

Ask Question

# P-value for point biserial correlation in R

Asked 4 years, 2 months ago Active 7 months ago Viewed 9k times



Does anybody know of an R package that produces a p-value for point biserial correlations?

3

I've tried all of the major packages that I know (with some help from Google) and haven't found any. If a package doesn't come to mind, is there some way that I can intuitively calculate the p-value?



correlation p-value

share cite improve this question follow

edited Apr 11 '18 at 13:48

asked Jul 28 '16 at 18:20

user2917781 **323** 🔲 3 📒 13

add a comment

# 3 Answers





The point-biserial correlation is equivalent to calculating the Pearson correlation between a continuous and a dichotomous variable (the latter needs to be encoded with 0 and 1). Therefore, you can just use the standard cor.test function in R, which will output the correlation, a 95% confidence interval, and an

This yields a correlation of r=0.202, which is not significant (t=1.429,  $\mathrm{df}=48$ , p=0.1595):

As @sal-mangiafico and @igor-p point out, the function <code>biserial.cor</code> from the <code>ltm</code> package produces slightly different results. This is because <code>cor.test</code> uses the population standard deviation, whereas <code>biserial.cor</code> uses the sample standard deviation. Furthermore, the result of <code>biserial.cor</code> has the opposite sign than the result of <code>cor.test</code>. This can be adjusted by specifying the argument <code>level=2</code> in <code>biserial.cor</code>.

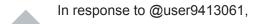
share cite improve this answer follow

edited Feb 18 at 9:00

answered Jul 13 '17 at 8:06

cbrnr
237 1 1 12

add a comment



- 2 I think I discovered the source of the problem.
- In the standard definition of biserial correlation, the population standard deviation is used.
- 1tm::biserial.cor uses the sample standard deviation.

In the following, a function is defined to calculate the population standard deviation. The function <code>biserial.cor.new</code> is defined, which is the same as <code>ltm::biserial.cor</code> with <code>sd.pop</code> used instead of <code>sd</code>.

I think biserial.cor.new will return the same result as cor.test.

```
sd.pop = function(x) {sd(x)*sqrt((length(x)-1)/length(x))}
biserial.cor.new =
function (x, y, use = c("all.obs", "complete.obs"), level = 1)
{
   if (!is.numeric(x))
       stop("'x' must be a numeric variable.\n")
   y <- as.factor(y)
   if (length(large colorele(x)) > 2)
```



```
stop(""x" and "y" do not have the same length")
use <- match.arg(use)
if (use == "complete.obs") {
    cc.ind <- complete.cases(x, y)
        x <- x[cc.ind]
    y <- y[cc.ind]
}
ind <- y == levs[level]
diff.mu <- mean(x[ind]) - mean(x[!ind])
prob <- mean(ind)
diff.mu * sqrt(prob * (1 - prob))/sd.pop(x)
}</pre>
```

#### And an example:

```
x = c(3,4,5,6,7,5,6,7,8,9)
y = c(0,0,0,0,0,1,1,1,1,1,1)
library(ltm)
### DIFFERENT RESULTS WITH ltm::biserial.cor
biserial.cor(x, y, level=2)
   ### [1] 0.5477226
cor.test(x,y)
   ### Pearson's product-moment correlation
   ### sample estimates:
   ### cor
   ### 0.5773503
### SAME RESULTS WITH new function
biserial.cor.new(x,y, level=2)
    ### [1] 0.5773503
cor.test(x,y)
   ### Pearson's product-moment correlation
   ### sample estimates:
   ### cor
   ### 0.5773503
```

share cite improve this answer follow

edited Apr 9 '18 at 23:41

answered Apr 9 '18 at 18:26



add a comment



For my understanding you don't have to code the dichotome variable with 0 and 1. Therefore using other values results in exactly the same output. Try for example:



Both gives you an r of 0.01732137. The only thing that can happen by coding the dichotome variable differently is that you get -0.01732137, which will be the case if the first number is bigger than the second, e.g.

```
y3 <- rep(c(0,1), 50)
cor.test(x, y3)
```

results in -0.01732137.

Furthermore, I read on different pages that "the point-biserial correlation is equivalent to calculating the Pearson correlation between a continuous and a dichotomous variable", but in fact I get different results if I conduct a Pearson and a point-biserial correlation on same data. An example:

```
x <- 1:100
y <- rep(c(0,1), 50)
cor.test(x, y)</pre>
```

gives me 0.01732137, but biserial.cor(x, y) results in -0.01723455.

I understand that it is okay to get positive and negative values, but the absolute value should be the same, which is not the case. The results are also different if I use other data, e.g. x <- rnorm(100, 100, 15) instead of x <- 1:100.

For this reason I am unsure whether it is acceptable to use <code>cor.test()</code> and report that you have conducted a point-biserial correlation.

Please state, which package you take biserial.cor from. I guess, it's the ltm package, but who knows. So the correlation is 0.017 in both cases, the rest probably some rounding error? If one function was superior, how do you know, which? – Bernhard Apr 9 '18 at 8:54 

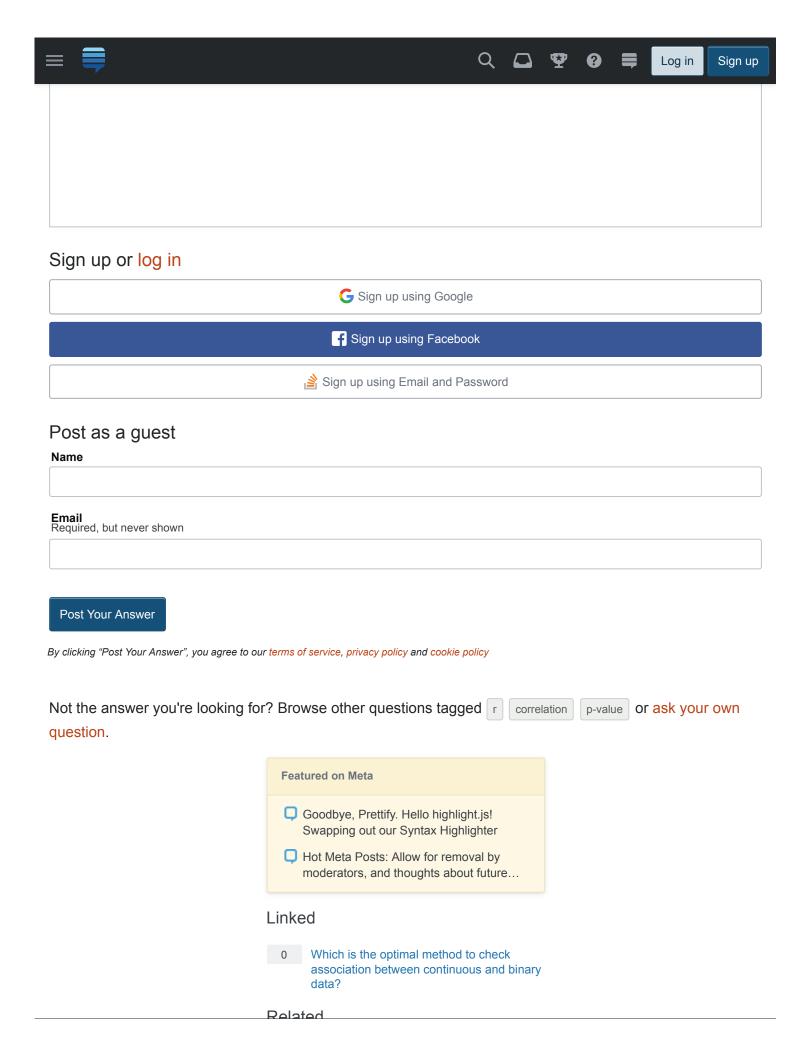
It appears that ltm::biserial.cor and stats::cor.test do not produce the same result. See ? ltm::biserial.cor for the formula used there. – Sal Mangiafico Apr 9 '18 at 15:43

Yes, I used the Itm package. – user203567 Apr 10 '18 at 9:41

### Your Answer

add a comment









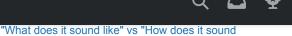
autoregressive parameters in R using GLS with ARMA(p=1) correlation

- Calculating average variance extracted (AVE) in R for checking discriminant validity (Fornell-Larcker criterion)
- What does correlation mean in error propagation?
- How to determine the cut-point of continuous predictor in survival analysis, optimal or median cut-point?
- C-index for survival analysis with confidence interval (R)
- 3 Understanding very high p value with Spearman's rank correlation
- 3 Correlation of random variables
- 0 Erdem correlation for time series

## Hot Network Questions

- What is the mean absolute difference between values in a normal distribution?
- Algebraic independence of shifts of the Riemann zeta function
- Is there a technical (or slang) term for triggering a game loss due to attempting to draw from an empty library?
- What book(s) would you recommend for structuring and pricing Exotic Products?
- Employer planning on making a change that I'm prepared to quit over. How should I tell manager?
- How can I identify the reason that makes a MILP model hard for solvers such as CPLEX?
- Why is Schenker so influential in US academia? Is it the same elsewhere?
- on this be formulated as one inequality
- How do I get rid of a mental barrier with respect to continuing research work that I have already done?
- Output the International Phonetic Alphabet
- Does having the (accurate) shape of an Aboleth give one all their ancestral-genetic memories?
- The randomness of modular squaring
- Why "Giraffe" as a name for the animal?
- Good name for object between squared and rounded
- Can employer legally stop paying time & 1/2 to exempt employee after stating in the offer that they would do so?





Log in

Sign up

- How can some USB 2.0 audio interfaces support phantom power through USB alone?
- What would happen if I don't replace worn drivetrain components?
- Did the switchblade in Se7en violate the Chekhov's gun principle?
- Why does the Amiga 500 have half the HSYNC pulses seperately compared to those in CSYNC?
- Grammar of Negative Verb + 限り
- What's the word for asking someone to deliver their promise?
- Why does the Quantum Realm behave different for Janet van Dyne than for Scott Lang?
- **Question feed**

like"

| CROSS VALIDATED        | COMPANY           | STACK EXCHANGE<br>NETWORK                                   | Blog Facebook Twitter LinkedIn Instagram  |
|------------------------|-------------------|---|---|
| Tour                   | Stack Overflow    | Technology > Life / Arts > Culture / Recreation > Science > |   |
| Help                   | For Teams         |   |   |
| Chat                   | Advertise With Us |   |   |
| Contact                | Hire a Developer  |   |   |
| Feedback               | Developer Jobs    |   |   |
| Mobile                 | About             | Other >   |   |
| Disable Responsiveness | Press             |   |   |
|                        | Legal             |   | site design / logo © 2020 Stack Exchange Inc;<br>user contributions licensed under <b>cc by-sa</b> .<br>rev 2020.9.25.37676 |
|                        | Privacy Policy    |   |   |
|                        |                   |   |   |
|                        |                   |   |   |
|                        |                   |   |   |