



STHDA 


Statistical tools for high-throughput data analysis

Licence: 

Search... 

[Home](#) | [Basics](#) | [Data](#) | [Visualize](#) | [Analyze](#) | [Resources](#) | [Our Products](#) | [Support](#) | [About](#)

[Home](#) / [Easy Guides](#) / [R software](#) / [R Basic Statistics](#) / [Comparing Proportions in R](#) / [Chi-square Goodness of Fit Test in R](#)

 Actions menu for module Wiki



Chi-square Goodness of Fit Test in R

 Tools

- [What is chi-square goodness of fit test?](#)
- [Example data and questions](#)
- [Statistical hypotheses](#)
- [R function: chisq.test\(\)](#)
 - [Answer to Q1: Are the colors equally common?](#)
 - [Answer to Q2 comparing observed to expected proportions](#)
 - [Access to the values returned by chisq.test\(\) function](#)
- [See also](#)
- [Infos](#)

What is chi-square goodness of fit test?

The **chi-square goodness of fit** test is used to compare the observed distribution to an expected distribution, in a situation where we have two or more categories in a discrete data. In other words, it compares multiple observed proportions to expected probabilities.





Compare Multiple Observed Proportions to Expected Probabilities

- + Definition
- + Research Questions & Statistics
- + Practical Examples in R
- + Interpret



© sthda.com 2016

Example data and questions

For example, we collected wild tulips and found that 81 were red, 50 were yellow and 27 were white.

1. Question 1:

? Are these colors equally common?

If these colors were equally distributed, the expected proportion would be 1/3 for each of the color.

2. Question 2:

Suppose that, in the region where you collected the data, the ratio of red, yellow and white tulip is 3:2:1 ($3+2+1 = 6$). This means that the expected proportion is:

- $3/6 (= 1/2)$ for red
- $2/6 (= 1/3)$ for yellow
- $1/6$ for white

? We want to know, if there is any significant difference between the observed proportions and the expected proportions.

Statistical hypotheses

- *Null hypothesis* (H_0): There is no significant difference between the observed and the expected value.
- *Alternative hypothesis* (H_a): There is a significant difference between the observed and the expected value.

R function: `chisq.test()`

The R function `chisq.test()` can be used as follow:

```
chisq.test(x, p)
```

- **x**: a numeric vector
- **p**: a vector of probabilities of the same length of x.

Answer to Q1: Are the colors equally common?

```
tulip <- c(81, 50, 27)
res <- chisq.test(tulip, p = c(1/3, 1/3, 1/3))
res
```

```
Chi-squared test for given probabilities
data: tulip
X-squared = 27.886, df = 2, p-value = 8.803e-07
```

The function returns: the value of chi-square test statistic ("X-squared") and a p-value.





 Note that, the chi-square test should be used only when all calculated expected values are greater than 5.


```
# Access to the expected values
res$expected
```

```
[1] 52.66667 52.66667 52.66667
```

Answer to Q2 comparing observed to expected proportions

```
tulip <- c(81, 50, 27)
res <- chisq.test(tulip, p = c(1/2, 1/3, 1/6))
res
```

```
Chi-squared test for given probabilities
data: tulip
X-squared = 0.20253, df = 2, p-value = 0.9037
```

 The **p-value** of the test is 0.9037, which is greater than the significance level $\alpha = 0.05$. We can conclude that the observed proportions are not significantly different from the expected proportions.

Access to the values returned by chisq.test() function

The result of **chisq.test()** function is a list containing the following components:

- **statistic**: the value the chi-squared test statistic.
- **parameter**: the degrees of freedom
- **p.value**: the **p-value** of the test
- **observed**: the observed count
- **expected**: the expected count

The format of the **R** code to use for getting these values is as follow:

```
# printing the p-value
res$p.value
```

```
[1] 0.9036928
```

```
# printing the mean
res$estimate
```

NULL

See also

- [One Proportion Z-Test in R: Compare an Observed Proportion to an Expected One](#)
- [Two Proportions Z-Test in R: Compare Two Observed Proportions](#)
- [Chi-Square Test of Independence in R: Evaluate The Association Between Two Categorical Variables](#)

Infos

 This analysis has been performed using **R software** (ver. 3.2.4).

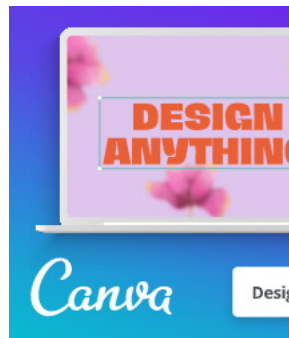
 Enjoyed this article? I'd be very grateful if you'd help it spread by emailing it to a friend, or sharing it on Twitter, Facebook or Linked In.

Show me some love with the like buttons below... Thank you and please don't forget to share and comment below!!

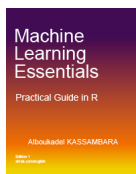




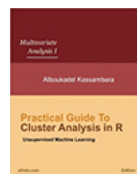
Share 10 Like 10 Tweet **Share** Save Share 17



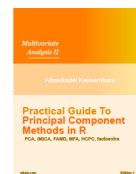
Recommended for You!



Machine Learning Essentials: Practical Guide in R



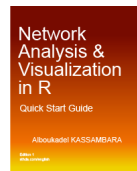
Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R



R Graphics Essentials for Great Data Visualization



Network Analysis and Visualization in R



More books on R and data science

Recommended for you

✓ This section contains best data science and self-development resources to help you on your path.

Coursera - Online Courses and Specialization

Data science

- [Course: Machine Learning: Master the Fundamentals](#) by Stanford
- [Specialization: Data Science](#) by Johns Hopkins University
- [Specialization: Python for Everybody](#) by University of Michigan
- [Courses: Build Skills for a Top Job in any Industry](#) by Coursera
- [Specialization: Master Machine Learning Fundamentals](#) by University of Washington
- [Specialization: Statistics with R](#) by Duke University
- [Specialization: Software Development in R](#) by Johns Hopkins University
- [Specialization: Genomic Data Science](#) by Johns Hopkins University

Popular Courses Launched in 2020

- [Google IT Automation with Python](#) by Google
- [AI for Medicine](#) by deeplearning.ai
- [Epidemiology in Public Health Practice](#) by Johns Hopkins University
- [AWS Fundamentals](#) by Amazon Web Services

Trending Courses

- [The Science of Well-Being](#) by Yale University
- [Google IT Support Professional](#) by Google
- [Python for Everybody](#) by University of Michigan
- [IBM Data Science Professional Certificate](#) by IBM





Books - Data Science

Our Books

- [Practical Guide to Cluster Analysis in R](#) by A. Kassambara (Datanovia)
- [Practical Guide To Principal Component Methods in R](#) by A. Kassambara (Datanovia)
- [Machine Learning Essentials: Practical Guide in R](#) by A. Kassambara (Datanovia)
- [R Graphics Essentials for Great Data Visualization](#) by A. Kassambara (Datanovia)
- [GGPlot2 Essentials for Great Data Visualization in R](#) by A. Kassambara (Datanovia)
- [Network Analysis and Visualization in R](#) by A. Kassambara (Datanovia)
- [Practical Statistics in R for Comparing Groups: Numerical Variables](#) by A. Kassambara (Datanovia)
- [Inter-Rater Reliability Essentials: Practical Guide in R](#) by A. Kassambara (Datanovia)

Others

- [R for Data Science: Import, Tidy, Transform, Visualize, and Model Data](#) by Hadley Wickham & Garrett Golemund
- [Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems](#) by Aurelien Géron
- [Practical Statistics for Data Scientists: 50 Essential Concepts](#) by Peter Bruce & Andrew Bruce
- [Hands-On Programming with R: Write Your Own Functions And Simulations](#) by Garrett Golemund & Hadley Wickham
- [An Introduction to Statistical Learning: with Applications in R](#) by Gareth James et al.
- [Deep Learning with R](#) by François Chollet & J.J. Allaire
- [Deep Learning with Python](#) by François Chollet

Want to Learn More on R Programming and Data Science?

Follow us [by Email](#)

[Subscribe](#)
by [FeedBurner](#)

On Social Networks:
on Social Networks



Get involved :



Click to **follow us** on [Facebook](#) and [Google+](#) :



Comment this article by clicking on "Discussion" button (top-right position of this page)

This page has been seen 230411 times

Sign in

Login

Password

Auto connect



Register



Forgotten password

Welcome!

[Want to Learn More on R Programming and Data Science?](#)

Follow us [by Email](#)

[Subscribe](#)
by [FeedBurner](#)





Click to see our collection of resources to help you on your path...



Course & Specialization

Recommended for You (on Coursera):

- Course: Machine Learning: Master the Fundamentals
- Specialization: Data Science
- Specialization: Python for Everybody
- Course: Build Skills for a Top Job in any Industry
- Specialization: Master Machine Learning Fundamentals
- Specialization: Statistics with R
- Specialization: Software Development in R
- Specialization: Genomic Data Science

[See More Resources](#)

Adobe

**Get 40% off
Creative Cloud.**

Claim our
introductory offer.
T&Cs apply.

[Buy now](#)


The advertisement features the Adobe logo at the top, followed by a bold headline and a sub-headline. A blue button with white text is positioned below the text. The background of the ad is a light peach color with a decorative illustration of a giraffe and plants at the bottom.

 `factoextra`

 `survminer`

 `ggpubr`

 `ggcorrplot`

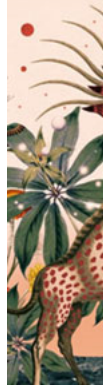
 `fastqcr`



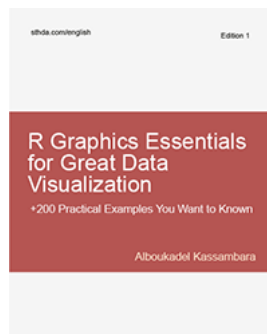
Get 40% off
Creative Cloud.

Claim our
introductory offer.
T&Cs apply.

Buy now

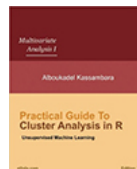


Our Books

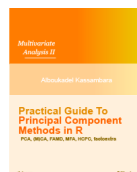


R Graphics Essentials for Great Data Visualization: 200 Practical Examples You Want to Know for Data Science

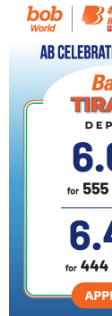
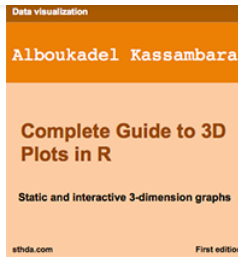
★ **NEW!!**



Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R



 [Datanovia: Online Data Science Courses](#)

 [R-Bloggers](#)

Newsletter

Email



Boosted by PHPBoost

Recommended for you



Reading Data From
Excel Files (xls | xlsx) in...

www.sthda.com



ggplot2 barplots : Quick
start guide - R softwar...

www.sthda.com



Correlation Test
Between Two Variable...

www.sthda.com



ggplot2 histogram plot :
Quick start guide - R s...

www.sthda.com

AddThis

