



R news and tutorials contributed by (573) R bloggers

- [Home](#)
- [About](#)
- [RSS](#)
- [add your blog!](#)
- [Learn R](#)
- [R jobs](#)
- [Contact us](#)

Welcome!

Follow @rbloggers { 28.8K }


Here you will find daily **news and tutorials about R**, contributed by over 573 bloggers. There are many ways to **follow us -**

[By e-mail:](#)

25779 readers


BY FEEDBURNER

[On Facebook:](#)



R blogg...
30k likes

Be the first of your friends to like this



If you are an R blogger yourself you are invited to [add your own R content feed to this site](#) (Non-English R bloggers should add themselves- [here](#))

[Jobs for R-users](#)

- [Root word and Noun Phrase Extraction](#)
- [Data analyst](#)
- [Statistician R developer for agriculture statistics](#)
- [Freelance Intern R and shiny developer](#)
- [R Shiny Developer](#)

Popular Searches

- [web scraping](#)
- [heatmap](#)
- [maps](#)
- [twitter](#)
- [alt=](#)

- [shiny](#)
- [boxplot](#)
- [hadoop](#)
- [time series](#)
- [animation](#)
- [ggplot2](#)
- [latex](#)
- [ggplot](#)
- [trading](#)
- [excel](#)
- [finance](#)
- [pca](#)
- [quantmod](#)
- [googlevis](#)
- [RStudio](#)
- [how to import image file to R](#)
- [market research](#)
- [rattle](#)
- [eclipse](#)
- [knitr](#)
- [SQL](#)
- [coplot](#)
- [rcmdr](#)
- [tutorial](#)
- [map](#)

Recent Posts

- [tweets of ggtree](#)
- [Latest Redmonk and Tiobe language rankings for R](#)
- [Business Intelligence and Data Science Groups in East Tennessee](#)
- [Nairobi Data Science Meet Up: Finding deep structures in data with Chris Orwa](#)
- [Advanced Analytics Software's Most Important Feature? Gartner Says it's VCF](#)
- [Calling RSiteCatalyst From Python](#)
- [Hillary 1993: Largest Drop in Girl Names EVER: Chelsea Distant Second](#)
- [Bind exercises](#)
- [Post your resume to R-users.com \(+6 new R jobs for 2016-02-22\)](#)
- [The R-Podcast Episode 18: Interviews with the RStudio Team](#)
- [more e's \[and R's\]](#)
- [Convert Car2go Austin into an EV Fleet](#)
- [Large scale eigenvalue decomposition and SVD with rARPACK](#)
- [inegiR version 1.2](#)
- [Working with the Clinton State Dept Email Dumps in R \(Part 1: Graphs\)](#)

Other sites

- [Jobs for R-users](#)
- [SAS blogs](#)
- [Statistics of Israel](#)

One-proportion z test in R

March 16, 2015

By [Nicole Radziwill](#)

Like Share {29} Tweet [Share](#) 5

(This article was first published on [Quality and Innovation » R](#), and kindly contributed to [R-bloggers](#))

For quite a while, I've been confused by the behavior of the `prop.test` function in R: 1) the p-values always seem to be slightly off, and 2) the confidence intervals never seem to be symmetric around my estimate of the population proportion. Today, I decided I was going to figure out what was up... and whether I should trust my own analytical calculations for this test or what I get in R. I started by checking [R Bloggers](#) to see if anyone had explored this problem, but [I only found one post about the two proportion z-test](#), and found none focusing on the one proportion z-test.

Here's the essence of the test: you observe x subjects out of a sample size of n observations. Perhaps you just want to create a confidence interval to get a sense of where the true population proportion lies. Or maybe, you'd like to test the null hypothesis that the true population proportion is some value p_0 , against the alternative that it is really greater than, less than, or not equal to that value.

To do this, you can use the normal approximation to the binomial distribution to calculate a test statistic z . The estimated population proportion, **p-hat**, is just x/n . You have to find the standard error (SE) of your estimate as well to get the value of z :

$$z = \frac{\hat{p} - p_0}{SE(\hat{p})} = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

The confidence interval is constructed from the estimated population proportion, and the margin of error, which is determined by the standard error of the estimate and a scaling factor (z^*) that establishes the width of the confidence interval:

CI : Estimate \pm Margin of Error (ME)

$$\hat{p} \pm z^* SE(\hat{p})$$

$$\hat{p} \pm z^* \sqrt{\frac{pq}{n}}$$

Here's an example. Let's say I know that nationally, 90% of home buyers would select a home with energy efficient features, even paying a premium of 2-3%, to reduce their utility costs long term. As the hypothetical president of an engineering firm that installs these energy efficient features, I want to know if less than 90% of buyers in my local area feel this way (because if so, I might want to launch an education campaign to help my business). I randomly sample 360 potential buyers and find that 312 of them would indeed select a home with energy efficient features.

I want to test my sample against the null hypothesis that the true population proportion is 0.90, with the alternative that the proportion is less than 0.90. I use an estimated population proportion (**p-hat**) of **312/360=0.867**, and using the equations above, find that my test statistic z turns out to be **-2.108**, with a corresponding **p-value** of **0.0175**. I reject the null hypothesis that the true population proportion is 0.90 in favor of the alternative, and start making plans to launch my education program.

(I don't get the same values when I use `prop.test`. But later in this post, I'll show you why *I shouldn't have expected that in the first place!!*)

I wrote an R function to automate my z test (and also, to double check my calculations):

```
z.test <- function(x,n,p=NULL,conf.level=0.95,alternative="less") {
```

```

ts.z <- NULL
cint <- NULL
p.val <- NULL
phat <- x/n
qhat <- 1 - phat
# If you have p0 from the population or H0, use it.
# Otherwise, use phat and qhat to find SE.phat:
if(length(p) > 0) {
  q <- 1-p
  SE.phat <- sqrt((p*q)/n)
  ts.z <- (phat - p)/SE.phat
  p.val <- pnorm(ts.z)
  if(alternative=="two.sided") {
    p.val <- p.val * 2
  }
  if(alternative=="greater") {
    p.val <- 1 - p.val
  }
} else {
  # If all you have is your sample, use phat to find
  # SE.phat, and don't run the hypothesis test:
  SE.phat <- sqrt((phat*qhat)/n)
}
cint <- phat + c(
  -1*((qnorm(((1 - conf.level)/2) + conf.level))*SE.phat),
  ((qnorm(((1 - conf.level)/2) + conf.level))*SE.phat) )
return(list(estimate=phat,ts.z=ts.z,p.val=p.val,cint=cint))
}

```

When I run it on my example above, I get what I expect:

```

> z.test(312,360,p=0.9)
$estimate
[1] 0.867

$ts.z
[1] -2.11

$p.val
[1] 0.0175

$cint
[1] 0.836 0.898

```

But when I run prop.test, this isn't what I see at all!

```

> prop.test(312,360,p=0.9)

1-sample proportions test with continuity correction

data: 312 out of 360, null probability 0.9
X-squared = 4.08, df = 1, p-value = 0.04335
alternative hypothesis: true p is not equal to 0.9
95 percent confidence interval:
 0.826 0.899
sample estimates:
 p
0.867

```

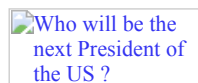
Why are there differences? Here's what I found out:

- The **prop.test** function doesn't even do a z test. It does a Chi-square test, based on there being one categorical variable with two states (success and failure)! Which I should have noticed immediately from the line that starts with "X-squared" 😊
- The [Yates continuity correction](#), which adjusts for the differences that come from using a normal approximation to the binomial distribution, is also automatically applied. This removes $0.5/n$ from the lower limit of the confidence interval, and adds $0.5/n$ to the upper limit.
- Most notably, I was *really* disturbed to see that the confidence interval given by **prop.test** wasn't even centered around the sample estimate, **p-hat**. Oh no! But again, this is no concern, since prop.test uses the [Wilson score interval](#) to construct the confidence interval. This results in an asymmetric, but presumably more accurate (with respect to the real population) confidence interval.

Moral of the story? **There are many ways to test an observed proportion against a standard, target, or recommended value. If you're doing research, and especially if you're using R, be sure to know what's going on under the hood before you report your results.**

Like Share 29 Tweet Share 5

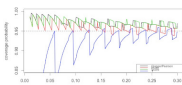
Related



Who will be the next President of the US ?
In "R bloggers"



Want to say one thing and the exact opposite with strong confidence ?
In "R bloggers"



Example 9.37:
(Mis)behavior of binomial confidence intervals
In "R bloggers"

29 Tweet 5
Like Share

To leave a comment for the author, please follow the link and comment on their blog: [Quality and Innovation » R](#)

R-bloggers.com offers [daily e-mail updates](#) about [R](#) news and [tutorials](#) on topics such as: [Data science](#), [Big Data](#), [R jobs](#), visualization ([ggplot2](#), [Boxplots](#), [maps](#), [animation](#)), programming ([RStudio](#), [Sweave](#), [LaTeX](#), [SQL](#), [Eclipse](#), [git](#), [hadoop](#), [Web Scraping](#)) statistics ([regression](#), [PCA](#), [time series](#), [trading](#)) and more...

If you got this far, why not **subscribe for updates** from the site?
Choose your flavor: [e-mail](#), [twitter](#), [RSS](#), or [facebook](#)...

Like Share 29 Tweet Share 5

Comments are closed.

Search & Hit Enter

Recent popular posts

- [Working with the Clinton State Dept Email Dumps in R \(Part 1: Graphs\)](#)
- [Advanced Analytics Software's Most Important Feature? Gartner Says it's VCF](#)

Most visited articles of the week

1. [The Simple Reason Sanders Is Winning](#)
2. [Working with the Clinton State Dept Email Dumps in R \(Part 1: Graphs\)](#)
3. [Installing R packages](#)
4. [In-depth introduction to machine learning in 15 hours of expert videos](#)
5. [Using apply, sapply, lapply in R](#)
6. [How to Learn R](#)
7. [How I build up a ggplot2 figure](#)
8. [Computing and visualizing PCA in R](#)
9. [How to Make a Histogram with Basic R](#)

Sponsors

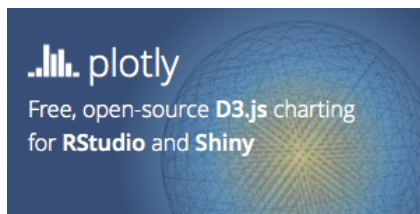




Learn R By Doing
www.DataCamp.com

Start Free Course

The banner features a blue background with a white grid and a line graph showing data trends.



plotly
Free, open-source **D3.js** charting
for **RStudio** and **Shiny**

The banner shows a dark blue background with a glowing, circular data visualization.


[Plotly: collaborative, publication-quality graphing.](#)



R training
R consulting

QUANTIDE
knowledge from data

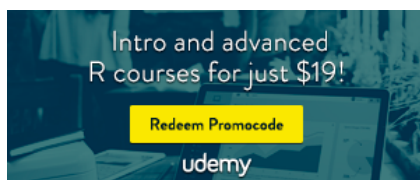
The banner has a blue background with a large, stylized 'Q' and the text 'QUANTIDE'.



New in **RStudio Server Pro**
Collaboration | Multiple Sessions | R Versions

RStudio Free Evaluation

The banner features the RStudio logo and a dark blue background.

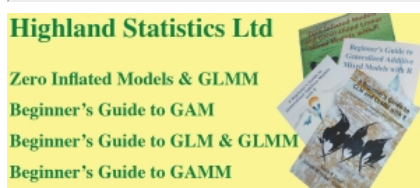


Intro and advanced
R courses for just \$19!

Redeem Promocode

udemy

The banner shows a laptop screen with a course listing and the Udemy logo.



Highland Statistics Ltd

Zero Inflated Models & GLMM
Beginner's Guide to GAM
Beginner's Guide to GLM & GLMM
Beginner's Guide to GAMM

The banner includes images of book covers for various statistical guides.



Werden Sie zum Expe[R]ten mit der
R-Akademie von

eoda
daten » wissen » nutzen

Beratung | Software
Training | Lösungen

The banner features the eoda logo and a green background.



STATISTICS
VIEWS

Bringing Statistics Together

The banner has a white background with a bar chart and the text 'STATISTICS VIEWS'.

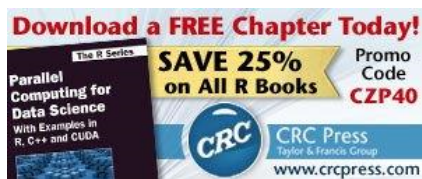
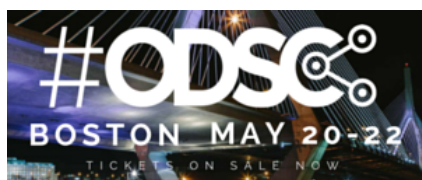


**BECOME A DATA
SCIENTIST**

12 Weeks to A New Career

NYC DATA SCIENCE
ACADEMY

The banner shows a city skyline at night with the NYC Data Science Academy logo.



[Contact us](#) if you wish to help support R-bloggers, and place your banner here.

[Jobs for R users](#)

- [Root word and Noun Phrase Extraction](#)
- [Data analyst](#)
- [Statistician R developer for agriculture statistics](#)
- [Freelance Intern R and shiny developer](#)
- [R Shiny Developer](#)
- [Data Visualization Intern R Shiny](#)
- [Senior Statistician/Statistician \(>\\$100/year\) @ Shinjuku-ku, Tōkyō-to, Japan](#)

Search & Hit Enter

[Full list of contributing R-bloggers](#)

R-bloggers was founded by [Tal Galili](#), with gratitude to the [R](#) community.

Is powered by [WordPress](#) using a [bavotasan.com](#) design.

Copyright © 2016 R-bloggers. All Rights Reserved. [Terms and Conditions](#) for this website