

Full Length Research Paper

Karl Pearson's chi-square tests

Nihan Sölpük Turhan

Department of Educational Sciences, Faculty of Education, Fatih Sultan Mehmet Foundation University, İstanbul, Turkey.

Received 6 September, 2019; Accepted 5 May, 2020

Statistical tests have been an important tool for interpreting the results of research correctly. The factors that influence the determination of the statistical test are research purpose, hypothesis and data. Today, statistical tests are used more frequently, and they aim to analyze whether statistical tests are used in accordance with research. For this purpose, frequently used chi-square tests are discussed and in this work and the research hypotheses are examined. The method of this research is qualitative and was developed according to the literature review. In the analysis of the data, it explains what the chi-square test is and the differences in practice according to the research hypothesis. In this study, a comparison was made between goodness of fit, homogeneity and independence chi-square tests. In the findings of the study, the differences between the studies using three different tests are presented according to the population and hypothesis. The differences between the studies using three different tests are presented according to the population, hypothesis, and statistical formulas. This area includes definitions in the literature and applications in the field of educational sciences. Simplified definitions and applications that can be adopted by researchers are presented.

Key words: Chi-square test, goodness of fit, independence, homogeneity.

INTRODUCTION

Statistical tests are important for scientific research. Statistical tests to be used in a research affect the comments related to the research results. The biggest problem here is not the theory, but the possibility that a research can change or change the analysis of the data (Black, 1993). The fact that the data selection and the decision process are not taken into account in the analysis are discussed in detail in the article by Gow et al. (2016). The power, reliability, quality, and significance of a scientific study depend on the study process to some extent. Hypothesis and statistical tests of a study may change the inferences from the study. At the same time, they draw attention to the problem of efficiency of the

statistical tests in revealing the truth in the study process depending on their purpose and condition of application. Data analysis is also essential in the study process in addition to statistical tests and hypotheses. Koul (1984) suggests that none of the statistical methods need to be applied unless data analysis and processing are explained and clarified. Scientific studies consist of statistical tests that are made based on categorical and non-categorical data. Such tests are called parametric and non-parametric tests. They help a researcher to analyze the data accurately and significantly. Thus, they are beneficial in summarizing the results in a significant and appropriate way and drawing general conclusions

E-mail: nsolpuk@fsm.edu.tr.

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

(Karagöz, 2010).

In parametric tests, actions are taken based on a parameter, a specific distribution and the concept of variance in the statistics. They are inflexible statistical methods related to population or sample. A parametric statistical test makes an assumption on the population parameters and the distributions where the data are obtained. Accordingly, some assumptions such as suitability of the data to normal distribution, their random selection and quantitative form are based on. Such tests contain Student-t tests and ANOVA test assuming that the data come from a normal distribution.

Non-parametric tests are dealt without being based on population distribution or population parameters. When the word "non-parametric" is used in statistics, it does not mean that nothing is known about the population. This generally means that we know that the population data do not have a normal distribution. A generalization cannot be made based on the results obtained from the hypotheses used in such tests (Gay, 1976). According to Onyango and Odebero (2009) as the population is not known to be normal, sample size is small and the variables expressed as nominal or ordinal.

Non-parametric tests include many tests such as Chi-square, Mann-Whitney U, Sign, Wilcoxon, Median, Kolmogorov-Smirnov and McNemar. Chi-square test is used in discrete data in the form of frequency. It is an independence test and is used to estimate the probability of some non-random factors to take account of the observed correlation. When the literature is examined, similar studies related to the distribution of p-value in small sample size were found in the majority of the studies (Mehotra et al., 2003). Chi-square test shall be taken into consideration in this study. The reason is that this test is commonly used by researchers compared to other non-parametric tests. This study deals with applications of Chi-square test and its use in educational sciences.

Chi-square test

Chi-square test is used to find if there is any correlation among nonnumeric variables that are frequently used in statistical studies (Kothari, 2007). It is symbolized as χ^2 . Kothari (2007) stated that the following requirements must be fulfilled before the test.

- (i) Observed and expected observations are to be collected randomly.
- (ii) All the members (or items) in the sample must be independent.
- (iii) None of the groups must contain very few items (less than 10).
- (iv) The number of total items must be quite large (at least 50).

The logic of hypothesis testing was first developed by

Karl Pearson (1857-1936) (Magnello, 2005). Chi-square goodness of fit tests, independence tests, and homogeneity tests that were developed by Pearson are the most significant contributions that he made to the modern statistics theory. The significance of Chi-square distribution of Pearson is that statisticians can use statistical methods that do not depend on normal distribution in order to interpret findings. The significance of Chi-square value is determined by using the suitable degree of freedom and degree of significance and consulting a Chi-square table (Moore, 1994). The two special purposes of Chi-square test are to test the hypothesis that there is no correlation among two or more groups, populations or criteria, and to test to what extent the observed data distribution fits to the expected distribution.

Purpose of the study

The purpose of this work is to compare chi-square tests, goodness of fit tests, independence tests and homogeneity tests of Karl Pearson. The difference among these three types of test and each type needs to be examined. Different article samples that were published by using three different test methods are presented. The purpose of this article is to provide a quick general overview of chi-square test.

METHODOLOGY

This study contains collection of current literature, its critical analysis and the data from various sources. Literature analysis that is a qualitative research data collection method is used in this study, and the aim is to do an in-depth analysis of the study problem. Findings obtained by including recent studies on a particular subject in literature reviews are presented under various concepts or themes (Grant and Booth, 2009).

Data collection

Literature review is generally used as a combination of the methodologies in the same study case. Literature review is used and the articles and scientific sources related to the study subject are utilized in this study. In this research, Google scholar and ScienceDirect databases were used in literature review. First of all, the definitions regarding the three different managements and the differences in the steps in the analysis process were revealed by the findings obtained from the literature review. Later, researches developed using chi-square goodness of fit, independence and homogeneity tests were handled according to the purpose, hypothesis and result.

Data analysis

The literature that can be used for systematic evaluation is analyzed by collecting them in different ways in the study. Data analysis is examined and reported as part of the previous literature researches. By this means, the data in the literature are compared,

Table 1. Chi-square tests and attributes.

Chi-square test attribute	Test of independence	Test of homogeneity	Test of goodness of fit
Sampling type	Single dependent sample	Two (or more) independent samples	Sample from population
Interpretation	Association between variables	Difference in proportions	Difference from population
Null hypothesis	No association between Variables	No difference in proportion between groups	No difference in distribution between sample and population

Source: Franke et al. (2012).

examined and synthesized. The factors to be taken into consideration in effective and efficient analysis of the study results by using Karl Pearson Chi-square tests are presented in the study. Definitions of various authors on these tests are examined. Attributes and application areas of chi-square goodness of fit, independence, and homogeneity tests and the conceptual model are also stated. Finally, different studies that were made by using these tests are dealt with.

FINDINGS

In this study, the differences by the purpose and condition of application of chi-square test and the sample analyses are obtained from the studies in the literature. Chi-square test is examined under three titles depending on its purpose and condition of application:

- (1) Chi-square goodness of fit test
- (2) Chi-square independence test
- (3) Chi-square homogeneity test

In Table 1, three different chi-square tests were compared. These tests were examined according to sampling type, interpretation and null hypothesis. In this regard, these distinctions can best be explained by the null hypothesis and sampling type tested in each of these tests. So, interpretation can change depending on these.

Chi-square of goodness of fit test

Chi-square of goodness of fit test is also called *single-sample goodness of fit test* or *Pearson's Chi-square of goodness of fit test*. This test is a single-sample non-parametric test.

- (i) Cases (for instance, participants) are obtained from a single categorical variable. For example, "educational background" consisting of two groups: "high school" and "university".
- (ii) Distribution is obtained from a known or hypothesized distribution. For example, a known distribution such as the rate of literate and illiterate persons in a country or a

hypothesized distribution such as the rate of men compared to women in the number of participants of the university admission exam in the next year.

While chi-square of goodness of fit test is applied, it is important to "hypothesize" if we can expect the cases in each group of the categorical variable to be "equal" or "unequal". Mutai (2000) suggests that goodness of fit test is far from comparing the data that are empirically derived to the results that are expected theoretically (expected to be frequencies).

In data analysis, it is required to find degrees of freedom, expected frequency counts, test statistics and P value related to the test statistic. Degrees of freedom (DF) are equal to (k) number of minus 1 levels of the categorical variable.

$$DF = k - 1$$

Expected frequency count is equal to the expected frequency counts at each level of the categorical variables, and the sample size of the rate is hypothesized by using the null hypothesis.

$$E_i = np_i$$

E_i , is the expected frequency for ith level of the categorical variable, n is the total sample size, and p_i , is the hypothesized rate of the observations at i level. The value of Chi-Square goodness of fit test is calculated by using the formula,

$$X^2 = \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

O_i , is the observed frequency count for ith level of the categorical variable, and E_i , is the expected frequency count for ith level of the categorical variable.

Hypothesis of Chi-square of goodness of fit test is stated thus,

H_0 : Data follow a specified distribution.

Table 2. Sample analysis related to chi-square of goodness of fit test.

Aim	H ₀ hypothesis	Result
"The constant error of the halo in educational outcomes research" (Pike, 1999)	"Traditional and halo models including both college experience and gain factors were specified and tested" (Pike, 1999)	"The chi-square goodness-of-fit statistic for the traditional model was statistically significant ($\chi^2=1180.25$; $df=360$; $p < 0.001$)" (Pike, 1999)

Table 3. Sample analysis related to Chi-square test of independence.

Aim	H ₀ hypothesis	Result
"In this study, the correlation between the proportional reasoning skill and the ratio and proportion problem posing skill of primary school students is studied" (Çelik and Özdemir, 2011).	"There is no significant correlation between proportional reasoning levels and problem posing skills of the students" (Çelik and Özdemir, 2011).	"Pearson chi-square value is found $\chi^2=185.63$. According to the values p, the value $p < 0.05$ is obtained, so it is found that there is a significant correlation between proportional reasoning levels and problem posing skills of the students" (Çelik and Özdemir, 2011).

H₁: Data do not follow the specified distribution.

The comment on rejection is that the sample is significantly different from the population by the correlation variable. Table 2 contains a sample study in which a sample related to chi-square of goodness of fit test is compared to a population that has known parameters in a correlation variable.

Chi-square test of independence

It evaluates if some categorical variables are correlated with some populations, because variables tend to be a bit different from their populations. However, it is not probable for the variables in a sample to have a strong correlation if the variables are independent in the whole population. Consequently, we conclude that the variables are probably not independent in the population.

In data analysis, it is required to find degrees of freedom, expected frequencies, test statistics and P value related to the test statistic.

Degree of freedom (DF) is equal to the following:

$$DF = (r - 1) * (c - 1)$$

In this formula, r is a variable and c is the level number for the other categorical variable.

Expected frequency counts are calculated separately for each level of a categorical variable at each level of the other categorical variable. According to the formula below, $r * c$ gives the expected frequencies.

$$E_{r,c} = (n_r * n_c) / n$$

$E_{r,c}$ is the expected frequency number for level r of the variable A and level c of the variable B.

n_r is the total sample observation number at the level r of the variable A.

n_c is the total sample observation number at the level c of the variable B

n is the total sample size.

Test statistic. Test statistic is a chi-square random variable that is defined by the equation below (χ^2).

$$\chi^2 = \sum [(O_{r,c} - E_{r,c})^2 / E_{r,c}]$$

$O_{r,c}$ is the observed frequency number at the level r of the variable A and the level c at the variable B. $E_{r,c}$ is the expected frequency number at the level r of the variable A and the level c of the variable B.

The value P is the observation probability of a sample statistic that is as extreme as test statistic. Test statistic is chi-square, so it is used to evaluate a probability that is related to the test statistics. The degrees of freedom that are calculated above are used in it.

Hypothesis of Chi-square test of independence is stated below.

H₀: Correlation variables are independent.

H₁: Significant variables are correlated.

In Table 3, the hypothesis sentence and the test result related to chi-square test of independence are given. Chi-square test of independence determines if the two categorical variables in a single sample are independent from each other.

Chi-square test of homogeneity

The test is applied to a single categorical variable from two or more different populations. It is used to determine if frequency counts are distributed among different

Table 4. Sample analysis related to chi-square homogeneity test.

Aim	H ₀ hypothesis	Result
“This relative distribution can be used to test for homogeneity between groups, and the groups having relatively higher distribution are favored by the students to make their decision” (Paul and Smith-Hunter, 2011).	“Drinkers and nondrinkers are not homogeneous” (Paul and Smith-Hunter, 2011).	“Chi-square test of homogeneity shows that these two groups (drinkers and nondrinkers) are not homogeneous in terms of citations” (Paul and Smith-Hunter, 2011).

populations in the same way. The only difference between the independence test and homogeneity test is the specification of null hypothesis. The homogeneity test tests the null hypothesis that claims homogeneity or equality based on some attributes.

Chi-square test of homogeneity is used to determine if two or more independent sample vary by distributions on a single variable. A common use of this test is to compare two or more groups or conditions on a categorical result. Formulation of omnibus test statistic is formed as independence test and homogeneity test. But, although they are the same, there are differences in these two test sampling hypotheses. Hypothesis of Chi-square homogeneity test is stated below.

H₀: Distribution of correlation variables is the same.

H₁: Distribution of the significant variables is not the same.

In Table 4, chi-square independence test was made in the study, and whether the groups vary by distribution of the significant variable as a result was analyzed.

DISCUSSION

This study emphasizes the definition and different applications of Chi-square test. Applications of chi-square homogeneity, goodness of fit and independence tests are examined in detail. A simplified conceptual model that can be applied in most cases was developed in order to understand different chi-square tests. Although formulations of the omnibus test statistic of chi-square independence and chi-square homogeneity test that are used in many scientific studies are the same, these two tests vary in sampling hypotheses, null hypotheses and the subsequent rejecting options. The main difference between these two tests is the manner of collection and sampling of data. Formulation of chi-square goodness of fit test statistic and its hypothesis are different from others.

Specifically, independence test collects data on a single sample and then compares the two variables in this sample in order to determine the correlation between them. When the data are collected by using only a single sample in chi-square independence test, only independence test is valid. In this test, comments can

be made only on the correlation between the variables.

In homogeneity test, the data from two or more different groups are collected. The two samples are then compared on a single significant variable in order to test if there is any difference between the rates. When the data from two or more samples are collected in chi-square homogeneity test, the homogeneity test is suitable and comparison of rates can be made among multiple groups. Wickens (1989) presents a delicate and concise description of these tests as well as their sampling assumptions and hypotheses. In addition to homogeneity and independence tests, Wickens presents an additional alternative in which both margins are constant „irrelevant classification test“.

In general, chi-square test is a strong statistics that enables the testing of the hypotheses related to the variables that are measured at nominal level. However, important factors to be considered while using chi-square test should be suitable for the purpose, hypothesis and data of the test. This method is frequently used especially in small sample quantitative research. In this line, the aim and hypothesis of the research should be clearly stated and it is important to make an appropriate analysis.

CONFLICT OF INTERESTS

The author has not declared any conflict of interests.

REFERENCES

- Black F (1993). Beta and return. *Journal of Portfolio Management* 20:8-18.
- Çelik A, Özdemir EY (2011). The relationship between the proportional reasoning skills of primary school students and their ability to form a ratio-proportion problem. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi* 30(1):1-11.
- Franke TM, Ho T, Christie CA (2012). The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation* 33(3):448-458.
- Gay LR (1976). *Educational research: Competencies for analysis and application*. Ohio, Bell & Howell Company.
- Gow ID, David FL, Peter CR (2016). Causal inference in accounting research. *Journal of Accounting Research* 54:477-523.
- Grant MJ, Booth A (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2):91-108.
- Karagöz Y (2010). Nonparametrik tekniklerin güç ve etkinlikleri. *Elektronik Sosyal Bilimler Dergisi* 33:1-23.
- Kothari CR (2007). *Quantitative techniques*. New Delhi, UBS Publishers Ltd.

- Koul L (1984). Methodologies of educational research. New Delhi, Vikas.
- Magnello ME (2005). Karl Pearson and the origins of modern statistics: An elastician becomes a statistician. *The New Zealand Journal for the History and Philosophy of Science and Technology* P 1.
- Mehotra DV, Chan IS, Berger RL (2003). A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics* 59:441-450.
- Moore JF (1994). *Research methods and data analysis 1* hull: Institute of Education, University of Hull UK. pp. 9-126.
- Mutai BK (2000). *How to write quality research proposal: A complete and simplified recipe*. New Delhi, Thelley Publications.
- Onyango JO, Odebero SO (2009). Data analysis for educational and social research: Quantitative and qualitative approaches. *Kenya Journal of Education Planning Economics and Management* 1(1):44-53.
- Paul M, Smith-Hunter A (2011). The impact of social, economic and genetic factors on students' alcohol consumption decisions. *Academy of Information and Management Sciences Journal* 14(2).
- Pike GR (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education* 40(1):61-86.
- Wickens TD (1989). *Multiple contingency tables analysis for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.