Testing $H_0 : \rho = 0$: Comparing A Single Correlation to Zero
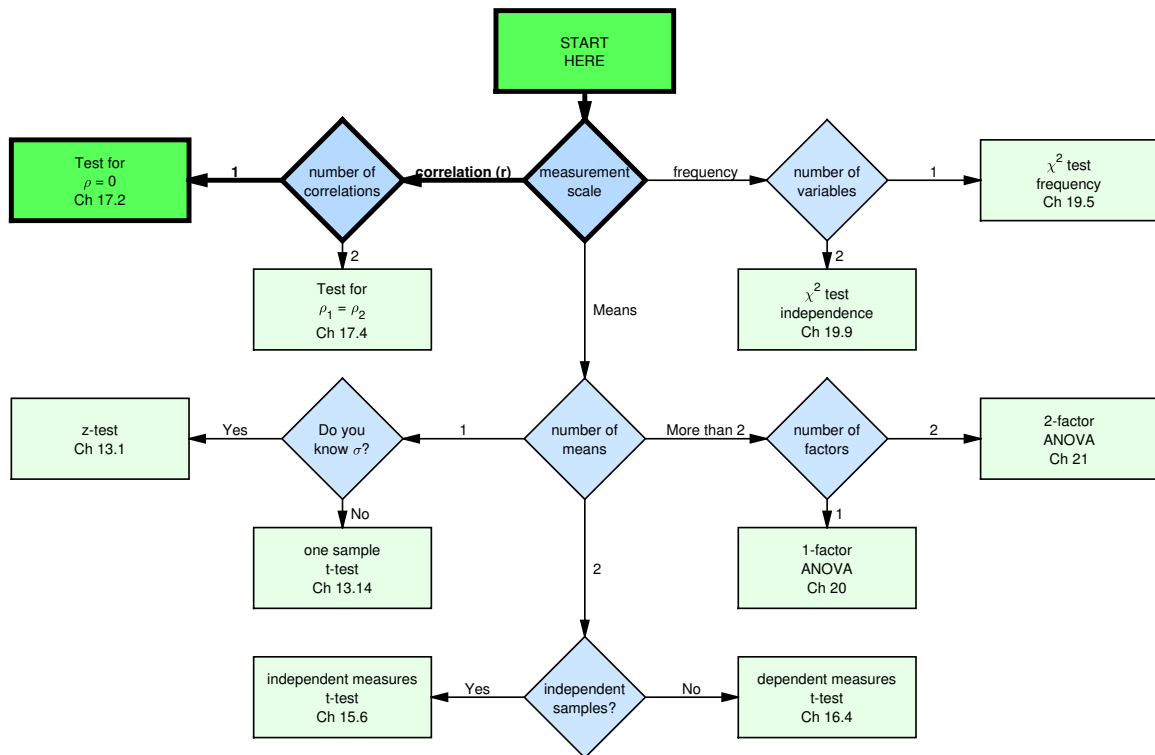
January 10, 2021

Contents

Testing H_0: $\rho = 0$

In this tutorial we'll be testing the null hypothesis that an observed correlation is drawn from a population of correlations that has a mean of zero ($\rho = 0$). This turns out to be a simple t-test. Here's how to get to the independent means t-test in the flow chart:

START HERE

Test for $\rho = 0$ — Ch 17.2

number of correlations

1 → Test for $\rho = 0$ Ch 17.2

2 → Test for $\rho_1 = \rho_2$ Ch 17.4

correlation (r)

measurement scale

frequency → number of variables

1 → χ^2 test frequency Ch 19.5

2 → χ^2 test independence Ch 19.9

Means → number of means

1 → Do you know σ?

Yes → z-test Ch 13.1

No → one sample t-test Ch 13.14

More than 2 → number of factors

2 → 2-factor ANOVA Ch 21

1 → 1-factor ANOVA Ch 20

2 → independent samples?

Yes → independent measures t-test Ch 15.6

No → dependent measures t-test Ch 16.4

Sir Ronald Fisher figured out that you can convert a test for correlations to a t-test using the 'Fisher's z transform':

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

The degrees of freedom is df = n-2 where n is the number of pairs in the correlation.

Here's some good news: Our stats tables and Excel spreadsheet have a table that provides the critical values of r so you don't even need to calculate t. Plus the Excel spreadsheet has a calculator that gives you a p-value for this t-test.

Example 1: The correlation between heights of your parents

Let's test the hypothesis that people tend to marry people of similar heights. Specifically, we'll use the survey from our class and test the hypothesis that the correlation between the heights of mothers and fathers is greater than zero. Let's choose an alpha value of $\alpha = 0.01$.

For the 143 students in our class, the correlation between the heights of mothers and fathers is 0.28.

Here's the scatterplot:

n = 143, r = 0.28

fathers heights (in)

mothers heights (in)
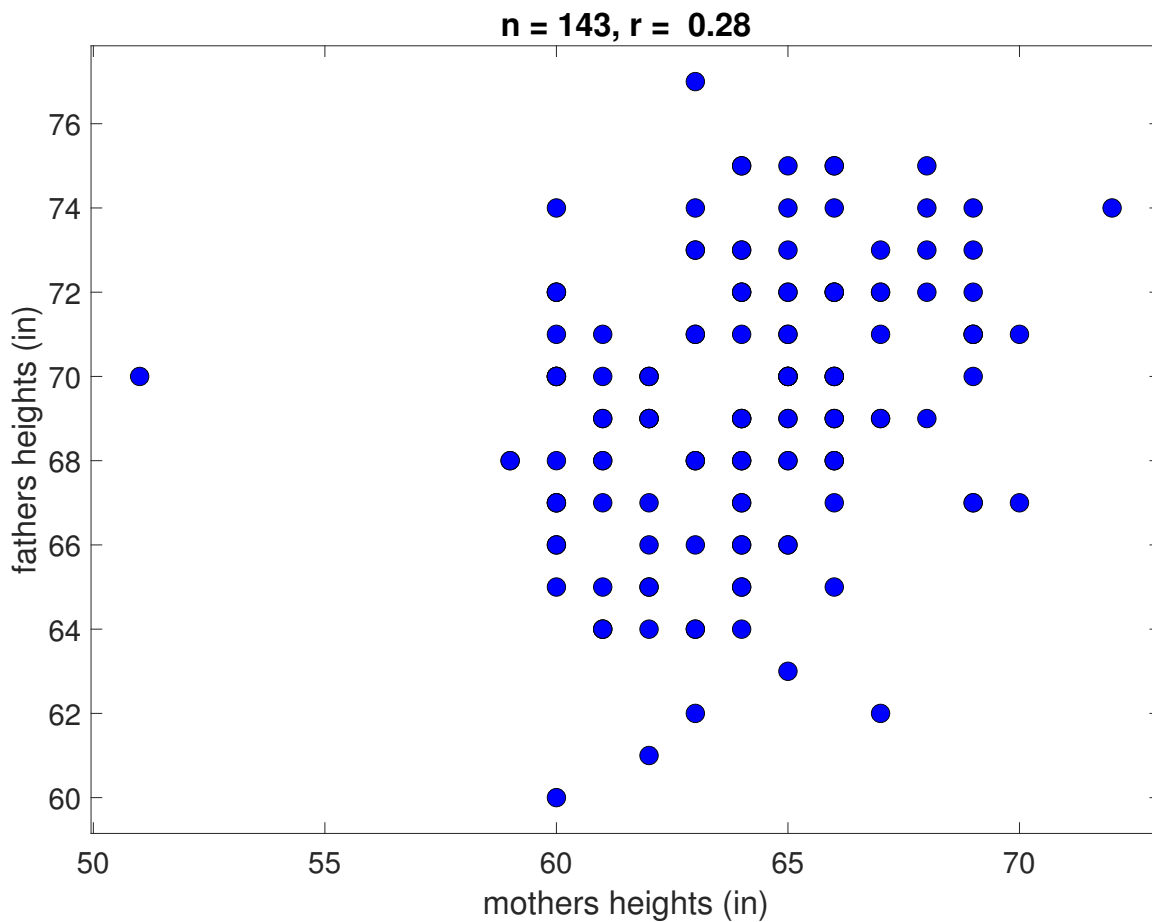
Since we're looking for a positive correlation, this will be a 1-tailed test.

All we need to do is look up the row in Table G for df = 143 - 2 = 141:

α: one tail	**0.05**	**0.025**	**0.01**	**0.005**
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
100	0.164	0.195	0.230	0.254
120	0.150	0.178	0.210	0.232
140	0.139	0.165	0.195	0.216
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

If we use the stats calculator we find that the critical value of r is 0.195.

Our observed value of r is 0.28 is greater than the critical value of 0.195. We therefore reject H_0 and conclude that the correlation between the heights of mothers and fathers is significantly greater than zero.

We can use the calculator on the Excel spreadsheet for table G to find the p-value:

r to α			
df	r	α (one tail)	α (two tail)
141	0.28	0.0004	0.0007
	t		
	3.4633		

α to r			
df	α	r_{crit} (one tail)	r_{crit} (two tail)
141	0.01	0.1944	0.2148
t_{crit}	t_{crit}		
2.3531	2.6111		

To state our conclusions using APA format, we'd state:

The correlation between mothers and fathers for heights is significantly greater than zero, r(141) = 0.28, p = 0.0004.

Example 2: Drinking and Facebook friends

For the 152 students in our class, the correlation between the number of of alcoholic drinks and the number of facebook friends is 0.04. Let's see if this correlation is significantly different than zero. We'll use $\alpha = 0.05$. This will be a 2-tailed test.

Looking up the row in Table G for df = 152 - 2 = 150, and the column for $\alpha = \frac{0.05}{2} = 0.025$ (since this is a two-tailed test):

α: one tail	0.05	0.025	0.01	0.005
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
100	0.164	0.195	0.230	0.254
120	0.150	0.178	0.210	0.232
140	0.139	0.165	0.195	0.216
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

If we use the calculator we find that the critical value of r is 0.159.

The calculator in the spreadsheet gives the p-value of 0.6246

r to α			
df	r	α (one tail)	α (two tail)
150	0.04	0.3123	0.6246
	t		
	0.4903		

α to r			
df	α	r_{crit} (one tail)	r_{crit} (two tail)
150	0.05	0.1339	0.1593
t_{crit}	t_{crit}		
1.6551	1.9759		

Our observed value of r =0.04 is not outside the range from -0.159 to 0.159. We therefore fail to reject H_0 and conclude that the correlation between the of alcoholic drinks and facebook friends is not significantly different than zero.

To present our conclusions using APA format, we'd state:

The correlation between alcoholic drinks and facebook friends for is not significantly different than zero, $r(150) = 0.04$, $p = 0.6246$.

Using R to run a hypothesis test for one correlation

The following R script covers how to run a test for $\rho = 0$ using the function 'cor.test':

The R commands shown below can be found here: ComparingOneCorrelation.R

```
# Comparing one correlation to the null hypothesis that rho = 0.
#
# If you have your raw x and y data, R's function 'cor.test' gives you a p-value for testing
# the hypothesis that a correlation was drawn from a population that has a correlation of zero.

# It takes in the x and y variables along with 'alternative' which can be "greater", "less"
# for a one-tailed test or "two.sided"for a two-tailed test.  Here are the examples from the
# comparing_one_correlation tutorial.

# By default, cor.test converts the correlation of x and y into a t-statistic and then computes
# a t-test.

# Load in the survey data
survey <-read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W21survey.csv")
# Example 1: The correlation between heights of your parents
x <- survey$mheight
y <- survey$pheight
goodvals = !is.na(x) & !is.na(y)
x <- x[goodvals]
y <- y[goodvals]
# cor.test runs the t-test for you:
out <- cor.test(x,y,alternative = "greater")
# 'estimate' is the correlation
out$estimate
```

```
      cor
0.283387
# 'p.value' is the p-value
out$p.value
[1] 0.0003021708

# 'statistic' is the t-statistic used in the test:
out$statistic
       t
3.508878

# with degrees of freedom:
out$parameter
 df
141

# Here's how to display your results in APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',out$parameter,out$estimate,out$p.value)
[1] "r(141) = 0.28, p = 0.0003"

# This is calculated directly from r and df.  So this should be the same
# thing (See the beginning of the compariong_one_correlation_tutorial.pdf
out$estimate/sqrt((1-out$estimate^2)/(length(x)-2))
      cor
3.508878

# Example 2: Two tailed test on the correlation between drinking and Facebook friends
x <- survey$drink
y <- survey$friends
goodvals = !is.na(x) & !is.na(y)
x <- x[goodvals]
y <- y[goodvals]
out <- cor.test(x,y,alternative = "two.sided")
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',out$parameter,out$estimate,out$p.value)
[1] "r(150) = 0.04, p = 0.6329"

# If you just have r and the sample size but not the raw data, you can calculate
# the p-value by running a t-test by hand:

# The correlation between x and y is:
r <- cor(x,y)
# The degrees of freedom is
df <- length(x) -2
# The t-statistic us computed using the 'Fisher's z transform':
t <- r/sqrt((1-r^2)/df)
# The p-value for a two-tailed for t not equal to 0 should give you the same p-value as for cor.t
2*pt(abs(t),df,0,lower.tail = FALSE)
```

```
[1] 0.6329406
```

Questions

It's random question time! Here are 10 practice questions followed by their answers. Some of them require you to calculate correlations from the actual scores. For these, you might want to copy the numbers and paste them into Excel.

1) You sample 25 economists from a population and measure both their quantity and their melancholy.
You calculate that their quantity correlates with melancholy with r = 0.6.
Using an alpha value of $\alpha = 0.05$, is this observed correlation significantly greater than zero?

2) For a 499 project you sample 21 brains from a population and measure both their size and their importance.
You calculate that their size correlates with importance with r = -0.52.
Using an alpha value of $\alpha = 0.01$, is this observed correlation significantly less than zero?

3) Your boss makes you sample 17 tired brains from a population and measure both their courage and their importance.
You calculate that their courage correlates with importance with r = -0.04.
Using an alpha value of $\alpha = 0.05$, is this observed correlation significantly less than zero?

4) Your boss makes you sample 8 psychedelic ping pong balls from a population and measure both their shopping and their amount.
You calculate that their shopping correlates with amount with r = 0.84.
Using an alpha value of $\alpha = 0.01$, is this observed correlation significantly greater than zero?

5) Because you don't have anything better to do you sample 19 gifted personalities from a population and measure both their conductivity and their warmth.
You calculate that their conductivity correlates with warmth with r = -0.24.
Using an alpha value of $\alpha = 0.01$, is this observed correlation significantly less than zero?

6) Your friend gets you to sample 23 psychedelic babies from a population and measure both their knowledge and their visual acuity.
You calculate that their knowledge correlates with visual acuity with r = 0.38.
Using an alpha value of $\alpha = 0.01$, is this observed correlation significantly greater than zero?

7) In your spare time you sample 11 acidic men from a population and measure both their grief and their test scores.
You calculate that their grief correlates with test scores with r = -0.24.
Using an alpha value of $\alpha = 0.05$, is this observed correlation significantly different than zero?

8) You sample 13 ashamed Americans from a population and measure both their mail and their experience.
You calculate that their mail correlates with experience with r = -0.47.

Using an alpha value of $\alpha = 0.05$, is this observed correlation significantly different than zero?

9) You get a grant to sample 6 democrats from a population and measure both their conductivity and their sadness.
You calculate that their conductivity correlates with sadness with r = 0.55.
Using an alpha value of $\alpha = 0.05$, is this observed correlation significantly greater than zero?

10) One day you sample 19 idiotic republicans from a population and measure both their recognition and their shopping.
You calculate that their recognition correlates with shopping with r = -0.43.
Using an alpha value of $\alpha = 0.01$, is this observed correlation significantly different than zero?

Answers

1) Correlation between quantity and melancholy of economists

$r_{obs} = 0.6$
$t = 3.5969$
$t_{crit} = 1.71 (df = 23)$

or $r_{crit} = 0.34$

We reject H_0.

The correlation between quantity and melancholy for economists is significantly greater than zero, r(23) = 0.6, p = 0.0008.

```
r <- 0.6
df <- 25-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] 3.596874
p <- pt(t,df,0,lower.tail=FALSE)
p
[1] 0.0007608285
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(23) = 0.60, p = 0.0008"
```

2) Correlation between size and importance of brains

$r_{obs} = -0.52$
$t = -2.6536$
$t_{crit} = -2.54 (df = 19)$

or $r_{crit} = -0.50$

We reject H_0.

The correlation between size and importance for brains is significantly less than zero, r(19) = -0.52, p = 0.0078.

```
r <- -0.52
df <- 21-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -2.653614
p <- pt(t,df,0,lower.tail=TRUE)
```

```
p
[1] 0.0078394
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(19) = -0.52, p = 0.0078"
```

3) Correlation between courage and importance of tired brains

$r_{obs} = -0.04$
$t = -0.155$
$t_{crit} = -1.75 (df = 15)$

or $r_{crit} = -0.41$

We fail to reject H_0.

The correlation between courage and importance for tired brains is not significantly less than zero, r(15) = -0.04, p = 0.4394.

```
r <- -0.04
df <- 17-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -0.1550434
p <- pt(t,df,0,lower.tail=TRUE)
p
[1] 0.4394271
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(15) = -0.04, p = 0.4394"
```

4) Correlation between shopping and amount of psychedelic ping pong balls

$r_{obs} = 0.84$
$t = 3.7922$
$t_{crit} = 3.14 (df = 6)$

or $r_{crit} = 0.79$

We reject H_0.

The correlation between shopping and amount for psychedelic ping pong balls is significantly greater than zero, r(6) = 0.84, p = 0.0045.

```
r <- 0.84
df <- 8-2
# Fisher's z transform
```

```
t <- r/sqrt((1-r^2)/df)
t
[1] 3.792154
p <- pt(t,df,0,lower.tail=FALSE)
p
[1] 0.004525261
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(6) = 0.84, p = 0.0045"
```

5) Correlation between conductivity and warmth of gifted personalities

$r_{obs} = -0.24$
$t = -1.0193$
$t_{crit} = -2.57 (df = 17)$

or $r_{crit} = -0.53$

We fail to reject H_0.

The correlation between conductivity and warmth for gifted personalities is not significantly less than zero, r(17) = -0.24, p = 0.1612.

```
r <- -0.24
df <- 19-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -1.019338
p <- pt(t,df,0,lower.tail=TRUE)
p
[1] 0.1611655
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(17) = -0.24, p = 0.1612"
```

6) Correlation between knowledge and visual acuity of psychedelic babies

$r_{obs} = 0.38$
$t = 1.8826$
$t_{crit} = 2.52 (df = 21)$

or $r_{crit} = 0.48$

We fail to reject H_0.

The correlation between knowledge and visual acuity for psychedelic babies is not significantly greater than zero, r(21) = 0.38, p = 0.0368.

```
r <- 0.38
df <- 23-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] 1.882599
p <- pt(t,df,0,lower.tail=FALSE)
p
[1] 0.03684203
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(21) = 0.38, p = 0.0368"
```

7) Correlation between grief and test scores of acidic men

$r_{obs} = -0.24$
$t = -0.7417$
$t_{crit} = \pm 2.26 (df = 9)$

or $r_{crit} = \pm 0.60$

We fail to reject H_0.

The correlation between grief and test scores for acidic men is not significantly different than zero, r(9) = -0.24, p = 0.4772.

```
r <- -0.24
df <- 11-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -0.7416771
p <- 2*pt(abs(t),df,0,lower.tail = FALSE)
p
[1] 0.4771901
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(9) = -0.24, p = 0.4772"
```

8) Correlation between mail and experience of ashamed Americans

$r_{obs} = -0.47$
$t = -1.766$
$t_{crit} = \pm 2.20 (df = 11)$

or $r_{crit} = \pm 0.55$

13

We fail to reject H_0.

The correlation between mail and experience for ashamed Americans is not significantly different than zero, r(11) = -0.47, p = 0.1051.

```
r <- -0.47
df <- 13-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -1.766028
p <- 2*pt(abs(t),df,0,lower.tail = FALSE)
p
[1] 0.1050928
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(11) = -0.47, p = 0.1051"
```

9) Correlation between conductivity and sadness of democrats

$r_{obs} = 0.55$
$t = 1.3171$
$t_{crit} = 2.13 (df = 4)$

or $r_{crit} = 0.73$

We fail to reject H_0.

The correlation between conductivity and sadness for democrats is not significantly greater than zero, r(4) = 0.55, p = 0.1291.

```
r <- 0.55
df <- 6-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] 1.317106
p <- pt(t,df,0,lower.tail=FALSE)
p
[1] 0.1290937
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(4) = 0.55, p = 0.1291"
```

10) Correlation between recognition and shopping of idiotic republicans

$r_{obs} = -0.43$
$t = -1.9638$

$t_{crit} = \pm 2.90 (df = 17)$

or $r_{crit} = \pm 0.57$

We fail to reject H_0.

The correlation between recognition and shopping for idiotic republicans is not significantly different than zero, r(17) = -0.43, p = 0.0661.

```
r <- -0.43
df <- 19-2
# Fisher's z transform
t <- r/sqrt((1-r^2)/df)
t
[1] -1.963756
p <- 2*pt(abs(t),df,0,lower.tail = FALSE)
p
[1] 0.06612768
# APA format:
sprintf('r(%g) = %4.2f, p = %5.4f',df,r,p)
[1] "r(17) = -0.43, p = 0.0661"
```