

# Categorical Data and The Chi-Squared Test: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2020

## Syntax

- The `pchisq()` function allows us to calculate the cumulative probability for a chi-squared distribution given a particular value and the degrees of freedom.

## Concepts

- The chi-squared test enables us to quantify the difference between sets of observed and expected categorical values to determine statistical significance.
- To calculate the chi-squared test statistic, we use the following formula:  $\chi^2 = \sum \frac{(O - E)^2}{E}$ .
- The formula above follows a chi-squared distribution, which is the square of a normal distribution.
- A p-value allows us to determine whether the difference between two values is due to chance, or due to an underlying difference.
- The chi-squared test is sensitive to small sample sizes. If our sample is too small, we run the risk of accidentally rejecting the null hypothesis when we shouldn't.
- A degree of freedom is the number of values that contribute to the test statistic, minus one.

## Resources

- [Chi-Square Test](#)
- [Degrees of Freedom](#)



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2020