<https://translate.google.com/translate?hl=pt-BR&sl=en&u=https://stats.stackexchange.com/q/173893&prev=search> (explica o que é 2.2e-16 )

variaveis no crivo

rendaDaFamilia = renda total

rendaDivPessoa renda por quantidade de pessoa

#### **O Significado e a Interpretação dos P-values (o que os dados dizem?)**

O P-value, que depende diretamente de uma dada amostra, tenta fornnecer uma medida da força dos resultados de um teste, em contraste a uma simples rejeição ou não rejeição. Se a hipótese nula for verdadeira e a chance da variação aleatória for a única razão para as diferenças amostrais, então o P-value é uma medida quantitativa para alimentar o processo de tomada de decisão como evidência. A tabela seeguinte fornece uma interpretação razoável dos P-values:

| P-value | Interpretação |
| --- | --- |
| P< 0,01 | evidência muito forte contra H0 |
| 0,01< = P < 0,05 | evidência moderada contra H0 |
| 0,05< = P < 0,10 | evidência sugestiva contra H0 |
| 0,10< = P | pouca ou nenhuma evidência real contra H0 |

Esta interpretação é largamente aceita, e muitos jornais científicos rotineiramente publicam artigos usando tal interpreatação do resultado do teste de hipótese.

Para amostra de tamanho fixo, quando o número de realizações é decidido antecipadamente, a distribuição de p é uniforme (assumindo a hipótese nula). Expressariamos isto como P(p < x) = x. Isto significa que o critério de p <0,05 atinge um de 0,05.

Um p-value é uma medida de quanta evidência você tem contra a hipótese nula. Quanto menor o p-value, mais evidência você tem. Deve-se combinar o p-value com o nível de significância para tomar decisão sobre um dado teste de hipótese. Em tal caso, se o p-value for menor que algum corte (usualmente 0,05, algumas vezes um pouco mais como 0,1 ou um pouco menos como 0,01) então você rejeita a hipótese nula.

Entendido que a distribuição dos p-values sob hipótese nula H0 é uniforme, e então não depender de uma forma particular do teste estatístico. Num teste de hipótese estatístico, o P value é a probabilidade de observar um teste estatístico no mínimo como extremo o valor realmente observado, assumindo que a hipótese nula seja verdadeira. O valor de p é definido com respeito a uma distribuição. Portanto, podemos chamá-lo "hipótese de modelo-distribucional " ao invés de "a hipótese nula".

Em resumo, ele simplesmente significa que se a nula tivesse sido verdadeira, o p value é a probabilidade contra a nula naquele caso. O p-value é determinado pelo valor observado, entretanto, isto torna difícil para estabelecer o inverso de p.

Ler o seguinte artigo:

Arsham H., Kuiper's P-value as a Measuring Tool and Decision Procedure for the Goodness-of-fit Test, *Journal of Applied Statistics*, Vol. 15, No.3, 131-135, 1988.

All contents copyright (c) 2012.

All rights reserved.

Esta página foi alterada pela última vez em 8/07/12

<http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis5.html>



Categorical Data Analysis

[print all](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis_print.html)

[skip navigation](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis_print.html)

* **[Prev](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis4.html)**
* **[Next](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis6.html)**
* [1](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/index.html)
* **|**  [2](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis2.html)
* **|**  [3](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis3.html)
* **|**  [4](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis4.html)
* **| 5**
* **|**  [6](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis6.html)
* **|**  [7](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis7.html)
* **Contents**

More Resources sidebar

**[Intro to R Contents](http://cran.r-project.org/doc/manuals/R-intro.html)**

**[Common R Commands](http://www.personality-project.org/r/r.commands.html)**

# **Tests of Single Proportions**

Calculating odds and risk ratios only gives an indication of whether a potential cause is related to the outcome. To be more specific, we can do tests on groups with different exposures with regard to their outcomes. First, let us introduce the idea of testing for proportions, from the simplest scenario.

Tests of single proportions are generally based on the binomial distribution with size parameter N and probability parameter p. For large sample sizes, this can be well approximated by a normal distribution with mean N\*p and variance N\*p(1 − p). As a rule of thumb, the approximation is satisfactory when the expected numbers of "successes" and "failures" are both larger than 5. The normal approximation can be somewhat improved by the Yates correction (aka continuity correction), which shrinks the observed value by half a unit towards the expected value when calculating the test statistic (by default, this correction is used; it can also be turned off by using "correct = F").

In the outbreak data set, 447 of the 998 individuals who ate beef curry were observed to have food poisoning symptoms, and one may want to test the hypothesis that the probability of a "random individual who ate beef curry" having food poisoning is 0.1.

* H0: The proportion of individuals who eat beefcurry and get sick is 0.1: true p = 0.1
* Ha: The proportion of individuals who eat beefcurry and get sick is not 0.1: true p ≠ 0.1

These hypotheses can be tested using prop.test. The three arguments to prop.test are the number of positive outcomes, the total number, and the (theoretical) probability parameter that you want to test for. The latter is 0.5 by default (OK for symmetric problems).

> prop.test(447, 998, .1)

1-sample proportions test with continuity correction

data: 447 out of 998, null probability 0.1

X-squared = 1338.242, df = 1, p-value < 2.2e-16

alternative hypothesis: true p is not equal to 0.1

95 percent confidence interval:

0.4168064 0.4793912

sample estimates:

p

0.4478958

**Conclusion:**

We reject the null hypothesis (χ12 = 1338.242, df = 1, p-value < 2.2e-16). The estimated proportion of people who ate beef curry is 0.448 (95% CI: 0.42, 0.49).

# **Tests for Two Independent Proportions**

The function prop.test can also be used to compare two or more proportions, which can help answer more interesting questions for the outbreak data. For comparing two proportions, the arguments are given as two vectors, where the first vector contains the number of positive outcomes in each group, and the second vector the total number for each group.

Suppose we want to test the hypothesis that gender is associated with developing food poisoning based on the outbreak data. Specifically, we are interested in determining whether men are at a higher risk for developing food poisoning than women (this should be our "test" hypothesis). The relevant hypotheses are as follows:

* H0: The proportion of males who have gastrointestinal illness is **less than or equal to** the proportion of females who have gastrointestinal illness.
* Ha: The proportion of males who have gastrointestinal illness is **greater than** the proportion of females who have gastrointestinal illness.

We need to construct two vectors first:

> male.cases = length(which(case == 1 & sex == 1))

> female.cases = length(which(case == 1 & sex == 0))

> people.cases = c(male.cases, female.cases)

> male.total = length(which(sex==1))

> female.total = length(which(sex==0))

> people.total= c(male.total, female.total)

Now we will do a two-sample test for proportions (note the one-sided alternative here!)

> prop.test(people.cases, people.total, alternative = "greater")

2-sample test for equality of proportions with continuity correction

data: people.cases out of people.total

X-squared = 8.3383, df = 1, p-value = 0.001941

alternative hypothesis: greater

95 percent confidence interval:

0.03998013 1.00000000

sample estimates:

prop 1 prop 2

0.4604716 0.3672922

**Conclusion:** We reject the null hypothesis, and conclude that the proportion of males who have gastrointestinal illness is greater than the proportion of females with gastrointestinal illness (χ12 = 8.34, p-value = 0.0019). The estimated proportion of males with gastrointestinal illness is 0.46, while the estimated proportion of females with gastrointestinal illness is 0.37. The 95% CI for the difference between the proportions is (0.04, 1.00). Note that this CI excludes 0, and so is concordant with our decision to reject the null based on the p-value.

The above test uses approximations, which may not be accurate if the sample sizes are small. If you want to be sure that at least the p-value is correct, you can use Fisher's exact test. The relevant function is **fisher.test**, which requires that data be given in matrix form. **The second column of the table needs to be the number of negative outcomes, not the total number of observations**. This is obtained as follows:

> cases.matrix = matrix(c(male.cases, female.cases, male.total - male.cases, female.total - female.cases),2,2)

> fisher.test(cases.matrix, alternative="greater)

Fisher's Exact Test for Count Data

data: cases.matrix

p-value = 0.001881

alternative hypothesis: true odds ratio is greater than 1

95 percent confidence interval:

1.176021 Inf

sample estimates:

odds ratio

1.469679

Notice that in this case the p-values from Fisher's exact test and the normal approximation are *very* close, as expected by the large sample sizes.

The standard chi-square (χ2) test in **chisq.test** performs chi-squared contingency table tests and goodness-of-fit tests. It works with data in matrix form, just as fisher.test does. For a 2×2 table the test is exactly equivalent to **prop.test** (except that this is always for a two-sided alternative!).

> chisq.test(cases.matrix)

Pearson's Chi-squared test with Yates' continuity correction

data: cases.matrix

X-squared = 8.3383, df = 1, p-value = 0.003882

| Based on the outbreak data, carry out an appropriate test and report the results for testing the hypothesis that people who drank water were more likely to get diarrhea than those who did not. |
| --- |

[return to top](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis5.html#) | [previous page](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis4.html) | [next page](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis6.html)

Content ©2016. Some Rights Reserved.

Date last modified: January 6, 2016.

Boston University School of Public Health

Creative Commons license Attribution Non-commercial

hidden textpopper

retirado de <http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R6_CategoricalDataAnalysis/R6_CategoricalDataAnalysis5.html>

correlação e regressão

<https://slideplayer.com.br/slide/4259024/#.XFjQEXnQmQ8.google>

<http://www.leg.ufpr.br/~paulojus/CE701/ce701/node8.html>