Hide menu

The chi-squared test

Video: Welcome to module 4
3 min

Video: Hypothesis testing with chi-squared
5 min

Reading: Chi-squared tests: Goodness of fit versus independence
20 min

Practice Quiz: Test your knowledge: The chi-squared test
6 min

Analysis of variance

ANCOVA, MANOVA, and MANCOVA

Review: Advanced hypothesis testing

Chi-squared tests: Goodness of fit versus independence

In the previous course, you learned how hypothesis tests are used to see significant differences among groups. Chi-squared tests are used to determine whether one or more observed categorical variables follow expected distribution(s). For example, you may expect that 50% more movie goers attend movies on weekends in comparison to weekdays. After observing movie goers attendance for a month, you then can perform a chi-squared test to see if your initial hypothesis was correct.

This reading will cover the two main chi-squared tests—goodness of fit and test for independence—which can be used to test your expected hypothesis against what actually occurred. Data professionals perform these hypothesis tests to offer organizations actionable insights that drive decision making.

The Chi-squared goodness of fit test

Chi-squared (χ^2) goodness of fit test is a hypothesis test that determines whether an observed categorical variable with more than two possible levels follows an expected distribution. The null hypothesis (H_0) of the test is that the categorical variable follows the expected distribution. The alternative hypothesis (H_a) is that the categorical variable does not follow the expected distribution. Consider the scenario in this reading that will define the null and alternative hypotheses based on the scenario, set up a Goodness of Fit test, evaluate the test results, and draw a conclusion.

Chi-squared goodness of fit scenario

Imagine that you work as a data professional for an online clothing company. Your boss tells you that they expect the number of website visitors to be the same for each day of the week. You decide to test your boss's hypothesis and pull data every day for the next week and record the number of website visitors in the table below:

Day of the Week	Observed Values
Sunday	650
Monday	570
Tuesday	420
Wednesday	480
Thursday	510
Friday	380
Saturday	490