

1.6: Chi-Squared Test

Hypothesis Testing

When evaluating social phenomena, especially regarding crime, people often make judgments. However, such judgments are not always correct. Researchers conducting empirical studies make tentative statements about phenomena and make decisions about the validity of those statements. These tentative statements are called hypotheses, and if one denies a hypothesis that happens to be true, they commit an error in judgment.

When making decisions based on sample statistics about attributes of a population, i.e., unknown facts, one must judge whether something is like that or not. However, such judgments are not always accurate and can sometimes be wrong. In other words, researchers may make errors in judgment.

Specifically, hypothesis testing entails comparing empirically observed sample findings with the theoretically expected outcomes if the null hypothesis were true. The null hypothesis represents the hypothesis that a researcher aims to reject, thereby supporting its alternative. This hypothesis often posits that two or more variables are not related. To compare the null and alternative hypotheses, the researcher calculates the probability of the observed outcome occurring solely due to chance or random error (Vogt & Johnson, 2011).

NHST Steps

This hypothesis testing is also known as null hypothesis significance testing, or NHST. In the context of NHST, it is recommended to follow these five steps:

- Step 1: Formulate the null and alternative hypotheses.
- Step 2: Calculate the test statistic.
- Step 3: Determine the probability (p-value) of obtaining a test statistic at least as extreme as the observed value, assuming no relationship exists.
- Step 4: If the p-value is very small, typically less than 5%, reject the null hypothesis.
- Step 5: If the p-value is not small, typically 5% or greater, retain the null hypothesis.

Chi-Squared Test

The first hypothesis testing covered in this book is the chi-squared test, commonly called the one-sample chi-square. It is frequently employed to compare the proportion of cases from a sample with either hypothesized values or those previously obtained from a comparison population. In the data file, only one categorical variable and a designated proportion against which to evaluate the observed frequency are required. This test may assess whether no difference exists in the proportion within each category (e.g., 50%/50%) or against a specific proportion derived from a previous study.

For instance, 44 male respondents used credit cards, while 62 female respondents used credit cards for online purchases in our example. From this tabulation, we may conclude that more female respondents use credit cards for online purchases than their male counterparts. There is a difference in the usage of credit cards for online purchases between male and female groups. However, we have no evidence to determine whether this difference is statistically significant or by accident. That's why we conduct a hypothesis test to confirm our findings with statistical evidence. Below, we will evaluate the usage of credit cards for online purchases in the past 12 months between men and women using the NCVS data we used in the last chapter.

NHST Steps for Chi-Squared Test

Step 1: Formulate the Null and Alternative Hypotheses.

The first step in conducting the chi-squared test is to write the null and alternative hypotheses.

- H_0 : The usage of credit cards for online purchases in the past 12 months is the same across men and women.
- H_A : The usage of credit cards for online purchases in the past 12 months is not the same across men and women.

Step 2: Calculate the Test Statistic.

The test statistic to use when examining a relationship between two categorical variables is the chi-squared statistics, χ^2 . First, you will download the [revised data from the 2016 NCVS from the shared Google Drive](#). Then, you will need to first load the data using the following syntax.

```
library(haven)
rNCVS2016 <- read_dta("C:/Users/75JCH0I/OneDrive - West Chester
University of PA/WCU Research/R/data/rNCVS2016.dta")
View(rNCVS2016)
```

We will perform a chi-squared test on a contingency table using R. The variable male was coded as 0 for women and 1 for men. The PayCred variable represents whether the respondent used credit cards for online purchases in the past 12 months. Those who used it were coded as 1, and those who did not were coded as 0.

```
chisq.test(x = rNCVS2016$Male,
           y = rNCVS2016$PayCred)
```

The test statistic was $\chi^2 = 0.54045$

Step 3: Determine the Probability (P-Value) of Obtaining a Test Statistic at Least as Extreme as the Observed Value, Assuming no Relationship Exists.

The probability of observing a chi-squared value of 0.54 in our sample—assuming no association between men and women in the population in using credit cards for online purchases in the past 12 months—is calculated to be 0.4622, indicating a p-value greater than 0.05.

Step 4: If the P-Value Is very small, Typically Less Than 5%, Reject the Null Hypothesis.

Step 4 is not relevant in this situation.

Step 5: If the P-Value is not small, Typically 5% or Greater, Retain the Null Hypothesis.

The probability that the null hypothesis, stating "The usage of credit cards for online purchases in the past 12 months is the same across men and women," holds true in the population, based on our sample data, is calculated to be 0.4622, indicating a p-value greater than 0.05. This relatively high probability suggests that the null hypothesis is likely true and should not be rejected.

Reporting the Results

So, how do we write up the results based on our test? We conducted a chi-squared test to examine the null hypothesis, which posited no association between using credit cards for online purchases in the past 12 months and gender. Our analysis failed to reject the null hypothesis, suggesting no statistically significant association between the two variables [$\chi^2(1) = 0.54$; $p > .05$]. For a demonstration of how a chi-square test is applied and reported, I recommend reviewing the work of Choi and Han (2022). Their study provides a clear example of applying and reporting a chi-square test using the data from NCVS.

Conclusion

In this chapter, we conducted first hypothesis testing using a chi-squared test. In the next chapter, we will use a different statistical test called t-test to compare different groups.

References

- Choi, J., & Han, S. (2022). Exploring gender disparity in capable guardianship against identity theft: A focus on internet-based behavior. *International Journal of Criminal Justice*, 4(1), 25-48.
- Vogt, W. P., & Johnson, R. B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (4th ed.). Sage.

This page titled [1.6: Chi-Squared Test](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jaeyong Choi \(The Pennsylvania Alliance for Design of Open Textbooks \(PA-ADOPT\)\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.