Chapter 2

Analytical Methods in Social Science

Statistics is a branch of science which deals with the collection, classification, description and interpretation of data obtained by conducting surveys and experiments. The essential purpose of it is to describe and draw valid inferences about numerical properties of populations" (Ferguson 1966). The experience in statistics application indicates that a single statistical method can be used in different research areas for dealing with different types of problems. Various possibilities which application of a statistical method provides in certain research areas should be considered as an adequate way of dealing with a problem of the research. However, it should be noted that statistics is not a method by which it is possible to solve all the problems in a research. Nachmias and Nachmias (2009) stated that there are basically two types of statistics which include descriptive and inferential statistics. Descriptive statistics enable the researcher to summarize and organize data in an effective and meaningful way. It involves the use of tables, charts, graphs, mean, modes, median, standard scores and correlation to treat collected data. Inferential statistics is concerned with making inferences from a unit of a population. Inferential

statistics is concerned with making inferences from a unit of a population. Inferential statistics allow the researcher to make decisions or inferences by interpreting data patterns. Researchers use inferential statistics to determine whether an expected pattern designated by the theory and hypotheses is actually found in the observations. To decide whether this hypothesis is true, researchers might survey the respondents and then use descriptive statistics to make comparison between these groups and would employ inferential statistics to determine whether the differences between the groups are significant or not.

Social research is a type of research conducted by researchers following a systematic plan. Social research methodologies can be classified along a quantitative/qualitative dimension. Quantitative designs approach social phenomena through quantifiable evidence, and often rely on statistical analysis of many cases (or across intentionally designed treatments in an experiment) to create valid and reliable general claims. Qualitative designs emphasize understanding of social phenomena through direct observation, communication with participants, or analysis of texts, and may stress contextual subjective accuracy over generality. When applying a statistical method, it is common to differentiate between quantitative and qualitative features and variables. Nominal and ordinal variables and data are usually considered as qualitative (attributive), while interval variables and ratio variables are considered as quantitative (Ferguson 1966, Krneta 1987). Also, it is common to apply nonparametric statistical methods on nominal and ordinal data, while parametric methods are used for the interval and ratio data (Ferguson 1966).

Statistics and statistical methods have highly significant application in sociology. Functions of statistics are numerous: the methods of descriptive statistics have an important application for describing natural phenomena; inferential statistics is used for inductive reasoning about unknown properties of a larger group using the known indicators of the causes; hypothesis testing most frequently refers to the results of one, two or more causes, on the basis of which it is possible to draw conclusions on the problem of the research, by accepting or refuting an initial hypothesis; regression and correlation analysis, in the most simple case, examines the influence and dependence between two or more variables.

2.1 Parametric Methods/Tests:

The two classes of statistical tests are called parametric and nonparametric. The word parametric comes from "metric," meaning to measure, and "para," meaning beside or closely related; the combined term refers to the assumptions about the population from which the measurements were obtained. Nonparametric data do not meet such rigid assumptions. Nonparametric tests sometimes are referred to as "distribution-free." That is, the data can be drawn from a sample that may not follow the normal distribution.

Before a parametric test can be undertaken, it must be ascertained that: 1) the samples are random (i.e., each member of the population has an equal chance of being selected for measurement); 2) the scores are independent of each other; 3) the experiments are repeatable with constancy of measurements from experiment to experiment; 4) the data are normally distributed; and 5) the samples have similar variances (1, 2). Parametric

statistics use mean values, standard deviation and variance to estimate differences between measurements that characterize particular populations.

The two major types of parametric tests are Student's t-tests and analyses of variance (ANOVA). Nonparametric tests use rank or frequency information to draw conclusions about differences between populations. Parametric tests usually are assumed to be more powerful than nonparametric tests. However, parametric tests cannot always be used to analyze the significance of differences because the assumptions on which they are based are not always met.

2.1.1 Basic concepts of Testing of Hypothesis:

a) Hypotheses

A *hypothesis* is a statement about the parameters of the underlying distribution of the observations (specified in the model). If parameters are unknown for any test, we do not know whether a hypothesis is true or false. But data are evidence, and we might have enough evidence against a hypothesis to allow us to reject it.

Any testing involves two hypotheses: the *null hypothesis*, H_0, and *the alternative hypothesis*, H_1. The null hypothesis is protected: it will only be rejected if there is strong evidence against it in favour of H_1.The alternative is only a benchmark to test against. The null hypothesis is what you want to prove false. The result of a statistical investigation is given in terms of the null hypothesis: if there is insufficient evidence against H_0 we say "we do not reject H_0" or "we accept H_0"; otherwise we say "we can reject H_0".

Examples:

- (Biological researcher) H_0 : $\theta = 0.3$, H_1 : $\theta > 0.3$
- (Weight observer) H_0 : $\mu = 5$kg, H_1 : $\mu < 5$kg

b) Errors

A type I error occurs if we reject H_0 when it is true. This is a serious error. The probability of making a type I error should be low. A *type II error* occurs if we fail to reject H_0 when it is in fact false. This is a less serious error.

c) Test statistics and critical regions

A test consists of a test statistic, such as the sample mean, sample variance or something more complicated and a critical region, specifying which values of the test statistic will result in rejection of H_0.

Examples:

- (Biological researcher) Reject H_0 if more than half the patients recover
- (Weight observer) Reject H_0 if the sample mean weight loss is no more than 3.1kg.
- Biased coin.

d) Significance and p-values

The distribution of the test statistic (the *reference distribution*) is determined by the model; the parameter value is specified by the hypothesis. So we can find the cut off point for the critical region which gives a type I error probability of exactly 5%, say.

If the observed value of the test statistic falls in this critical region, we say it is *significant* and then we reject H_0 *at the 5% level*.

We can do the same for 1% (*highly significant*) and 0.1% (*very highly significant*) as well. The p-value measures the significance level at which the observed value of the test statistic first becomes significant. In other words, it answers the question "How unlikely is this result if H_0 is true?"

e) One-sided and two-sided tests

Whether a test is one-or two-sided depends on the alternative H_1. An alternative involving ">" or "<" is one-sided; one involving both is two-sided. (The words *one-tailed* and *two-tailed* are also used).The critical region for a two-sided test at 5% significance level involves 2½% in each tail of the distribution. For this reason, when you are dealing with a two-sided test, the p-value which you calculate from one-sided tables must be doubled.

2.1.2 Different types of Parametric Methods/Tests:

a) Testing the mean when the variance is known

Since the sample mean has $N(\mu, \delta^2/n)$ distribution, we can define

$$Z=\frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}}$$

and state that $Z \sim N(0,1)$.Now δ is known, and μ is specified by H_0, so we can evaluate Z from the data *under the assumption that H_0 is true*.

b) Testing the mean when the variance is unknown

The value of *Z* used above still has N (0, 1) distribution. But it is no longer a statistic, as s is unknown. We must use $S^2$to estimate s^2 to give:

$$T = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

T is not Normal; it has a *Student's t* distribution on *n* - 1 degrees of freedom. The testing procedure is as above, but using *T* instead of *Z* and finding the cut off points of the critical regions in t tables instead of Normal tables.

Large sample approximation

When *n* is large the t distribution approaches the Normal. t tables usually stop at 120 degrees of freedom; for $n > 120$ it is acceptable to use Normal instead of t. Some books advocate a lower cut off point, such as 30 d.f.

Degrees of freedom

The term "degrees of freedom" makes sense in the context of testing goodness of fit. Loosely, start with *n* d.f., then lose 1 for each unknown parameter estimated in the model for the mean.

c) Testing the variance

If $Z_1, .., Z_n$ are independent standard Normal variables, then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

χ_n^2 is the *chi-squared* distribution on *n* degrees of freedom.

Therefore, for $X_1 ... X_n \sim N(\mu, \delta^2)$,

$$\sum \frac{(X_i - \mu)^2}{\delta^2} \sim \chi^2_n$$

If μ is known, we can use a test based on χ^2_n . Usually μ is unknown and must be estimated.

$$\frac{(X_i - \bar{X})^2}{\delta^2} \sim \lambda^2_{n-1}$$

Note that the LHS is $\frac{(n-1)S^2}{\delta^2}$ .In addition; sample variance and sample mean are independent random variables. To test H_0: δ^2=16 (say) against H_1: δ^2>16, calculate (n-1) S^2/16 and compare with λ^2_{n-1} tables: if the statistic is too large, reject the null hypothesis.

Robustness

All tests are based on a model. Some work well even when the assumptions of the model are not satisfied (eg, underlying distribution is not really Normal): they are *robust tests*. The t-test is robust. The test based on λ^2 is not.

d) Comparative tests based on two Normal samples

We assume that there are two independent simple random samples X_1.., $X_n \sim N(\mu_x, \delta_x^2)$and Y_1, .., $Y_m \sim N(\mu_y, \delta_y^2)$.We may want to test whether $\mu_x = \mu_y$ or whether $\delta_x^2 = \delta_y^2$

i) Two-sample t test

Here we test H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$ (or a one-sided alternative).

The test presented only works if $\delta_x^2 = \delta_y^2$. (Later we see how to test this.) Let δ^2 denote the common variance. Then, if H_0 is true,

$$\bar{X}\text{-}\bar{Y} \sim N(0, \delta^2 (\frac{1}{n} + \frac{1}{m}))$$

which allows us to deduce that $Z \sim N(0,1)$. In general δ is unknown and must be estimated. We use a combination of S^2_X and S^2_Y, called the *pooled estimate of the variance*,

$$S^2_p = \frac{(n\text{-}1)S^2_X + (m\text{-}1)S^2_Y}{n+m\text{-}2}$$

Now we can say that $t \sim t_{m+n-2}$ and use a t test.

ii) Paired sample test

Here the two samples have the same size and there is a natural correspondence between individual observations from the two samples. For example, Before and After. In these cases we calculate the differences, After - Before, or the ratios, After/Before, and treat that as a single sample. A t test will usually be appropriate.

e) Testing equality of variance

We test H_0: $\delta_x^2 = \delta_y^2$ against H_1: $\delta_x^2 \neq \delta_y^2$ (two-sided alternative).The test used is the F test. If H_0 is true, then the ratio S_X^2 / S_Y^2 has an F distribution with (n-1, m-1) degrees of freedom.

It is equally true to say that $S_Y^2 / S_X^2 \sim F_{m-1,\, n-1}$.Because of the way the F tables are organized, we always put the largest sample variance on top.

What if variances aren't equal?

Various possibilities: Welch test (distribution is approximately t), Behrens' test (exact, but tables are hard to read), or a non-parametric test such as the Wilcoxon rank sum test (see later).

f) Tests based on Binomial samples

- ***Single binomial sample***

Example: A course review claims that on average at least 60% of the students end up with a first or upper second. Of the 60 graduates in the last 3 years, 32 have been awarded firsts or II-1s. Does this contradict the claims of the review?

Basic idea as before: specify a model, write down the hypotheses, find a test statistic, calculate its distribution under H_0, and look at the observed value to see how unusual it is. In the example the model is that each student, independently, gets a first or II-1 with probability θ.Hypotheses are H_0: θ= 0.6, H_1: $\theta < 0.6$. Test statistic, X, is the number of students who got firsts and II-1s.If H_0 is true, $X \sim$ Bin (60, 0.6).This is roughly N (36, 14.4), so define

$$Z = \frac{X - 36}{\sqrt{14.4}} = -1.054.$$

Not significant.

- ***Comparison of two binomial samples***

Here we are testing equality of proportions. The model is that each of the n_X units in the first sample has probability θ_x of possessing a particular property; each of the n_Y in the second sample has probability θ_y .Usually H_0 is that $\theta_x = \theta_y$ with a 1-sided or 2-sided alternative.

If H_0 is true, then

$$\frac{X}{n_X} - \frac{Y}{n_Y} \text{ has mean 0,}$$

$$\text{variance } \theta(1 - \theta)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right),$$

where θ is the common value of θ_X and θ_Y.But θ is unknown, so must be estimated using both samples: we use

$$\frac{X + Y}{n_X + n_Y}$$

With this estimate of θ we employ a Normal approximation. Tests using Normal approximations can be used in more cases than just Binomial, though they are not the "best" tests.

A statistical test can never establish the truth of a hypothesis with 100-percent certainty. Typically, the hypothesis is specified in the form of a "null hypothesis," i.e., the score characterizing one group of measurements does not differ (within an acceptable margin of measurement error) from the score characterizing another group. Note the hypothesis does not state the two scores are the same; rather, it states no significant difference can be detected. Performing the statistical procedure yields a test result that helps one reach a decision that the scores are not different (the hypothesis is confirmed) or the difference in the scores is too great to be explained by chance (the hypothesis is rejected).Rejecting the hypothesis when it actually is true is called a Type-I error. Failure to reject the hypothesis when it is false is termed a Type-II error. For convenience and simplicity, a 5-percent risk of making a Type-I error has become conventional; one should be

correct 95 out of 100 times when using the listed value in the probability tables to accept or reject the hypothesis. In statistics, robustness is the degree to which a test can stray from the assumptions before changing the confidence you have in the result of the statistical test you have used. Choosing a nonparametric test trades the power, or large sample size, of the parametric test for robustness. Further, a method requiring few and weak assumptions about the population(s) being sampled is less dependable than the corresponding parametric method and increases the chances of committing a Type-Il error.

2.1.3 Independence or Dependence of Samples

Independence or dependence of samples concerns whether the different sets of numbers being compared are independent or dependent of each other. Sets are independent when values in one set tell nothing about values in another set. When two or more groups consist of different, unrelated individuals, the observations made about the samples are independent. When the sets of numbers consist of repeated measures on the same individuals, they are said to be dependent. Similarly, if male and female characteristics are compared using brother-sister pairs, the samples are dependent. Matching two or more groups of individuals on factors such as income, education, age, height and weight also yields dependent samples.

2.1.4 Random Selection from Normally Distributed Populations

This type of selection sometimes is difficult to confirm. However, so long as the data sets used in the analysis are relatively normally distributed, the

robustness of most parametric tests still provides an appropriate level of rejection of the null hypothesis.

2.1.5 Homogeneity of Variance

Homogeneity of variance of the data from each group being compared must be equal (homogeneous) and can be tested statistically. If it is found to differ significantly, then nonparametric tests must be used.

Parametric tests require data from which means and variances can be calculated, i.e., interval and ratio data. Some statisticians also support the use of parametric tests with ordinal-scaled values because the distribution of ordinal data often is approximately normal. As long as the actual data meet the parametric assumptions, regardless of the origin of the numbers, then parametric tests can be conducted. As is the case with all statistical tests of differences, the researcher must interpret parametric statistical conclusions based on ordinal data in light of their clinical or practical implications.

2.2 Non Parametric Methods:

Besides the Parametric methods, another method commonly used in statistics to model and analyze ordinal or nominal data with small sample sizes. Unlike parametric models, nonparametric models do not require the modeler to make any assumptions about the distribution of the population, and so are sometimes referred to as a distribution-free method.

2.2.1 Uses of Nonparametric Tests:

Nonparametric tests are used in the behavioural sciences when there is no basis for assuming certain types of distributions. Siegel has advocated that nonparametric tests be used for nominal and ordinal levels of measurements while parametric tests are used for analyzing interval and ratio data.

On the other hand, Williamsen has argued that statistical tests are selected to meet certain goals or to answer specific questions rather than to match certain levels of measurement with parametric or nonparametric procedures. This view currently is prevalent among many statisticians.

In practice, levels of measurement sometimes are "downgraded" from ratio and interval scales to ordinal or nominal scales for the convenience of a measuring instrument or interpretation. For example, muscular strength (measured with a force gauge and considering the length of the lever arm through which the force is acting) is a variable that yields ratio data because a true zero point exists in the level of measurement. Muscular strength is absent with paralysis (true zero point). The manual muscle test converts the ratio characteristic of force into an ordinal scale by assigning grades of relative position (normal, good, fair, poor, trace, zero; or 5,4, 3,2, 1,0).

Nonparametric tests should not be substituted for parametric tests when parametric tests are more appropriate. Nonparametric tests should be used when the assumptions of parametric tests cannot be met, when very small numbers of data are used, and when no basis exists for assuming certain types or shapes of distributions.

Nonparametric tests are used if data can only be classified, counted or ordered-for example, rating staff on performance or comparing results from manual muscle tests. These tests should not be used in determining precision or accuracy of instruments because the tests are lacking in both areas.

2.2.2 Advantages

Nonparametric tests usually can be performed quickly and easily without automated instruments (calculators and computers). They are designed for small numbers of data, including counts, classifications and ratings. They are easier to understand and explain.

Calculations of nonparametric tests generally are easy to perform and apply, and they have certain intuitive appeal as shortcut techniques. Nonparametric tests are relatively robust and can be used effectively for determining relationships and significance of differences using behavioral research methods.

2.2.3 Disadvantages

Parametric tests are more powerful than nonparametric tests and deal with continuous variables whereas nonparametric tests often deal with discrete variables. Using results from analyses of nonparametric tests for making inferences should be done with caution because small numbers of data are used, and no assumptions about parent populations are made. The ease of calculation and reduced concern for assumptions have been referred to as "quick and dirty statistical procedures”.

2.2.4 Nonparametric Methods for Descriptive Statistics

Descriptive statistics involve tabulating, depicting and describing collections of data. These data may be either quantitative, such as measures of leg length (variables that are characterized by an underlying continuum) or representative of qualitative variables, such as gender, vocational status or personality type.

Collections of data generally must be summarized in some fashion to be more easily understood. Descriptive statistics serve as the means to describe, summarize and reduce to manageable form the properties of an otherwise unwieldy mass of data. Descriptive statistics used to characterize data analysed by parametric tests include the mean, standard deviation and variance. Those descriptive statistics used to characterize data analyzed by nonparametric tests include the mode, median and percentile rank:

- Mode is the most frequently occurring score within a distribution. It is possible to have a sample of scores that has no mode, and it is possible to have two or more modes; if there are two modes, the distribution is called bimodal.
- Median is the middle score of a ranked distribution of measurements.
- Percentile or percentile rank of an observation is the percentage of a distribution that falls below that observation. Percentile rank is calculated with below Equation:

$$PR_i = 100\left(1 - \frac{R_i - 0.5}{n}\right)$$

where R_i is the rank of the observation X_i (ranked from highest to lowest), and *n* is the number of observations in the distribution. The median is the 50th percentile.

In statistics, the mean or median commonly is used when dealing with measurement data. The mode most often is useful when dealing with data more appropriately handled with classification procedures (e.g., mild, moderate, severe).

ISBN: 978-93-85822-53-7

2.2.5 Measures of Association (Correlation)

Correlation coefficients are used to reveal the nature and extent of association between two variables. Each method used to determine a correlation coefficient has conditions that must be met for its use to be appropriate. The first step in analyzing a relationship always is selection of the proper measure of association based on the conditions of the study and the hypothesis to be tested.

Measures of association are useful for a variety of studies. Correlation coefficients are used in exploratory studies to determine relationships among variables in new study areas. The results of such studies allow investigators to formulate further research questions or hypotheses to delve more deeply into the study area. In some studies, the hypotheses focus on associations between selected variables, and the correlation coefficients serve to test these hypotheses.

Similarly, hypotheses based on expected associations among variables make important contributions to theory building.

Finally, correlation coefficients are used to manage threats to validity in experimental and quasi-experimental studies. They can be used to test the credibility of findings when groups have been compared by checking on the association of independent and extraneous variables with the dependent variable.

2.2.6 Spearman's Rank Order Correlation Coefficient

Spearman's rank order correlation coefficient *rho* is a nonparametric method of computing correlation from ranks. The method is similar to that used to compute Pearson's correlation coefficient (a parametric test), with the

computed value *rho* providing an index of relation between two groups of ranks.

If the original scores are ranks, the computed index will be similar in value to that computed by the Pearson (product moment) method. The is the product moment method assigns weight to the magnitude of each score whereas the rank method focuses on the ordinal position of each score. The coefficient of rank correlation (*rho*) ranges from +1, when paired ranks are changing in the same order, to -1, when ranks are changing in reverse order. A score of zero indicates the paired ranks are occurring at random. The equation for rank correlation is:

$$\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

where d is the difference between each subject's rank for the two variables being studied, $\sum d^2$ is the sum of squared differences between ranks, 6 is a constant and n is the number of paired scores.

Suppose 10 students are drawn at random from a large class; each student has been rated on a 10-point scale for a recent clinical experience, and each student has a grade-point average (GPA) on file. The coefficient of ranks can be computed to determine the extent of agreement between the two sets of scores (clinical experience ratings and GPA). In the rank correlation method, the raw scores are replaced by assigning an appropriate rank to each score of each set. Ranks for each set correspond to the total number of scores in that set.

Step 1. Make a table of the subjects' scores and ranks for the two variables of interest and subtract the ranks to determine the difference (diff)

ISBN: 978-93-85822-53-7

for each pair of ranks. Square each of these differences and sum the squared values.

This example illustrates what happens when scores are similar (tied ranks). When tied ranks occur, each score is assigned the average rank the tied scores occupy (a higher rank is better). The GPA of 3.2, for example, had two scores occupying ranks 5 and 6. The average rank for the score 3.2 is obtained by adding ranks (5 +6) and dividing by the number of ranks occupied (e.g., 5 +6 + 2= 5.5 ranking).

Step 2. Substitute the calculated value of $\sum d^2$ and solve for ρ:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{(6)(22)}{1000-10} = 0.867$$

Consulting a textbook of statistics that provides a table of values for ρ, one finds a minimum ρ value of 0.746 is needed to be considered significant at the .05 level of significance. Thus, the correlation coefficient ρ of 0.867 confirms a statistically significant correlation between the two sets of rankings.

2.2.7 Kendall's Rank Correlation

Kendall's rank correlation tau ('r) is another nonparametric measure of association. When relatively large numbers of ties exist in a set of ranking, Kendall's tau is preferred over Spearman's *rho*. The formula and procedures for calculating it have been adapted from Siegel (33).

$$\tau = \frac{\text{actual score}}{\text{maximum possible score}} = \frac{S}{\frac{1}{2}N(N-1)}$$

ISBN: 978-93-85822-53-7

where N = the number of objects or individuals ranked on both X and Y characteristics.

The value of S can be determined by arranging the first set of measurements into their natural order (e.g., 1, 2, 3, 4, 5) and aligning the second set of measurements under them (e.g., 2, 1,4, 5, 3). Starting with the first number on the left in the bottom row, the number of ranks on the right which are larger are counted. The derivations of the actual score and the maximum possible score are illustrated in the example that follows: Two orthotists rank the fit of an "off-the-shelf" ankle-foot orthosis (AFO) on five different patients.

Step 1. Rearrange the data so the first orthotists rankings fall in a natural (increasing) order and the second orthotists rankings are tabulated in the same order.

Step 2. Compare the first ranking of orthotist 2 with every ranking to its right, assigning a +1 to every pair in which the order is natural and a -1 to every pair in which the order is unnatural:

- Comparing 2 with 1 produces a -1
- Comparing 2 with 4 produces a +1
- Comparing 2 with 5 produces a +1
- Comparing 2 with 3 produces a +1

Repeat for each subsequent ranking of orthotists 2:

- comparing 1 with 4 produces a +1
- comparing 1 with 5 produces a +1
- comparing 1 with 3 produces a +1
- comparing 4 with 5 produces a +1

- comparing 4 with 3 produces a -1
- comparing 5 with 3 produces a -1
- The sum of the comparisons : 4

Step 3. Add these measures of "disarray" (sum = 4) and enter this sum in the above formula as a substitute for S.

Step 4. The value of N = 5. Thus, it becomes:

$$r = \frac{S}{(N+2)(N-1)} = \frac{4}{(2.4)(4)} = .4$$

Step 5. The statistical significance can be determined by two procedures, depending on sample size.

If N is equal to or less than 10, use a probability table such as that found in the appendix of a textbook on statistics to find the statistical significance of 'r. In this example, the table of probability indicates a probability score (p-value) of 0.242 for a value of 0.400. Thus, this test supports the conclusion that the ratings of the two orthotists are not significantly correlated.

For situations in which N is greater than 10, a z score can be computed for the 'r obtained and the statistical significance of the correlation read from a corresponding table of z scores:

$$z = \frac{r}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}}$$

2.2.8 Chi -square Test of Independence

The Chi-square test of independence is a nonparametric test designed to determine whether two variables are independent or related. This test is

ISBN: 978-93-85822-53-7

designed to be used with data that are expressed as frequencies; it should not be used to analyze data expressed as proportions (percentages) unless they are first converted to frequencies.

The application of Chi-square to contingency tables can best be illustrated by working through an example. Suppose a sample of new graduates of an orthotic educational program and orthotists with more than five years of clinical experience were asked whether research should be a part of every orthotist's practice. The replies were recorded as "Agree" or "Disagree."

Step 1. Organize the data into the form of a 2 x 2 contingency table. Note the table includes row totals, column totals and the grand total of subjects included in the sample.

The actual numbers of "Agree" responses were 82 from recent graduates and 30 from experienced orthotists. The numbers disagreeing with the statement were 12 and 66, respectively.

The rationale that underlies Chi-square is based on the differences between the observed and the expected frequencies. The observed frequencies are the data produced by the survey. The expected frequencies are computed on the assumption that no difference existed between the groups except that resulting from chance occurrences.

Step 2. The expected frequencies are computed as follows:

$$\frac{(a+b)}{n} \times (a+c) = \text{expected } a' = \frac{(82+12)}{190} \times (82+30) = 55.41$$

$$\frac{(a+b)}{n} \times (b+d) = \text{expected } b' = \frac{(82+12)}{190} \times (12+66) = 38.59$$

$$\frac{(c+d)}{n} \times (a+c) = \text{expected } c' = \frac{(30+66)}{190} \times (82+30) = 56.54$$

$$\frac{(c+d)}{n} \times (b+d) = \text{expected } d' = \frac{(30+66)}{190} \times (12+66) = 39.41$$

Cross-tabulation and the computation of Chi-square can be made when the variables are nominal as well as ordinal, interval or ratio, and the Chi-square statistic is useful for discrete or continuous variables. However, it is assumed that data occur in every category; thus, no cell may have an observed frequency of zero. The formula for the degrees of freedom for calculating Chi-square and the contingency coefficient is:

$df=(k-1)(r-1)$ (5)

where k = number of columns in the contingency table and r = number of rows in the contingency table.

Step 3. The Chi-square (X^2) is calculated as below:

$$x^2 = \sum \frac{(O-E)^2}{E}$$

where O is the observed number of cases found in the ith row of the ith column, and E is the expected frequency obtained by multiplying the two marginal totals for each cell and dividing the product by N

ISBN: 978-93-85822-53-7

Step 4. The Chi-square is computed by finding the difference between the observed and expected frequencies in each cell, squaring that difference and dividing by the expected frequency of that cell. The result for each cell is then added, and the total is the value of the Chi-square. (Chi-square = X^2 = 61.84.)

Step 5. Consulting a table of Chi-square values in a textbook of statistics, using 1 degree of freedom and the 0.05-level of significance, we find that a minimum value of 3.84 is needed for the observed frequency to be considered significantly different from the expected frequency. In this example, the value of X greatly exceeds that minimum value; thus, the observed values are significantly different from the expected values.

The use of the Chi-square statistic has important limitations. Although no association is indicated by a zero, a perfect association is not indicated by a 1.00. Moreover, the size of Chi-square is influenced by both the size of the contingency table and the size of the sample.

The addition of rows and columns as a table grows is accompanied by larger and larger values of Chi-square- even when the association remains essentially constant. If the sample size is tripled, the value of Chi-square is tripled, and everything else remains the same. Degrees of freedom depend on the number of rows and columns, not the sample size; thus, inflated values of Chi-square occur for large samples, leading the investigator to conclude the differences between observed and expected frequencies are more significant than warranted. The Chi-square is designed for use with relatively small samples and a limited number of rows and columns.

2.2.9 Phi

The correlation coefficient *phi* corrects for the size of the sample when the table size is 2 x 2. The equation is:

$$\Phi = \sqrt{\frac{\chi^2}{N}}$$

Phi is 0 when no relationship exists and 1 when variables are related perfectly. When tables are greater than 2 x 2, *Phi* has no upper limit and is not a suitable statistic to use. The statistical significance of *Phi* may be tested by calculating a corresponding Chi-square value and assigning 1 degree of freedom to it:

$$\chi^2 = N(\Phi)$$

2.2.10 Cramer's V

Cramer's V is an adjusted CF, modified to be suitable for tables larger than 2 x 2. The value of V is zero when no relationship exists and 1 when a perfect relationship exists. The equation for Cramer's V is:

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

Thus, when 2 x 2 tables are involved, Phi may provide a more useful measure of the relationship between the two variables than that provided by Chi-square. For tables larger than 2 x 2, Cramer's V is the statistic of choice (k = number of columns in the table).

Tests for Significance of Differences

2.2.11 Two-Group Design: Chi-square (2 x 2)

The Chi-square comparison of differences between two groups is one of the better known and commonly used statistical procedures. The same procedure for Chi-square (x^2) as described above can be used to test for the significance of differences in two groups of data that are expressed as frequencies.

Suppose researchers wanted to determine if the proportion of trauma patients being referred for orthotic services in a particular hospital was significantly different than the number being referred for orthotic services in another hospital with a similar mix of patients. During a specific 12-month period, the orthotic department in Hospital A filled 238 requests for orthoses from a pool of 2,222 patients, and the orthotic department in Hospital B filled 221 requests for orthoses from a pool of 1,238 patients. First, the data are organized into a 2 x 2 contingency table.

As before, the general equation for Chi-square is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

To compute X for a contingency table, simply square the difference between the observed and expected frequencies in each cell and divide by the expected frequency of that cell. Finally, total the cells to obtain the X^2value. X^2= 23.72, which is evidence that the experiences of the two hospitals are significantly different.

2.2.12 Two-Group Design: Chi -square Median Test

The Chi-square median test can be used to determine if the medians of two groups are different. For example, all of the male patients fitted with an AFO to correct foot-drop following the onset of hemiplegia were asked to rate the comfort of their footwear when walking with the AFO. Forty-four patients were evaluated; 32 wore normal leather shoes and 12 wore tennis shoes.

Comfort was rated on a nine-point scale (the larger the score, the greater the comfort in walking), and the evaluation was made six months after fitting the AFO and with the subject walking 50 yards. The median comfort rating of the 44 patients was 7.3. The number of subjects rating their comfort above or below is the grand median.

The Chi-square computation viewing the leather shoe and tennis shoe wearers as random samples is shown below. The ratings are discrete units; each patient's rating appears only once, and the ratings are independent.

$$\chi^2 = n_1\left[\sum_{r=1}^{R} + \sum_{c=1}^{C} + \frac{n_{rc}^2}{n_r n_c} - 1\right]$$

where n_1 the total number of observations, n_{rc}^2 is the number of observations in the rc^{th} cell of the contingency table, n_r is the number of observations in the r^{th} row of the table, and n_c is the number of observations in the c^{th} column of the table.

$$\chi^2 = 44\left[\frac{4^2}{19\times(12)} + \frac{15^2}{19\times(32)} + \frac{8^2}{25\times(12)} + \frac{17^2}{5\times(32)} - 1\right]$$

$$=44(0.070175+0.370066+0.213333+0.361250-1)$$

$$=44(1.014825-1)$$

$$=0.652$$

A table of Chi-square values is then consulted to determine if this calculated value of Chi-square is sufficiently large to represent a statistically significant difference of the mean scores. The degree of freedom is (R - 1) (C - 1) = (1)(1) = 1. In this example, a Chi-square of 3.84 or larger would be needed (at a .05 level of significance) to justify the conclusion that the comfort levels of the two different types of footwear were significantly different.

2.2.13 Two-Group Design: Tukey's Quick Test

Tukey's quick test is used to determine if the results of two different interventions produced the same or different effects. Suppose a sample of 20 patients with limitation in elbow extension on one side that exceeded 50 degrees was treated with one of two methods for reducing contractures. Subgroup A, consisting of 10 patients, was treated with serial casting over a period of one month; Subgroup B, also consisting of 10 patients, was treated with an adjustable splint worn 18 hours a day for one month. The increased range of motion for subjects were in the two groups.

Tukey's quick test is applied by identifying the group containing the largest value and the group containing the smallest value in the two groups. In this example, Group B contains the largest value of either group (41), and Group A contains the smallest value (14). The number of values in Group B that are larger than the largest value in Group A (36) are counted and recorded (in this example, there are 2). Next, the number of values in Group A that are smaller than the smallest value in Group B (18) are counted (there are 2). The two counts are added, and, if the sum is equal to or greater than 7, we conclude the effects of the two treatments are different. If the

sum is less than 7, we conclude the effects of the two treatments are not different. In the present example, the sums of the two counts equal 4; therefore, we conclude the effects of the two interventions are not different. In the event the largest and the smallest values occurred in the same group, we conclude automatically that the two treatments did not have different effects. In Tukey's test the number 7 is a constant and is the criterion value to be used with any set of data.

2.2.14 Two-Group Design: Mann-Whitney U-Test

The Mann-Whitney U-test is a rank test for two independent samples, each with a small number of subjects. This test is a good alternative to the parametric t-test. Suppose measurements of the height of the ankle joint axis (in millimeters) in a group of patients receiving services in Orthotic Clinic A are compared with measurements taken from a group of patients in Orthotic Clinic B to determine if they are comparable. Because of the small number of cases, a nonparametric test is selected. The measurements are assigned a rank in ascending order of height, with a rank of 1 being the smallest value:

$$\chi^2 = 44\left[\frac{4^2}{19\times(12)}+\frac{15^2}{19\times(32)}+\frac{8^2}{25\times(12)}+\frac{17^2}{5\times(32)}-1\right]$$

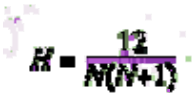
$$=44(0.070175+0.370066+0.213333+0.361250-1)$$

$$=44(1.014825-1)$$

$$=0.652$$

The ranks then are ordered according to their identity.

ISBN: 978-93-85822-53-7



The value of the Mann-Whitney U-test is found by determining the number of A scores preceding each B score. The U is: 1 + 2 +3 + 4 + 5 = 18 (rank 2A precedes 3B = +1; ranks 2A and 4A precede 5B = +2; ranks 2A, 4A and 6A precede 7B = + 3 and so on). Consulting a Mann-Whitney U-test table for nB = 7 (larger sample size), locate the U value (18) on the left-hand margin and nA = 5 at the top of the table. The probability that these two samples are equivalent is 0.562, which is not statistically significant (i.e., the distribution of ankle heights is not different). This procedure is appropriate only when the larger sample size is 8 or smaller. Different procedures and tables are used for samples ranging between 9 and 20 and larger than 20, respectively. Procedures and tables for the Mann-Whitney

U-test can be found in Siegel and Castellan.

2.2.15 Two-Group Design: Wilcoxon Matched Pairs/Rank Test

The Wilcoxon matched pairs/rank test is an alternate form of the Mann-Whitney test that is used when the samples are dependent. For purposes of illustration, presume the time to ambulate 25 meters is measured with a stopwatch when the patient is wearing a new type of lightweight KAFO and again when wearing a conventional metal KAFO. The ambulation times for each patient are tabulated, and the absolute difference between each pair of numbers is calculated. The nonzero differences then are ranked according to their absolute values and separated into ranks associated with positive and negative differences.

As in the case with the Mann-Whitney procedures for analyzing differences between independent samples, the resulting score, in this case called a T value, is used to look up the statistical significance of the differences in a table. In this example, a T value of 8 or more indicates that the two situations are significantly different, and the subjects walked more quickly when wearing the lightweight KAFO.

Multi-Groups Design with Independent Samples: The Kruskal-Wallis One-Way Analysis of Variance by Ranks. This method functions like the conventional one-way analysis of variance. The null hypothesis is tested to determine if the differences among samples show true population differences or whether they represent chance variations to be expected among several samples from the same population. The test is based on the assumptions that ranks within each sample constitute a random sample for the set of numbers 1 through N (15) and that the variable being tested has a continuous distribution (4). Scores in all samples are combined and arranged in order of magnitude so that ranks can be given to each score. The lowest score is assigned the rank of 1. The scores then are replaced in their respective samples with appropriate ranks. The ranks for each sample are summed. The assumption is that mean rank sums (R) are equal for all samples and equal to the mean of the N ranks, (N + 1)12, if the samples (K) are from the same population (16). Both equal- and un-equalsized samples can be used in this test because the sums of sample ranks (>R) are pooled in the equation. The statistic H used in this test can be defined by the equation:

$$H = 12/[N(N+1)]$$

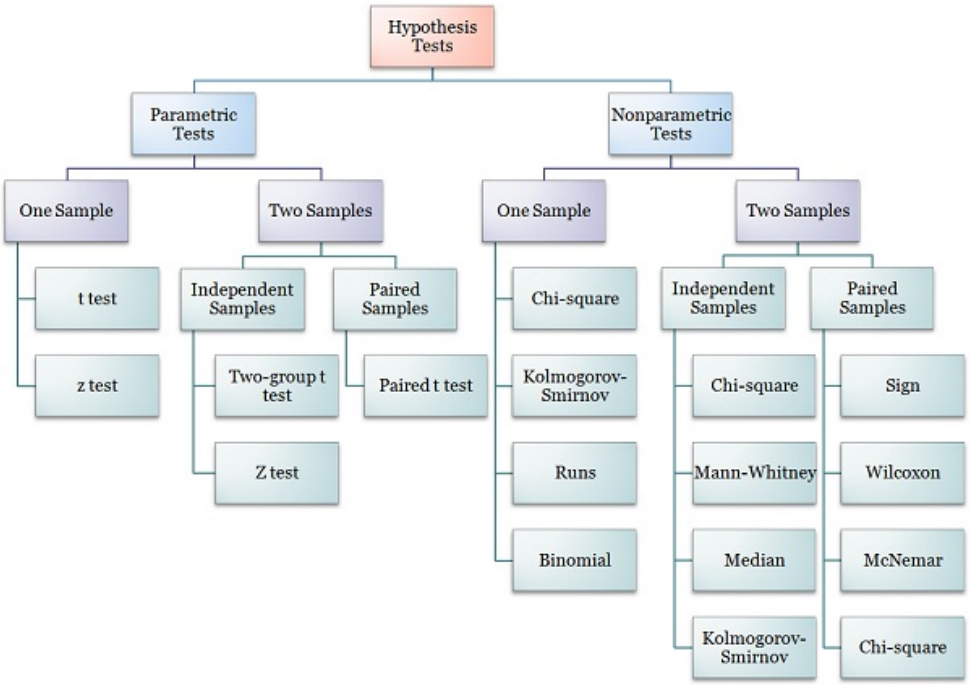
where N is the number of scores in all samples combined. The random sample distribution of H is approximated by a Chi-square distribution of K-1 degrees of freedom, where K is the number of samples. The Chi-square probability can be found in appendix tables published in Reference 11. The Kruskal-Wallis One-Way Analysis of Variance by Ranks is used when assumptions for the parametric Analysis of Variance are not suitable for the data, or when the level of data is less than interval measures.

Appendix 1: Common Statistical tests and their use

Type of Test:	**Use:**
Correlation	These tests look for an association between variables
Pearson correlation	Tests for the strength of the association between two continuous variables
Spearman correlation	Tests for the strength of the association between two ordinal variables (does not rely on the assumption of normal distributed data)
Chi-square	Tests for the strength of the association between two categorical variables
Comparison of Means: look for the difference between the means of variables	
Paired T-test	Tests for difference between two related variables
Independent T-test	Tests for difference between two independent variables
ANOVA	Tests the difference between group means after any other variance in the outcome variable is accounted for
Regression: assess if change in one variable predicts change in another variable	
Simple regression	Tests how change in the predictor variable predicts the level of change in the outcome variable
Multiple regression	Tests how change in the combination of two or more predictor variables predict the level of change in the outcome variable

Non-parametric: are used when the data does not meet assumptions required for parametric tests	
Wilcoxon rank-sum test	Tests for difference between two independent variables - takes into account magnitude and direction of difference
Wilcoxon sign-rank test	Tests for difference between two related variables - takes into account magnitude and direction of difference
Sign test	Tests if two related variables are different – ignores magnitude of change, only takes into account direction

Appendix 2: Hypothesis Tests Hierarchy



ISBN: 978-93-85822-53-7

Appendix 3: Parametric and Nonparametric Equivalencies

The table below outlines some common research designs and their appropriate parametric and nonparametric equivalents.

Analysis Type	Parametric Test	Nonparametric Test
Compare means of two independent groups	Independent sample t-test	Wilcoxon rank-sum test
Compare means of the same group at two time points	Paired sample t-test	Wilcoxon signed-rank test
Compare means between three or more independent groups	Analysis of variance (ANOVA)	Kruskal-Wallis test
Estimate degree of association	Pearson's Product Moment correlation	Spearman's rank correlation

REFERENCE

[1] Agbonifoh, B. A. and Yomere, G. O. (1999).Research Methodology in the Social Sciences and Education, Benin City, Centerpiece Publishers.American Statistician. Vol. 61, No. 1: 47 – 55.

[2] Aweriale, P. E. and Dibua, V. A. (1997).Research methods and Statistics, Benin City, Olive Publishers.

[3] Baccheti, F. (2002). Peer Review of Statistics in Medical Research: The Other Problem. British Medical

[4] Cohen, J. (1998). Statistical Power Analysis for the Behavioural Science, New Jersey.Hillsdale Publishers.

[5] Conover, W.J. (1980). Practical Nonparametric Statistics, New York: Wiley & Sons.

[6] Currier DP. Elements of research in physical therapy. 3rd ed. Baltimore: Williams & Wilkins, 1990.

[7] Domholdt B. Physical therapy research: principles and applications. Philadelphia: WB. Saunders Co., 1993.

[8] Egbule, J. F. and Okobia, D. O, (2001).Research Methods in Education for Colleges and Universities, Agbor, Kmensuo Publishers.

[9] Ferguson GA. Statistical analyses in psychology and education. 5th ed. New York: McGraw-Hill Book Co., 1981.

[10] Fletcher, K. E, French, C. T, Corapi, K. M, Irowins, R. S. and Norman, G. R. (2010). Prospective measures

[11] Freund JE. Modern elementary statistics, 5th ed. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1979.

[12] Gaito, J. (1980). Measurement Scale and Statistics: Resurgence of an Old Misconception, Psychological Bulletin.87, 564 - 567.

[13] Glass G~ Hopkins KD. Statistical methods in education and psychology. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1984.

[14] Jameison, S. (2004). Likert Scales: How to Abuse Them. Medical Education, 38, 1217 – 1218Kim, M. (2006). Statistical Methods in Arthritis and Rheumatism, Arthritis and Rheumatism Vol. 54, No. 12:Journal, 234, 1271 – 1273.

[15] Kress, G. (1982). Marketing Research, U.S.A, Prentice Hall Publishers.

[16] Kuzon, W. M., Urbanchek, M. G. & McCabe, S. (1996). The Seven Deadly Sins of Statistical Analysis.Annals of Plastic Surgery. 37, 265 – 272.

[17] Lunsford BR. Statistics: screening and data summary. JPO 1993; 5:4:125-30.

[18] Lunsford TR, Lunsford BR. The research sample, part III: sample size. JPO 1995; 7:4:137-41.

[19] McCollough C. Introduction to statistical analyses: a semi-programmed approach. New York: McGraw-Hill Book Co., 1974.

[20] Motulsky, H. (1995). Intuitive Biostatistics, New York: Oxford University Press.

[21] Nachmias, C. F. and Nachmias, D. (2009).Research Methods in the Social Sciences, London, Replika Press Ltd.

[22] Nalou, A. A. (2011). Assessing the Statistical Methodologies of Business Research in South Africa.

[23] Nickerson, R. S. (2000). “Null Hypothesis Significance Testing: A Review of an Old and Continuing

[24] Norusis M. The SPSS guide to data analysis for SPSS-X. Chicago: SPSS Inc., 1987.

[25] Olannye, P. A. (2006). Research Methods for Business: A Skill Building Approach. Lagos, Pee Jen Publishers.

[26] Pearson, E. S. (1981). The Test of Significance for the correction coefficient. Journal of the American Statistical Association 27, 128 – 134.

[27] Rosner, B. (2000). Fundamentals of Biostatistics, California: Duxbury Press.

[28] Rousseau, G. G. (1992). Identifying Criteria of Consumer Awareness in a New South Africa, Management Dynamics. Vol. 1, No.: 67 – 86.

[29] Schmudt, M. J. and Svend, H. (2006).Marketing Research.New York, Prentice Hall Publishers.

[30] Schor S. Fundamentals of biostatistics. New York: GP Putnam Sons, 1969.

[31] Siegel 5, Castallan NJ. Nonparametric statistics for the behavioral sciences, 2nd ed. New York: McGraw-Hill Book Co., 1988.

[32] Siegel S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Co., 1956.

[33] Stresak, A. A, Zaman, Q. Marinell, G. Pfeifer. Kp. And Ulmer, H. (2007). The Use of Statistics in Medical

[34] Thorndike R. Correlational Procedures for Research. New York: Gardner Press Inc., 1976.

[35] Tukey JW Quick and dirty methods in statistics: 2. simple analyses for standard designs. In: Quality control conference papers. New York: American Society of Quality Control, 1951.

[36] Walsh, J.E. (1962) Handbook of Nonparametric Statistics, New York: D.V. Nostrand.

[37] Williamsen EW. Statistical reasoning. San Francisco: WH. Freeman and Co., 1974.