include:

- Formulating null and alternative $(H_0 \text{ and } H_a)$ hypotheses.
- Identifying the appropriate test technique by determining a test-statistic and its sampling distribution, usually $Z,\ t,\ \chi^2,\ F$ *etc.*. for parametric techniques. However, the common test-statistics for nonparametric techniques are: Binomial B, Chi-square, Kolmogorov-Smirnov, Wilcoxon Signed Ranks T and Mann-Whitney U. Others are Kruskal-Wallis H, Friedman, Runs R, and correlation and PRE-based statistics such as Goodman and Kruskal lambda-p, tau-p, Kendall's phi-p, Light-Margolin, and Kappa for nominal data. Others are Spearman rank correlation, gamma coefficient, Kendall's tau-a, tau-b, and tau-c for ordinal data.
- Specifying the level of significance – usually $\alpha = 0.05$.
- Specifying the ***accept – reject H_0 criterion*** – usually a condition under which H_0 is accepted.

- Computing the test-statistic based on sample data and comparing it with the theoretical value of its sampling distribution and presenting findings.
- Making informed conclusion about the null hypothesis being tested based on the results and findings.

1.7 *Concluding Remarks on Inference Techniques*

Test of significance plays two roles, which are first, to prevent the researcher from making a fool of him/her self and second, not to make un-publishable results publishable. Theoretically, parametric techniques involve estimating population parameters on the basis of sample observations. The basic assumptions of the use of the techniques are that: the form of distribution of errors is known to be normal distribution or can be made to be known; the sample of observations is random; sampled populations are homogeneous; and more important, measurement level for the sample observations is interval or ratio – that is sample data are quantitative. Experimental research typically fulfils these assumptions or conditions.

It is useful to note that when requirements of a normal population distribution are satisfied, non-parametric tests are generally less efficient than their parametric counterparts, but the reduced efficiency can be compensated for by an increased sample size. Below is a summary of efficiency rates of some selected nonparametric tests applied to normal populations[3].

[3] See Triola, Mario F. (1998), *Elementary Statistics*, 7e, Addison Wesley.

Application	Parametric	Non-parametric	Efficiency rate
Two dependent samples	Z or t test	Sign test	0.63
Two independent samples	Z or t test	Wilcoxon signed-rank test	0.95
Several independent samples	F test for ANOVA	Kruskal-Wallis and/or Friedman tests	0.95
Correlation	Pearson correlation test	Rank and PRE-based correlation tests	0.95

Observational or non-experimental research characterizes most scientific research projects in business, economics and other social sciences. The conditions surrounding the observational research environment have fewer tested assumptions. For instance, business, economics and other social sciences are characterised by the problem of isomorphism in which measurement procedures in reality may not correspond to theory and/or empiricism. As such the assumptions of normal population and quantitative data obtained, which together imply measurement, are not always met. Consequently, research projects in these disciplines are largely qualitative and therefore not susceptible to parametric inference techniques.

Thus, non-parametric techniques are useful under two conditions: first, when experimental observations are susceptible to classification and/or ordering but cannot be measured on a quantitative scale; and second, when uncertainty exists about validity of the assumptions in testing hypotheses that are associated with populations of quantitative data. Nonparametric techniques use both types of data that are measured at nominal or ordinal levels such as categorical data. Besides, nonparametric statistical tests are usually easy to understand and use. In this respect, if a non-parametric technique is available it should be used in preference to parametric technique unless there is experimental evidence about distribution of sampling errors being normal.

The next chapter of this book is focusing on sign-test-inference techniques that are appropriate with nominal data. Chapters three and four are respectively focusing on inference analysis of ordinal data, starting first with sample median test and then follow the rank-sum non-parametric statistical techniques. Randomness techniques are covered in chapter five while correlation techniques are presented in chapter six.

Problem Set One

Question 1.1

Distinguish the following concepts as they are used in inference statistical methods: a *hypothesis* and a *proposition*; *null* and *alternative hypothesis*; *type-one error* and *type-two error*, *simple hypothesis* and *composite hypothesis*; *power of a test* and *level of significance*; *substantive significance* and *statistical significance*; *randomized test* and *non-randomized test*; *parametric* and *non-parametric techniques*.

Question 1.2

Distinguish, with examples, between: (a) *conceptualization* and *operationalization* and (b) *operationalization* and *measurement* as used in social science research.

Question 1.3

Briefly, describe the steps involved in conducting a statistical hypothesis testing.

Question 1.4

Distinguish between qualitative and quantitative research as used in research methods.

2

INFERENCE TECHNIQUES FOR NOMINAL DATA: THE SIGN-TESTS

Sign tests are inference statistical methods for data that are capable of being categorized into two groups, such as, *bad versus good, high versus low, positive versus negative or plus versus minus.* In short, sign tests are applied whenever data have been *signed*; so that analysis is carried out on ***signs*** rather than numerical values of the observations. In non-parametric statistical techniques, the number of signs or counts or frequencies of sampled observations is the subject of analysis. Thus, the sampling distribution of the number of positive or negative signs that are obtained in a random trial is studied and decision on the null hypothesis reached upon.

2.1 One-Sample Sign Test

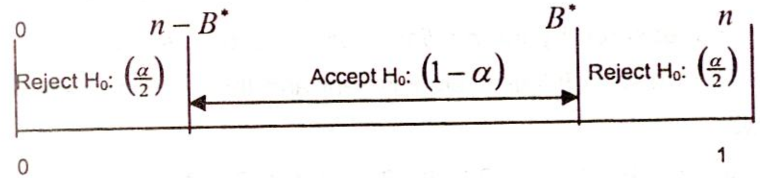
The purpose of one-sample sign test is to determine the extent to which a given random sample comes from a population with specified characteristics or attributes. Thus, the null hypothesis (H_0) may be stated as: *The sample comes from a population with specified characteristics or attributes*. The concern is whether or not a sample observation possesses characteristics or attributes of a given population. Hence, realization of each random observation in a sample is a *binomial process* in which an observation possesses the specified population attributes or it does not. Consequently, the appropriate measurement of an observation in a binomial process is labelling it as $+ve$, if it has the population attributes or labelling it as $-ve$, if it does not have the attribute.

If the null hypothesis (H_0) is true, the expected number of positive signs $(+ve)$ will be statistically equal to the number of negative signs $(-ve)$ in a random sample. Put differently, given a null hypothesis that sample observations come from a population with specified characteristics or attributes, the probability of obtaining either a $+ve\ (or\ -ve)$ signed-observation in any random sample is $p = 50\%\ or\ p = \frac{1}{2}$. Since the process is binomial, the number of trials n, in this binomial exercise is the sum of both positive and negative signs, which is $n = B^+ + B^-$. The critical value is determined by the maximum between number of +ve or –ve signs obtained in any single trial or more formally $B^* = \max(B^+, B^-)$. Given that B^+ = number of $+ve$ signs and B^- =number of $-ve$ signs, then the decision to accept or reject H_0 will depend on B^+, being statistically larger than B^- in the total of n valid trials.

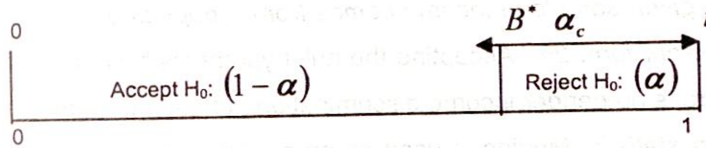
In statistical inference language, the decision to accept H_0 is contingent on the probability of getting $B^+\ (or\ B^-)$ being less or greater than a specified level of significance. Thus, the condition for rejecting the null hypothesis is: **reject H_0 iff the number of +ve (-ve) signs, is statistically larger than the number of –ve (+ve) signs.** Given a level of significance of α, the critical value of B for accepting rejecting H_0 is $B^* = \max(B^+, B^-)$. When translated into probability terms, the above condition implies: $\alpha_c = \Pr ob(B^* \le B \le n - B^* : B \approx binomial(n, p))$. Thus, for a two-tailed test, the null hypothesis is accepted iff:

$$\alpha_c = \Pr ob(B^* \le B \le n - B^* : B \approx binomial(n, p)) > \alpha .$$

The accept-reject area for a two-tailed test is indicated in the accompanying diagram.



A one-tailed test may also be considered, in which a null hypothesis is accepted if: $\alpha_c = \Pr ob(B \ge B^* : B \approx binomial(n, p)) > \alpha$. In the present case, $n = B^+ + B^-$ and $p = \frac{1}{2}$; and thus, $\alpha_c = \sum_{B=B^*}^{n} \binom{n}{B} p^B (1-p)^{n-B}$. The diagram below describes the *accepting-rejecting* H_0 situation for a one-tailed test.



The binomial test is in practice approximated by the standard normal test-statistic as $Z_c = \dfrac{B^* - np}{\sqrt{np(1-p)}}$ if the sample size is large enough. In such situation, the null hypothesis is accepted iff $-Z_{\alpha/2} < Z_c < Z_{\alpha/2}$ for a **two-tailed test** or $Z_c < Z_\alpha$ for a **one-tailed test.**

Example 2.1.1

The average income $(INCOME)$ of full-time women employees in a given industry as measured by the *median* is 25. A random sample of 13 men is chosen from those in full-time employment and their incomes are recorded as follows:

$$30, 28, 36, 34, 27, 24, 29, 33, 31, 98, 26, 23, and\ 32.$$

Do these data provide sufficient evidence to indicate gender income discrimination?

Solution 2.1.1

Following statistical test framework, we have:

(1) The null hypothesis H_0: *the sample of income comes from a population with median of 25*, and the alternative hypothesis is Ha: *the given sample of income comes from a population with a median different form 25.* Accepting the null hypothesis has implication that there is no gender income discrimination in full-time employment.

(2) *Test statistic*: Median is used as an average measure of income of the full-time women-employee population, and testing this parameter requires knowledge of Order Statistics[4]. Hence, the null hypothesis concerning population median cannot be tested using the conventional parametric test techniques. Given the above discussion, nonparametric test that is based on a single sample is considered appropriate for testing the above null hypothesis. Specifically, binomial or location test tests the above null hypothesis, and hence the test statistic is the **binomial variable** $\boldsymbol{B}$, which is the total number of $+ve\ and - ve$ signs.

[4] The subject of Order Statistics is beyond the scope of this book.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: The critical value of the test statistic is $B^* = \max(B^+, B^-)$. The decision rule, for a two-tailed test, is to **accept H_0 if and only if** $\Pr ob(B^* \le B \le n - B^*) > \alpha$ **and reject H_0 otherwise.**

(5) *Computations*: Since the studied population is women on fulltime employment, then sample of incomes of men on fulltime employment are signed or located in the women population – the specified population. If average incomes of men and women are the same $(H_0 : is\ true)$, then the proportion of men having incomes above (below) the women median income is 50%. Thus, an income of male employee is assigned as **plus** (+) if it is greater than the median income of women population of 25 and **minus** (–) if it is equal to or less.

Thus, $income = \begin{cases} +if\ greater\ than\ 25 \\ -if\ equal\ to\ or\ less\ than\ 25 \end{cases}$

Income	30	28	36	34	27	24	29
Sign	+	+	+	+	+	-	+
Income		33	31	98	26	23	32
Sign		+	+	+	+	-	+

Basing on the above table, the numbers of positive and negative signs are summarized as: $B^+ = 11,\ B^- = 2\ and\ hence\ n = 13$. In the language of binomial probability distribution, there are $n = 13$ independent trials; and each trial may result in either + or –. Besides, since expected numbers of positive and negative signs are the same under the null hypothesis (H_0), the proportion of obtaining a positive (negative) number of signs is p = ½.

Consequently,

$$\alpha_c = \Pr ob\left(B^* \le B \le n - B^* : B \approx binomial\left(n = 13,\ p = \tfrac{1}{2}\right)\right) \text{ or}$$

$$\alpha_c = \sum_{B=0}^{n-B^*} \binom{13}{B} \left(\tfrac{1}{2}\right)^B \left(\tfrac{1}{2}\right)^{13-B} + \sum_{B=B^*}^{n} \binom{13}{B} \left(\tfrac{1}{2}\right)^B \left(1 - \tfrac{1}{2}\right)^{13-B}$$

$$\alpha_c = \left[\binom{13}{0} + \binom{13}{1} + \binom{13}{2}\right]\left(\tfrac{1}{2}\right)^{13} + \left[\binom{13}{11} + \binom{13}{12} + \binom{13}{13}\right]\left(\tfrac{1}{2}\right)^{13}$$

$$\alpha_c = 0.011 + 0.011 = 0.022$$

(6) *Conclusion*: Since $\alpha_c = 0.022$ is smaller than $\alpha = 0.05$ then H_0 is rejected, implying that there is gender income discrimination among fulltime employees in the industry understudy[5].

[5] The computations for a binomial test approximation (one tailed test) implies are respectively equal to: $Z_c = \frac{B^* - np}{\sqrt{np(1-p)}} = \frac{11 - 13 \times \frac{1}{2}}{\sqrt{13 \times \frac{1}{2} \times \frac{1}{2}}} = 2.496$ *and* $z_\alpha = 1.96$ and this implies that Ho is rejected.

2.2 Two-paired Sample Sign Test

Two-paired sample sign test is applicable in a research environment that is characterized by repeated measurements taken from same/similar subjects or different treatments being managed to the same/similar subjects at a point in time. Included here are aspects of before-after and with-without research designs. The conventional focus of the hypothesis testing in such research designs is on determining the extent to which a study variable displays the difference under the before and after or with and without treatment conditions. Ideally, this is about testing for a statistical difference in the response-variable phenomenon being studied. In particular, the general form of the null hypothesis in such research design is H_0: *The two samples come from identical populations*; so that there is no difference in the response-variable phenomenon before-after or with-without treatment conditions.

Since measurement of the observation in the population understudy focuses on *signs* rather than *numerical values*, the appropriate statistical testing is nonparametric based on two correlated paired-samples. Hence, the differences of the paired observations are signed rather than being numerically recorded. Consequently, binomial test is an appropriate statistical technique for testing the null hypothesis on the significance of differences between the populations understudy.

Example 2.2.1

Two kinds of feed (***FEEDA***; ***FEEDB***) are fed to samples of homogenous pigs and the following are weights gained (Δkg) after a fixed period of time. Test the claim that the true average gain in weight of pigs is the same for both feeds.

FEED A	14	16	12	15	17	18	11
FEED B	14	15	13	11	16	19	10
FEED A		13	15	19	26	22	10
FEED B		18	8	12	25	20	9

Solution 2.2.1

Again following the statistical testing framework we have:

(1) The null hypothesis is H_0: *the two samples come from identical populations*, so that the two feeds have equal nutritional ingredients with respect to gaining weight. The alternative hypothesis is H_a: *the two samples come from different populations*, which imply that the two types of animal-feeds have different nutritional ingredients.

(2) *Test statistic*: Since the null hypothesis does not claim about knowledge of the true value of the difference in the gain of weight, the change in weight cannot be meaningfully subtracted to get a difference in gain as data on initial weight are missing[6]. Consequently, it is appropriate to use nonparametric technique, in which differences in weight gained are signed as positive or negative. Given that $B^{+} = number\ of\ positive\ signs$ and $B^{-} = number\ of\ negative\ signs$ then, the sum $\boldsymbol{B} = \boldsymbol{B}^{+} + \boldsymbol{B}^{-}$ is a **binomial variable**, which tests the null hypothesis.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: The critical value of the binomial test statistic *B* is $\boldsymbol{B}^{*} = \max\left(\boldsymbol{B}^{+}, \boldsymbol{B}^{-}\right)$. The decision rule for a two-tailed test is to accept H_0 iff $\Pr ob\left(B^{*} \leq B \leq n - B^{*}\right) > \alpha$ and reject H_0 otherwise.

[6] Note that for a meaningful difference of weight gained, initial weight must be the same for all subjects considered.

(5) *Computations*: If the null hypothesis is true, the expected gain of weight that is associated with the two types of feed is equal. In that respect the difference in weight gained between the two types of feeds would be zero. Testing the null hypothesis is accomplished by signing the differences in weight gained using the following rule:

$$dweightAB = \begin{cases} + \ if\ \ gain\ from\ A\ exeeds\ that\ \ from\ B \\ - \ if\ gain\ from\ A\ is\ equal\ to\ or\ less\ than\ from\ B \end{cases}$$

The above assignment rule is summarized in the table below:

FEED A	14	16	12	15	17	18	11
FEED B	14	15	13	11	16	19	10
Sign	–	+	–	+	+	–	+
FEED A		18	15	19	26	22	10
FEED B		13	8	12	25	20	9
Sign		+	+	+	+	+	+

In the above assignments, it is noted that $B^{+} = 10,\ B^{-} = 3$ and $n = B^{+} + B^{-} = 13$. If H_0 is true, $\boldsymbol{B}^{+}$ and $\boldsymbol{B}^{-}$ should be about the same. This implies that the probability of obtaining positive or negative signed difference is $\boldsymbol{p} = \frac{1}{2}$. However, if H_0 is not true then either $\boldsymbol{B}^{+}$ is larger (or smaller) than $\boldsymbol{B}^{-}$. Thus, the critical value of $\boldsymbol{B}$ for rejecting H_0 is $B^{*} = \max\left(B^{+},\ B^{-}\right) = \max(10,\ 3) = 10$.

Consequently, given that $B^{*} = 10,\ and\ \ n = 13$ then the computed probability of alpha, which is level of significance, is equal to:

$\alpha_c = \Pr ob\left(B^{*} \leq B \leq n - B^{*} : B \approx binomial\left(n = 13,\ p = \frac{1}{2}\right)\right)$ *or*

$$\alpha_c = \sum_{B=0}^{n-B^*}\binom{13}{B}\left(\tfrac{1}{2}\right)^B\left(\tfrac{1}{2}\right)^{13-B} + \sum_{B=B^*}^{n}\binom{13}{B}\left(\tfrac{1}{2}\right)^B\left(1-\tfrac{1}{2}\right)^{13-B}$$

$$\alpha_c = \left[\binom{13}{0}+\binom{13}{1}+\binom{13}{2}+\binom{13}{3}\right]\left(\tfrac{1}{2}\right)^{13}$$

$$+\left[\binom{13}{10}+\binom{13}{11}+\binom{13}{12}+\binom{13}{13}\right]\left(\tfrac{1}{2}\right)^{13}$$

$$\alpha_c = 0.0461 + 0.461 = 0.0922$$

(6) *Conclusion*: Since $\alpha_c = 0.0922$ is larger than $\alpha = 0.05$ then the H_0: is accepted in favour of the alternative hypothesis. This result indicates that on average, gain in weight is the same. The conclusion is that the two types of feeds have the same nutritional ingredients for the kind of animals considered. The same conclusion is attained when approximation to binomial test is used, that is

$$Z_c = \frac{B^* - np}{\sqrt{np(1-p)}} = \frac{10 - 13\times\frac{1}{2}}{\sqrt{13\times\frac{1}{2}\left(1-\frac{1}{2}\right)}} = 1.941 \; and \; z_\alpha = 1.96$$

implying that Ho is accepted.

Exercise 2.2.1

A study was conducted to determine the effect of TV advertisement on changed attitude of young people using a 0–20 attitudinal scale, with 20 being strongest. Do the data present sufficient evidence to indicate changed attitude?

Individual		1	2	3	4	5	6
Attitude	*Before*	14	16	15	18	15	17
	After	14	18	16	17	16	19
Individual		7	8	9	10	11	12
Attitude	*Before*	19	17	17	16	19	15
	After	20	18	19	15	18	16

Exercise 2.2.2

A market research of a soft drink producer tested 12 subjects' preference before and after a period of strenuous work. Preference was measured on a 0-10 scale, with 10 being the strongest. Do these data present sufficient evidence to indicate changed preference?

Subject		1	2	3	4	5	6
Preference	*Before*	4	8	2	4	5	3
	After	6	8	6	5	4	3
Subject		7	8	9	10	11	12
Preference	*Before*	6	3	7	6	2	4
	After	8	7	3	9	3	4

Exercise 2.2.3 (*optional*)

To what extent does use of parametric t-test change the conclusions in exercises 2.2.1 and 2.2.2 above?

2.3 *SPSS Tutorial on Sign or Binomial Test Technique*

The SPSS output for the problem in example 2.1.1 using of sign or binomial test is presented below.

	Category	N	Observed prop	Test prop	Exact sign (2-tailed)
INCOME Group1	<=25	2	.15	.50	0.022
Group 2	>25	11	.85		
Total		13	1.00		

To produce the above output, choose the following commands from the menu:

Analyze

Nonparametric tests

Binomial (test proportion is by default set at $0.50 = \frac{1}{2}$)

Test variable: ***INCOME*** (average income of full-time women employees)

Cut-off point = **25**

To produce the SPSS output for the problem in example 2.1.2, choose the following commands from the menu:

Analyze

Nonparametric tests

Binomial (test proportion is by default set at $0.50 = \frac{1}{2}$)

Test variable: ***DFEED***[7]

Cut-off point: **0**

The output for this problem is:

		Category	N	Observed prop	Test prop	Exact sign (2 tailed)
DFEED	Group 1	<= 0	3	.23	.50	0.092
	Group 2	> 0	10	.77		
	Total		13	1.00		

[7] This variable is computed from the data editor as the difference between change in weight from feed A and feed B so that $DFEED = FEEDA - FEEDB$

Problem Set Two

Question 2.1

A hot debate exists regarding reported Government Environmental Management Agency (GEMA) gasoline-mileage figures for new cars entering the country. This is because the government tests only simulate actual driving conditions while ignoring such factors as wind, road conditions, traffic intensity, and driver differences. A Private Testing Agency (PTA) compared the GEMA mileage rating of 12 new makes of automobiles with mileage recorded on a test run under actual driving conditions. The results are shown in the accompanying table (*P indicates that the mileage obtained under actual conditions was greater than the GEMA figure; G indicates that the GEMA mileage rating was greater; 0 indicates that the two ratings were equal*)

Car make	1	2	3	4	5	6	7	8	9	10	11	12
Test results	G	G	0	P	G	G	G	G	0	G	P	G

Do the findings of the PTA indicate that the GEMA ratings are, on average, higher than what you can expect to obtain for gasoline mileage while driving under normal conditions?

Question 2.2

Business analysts and economists are yet to agree on the effect, in terms of magnitude and direction, of economic reforms programme undertaken by Tanzanian Government. To explore this issue, a researcher has collected data on average industrial production of key sectors before and after the introduction of reform programme.

Production sector (GDP at factor cost)	Average production index (%)	
	Before reform	After reform
Manufacturing	11.6	48.0
Construction	9.4	8.7
Mining	4.9	13.9
Agriculture	55.1	3.4
Trade, restaurants, and hotels	9.7	14.5
Transport and communications	7.3	6.1
Others	2.0	5.4

Do these data provide sufficient evidence to indicate a difference in industrial production between the two periods?

Question 2.3

Organizational Management can fight cardiovascular disease by adopting fitness programmes for employees, and thereby reduce the social and economic costs of the disease. To improve employee' physical and mental fitness, many firms are now offering work release for voluntary exercise programme. To measure the effect of such a programme, a manufacturing firm recorded the time spent per week in vigorous exercise undertaken by 85 of its employees two years after the initiation of the volunteer exercise programme. These data were then compared with the level of participation by the firm's employees in various exercise activities prior to the firm-sponsored exercise programme.

Hours in Vigorous exercise	Number Before Programme	Number after Programme
No exercise	52	33
Some but less than 3 hours	29	23
3 hours to 6 hours	11	14
6 hours to 9 hours	6	9
Over 9 hours	2	4

Do the data provide sufficient evidence to indicate that the level of participation in vigorous exercise activities has changed since the initiation of the firm-sponsored exercise programme?

Question 2.4

Is there a change in the share to GDP (factor cost) among the sectors of Tanzania's economy before and after economic reform programme? To explore this issue, a researcher has collected data on the average share of each sector to the GDP before and after reforms. The data are summarized in the accompanying table. Do these data provide sufficient evidence to indicate a change in the sectoral share to GDP?

Sector	**% Share GDP** (factor cost current prices)	
	Before reform	*After reform*
Agriculture	52.40	27.20
Mining	0.50	2.50
Manufacturing	7.95	8.30
Electricity and water	0.74	1.70
Construction	3.03	3.80
Trade	14.00	16.50
Communication and transport	6.42	5.40
Public administration	5.90	10.30

3

INFERENCE TECHNIQUES FOR ORDINAL DATA: THE MEDIAN TESTS

Median is one of the measures of averages in statistics that play a central role in inference methods. By definition a median is an observation half-way when data set is arranged in order of magnitude. For this to be a possible possibility, the sample observations must be capable of being ranked and thus measured at ordinal level. There are two non-parametric techniques for analysing ordinal data, namely, the sample median tests and the rank-sum tests. This chapter focuses on the sample median tests and Kolmogorov-Smirnov test. Both test techniques are based on frequency distributions, with the latter focusing on goodness-of-fit test. Rank-sum tests that are based on ranked data are covered in the next chapter.

3.1 *Two–unpaired Sample Median-Test*

This test is applicable when two samples are neither paired nor correlated, and therefore $n_1 \neq n_2$. The null hypothesis being tested is H_0: *the two samples come from identical population with equal median* and the alternative hypothesis is H_a: *the two samples come from different populations*. To test the null hypothesis, we rely on the number of counts or frequencies of each sample observations that are below and above a pooled median. The pooled median is considered as an unbiased estimate of the hypothesized common median.

The distribution of the number of counts or frequencies is summarized in a $2 \, by \, 2$ contingency table as follows.

Sample	Below-median	Above-median	Totals (R_i)
Sample 1	o_{11}	o_{12}	$R_1 = \sum o_{1j}$
Sample 2	o_{21}	o_{22}	$R_2 = \sum o_{2j}$
Totals (C_j)	$C_1 = \sum o_{i1}$	$C_2 = \sum o_{i2}$	$N = \sum \sum o_{ij}$

If the null hypothesis is true, the expected number of counts or frequencies will not differ significantly from the observed number of counts or frequencies in each cell of the contingency table. A test statistic based on Chi-square detects any significant difference between observed and expected number of counts or frequencies in the contingency table. Given that o_{ij} = observed frequencies, e_{ij} = expected frequencies such that $e_{ij} = \frac{R_i C_j}{N}$ then the test statistic for two-unpaired sample median test is:

$$\chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

The following computational procedural steps accomplish the median test described above[8].

- Rank observations from the two samples in ascending order of magnitude while retaining the sample identity.
- Find median of a combined/pooled sample.
- Set up a $2 \, by \, 2$ contingency table with rows representing the two samples and columns representing the below-median and above-median attribute.

[8] Note that these steps are associated with statistical hypothesis-testing framework.

- Compute row total (R_i); column total (C_j); grand total (N); and the expected number of counts or frequencies in each cell of the contingency table using the formulae: $e_{ij} = \frac{R_i C_j}{N}$

- Compute the Chi-square: $\chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$ and compare it with the theoretical $\chi^2_{\alpha,\nu}$ at an appropriate degree of freedom, $\nu = (r-1)(c-1)$. Table 3 appended is used to obtain theoretical values of Chi-square at appropriate degree of freedom.

The null hypothesis will be rejected if computed Chi-square value is larger than the theoretical value, that is $\chi_c^2 > \chi^2_{\alpha,\nu}$. This would imply that observed frequencies are significantly different from expected. The conclusion that is being inferred here is that the two samples come from different populations.

Example 3.1.1

Best practices in pricing strategy point out two models that are widely used by business entrepreneurs to enter a new market. These are low pricing strategy and high pricing strategy. A business analyst has examined the effect of pricing strategies on long-term performance, by comparing average annual sales, of enterprises that are known to have practiced the two common pricing strategies. Data on the average annual sales (in million TAS) are summarized in the table below.

Low-pricing	20	25	15	9	41	60	22	18
	65	90	180	280				
High-pricing	10	55	42	60	70	120	220	340

Do the data provide sufficient evidence ($\alpha = 0.05$) that high pricing strategy outperforms low-pricing strategy with respect to average annual sales?

Solution 3.1.1

(1) The null hypothesis is H_0: *the two pricing strategies have identical effect on average annual sales*. The alternative hypothesis is H_a: *the two pricing strategies have different effect on average annual sales populations*.

(2) *Test statistic*: Since the null hypothesis does not claim about knowledge of the true-value of the difference in the average annual sales realized by firms, a 2×2 contingency table is set up, with rows representing the two samples and columns representing the below-median and above-median attribute.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: If the null hypothesis is true the expected frequencies of annual sales that are associated with the two pricing strategies types of feed are equal to the observed frequencies. In that respect the difference in annual sales between the two types of pricing strategies would be zero. Thus, given that o_{ij} *and* e_{ij} are observed and expected frequencies respectively, then **the null hypothesis** is accepted when computed Chi-square value is less than the theoretical values. Formally this means that the H_0: is accepted iff: $\chi_c^2 < \chi^2_{\alpha,\nu} : \chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$. And since $\alpha = 0.05$ and $\nu = (k-1) = 1$, then using Chi-square table appended $\chi^2_{\alpha,\nu} = \chi^2_{c0.05,1} = 3.841$ (Table 3).

(5) *Computations*: Testing the null hypothesis is accomplished by preparing a combined signed-ranking, finding a common median, and setting up a 2 by 2 contingency table as follows.

Sign-ranked sales	9L	10H	15L	18L	20L	22L	25L
Ranks	1	2	3	4	5	6	7
Sign-ranked sales	41L	42H	55H	60L	60H	65L	70H
Ranks	8	9	10	11.5	11.5	13	14
Sign-ranked sales		90L	120H	180L	220H	280L	340H
Ranks		15	16	17	18	19	20

Given the above signed-ranking, the combined median-sales is $Median = 57.5$, and the contingency table is presented in the table below (expected frequencies in brackets): The computed value of the test statistic is $\chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 0.833$.

Price Strategy	Below-median	Above-median	Totals (R_i)
Low-priced	7 (6)	5 (6)	12
High-priced	3 (4)	5 (4)	8
Totals (C_j)	10	10	20

(6) *Conclusion*: *Conclusion*: Given the decision rule, which is $\chi_c^2 < \chi_{\alpha,v}^2$, and the sample results as $\chi_c^2 = 0.833$, the null hypothesis H$_0$ is accepted. The two pricing strategies have the same effect on average annual sales for firms reported in the study. This result leads to the finding that other factors being equal, pricing strategy is independent of firm performance, measured by average annual sales.

Exercise 3.1.1

Is there a difference in production levels among the major agricultural crops in Tanzania over time? To explore this issue, a researcher collected data on production levels of major crops for eight years. The data are summarized in the accompanying table. Use k-sample median test to ascertain the proposition of no difference in production using monetary value and percentage.

Years		1992	1993	1994	1995
Cereal	TAS 000	3687	3313	4362	4464
	Percentage	100	90	118	121
Non-cereal	TAS 000	6861	6403	7334	7498
	Percentage	100	93	107	109
Trad. exports	TAS 000	229	222	220	303
	Percentage	100	97	96	120
Years		**1996**	**1997**	**1998**	**1999**
Cereal	TAS 000	3111	4371	3996	3367
	Percentage	84	116	103	91
Non-cereal	TAS 000	5886	7971	7441	7323
	Percentage	86	116	109	107
Trad. exports	TAS 000	272	299	273	282
	Percentage	131	119	123	129

Do the data provide sufficient evidence to indicate a changed policy regarding production imperatives in the agricultural sector? What reservations do you consider critical in interpreting the results you have obtained above?

3.2 *The K-Sample Median Test*

In its structure the k-sample median test is an extension of the two-unpaired sample median test, only that there are more than two unpaired samples. The null hypothesis being tested is H_o: *the k-samples come from k identical populations*, and the alternative is Ha: *at least one sample comes from a different population*. To test the above null hypothesis we rely on the number of counts or frequencies of observations from each sample that are below or above the pooled sample median.

The distribution of the number of counts or frequencies in each sample that are below or above median is presented in the $k \times 2$ contingency table below.

Sample	Below-median	Above-median	Totals (R_i)
Sample 1	o_{11}	o_{12}	$R_1 = \sum o_{1j}$
Sample 2	o_{21}	o_{22}	$R_2 = \sum o_{2j}$
• •			• •
Totals (C_j)	$C_1 = \sum o_{i1}$	$C_2 = \sum o_{i2}$	$N = \sum\sum o_{ij}$

If H_0 is true the expected number of counts or frequencies (e_{ij}) in each cell of the contingency table will not differ significantly from the observed number of outcomes or frequencies (o_{ij}). Any significant difference will be detected by a test-statistic based on Chi-square given by $\chi^2 = \sum_i^k \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$; the appropriate degree of freedom of the Chi-square is $\nu = (k-1)$. The computational procedures are as indicated above for a two-unpaired sample median test.

Example 3.2.1

Data in the table below pertains to performance, measured by gross sales in TAS, of four advertising channels which were adopted by selected whole-sale businesses in Tanzania.

Channel	Performance (in mil. TAS)						
A	57	50	36	28	90	64	70
B	100	89	80	53	40	25	
C	75	45	85	60	90	30	20
D	105	35	65	150	285		

Do these data present sufficient evidence that advertising channels differ in promoting performance as measured by sales in this particular instance?

Solution 3.2.1

To solve the above problem, k-sample median test is considered as an appropriate non-parametric technique.

(1) The null hypothesis to be tested is: *Ho: The four advertising channels have same effect on business performance*; the alternative *Ha: Some advertising channels have superior performance*.

(2) K-sample median test based on Chi-square with degree of freedom that is equal to $\nu = (k-1)$.

(3) Level of significance is $\alpha = 0.05$.

(4) Decision rule: **Accept the null hypothesis** (Ho: equal effect for the four channels) **iff**; $\chi_c^2 < \chi_{\alpha,\nu}^2 : \chi_c^2 = \sum_i^k \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$. Note that o_{ij} *and* e_{ij} are observed and expected frequencies respectively. Given that $\alpha = 0.05$ and $\nu = (k-1) = 4-1 = 3$, then as per Table 3, $\chi_{\alpha,\nu}^2 = \chi_{c0.05,3}^2 = 7.81$.

(5) *Computations*: The signed ranking of combined sample is summarized below :

Signed-scores	20c	25b	28a	30c	35d	36a	40b
Rank	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Signed-scores	**45c**	**50a**	**53b**	**57a**	**60c**	**64a**	**65d**
Rank	*8*	*9*	*10*	*11*	*12*	*13*	*14*
Signed-scores	**70a**	**75c**	**80b**	**85c**	**89b**	**90a**	**90c**
Rank	*15*	*16*	*17*	*18*	*19*	*20.5*	*20.5*
Signed-scores	**100b**	**105d**	**150d**	**285d**			
Rank	*22*	*23*	*24*	*25*			

The combined sample Median is 64; and based on this result, a contingency table of observed frequencies is prepared as follows:

Advertising channels	Below Median	Above Median	Total
A	4 (3.36)	3 (3.64)	7
B	3 (2.88)	3 (3.12)	6
C	4 (3.36)	3 (3.64)	7
D	1 (2.40)	4 (2.60)	5
Total	12	13	25

The computed Chi-square is $\chi_c^2 = \sum_i^k \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 2.049$,

which is less than the theoretical value $\chi^2_{0.05,\ 3} = 7.81$

(6) *Conclusion*: Given the decision rule, which is $\chi_c^2 < \chi^2_{\alpha,\ v}$, and the sample results as $\chi_c^2 = 2.049$, the null hypothesis H_0 is accepted. The four advertising channels have the same effect on business sales performance. The choice of a channel should

depend on other factors such as costs, etc, rather than business performance.

Exercises 3.2.1

Is there a change in the share to GDP (factor cost) among the sectors of Tanzania's economy before and after economic reform programme? To explore this issue, a researcher has collected data on the average share of each sector to the GDP before and after reforms. The data are summarized in the accompanying table. Use median test to ascertain a claim that there is a change in the sectoral share to GDP? What is the policy implication of the results you have obtained?

Sector	**% Share GDP** (factor cost current prices)	
	Before reform	*After reform*
Agriculture	52.40	27.20
Mining	0.50	2.50
Manufacturing	7.95	8.30
Electricity and water	0.74	1.70
Construction	3.03	3.80
Trade	14.00	16.50
Communication and transport	6.42	5.40
Public administration	5.90	10.30

3.3 Hyper-geometric Probability Model as Alternative to the Median Test

Let $x_1, x_2, \cdots, x_m$ be a random sample of size m from cumulative density function $F_x(\bullet)$ with corresponding density function $f(x)$. Let also $y_1, y_2, \cdots, y_n$ be another random sample of size n from cumulative density function $F_y(\bullet)$ with density function given by $f(y)$. Assume that observations from $F_x(\bullet)$ are independent of the observations from $F_y(\bullet)$. Using median test, it is required to test the extent to which the two populations are the same, or in a more formal way, it is desired to test a null hypothesis $H_0 : F_x(Z) = F_y(Z)\ \forall Z$ against the alternative hypothesis $H_a : F_x(Z) \neq F_y(Z)\ for\ at\ least\ one\ Z$.

Let Z represent observations of a combined random sample so that the ranked set of observations from the two random samples are such that: $Z_1 \leq Z_2 \leq \cdots \leq Z_{m+n}$. The median of the combined sample is $\tilde{Z}$, which is equal to the $\frac{m+n}{2}th$ observation. Let M_x be the number of $X-values$ above $\tilde{Z}$ and N_y be the number of $Y-values$ above $\tilde{Z}$. Given that random sampling procedure is used to select observations from the two populations, $M_x\ (or\ N_y)$ is a random variable.

In a sample of size $\frac{m+n}{2}$, the probability distribution of $M_x\ (or\ N_y)$ is hyper-geometric density function described by:

$$P(M_x = m_x) = g(m_x) = \frac{\binom{m}{m_x}\binom{n}{\frac{m+n}{2} - m_x}}{\binom{m+n}{\frac{m+n}{2}}}.$$

Thus, if H_0: is true, there will be m_x X-values (or n_y Y-values) in a random sample of size $\frac{m+n}{2}$, and $m_x\ (or\ n_y)$ will be approximately $\frac{m}{2}\ (or \frac{n}{2})$. Thus, given a level of significance of α, a critical value k of $M_x\ (or\ N_y)$ can be determined such that $P(|M_x - \frac{m}{2}| \geq k) = \alpha$. Having determined the critical value then H_0: is accepted if $|M_x - \frac{m}{2}| \geq k$ and rejected otherwise[9].

The hyper-geometric probability test for median can be extended to three samples from cumulative density functions $F_x(\bullet)$, $F_y(\bullet)$ and $F_r(\bullet)$. Let $m, n, and\ p$ be sample sizes respectively and $M_x, N_y, and\ P_r$ represent numbers of observations from the three respective samples that are above the combined median. Considering that the combined sample is random, $M_x\ (N_y\ or\ P_r)$ has a hyper-geometric probability density function of a multinomial form:

[9] To allow for finite computations, a restriction is imposed that suppresses $\frac{m}{2}$, $\frac{n}{2}$, and $\frac{m+n}{2}$ to an integer, that is, $m_x = \text{int}(\frac{m}{2})$, $n_y = \text{int}(\frac{n}{2})$ and $\text{int}(\frac{m+n}{2})$

$$P(M_x = m_x) = g(m_x) = \frac{\binom{m}{m_x}\binom{n}{n_y}\binom{p}{p_r}}{\binom{m+n+p}{m_x+n_y+p_r}}$$

If H_0 : is true, m_x $(n_y \text{ or } p_r)$ will be approximately $\frac{m}{2}$ $(\frac{n}{2} \text{ or } \frac{p}{2})$ and thus given a level of significance of α, a critical value k of M_x $(N_y \text{ or } P_r)$ can be determined such that $P(|M_x - \frac{m}{3}| \geq k) = \alpha$. Having determined the critical value then H_0 : is accepted if $|M_x - \frac{m}{3}| \geq k$ and rejected otherwise.

Example 3.3.1

The following data pertains to average scores obtained by students who were exposed to two different pilot teaching methods:

Method A	45	89	36	28	57	90		
Method B	67	70	54	25	18	65	36	95

Use hyper-geometric probability as alternative to sample median test to verify the proposition that the two teaching methods have the same effect on scores obtained by students $(\alpha = 0.05)$.

Solution 3.3.1

The null hypothesis to be tested is, Ho: *the two teaching methods have same effect on scores obtained by students.* In this example $m = 6,\ and\ n = 8$ and therefore the median is the $\frac{m+n}{2}th\ or\ 7th$ score when the data set is arranged in order of magnitude. If the hypothesis is true, the number of scores of students under methods A and B that are below

Median must be equal to 3 and 4 respectively in a sample of size 7. The probability of selecting m_x scores from Method A is given by:

$$P(M_x = m_x) = g(m_x) = \frac{\binom{6}{m_x}\binom{8}{7-m_x}}{\binom{14}{7}}$$

The following table summarizes the corresponding probabilities for all possible values of m_x.

m_x	n_y	$\binom{6}{m_x}\binom{8}{7-m_x}$	$P(M_x = m_x)$
0	**7**	**1 x 8 = 8**	**0.002331**
1	**6**	**6 x 28 = 168**	**0.048951**
2	5	15 x 56 = 840	0.244755
3	4	20 x 70 = 1,400	0.407925
4	3	15 x 56 = 840	0.244755
5	**2**	**6 x 28 = 168**	**0.048951**
6	**1**	**1 x 8 = 8**	**0.002331**
Total		3,432	1.000000

Based on the computations above, **the number of m_x scores, in a sample of size 7, must at least be 2 and at most 4 for Ho: to be accepted**.

The ranked scores are:

Signed-scores	**18a**	**25b**	**28a**	**36a**	**36b**	**45a**	**54b**
Rank	**1a**	**2b**	**3a**	**4a**	**5b**	**6a**	**7b**
Signed-scores	57a	65b	67b	70b	89a	90a	95b
Rank	8a	9b	10b	11b	12a	13a	14b

In the present case, the number of scores from method A that are below median in the sample of size7 is four (4); hence the null hypothesis is accepted.

Example 3.3.2

The accompanying table gives data on performance of SME's in terms of sales volume under two advertising channels. Use hyper-geometric probability model as alternative to sample median test to verify the proposition that the two advertising channels have the same effect on SME performance $(\alpha = 0.05)$.

Channel	Performance (in mil. TAS)											
A	57	50	36	28	90	64	70	45	80	53	40	25
B	100	105	75	89	85	60	90	30	-	-	-	-

Solution 3.3.2

The null hypothesis to be tested is, *Ho: SME'S do not differ in their pricing strategies*. In the above example $m = 12, \text{ and } n = 8$ and therefore the median is $\frac{m+n}{2} th \text{ or } 10th$ gross-sales when the combined data set is arranged in order of magnitude. Thus, if a sample of size 10 is randomly obtained there is expected to be $\frac{m}{2} = 6 \text{ and } \frac{n}{2} = 4$ observations in the sample representing each pricing category. The probability of selecting m_x gross-sales-items from the first pricing category is given by:

$$P(M_x = m_x) = g(m_x) = \frac{\binom{12}{m_x}\binom{8}{10-m_x}}{\binom{20}{10}}$$

The following table summarizes the corresponding probabilities for all possible values of m_x.

m_x	n_y	$\binom{12}{m_x}\binom{8}{10-m_x}$	$P(M_x = m_x)$
2	**8**	**66 x 1 = 66**	**0.0003572**
3	**7**	**220 x 8 = 1760**	**0.0095260**
4	6	495 x 28 = 13860	0.0750178
5	5	792 x 56 = 44352	0.2400571
6	4	924 x 70 = 64680	0.3500833
7	3	792 x 56 = 44352	0.2400571
8	2	495 x 28 = 13860	0.0750178
9	**1**	**220 x 8 = 1760**	**0.0095260**
10	**0**	**66 x 1 = 66**	**0.0003472**
Total		184,756	1.0000000

Based on the computations above, **the number of m_x gross-sales-items in a sample of size 10 must at least be 3 and at most 7 for Ho: to be accepted**.

The signed ranks of the sales volumes are:

Signed sales	**25a**	**28a**	**30b**	**36a**	**40a**	**45a**	**50a**	**53a**	**57a**	**60b**
Signed-ranks	**1a**	**2a**	**3b**	**4a**	**5a**	**6a**	**7a**	**8a**	**9a**	**10b**
Signed sales	64a	70a	75b	80a	85b	89b	90b	90a	100b	105b
Signed-ranks	11a	12a	13b	14a	15b	16b	17b	18a	19b	20b

The number of times that sales volume under Channel-A advertising strategy are below median in the sample of size 10 is eight (8); hence the null hypothesis is rejected.

Exercise 3.3.1 (*optional*)

What conclusion is reached if a parametric t-test is used to test the null hypothesis that the two advertising channels have the same effect on SME performance $(\alpha = 0.05)$.

3.4 Kolmogorov-Smirnov Goodness-of-fit Test

Researchers often confront a problem of ascertaining the extent to which observed cumulative frequency distribution fits well to a hypothesized theoretical cumulative frequency distribution. In such situations, the problem is about testing a hypothesis that there is no difference between observed and theoretical distribution. This problem lends itself to the testing of goodness-of-fit, which is about ascertaining the extent to which the difference between observed and expected frequencies is zero. It is therefore about testing variation of observed frequency with respect to expected frequencies. If the difference is small across all the possible ranges of the factor under consideration, there is a good fit.

Given that goodness-of-fit test is about testing variation of observed frequencies, Chi-square test is the appropriate test statistic. However, Kolmogorov-Smirnov test is considered more appropriate to answer the question of the form: *how close is observed cumulative frequency distribution to a theoretical cumulative distribution*? This is because Kolmogorov-Smirnov tests the extent to which there is a difference between observed and theoretical **cumulative frequencies**, whereas Chi-square tests the extent to which there is a difference between observed and theoretical **frequencies** for nominal data.

Given that F_o = observed **relative** cumulative frequency and F_e = expected **relative** cumulative frequency, then the statistic that is used by the K-S test is: $D_n = \max|F_e - F_o|$. The null hypothesis under the Kolmogorov-Smirnov test is H₀: *there is good fit between observed and theoretical cumulative frequency distributions*. Critical values of D_n for the Kolmogorov-Smirnov Goodness-of-fit Test are obtained from Table 5 appended. For instance, given a level of significance of $\alpha = 0.05$ (one-

tailed) and a sample size of $n = 20$, the critical value of $D_{20,0.05} = 0.294$. When sample size is larger than 35, the last row of Table 5 enables one to compute the critical value of D_n. For instance, if $\alpha = 0.05$ and $n = 64$, the critical value is $D_{64,0.05} = \frac{1.36}{\sqrt{64}} = 0.17$. The null hypothesis of goodness-of-fit is rejected if computed value, $D_{nc} = \max|F_e - F_o|$, is greater than the critical value of $D_{n,\alpha}$. Alternatively, a normal approximation of the K-S statistic, $Z_c = D\sqrt{n}$, may be used, and if it is used, the null hypothesis is rejected iff $|Z_c = D\sqrt{n}| > Z_{\alpha/2}$.

Example 3.4.1

The arrival rate of cars requiring repair and general service at a garage in a city is hypothesized to follow a Poisson probability distribution with a mean of $\lambda = 8$ cars per week. A study was conducted to determine the extent to which Poisson probability distribution best describes the arrival process, and the results are summarized below. Use Kolmogorov-Smirnov Goodness-of-fit test to check the claim on the Poisson arrival process $(\alpha = 0.05)$.

No Cars	Frequency
0 – 4	6
5 – 9	29
10 – 14	45
15 – 19	30
20 and above	10

Solution 3.4.1

The null hypothesis is *Ho: Observed arrival rate distribution of cars requiring maintenance and general repairs follows a Poisson distribution*. Critical value of the Kolmogorov-Smirnov Goodness-of-fit statistic is obtained from Table 5 appended as follows:

$$D_n = \begin{cases} D_{n,\,\alpha} & \text{for } n \le 35 \\ \dfrac{1.36}{\sqrt{n}} & \text{otherwise} \end{cases}$$

In the present case, given that $n = 120,\ and\ \alpha = 0.05$ then, $D_{120,\,0.05} = \dfrac{1.36}{\sqrt{120}} = 0.124$. The goodness-of-fit hypothesis is accepted if computed value of D_n is less than this critical value.

The computational procedures for conducting the Kolmogorov-Smirnov test, as summarized in the table below, involves the following steps:

- Computing probabilities, $P(X;\ \lambda = 8)$ of realizing the number of cars described by a random variable X under the null hypothesis. This is done by using the cumulative Poisson probability formula: $P(X;\lambda = 8) = \sum \dfrac{e^{-\lambda}\lambda^x}{x!}$ or reading these probabilities from appropriate cumulative Poisson probability table.
- Computing expected frequencies using the formulae $f_e = n\ P(X;\ \lambda = 8)$,
- Computing relative cumulative observed and expected frequencies and denote these as F_o and F_e respectively.
- Computing the K-S test statistic; $D_n = \max|F_e - F_o|$.

X	$P(X;\ \lambda = 8)$	f_o	f_e	F_o	F_e	$\lvert F_e - F_o \rvert$
0 – 4	0.0996	6	11.952	0.0500	0.0996	0.0496
5 – 9	0.6170	29	74.040	0.2917	0.7166	**0.4249**
10 – 14	0.2661	45	31.932	0.6667	0.9827	0.3160
15 – 19	0.0170	30	2.040	0.9167	0.9997	0.0830
20 above	0.0003	10	0.036	1.0000	1.0000	0.0000
Total	1.0000	120	120	-	-	-

From computations in the table above, $D_{nc} = \max|F_e - F_o| = 0.4249$, and this is larger than the critical value of $D_{120,\,0.05} = \dfrac{1.36}{\sqrt{120}} = 0.124$. This means that the null hypothesis of the goodness-of-fit is rejected, and thus it is concluded that the observed frequency distribution of arrival rate of cars at this garage is not well described by Poisson probability distribution with mean of 8 cars per week reported in this study. A similar conclusion is reached when a normal approximation of the K-S statistic is used. In this example the computed value is $Z_c = 0.4249\sqrt{120} = 4.65$.

Example 3.4.2

The relative frequencies of the number of vacant rooms at Zeta Motel recorded over the years enabled its manager to determine the probability distribution of the number of vacant rooms given in the accompanying table.

Number of vacant rooms	0	1	2	3	≥ 4
Probability	.10	.25	.35	.20	.10

However, since the manager recorded these data, a new motel has been built at a nearby location. In the first 120 days since completion of the new motel, the manager recorded the number of room vacancies per day in his Zeta Motel. These data are shown in the second table.

Number of vacant rooms	0	1	2	3	≥ 4
Number of days	15	30	50	18	7

Do these data, based on Kolmogorov-Smirnov goodness-of-fit test, present sufficient evidence, at $\alpha = 0.05$ level of significance, to indicate that the pattern of room vacancies in the Zeta Motel has changed since the opening of the new motel?

Solution 3.4.2

The null hypothesis being tested is that *Ho: Probability distribution of the vacant rooms at Zeta Motel has not changed on account of the new motel opening*. The computational procedures for conducting the Kolmogorov-Smirnov test, as summarized in the table below, involves the following steps:

- The probabilities, $P(X)$ of realizing the number of vacant rooms described by a random variable X under the null hypothesis are given in the first table.
- Computing expected frequencies using the formulae $f_e = n\ P(X)$.
- Computing relative cumulative observed and expected frequencies and denote these as F_o and F_e respectively.
- Computing the K-S test statistic; $D_n = \max|F_e - F_o|$.

X	$P(X)$	f_o	f_e	F_o	F_e	$\lvert F_e - F_o \rvert$
0	0.10	15	12	0.125	0.10	0.025
1	0.25	30	30	0.375	0.35	0.025
2	0.35	50	42	0.792	0.70	**0.092**
3	0.20	18	24	0.942	0.90	0.042
4 and above	0.10	7	12	1.000	1.00	0.000
Total	1.00	120	120	-	-	-

From computations in the table above, $D_{nc} = \max|F_e - F_o| = 0.092$, and this is smaller than the critical value of $D_{100,\ 0.05} = \frac{1.36}{\sqrt{100}} = 0.136$. This means that the null hypothesis of the goodness-of-fit is accepted, and thus it is concluded that the probability distribution of vacant rooms of Zeta Motel has not changed on completion of the new motel at the nearby location. This is not a surprise, as in a service industry, on average; each service place or facility has its own clientele. For instance, visitors tend to stick to service facility they have knowledge about or established relationship long before. Besides, the expected quality of service may be lower at the new motel compared to that currently offered at Zeta Motel. The same conclusion is reached using normal approximation of the K-S statistic, which in this particular case is: $Z_c = 0.092\sqrt{120} = 1.01$.

Exercise 3.4.1

In theory at least, age is considered the most important demographic factor that has strong correlation with fertility, divorce rates, and life-styles in a society. By predicting fertility and divorce rates as well as life-styles, economic demographers attempt to forecast an economy, predicting which industries will flourish and which ones will falter. Hence, knowledge of age-group distribution is of great help towards forecasting economic performance and/or structure. In attempting to describe the distribution of age, a researcher has prepared, based on population census statistics, a frequency distribution of age-groups for 10,000 randomly selected respondents in a district (see the accompanying table). Do these data, based on Kolmogorov-Smirnov goodness-of-fit test, provide sufficient evidence $(\alpha = 0.10)$ to indicate that age-group is a normal variable? (Hint: compute sample mean and variance of age and use these estimates as parameters of the normal distribution).

Age-group	Frequency
00 – 09	1210
10 – 19	1390
20 – 29	1300
30 – 39	1560
40 – 49	1680
50 – 59	1220
60 – 69	780
70 – 79	600
80 and above	320

3.5 *SPSS Tutorial on Median and Kolmogorov-Smirnov Tests*

The SPSS output for Median and Kolmogorov-Smirnov Goodness-of-fit Tests are presented below, starting with problem **3.1.1** followed by problem **3.4.1**.

3.5.1 The Median Test

Problem 3.1.1 is on the relationship between pricing strategy and performance as measured by sales. The objective is to ascertain the extent to which low-price and high-price strategies are different in terms of performance as measured by sales.

LOW-HIGH PRICING * BELOW-ABOVE MEDIAN Cross tabulation

Count

		BELOW-ABOVE MEDIAN		Total
		Below-	Above-	
LOW-HIGH PRICING	Low-priced	7	5	12
	High-priced	3	5	8
Total		10	10	20

Chi-Square Tests

	Value	df	Asy Sig (2-sided)	Exact Sig (2-sided)	Exact Sig (1-sided)
Pearson Chi-Square [a]	.833	1	.361		
Continuity Correction [b]	.208	1	.648		
Likelihood Ratio	.840	1	.359		
Fisher's Exact Test				.650	.325
Linear-by-Linear Association	.792	1	.374		
N of Valid Cases	20				

a Computed only for a 2x2 table

b 2 cells (50.0%) have expected count less than 5. The minimum expected count is 4.00.

To produce this output, first, enter sales data from the low and high priced strategies as one variable in the Data editor and compute the **combined median**. Second, create two nominal variables, ***LOW-HIGH PRICE*** (11=low-priced; 12=high-priced) and ***BELOW-ABOVE MEDIAN*** (21 = below median; 22 = above median). Then, choose the following commands from the menu:

Analyze

Descriptive

Crosstabs ...

Test variable List: ***LOW-HIGH PRICE*** (rows); ***BELOW-ABOVEMEDIAN*** (columns)

Statistics: **Chi-square**

To conduct a k-sample median test using SPSS, the row-categorical variable is extended to k-levels. For instance, in example 3.2.1, there are now four samples. A row-variable ***CHANNEL*** is created with four levels as follows; 11 = channel A; 12 = channel B; 13 = channel C, and 14 = channel D. The corresponding column-variable, ***BELOW-ABOVE MEDIAN*** retains two levels, below median and above median.

CHANNEL * BAMEDIAN Cross tabulation

Count

		BELOW-ABOVE MEDIAN		Total
		Below-median	Above-median	
CHANNEL	Channel A	4	3	7
	Channel B	3	3	6
	Channel C	4	3	7
	Channel D	1	4	5
Total		12	13	25

Chi-Square Tests

	Value	df	Asy Sig (2-sided)
Pearson Chi-Square [a]	2.049	3	.562
Likelihood Ratio	2.174	3	.537
Linear-by-Linear Association	1.005	1	.316
N of Valid Cases	25		

a 8 cells (100.0%) have expected count less than 5. The minimum expected count is 2.40.

3.5.2 Kolmogorov-Smirnov Goodness-of-fit Test

The outputs of Kolmogorov-Smirnov good-ness-of-fit test using SPSS, on ***CARSARRIVAL*** (arrival rate of cars); ***VACANTROOMS*** (number of vacant rooms) with test distributions being Normal, Uniform, Poisson, and Exponential, for examples **3.4.1** and **3.4.2**, are presented below.

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
CARARRIVAL	120	12.3750	5.06937	2.00	22.00
VACANTROOMS	120	1.7667	1.04305	.00	4.00

One-Sample Kolmogorov-Smirnov Test 1

		CARARRIV	VACTROOMS
N		120	120
Normal Parameters [a, b]	Mean	12.3750	1.7667
	Std D	5.06937	1.04305
Most Extreme Differences	**Absolute**	**.196**	**.214**
	Positive	.196	.203
	Negative	-.179	-.214
Kolmogorov-Smirnov Z		**2.149**	**2.339**
Asy Sig (2-tailed)		.000	.000

a Test distribution is Normal.
b Calculated from data.

One-Sample Kolmogorov-Smirnov Test 2

		CARARRIV	VACTROOMS
N		120	120
Uniform Parameters [a, b]	Minimum	2.00	.00
	Maximum	22.00	4.00
Most Extreme Differences	**Absolute**	**.208**	**.292**
	Positive	.167	.292
	Negative	-.208	-.125
Kolmogorov-Smirnov Z		**2.282**	**3.195**
Asymp. Sig. (2-tailed)		.000	.000

a Test distribution is Uniform.
b Calculated from data.

One-Sample Kolmogorov-Smirnov Test 3

		CARARRIV	VACTROOMS
N		120	120
Poisson Parameter [a, b]	Mean	12.3750	1.7667
Most Extreme Differences	**Absolute**	**.217**	**.098**
	Positive	.217	.052
	Negative	-.210	-.098
Kolmogorov-Smirnov Z		**2.381**	**1.072**
Asymp. Sig. (2-tailed)		.000	.201

a Test distribution is Poisson.
b Calculated from data.

One-Sample Kolmogorov-Smirnov Test 4

		CARARRIV	VACTROOMS
N		120	120
Exponential parameter [a, b, c]	Mean	12.3750	2.0190
Most Extreme Differences	**Absolute**	**.382**	**.303**
	Positive	.170	.303
	Negative	-.382	-.248
Kolmogorov-Smirnov Z		**4.185**	**3.100**
Asymp. Sig. (2-tailed)		.000	.000

a Test Distribution is Exponential.
b Calculated from data.
c There are 4 values outside the specified distribution range. These values are skipped.

To produce these outputs, choose the following commands from the menu:

Analyze

Nonparametric tests

1 Sample K-S ... (Test proportion is by default set at $0.50 = \frac{1}{2}$)

Test variable List: ***CARSARRIVAL; VACANTROOMS***[10]

Test distribution[11]: **Normal, Uniform, Poisson or Exponential**

Problem Set Three

Question 3.1

The two common brands of colour television picture tube in a city are alpha and beta. Recently, there are claims that the two brands are different in terms of the life in months of service before failure. To investigate this claim, a researcher collected data on the life twelve television sets of alpha-brand and eight of beta-brand. The results of the study are shown in the table.

[10] For grouped distributions midpoints are entered in the Data editor.
[11] Given that the two examples are dealing with arrival of cars or visitors, it is prudent to assume a test distribution as being Poisson. The other test distributions are displayed here for the purpose of comparison.

Determine if the life of television picture tube differs between the two brands $(\alpha = 0.05)$.

Brand	**Life in months of television picture tube**											
Alpha	33	36	39	41	46	48	42	34	35	40	28	50
Beta	24	40	30	36	37	28	19	49				

Question 3.2

The accompanying table gives average production records, based on a completely randomized design study that was undertaken to compare productivity of operators of four identical assembly machines. Is there a statistical support that the four assembly machine operators differ in average daily productivity?

Operator A	*Operator B*	*Operator C*	*Operator D*
225	225	220	224
222	219	215	210
230	225	218	221
205	213	206	212
235	228	227	230

4

INFERENCE TECHNIQUES FOR ORDINAL DATA: THE RANK-SUM TESTS

Rank-sum nonparametric tests rely on ranking of sample observations. Consequently, the rank-sum tests are applicable when data are measured at ordinal level or above. The statistic to be studied is the sum of ranks (rather than signs or values) of the sampled observations. Thus, the focus in rank-sum tests is *sampling distribution of the sum of ranks*. The analytical framework for using rank-sum nonparametric tests varies from a completely randomized design (one-way ANOVA) to a randomized block design (two-way ANOVA). The null hypothesis that is being tested also varies with the research design used. Four types of non-parametric tests are distinguished within the rank-sum category, and these are: Wilcoxon T-test for two-paired and correlated samples, Mann-Whitney U-test, Kruskal-Wallis H-test and Friedman test. These nonparametric tests are presented along with some practical examples.

4.1 Wilcoxon Signed-rank Test for Two-paired and Correlated Samples

The Wilcoxon non-parametric statistical test is typically used to determine the extent to which two samples are correlated and thus, come from identical populations. The focus of the test is on analysis of data generated under a *before-after*, or *with-without research designs*; hence the terms *paired* and *correlated samples*. Note that analysis of such research designs can be done by binomial test discussed under the sign tests. In this respect, Wilcoxon T-test provides an alternative way to binomial test for data that are measured at ordinal level and above.

The null hypothesis being tested is H_0: *the two samples come from identical populations*, and the alternative hypothesis is H_a: *the two samples come from different populations*. Wilcoxon T-test is used to ascertain hypothesis or propositions in research designs that are typically being associated with the before-after, with-without, or two treatments applied to homogenous experimental units. Thus, the focus of the test is to answer the following questions:

- Are there differences in the subjects' response before and after the treatment?
- Are there differences in the subjects' response with and without treatment?
- Are two treatments statistically different?

The statistic that is used to test the null hypothesis is the sum of ranks of the difference of the paired observations. The difference of the paired observations are ranked while retaining the positive or negative sign identity so that $R^+ = sum\ of\ positive\ ranks;\ and\ R^- = sum\ of\ negative\ ranks$. If the observations come from identical populations, i.e., $F_x(\bullet) = F_y(\bullet)$, then the number of observations with positive ranks will be equal to the number of observations with negative ranks. In this case, the difference between sum of positive ranks (R^+) and sum of negative ranks (R^-) will be small. Thus, if the null hypothesis is true, the statistic $T = \min(R^+, R^-)$ has the following asymptotical parametric characteristics with respect to mean and variance $(n = sample\ size = number\ of\ pairs\ without\ ties\ in\ ranking)$.

$$E(T) = \frac{n(n+1)}{4} \qquad and \qquad V(T) = \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Hence, the test statistic for the Wilcoxon non-parametric test is $Z = \frac{T - E(T)}{\sqrt{V(T)}} = \frac{T - E(T)}{\sigma_T}$. The accept area for the null hypothesis is $-1.96 < Z < 1.96$. The decision rule is: ***reject*** H_0 ***iff*** $|Z_c| > 1.96$.

However, for small samples critical values of T, above which H_0 is accepted, are tabulated (Table 6 appended) for each relevant level of significance and sample size n. For instance, a critical value of T for a two-tailed test $(n = 10,\ \alpha = 0.05)$ is $T_{.05,10} = 8$. Furthermore, a critical value of T for a one-tailed test $(n = 30,\ \alpha = 0.05)$ is $T_{.05,30} = 152$. In all the cases, the null hypothesis is accepted if the computed T-value is greater than the theoretical T-value, i.e., $T_c > T_{n,\alpha}$.

The computational steps for conducting Wilcoxon T-test are:

- Consider the differences in the subjects' response before and after (with and without treatment; under treatment 1 and treatment 2).

- Rank the *non-zero differences* while retaining the sign (positive or negative) identity; a difference of zero is ignored in the ranking process.
- Find sum of positive ranks and assign it as R^+.
- Find sum of negative ranks and assign it as R^-.
- Find $T = \min(R^+, R^-)$.
- Compute mean $E(T)$ and Variance $V(T)$ using the sample data.
- Obtain computed value of $Z_c = \frac{T - E(T)}{\sqrt{V(T)}} = \frac{T - E(T)}{\sigma_T}$ using the sample data; Accept H_0 iff $|Z_c| < 1.96$.

Example 4.1.1

An experiment is conducted to assess the effect of a particular type of diet on gain or weight of pigs. Weight measurements are recorded before and after management of the new diet. Use an appropriate nonparametric test to examine the claim that the new diet has significant effect on weight of pigs.

Before	70	63	68	65	61	71	67	75	90
After	66	65	68	67	65	73	66	71	93

Solution 4.1.1

(1) The null hypothesis is H_0: *New diet has no effect on weight of pigs*, and the alternative hypothesis is H_a: *New diet has effect on weight of pigs*.

(2) *Test statistic*: The appropriate nonparametric test is Wilcoxon test that is based on sum of ranks. The Wilcoxon T-test statistic is

defined as: $T = \min(R^+, R^-)$. For large samples this statistic has a normal distribution with respective mean and variance being equal to: $E(T) = \frac{n(n+1)}{4}$, and $V(T) = \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$. In that respect, the test statistic is $Z = \frac{T - E(T)}{\sigma_T}$.

(3) The level of significance is $\alpha = 0.05$; two-tailed.

(4) *Decision rule*: In the present case $n = 8$, which is statistically considered a small sample size. And given the small sample size on hand, critical value of the Wilcoxon T-statistic, obtained from Table 6, is $T_{\alpha,8} = 4$. The null hypothesis H_0: is accepted if $T_c > 4$. However, for large samples, the decision rule is to accept H_0 if computed value of the test statistic is $|Z_c| < 1.96$.

(5) The results of computational steps are summarized in the following table:

Before	70	63	68	65	61	71	67	75	90
After	66	65	68	67	65	73	66	71	93
Difference	-4	+2	0	+2	+4	+2	-1	-4	+3
Signed ranks	7^-	3^+	Na	3^+	7^+	3^+	1^-	7^-	5^+

- Ranks of positive differences are: $3,\ 3,\ 7,\ 3\ and\ 5$; Sum of positive ranks: $R^+ = 3 + 3 + 7 + 3 + 5 = 21$.
- Ranks of negative differences are: $1,\ 7\ and\ 7$; Sum of negative ranks: $R^- = 1 + 7 + 7 = 15$.
- $T = \min(21, 15) = 15$ and the valid sample size is $n = 8$.
- $E(T) = \frac{8(8+1)}{4} = 18$

- $V(T) = \frac{8(8+1)(2 \times 8 + 1)}{24} = 51 \Rightarrow \sigma_T = 7.14$
- $Z_c = \frac{15 - 18}{7.14} = -0.424$

(6) *Conclusion*: Given the decision rule and the sample results above, the hypothesis H_0 is accepted. It is, therefore, concluded that the new diet has no significant effect on weight of pigs.

Exercise 4.1.1

Has liberalized trade regime enhanced performance of small and medium enterprises? To answer this question, a study was conducted to determine the effect of liberalized trade regime on the performance of small and medium enterprises (SME's) in some selected African countries. The data on contribution to GDP and employment across eight sectors are summarized in the table below. Do these data provide sufficient evidence to indicate a changed operating characteristics and/or performance of SME's?

Sector	GDP Contribution (%)		Employment (%)	
	Before	After	Before	After
1	10.0	5.0	6.0	5.0
2	9.0	21.0	11.0	7.0
3	8.0	17.0	10.0	14.0
4	15.0	16.0	17.0	22.0
5	4.0	2.0	6.0	8.0
6	3.0	8.0	0.5	3.0
7	25.0	26.0	13.0	5.0
8	12.0	19.0	14.0	20.0

Exercise 4.1.2 (*optional*)

Does the use of parametric t-test on Exercise 4.1.1 change the results, findings and conclusions?

4.2 Mann-Whitney U Test

The Mann-Whitney U test is applicable when there are only two independent samples with sample sizes equal to n_1 *and* n_2 respectively. The objective is to ascertain the extent to which some specified set of attributes of the two samples are similar in which case the samples come from identical populations. The Mann-Witney U test uses the sum of ranks $R_1\ (or\ R_2)$ assigned to observations in the first (or second) sample. Thus, the useful statistics in the U-test are:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \ \ and \ \ U_2 = n_2 n_1 + \frac{n_2(n_2+1)}{2} - R_2$$

If the null hypothesis, H_0: *the two samples come from identical populations* is true, the test statistic is $U = \begin{cases} U_1 \ if \ n_1 \le n_2 \\ U_2 \ otherwise \end{cases}$

This test-statistic has the following asymptotical mean and variance given by:

$E(U) = \frac{n_1 n_2}{2}$ *and* $V(U) = \sigma_U^2 = \frac{n_1 n_2(n_1+n_2+1)}{12}$. Based on Central Limit Theorem, the statistic $\frac{U - E(U)}{\sigma_U} = Z$ has a standard normal distribution; $Z \approx N(0,1)$.

Thus, given two samples the computed U *and* Z values are obtained as follows:

- Define U statistic as: $U = \begin{cases} U_1 \ if \ n_1 \le n_2 \ (first\ sample) \\ U_2 \ otherwise \end{cases}$
- Rank all observations from the two samples while retaining sample identity.
- Compute sum of ranks R_1 *and* R_2 assigned to observations in the first and second sample respectively.
- Compute

  $$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \ \ and \ \ U_2 = n_2 n_1 + \frac{n_2(n_2+1)}{2} - R_2$$

- Compute the mean and variance of the U statistic – $E(U)$ *and* $V(U) = \sigma_U^2$
- Compute the value of test-statistic: $Z_c = \frac{U - E(U)}{\sigma_U}$ and this computed value is compared with the theoretical value $Z_{\alpha/2}$ at the appropriate level of significance.

For small samples, critical value U_0 can be obtained from the table of sampling distribution of U (Table 7 appended) for sample sizes n_1 and n_2 $(n_1 \le n_2)$; such that, $\Pr ob(U \le U_0) \le \alpha$. For instance, $U_0 = 10$ is a critical value of U for $n_1 = 6, n_2 = 8$ because of the fact that $\Pr ob(U \le 10) = 0.0406,$ ***which is less than*** $\alpha = 0.05$. Note also that $\Pr ob(U \le 11) = 0.0539$.

An alternative to the Mann-Whitney test for small samples is *Wilcoxon rank-sum T-test for two independent samples*. The Wilcoxon rank-sum T-test statistic is defined as: $T = \begin{cases} R_1 \text{ if } n_1 \le n_2 \\ R_2 \text{ otherwise} \end{cases}$ and the critical values, T_L and T_U, are obtained based on a sampling distribution of T (Table 8 appended). For instance, given that $n_1 = 7,\ n_2 = 10,\ \alpha = 0.05$, then critical value of T_α (one-tailed) is either $T_L = 46$ when reject area is the left tail or $T_U = 80$ when reject area is the right tail. However, the interval bounded by the critical values $T_L = 46$ and $T_U = 80$ would constitute an accept area for a two-tailed test at $\alpha = 0.10$ level of significance.

Example 4.2.1

The following data are on gains in weight of turkey fed with two types of diets A and B. Test the claim that the two animal feeds have unequal nutritional ingredients for the animals considered.

Diet A	14	13	12	15	10	16	
Diet B	15	12	14	9	8	17	11

Solution 4.2.1

(1) The claim being advanced by the researcher is that the two feeds have the different nutritional ingredients implying that the null hypothesis to be tested is: H_0: *the two sample come from the same population.* The alternative hypothesis is H_a: *the two samples come from different populations.*

(2) The *test-statistic* is Mann-Whitney defined as: $U = \begin{cases} U_1 \text{ if } n_1 \le n_2 \\ U_2 \text{ otherwise} \end{cases}$.

The statistic is based on the sum of ranks, R_1 and R_2 such that:

$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ and $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$. For large samples; $\frac{U - E(U)}{\sigma_U} = Z$ and this is a two- tailed test.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: Since small samples are being considered, a critical value for U given $n_1 = 6$ *and* $n_2 = 7$ is obtained from a sampling distribution of U such that: $\Pr ob(U \le U_0) \le \alpha$. In the present case $U_0 = 8$ because of the fact that, based on Table 7 appended, $\Pr ob(U \le 8) = 0.0367;$ ***which is less than*** 0.05. Thus, the null hypothesis H_0 is accepted iff $U_c \le 8$. In the absence of tables for the probability distribution of U the critical accept area for the null hypothesis is $-1.96 < Z < 1.96$, and the decision rule is: accept H_0 if $|Z_c| < 1.96$.

(5) *Computations*: Ranking the observations of a combined sample while retaining sample identify of each observation leads to the following results.

Scores	8_B	9_B	10_A	11_A	12_B	12_B	13_A
Rank	1	2	3	4	5.5	5.5	7
Scores	14_A	14_B	15_A	15_A	16_B	17_B	
Rank	8.5	8.5	10.5	10.5	12	13	

The ranks of observations of first sample (identified by the letter A) are: $3,\ 5.5,\ 7,\ 8.5,\ 10.5$ ***and*** 12, and hence, sum of ranks: $R_1 = 46.5$. The ranks of observations of the second sample (identified by the letter B) are: $1,\ 2,\ 4,\ 5.5,\ 8.5,\ 10.5$ ***and*** 13 so that the sum of ranks $R_2 = 44.5$. Given that the sample sizes are n_1 *and* n_2, then $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 16.5$, and thus, the computed value of ***U*** is $U_c = 16.5$, which is larger than $U_0 = 8$. Thus, H_0 is accepted.

Alternatively, the computation for the Z-test is as follows:

$$E(U) = \frac{n_1 n_2}{2} = 21 \; and \; V(U) = \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = 49 \Rightarrow \sigma_U =$$

. Finally, $Z_c = \frac{16.5 - 21}{7} = -0.64$, which lies in the accept area.

(6) *Conclusion*: The null hypothesis H_0: the two samples come from the same population is accepted. This result indicates that there is no difference in the gain of weight of turkey fed with the two types of feed. The conclusion is that the two animal feeds have equal nutritional ingredients. The reader can verify that the same conclusion is reached if sum of ranks of observations of the second sample is used. Note that: $R_2 = 44.5 \Rightarrow U_2 = U_c = 25.5$, and consequently, $Z_c = \frac{25.5 - 21}{7} = 0.643 < 1.96$; the null hypothesis is accepted.

Exercise 4.2.1

Use Wilcoxon Rank Sum T-test for Independent Samples to test the effectiveness of the animal feeds in the diet problem 4.2.1 above.

Exercise 4.2.2 (*optional*)

Does the use of parametric t-test on Exercise 4.2.1 change the results, findings and conclusions?

4.3 Kruskal-Wallis H test

The objective of Kruskal-Wallis H test is to ascertain the extent to which three or more $(k \geq 3)$ independent random samples possess a specified set of attributes and thus considered to come from identical populations. The null hypothesis being tested is H_0: *k-samples come from identical populations* and the alternative is H_a: *k-samples come from different populations*. In a way the test is an alternative to the one-way analysis of variance test for completely randomized research designs. Like in the Mann-Whitney U test, Kruskal-Wallis H test is based on the combined ranking procedure while retaining the sample identity. However, the test-statistic with Kruskal-Wallis test procedure is the variation of the sum of ranks. Given that R_i = sum of ranks assigned to observations in the i[th] sample, n_i = sample size, and $n = \sum n_i$ then the above null hypothesis is tested by H-statistic, which is computed as:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{n_i} \right) - 3(n+1)$$

If the null hypothesis H_0: is true, the statistic ***H*** has a Chi-square sampling distribution with $k-1$ degrees of freedom. The computational steps for conducting the above test statistic are summarized below as follows:

- Rank the observations of the two samples while retaining the sample identity.
- Find the sum of ranks $R_1, R_2 \ldots,$ and R_k assigned to observations of sample one, two, …, and k respectively.
- Compute size of combined sample: $n = \sum n_i$.
- Compute the H statistic based on the sample data – thereby obtaining the computed value of Chi-square - χ_c^2 .
- Compare the computed value of Chi-square with the critical theoretical value, which is $\chi^2_{\alpha.k-1}$. Accepted the null hypothesis if $\chi_c^2 < \chi_\alpha^2$.

Example 4.3.1

Three methods are adopted in teaching a programme, and at the end of the academic year the following average scores are obtained for the schools that were considered in the study:

Method A	94	87	91		
Method B	85	84	79	61	80
Method C	89	69	72	69	

Use appropriate rank-sum nonparametric test to determine the truth of the claim that the three methods are different.

Solution 4.3.1

(1) The null hypothesis H_0: *the three methods have the same effect on average academic achievement*, and the alternative H_a: *at least one of the methods has different effect on academic achievement*.

(2) *Test statistic*: The appropriate nonparametric technique is the Kruskal-Wallis *H* test-statistic. If the null hypothesis is true the *H* statistic has a Chi-square with $k-1$ degree of freedom; in the present case, $k = 3$.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: The critical value of the Chi-square is $\chi^2_{.05,2} = 5.991$.

The null hypothesis is rejected if the computed Chi-square value is greater than 5.991.

(5) *Computations*: The ranked scores and sum of ranks that are associated with each sample are summarized in the table below:

Signed-scores	61_B	69_C	69_C	72_C	79_B	80_B
Rank	1	2.5	2.5	4	5	6
Signed-scores	84_B	85_B	87_A	89_C	91_A	94_A
Rank	7	8	9	10	11	12

The sums of ranks assigned to observations in each sample are as follows:

Sample A: $R_A = 9 + 11 + 12 = 32$

Sample B: $R_B = 1 + 5 + 6 + 7 + 8 = 27$

Sample C: $R_C = 2.5 + 2.5 + 4 + 10 = 19$

From the sample data: $n = n_A + n_B + n_C = 3 + 5 + 4 = 12$, thus

$$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\left(\frac{R_i^2}{n_i}\right) - 3(n+1)$$

$$H = \frac{12}{12(12+1)}\left[\frac{32^2}{3} + \frac{27^2}{5} + \frac{19^2}{4}\right] - 3(12+1)$$

$$H = 44.41 - 39 = 5.41 = \chi_c^2 < \chi_\alpha^2$$

(6) *Conclusion*: Since the computed value of the Chi-square test statistic is smaller that the critical value, the null hypothesis cannot be rejected. This result implies that the three teaching methods are identical in terms of academic achievement for the schools considered in the pilot teaching programme.

Example 4.3.2

A study is conducted to compare auto-gasoline mileage (y_{ij}) for different brands of gasoline using six automobiles. Each gasoline brand was tested in each automobile, so as to eliminate (blocking out) the auto-to-auto variability and the resulting data, in kilometre per litre, are presented below.

y_{ij}	Automobile							
Gasoline	1	2	3	4	5	6	7	8
A	35	31	33	37	37	33	35	30
B	30	32	31	32	35	33	38	32
C	38	33	34	37	37	38	29	31
D	35	36	35	38	40	36	34	37
E	32	29	28	30	34	35	28	27
F	39	41	36	30	42	45	40	25

Do these data provide sufficient evidence to indicate a difference in the mean kilometre per litre for the six brands of gasoline?

Solution 4.3.2

(1) The null hypothesis H_0: *the mean kilometre per litre is the same for the six gasoline brands*, and the alternative H_a: *at least one of the gasoline brands has a different mean kilometre per litre.*

(2) *Test statistic*: The appropriate nonparametric technique is the Kruskal-Wallis ***H*** test-statistic. If the null hypothesis is true the ***H*** statistic has a Chi-square with $k-1$ degree of freedom; in the present case, $k=6$.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: The critical value of the Chi-square is $\chi^2_{.05,5} = 11.07$.

The null hypothesis is rejected if computed Chi-square value is greater than 11.07.

(5) *Computations*: The ranked observations/scores and ranks that are associated with each sample are summarized in the tables below.

Signed-score	25f	27e	28e	28e	29e	29c	30a	30b
Rank	1	2	3.5	3.5	5.5	5.5	8.5	8.5

Signed-score	30e	30f	31a	31b	31c	32b	32b	32b
Rank	8.5	8.5	12	12	12	15.5	15.5	15.5

Signed-score	32e	33a	33a	33b	33c	34c	34d	34e
Rank	15.5	19.5	19.5	19.5	19.5	23	23	23

Signed-score	35a	35a	35d	35d	35e	35b	36d	36d
Rank	27.5	27.5	27.5	27.5	27.5	27.5	32	32

Signed-score	36f	37d	37c	37c	37a	37a	38b	38c
Rank	32	36	36	36	36	36	40.5	40.5

Signed-score	38c	38d	39f	40d	40f	41f	42f	45f
Rank	40.5	40.5	43	44.5	44.5	46	47	48

The sums of ranks assigned to observations in each gasoline-sample are as follows:

$R_A = 8.5 + 12 + 19.5 + 19.5 + 27.5 + 27.5 + 36 + 36 = 186.5$

$R_B = 8.5 + 12 + 15.5 + 15.5 + 15.5 + 19.5 + 27.5 + 40.5 = 154.5$

$R_C = 5.5 + 12 + 19.5 + 23 + 36 + 36 + 40.5 + 40.5 = 213$

$R_D = 23 + 27.5 + 27.5 + 32 + 32 + 36 + 40.5 + 44.5 + = 263$

$R_E = 2 + 3.5 + 3.5 + 5.5 + 8.5 + 15.5 + 23 + 27.5 = 89$

$R_F = 1 + 8.5 + 32 + 43 + 44.5 + 46 + 47 + 48 = 270$

From the sample data: $n = n_A + n_B + n_C + n_D + n_E + n_F = 48$, thus

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{n_i} \right) - 3(n+1)$$

$$H = \frac{12}{197} \left[\frac{186.5^2}{8} + \frac{154.5^2}{8} + \frac{213^2}{8} + \frac{263^2}{8} + \frac{89^2}{8} + \frac{270^2}{8} \right] - 3(49)$$

$$\therefore\ H = 161.17 - 147 = 14.17 = \chi_c^2 > \chi_\alpha^2$$

(6) *Conclusion*: Since the computed value of the Chi-square test statistic is larger than the critical value, the null hypothesis cannot be accepted. This result implies that the six gasoline brands are not identical in terms of the mean kilometre per litre brands reported in the study.

Exercise 4.3.1

Use k-sample median test to test the hypothesis that the six gasoline brands have the same mean kilometre per litre.

Exercise 4.3.2 *(optional)*

Does use of F-test on Exercise 4.3.1 above change the conclusions regarding the null hypothesis?

4.4 *Friedman Test for k Correlated Samples*

The Friedman nonparametric test is used or applied when observations, from randomized-block designed experiments, are measured at ordinal level. Randomized block designed experiments are by definition characterized by k levels of treatment and p levels of blocking factor $(k \geq 3,\ p \geq 2)$. Thus, the Friedman test is an alternative to the two-way analysis of variance test, in which two null hypotheses are being tested. First, a null hypothesis (H_{01}) that is being associated with the extent to which k-level treatment has effect on a response variable-phenomenon being investigated. Second, a null hypothesis tested is (H_{02}), and this is being associated with the extent to which p-levels block affects the response variable.

Thus, the typical null hypotheses being tested are; H_{01}: *Treatment levels have no effect on the response variable-phenomenon* and H_{02}: *Block levels have no effect on the response variable-phenomenon.* The test is based on ranking of observations; being based on treatment if H_{01} is tested and on blocking if H_{02} is tested. There are, therefore, two kinds of rankings, namely, ranking of treatment that is curly bracketed by (•) and whose row total is $R_i = \sum(\bullet)$; and ranking within block that is curly bracketed by [•] and whose column total is $R_j = \sum[\bullet]$.

Typically data from randomized-block designed experiments are tabulated as follows:

Treatments	Blocks			$R_i=\sum(\bullet)$
	X	Y	Z	
A	x_{11} (•) [•]	y_{12} (•) [•]	z_{13} (•) [•]	$R_{1.}$
B	x_{21} (•) [•]	y_{22} (•) [•]	z_{23} (•) [•]	$R_{2.}$
C	x_{31} (•) [•]	y_{32} (•) [•]	z_{33} (•) [•]	$R_{3.}$
D	x_{41} (•) [•]	y_{42} (•) [•]	z_{43} (•) [•]	$R_{4.}$
$R_j=\sum[\bullet]$	$R_{.1}$	$R_{.2}$	$R_{.3}$	

The computational steps that are involved for carrying out the Friedman test are:

- Rank the k observations within block levels – curly bracketed by (•).
- Rank the p observations within each treatment levels – curly bracketed by [•].
- Compute sum of ranks for each treatment level and assign the sum as R_i.
- Compute sum of ranks for each block and assign the sum as R_j.
- Compute variation of treatment rankings: $S_T=\sum_{i=1}^{k}R_i^2-\frac{\left(\sum_{i=1}^{k}R_i\right)^2}{k}$.
- Compute variation of block rankings: $S_B=\sum_{j=1}^{p}R_j^2-\frac{\left(\sum_{j=1}^{p}R_j\right)^2}{p}$.
- If the null hypotheses are true, the statistics $\frac{12S_T}{pk(k+1)}$ and $\frac{12S_B}{kp(p+1)}$ have a Chi-square sampling distribution with $k-1$ and $p-1$ degrees of freedom respectively. The statistics

$F_{rT}=\frac{12S_T}{pk(k+1)}$ and $F_{rB}=\frac{12S_B}{kp(p+1)}$ are also computed by using the following respective expressions:

$$F_{rT}=\frac{12}{pk(k+1)}\left[\sum R_i^2\right]-3p(k+1)$$

$$F_{rB}=\frac{12}{kp(p+1)}\left[\sum R_j^2\right]-3k(p+1)$$

Example 4.4.1

Four varieties of wheat were planted randomly under three types of fertilizers with the following average yield results presented below. Use appropriate nonparametric statistical test to ascertain the claim that wheat variety and fertilizer type have effect on yield.

Wheat\Fertilizer	X	Y	Z
A	51	52	47
B	50	43	42
C	46	46	42
D	49	49	46

Solution 4.4.1

(1) The wheat represents the treatment and the fertilizer type typifies the block. Thus, the null hypotheses are: H_{01}: *wheat variety has no effect on yield and:* H_{01}*: fertilizer type has no effect on yield.*

(2) *Test statistic*: The envisaged research design of the experiment is randomized-block design, and hence, the appropriate nonparametric technique is Friedman test based on Chi-square.

(3) The level of significance is $\alpha=0.05$

(4) *Decision rule*: Given that treatment factor has $k=4$ levels and the fertilizer type has $p=3$ levels, then the critical value of the Chi-

square is $\chi^2_{\alpha,k-1} = 7.81$ for testing the wheat type effect (H_{01}) and $\chi^2_{\alpha,p-1} = 5.99$ for testing the fertilizer effect (H_{02}).

(5) *Computations*: Results of computations for the Friedman nonparametric test are summarized in the table below {treatment ranks are curly bracketed as $R_i = \sum(\bullet)$, while blocking ranks are embraced by $R_j = \sum[\bullet]$

Wheat	Fertilizer X	Y	Z	$R_i = \sum(\bullet)$
A	51 (4) [2]	52 (4) [3]	47 (4) [1]	12
B	50 (3) [3]	43 (1) [2]	42 (1.5) [1]	5.5
C	46 (1) [2.5]	46 (2) [2.5]	42 (1.5) [1]	4.5
D	49 (2) [2.5]	49 (3) [2.5]	46 (3) [1]	8
$R_j = \sum[\bullet]$	10	10	4	NA

- Variation of wheat-rankings is computed by:

$$S_T = \sum_{i=1}^{k} R_i^2 - \frac{\left(\sum_{i=1}^{k} R_i\right)^2}{k} = 258.5 - 225 = 33.5.$$

- Variation of fertilizer rankings is computed by:

$$S_B = \sum_{j=1}^{p} R_j^2 - \frac{\left(\sum_{j=1}^{p} R_j\right)^2}{p} = 216 - 192 = 24.$$

- The computed values of the Chi-square for H_{01} and H_{02} are respectively equal to $\chi^2_T = \frac{12S_T}{pk(k+1)} = \frac{12 \times 33.5}{3 \times 4(4+1)} = 6.7$, a value that is smaller than the theoretical value of $\chi^2_{.05,3} = 7.81$; and $\chi^2_B = \frac{12S_B}{kp(p+1)} = \frac{12 \times 24}{4 \times 3(3+1)} = 6.0$, which is greater than the theoretical value of $\chi^2_{.05,2} = 5.99$.

(6) *Conclusion*: Results indicate that H_{01}: is accepted while H_{02}: is rejected, implying that the *wheat variety has no effect on yield*; while the *fertilizer type has effect on yield.*

Example 4.4.2

Refer to Example 4.3.2 reported in section 4.3.

A study is conducted to compare auto-gasoline mileage for different brands of gasoline using six automobiles. Each gasoline brand was tested in each automobile, so as to eliminate (blocking out) the auto-to-auto variability and the resulting data, in kilometre per litre, are presented below.

y_{ij} Gasoline	Automobile 1	2	3	4	5	6	7	8
A	35	31	33	37	37	33	35	30
B	30	32	31	32	35	33	38	32
C	38	33	34	37	37	38	29	31
D	35	36	35	38	40	36	34	37
E	32	29	28	30	34	35	28	27
F	39	41	36	30	42	45	40	25

Suppose the researcher reveals later that autos 1 – 3 were run on a rough but dry road, autos 4 and 5 on a tarmac-slippery road, and autos 6 – 8 on a tarmac but mountainous road. Use appropriate technique to analyse the data given this additional information.

Solution 4.4.2

The reader will note that with the additional information, the problem is now a randomized-block design, because of the differences in **gasoline technology** as well as the **road attribute**. Hence, mean kilometre per litre could as well be influenced by road attribute.

(1) The gasoline brand represents the treatment and the road attribute typifies the block. Thus, the null hypotheses are; H_{01}: *mean kilometre is the same for six gasoline brands* and H_{01}: *road attribute has no effect on mean kilometre per litre for the gasoline brands.*

(2) *Test statistic*: The envisaged research design of the experiment is randomized-block design, and hence, the appropriate nonparametric technique is Friedman test based on Chi-square.

(3) The level of significance is $\alpha = 0.05$

(4) *Decision rule*: Given that treatment factor has $k = 6$ levels and road has three attributes, then $p = 3$ levels, then the critical value of the Chi-square is $\chi^2_{\alpha,k-1} = 11.07$ for testing the gasoline brand type effect (H_{01}) and $\chi^2_{\alpha,p-1} = 5.99$ for testing the road effect (H_{02}).

(5) *Computations*: The mean kilometre per litre for each gasoline-auto combination is recomputed taking into account the road attribute. For instance, the mean for (A, 1-3) combination is $\frac{35+31+33}{3} = 33$. Results of computations for the Friedman nonparametric test are summarized in the table below {treatment ranks are curly bracketed as (•), while blocking ranks are embraced by [•]}

Gas	Road			$R_i = \sum(\bullet)$
	1 – 3	4 – 5	6 – 8	
A	33 (3) [2]	37 (4.5) [3]	32.7 (2.5) [1]	10
B	31 (2) [1]	33.5 (2) [2]	34.3 (4) [3]	8
C	35 (4) [2]	37 (4.5) [3]	32.7 (2.5) [1]	11
D	35.3 (5) [1]	39 (6) [3]	35.7 (5) [2]	16
E	29.7(1) [1]	32(1) [3]	30 (1) [2]	3
F	38.3(6) [3]	36(3) [1]	36.7(6) [2]	15
$R_j = \sum[\bullet]$	10	15	11	NA

- Variation of gasoline-rankings is computed by:

$$S_T = \sum_{i=1}^{k} R_i^2 - \frac{\left(\sum_{i=1}^{k} R_i\right)^2}{k} = 775 - 661.5 = 113.5$$

- Variation of road-rankings is computed by:

$$S_B = \sum_{j=1}^{p} R_j^2 - \frac{\left(\sum_{j=1}^{p} R_j\right)^2}{p} = 446 - 432 = 14$$

- The computed values of the Chi-square for H_{01} and H_{02} are respectively equal to: $\chi^2_T = \frac{12S_T}{pk(k+1)} = \frac{12 \times 113.5}{3 \times 6(6+1)} = 10.81$, which is smaller than the theoretical value of $\chi^2_{.05,5} = 11.07$; and $\chi^2_B = \frac{12S_B}{kp(p+1)} = \frac{12 \times 14}{6 \times 3(3+1)} = 2.33$, again it is smaller than the theoretical value of $\chi^2_{.05,2} = 5.99$.

(6) *Conclusion*: Results indicate that both the null hypotheses, H_{01} and H_{02}, are accepted, implying that the *mean kilometre per litre is*

independent of gasoline brands or road attribute for the auto mileage data reported in this study.

Exercise 4.4.1

Taking example 4.4.2 above, suppose further records indicate that gasoline brands A, B, C and D are extracted by a technology that is different from the one used to extract brands E and F. How will this piece of new information affect the analysis and results?

Exercise 4.4.2

The typing speeds, in words per minute, for each secretary-typewriter combination, based on a study undertaken to determine the relative typing speeds of four different brands of electronic typewriters are summarized in the accompanying table. Each typewriter was assigned to each of eight secretaries, the order of assignment conducted in a random manner.

Typewriter Brand	Secretary							
	1	2	3	4	5	6	7	8
A	79	80	77	75	82	77	78	76
B	74	79	73	70	76	78	72	74
C	82	86	80	79	81	80	80	84
D	79	81	77	78	82	77	77	78

(a) Do the data provide sufficient evidence to indicate that the mean typing speed for the secretaries varies from brand to brand of typewriters? Test by using 5% level of significance.

(b) Suppose it is later revealed that the secretaries are graduates from four different VETA Colleges such that: 1-2 are from College I, secretaries 3-4 from College II, secretaries 5-6 from College III, and secretaries 7-8 from College IV. How would this additional information change the results you have performed in (a) above?

Exercise 4.4.3 *(optional)*

To what extent does use of two-way ANOVA change the conclusions regarding the null hypotheses on the wheat and the gasoline problems documented in the section?

4.5 SPSS Tutorial on Rank-sum Test Techniques

The SPSS output for each example worked in this chapter is presented along with the data entry in the Data Editor. This is followed by demonstrating the sequence of commands from the menu that generates the output.

4.5.1 The Wilcoxon Signed Ranks Test

The SPSS output for the problem of example **4.1.1** on animal diet and weight of pigs, using **Wilcoxon Signed Ranks Test** for two correlated samples is:

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
WEGHTA	9	70.0000	8.61684	61.00	90.00
WEIGHTB	9	70.4444	8.88976	65.00	93.00

Ranks

		N	Mean rank	Sum of ranks
WEIGHTB - WEGHTA	**Negative Ranks**[a]	3	5.00	15.00
	Positive Ranks[b]	5	4.20	21.00
	Ties[c]	1		
	Total	9		

a WEIGHTB < WEGHTA
b WEIGHTB > WEGHTA
c WEGHTA = WEIGHTB

Test Statistics [b]

	WEIGHTB - WEGHTA
Z	**-.424** [a]
Asymp. Sig. (2-tailed)	.671

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

To produce this output, choose the following commands from the menu:

Analyze

Nonparametric

2 Related samples

Test pair list; current selections: ***WEIGHTA*** (weight after as variable 1); ***WEIGHTB*** (weight before as variable 2)

Test type: **Wilcoxon** (**Sign** or **McNemar** as the case may be)

Statistics: **Descriptive**

4.5.2 The Mann-Whitney U Test and Wilcoxon T Test for Independent Samples

Based on **Mann-Whitney U** test and **Wilcoxon T** test for independent samples, and data on weight of turkey feed with two types of diets (problem 4.2.1), the SPSS output is:

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
WEIGHTGAIN	13	12.7692	2.74329	8.00	17.00
DIETAB	13	1.54	.519	1	2

Ranks

	DIETAB	N	Mean Rank	Sum of Ranks
WEIGHTGAIN	1	6	7.75	46.50
	2	7	6.36	44.50
	Total	13		

Test Statistics [b]

	WEIGHTGAIN
Mann-Whitney U	**16.500**
Wilcoxon W	**44.500**
Z	**-.646**
Asymp. Sig. (2-tailed)	.519
Exact Sig. [2*(1-tailed Sig.)]	.534 [a]

a Not corrected for ties.
b Grouping Variable: DIETAB

To produce this output, choose the following commands from the menu:

Analyze

Nonparametric

2 Independent Samples

Test Variable List: ***WEIGHTGAIN*** (data on weight of turkey with diets A and B are entered as one variable 1); ***DIETAB*** (entered as grouping variable; 1 = for diet A and 2 = for diet B)

Test type: **Mann-Whitney** (**Kolmogorov-Smirnov** as the case may be)

Statistics: **Descriptive**

4.5.3 Kruskal-Wallis H-Test

Below are the SPSS output for problems on the choice of teaching (problem 4.2.1) and mileage-gasoline (problem 4.2.2) based on Kruskal-Wallis H-Test.

Teaching Method Problem

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
SCORES	12	80.00	10.251	61	94
TEACHMETHOD	12	2.08	.793	1	3

Ranks

	TEACHMETHOD	N	Mean Rank
SCORES	1	3	10.67
	2	5	5.40
	3	4	4.75
	Total	12	

Test Statistics [a, b]

	SCORES
Chi-Square	**5.433**
Df	2
Asymptotic Significant	.066

a Kruskal-Wallis Test
b Grouping Variable: TEACHMETHOD

Mileage-gasoline Problem

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
MILEAGE	48	34.23	4.173	25	45
GASOLINE TYPE	48	3.50	1.726	1	6

Ranks

	GASOLINE TYPE	N	Mean-Rank
MILEAGE	1	8	23.31
	2	8	19.31
	3	8	26.63
	4	8	32.88
	5	8	11.13
	6	8	33.75
	Total	48	

Test Statistics [a, b]

	MILEAGE
Chi-Square	**15.087**
Df	5
Asymptotic Significant	.010

a Kruskal-Wallis Test
b Grouping Variable: GASOLINE TYPE

To produce the above output, choose the following commands from the menu:

Analyze

Nonparametric

K-independent Samples

Test Variable List: ***SCORES (MILEAGE)*** (data on weight of turkey with diets A and B are entered as one variable); ***TEACHMETHOD*** (entered as grouping variable; 1 = minimum; 3 = maximum)

Test type: **Kruskal-Wallis**

Statistics: **Descriptive**

4.5.4 The Friedman Test

There are two null hypotheses under the Friedman Test; H_{01} : *on the treatment effect* and H_{02} : *on the blocking effect*. Thus, when testing the effect of the treatment, there are k-levels and hence k test variables for the purpose of testing treatment effect. The four wheat types in example 4.4.1 and the six gasoline brands in example 4.4.2 represent the treatments levels and thus the test variables. In testing the blocking there are p-levels of blocking. The three levels of fertilizer factor in example 4.2.1 or the three road attributes in example 4.2.2 represent the test variables for the purpose of testing the null hypotheses that are associated with blocking effect. The treatment and block variable list for both examples is summarized in the accompanying table.

Variable list	Example 4.2.1	Example 4.2.2
Treatment	***WHEATA***	***GASOLINEA13***
	WHEATB	***GASOLINEB13***
	WHEATC	***GASOLINEC13***
	WHEATD	***GASOLINED13***
		GASOLINEE13
		GASOLINEF13
Block	***FERTILIZERX***	***AUTO13***
	FERTILIZERY	***AUTO45***
	FERTILIZERZ	***AUTO68***

The SPSS output of problems 4.4.1 and 4.4.2 using Friedman Test for k correlated samples are presented below:

The wheat yield problem

Ranks

	Mean Rank
WHEATA	4.00
WHEATB	1.83
WHEATC	1.50
WHEATD	2.67

Test Statistics [a]

N	3
Chi-Square	**6.931**
df	3
Asymp. Sig.	.074

a Friedman Test

Ranks

	Mean Rank
FERTILIZERX	2.50
FERTILIZERY	2.50
FERTILIZERZ	1.00

Test Statistics [a]

N	4
Chi-Square	**6.857**
df	2
Asymp. Sig.	.032

a Friedman Test

The mileage-gasoline-road problem

Ranks

	Mean Rank
GASOLINEA13	3.33
GASOLINEB13	2.67
GASOLINEC13	3.67
GASOLINED13	5.33
GASOLINEE13	1.00
GASOLINEF13	5.00

Test Statistics [a]

N	3
Chi-Square	**11.019**
df	5
Asymp. Sig.	.051

a Friedman Test

Ranks

	Mean Rank
AUTO13	1.67
AUTO45	2.50
AUTO68	1.83

Test Statistics [a]

N	6
Chi-Square	**2.333**
df	2
Asymp. Sig.	.311

a Friedman Test

To produce the above output, choose the following commands from the menu:

Analyze

Nonparametric

K-Related Samples

Test Variable List: (see in the table above)

Test type: **Friedman**

Statistics: **Descriptive**

Problem Set Four

Question 4.1

It is being suggested that consumer protection legislation in Tanzania is ineffective because consumers do not possess sufficient understanding of the law to recognize illegal activities and do not have sufficient confidence that legal system will punish violators. To explore this aspect, a researcher has examined consumer attitudes towards the cooling-off law, a provision that provides consumers time during which they may consider certain purchases made from direct-to-home salespeople.

Since consumers in low-income areas are not likely to be in contact with direct-to-home salespeople, the researcher decided to compare the attitudes toward the cooling-off law by contacting $n_1 = 10$ residents from low-income area and $n_2 = 14$ residents from high-income area. Each respondent was asked to indicate, on a 10-point scale, the extent to which they agree or disagree with the following statement: "*Consumer protection laws can help stop unfair practices used in business.*" The consumer responses, with 1 indicating complete agreement with the statement and 10 indicating complete disagreement, are given in the table.

Low-income	8.5	9.5	9.0	7.5	8.0	8.5	10.0	5.0	6.5	7.5
High-income	4.0	3.5	5.5	6.5	7.0	6.0	2.0	5.0	4.5	5.5
	3.0	6.5	9.0	5.0						

Do these data suggest that residents from low-income areas have less confidence in consumer protection laws than those from high-income areas?

Question 4.2

Research into managerial and employee preferences for alternative forms of compensation (fringe benefits) has produced some intriguing results. Some researchers have found that preferences often vary according to size of the employee benefit package offered by management. To explore this possible possibility, a researcher asked employees of an electronics manufacturer to record their preferences for the percentage distribution of compensation in an anticipated 5.0% compensation increase and a 10.0% increase. Their average responses are shown in the following table.

Form of compensation	Percentage distribution	
	5.0% increase	*10.0% increase*
Salary	35	49
Improved medical plan	14	20
Fully paid retirement plan	22	13
Increase in life insurance benefits	8	5
Full cover of social security	17	11
Increased annual leave period	2	1
Paid disability income plan	2	1

Is there a sufficient evidence to indicate a difference in the employee's compensation preferences when total amount of compensation is increased from 5% to 10%?

Question 4.3

Organizations are increasingly placing greater demands on their employees to be adaptive to a changing work environment. According to many students of organizational behaviour, these demands may be contributing to increased incidence of employee stress reactions. To measure the effects of stress in two different settings, an analyst asked several employees from private firms and several others from public service to indicate, on a seven-point Likert scale, the degree to which organizational factors have caused their personal stress reactions (1= no reaction; 7= extreme reaction). The

average responses of those interviewed are shown in the accompanying table. Do these data suggest that sufficient evidence exists to indicate a difference in personal stress reactions between employees in private firms and those of public service?

Personal stress reaction	Employees in private firms	Employees in public service
Emotional stress	5.52	3.76
Low self-esteem	5.08	3.48
Low trust	4.98	4.45
Reduced personal productivity	4.96	3.78
Cardiovascular	3.75	5.43
Frequent absenteeism	2.53	2.47
Low job satisfaction	5.62	3.65
Gastrointestinal	2.19	4.92

Question 4.4

A corporation has agreed to offer a new health insurance plan to its employees. Two large insurance companies have each presented a health insurance plan to management; both are within the guidelines specified in the collective-bargaining agreement and should equally be costly to the corporation. To obtain worker's attitudes toward the competing plans, the personnel officer randomly selected twelve employees from the employee registry. Both insurance plans were explained in detail to each employee in the sample. The employees were then asked to rate each plan on a ten-point Likert scale[12], with 1 indicating the plan is completely unacceptable and 10 indicating perfectly acceptable. The recorded responses are shown in the table. Do these data suggest that the employees' attitudes differ significantly toward these two insurance plans?

---

[12] A Likert scale is an instrument that associates ordinal values with qualitative attributes. In this case the attribute is measure of acceptability for which the associated Likert scale appears as follows:

Opinion: *completely unacceptable* *acceptable* *completely acceptable*

0 1 2 3 4 5 6 7 8 9 10

Insurance plan	Employee					
	1	*2*	*3*	*4*	*5*	*6*
Plan I	6.0	5.5	7.0	7.5	9.5	9.0
Plan II	4.0	3.0	9.0	8.5	8.0	7.0
Insurance Plan	Employee					
	7	*8*	*9*	*10*	*11*	*12*
Plan I	8.5	8.0	9.0	5.5	3.5	9.5
Plan II	6.0	8.0	6.5	7.0	4.5	8.5

Question 4.5

There are variations in opinions, regarding organizational pay increases, between management and workers. Some suggest that they are signs of organizational recognition, while others contend that pay increases are primarily valued as inflation adjustments. Eight line employees and nine managers of an industrial organization were asked to respond to the following seven-point Likert scale regarding their interpretation of the meaning of company pay increases. The survey provided the results shown in the table, with 1 indicating *totally a sign of organizational recognition*; and 7 indicating *totally an inflation adjustment*. Do these data present sufficient evidence to indicate a difference between the perceptions of line employees and management of the intent of pay raises?

Line Employees	4.8	5.1	6.1	5.5	7.0	6.0	6.6	4.9	4.5
	5.0	6.5	3.3						
Managers	3.0	2.4	3.5	4.3	5.4	2.8	5.4	3.8	7.0

Question 4.6

Corporate management studies suggest that when workers feel that they have a stake in the ownership, and management, worker productivity is enhanced. A study designed to examine this aspect measured perceptions by worker-owners and managers of their role as owners, financial partners, and joint decision makers. Interviews conducted in several firms among

workers and managers who owned at least 100 shares of stock in the firm in which they were employed. Each person was asked to respond on a six-point scale to the following statement: "*Management and workers are equal partners in this company.*" The responses of worker-owners and managers in one firm are shown in the accompanying table, with a response of 1 implying strong disagreement and 6 implying strong agreement.

Worker-owners	4.5	3.0	3.2	3.7	2.4	1.8	4.0
Managers	5.0	4.0	4.6	5.5	4.4		

Do these data suggest that the perceptions of worker-owners and managers differ regarding the distribution of power within the firm?

Question 4.7

Companies requiring extreme precision in assembly operations, especially those in high technology industries, often find that the rate of productivity differs considerably between male and female employees involved in assembly work. To examine this issue, an electronics manufacturer observed the time for completion of an intricate electronics component by 13 women and 12 men employed as assemblers. The order in which the 25 employees successfully assembled the component was, a woman was first to successfully assemble the product, another woman completed the assembly in the second shortest time, and so on, as follows:

WWWMWMMWWW WMWMMMWMMW WWMMM

Is there a difference in assembly times required by men and women to assemble the electronic component?

Question 4.8

It is sometimes remarked that brand loyalists possess blind loyalty to a brand even though they may agree that a competing brand is more efficient in terms of the primary use of the product. To investigate this theory, a researcher contacted nine housewives and asked them to rank the seven leading brands of powdered laundry detergent in terms of the cleaning power they perceived the brands to possess. Five of the nine housewives were loyalists to brand A, while the other four possessed no brand loyalty among the five competing brands. The ranks given to brand A by the nine housewives are listed in the table, with a rank of 1 implying greatest cleaning power; and 7 implying the least cleaning power.

Brand Switchers	3	4	1	6	
Brand Loyalists	5	1	2	3	4

Do these data suggest that loyalists to brand A rank it higher in cleaning power, on the average, than those who possess no brand loyalty?

Question 4.9

The accompanying table summarizes data, on average variable cost per km, obtained by a trucking company that wished to compare three makes of trucks before ordering an entire fleet of one of the makes. The experiment involved running five trucks of each make for 3,000 kilometres. However, because of tyre failure, accidents, and driver illness, two trucks for makes A and B did not complete the 3,000-km test. Do these data provide sufficient evidence to indicate a difference in the average variable cost per km of operation for the three makes of trucks?

Make A	25.4	27.9	27.2		
Make B	27.9	29.5	28.3	29.8	
Make C	28.0	27.6	28.3	26.9	30.2

Question 4.10

The Banking and Financial Institutions, Act 1992, allows banks to become dual issuers of credit cards, and thus offering banks the opportunity to serve a broader variety of their customers' needs. But to develop proper policy regarding the distribution of credit cards, bankers must first understand the credit usage rate. In response to this issue, a major Bank has commissioned a study to compare the use of credit by those who hold master cards, visa cards, and those who hold both. The credit usage over a three-month period by 18 subjects involved in the study is given in the table. Do these data provide sufficient evidence to indicate a difference in the average credit usage by the three different groups of card holders?

Master Card	*Visa*	*Both Cards*
Shs. 2,500,000	*Shs.* 866,500	*Shs.* 3,800,300
1,600,000	2,450,500	2,000,900
1,005,000	4,900,000	5,850,000
980,500	500,000	250,000
3,600,000	600,000	
5,200,400	3,015,000	
400,000		
4,530,000		

Question 4.11

Read the following Case Study and then answer the questions that follow.

CASE STUDY

Richmond's Response to the Regulators

The Government Environment Management Agency (GEMA) has issued regulations mandating a gradual phase-out of tetraethyl lead (TEL) for gasoline sold in the United Republic of Amani. This has come about as a result of several major studies conducted by GEMA during the past five years. Separate legislation developed by other Government agencies provides even stiffer limits for the TEL content of gasoline.

For years, refiners had used TEL as an addition to gasoline as a cheap and convenient way to improve the octane rating of gasoline, thus reducing the potential for harm to motor vehicle engines. In the absence of TEL a refinery must reprocess some of the low-octane components of gasoline to increase their octane ratings. This can be accomplished either by breaking apart the hydrocarbon chains, through processes known in the trade as cat cracking or hydro-cracking or by rearranging the bonding in the chains, through processes called reforming or alkylation. All four processes are very costly and efficient but provide more variability in results than simple addition of TEL to improve the octane ratings of a blend.

Faced with the dual impact of federal regulations on TEL and the more stringent conditions of other Government agencies, officials at Richmond Refinery initiated a crash programme to determine the most efficient means of addressing the new requirements. Experiments were undertaken using each of the four known methods of increasing octane ratings without exceeding mandated limits on the use of TEL. In their study, gasoline from Richmond refinery was reprocessed in a way such that the costs were equalized by using each experimental procedure. The data in the table show the octane ratings resulting from application of the four reprocessing

procedures to gasoline derived from each of the eight storage tanks, distributed as follows: storage tanks 1, 2, 3, and 4 had crude oil from the newly discovered oil-fields, tanks 5, 6, and 7 had imported crude oil; and tank 8 had strategic crude oil reserve.

Storage Tank	**Octane Rating when Reprocessed**			
	Cat cracking	*Hydro-cracking*	*Reforming*	*Alkylation*
1	88	98	95	92
2	87	96	94	88
3	86	85	90	84
4	90	79	93	87
5	95	87	81	80
6	85	91	86	94
7	92	90	88	90
8	91	88	89	79

(a) Do these data suggest that differences exist in the ability of the four reprocessing procedures?

(b) Suppose it is later revealed that the storage tanks 1 – 3 are of same capacity – each carrying 1,000,000 litres; tanks 4 – 6 each carry 500,000 litres; and that tanks 7 – 8 are smaller ones each with a capacity of 100,000 litres. How would this additional information change the analytical inference methods you have performed in (a) above?

5

THE WALD-WOLFOWITZ RUNS TEST OF RANDOMNESS

Statistical inference methods and techniques are all characterized by one major assumption, which is that the study sample is random. How do we ascertain that the sample is random? That is, given a sequence of events or observations, how do we check that they constitute a random phenomenon? The answers to such questions are based on a number of times there is a change in sequence of the events or attributes of sampled observations. The change in sequence of events or attributes of sampled observations is called a run.

5.1 The Runs Test

A run is a change of sequence of identical events or attributes of sampled observations. Runs-test determines a random sequence of events or attributes of observations where each element in the sequence may assume one of two outcomes such as, *success* versus *failure*, *good* versus *bad*, *large* versus *small*, *non-defective* versus *defective*, or *below* versus *above* a level of some specified attribute index. The following definition of the concept of a run is adopted.

Definition

A run is a succession of identical events or attributes that may be represented by letters or other symbols, followed by different successions of events or attributes or no events at all.

The runs-test is thus applied to a sequence of n_1 successes and n_2 failures. In a manufacturing process, for instance, it is of great interest to know the extent to which sequence of defective and/non–defective items are randomly produced. Non-randomness of the sequence of defective items in a manufacturing process indicates that there is lack of process control, and if this is revealed, appropriate corrective actions may be taken immediately.

Theoretically, frequent changes of identical succession of events or attributes in sampled observations indicate presence of a random phenomenon being in place. Frequent changes of identical succession of attributes leads to a large number of runs (R), *and hence the number of runs is positively associated with a random phenomenon being in place.*

The null hypothesis that is being tested in a runs test is H_0: *the sample is random* and the alternative hypothesis is H_a: *the sample is non-random.* The number of runs (or identical succession of events or attributes of sampled observations) tests randomness of the sample. If the null hypothesis is true, the number of runs R has the following respective asymptotical expected mean and variance:

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 \text{ and } V(R) = \sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

For large samples, the statistic $Z = \frac{R - E(R)}{\sigma_R}$ follows a standard normal distribution. Consequently, smaller values of the test-statistic Z in absolute terms indicate that the sequence of events or observations in a sample constitute a random phenomenon.

However, for small samples probability distribution of the total number of runs R for sample sizes n_1 and n_2 is found in Table 9 appended, which is limited to samples with $n_1, n_2 \le 10$. For instance, given that $n_1 = 8,\ n_2 = 10,\ and\ \alpha = 0.05$ then the critical value of ***R*** such that $\Pr ob(R \le R_0) < \alpha$ is $R_0 = 6$. Thus, the null hypothesis is accepted iff $R_c > 6$.

Example 5.1.1

The following arrangement of healthy (H) and diseased (D) pine trees observed next to each other in a given order is in the records of Forestry Service Crew. The arrangement is:

$$HHHHHDDDHHHHHHHHHDDHHDDDDDHHHH.$$

Is there a statistical evidence to show that the observed distribution of pine trees constitutes a random phenomenon?

Solution 5.1.1

(1) The null hypothesis is H_0: *Distribution of trees is random* and the alternative hypothesis is H_a: *Distribution of pine trees is non-random.*

(2) *Test statistic*: The appropriate nonparametric test is the number of runs ***R*** whose mean value and variance are respectively equal to:

$$E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1\ and\ V(R) = \sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Hence, given large samples, the statistic $Z = \frac{R - E(R)}{\sigma_R}$ tests the above null hypotheses.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: for small samples critical value of ***R*** is obtained from statistical table 9 **(appended)** of probability distribution of the total number of runs ***R*** such that $\Pr ob(R \le R_0) < \alpha$. In that case the null hypothesis is accepted iff $R_c > R_0$. For large samples the statistic $Z = \frac{R - E(R)}{\sigma_R}$ has a standard normal distribution, and thus the null hypothesis is accepted iff $|Z_c| < 1.96$.

(5) *Computations*: the number of runs ***R*** for the present problem is $R_c = 7$ counted as follows:

Run	*HHHHH*	*DDD*	*HHHHHHHHH*	*DD*	*HH*	*DDDDD*	*HHHH*
Count	1	2	3	4	5	6	7

Given that $n_1 = 20, n_2 = 10$ then: $E(R) = 14.33,\ \sigma_R = 2.38,$ and consequently $Z = -3.08$.

(6) *Conclusion*: The null hypothesis (H_0) is rejected implying that distribution of pine trees is non-random. There is a real systematic factor to be observed in the distribution of pine trees observed. One such real systematic factor may be presence of a tree disease.

5.2 *The Median Test for Randomness*

In example 5.1.1 above, a decision to designate pine trees as *healthy* (H) or *diseased* (D) depended on the knowledge or experience of the Forestry crew. Thus, the dividing line between attributes (in this case healthy or diseased) was subjective. There are cases when median or some other test-value index is used to classify the two outcomes in a runs test, for instance, below median versus above median. Thus, we may use event **A** for an outcome that is below the median and event **B** for an outcome that is above median. The sequence of these two designated events now becomes the subject of analysis in conducting the randomness testing.

Example 5.1.2

The following data pertain to increase in pulse rate of astronauts before a scheduled flight. The rates are:

26, 20, 17, 22, 23, 21, 25, 30, 14, 12, 21, 19, 25, 14 *and* 18.

Use median-test of randomness to ascertain the claim that the pulse rate of astronauts is random.

Solution 5.1.2

(1) The null hypothesis H_0: *the distribution of pulse rate is random* and the alternative hypothesis is H_a: *the distribution of pulse rate is non-random.*

(2) *Test-statistic* is based on run **R** – the number of changes of pulse rate from below to above median. If H_0 is true and given a large sample, the statistic $Z = \frac{R - E(R)}{\sigma_R}$ has a standard normal distribution.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: Given that **Z** has a standard normal distribution, a decision rule is to: Reject H_0 if $|Z_c| < 1.96$. Alternatively, since $n_1 = 7$, $n_2 = 6$ and considering the probability distribution of **R** in Table 9; $\Pr ob(R \le 4) = 0.043$ (which is less than $\alpha = 0.05$). Thus,

the critical value is $R_0 = 4$. This implies that H_0 is accepted if $R_c > 4$.

(5) *Computations*: Median pulse rate is $Median = 21$. Let **A** represent an event that pulse rate is below the median and **B** represent an event that pulse rate is above the median. The given data on pulse rate are represented in terms of sequence of events **A** and **B** as follows: $BAABBBBBAABABAA$. The number of runs in this sequence of events **A** and **B** is **R = 6** counted as follows:

Run	B	AA	BBBBB	AA	B	A	B	AA
Count	1	2	3	4	5	6	7	8

Given that the number of $A's$ is less than or equal to the number of $B's$, then $n_1 = 7, n_2 = 8,\ E(R) = 8.47$ and $\sigma_R^2 = 3.449$, and therefore, $Z_c = -0.253$. The alternative is to compare the computed value of **R** and its critical value based on probability distribution of **R**. Given the small sample sizes above it is found from Table 9 that $\Pr ob(R \le 4) = 0.043,\ \textit{which is less}\ \alpha = 0.05$; implying that $R_0 = 4$ is the critical value of **R** above which the above null hypothesis is accepted. In the present case $R_c = 8$.

(6) *Conclusion*: Results from sample data indicate that $|Z_c| < 1.96;\ \textit{and}\ R_c > R_0$. Thus, the null hypothesis (H_0) is not rejected; implying that the pulse-rate data are random. It is, therefore, concluded that there is no systematic factor surrounding the pulse rate of the astronauts.

Example 5.1.3

Items emerging from a continuous production process were classified as defective (D) or non-defective (N). A sequence of items observed over time was as follows:

$$DNNNNNDDNN\,NNNNDDDNNN\,NNDNNNDDNNNDD$$

Do these data suggest lack of process control in the production system?

Solution 5.1.3

(1) The null hypothesis H$_0$: *the sequence of items observed over time is random and* the alternative hypothesis is H$_a$: *the sequence is non-random.*

(2) *Test-statistic* is based on run ***R*** – the number of changes of defective/non-defective items. If H$_0$ is true and given a large sample, the statistic $Z = \dfrac{R - E(R)}{\sigma_R}$ has a standard normal distribution, which is the situation in the present case, since $n_1 = 11$, $n_2 = 22$.

(3) Level of significance is $\alpha = 0.05$.

(4) *Decision rule*: Given that ***Z*** has a standard normal distribution, a decision rule is to: Reject H$_0$ if $|Z_c| < 1.96$.

(5) *Computations*: The number of runs in the sequence of events ***D*** and ***N*** is $R = 11$ counted as follows:

$$DNNNNNDDNN\,NNNNDDDNNN\,NNDNNNDDNNNDD$$

Run	*D*	*NNNNN*	*DD*	*NNNNNN*	*DDD*	*NNNN*
Count	1	2	3	4	5	6
Run	*D*	*NNN*	*DD*	*NNN*	*DD*	
Count	7	8	9	10	11	

Given the above number of $D's$ and $N's$:

$$n_1 = 11,\ n_2 = 22,\ E(R) = 15.67,\ V(R) = \sigma_R^2 = 6.2639 \Rightarrow Z_c = -1.865$$

(6) *Conclusion*: Results from sample data indicate that $|Z_c| > 1.96$. Thus, the null hypothesis (H$_0$) is rejected; implying that the sequence of items is not random. It is therefore concluded that there is a systematic factor surrounding the production system – there is lack of control in the production system reported in the study.

Exercise 5.2.1

The accompanying table summarizes monthly average share prices of a stock at the Dar es Salaam Stock Exchange (DSE) for the years 2000 – 2003. To what extent do these data on prices reflect the random walk hypothesis so much laboured in capital market theory?

Years	Month					
	J	F	M	A	M	J
2000	555	555	550	550	575	575
2001	550	560	560	565	585	595
2002	1135	1250	1337	1513	1637	1600
2003	1575	1337	1514	1420	1630	1635
Years	Month					
	J	A	S	O	N	D
2000	636	560	550	550	550	550
2001	595	685	765	780	1000	1063
2002	1475	1337	1425	1600	1688	1625
2003	1600	1650	1615	1640	1650	1670

5.3 *SPSS Tutorial on Randomness Test Technique*

The table below is the SPSS output for randomness test for the examples **5.1.1**, **5.1.2** and **5.1.3** for the variables, ***PINETREE***, ***PULSERATE*** and ***VARITEMS***

Descriptive Statistics

	N	Mean	Std D	Minim	Maxim
PINETREE	30	.6667	.47946	.00	1.00
PULSERATE	15	20.4667	4.98378	12.00	30.00
VARITEMS	33	.6667	.47871	.00	1.00

Runs Test

	PINETREE	PULSERATE	VARITEMS
Test Value [a]	1.0000	21.0000	1.0000
Cases < Test Value	10	7	11
Cases >= Test Value	20	8	22
Total Cases	30	15	33
Number of Runs	**7**	**8**	**11**
Z	**-2.870**	**.000**	**-1.665**
Asymp. Sig. (2-tailed)	.004	1.000	.096

a Median

To produce this output, enter the data on tree-pine distribution as ***PINETREE*** (0 = ***D***iseased, 1 = ***H***ealthy); the pulse-rates of astronauts as ***PULSERATE***; and data on production of items as ***VARITEMS*** (0 = ***D***efective, 1 = ***N***on-defective). Choose the following sequence of commands from the menu:

Analyze

Nonparametric tests

Runs test

Cut Point: **Median** for all variables, ***PINETREE***, ***PULSERATE*** and ***VARITEMS***

Statistics: **Descriptive**

Problem Set Five

Question 5.1

The Random Walk Hypothesis in Financial Economics suggests that the movement of certain security prices is completely random and follows no discernable pattern over time. Listed in the table are daily closing prices over 30 consecutive weeks for a certain security listed on the Dar es Salaam Stock Exchange (DSE). Do these data support the random walk hypothesis?

Week	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Price	255	267	240	235	270	310	305	280
Week	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
Price	290	275	300	310	320	315	305	300
Week	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
Price	295	330	280	285	260	290	300	325
Week	*25*	*26*	*27*	*28*	*29*	*30*		
Price	280	285	275	270	260	290		

Question 5.2

A financial analyst set to investigate the effects of isolated factors that may systematically affect stock prices and, as a result, cause market model residuals to vary in a non-random manner. The observed market model residuals over a twenty-day period are shown in the accompanying table. Do the market model residuals indicate lack of randomness over time and, hence, the presence of some isolated factor that systematically affect stock prices?

Day	1	2	3	4	5	6	7
Residual	-.008	-.008	-.017	-.006	.013	-.019	-.014
Day	8	9	10	11	12	13	14
Residual	.003	-.003	.008	.006	.015	.022	.004
Day	15	16	17	18	19	20	
Residual	.010	-.007	.005	.002	-.012	.003	

Question 5.3

A quality control chart has been maintained for a certain measurable characteristic of items taken from a conveyor belt at a certain point in a production line. The measurements obtained today, in order of time, are:

68.2	71.7	68.3	71.6	70.4	65.0	63.6	64.7	65.3
64.2	67.6	68.6	66.8	68.9	66.8	70.1	69.2	72.1

Do these data suggest lack of stability in the production process?

Question 5.4

It is now over fifteen years since Tanzania economy embraced the financial and capital market operations. Following the rapid advancement of information, communication and technology in most economic sectors, financial experts posit that Dar es Salaam Stock Exchange (DSE) operations are not far from being efficient in the informational sense. To underscore this belief a researcher has conducted a study of stock prices, and the monthly average share prices of a particular stock for the years 2004 – 2007 is summarized in the accompanying table. To what extent do these data on prices reflect the random walk hypothesis so much laboured in capital market theory?

Years	Month					
	J	F	M	A	M	J
2004	555	555	550	550	575	575
2005	550	560	560	565	585	595
2006	1135	1250	1337	1513	1637	1600
2007	1575	1337	1514	1420	1630	1635
Years	**Month**					
	J	A	S	O	N	D
2004	636	560	550	550	550	550
2005	595	685	765	780	1000	1063
2006	1475	1337	1425	1600	1688	1625
2007	1600	1650	1615	1640	1650	1670

6

CORRELATIONAL INFERENCE METHODS

As noted in the introductory remark of this book, correlation techniques enable one to unveil causal relationship between two sets of population attributes, which may be reconstructed as independent on the one hand and dependent on the other. The focus in this book is on correlation indices that are based on nominal and/or ordinal data. Given this broad focus, this chapter presents statistical correlation indices that are computed from *contingency tables or cross-tabulations.* Contingency tables or cross-tabulations are most widely used techniques for summarizing and analysing nominal and/or data in research.

6.1 Measures of Correlation for Cross tabulation

Common statistical measures of association for contingency tables are divided into three categories. First, measures that enable to discern the degree of correlation between two variables, X_i and Y_i. Second are measures that enable to ascertain the strength of a relationship between observed group membership Y_i and expected or predicted group membership $E(Y_i)$. Statistical measures in the first case are *Chi-square-based correlations* and those in the second case are referred to as *proportional reduction in error* and thus *PRE-based correlations.* These PRE-based indices are analogous to coefficient of determination R^2.

A third category of correlation measures include; the Spearman rank correlation that is based on ordinal data, Kendall's tau, Goodman and Kruskal gamma, and Sommer's d_y- coefficients, which are computed on the basis of ordered contingency tables.

A contingency table is a two-way cross-classification of qualitative data. It is a table in which study observations are classified according to two criteria of classification; vis-à-vis criterion A with rows representing its dimensions $A_1, A_2, \cdots, A_r$ and criterion B with dimensions $B_1, B_2, \cdots, B_c$ represented as columns. Dimensions on each factor may *differ in kind* implying that the two factors are nominal scale-measured, thereby leading to **unordered** contingency table. However, if the dimensions *differ in degree* that characterizes ordinal-scaled measurement, an **ordered** contingency table is obtained.

Performing statistical analysis using contingency tables or cross-tabulations often take the following forms.

- Testing homogeneity of proportions while answering the research question: *Are the counts consistent with the hypothesis that true proportion of individuals from group A_1 having characteristics B_1 is the same from the A_2 group?*
- Testing independence in which a general research question posed is: *To what extent are two factors A and B correlated?*
- Testing association that enable to understand the nature and extent of a relationship being studied, while answering a general research question: *How strong is the observed relationship between factors A and B?*

In the first two cases, Chi-square statistic is used to test both homogeneity of proportions and independence of two factors, and hence the use of chi-square-based statistical indices. PRE-based statistical indices are used to test degree of association and/or relationship between two study factors, while symmetric measures of correlation, which are; tau, gamma, and sommer's d, determine the degree of monotonic relationships between two contingency table factors.

6.2 *Chi-square-based Indices of Associations*

Data presented in contingency tables enable researchers to perform descriptive analysis. Apart from performing descriptive analysis, researchers often perform quantitative analysis using contingency or cross-tabulations. This is done by computing correlation coefficients that seek to ascertain the extent to which factors/variables, with ordered or unordered categories, are associated or correlated.

In applied work, Chi-square is commonly used to test homogeneity of proportions or independence between two factors. High values of χ^2 are associated with statistical significance, and this implies rejecting the null hypothesis of homogeneity of proportions or independence and thus accepting an alternative hypothesis. Rejecting the independence proposition in favour of the dependence is indeed accepting that the two factors under consideration are statistically correlated. It is in this respect that χ^2 is a measure of correlation, independence, and goodness-of-fit.

The foundation of Chi-square-based measures of association is distribution theory in statistics, which states that sampling distribution of second moment sample estimators, is a Chi-square. Since correlation in squared-form, just like variance, is a second moment property of estimators, Chi-square is used to ascertain its statistical significance. Given a contingency table with the following characteristics: o_{ij} = observed

frequencies, R_i = row totals, C_j = column totals and e_{ij} = expected frequencies such that: $e_{ij} = \frac{R_i C_j}{GT}$ then the computed Chi-square value is:

$$\chi_c^2 = \sum_i^r \sum_j^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

6.2.1 Testing Homogeneity or Equality of Proportions

The Pearson χ^2 statistic is used to ascertain homogeneity or equality of proportions in a 2×2 contingency table. The basic question posed is: Are observed frequencies or counts consistent with the hypothesis that proportion of individuals from group A_1 having characteristic-property B_1 is the same for those from A_2 group? Thus, if P_1 probability of individual from A_1 having characteristic-property B_1, and P_2 is probability of individual from A_2 having property B_1, then the null hypothesis to tested is $Ho: \mathrm{P}_1 = \mathrm{P}_2 = \mathrm{P}$. If independent samples from A_1 and A_2 are available, such that $P_1 = \frac{o_{11}}{n_1}$, $P_2 = \frac{o_{21}}{n_2}$ and $P = \frac{o_{11} + o_{21}}{n_1 + n_2}$; then the appropriate normal approximation to test *Ho* is: $Z = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \Rightarrow \mathrm{N}(0,1)$.

The alternative to the above homogeneity or equality of proportions test-statistic is the Pearson χ^2, computed as $\chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$.

Equality of proportions using Chi-square measures the relative risk for a prospective research design. However, for a retrospective research design odds ratio approach is used.

Testing equality of proportions also emerges when before-after research designs are used. The purpose of such designs is to test the extent to which there are changes in proportions for a variable considered before and another considered after treatment applications for a 2×2 contingency table as follows.

Variable 1		Variable 2 Yes	No	Total
	Yes	o_{11}	o_{12}	$o_{11} + o_{12}$
	No	o_{21}	o_{22}	$o_{21} + o_{22}$
Total		$o_{11} + o_{21}$	$o_{12} + o_{22}$	N

The proportion of YES on the first variable is $PS_1 = \frac{o_{11} + o_{12}}{N}$ and the corresponding proportion on the second variable is $PS_2 = \frac{o_{11} + o_{21}}{N}$. The null hypothesis being tested is $H_0 : PS_1 = PS_2$ against the alternative hypothesis $H_a : PS_1 \neq PS_2$. If the null hypothesis is true, there is no change in proportions, the test is accomplished by Z- test (with $P_{12} = \frac{o_{11} + o_{12} + o_{11} + o_{21}}{N + N}$) as follows: $Z = \frac{PS_1 - PS_2}{\sqrt{P_{12}(1 - P_{12})\left(\frac{1}{N} + \frac{1}{N}\right)}}$.

Example 6.2.1

To take an example, a study was conducted to determine the extent to which political awareness of ordinary citizens has changed after implementing a political awareness programme aired by the National Broadcasting Corporation (NBC). A question posed to respondents, before and after

implementing the awareness programme, was "***Is political pluralism better for our economic development***?" The accompanying table gives a summary of the "$YES - NO$" responses of 1600 randomly selected people across a multitude of the population. Do these data present sufficient evidence to indicate a changed perception towards political pluralism?

Before political awareness		**After political awareness** Yes	No	Total
	Yes	350	150	500
	No	630	470	1100
Total		980	620	1600

Solution 6.2.1

The proportion of YES on the before-political-awareness and the corresponding proportion of the same on the after-political-awareness are respectively equal to:

$$PS_1 = \frac{o_{11} + o_{12}}{N} = \frac{500}{1600} = 0.3125 \text{ and } PS_2 = \frac{o_{11} + o_{21}}{N} = \frac{980}{1600} = 0.6125.$$

The null hypothesis being tested is $H_0 : PS_1 = PS_2$ against the alternative hypothesis of the form $H_a : PS_1 \neq PS_2$. Given that the overall proportion is $PS_{12} = \frac{350 + 150 + 350 + 630}{1600 + 1600} = 0.4625$, the test-statistic applicable is:

$$Z = \frac{0.6125 - 0.3125}{\sqrt{0.4625(1 - 0.4625)\left(\frac{1}{1600} + \frac{1}{1600}\right)}} = \frac{0.3000}{\sqrt{0.0003107}} = 17.02.$$

This result indicates that the null hypothesis, $H_0 : PS_1 = PS_2$ is rejected and this implies that there is a change in proportions in awareness before and after a political awareness programme. This means that the political awareness programme did influence the understanding and ultimate awareness of ordinary citizens regarding political pluralism reported in the study.

Based on Chi-square test of equal proportions, leads to the following:

$\chi_c^2 = \sum_i^2 \sum_j^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 23.46$, which is larger than $\chi_{.05,1}^2 = 7.88$. This result supports that alternative proposition that there is a change in perception regarding politics and development after the awareness programme.

6.2.2 Testing Independence between Two Factors

In addition to testing equality of proportions, Pearson χ^2 also tests the extent to which two factors displayed in a $r \times c$ contingency table are statistically independent or not. The concept of statistical independence means that probability of realizing one factor does not affect the probability of realizing another factor. In the language of probability theory, A and B are independent if $\Pr ob(A, B) = \Pr ob(A) \Pr ob(B)$. The computed Chi-square under the hypothesis that two given variables or factors are independent is: $\chi_c^2 = \sum_i^r \sum_j^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$.

The Chi-square-based correlation coefficients are considered appropriate than the usual χ^2 for comparing tables with varying sample sizes and dimensions. Since causal association is not implied, the Chi-square based correlations are referred to as **symmetric measures of correlations**. The most common chi-square-based correlation coefficients

are: **phi** (ϕ), **Tschuprows** (T), **Cramer's** (υ), **contingent coefficient** (C), **correlation of attributes** (r), and **Yule's** (Q) coefficient.

Given that $\chi^2 = \sum_i^r \sum_j^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$, the chi-square based correlation coefficients are computed as follows:

- The phi coefficient[13] $\phi = \sqrt{\frac{\chi^2}{N}}$.
- Contingent coefficient $c = \sqrt{\frac{\chi^2}{\chi^2 + N}}$
- Correlation of attributes coefficient[14] $r = \sqrt{\frac{\chi^2}{N(r-1)}}$
- Cramer's $\upsilon = \sqrt{\frac{\chi^2}{N \min(r-1, c-1)}} = \sqrt{\frac{\phi^2}{\min(r-1, c-1)}}$
- Tschprows coefficient $\mathrm{T} = \frac{\chi^2}{N\sqrt{(r-1)(c-1)}} = \frac{\phi^2}{\sqrt{(r-1)(c-1)}}$
- Yule's Q coefficient[15] $Q = \frac{o_{11}o_{22} - o_{12}o_{21}}{o_{11}o_{22} + o_{12}o_{21}}$

The above, chi-square based, correlation coefficients are interpreted as *estimated probabilities* of getting observed frequencies, under the null

[13] For a 2×2 table, this coefficient is equal to

$$\phi = \frac{o_{11}o_{22} - o_{21}o_{12}}{\sqrt{(o_{11} + o_{12})(o_{21} + o_{22})(o_{11} + o_{21})(o_{12} + o_{22})}}.$$

[14] This is applicable for a $r \times r$ contingency table.

[15] This is a special case of Goodman and Kruskal gamma coefficient $\left(\gamma = \frac{C-D}{C+D}\right)$ that is associated with a $r \times c$ contingency table for ordinal data

hypothesis that there are no differences in population proportions. In this respect, Chi-square based correlations have probabilistic interpretations. However, the *exact probability* of obtaining the observed frequencies under the null hypothesis is given by **Fisher's Exact Test**. For instance, the probability of getting exactly a *and* b frequencies in the two cells of the top row is given by:

$$P(a,b) = \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{N}{a+b}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{N!a!b!c!d!}$$

Example 6.2.1

Medical science experts have, for a long time, associated cancer with smoking behaviour. In a health survey study, a researcher has recorded, in a contingency table, the following data on cancer-infected status and smoking behaviour of adult population in a given district (N=1360). Use non-parametric measures of association to determine the relationship between cancer and smoking behaviour.

	Smoking	Non-smoking
Infected cancer	410	230
No- cancer	450	270

Solution 6.2.1

The following table summarizes the data, in a contingency table, required for computing measures of associations (numbers in brackets are expected frequencies - e_{ij}).

$o_{ij}\ (e_{ij})$	Smoking	Non-smoking	Total
Infected cancer	410 (405)	230 (235)	640
No- cancer	450 (455)	270 (265)	720
Total	860	500	1360

Given $N = 1360$ and $\chi^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = .317$, then the Chi-square based correlation coefficients are:

- $\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{.317}{1360}} = 0.01527$
- $C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{.317}{.317 + 1360}} = 0.015267$
- $r = \sqrt{\frac{\chi^2}{N(r-1)}} = \sqrt{\frac{.317}{1360(2-1)}} = 0.0152654$
- $\upsilon = \sqrt{\frac{\chi^2}{N\min(r-1,c-1)}} = \sqrt{\frac{.317}{1360 \times 1}} = 0.0152654$
- $\mathrm{T} = \frac{\chi^2}{N\sqrt{(r-1)(c-1)}} = \frac{.317}{1360 \times \sqrt{(1)(1)}} = 0.000233$
- $Q = \frac{o_{11}o_{22} - o_{12}o_{21}}{o_{11}o_{22} + o_{12}o_{21}} = \frac{410 \times 270 - 450 \times 230}{410 \times 270 + 450 \times 230} = 0.0336$

The results of the above measures of association indicate that there is insignificant substantive correlation between cancer and smoking behaviour in the studied population. In probability terms, these results imply that there is a small chance or probability that smoking and cancer can occur jointly. Differently put, the probability that smoking does lead to cancer is very small. It is unlikely that cancerous patients are necessarily smokers.

However, evaluating the *exact probability* of obtaining the observed frequencies under the null hypothesis using **Fisher's Exact Test** has arithmetic intricacies, particularly with large observed frequencies.

Exercise 6.2.1

The following table summarizes data on subjects who were not using Insecticides Treated Nets (ITN) and therefore exposed to malaria attack, and those who used ITN and thus not exposed to the deadly tropical disease, under The Kilimanjaro Intermittent Preventive Treatment of Malaria in Infants (KILI-IPTi) Project. These subjects, chosen from high malaria transmission area, were classified as being **diseased** if they contracted malaria or **not diseased** if they did not contract the disease over the project period.

	Diseased	Not diseased	Total
Exposed (No ITN use)	305	294	599
Not Exposed (ITN use)	98	583	681

Source: KILI-IPTi Study Reports, Same District. TANZANIA

Use Chi-square based techniques to test *equality of proportions* or *independence* between **ITN use** and **malaria incidence** for the subjects reported in the KILI IPTi Project.

Exercise 6.2.2

A business firm wishes to determine whether quality of its product is the same at all the five of its production units. The results of 100 items from each line that were tested for quality are shown in the accompanying table.

	Production unit				
Quality	1	2	3	4	5
Satisfactory	80	93	89	92	81
Below-quality	20	7	11	8	19
Totals	100	100	100	100	100

Test the claim that the proportions of satisfactory quality items produced by each of the five production units are equal.

6.3 *PRE-based measures of associations*

As noted above, PRE-based measures of association enable to ascertain the strength of a relationship between observed group-membership Y_i and expected or predicted group membership $E(Y_i)$. It is about improvement of a model in predicting a given factor, considered as dependent, upon knowledge of the other, considered as independent. If improvement is possible, there is a proportional reduction in error. Thus, PRE-based correlation indices enable to determine a gain in predicting one categorical table factor when value of another factor is known or made to be known relative to when it is unknown. In that respect, PRE-based coefficients are **directional measures of correlation** as they indicate direction of one factor upon knowing another.

PRE-based correlation coefficients measure degree of accuracy of prediction in **predictive, classificatory** or **selective** models. In **prediction** models cases may be homogeneous; and an attempt is made to predict or classify cases on the basis of knowledge of some criterion factor. For instance, predicting political orientation with a stated proportional reduction in error on the basis of knowledge of ethnicity. Given a population of homogeneous cases, identical treatment is permissible in predictive research designs. **Probability models** such as **probit**, **normit** and **logistic regression models** are the typical predictive research models.

In **classification** models cases are truly heterogeneous so that a model has to classify as many cases into each category as are actually observed. Put differently, the number of cases observed to be in each category should be the same as the number of cases predicted to be in each category. Identical treatment is not a viable option in classification research designs. Typical classificatory models in social science research include **discriminant analysis models**.

Cases in **selection** models are "accepted" or "rejected" for inclusion in a group based both on whether they satisfy some criterion for success in the group and on the minimum required, maximum allowable, or specified number of cases that may (or must) be included in the group. **Factor, cluster and multi-dimensional scaling analysis models** are the typical selective research models in social sciences.

There is a link between the conventional R^2 and PRE-based correlation coefficients. By definition, R^2 is explained variation, which measures the proportion or percentage by which use of regression equation reduces the error of prediction relative to predicting using mean $(\bar{Y})$. As such, R^2 is a PRE-based statistic and on account of this, PRE-based correlation coefficients from contingency tables are *pseudo-coefficients of determination*.

The common PRE-based measures of association are **Goodman and Kruskal lambda-p** (λ_p) for prediction models, **Goodman and Kruskal tau** (τ_p) for classification models, and **Kendall's phi-coefficient** (ϕ_p) for

selection models[16]. A proportional change in error indicator is computed using the following expression:

$$Predictive\,efficiency = \frac{Errors\ without\ model - Errors\ with\ model}{Errors\ without\ model}$$

The errors with model are simply the cases for which predicted value of the dependent variable is incorrect when a model is used. The number of such errors is analogous to the error sum of squares due to treatment in regression analysis. Furthermore, the errors without model are the cases misclassified when no model is used, and the number of these errors is analogous to total sum of squares. The number of errors without model will differ depending on whether prediction, classification, or selection model is used.

Improved prediction of a dependent variable based on a model implies that there is ***proportional reduction in error*** or PRE. If the model does worse in predicting value of the dependent variable, the predictive efficiency is negative, in which case, there is a ***proportional increase in error***. Thus, PRE-based correlation coefficients are interpreted as the proportion of expected errors reduced by a predictive, classificatory or selective model. Put differently, PRE-based correlation coefficients indicate the proportion by which a model improves prediction. Both cases imply that there is a correlation between the two contingency table factors under consideration.

In literal terms, predictive efficiency of a model is about reducing the number of expected errors when using a model. A predictive efficiency of zero indicates that the model enables to reduce all the expected errors and this imply that all cases are correctly predicted, classified or selected.

[16] Other PRE-based statistical indices of predictive efficiency are Light-Margolin measure, Cross-product ratio (θ), and kappa (κ) and eta (η) coefficients. The first three are presented in the next section.

Ordinarily, a PRE-based coefficient of 0.50 imply that the model reduces the number of expected errors to half, while a coefficient of 0.75 means that number of errors are reduced to one-fourth of the expected. Correspondingly, a coefficient equal to unity indicates that no errors are reduced by having knowledge of the factor or by using a model and in that case the model is useless.

PRE-based coefficients are analogous to R^2 as a measure of *substantive significance*. As a rule of thumb, substantive significance is associated with R^2 being at least 0.30, while $p \leq 0.0001$ indicates that the coefficient of determination is statistically significant. For a *statistical significance* of the pseudo R^2, an analogue to F statistic is the normal approximation to binomial test, which is computed as:

$$d = \frac{(P_e - p_e)}{\sqrt{\frac{P_e(1-P_e)}{N}}}$$

Where; $P_e = \frac{errors\ without\ model}{N}$ and $p_e = \frac{errors\ with\ model}{N}$

The d-statistic tests the proportion of cases that are correctly or incorrectly classified by the model. It indicates whether the proportion incorrectly predicted with model (which is by assumption, dependent on the model, and thus variable) differs significantly from the proportion incorrectly predicted without the model (which is dependent only on the marginal distributions, not on a model, and thus assumed fixed). A final word on the relationship between goodness-of-fit and predictive efficiency indices is that the latter provide quantitative estimate of how well the cases are classified by the model. They indicate the extent to which an independent variable

allows to classify cases into categories of a dependent variable with a *high degree of accuracy*.

Researchers, however, are more often interested in *good-ness-of-fit of a model* $(\chi^2\ or\ R^2)$ than in *accuracy* of prediction, classification or selection of a model as indicated by $\lambda_p, \tau_p, or\ \phi_p$. This interest by researchers is no accident especially for theory-testing in which goodness-of-fit is more important than accuracy of prediction, classification, or selection. Certainly *accuracy* and therefore, *predictive, classificatory, or selective efficiency of a model is very pertinent in theory-building*, which mainly uses qualitative research data.

Thus, the basic and fundamental research questions in predictive, classificatory or selective design models are: *Does it pay to have a predictive, classificatory or selection model when predicting a variable on the basis of knowledge of a given variable considered as independent? What is the proportional reduction in error when a model (predictive, classificatory or selection factor) is used to predict a dependent factor? How much does a predictive, classificatory, or selective model improve prediction of the dependent factor based on knowledge of an independent factor?*

6.3.1 Goodman and Kruskal lambda index for prediction models

This statistical index is used to test the degree of association between two factors in a $r \times c$ contingency table. The number of errors without model is computed on the basis of the *mode of dependent* variable or factor as the predicted value for all cases. A value of $\lambda_p = 1$ indicates that all cases are correctly classified; and positive values have the same interpretation of R^2; meaning the percentage of cases correctly classified. The mode of independent factor $(MODIV)$ is defined as the sum of modal-row frequency, and the mode of the dependent factor $(MODDV)$ is the modal column-total or max (column-total). Thus, the lambda-p coefficient is computed as follows:

$$\lambda_p = \frac{MODIV - MODDV}{Total\ frequency - MODDV}$$

To ascertain the statistical significance of the lambda-p coefficient, the following computations are considered. First, to compute the variance of lambda-p, $\sigma^2(\lambda_p)$, and second is to determine the $(1-\alpha)\%$ confidence interval, which is $\lambda_p \pm Z_{1-\frac{\alpha}{2}}\sigma(\lambda_p)$. The variance of the lambda-p is:

$$\sigma^2(\lambda_p) = \frac{(N - MODIV)(MODIV + MODDV - 2\times \max(DVColFrq))}{(N - MODDV)^3}$$

Given that: $N = total\ frequency$; $o_{ij} = frequency$; $R_i = rowtotal$; $C_j = column\,total$; $o_{im} = \max(o_{ij})\ \forall i$ and $o_{jm} = \max(o_{ij})\ \forall_j$, then the general computational procedure of the Goodman and Kruskal lambda is given by the following expression:

- $$\lambda_p = \frac{\sum_{i=1}^{r} o_{im} - \max(C_j)}{N - \max(C_j)}$$

- $$\sigma^2(\lambda_p) = \frac{\left(N - \sum_{i=1}^{r} o_{im}\right)\left(\sum_{i=1}^{r} o_{im} + \max(C_j) - 2o_{jm}\right)}{(N - \max(C_j))^3}$$

The following example from marketing field illustrates the use of Goodman and Kruskal lambda index. A market survey study is investigating the extent to which there is a relationship between age of a buyer and choice of option package of a new product. Below is a summary of the observed number of respondents (o_{ij}) from the survey.

Age	**Option package** A	B	C	Total
Under 30	200	180	325	705
30 or above	255	320	230	805
Total	455	500	555	1510

The problem at hand is about testing a null hypothesis: $Ho: \lambda_p = 0$; against the alternative, $Ha: \lambda_p > 0$. Based on the above data, the following computations are in order $(\alpha = 0.05)$[17]:

- $$\lambda_p = \frac{\sum_{i=1}^{r} o_{im} - \max(C_j)}{N - \max(C_j)} = \frac{325 + 320 - 555}{1510 - 555} = \frac{90}{955} = 0.094$$

- $$\sigma^2(\lambda_p) = \frac{\left(N - \sum_{i=1}^{r} o_{im}\right)\left(\sum_{i=1}^{r} o_{im} + \max(C_j) - 2o_{jm}\right)}{(N - \max(C_j))^3}$$

$$= \frac{[1510 - (325 + 320)][(325 + 320) + 555 - 2\times 325]}{(1510 - 555)^3} = 0.000546$$

[17] Note that mode of the dependent variable/factor (choice of option package in this case), is $\max(C_j) = \max(455,\ 500,\ 555) = 555$

- The 95% confidence interval for the Goodman and Kruskal lambda is: $\lambda_p \pm Z_{1-\frac{\alpha}{2}}\sigma(\lambda_p) \Rightarrow 0.094 \pm 1.96 \times \sqrt{0.000546}$.
- The 95% confidence interval for λ_p is $(0.0482,\ 0.1398)$, which does not include the zero value specified in the null hypothesis, and on account of this result, the null hypothesis is rejected.
- **Conclusion**: The result indicates that the proportional reduction in error in predicting choice of package option based on knowledge of age is 9.4%. This result provides sufficient statistical evidence that there is a correlation between age and choice of package reported in this study. However, the proportion reduction in error in predicting option package from knowledge of buyer-age is small $(\lambda_p = 0.094 < 0.30)$ and thus considered substantively insignificant.

Note that the alternative test of statistical significance of λ_p is ascertained by computing a $d-statistic$, which is analogous to F-test for R^2, as follows: $P_e = \frac{645}{1510} = 0.43$; $p_e = \frac{555}{1510} = 0.38$ and hence $d_\lambda = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 3.92$. Thus, although, the value of $\lambda_p = 0.094$ is substantively insignificant it is however statistically significant with a probability value of $p < 0.05$.

6.3.2 The Goodman and Kruskal tau index for classification models

As a PRE-based index Goodman and Kruskal tau (τ_p) index enables to evaluate the extent to which it is possible to predict one factor upon knowledge of the other. The index has a probabilistic interpretation and the basis of prediction is number of errors without model versus number of errors with model.

In literal terms, tau-p is about determining the number of errors reduced when using a classificatory model to make prediction. A value of zero for tau-p $(\tau_p = 0)$ indicates that the model helps to reduce all the expected errors and this implies that all cases are correctly classified. Ordinarily, $\tau_p = 0.50$ implies that the model cuts the number of errors to half, while $\tau_p = 0.75$ means that number of errors are reduced to one-fourth of the expected. Correspondingly, $\tau_p = 1.00$ indicate that no errors are reduced and in that case the model is useless. As noted above, the Goodman and Kruskal tau index is computed as:

$$\tau_p = \frac{Errors\ without\ model - Errors\ with\ model}{Errors\ without\ model}$$

In applied work, if $R_i = row\,total$ representing the number of cases observed in category i of the independent factor, $C_j = column\,total$ for the dependent factor and o_{ij} is the frequency of cell C_{ij}, then:

- The number of **errors without model** is $\varepsilon = \sum_{j=1}^{c} C_j\left(\frac{N - C_j}{N}\right)$
- The number of **errors with model** is $\delta = \sum_{i=1}^{r}\sum_{i=1}^{c} o_{ij}\left(\frac{R_i - o_{ij}}{R_i}\right)$.

Hence, Kendall's tau-p is computed as follows:

$$\tau_p = \frac{\sum_{j=1}^{c} C_j\left(\frac{N-C_j}{N}\right) - \sum_{i=1}^{r}\sum_{j=1}^{c} o_{ij}\left(\frac{R_i - o_{ij}}{R_i}\right)}{\sum_{i=1}^{c} C_j\left(\frac{N-C_j}{N}\right)}$$

To take an example, suppose that it is desired to evaluate the extent to which knowledge of religion (independent factor - row) allows to predict political orientation (dependent factor - column), based on data summarized in the table below.

Religion	Political orientation				
	o_{ij}	Conservatism	Liberal	Cosmopolitan	Total
	Christianity	300	600	300	1200
	Judaism	600	100	100	800
	Total	900	700	400	2000

The number of errors without model and the number of errors with model are respectively equal to 1270 and 1075 computed as follows:

$$\varepsilon = \sum_{j=1}^{c} C_j\left(\frac{N-C_j}{N}\right)$$

$$= 900\left(\frac{2000-900}{2000}\right) + 700\left(\frac{2000-700}{2000}\right) + 400\left(\frac{2000-400}{2000}\right) = 1270$$

$$\delta = \sum_{i=1}^{r}\sum_{j=1}^{c} o_{ij}\left(\frac{R_i - o_{ij}}{R_i}\right)$$

$$= 300\left(\frac{1200-300}{1200}\right) + 600\left(\frac{1200-600}{1200}\right) + 300\left(\frac{1200-300}{1200}\right)$$

$$+ 600\left(\frac{800-600}{800}\right) + 100\left(\frac{800-100}{800}\right) + 100\left(\frac{800-100}{800}\right) = 1075$$

The Goodman and Kruskal tau is $\tau_p = \frac{\varepsilon - \delta}{\varepsilon} = \frac{1270-1075}{1270} = 0.154$. This result indicates that we have saved ourselves 195 errors out of the expected 1270, and thus we have reduced our errors by 15.4%. The proportion of cases that are correctly classified by the model has increased by 15.4%, indicating that knowledge of **religion** helps to predict **political orientation** by about 15.4%.

To ascertain statistical significance of this pseudo coefficient of determination $d-statistic$ is used. The computations for the statistic are as follows: $P_e = \frac{1270}{2000} = 0.6350;\ p_e = \frac{1075}{2000} = 0.5375$ and hence,

$$d_\tau = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 9.06$$ and its probability value is $p < 0.05$.

6.3.3 The Kendall's phi coefficient for selection models

The Kendall's phi coefficient[18] is appropriately applied with selection research design models for a 2×2 contingency table. *The number of errors without model is the sum of expected frequencies in cells* C_{12} *and* C_{21},

[18] Note the difference between ϕ as a measure of symmetric measure of correlation and ϕ_p as a predictive or directional measure of correlation

while the number of errors with model is the sum of observed frequencies in the two cells. Thus, Kendall's phi coefficient is equal to:

$$\phi_p = \frac{\frac{(o_{11}+o_{12})(o_{12}+o_{22})}{N} + \frac{(o_{11}+o_{21})(o_{21}+o_{22})}{N} - (o_{12}+o_{21})}{\frac{(o_{11}+o_{12})(o_{12}+o_{22})}{N} + \frac{(o_{11}+o_{21})(o_{21}+o_{22})}{N}}$$

$$\phi_p = \frac{o_{11}o_{22} - o_{12}o_{21}}{\frac{1}{2}[(o_{11}+o_{12})(o_{12}+o_{22}) + (o_{11}+o_{21})(o_{21}+o_{22})]}$$

Higher orders of contingency tables may be used in computing Kendall's phi coefficient provided that the definitions of the number of errors with and without model are extended to include all off-diagonal cells. A value of unity for ϕ_p indicates that all cases are correctly predicted or classified given the specified attributes.

Example 6.3.1

A bank lending system uses discriminant model to assess its borrowers and thus classify them as either potential defaulters or non-defaulters. The following is a summary, in a contingency table, of performance of the bank-lending assessment model in terms of actual versus predicted number of non-defaulters (ND) and defaulters (D).

o_{ij}		Predicted		Total
		ND	D	
Actual	ND	5	4	9
	D	7	9	16
Total		12	13	25

Use appropriate cross-tabulation technique to evaluate the predictive efficiency of the bank-lending assessment system.

Solution 6.3.1

The problem above is a classic application of PRE-based correlation technique, in which it is desired to evaluate the strength of relationship between observed and predicted group membership of bank loan applicants. Specifically, it is desired to evaluate the extent to which discriminant model is able to discriminate non-defaulters (ND) and defaulters (D). This in essence is a selection research designed model. The essential computations for a 2×2 contingency table are as follows:

The essential computations for computing the Kendall's phi coefficient (ϕ_p), using data on the performance of bank lending model, are given below (numbers in brackets are expected frequencies).

o_{ij}		Predicted		Total
		ND	D	
Actual	ND	5(4.32)	4(4.68)	9
	D	7 (7.68)	9 (8.32)	16
Total		12	13	25

Thus, the proportion in error reduction when bank lending model is used to predict loan repayment performance is given by:

$$\phi_p = \frac{Expected\,errors - observed\,errors}{Expected\,errors} = \frac{(4.68+7.68)-(4+7)}{(4.68+7.68)} = 0.11.$$

This result indicates that the bank lending model improves prediction of loan repayment performance by 11%.

The significance of this pseudo-coefficient of determination is ascertained using $d - statistic$. Given that $P_e = \frac{12.36}{25} = 0.4944$ and $p_e = \frac{11}{25} = 0.4400$, then $d_\phi = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 0.544$ with a probability value of $p < 0.05$. It is thus concluded that although the bank lending model improves prediction by about 11%, statistically it has limited predictive or classification efficiency.

However, based on a classificatory model, numbers of errors with and without model are respectively equal to:

$$\delta = 5\left(\tfrac{12-5}{12}\right)+7\left(\tfrac{12-7}{12}\right)+4\left(\tfrac{13-4}{13}\right)+9\left(\tfrac{13-9}{13}\right)=11.371 \text{ and;}$$

$$\varepsilon = 9\left(\tfrac{25-9}{25}\right)+16\left(\tfrac{25-16}{25}\right)=11.520\text{, so that } \tau_p = \tfrac{\varepsilon-\delta}{\varepsilon} = 0.0129.$$

The value of tau-p, being less than 0.30, is substantively insignificant. The d-statistic is used to ascertain its statistical significance. This is done by computing proportion of errors with and without model as follows. $P_e = \frac{11.520}{25} = 0.46$; $p_e = \frac{11.371}{25} = 0.458$. Therefore, the test-statistic of τ_p is

$$d_\tau = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 0.02$$; whose probability value is $p < 0.05$. The conclusion is that the lending model is statistically inadequate in classifying default and non-default in this particular reported case.

Exercise 6.3.1

To what extent is it possible to predict political orientation based on knowledge of ethnicity as summarized in the table below? Use $\alpha = 0.05$.

o_{ij}	African	Non-African
Conservative	20	30
Radical	50	70

Exercise 6.3.2

MBA graduates were cross-classified on the basis of undergraduate liberal arts major, humanities versus social/natural sciences, and their orientation of the graduate programme, quantitative versus qualitative.

Undergraduate major	MBA Programme orientation	
	Quantitative	Qualitative
Social/natural	36	14
Humanities	12	18

Test the claim that nature of undergraduate major is independent of the selected MBA Programme orientation (use $\alpha = 0.05$).

Exercise 6.3.3

Basing on data reported in **Exercise 6.2.1**, use PRE-based indices of your choice to check efficiency of Insecticides Treated Nets in the prevention of malaria in infants in higher transmission area

Exercise 6.3.4

A large corporation was interested in determining whether there is an association between commuting time of employees and stress-related problems observed on the job. The study was restricted to assembly line employees who work an 8-hour shift of mostly repetitive tasks. Below is a summary of the survey of 120.

Commuting time	Stress-related	
	Observed	Not observed
Under 15 minutes	12	21
15 – 45 minutes	20	33
Over 45 minutes	24	10

Use PRE-based index of your choice to predict stress based on commuting time.

6.4 *Other PRE-based Coefficients*

In addition to the three common indices of predictive efficiency, *Light-Margolin measure, cross product-ratio*, and *kappa index* are also being used to evaluate predictive efficiency of predictive, classificatory, or selective designed models.

6.4.1 The Light-Margolin measure

This PRE-based statistical index is a pseudo R^2, which determines the extent to which variation of frequency in one factor considered as dependent (column) is explained by the other considered as independent (row). The computational procedure has the same structure as with linear regression analysis. The numerator is synonymous to the sum of squares due to regression (SSR), and the denominator resembles the total sum of squares (SST).

In computing the Light-Margolin measure, however, the following definitions are adopted. SST = total frequency *minus* average of squared column totals; SSR = sum of average of row-squared frequency; R_i = row-total; C_j = column-total; and N = total frequency. The Light-Margolin statistical index is computed with the following formula:

$$R^2_{psdo} = \frac{SSR}{SST} = \frac{\left(\sum_{i=1}^{r} \frac{1}{R_i} \sum_{j=}^{c} o_{ij}^2\right) - \frac{1}{N}\sum_{j=1}^{c} C_j^2}{N - \frac{1}{N}\sum_{j=1}^{c} C_j^2}$$

To take an example, a summary of the observed number of respondents (o_{ij}) from the survey of age and choice package option considered is given below.

Age	Option package			Total
	A	B	C	
Under 30	200	180	325	705
30 or above	255	320	230	805
Total	455	500	555	1510

Is variation of the choice of package explained by the age factor? The Light-Margolin measure is used to answer this question as follows:

$$R^2_{psdo} = \frac{\left(\sum_{i=1}^{r} \frac{1}{R_i} \sum_{j=1}^{c} o_{ij}^2\right) - \frac{1}{N}\sum_{j=1}^{c} C_j^2}{N - \frac{1}{N}\sum_{j=1}^{c} C_j^2}$$

$$\sum_{i=1}^{2} \frac{1}{R_i} \sum_{j=1}^{3} o_{ij}^2 = \frac{\left(200^2 + 180^2 + 325^2\right)}{705} + \frac{\left(255^2 + 320^2 + 230^2\right)}{805}$$

$$= 253 + 274 = 527$$

$$\frac{1}{N}\sum_{j=1}^{3} C_j^2 = \frac{455^2 + 500^2 + 555^2}{1510} = 507$$

$$\therefore \quad R^2_{psdo} = \frac{527 - 507}{1510 - 507} = \frac{20}{1003} = 0.01994$$

The results indicate that only about 1.99% of the variation in choice of package is explained by the age factor, a result that is hardly considered as being substantively significant. It is thus concluded that age factor does not determine choice of option package for the respondents reported in this study. The significance of this pseudo coefficient of determination is ascertained using $d - statistic$ as follows:

$P_e = \frac{527}{1510} = 0.349$, $p_e = \frac{507}{1510} = 0.336$; hence, $d = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 1.06$.

The probability value, corresponding to the $d-statistic$, is $p>0.05$. It is thus concluded that the knowledge of buyer-age has limited predictive efficiency regarding choice of option package.

6.4.2 Cross-product ratio

The cross-product ratio (or ratio of odds) is a PRE-based statistical index that is used to explore the nature and extent of a relationship between two factors in a 2×2 contingency table. This statistical index is appropriate for a retrospective and/or cross-sectional study designs[19]. The design **starts** with **an event** and looks backwards for a **factor** of interest that is purported to have correlation with the event. The purpose is to determine the extent to which there is association between the factor and the event. The research question is *"Does knowledge of a factor help to predict an event?"* This design is different from the prospective study design that **starts with a factor** and follows it, in order to determine its association with a given **event**. The corresponding research question in a prospective study design is *"Does knowledge of an event help to predict a factor?"*

The cross-product ratio is computed as $\theta=\frac{o_{11}o_{22}}{o_{12}o_{21}}$ and its variance in logarithmic form is $\sigma^2(\ell n\theta)=\frac{1}{\sum\sum o_{ij}}=\frac{1}{o_{11}+o_{12}+o_{21}+o_{22}}$. Given the null hypothesis that $Ho:\ell n\theta=0$ and $\alpha=0.05$, there is sufficient evidence that knowledge of an event helps in predicting a factor if the confidence interval $\ell n\theta\pm Z_{1-\frac{\alpha}{2}}\sigma(\ell n\theta)$ does not cover the number zero,

[19] The cross-product ratio is the ratio of odds $\left(\theta=\left(\frac{o_{11}}{o_{12}}\div\frac{o_{21}}{o_{22}}\right)\right)$ and the relative risk is measured by the ratio of incidence rates of an event $\left(risk=\left(\frac{o_{11}}{o_{11}+o_{12}}\right)\div\left(\frac{o_{21}}{o_{21}+o_{22}}\right)\right)$

implied in the null hypothesis. This literally means a factor can be predicted upon knowledge of another factor called the event.

To give an example, suppose that opinions were sought, from both rural and urban citizens, regarding introduction of a national policy strategy on poverty reduction. A study of the type was considered appropriate given that there is a marked difference in the incidence of poverty between urban and rural areas. Accordingly, such difference is thought a priori to translate itself into difference in opinions between urban and rural population regarding poverty reduction strategies. The results of an opinion poll, in terms of the number of respondents (o_{ij}), are summarized in a contingency table below.

	Domicile		
Attitude	Urban	Rural	Total
Favour	672	588	1260
Oppose	928	612	1540
Total	1600	1200	2800

The problem at hand is about testing the extent to which attitude of citizenry regarding incidence of poverty and its reduction strategy is correlated with domicile. More formally, it is about answering a research question: "*Who are likely to support (oppose) poverty reduction strategy policy, urban or rural population?*" Operationally, it is about testing a null hypothesis $Ho:\ell n\theta=0$; against the alternative hypothesis $Ha:\ell n\theta>0$. Based on the data above on domicile and attitude of citizenry towards poverty reduction strategy, the following computations and conclusions are made $(\alpha=0.05)$:

- $\theta=\frac{o_{11}o_{22}}{o_{12}o_{21}}=\frac{672\times612}{588\times928}=0.7537$

- $\sigma^2(\ell n\,\theta) = \frac{1}{\sum\sum o_{ij}} = \frac{1}{o_{11}+o_{12}+o_{21}+o_{22}} = \frac{1}{2800} = 0.000357$
- $\ell n\,\theta \pm Z_{1-\frac{\alpha}{2}}\sigma(\ell n\,\theta) \Rightarrow -0.2828 \pm 1.96 \times \sqrt{0.000357}$
- The 95% confidence interval for θ is $(-0.3198, -0.2458)$ and this does not include the number zero specified in the null hypothesis or $Ho: \ell n\,\theta = 0$.
- **Conclusion**: The results provide sufficient evidence that there is a *negative* correlation between domicile citizenry and attitude on the poverty reduction strategy for the respondents reported in this study. It is thus concluded that attitude can be predicted on the basis of knowledge of domicile.

6.4.3 Kappa index

The kappa index (κ) ascertains the extent to which multiple ratters display perfect agreement or disagreement. The research question posed is: *Are those receiving the same rating from multiple ratters differing from the expected by chance alone?* Given that; $P =$ observed (expected) proportions in the main diagonal cells, the computation of the index is:

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{\sum o_{ii} - \sum e_{ii}}{\sum\sum o_{ij} - \sum e_{ii}}$$

As noted earlier, the directional coefficients range from 0 to 1. The interpretations of the coefficients differ from one another. For instance, values of kappa (κ) greater than 0.75 indicate that there is excellent agreement beyond chance, and values less than 0.40 indicate insignificant substantive agreement. Thus, as a PRE-based correlation coefficient, kappa indicates the extent to which a model, based on several attributes acting as ratters, is able to correctly or incorrectly classify cases or objects.

Examples 6.3.1

Use correlational techniques that are based on Goodman and Kruskal lambda, cross product ratio, and kappa index to evaluate the predictive efficiency of the bank-lending model presented in *Example 6.2.1*.

Solution 6.4.1

(a) **Goodman and Kruskal lambda index**

The computations for the Goodman and Kruskal lambda index are as follows.

- $\lambda_p = \frac{\sum_{i=1}^{2} o_{im} - \max(C_j)}{N - \max(C_j)} = \frac{5+9-13}{25-13} = \frac{1}{12} = 0.083$
- $\sigma^2(\lambda_p) = \frac{[25-(5+9)][(5+9)+13-2\times 9]}{(25-13)^3}$
  $= \frac{11\times 9}{12^3} = 0.005635;\ \sigma(\lambda_p) = 0.0751$
- The 95% confidence interval for the Goodman and Kruskal lambda is: $\lambda_p \pm Z_{1-\frac{\alpha}{2}}\sigma(\lambda_p) \Rightarrow 0.083 \pm 1.96 \times \sqrt{0.0751}$

- The 95% confidence interval is $(-0.0642,\ 0.2302)$ which includes zero specified by an implied null hypothesis $H_0 : \lambda_p = 0$.

- **Conclusion**: The results indicate that bank-lending model reduces errors in predicting default or non-default by 8.3%. In this respect, the bank lending model improves prediction. It is however concluded that the proportion by which bank lending model reduces errors in predicting defaulters or non-defaulters is statistically insignificant.

(b) Cross-product-ratio

The steps involved are as follows.

- $\theta = \dfrac{o_{11}o_{22}}{o_{21}o_{12}} = \dfrac{5\times 9}{7\times 4} = 1.6071;\quad \ell n(\theta) = 0.4744$; and

- $\sigma^2(\ell n\theta) = \dfrac{1}{\sum\sum o_{ij}} = \dfrac{1}{25} = 0.04 \Rightarrow \sigma(\ell n\theta) = 0.20$.

- Consequently, the 95% confidence interval for the cross-product ratio is $\ell n\theta \pm Z_{1-\frac{\alpha}{2}}\sigma(\ell n\theta) \Rightarrow 0.4744 \pm 1.96\times\sqrt{0.04}$. The number zero specified in the implied null hypothesis of $Ho : \ell n\theta = 0$, lies outside the 95% confidence interval of $(0.0824,\ 0.8664)$.

- The conclusion is the bank-lending system model is able to discriminate between non-defaulters and the defaulters.

Example 6.4.2

Use cross product ratio to ascertain the correlation between cancer infection and smoking behaviour given a survey data summarized in the table below?

o_{ij}	Smoking	Non-smoking	Total
Infected cancer	410	230	640
No- cancer	450	270	720
Total	860	500	1360

Solution 6.4.2

Cross-product is used given that it is a 2×2 contingency table

- $\theta = \dfrac{o_{11}o_{22}}{o_{21}o_{12}} = \dfrac{410\times 270}{230\times 860} = 1.0696;\qquad \ell n(\theta) = 0.0673$; and

- $\sigma^2(\ell n\theta) = \dfrac{1}{\sum\sum o_{ij}} = \dfrac{1}{1360} = 0.000735$.

- The 95% confidence interval for the cross-product ratio in log-form is $\ell n\theta \pm Z_{1-\frac{\alpha}{2}}\sigma(\ell n\theta) \Rightarrow 0.0673 \pm 1.96\times\sqrt{0.000735}$, leading to the interval of $(0.0142,\ 0.1204)$. This interval does not include the number zero specified in an implied null hypothesis $Ho : \ell n\theta = 0$.

- The conclusion is that there is a correlation between cancer infection and smoking behaviour for the respondents in reported in the study.

Light-Margolin measure may be used to answer the implied research question: what proportion of cancerous disease is explained by smoking behaviour? In this case, rows represent independent factor and columns represent the dependent factor in the contingency table. The result of computations for the Light-Margolin measure is:

$$R^2_{psdo} = \frac{\left(\sum_{i=1}^{r}\frac{1}{R_i}\sum_{j=}^{c}o_{ij}^2\right)-\frac{1}{N}\sum_{j=1}^{2}C_j^2}{N-\frac{1}{N}\sum_{j=1}^{2}C_j^2}$$

$$= \frac{\frac{(410^2+450^2)}{860}+\frac{(230^2+270^2)}{500}-\frac{(640^2+720^2)}{1360}}{1360-\frac{(640^2+720^2)}{1360}}$$

$$R^2_{psdo} = \frac{431+252-682}{1360-682} = \frac{1}{678} = 0.001475$$

The result indicates that only about 0.15% of cancerous disease is explained and/or is better predicted by incidence of smoking behaviour. Given that

$P_e = \frac{683}{1360} = 0.502;\ p_e = \frac{682}{1360} = 0.501$ and hence, $d = \frac{P_e - p_e}{\sqrt{\frac{P_e(1-P_e)}{N}}} = 0.074$

and its probability value is $p > 0.05$. It is, therefore, concluded that the Light-Margolin measure of 0.15% is considered to be insignificant both substantively and statistically.

Exercise 6.4.1

Use Cross-product ratio technique to check the claims that there is association between: smoking behaviour and cancer as described in **example 6.2.1**; political orientation and ethnicity described in **exercise 6.3.1**; undergraduate major and MBA programme orientation described in **exercise 6.3.2** $(\alpha = 0.05)$.

Exercise 6.4.2

Use Light and Margolin R^2 to evaluate the extent to which it is possible to predict dependent factor upon knowledge of the independent factor using data presented in **exercises 6.3.1 to 6.3.4** above.

Exercise 6.4.3

A large consumer testing organization is interested in evaluating the durability of different brands of batteries used in autos. Three different brands of batteries are tested: Yuasa, GS, and Chloride-oxide batteries. In addition to studying differences among the battery-brands, the testing agency wished to determine whether durability of the various brands of batteries differ for various auto models: saloon car, pick-up car, and a truck. Table below reports data on the number of respondents in a survey of city car-owners.

	Type of auto			Life of battery	
Battery brand	Saloon	Pick-up	Truck	Short	Long
Yuasa	180	166	140	203	258
GS	125	102	94	300	102
Chloride-oxide	150	142	120	47	165

(a) To what extent does type of auto enable a person to predict preference of battery type?

(b) To what extent does knowledge of battery brand enable a person to predict life of the battery purchased?

Exercise 6.4.4

The nutritional heath status of Tanzania's population has failed to make great strides over the years 1995 to 2007, particularly in regard to the most vulnerable within society – the children. The table below reports data on child nutritional status between rural and urban areas for a survey of 250 children in primary schools in selected regions. To what extent does knowledge of place of residence enable a person to predict heath status of children?

	Rural	Urban
Height-for-age	60	40
Weight-for-height	30	70

Exercise 6.4.5

The table below reports data on individuals, who were cross-classified on the basis of sex (A), place of residence (B), and preference for domestic or foreign cars.

Sex	Residence	Domestic	Foreign
Male	Urban	218	29
	Rural	155	14
Female	Urban	65	16
	Rural	26	7
Total		464	66

(a) To what extent does knowledge of sex enable a person to predict residence?

(b) To what extent does knowledge of residence enable someone to predict car preference?

Exercise 6.4.6

Understanding the relationship between I (percent of consumers who express an intention to buy branded goods) and U (percent of consumers using branded goods) pre-occupies marketers of branded goods. The central concern is to provide an answer to the question: *Do consumers' expressed intentions-to-buy branded goods predict future changes in their usage behaviour*? An answer to this pertinent marketing question has implication for the kind and type of marketing strategies that a company adopts for branded goods. To explore the $I-U$ relationship, a market researcher has collected data on users and non-users of a brand, and the survey data are summarized in the accompanying table.

Percent who in 2007expressed an intention to buy branded goods	**Users in 2007**	**Non-users in 2007**
Users in 2008	*75*	*27*
Non-users in 2008	*77*	*15*

6.5 Correlational techniques for ordinal data

So far, non-parametric techniques with ordinal data have focused on ascertaining relative magnitude of sample observations for the purpose of determining effects of treatments. The rank-sum techniques in chapter three are important in this regard. The cardinal purpose of the said chapter is to test the extent to which sample observations display attributes of a hypothesized target population.

Ordinal data are also used for testing monotonic relationships between two-ranked variables. A rank correlation provides a measure of the degree of linearity or a coefficient of agreement for preference data. These correlations are interpreted as follows: correlation equal to unity is for a perfect agreement in ranking system; correlation of negative unity is for perfect disagreement in the ranking system; and a correlation of zero for a completely unrelated ranking system. The common rank correlation coefficients are **Spearman** (r_s), **Kendall's tau** $(\tau_a, \tau_b \text{ and } \tau_c)$[20], **Goodman and Kruskal gamma** (γ), and **Somer's** (d_y) coefficients.

6.5.1 The Spearman rank correlation

The Spearman rank correlation is calculated by using ranks as the paired measurements on two variables, say, X and Y. Thus, if x_r and y_r are the respective ranks of X and Y observations, then the Spearman rank correlation coefficient is:

$$r_s = \frac{N\sum x_r y_r - \sum x_r \sum y_r}{\sqrt{\left(N\sum x_r^2 - (\sum x_r)^2\right)\left(N\sum y_r^2 - (\sum y_r)^2\right)}}$$

When there are no ties in the X observations or the Y observations, the above expression for r_s algebraically reduces to a simpler expression given below.

$$r_s = 1 - \frac{6\sum d_i^2}{N(N^2-1)} \qquad where: d_i = x_r - y_r$$

As a test of the null hypothesis of no relationship in the population, a standard normal statistic is computed as follows: $Z = \frac{r_s - 0}{1/\sqrt{N-1}}$ provided that $N \geq 10$. Alternatively, a critical correlation $r = r_0$ based on sampling distribution of Spearman rank correlation can be obtained from Table 10 appended, and thus accept Ho if computed value is less the critical value, i.e., $r_c < r_0$.

[20] Kendall's *tau-a* and *tau-b* are for ungrouped ordinal data and Kendall's *tau-c* is for grouped ordinal data presented in an $r \times c$ ordered contingency table. Note that Kendall's tau coefficients are for ordinal data whereas, Goodman and Kruskal tau coefficient (τ_p) are for nominal data.

Example 6.5.1

The wisdom from capital market theory dictates that there is a positive relationship between investment return and risk inherent in the investment. In business, earnings-per-share $\left(EPS = \frac{earnings}{no.\ shares}\right)$ and its coefficient of variation $\left(CV = \frac{EPS}{\sigma_{EPS}}\right)$ are respectively used to measure return and inherent risk of corporate investment. The accompanying table reports data on ten companies for the year 2008.

Company	A	B	C	D	E
EPS	9.96	3.25	3.77	9.50	2.30
CV_{EPS}	5.82	25.30	1.50	35.20	8.00
Company	F	G	H	I	J
EPS	4.80	3.58	12.65	1.08	10.25
CV_{EPS}	9.90	14.70	32.40	2.50	18.90

(a) Calculate the Spearman rank correlation between earnings-per-share and the EPS coefficient of variation.

(b) Do these data present sufficient evidence to support the capital market theory that the higher the return the higher the risk?

Solution 6.5.1

The intermediate computations are summarized in the table below.

Company	EPS	CV_{EPS}	Ranks EPS	Ranks CV	Difference d_i	Difference d_i^2
A	9.96	5.82	8	3	5	25
B	3.25	25.3	3	8	-5	25
C	3.77	1.5	5	1	4	16
D	9.50	35.2	7	10	-3	9
E	2.30	8.0	2	4	-2	4
F	4.80	9.9	6	5	1	1
G	3.58	14.7	4	6	-2	4
H	12.65	32.4	10	9	1	1
I	1.08	2.5	1	2	-1	1
J	10.25	18.9	9	7	2	4
Total					0	90

(a) The Spearman rank correlation coefficient is:

$$r_s = 1 - \frac{6\sum d_i^2}{N(N^2-1)} = 1 - \frac{6\times 90}{10(100-1)} = 0.455$$

(b) Referring to Table 10, on Spearman rank correlation test, the critical correlation is $r_0 = 0.564$ and hence Ho is rejected for all values greater that 0.564. Since in the present case, $r_0 = 0.455$ then Ho is accepted, thereby supporting the capital market theory that investors must be compensated for assuming additional investment risk.

6.5.2 Other correlational techniques for ordinal data

The Kendall's tau-a, tau-b and tau-c, Goodman and Kruskal gamma, and Somer's coefficient $(\tau_a, \tau_b, \tau_c, \gamma, and\ d_y)$ are computed on the basis of concordant pairs[21]. A pair (X, Y) is said to be **concordant** (*discordant*) if the **rank of X exceeds** (*less than*) the rank of Y. Thus, let:

C = Number of concordant pairs in which rank of X exceeds rank of Y

D = Number of discordant pairs in which rank of Y exceeds rank of X

T_x = Number of pairs tied on X but not on Y

T_y = Number of ties on Y but not on X

$m = \min(r, c)$ for an $r \times c$ ordered cross-tabulation

N = Total number of paired observations

$$\tau_a = \frac{C-D}{\frac{1}{2}N(N-1)} = \frac{2(C-D)}{N(N-1)}$$

$$\tau_b = \frac{C-D}{\sqrt{(C+D+T_x)(C+D+T_Y)}}$$

$$\tau_c = \frac{(C-D)}{\frac{1}{2}N^2[(m-1)/m]} = \frac{2m(C-D)}{N^2(m-1)}$$

$$\gamma = \frac{C-D}{C+D}$$

$$d_y = \frac{C-D}{C+D+T_y} \quad or \quad d_x = \frac{C-D}{C+D+T_x}$$

The Kendall's *tau-a* and *tau-b* are applicable for ungrouped ordinal data, with *tau-a* being appropriate when there are no ties in the rankings of

[21] Somer's d_y - coefficient is a directional correlation, while the others are symmetric measures of correlation

both factors. Kendall's *tau-c* is applicable for grouped ordinal data, i.e., for an $r \times c$ ordered cross-tabulation.

The sampling distribution of $S = C - D$ under a null hypothesis $(Ho: S = 0 \Rightarrow \tau_a = 0)$ is approximately normal provided that, $N \geq 10$ and there are no ties. Given these two conditions, the test-statistic for Kendall's *tau* coefficient is $Z = \frac{S - 0}{\sigma_S}$; where $\sigma_S = \sqrt{\frac{1}{18} N(N-1)(2N+5)}$. Although the normal approximation holds only when they are no ties, it can be safely applied when the number of ties is relatively small.

The procedure of computing number of concordant and discordant pairs is as follows. First; ranks of the first factor are arranged in ascending order and considered fixed. This is followed by arranging ranks of the second factor alongside with their corresponding ranks of the first factor. Then, count number of ranks that are **above** (*below*) to the right of the first-rank-entry, second-rank-entry, and so on up to the last-rank-entry of the second factor. The sum of such counts represents the number of **concordant** (*discordant*) pairs.

The above computational procedure is also applicable with an $r \times c$ ordered cross-tabulation or contingency table. When a contingency table is unfolded into two marginal frequency distributions, the ordered entries of the first factor represent a ranked-fixed-set and those of the second factor represent the corresponding ranked entries. Thus, the number of **concordant** (*discordant*) pairs can be determined for the purpose of computing Kendall's tau and Goodman and Kruskal gamma based on contingency table.

However, a computational algorithm that is considered convenient for an $r \times c$ ordered cross-tabulation is in place. For instance, given that o_{ij} are

frequencies in a $r \times c$ contingency table, then the numbers of concordant (C) and discordant (D) pairs are respectively equal to:

$$C = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} o_{ij} \left(\sum_{k=i+1}^{r} \sum_{m=j+1}^{c} o_{km} \right) \text{ and } D = \sum_{i=1}^{r-1} \sum_{j=2}^{c} o_{ij} \left(\sum_{k=i+1}^{r} \sum_{m=1}^{j-1} o_{km} \right).$$

Note that in computing the number of concordant pairs start at the **north-west corner** cell (o_{11}) in the first column and move forward to the right. The starting point in computing the number of discordant pairs is the **north-east corner** cell (o_{1c}) in the third column and then moving backward to the left. The following exercise illustrates the use of the above formulas for computing the number of concordant and/or discordant pairs for a 3×3 contingency table given below.

o_{ij}		*j*		
		1	2	3
i	1	5	9	27
	2	10	13	16
	3	10	8	7

$$C = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} o_{ij} \left(\sum_{k=i+1}^{r} \sum_{m=j+1}^{c} o_{km} \right) = o_{11}(o_{22} + o_{23} + o_{32} + o_{33})$$
$$+ o_{21}(o_{32} + o_{33}) + o_{12}(o_{23} + o_{33}) + o_{22}(o_{33})$$
$$= 5(13 + 8 + 16 + 7) + 10(8 + 7) + 9(16 + 7) + 13(7)$$
$$= 220 + 150 + 207 + 91 = 668$$

$$D = \sum_{i=1}^{r-1} \sum_{j=2}^{c} o_{ij} \left(\sum_{k=i+1}^{r} \sum_{m=1}^{j-1} o_{km} \right) = o_{13}(o_{22} + o_{32} + o_{21} + o_{31})$$
$$+ o_{23}(o_{32} + o_{31}) + o_{12}(o_{21} + o_{31}) + o_{22}(o_{31})$$
$$= 27(13 + 8 + 10 + 10) + 16(8 + 10) + 9(10 + 10) + 13(10)$$
$$= 1107 + 288 + 180 + 130 = 1705$$

Example 6.5.2

Consider the ranks of the earnings per share and coefficient of variation factors in example 6.4.1.

Solution 6.5.2

To compute Kendall's *tau* coefficient, we arrange the ranks of EPS in ascending order along side with their corresponding CV ranks as follows:

EPS	1	2	3	4	5	6	7	8	9	10
CV	2	4	8	6	1	5	10	3	7	9

We compute the number of concordant pairs by considering the number of ranks that are above (or below) to the **right** of 2nd, 4th, 8th, and soon up to 9th rank. For instance, the number of CV-ranks that are **above** 2 is 8; **above** 4 is 6; **above** 8 is 2, and so on, **above** 9 is 0. The total number of concordant pairs is $C = 8+6+2+3+5+3+0+2+1+0 = 30$.

We also compute the number of ranks that are below CV-ranks to the **right** of 2nd, 4th, 8th, and so on to 9th rank. For instance, the number of CV-ranks that are **below** 2 is 1; **below** 4 is 2; **below** 8 is 5; and so on, **below** 9 is 0. Then, $D = 1+2+5+3+0+1+3+0+0+0 = 15$ is the total number of discordant pairs. Given that there are no tied ranks on EPS or CV, then Kendal's *tau-a* and Goodman and Kruskal gamma coefficients are:

$$\tau_a = \frac{C-D}{\frac{1}{2}N(N-1)} = \frac{30-15}{\frac{1}{2}\times 10\times 9} = 0.33$$

$$\gamma = \frac{C-D}{C+D} = \frac{30-15}{30+15} = 0.33$$

The normal approximation for $S = C - D$ under a null hypothesis that $Ho: S = 0$ is $Z = \frac{S-0}{\sigma_S}$; where $\sigma_S = \sqrt{\frac{1}{18}N(N-1)(2N+5)}$. In the present case $\sigma_S = 11.18$ and $Z = 1.34$; indicating clearly that $\tau_a = 0.33$ is statistically insignificant at the conventional level of $\alpha = 0.05$.

Example 6.5.3

Does security price risk vary according the size of issuing firms? To answer this question, a researcher randomly selected 25 small, 30 medium, and 50 large manufacturing enterprises from the manufacturing industry database. The average weekly price changes of the selected enterprises were recorded over a period of six months. The results on the number of observed frequencies (o_{ij}) are shown in the table.

o_{ij}	**Size of Manufacturing Enterprises**		
Security Price Change	*Small*	*Medium*	*Large*
Low-risk: < 2 points	5	9	27
Moderate-risk: 2 to 5 points	10	13	16
High-risk: > 5 points	10	8	7
Total	25	30	50

Solution 6.5.3

The two factors, presented in the contingency table above are ordered from small to large along the size of manufacturing enterprises factor; and from low to high along the price-risk factor. Kendall's *tau-c* and Goodman and Kruskal gamma are used to determine the extent to which price-risk varies with size. The intermediary computations are:

Number of Concordant pairs (C)		*Number of Discordant pairs (D)*	
5 (13+8+16+7) =5 x 44	220	27 (13+8+10+10) =27 x 41	1107
10 (8+7) = 10 x 15	150	16 (8+10) = 16 x 18	288
9 (16+7) = 9 x 23	207	9 (10+10) = 9 x 20	180
13 (7) = 13 x 7	91	13(10) = 13 x 10	130
Total	668	Total	1705

$$\tau_c = \frac{2m(C-D)}{N^2(m-1)} = \frac{2\times 3(668-11705)}{105^2(3-1)} = \frac{6\times(-1037)}{11025\times 2} = \frac{-6222}{22050} = -0.282$$

$$\gamma = \frac{C-D}{C+D} = \frac{-1037}{2373} = -0.437$$

The results indicate that there is a negative correlation between price risk and size of a manufacturing enterprises reported in the study. The normal approximation under a null hypothesis that $Ho: S = 0$ is $Z = \frac{S-0}{\sigma_S}$; where $\sigma_S = \sqrt{\frac{1}{18}N(N-1)(2N+5)}$. In the present case $\sigma_S = 361.16$ and $Z = -2.87$; indicating clearly that $\tau_c = -0.282$ is statistically significant at the conventional level of $\alpha = 0.05$.

Example 6.5.4

A researcher investigated whether there is really a relationship between a person's performance in a company training programme and ultimate success in the job. The accompanying table summarizes data on *performance in training programme* and *success in the job* from a random sample of 400 cases taken from Company's very extensive files of employees who completed the programme.

Success in the job	**Performance in training programme**			
	Below average	Average	Above average	Total
Poor	23	61	29	113
Average	28	78	60	166
Very good	9	49	63	121
Total	60	188	152	400

Use appropriate correlational techniques for ordinal data to ascertain the extent to which success in the job is associated with performance in training programme.

Solution 6.5.4

Given an ordered contingency table, the appropriate correlational technique is to compute Kendall's tau-c and Goodman and Kruskal gamma coefficients as follows:

$N = 400, \quad m = 3$

$C = 23(78+60+49+63)+28(49+63)+61(60+63)+78(63) = 21303$

$D = 29(28+9+78+49)+61(28+9)+78(9)+60(9+49) = 11195$

$$\tau_c = \frac{2m(C-D)}{N^2(m-1)} = \frac{2\times 3(21303-11195)}{400^2(3-1)} = \frac{60648}{320000} = 0.1895$$

$$\gamma = \frac{21303-11195}{21303+11195} = \frac{10108}{32498} = 0.311$$

There is a positive correlation between success in the job and performance in the training programme reported in this study. The significance of this result is checked by the normal approximation under a null hypothesis that $Ho: S = 0$, which is $Z = \frac{S-0}{\sigma_S}$. In the present case $\sigma_S = \sqrt{\frac{1}{18}N(N-1)(2N+5)} = 2671.64$ and thus $Z = 3.78$; indicating clearly that $\tau_c = 0.1895$ is statistically significant at the conventional level of $\alpha = 0.05$.

Exercise 6.5.1

Pro-rich and pro-poor political parties are vigorously marketing their policies to the public in the wake of democratic elections to be held next year. Evaluating the extent to which public opinion is supporting a given economic policy orientation is critical at this stage. A survey study is seeking public opinion on the acceptability of the pro-rich political party policies before the elections. The following table reports data on the number of people surveyed from four different economic regions: very poor, poor, rich and very rich who support its economic policies. Do these data on opinion polls provide evidence on the support of the pro-rich policies of the party?

	Low support	Moderate support	High support
Very poor	400	450	250
Poor	700	1000	1500
Rich	1000	500	1800
Very rich	900	2050	550

Exercise 6.5.2

Is it possible to predict marriage adjustment by individuals through knowledge of their educational level? To answer this research question, data on education and marriage adjustment were collected and the results are summarized in the following table:

Education	Marriage adjustment				
	Very low	Low	High	Very High	Total
Primary	18	29	70	115	232
Secondary	17	28	30	41	116
University	11	10	11	20	52
Total	46	67	111	176	400

Exercise 6.5.3

Book-reading habit is often being associated with educational level. The contention held is that the higher the educational level, the higher the desire for more knowledge and this desire translates into book-reading behaviour. To examine this aspect, a random sample of adults was obtained, and each member of the sample classified according to educational level and according to the number of books one has read within the past year. The results of the classification are summarized in the accompanying table. Do these data support, at 5% level of significance, the proposition that education is associated with book-reading habit?

	Educational level		
Books read	12 years	12 – 16 years	More than 16 years
None	330	50	20
One	50	100	50
Two	100	150	50
3 or more	20	30	50

6.6 *SPSS Tutorial on Correlation Techniques*

The following set of SPSS output are for problems that have been solved in the chapter. These problems focus on the degree of association or correlation between two categorical variables that are associated with: political awareness programme; cancer-smoking; age-package option; religion-political orientation; earnings per share and its variation; risk and firm size; and job success and performance in training programme. Data on variables considered for analysis are entered in the data editor as summarized in the table below.

No	Variable List		Description of measurement
1	*AWARE BEFORE*	(row)	0 = yes; 1 = no
	AWARE AFTER	(column)	0 = yes; 1 = no
2	*CANCER*	(row)	0 = infected; 1 = not infected
	SMOKING	(column)	0 = smoking; 1 = no smoking
3	*AGE*	(row)	0 = under 30; 1 = 30 and above
	PACKAGE	(column)	0 = option A; 1 = option B; 2= = option C
4	*POLITICS*	(row)	0 = conservative; 1 = liberal; 2 = cosmopolitan
	RELIGION	(column)	0 = Christianity; 1 = Judaism
5	*ACTUAL PAYMENT*	(row)	0 = non-default; 1 = default
	PREDICTED PAYMENT	(column)	0 = non-default; 1 = default
6	*EPS*	(row)	Scale-measured
	CV_{EPS}	(column)	Scale-measured
7	*RISK*	(row)	0 = low; 1 = moderate; 2 = high
	FIRM SIZE	(column)	0 = small; 1 = medium; 2 = large
8	*JOB SUCCESS*	(row)	0 = poor; 1 = average; 2 = very good
	PERFORMANCE	(column)	0 = below average; 1 = average; 2 = above average

To produce SPSS output, invoke the following commands from the menu:

Analyze

Descriptive

Cross tabs

Test variables: ***CANCER*** (row-variable), ***SMOKING*** (column-variable)

Statistics: *Preferred choice*

6.6.1 Political awareness before-after a programme

AWARE BEFORE * AWARE AFTER Cross-tabulation

Count

		AWARE AFTER		Total
		Yes	No	
AWARE BEFORE	Yes	350	150	500
	No	630	470	1100
Total		980	620	1600

Chi-Square Tests [a, b]

	Value	df	Asy Sig (2-sided)	Exact Sig (2-sided)	Exact Sig (1-sided)
Pearson Chi-Square[c]	**23.460**	**1**	**.000**		
Continuity Correction	22.927	1	.000		
Likelihood Ratio	23.941	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.446	1	.000		
N of Valid Cases	1600				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 193.75.

c Pearson Chi-square for linear models; Likelihood ratio for log-linear models; Linear-by-linear for quantitative variables

6.6.2 Cancer Infection and Smoking Behaviour

CANCER * SMOKING Cross tabulation

Count

		SMOKING		Total
		Smoking	Not smoking	
CANCER	Infected	410	230	640
	Not infected	450	270	720
Total		860	500	1360

Chi-Square Tests

	Value	df	Asymp Sig (2-sided)	Exact Sig (2-sided)	Exact Sig (1-sided)
Pearson Chi-Square	**.356**	**1**	**.551**		
Continuity Correction	.292	1	.589		
Likelihood Ratio	.356	1	.551		
Fisher's Exact Test				.573	.295
Linear-by-Linear Association	.356	1	.551		
N of Valid Cases	1360				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 235.29.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	**Phi**	**.016**	**.551**
	Cramer's V	**.016**	**.551**
	Contingency Coefficient	**.016**	**.551**
N of Valid Cases		1360	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

6.6.3 Age and package option choice

AGE * PACKAGE Cross-tabulation
Count

		PACKAGE			Total
		Option A	Option B	Option C	
AGE	Under 30	200	180	325	705
	30 and above	255	320	230	805
Total		455	500	555	1510

Chi-Square Tests

	Value	df	Asymp Sig (2-sided)
Pearson Chi-Square	**55.732**	**2**	**.000**
Likelihood Ratio	56.108	2	.000
Linear-by-Linear Association	24.539	1	.000
N of Valid Cases	1510		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 212.43.

Directional Measures

			Value	Asy s.e	Approx T	Approx Sig
Nominal by Nominal	Lambda	Symmetric	.111	.023	4.711	.000
		AGE Dependent	.135	.031	4.054	.000
		PACKAGE Dependent	**.094**	**.023**	**3.856**	**.000**
	Goodman and Kruskal tau	AGE Dependent	.037	.010		.000
		PACKAGE Dependent	.019	.005		.000
	Uncertainty Coefficient	Symmetric	.021	.005	3.784	.000
		AGE Dependent	.027	.007	3.784	.000
		PACKAGE Dependent	.017	.004	3.784	.000

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on chi-square approximation

d Likelihood ratio chi-square probability.

6.6.4 Religion and Political Orientation

POLITICS * RELIGION Cross-tabulation
Count

		RELIGION		Total
		Christianity	Judaism	
POLITICS	Conservative	300	600	900
	Liberal	600	100	700
	Cosmopolitan	300	100	400
Total		1200	800	2000

Chi-Square Tests [a]

	Value	df	Asy Sig (2-sided)
Pearson Chi-Square	**497.024**	**2**	**.000**
Likelihood Ratio	522.290	2	.000
Linear-by-Linear Association	318.989	1	.000
N of Valid Cases	2000		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 160.00.

Directional Measures

			Value	Asy s.e	Approx T	Approx Sig
Nominal by Nominal	Lambda	Symmetric	.316	.023	12.734	.000
		RELIGION Dependent	.273	.023	10.260	.000
		POILITICS Dependent	.375	.030	10.260	.000
	Goodman and Kruskal tau	**POLITICS Dependent**	**.154**	**.012**		.000
		RELIGION Dependent	.249	.019		.000
	Uncertainty Coefficient	Symmetric	.152	.012	12.373	.000
		POLITICS Dependent	.125	.010	12.373	.000
		RELIGION Dependent	.194	.016	12.373	.000

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on chi-square approximation

d Likelihood ratio chi-square probability.

6.6.5 Bank lending evaluation

ACTUAL * PREDICTED Cross-tabulation

Count

		PREDICTED		Total
		Non-default	Default	
ACTUAL	Non-default	5	4	9
	Default	7	9	16
Total		12	13	25

Chi-Square Tests [a, b]

	Value	df	Asy Sig (2-sided)	Exact Sig (2-sided)	Exact Sig (1-sided)
Pearson Chi-Square	**.322**	**1**	**.571**		
Continuity Correction	.023	1	.881		
Likelihood Ratio	.322	1	.570		
Fisher's Exact Test				.688	.440
Linear-by-Linear Association	.309	1	.578		
N of Valid Cases	25				

a Computed only for a 2x2 table

b 2 cells (50.0%) have expected count less than 5. The minimum expected count is 4.32.

Symmetric Measures

		Value	Approx Sig
Nominal by Nominal	**Phi**	**.113**	**.571**
	Cramer's V	.113	.571
	Contingency Coefficient	.113	.571
N of Valid Cases		25	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

6.6.6 Investment returns and risk[22]

Correlations

			EPS	CVEPS
Kendall's tau_b	**EPS**	**Correlation Coefficient**	**1.000**	**.333**
		Sig. (2-tailed)	.	.180
		N	10	10
Spearman's rho	**EPS**	**Correlation Coefficient**	**1.000**	**.455**
		Sig. (2-tailed)	.	.187
		N	10	10

[22] The sequence of SPSS commands for this output is: **analyze → correlate → bivariate correlations →** variables (***EPS, CVEPS***) → correlation coefficients (**Kendall's tau-b, Spearman**)

6.6.7 Price risk and firm size

RISK * FIRM SIZE Cross-tabulation

Count

		FIRM SIZE			Total
		Below average	Average	Above average	
RISK	Low	5	9	27	41
	Moderate	10	13	16	39
	High	10	8	7	25
Total		25	30	50	105

Chi-Square Tests

	Value	df	Asy Sig (2-sided)
Pearson Chi-Square	**11.411**	**4**	**.022**
Likelihood Ratio	11.548	4	.021
Linear-by-Linear Association	10.641	1	.001
N of Valid Cases	105		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.95.

Directional Measures

			Value	Asy s.e	Approx T	Approx Sig
Ordinal by Ordinal	Somers' d	Symmetric	-.292	.082	-3.555	.000
		RISK Dependent	-.296	.083	-3.555	.000
		FIRM SIZE Dependent	-.288	.081	-3.555	.000

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asy s.e	Approx T	Approx Sig
Ordinal by Ordinal	Kendall's tau-b	-.292	.082	-3.555	.000
	Kendall's tau-c	**-.282**	**.079**	**-3.555**	**.000**
	Gamma	**-.437**	**.114**	**-3.555**	**.000**
Valid Cases		105			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

6.6.8 Success in job and performance in training programme

SUCCESS IN JOB * PERFORM TRAIN Cross-tabulation

Count

		PERFORM-TRAIN			Total
		Average	Moderate	Above-average	
SUCCESS IN JOB	Poor	23	61	29	113
	Average	28	78	60	166
	Very good	9	49	63	121
Total		60	188	152	400

Chi-Square Tests

	Value	df	Asy Sig (2-sided)
Pearson Chi-Square	**20.395**	4	**.000**
Likelihood Ratio	21.155	4	.000
Linear-by-Linear Association	19.051	1	.000
N of Valid Cases	400		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.95.

Directional Measures

			Value	Asy s.e	Approx T	Approx Sig
Ordinal by Ordinal	Somers' d	Symmetric	.199	.042	4.666	.000
		SUCCESS IN JOB Dependent	.206	.044	4.666	.000
		PERFORM-TRAIN Dependent	.192	.041	4.666	.000

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asy s.e	Approx T	Approx Sig
Ordinal by Ordinal	Kendall's tau-b	.199	.042	4.666	.000
	Kendall's tau-c	**.190**	**.041**	**4.666**	**.000**
	Gamma	**.311**	**.064**	**4.666**	**.000**
N of Valid Cases		400			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Problem Set Six

Question 6.1

Researches in marketing management have revealed that lower-income shoppers do not appear to be taking advantage of unit pricing, perhaps because they do not understand the unit-price labelling system compared to middle-income and upper-income shoppers. In a follow-up study to provide a check on this expert opinion, an economist observed the purchase selection of $n = 1000$ shoppers in large super-markets. The supermarkets were located in three different areas of a city where the buyers were respectively, of lower-income, middle-income and upper-income households. Packages of the same brand but with different unit prices were placed adjacent to one another on the store shelves. The data were classified according to decision and income group as shown below.

Number of Shoppers	**Income group**			**Total**
	Lower	*Middle*	*Upper*	
Understand unit-price label	249	494	202	944
Do not understand unit-price label	26	26	4	56
Total	275	520	205	1000

Do the above data present sufficient evidence to support the marketing management research findings?

Question 6.2

Worker participation in management decision-making is more of rule than exception in Tanzania. To study the effects of worker satisfaction with worker participation in managerial decision-making, a researcher interviewed 200 workers in each of two separate Tanzanian firms. One firm had active worker participation in management decision-making; the other did not. Each selected worker was asked whether s/he generally approved managerial

decisions made in the firm. The results of the interviews are shown in the table. Do these data support the hypothesis that those workers in a firm with participative decision making more generally approve firm's managerial decisions than those employed by firms without participative decision-making?

Approval of Managerial Decisions	*Participative Decision Making*	*No Participative Decision Making*
Generally Approve	103	71
Do Not Approve	52	74

Question 6.3

Many researchers have reported that the amount of television clutter may impact unfavourably on advertising effectiveness[23]. Reduced attention, recall, and brand awareness have been found to be caused by an increase in TV clutter[24]. To explore the effects of TV clutter on brand awareness, an agency conducted telephone surveys among 300 households following each of three different 60-minute television programmes. The results of the survey are reported in the table.

Brand Recall	**Level of Clutter**		
	Light (*less than 10 minutes*)	*Moderate* (*10 to 15 minutes*)	*Heavy* (*15 minutes or more*)
At least half advertised brands	139	133	127
Less than half advertised brands	161	167	173
Total	300	300	300

Do these data suggest that brand recollection decreases with increased levels of TV clutter?

[23] TV Clutter includes all non-programme materials such as public service announcements and commercials.
[24] Webb, P. and M. Ray (1979), "Effects of TV Clutter," *Journal of Advertising Research*

Question 6.4

A survey of the opinions of the stockholders of a corporation about a proposed merger was undertaken to determine whether the resulting opinion was independent of the number of shares held. Are the results that are summarized in the accompanying table providing sufficient evidence to indicate that opinions concerning merger are dependent on number of shares held by a stockholder?

Shares Held	Opinion		
	In favour	*Opposed*	*Undecided*
Under 1000	370	160	50
1000 – 5000	300	220	80
Over 5000	320	440	60

Question 6.5

Researchers on consumerism and marketing management have noted that the level of support for consumerism and perceptions about the movement may vary among different segments of a society[25]. In examining this aspect, the researchers conducted a survey among 340 students, 150 housewives, and 170 businessmen to measure the strength of their conviction regarding product information provided on common consumer sundry items. The results of the study were that 290 students, 110 housewives, and 120 businessmen rated this aspect as extremely important. Do these data present sufficient evidence to indicate that the fraction of people favouring explicit product information on sundry items differs among students, housewives, and businessmen?

Question 6.6

Women, in many parts of the country, are launching their own savings and lending institutions that are intended to provide women with easier access to credit and banking careers. The move is seen as a search for fair treatment and to eliminate discrimination in banking practices against women. The CEO of a large women's bank conducted a survey to determine the attitudes of women of different ages toward the concept of a bank owned and operated by women. The results of the survey are shown in the accompanying table. Do these data present sufficient evidence to assume that the proportion of women favouring the concept of a bank owned and operated by women differs among women from different age groups?

Attitude	Age Group				
	21 – 30	*31 - 40*	*41 - 50*	*51 - 60*	*Over 60*
Support women bank concept	50	60	40	35	30
Do not support women bank concept	70	40	50	75	55

Question 6.7

The City Authority conducted a survey to determine whether the incidence of various types of crime varied from one part of the city to another. Presently the city has four administrative districts and crimes are classified as car-theft, homicide, larceny, sex-offences, and other. An analysis of cases showed results that are given in the accompanying table. Do these data present sufficient evidence to indicate that the occurrence of various types of crime depends on city district?

City District	Car-theft	Homicide	Larceny	Sex offences	Other
North-east	250	15	190	100	40
North-west	160	20	110	130	60
Central	60	7	280	40	20
South-east	90	3	310	200	70
South-west	110	30	40	35	10

Question 6.8

The relationship between worker productivity and quality of work-life environment is well recognized in business management theories. For

[25] Kangun, N., K. Cox, J. Higginbotham, and J. Burton (1975), "Consumerism and Marketing Management," *Journal of Marketing*

instance, management concessions such as improved lighting, attractive office facilities, and positive attention to worker grievances have often enhanced worker productivity. To determine management attitudes on this subject, a researcher questioned managers and executives from several different organizations. The results of the study are summarized in the accompanying table. Do these data provide sufficient evidence to indicate that the probability that a manager or executive favours improving the quality of employee work-life to increase productivity is independent on the type of organization?

Type of Organization	Number surveyed	Percentage favouring Quality-of-work-life Improvement
Public service	180	30
Manufacturing	220	35
Wholesale	300	32
Retail business	140	25
Tourism	160	40

Question 6.9

The usual assumption in marketing management is to identify similar market segments for a wide variety of consumer durables and services on the basis of ability to buy and social class. This assumption has been questioned on account of empirical evidence. For instance, members of lower social classes tend to buy large cars and other durables similar to those purchased by members of a higher income and social class. In exploring this aspect, recent purchases of new autos were categorized according to their social class and the price class of the auto purchased. The results of the study are shown in the accompanying table. Do these data present sufficient evidence to indicate that the price of auto purchased depends on social class?

Social Class	Auto Price Class		
	Low-priced	*Moderately-priced*	*High-priced*
Professionals	40	25	15
Executives	170	100	20
Politicians	200	220	60
Businessmen	20	40	50
Commercial Farmers	10	35	25

Question 6.10

Employee ownership of company shares as a way of increased employee commitment to a firm is widely encouraged. However, workers and management often perceive the relationship between employee ownership and commitment differently. To examine this aspect, a researcher conducted a survey of employees of a large corporate body who owned shares of the company. Each respondent was asked to state the reason for owning company shares. The results are summarized in the accompanying table. Is there a sufficient evidence to indicate that the reasons for employee ownership of a share differ among employees groups?

Reason	*Blue-collar*	*White-collar*	*Middle Management*	*Top Management*
Believe in employee ownership	40	25	20	7
To save my job	80	30	15	5
As investment	35	15	10	8

Question 6.11

Many researchers have reported that the amount of television clutter may impact unfavourably on advertising effectiveness. Reduced attention, recall, and brand awareness have been found to be caused by an increase in TV clutter. To explore the effects of TV clutter on *ordered-brand awareness*, an agency conducted telephone surveys among 300 households following each of three different 60-minute television programmes. The results of the survey are reported in the table.

Brand Recall	Level of Clutter		
	Light (less than 10 minutes)	Moderate (10 - 15 minutes)	Heavy (15 minutes or more)
Less than half advertised brands	161	167	173
At least half advertised brands	139	133	127
Total	300	300	300

Do these data suggest that brand recollection decreases with increased levels of TV clutter? How does your answer differ from that on question 6.3 posed previously?

Question 6.12

Women, in many parts of the country, are launching their own savings and lending institutions that are intended to provide women with easier access to credit and banking careers. The move is seen as a search for fair treatment and to eliminate discrimination in banking practices against women. The CEO of a large women's bank conducted a survey to determine the *ordered-support* of women of different ages toward the concept of a bank owned and operated by women. The results of the survey are shown in the accompanying table. Do these data present sufficient evidence to assume that there is correlation between type of *support* and *age* of women?

Support of women-owned bank	Age Group				
	21 – 30	*31 - 40*	*41 - 50*	*51 - 60*	*Over 60*
Low-support women bank concept	50	60	40	35	30
High-support women bank concept	70	40	50	75	55

Question 6.13

The usual assumption in marketing management is to identify similar market segments for a wide variety of consumer durables and services on the basis of ability to buy and social class. This assumption has been questioned on account of empirical evidence. For instance, members of lower social classes tend to buy large cars and other durables similar to those purchased by members of a higher income and social class. In exploring this aspect, recent purchases of new autos were *ordered-categorized* according to their income-level and the price class of the auto purchased. The results of the study are shown in the accompanying table. Do these data present sufficient evidence to indicate that the price of auto purchased depends on social class?

Social Class	Auto Price Class		
	Low-priced	*Moderately-priced*	*High-priced*
Low-income	220	125	55
Middle-income	200	220	60
High-income	50	80	120

2.2: Binomial test is appropriate, with the following characteristics: the target variable is the difference in production index after and before (***DGDP***) with a cut-off point of zero, so that $n = 8,\ p = \frac{1}{2},\ B^+ = 4,\ B^- = 4,\ B^* = 4$. Given that $\alpha = 0.05$, and that $\alpha_c = \Pr ob\left(B^* \le B \le n - B^* : B \approx binomial(n, p)\right) = 1.00$ =1.00, then the data provide sufficient evidence that the economic reforms have had a significant impact on average industrial production in the key sectors considered in this study.

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
DIFGDP	Group 1	<= 0	4	.50	.50	1.000
	Group 2	> 0	4	.50		
	Total		8	1.00		

3.2: The k-sample median test is used: The 28 observations are ranked while retaining sample identity. The Median = 222; and the contingency table is:

	Operator A	Operator B	Operator C	Operator D	Total
Above Median	4	4	2	3	13
Equal and below Median	3	3	5	4	15
Total	7	7	7	7	28

Chi-square = 1.579 and given that $\chi^2_{.05,6} = 12.59$, a null hypothesis of operators being equally productive is accepted.

4.2: The differences between Compensation plans 2 and 1 ranked; the sums of ranks are: negative ranks = 16.50; positive ranks = 11.50, $n = 7,\ T_{.05,7} = 2,\ T_c = 11.50$, since $T_c > T_{.05,7}$. The normal approximation for the Wilcoxon Signed Ranks Test indicates that equal productivity hypothesis is accepted as $Z = \dfrac{T - E(T)}{\sigma_T} = \dfrac{11.50 - 14}{5.916} = -0.424$. Both results indicate that there is no difference in the employee's compensation preferences on the two plans.

4.4: The differences between Insurance plans 2 and 1 are ranked; the sums of ranks are: negative ranks = 50.50; positive ranks = 15.50, $n = 11,\ T_{.05,11} = 11,\ T_c = 15.50$, a null hypothesis of no difference in employee attitudes towards the two insurance plans is accepted since $T_c > T_{.05,11}$. The normal approximation for the Wilcoxon Signed Ranks Test indicates that equal productivity hypothesis is accepted as

$$Z = \frac{T - E(T)}{\sigma_T} = \frac{15.50 - 33}{11.25} = -1.56.$$

4.6: Responses on perception of workers and managers who own at least 100 shares of stock are ranked while retaining sample identity. The sums of ranks are: $R_1 = 30.50$ for workers and $R_2 = 47.50$ for managers. Given that $n_1 = 7,\ n_2 = 5\ then\ U_0 = 6$ (see Table 6 appended). This value is the critical below which H_0 is rejected at 5% level of significance. Based on R_2 that corresponds to the smaller sample,

$U_c = n_2 n_1 + \dfrac{n_2(n_2 + 1)}{2} - R_2 = 2.50$. This result indicates that the null hypothesis on perceived equal distribution of corporate power is rejected. Based on normal approximation, the above null hypothesis is again rejected given that: $E(U) = 17.50,\ \sigma_U = 6.16\ and\ Z = \dfrac{2.50 - 17.50}{6.16} = -2.44$ $\to p < 0.05$.

4.8: Responses on the nine housewives are ranked while retaining identity. The sums of ranks are: $R_1 = 21.50$ for brand switchers and $R_2 = 23.50$ for brand loyalists. Given that $n_1 = 4,\ n_2 = 5$ and $\alpha = 0.05$ then $U_0 = 2$ (see Table 6 appended). This value is the critical below which H_0 is rejected. Based on R_1 that corresponds to the smaller sample, $U_c = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 8.50$. This result indicates that the null hypothesis on perceived cleaning power of the laundry detergent being equally effective is accepted at 5% level of significance. Based on normal approximation, the above null hypothesis is again accepted given that:

$$E(U) = 10,\ \sigma_U = 4.08\ and\ Z = \frac{8.50-10}{4.08} = -0.37.$$

4.10: The average credit usage for the three types of credit cards is ranked and Kruskal-Wallis H-Test used as summarized in the table below.

Test type: Kruskal-Wallis Test

TYPE OF CARD	N	Mean Rank	Total rank
Master card	8	9.88	79
Visa	6	8.33	50
Both cards	4	10.50	42
Total	18		

$$\chi_c^2 = \frac{12}{n(n+1)}\sum_{i=1}^{3}\left(\frac{R_i^2}{n_i}\right) - 3(n+1)$$

$$= \frac{12}{18\times 19}\left(\frac{79^2}{8} + \frac{50^2}{6} + \frac{42^2}{4}\right) - 3\times 19 = 0.466$$

Basing on this result, the null hypothesis of equal use of credit cards is accepted since $\chi^2_{.05,2} = 5.99$.

5.2: The tables below give a summary of the Runs Tests based on median and mean of the market model residuals. In both cases, the null hypothesis of randomness is accepted and thus there are no systematic factors in place that influence stock prices.

Runs Test 1

	RESIDUAL
Test Value	.002500
Cases < Test Value	10
Cases >= Test Value	10
Total Cases	20
Number of Runs	10
Z	**-.230**
Asymp. Sig. (2-tailed)	.818

Runs Test 2

	RESIDUAL
Test Value	-.000150
Cases < Test Value	9
Cases >= Test Value	11
Total Cases	20
Number of Runs	**10**
Z	**-.186**
Asymp. Sig. (2-tailed)	.853

6.2: The contingency table for the problem (expected frequencies are bracketed; independent= decision-making; dependent = approval) is:

Frequencies	Participation in decision making	No participation in decision making	Total
Approve	*103 (90)*	*71 (84)*	*174*
Do not approve	*52 (65)*	*74 (61)*	*126*
Total	*155*	*145*	*300*

$$\chi_c^2 = \frac{(103-90)^2}{90} + \frac{(71-84)^2}{84} + \frac{(52-65)^2}{65} + \frac{(74-61)^2}{61} = 9.26$$

The no-difference in decision-making hypothesis among workers of both firms is rejected in favour of the alternative since $\chi^2_{.05,1} = 3.84$.

6.4: This problem is about predicting shareholder opinions concerning a merger proposal on the basis of known number of shares held by shareholders. Considering that the number of shares held is an independent variable, Kendall's tau-p and Light-Margolin index are the appropriate statistical index for ascertaining the extent to which shareholder opinions can be predicted or made to be predicted from knowledge of the shareholding status. In the present case, $\tau_p = \frac{\varepsilon - \delta}{\varepsilon} = \frac{1156 - 954}{1156} = 0.175$, indicating that knowledge of shareholding status enables to reduce prediction errors on opinion about the merger proposal by about 17.5%. The Light-Margolin measure is $R^2_{pseudo} = \frac{284 + 241 + 365 - 844}{2000 - 844} = \frac{46}{1156} = 0.04$, which indicate that 4% of shareholder opinion about the merger proposal is explained by shareholding.

6.6: The problem is about testing equality of proportions, and the appropriate statistical index is the Chi-square. The null hypothesis is that proportions of those supporting the women bank concept is equal irrespective their age group, which if formally stated as: $H_0 : P_1 = P_2 = P_3 = P_4 = P_5$

Frequencies	21-30	31-40	41-50	51-60	Over 60	Total
Support	50 (51.1)	60 (42.6)	40 (38.3)	35 (46.8)	30 (36.2)	215
Do not support	70 (68.9)	40 (57.4)	50 (51.7)	75 (63.2)	55 (48.8)	290
Total	120	100	90	110	85	505

$$\chi^2_c = \frac{(50-51.1)^2}{51.1} + \frac{(60-42.6)^2}{42.6} + \cdots + \frac{(55-48.8)^2}{48.8} = 19.58,$$ which is greater than $\chi^2_{0.05,4} = 9.24$. This result indicates that the null hypothesis is rejected.

6.8: This problem is a direct application of Chi-square test of equality of proportions, testing the null hypothesis of equal proportion of managers or executives favouring improved quality of work-life for employees irrespective of the type of organizations they belong to. More formally: $H_0 : P_1 = P_2 = P_3 = P_4 = P_5$.

Frequencies	Favouring	Not favouring	Total
Public service	30 (29.16)	150 (150.84)	180
Manufacturing	35 (35.64)	185 (184.36)	220
Whole sale	32 (48.60)	268 (251.40)	300
Retail business	25 (22.68)	115 (117.32)	140
Tourism	40 (25.92)	120 (134.08)	160
Total	62	838	1000

$$\chi^2_c = \frac{(30-29.16)^2}{29.16} + \frac{(150-150.84)^2}{150.84} + \cdots + \frac{(120-134.08)^2}{134.08} = 16.2189$$

which is greater than $\chi^2_{0.05,4} = 9.24$. This result indicates that the null hypothesis is rejected at $\alpha = 0.05$.

6.10: The problem is an application of chi-square test of independence. PRE-based correlation indices $(\lambda_p, \tau_p, and\ R^2_{pseudo})$ may also be used to ascertain the extent to which employee commitment (dependent) can be predicted on the basis of employee-category (independent) as perceived by respondents in this study.

Frequencies	Blue-collar	White-collar	Middle management	Top management	Total
Believe in employee ownership	40 (49.17)	25 (22.21)	20(14.28)	7 (6.34)	92
To save my job	80 (69.48)	30 (31.38)	15 (20.17)	5 (8.97)	130
As investment	35 (36.35)	15 (16.41)	10 (10.55)	8 (4.69)	68
Total	155	70	45	20	290

$$\chi_c^2 = \frac{(40-49.17)^2}{49.17} + \frac{(25-22.21)^2}{22.21} + \cdots + \frac{(8-4.69)^2}{4.69} = 11.483$$

$$\lambda_p = \frac{80+30+20+8-130}{290-130} = \frac{8}{160} = 0.05$$

$$\varepsilon = \frac{92(290-92)+130(290-130)+68(290-68)}{290} = 187$$

$$\delta = \frac{40(155-40)+80(155-80)+35(155-35)}{155}$$

$$+\frac{25(70-25)+\cdots}{70} + \frac{7(20-7)}{20}\cdots = 182$$

$$\tau_p = \frac{\varepsilon-\delta}{\varepsilon} = \frac{187-182}{187} = \frac{5}{187} = 0.027$$

$$R^2_{pseudo} = \frac{\frac{9225}{155}+\frac{1750}{70}+\frac{725}{45}+\frac{138}{20}-\frac{29988}{290}}{290-\frac{29988}{290}} = 0.027$$

The result on Light-Margolin measure indicates that 2.7% of worker-commitment is accounted for by employee ownership.

6.12: The problem is a direct application of correlation analysis with ordinal grouped data, and thus the appropriate statistical indices are the Kendall's tau-c and the Goodman and Kruskal gamma:

$$C = 50(40+50+75+55)+60(50+75+55)+40(75+55)+35(55)$$
$$= 11000+10800+5200+1925 = 28925$$

$$D = 30(75+50+40+70)+35(50+40+70)+40(40+70)+60(70)$$
$$= 7050+5600+4400+4200 = 21250$$

$$\tau_c = \frac{2m(C-D)}{N^2(m-1)} = \frac{2\times 2(28925-21250)}{505^2(2-1)} = \frac{4\times 7675}{255025} = 0.1204$$

$$\gamma = \frac{C-D}{C+D} = \frac{28925-21250}{28925+21250} = \frac{7675}{50175} = 0.153$$

Given that: $N = 505,\quad \sigma_s = \sqrt{\frac{1}{18}N(N-1)(2N+5)} = 3788.4$ and since $S = C - D = 7675$, then $Z = \frac{S-0}{\sigma_s} = \frac{7675}{3788.4} = 2.026$.

The result is considered insignificant substantively $(\tau_c < 0.30)$, but statistically the result is significant at the conventional level of 5%. Overall it can be concluded that there is a positive correlation between *type of support* and *age* of women regarding the establishment of women-owned bank.

APPENDIX II: STATISTICAL TABLES

Table 1: Areas under the Normal Curve

This table gives the 100α percentage point of a standardized Normal distribution defined as $Z = \frac{X - \mu}{\sigma}$. The tabulation is for one tail only.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4525	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0022	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Table 2: Percentage Points of the t Distribution

This table gives the value of $t_{\alpha;\nu}$; the 100α percentage point of the t distribution for ν degrees of freedom. The tabulation is for one-tail only.

$\nu \backslash \alpha$	.25	.20	.10	.05	.025	.01	.005	.0005
1	1.000	1.376	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.061	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	.941	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	.906	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	.896	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	.889	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	.879	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	.860	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	.859	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	.857	1.318	1.711	2.064	2.492	2.397	3.745
25	.684	.856	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	.855	1.314	1.703	2.52	2.473	2.771	3.690
28	.683	.855	3.313	1.701	2.048	2.467	2.763	3.674
29	.683	.854	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	.854	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	.853	1.308	1.684	2.021	2.423	2.704	3.551
60	.679	.848	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	.843	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	.842	1.282	1.645	1.960	2.326	2.576	3.291

Source: This table is abridged from Table III of Fisher and Yates; *Statistical tables for Biological, Agricultural and Medical Research*, published by Oliver & Boyd Ltd Edinburgh and by permission of the authors and publishers.

Table 3: Percentage Points of the χ^2 Distribution

This table gives the value of χ^2; 100α percentage point of the Chi-square distribution for ν degrees of freedom

$\nu \backslash \alpha$	.995	.99	.95	.10	.05	.025	.01	.005
1	0.0^4393	0.0^3157	0.0^2393	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.10259	4.61	5.99	7.38	9.21	10.60
3	0.0717	0.115	0.352	6.25	7.91	9.35	11.34	12.84
4	0.207	0.297	0.711	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	1.145	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.635	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	2.17	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.73	13.36	15.51	17.53	20.09	21.96
9	1.735	2.09	3.33	14.68	16..92	19.02	21.67	23.59
10	2.16	2.56	3.94	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	4.57	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	5.23	18.55	21.03	23.34	26.22	27.30
13	3.57	4.11	5.89	19.81	22.36	24.73	27.69	29.82
14	4.07	4.66	6.57	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	7.26	22.31	24.99	27.49	30.58	32.80
16	5.14	5.81	7.96	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	8.67	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	9.39	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	10.12	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	10.85	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	11.59	29.61	32.67	35.48	38.93	41.40
22	8.64	9.54	12.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	13.09	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	13.85	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	14.61	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	15.38	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	16.15	36.74	40.11	42.19	46.96	49.65
28	12.46	13.56	16.93	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	17.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	18.49	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	26.50	51.80	55.75	59.34	63.69	66.76
50	27.99	29.70	34.76	63.16	67.50	71.42	76.15	79.49
60	35.53	37.48	43.18	74.39	79.08	83.29	88.37	91.95
70	43.27	45.44	51.73	85.52	90.53	95.02	100.42	104.21
80	51.17	53.54	60.39	96.57	101.87	106.62	112.32	116.32
90	59.20	61.75	69.12	107.56	113.14	118.13	124.11	128.29
100	67.33	70.06	77.92	118.49	124.34	129.56	135.80	140.16

Source: "Tables of Percentage Points of the χ^2- Distribution," *Biometrika Tables for Statisticians*, Vol. 1 3rd edition (1966). Reproduced by permission of the Biometrika Trustees.

Table 4: Values of $F_{0.05, \nu_1, \nu_2}$

This table gives the value of the 5-percentage point of F– distribution having ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. The tabulation is for one tail only.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18.5	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.92	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	254	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Table 4: Values of $F_{0.05, v_1, v_2}$ *(continued)*

$v_2 \setminus v_1$	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	244
2	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.12	2.04	1.96	1.91	1.87	1.825	1.77	1.71	1.65
29	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Source: Fisher and Yates; *Statistical tables for Biological, Agricultural and Medical Research*, published by Oliver & Boyd Ltd Edinburgh and by permission of the authors and publishers.

Table 5: Critical Values of D in the Kolmogorov-Smirnov Goodness-of-Fit Test

Sample size n	Level of Significance for $D = \max\lvert F_e - F_o\rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.210	.220	.240	.270	.320
30	.190	.200	.220	.240	.290
35	.180	.190	.210	.230	.270
Over 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Note: The values of D given in the table are critical values associated with selected values of n. Any value of D that is greater than or equal to the tabulated value is significant at the indicated level of significance.

Table 6: Critical values of T in the Wilcoxon signed-rank test: $n = 5(1)50$

α One-sided	α Two-sided	n = 5	6	7	8	9	10
.05	.10	1	2	4	6	8	11
.025	.05		1	2	4	6	8
.01	.02			0	2	3	5
.005	.01				0	2	3

α One-sided	α Two-sided	n = 11	12	13	14	15	16
.05	.10	14	17	21	26	30	36
.025	.05	11	14	17	21	25	30
.01	.02	7	10	13	16	20	24
.005	.01	5	7	10	13	16	19

α One-sided	α Two-sided	n = 17	18	19	20	21	22
.05	.10	41	47	54	60	68	75
.025	.05	35	40	46	52	59	66
.01	.02	28	33	38	43	49	56
.005	.01	23	28	32	37	43	49

α One-sided	α Two-sided	n = 23	24	25	26	27	28
.05	.10	83	92	101	110	120	130
.025	.05	73	81	90	98	107	117
.01	.02	62	69	77	85	93	102
.005	.01	55	68	68	76	84	92

Table 6: *Continued*
Critical values of T in the Wilcoxon signed-rank test: $n = 5(1)50$

α One-sided	α Two-sided	n = 29	30	31	32	33	34
.05	.10	141	152	163	175	188	201
.025	.05	127	137	148	159	171	183
.01	.02	111	120	130	141	151	162
.005	.01	100	109	118	128	138	149

α One-sided	α Two-sided	n = 35	36	37	38	39	40
.05	.10	214	228	242	256	271	287
.025	.05	195	208	222	235	250	264
.01	.02	174	186	198	211	224	238
.005	.01	160	171	183	195	208	221

α One-sided	α Two-sided	n = 41	42	43	44	45
.05	.10	303	319	336	353	371
.025	.05	279	295	311	327	344
.01	.02	252	267	281	297	313
.005	.01	234	248	262	277	292

α One-sided	α Two-sided	n = 46	47	48	49	50
.05	.10	389	408	427	446	466
.025	.05	361	379	397	415	434
.01	.02	329	345	362	380	398
.005	.01	307	323	339	356	373

Source: Wilcoxon F. and R. A. Wilcox (1964:28), "Some Rapid Approximate Statistical Procedures," Reproduced with the kind permission of American Cyanamid Company.

Table 7: Distribution function of U

$P(U \le U_0): n_1 \le n_2;\ 3 \le n_1 \le 10$

$n_2 = 3$

$U_0 \backslash n_1$	1	2	3
0	.25	.10	.05
1	.50	.20	.10
2		.40	.20
3		.60	.35
4			.50

$n_2 = 4$

$U_0 \backslash n_1$	1	2	3	4
0	.2000	.0667	.0286	.0143
1	.4000	.1333	.0571	.0286
2	.6000	.2667	.1143	.0571
3		.4000	.2000	.1000
4		.6000	.3143	.1714
5			.4286	.2429
6			.5714	.3429
7				.4429
8				.5571

$n_2 = 5$

$U_0 \backslash n_1$	1	2	3	4	5
0	.1667	.0476	.0179	.0079	.0040
1	.3333	.0952	.0357	.0159	.0079
2	.5000	.1905	.0714	.0317	.0159
3		.2857	.1250	.0556	.0278
4		.4286	.1964	.0952	.0476
5		.5714	.2857	.1429	.0754
6			.3929	.2063	.1111
7			.5000	.2778	.1548
8				.3651	.2103
9				.4524	.2738
10				.5476	.3452
11					.4206
12					.5000

Table 7: *Continued*
Distribution function of U

$P(U \le U_0): n_1 \le n_2;\ 3 \le n_1 \le 10$

$n_2 = 6$

$U_0 \backslash n_1$	1	2	3	4	5	6
0	.1429	.0357	.0119	.0048	.0022	.0011
1	.2857	.0714	.0238	.0095	.0043	.0022
2	.4286	.1429	.0476	.0190	.0087	.0043
3	.5714	.2143	.0833	.0333	.0152	.0076
4		.3214	.1310	.0571	.0260	.0130
5		.4286	.1905	.0857	.0411	.0206
6		.5714	.2738	.1286	.0628	.0325
7			.3571	.1762	.0887	.0465
8			.4524	.2381	.1234	.0660
9			.5476	.3048	.1645	.0898
10				.3810	.2143	.1201
11				.4571	.2684	.1548
12				.5429	.3312	.1970
13					.3961	.2424
14					.4654	.2944
15					.5346	.3496
16						.4091
17						.4686
18						.5314

Table 7: *Continued*
Distribution function of U
$P(U \le U_0): n_1 \le n_2;\ 3 \le n_1 \le 10$

$n_2 = 7$

$U_0 \setminus n_1$	1	2	3	4	5	6	7
0	.1250	.0278	.0083	.0030	.0013	.0006	.0003
1	.2500	.0556	.0167	.0061	.0025	.0012	.0006
2	.3750	.1111	.0333	.0121	.0051	.0023	.0012
3	.5000	.1667	.0583	.0212	.0088	.0044	.0020
4		.2500	.0917	.0364	.0152	.0070	.0035
5		.3333	.1333	.0545	.0240	.0111	.0055
6		.4444	.1917	.0818	.0366	.0175	.0087
7		.5556	.2583	.1152	.0530	.0256	.0131
8			.3333	.1576	.0745	.0367	.0189
9			.4167	.2061	.1010	.0507	.0265
10			.5000	.2636	.1338	.0688	.0364
11				.3242	.1717	.0903	.0487
12				.3939	.2159	.1171	.0641
13				.4636	.2652	.1474	.0825
14				.5364	.3194	.1830	.1043
15					.3775	.2226	.1297
16					.4381	.2669	.1588
17					.5000	.3141	.1914
18						.3654	.2279
19						.4178	.2675
20						.4726	.3100
21						.5274	.3552
22							.4024
23							.4508
24							.5000

Table 7: *Continued*
Distribution function of U
$P(U \le U_0): n_1 \le n_2;\ 3 \le n_1 \le 10$

$n_2 = 8$

$U_0 \setminus n_1$	1	2	3	4	5	6	7	8
0	.1111	.0222	.0061	.0020	.0008	.0003	.0002	.0001
1	.2222	.0444	.0121	.0040	.0016	.0007	.0003	.0002
2	.3333	.0889	.0242	.0081	.0031	.0013	.0006	.0003
3	.4444	.1333	.0424	.0141	.0054	.0023	.0011	.0005
4	.5556	.2000	.0667	.0242	.0093	.0040	.0019	.0009
5		.2667	.0970	.0364	.0148	.0063	.0030	.0015
6		.3556	.1394	.0545	.0225	.0100	.0047	.0023
7		.4444	.1879	.0768	.0326	.0147	.0070	.0035
8		.5556	.2485	.1071	.0466	.0213	.0103	.0052
9			.3152	.1414	.0637	.0296	.0145	.0074
10			.3879	.1838	.0855	.0406	.0200	.0103
11			.4606	.2303	.1111	.0539	.0270	.0141
12			.5394	.2848	.1422	.0709	.0361	.0190
13				.3414	.1772	.0906	.0469	.0249
14				.4040	.2176	.1142	.0603	.0325
15				.4667	.2618	.1412	.0760	.0415
16				.5333	.3108	.1725	.0946	.0524
17					.3621	.2068	.1159	.0652
18					.4165	.2454	.1405	.0803
19					.4716	.2864	.1678	.0974
20					.5284	.3310	.1984	.1172
21						.3773	.2317	.1393
22						.4259	.2679	.1641
23						.4749	.3063	.1911
24						.5251	.3472	.2209
25							.3894	.2527
26							.4333	.2869
27							.4775	.3227
28							.5225	.3605
29								.3992
30								.4392
31								.4796
32								.5204

Table 7: *Continued*

Distribution function of U

$$P(U \leq U_0): n_1 \leq n_2;\ 3 \leq n_1 \leq 10$$

$$n_2 = 9$$

$U_0 \backslash n_1$	1	2	3	4	5	6	7	8	9
0	.1	.0182	.0045	.0014	.0005	.0002	.0001	.0000	.0000
1	.2	.0364	.0091	.0028	.0010	.0004	.0002	.0001	.0000
2	.3	.0727	.0182	.0056	.0020	.0008	.0003	.0002	.0001
3	.4	.1091	.0318	.0098	.0035	.0014	.0006	.0003	.0001
4	.5	.1636	.0500	.0168	.0060	.0024	.0010	.0005	.0002
5		.2182	.0727	.0252	.0095	.0038	.0017	.0008	.0004
6		.2909	.1045	.0378	.0145	.0060	.0026	.0012	.0006
7		.3636	.1409	.0531	.0210	.0088	.0039	.0019	.0009
8		.4545	.1864	.0741	.0300	.0128	.0058	.0028	.0014
9		.5455	.2409	.0993	.0415	.0180	.0082	.0039	.0020
10			.3000	.1301	.0559	.0248	.0115	.0056	.0028
11			.3636	.1650	.0734	.0332	.0156	.0076	.0039
12			.4318	.2070	.0949	.0440	.0209	.0103	.0053
13			.5000	.2517	.1199	.0567	.0274	.0137	.0071
14				.3021	.1489	.0723	.0356	.0180	.0094
15				.3552	.1818	.0905	.0454	.0232	.0122
16				.4126	.2488	.1119	.0571	.0296	.0157
17				.4699	.2592	.1361	.0708	.0372	.0200
18				.5301	.3032	.1638	.0869	.0464	.0252
19					.3497	.1942	.1052	.0570	.0313
20					.3986	.2280	.1261	.0694	.0385
21					.4491	.2643	.1496	.0836	.0470
22					.5000	.3035	.1755	.0998	.0567
23						.3445	.2039	.1179	.0680
24						.3878	.2349	.1383	.0807
25						.4320	.2680	.1606	.0951
26						.4773	.3032	.1852	.1112
27						.5227	.3403	.2117	.1290
28							.3788	.2404	.1487
29							.4185	.2707	.1701
30							.4591	.3029	.1933
31							.5000	.3365	.2181
32								.3715	.2447
33								.4074	.2729
34								.4442	.3024
35								.4813	.3332
36								.5187	.3652
37									.3981
38									.4317
39									.4657
40									.5000

Table 7: *Continued*

Distribution function of U

$$P(U \leq U_0): n_1 \leq n_2;\ 3 \leq n_1 \leq 10$$

$$n_2 = 10$$

$U_0 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
0	.0909	.0152	.0035	.0010	.0003	.0001	.0001	.0000	.0000	.0000
1	.1818	.0303	.0070	.0020	.0007	.0002	.0001	.0000	.0000	.0000
2	.2727	.0606	.0140	.0040	.0013	.0005	.0002	.0001	.0000	.0000
3	.3636	.0909	.0245	.0070	.0023	.0009	.0004	.0002	.0001	.0000
4	.4545	.1364	.0385	.0120	.0040	.0015	.0006	.0003	.0001	.0001
5	.5455	.1818	.0559	.0180	.0063	.0024	.0010	.0004	.0002	.0001
6		.2424	.0804	.0270	.0097	.0037	.0015	.0007	.0003	.0002
7		.3030	.1084	.0380	.0140	.0055	.0023	.0010	.0005	.0002
8		.3788	.1434	.0529	.0200	.0080	.0034	.0015	.0007	.0004
9		.4545	.1853	.0709	.0276	.0112	.0048	.0022	.0011	.0005
10		.5455	.2343	.0939	.0376	.0156	.0068	.0031	.0015	.0008
11			.2867	.1199	.0496	.0210	.0093	.0043	.0021	.0010
12			.3462	.1518	.0646	.0280	.0125	.0058	.0028	.0014
13			.4056	.1868	.0823	.0363	.0165	.0078	.0038	.0019
14			.4685	.2268	.1032	.0467	.0215	.0103	.0051	.0026
15			.5315	.2697	.1272	.0589	.0277	.0133	.0066	.0034
16				.3177	.1548	.0736	.0351	.0171	.0086	.0045
17				.3666	.1855	.0903	.0439	.0217	.0110	.0057
18				.4196	.2198	.1099	.0544	.0273	.0140	.0073
19				.4725	.2567	.1317	.0665	.0338	.0175	.0093
20				.5275	.2970	.1566	.0806	.0416	.0217	.0116
21					.3393	.1838	.0966	.0506	.0267	.0144
22					.3839	.2139	.1148	.0610	.0326	.0177
23					.4296	.2461	.1349	.0729	.0394	.0216
24					.4765	.2811	.1574	.0864	.0474	.0262
25					.5235	.3177	.1819	.1015	.0564	.0315
26						.3564	.2087	.1185	.0667	.0376
27						.3962	.2374	.1371	.0782	.0446
28						.4374	.2681	.1577	.0912	.0526
29						.4789	.3004	.1800	.1055	.0615
30						.5211	.3345	.2041	.1214	.0716

Table 7: *Continued*
Distribution function of U
$P(U \le U_0): n_1 \le n_2;\ 3 \le n_1 \le 10$

$n_2 = 10$ *continued*

$U_0 \setminus n_1$	1	2	3	4	5	6	7	8	9	10
31							.3698	.2299	.1388	.0827
32							.4063	.2574	.1577	.0952
33							.4434	.2863	.1781	.1088
34							.4811	.3167	.2001	.1237
35							.5189	.3482	.2235	.1399
36								.3809	.2483	.1575
37								.4143	.2745	.1763
38								.4484	.3019	.1965
39								.4827	.3304	.2179
40								.5173	.3598	.2406
41									.3901	.2644
42									.4211	.2894
43									.4524	.3153
44									.4841	.3421
45									.5159	.3697
46										.3980
47										.4267
48										.4559
49										.4853
50										.5147

Table 8: Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test: Independent Samples

(a) $\alpha = .025$ one-tailed; $\alpha = .05$ two-tailed

n_2	3		4		5		6	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23
4	6	18	11	25	12	28	12	32
5	6	21	12	28	18	37	19	41
6	7	23	12	32	19	41	26	52
7	7	26	13	35	20	45	28	56
8	8	28	14	38	21	49	29	61
9	8	31	15	41	22	53	31	65
10	9	33	16	44	24	56	32	70

n_2	7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	7	26	8	28	8	31	9	33
4	13	35	14	38	15	41	16	44
5	20	45	21	49	22	53	24	56
6	28	56	29	61	31	65	32	70
7	37	68	39	73	41	78	43	83
8	39	73	49	87	51	93	54	98
9	41	78	51	93	63	108	66	114
10	43	83	54	98	66	114	79	131

Table 8: *Continued*

Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test: Independent Samples

(b) $\alpha = .05$ one-tailed; $\alpha = .10$ two-tailed

n_2	3		4		5		6	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22
4	7	17	12	24	13	27	14	30
5	7	20	13	27	19	36	20	40
6	8	22	14	30	20	40	28	50
7	9	24	15	33	22	43	30	54
8	9	27	16	36	24	46	32	58
9	10	29	17	39	25	50	33	63
10	11	31	18	42	26	54	35	67

n_2	7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	9	24	9	27	10	29	11	31
4	15	33	16	36	17	39	18	42
5	22	43	24	46	25	50	26	54
6	30	54	32	58	33	63	35	67
7	39	66	41	71	43	76	46	80
8	41	71	52	84	54	90	57	95
9	43	76	54	90	66	105	69	111
10	46	80	57	95	69	111	83	127

Source: Wilcoxon F. and R. A. Wilcox (1964:28), "Some rapid Approximate Statistical Procedures," Reproduced with the kind permission of American Cyanamid Company.

Table 9: Distribution of the number of runs R in samples of size $(n_1, n_2) : P(R \le a)$

(n_1, n_2)	*a*: 2	3	4	5	6	7	8	9	10
(2, 3)	.200	.500	.500	1.00					
(2, 4)	.133	.400	.800	1.00					
(2, 5)	.095	.333	.714	1.00					
(2, 6)	.071	.286	.643	1.00					
(2, 7)	.056	.250	.583	1.00					
(2, 8)	.044	.222	.533	1.00					
(2, 9)	.036	.200	.491	1.00					
(2, 10)	.030	.182	.455	1.00					
(3, 3)	.100	.300	.700	.900	1.00				
(3, 4)	.057	.200	.543	.800	.971	1.00			
(3, 5)	.036	.143	.429	.714	.929	1.00			
(3, 6)	.024	.107	.345	.643	.881	1.00			
(3, 7)	.017	.083	.283	.583	.833	1.00			
(3, 8)	.012	.067	.236	.533	.788	1.00			
(3, 9)	.009	.055	.200	.491	.745	1.00			
(3, 10)	.007	.045	.171	.455	.706	1.00			
(4, 4)	.029	.111	.371	.629	.886	.971	1.00		
(4, 5)	.016	.071	.262	.500	.786	.929	.992	1.00	
(4, 6)	.010	.048	.190	.405	.690	.881	.954	1.00	
(4, 7)	.006	.033	.142	.333	.606	.833	.954	1.00	
(4, 8)	.004	.024	.109	.279	.533	.788	.929	1.00	
(4, 9)	.003	.018	.085	.236	.471	.745	.902	1.00	
(4, 10)	.002	.014	.068	.203	.419	.706	.874	1.00	
(5, 5)	.008	.040	.167	.357	.643	.833	.960	.992	1.00
(5, 6)	.004	.024	.110	.262	.522	.738	.911	.976	.998
(5, 7)	.003	.015	.076	.197	.424	.652	.854	.955	.992
(5, 8)	.002	.010	.054	.152	.347	.576	.793	.929	.984
(5, 9)	.001	.007	.039	.119	.287	.510	.734	.902	.972
(5, 10)	.001	.005	.029	.095	.239	.455	.678	.874	.958

Table 9: *Continued*

Distribution of the number of runs R in samples of size $(n_1, n_2): P(R \le a)$

(n_1, n_2)	a = 2	3	4	5	6	7	8	9	10
(6, 6)	.002	.013	.067	.175	.392	.608	.825	.933	.987
(6, 7)	.001	.008	.043	.121	.296	.500	.733	.879	.966
(6, 8)	.001	.005	.028	.086	.226	.413	.646	.821	.937
(6, 9)	.000	.003	.019	.063	.175	.343	.566	.762	.902
(6, 10)	.000	.002	.013	.047	.137	.288	.497	.706	.864
(7, 7)	.001	.004	.025	.078	.209	.383	.617	.791	.922
(7, 8)	.000	.002	.015	.051	.149	.296	.514	.704	.867
(7, 9)	.000	.001	.010	.035	.108	.231	.427	.622	.806
(7, 10)	.000	.001	.006	.024	.080	.182	.355	.549	.743
(8, 8)	.000	.001	.009	.032	.100	.214	.405	.595	.786
(8, 9)	.000	.001	.005	.020	.069	.157	.319	.500	.702
(8, 10)	.000	.000	.003	.013	.048	.117	.251	.419	.621
(9, 9)	.000	.000	.003	.012	.044	.109	.238	.399	.601
(9, 10)	.000	.000	.002	.008	.029	.077	.179	.319	.510
(10, 10)	.000	.000	.001	.004	.019	.051	.128	.242	.414

Table 9: *Continued*

Distribution of the number of runs R in samples of size $(n_1, n_2): P(R \le a)$

(n_1, n_2)	a = 11	12	13	14	15	16	17	18	19	20
(4, 4)										
(4, 5)										
(4, 6)										
(4, 7)										
(4, 8)										
(4, 9)										
(4, 10)										
(5, 5)										
(5, 6)	1.00									
(5, 7)	1.00									
(5, 8)	1.00									
(5, 9)	1.00									
(5, 10)	1.00									
(6, 6)	.998	1.00								
(6, 7)	.992	.999	1.00							
(6, 8)	.984	.998	1.00							
(6, 9)	.972	.994	1.00							
(6, 10)	.958	.990	1.00							
(7, 7)	.975	.996	.999	1.00						
(7, 8)	.949	.988	.998	1.00	1.00					
(7, 9)	.916	.975	.994	.999	1.00					
(7, 10)	.879	.957	.990	.998	1.00					
(8, 8)	.900	.968	.991	.999	1.00	1.00				
(8, 9)	.843	.939	.980	.996	.999	1.00				
(8, 10)	.782	.903	.964	.990	.998	1.00	1.00			
(9, 9)	.762	.891	.956	.988	.997	1.00	1.00	1.0		
(9, 10)	.681	.834	.923	.974	.992	.999	1.00	1.0	1.0	
(10, 10)	.586	.758	.872	.949	.981	.996	.999	1.0	1.0	1.0

From "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," F. Swed and C. Eisenhart. *Annals of Mathematical Statistics*, Volume 14(1943). Reproduced with the kind permission of the Authors and of the Editor.

Table 10: Critical values of Spearman's rank correlation coefficient

$n \setminus \alpha$	.05	.025	.01	.005
5	0.900	-	-	-
6	0.829	0.886	0.943	-
7	0.714	0.786	0.893	-
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

From "Distribution of Sums of Squares of Rank Differences for Small Samples", E. G. Olds, *Annals of Mathematical Statistics*. Volume 9(1938). Reproduced with the kind permission of the Editor, *Annals of Mathematical Statistics*.

SELECTED REFERENCES

Berenson, Mark L., David M. Levine and Mathew Goldstein (1983) *Intermediate Statistical Methods and Applications: A Computer Package Approach*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Blalock, Hubert (1972) *Social Statistics*, International Student Edition, Kokakusha Ltd, Tokyo: McGraw-Hill.

Canover, W. J. (1980); *Practical Nonparametric Statistics*, Second edition, New York: John Wiley & Sons.

Freund, John E. and Garry A. Simon (1992), *Modern Elementary Statistics*, 8th Edition, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Gummesson, Evert (2000) *Qualitative Methods in Management Research*, 2nd Edition, Beverly Hills, California: Sage Publications.

Kendall, M. G. and A Stuart (1987) *The Advanced Theory of Statistics*, Fifth edition, London: Charles Griffin and Company Limited.

Menard, Scott (1995) "Applied Logistic Regression Analysis", *Quantitative Applications in the Social Sciences*, Sage University Paper, No. 106.

Mendenhall, William and James E. Reinmuth (1982) *Statistics for Management and Economics*, Fourth edition, Belmont, California: PWS Publishers, Duxbury Press.

Mihanjo, Adolf (2005), *Falsafa na Ufunuo wa Maarifa*, Morogoro: Salvatorian Institute of Philosophy and Theology.

Mood, Alexander M., Franklin A Craybill and Duan C Boes (1974) *Introduction to the Theory of Statistics*, Third edition, Kokakusha Ltd, Tokyo: Mc Graw-Hill.

Ndunguru, P. C (2001) *Basic Concepts of Probability Theory*, Mzumbe University: Research and Publications.

Ndunguru, P. C. (2007) *Econometrics: A Science of Non-experimental Data Analysis*, Mzumbe University: Research and Publications.

Ndunguru, P. C. (2007) *Lectures on Research Methodology for Social Sciences*, Mzumbe University: Research and Publications.

Pierce, A. (1970); *Fundamentals of Nonparametric Statistics*, Belmont, California: Dichenson.

Rees, D.G. (1987) *Foundations of Statistics*, London: Chapman and Hall.

Stigler, S. M., (1986) *The History of Statistics*, Cambridge, Massachusetts: Harvard University Press.

Triola, Mario F. (1998), *Elementary Statistics*, Seventh edition, Reading, Massachusetts: Addison-Wesley Longman, Inc..

Wilks, S. S. (1962) *Mathematical Statistics*, New York: John Wiley & Sons.

INDEX

Inference problems in qualitative research focus on theory-building that emphasizes on describing and determining relationships among study variables on the basis of data rather than reviewed theories. In testing descriptive and relational hypotheses, theory-building starts with identifying a difference in kind and/or in degree regarding presence or absence of a property or set of properties in study objects or events. Both aspects of difference are within the realm of measurement levels that are associated with classification and/or ordering. These measurement levels lend themselves to non-parametric statistical analysis. Non-parametric Statistics: Inference Methods for Qualitative Research, is a book that is solely dedicated to this specific theme, focusing mainly on business and management research. The inclusion of SPSS tutorial sections in solving problems makes the book a unique one, enabling readers to practise their skills in applying the statistical package.

Professor Philibert Caspar Ndunguru has B.A. (Hons) Degree in Statistics and Econometrics from University of Dar es Salaam, and an M.B.A Degree in Financial Economics, Econometrics, and Business Management from Katholieke Universiteit Leuven in Belgium, and a Ph.D. (Business Administration) Degree from University of Dar es Salaam. He is lecturing on Statistics, Investment and Portfolio Theory, Econometrics, Business Policy, and Social Science Research Methods at Mzumbe University. Professor Ndunguru has also lectured on Biometry and Econometrics at Sokoine University of Agriculture. Other published books by Professor Ndunguru include: Corporate Strategic Planning: A Guide for Managers in Tanzania (1999), Basic Concepts of Probability Theory (2001), Lectures on Research Methodology for Social Sciences (2007) and Econometrics: A Science for Non-experimental Data Analysis (2007).