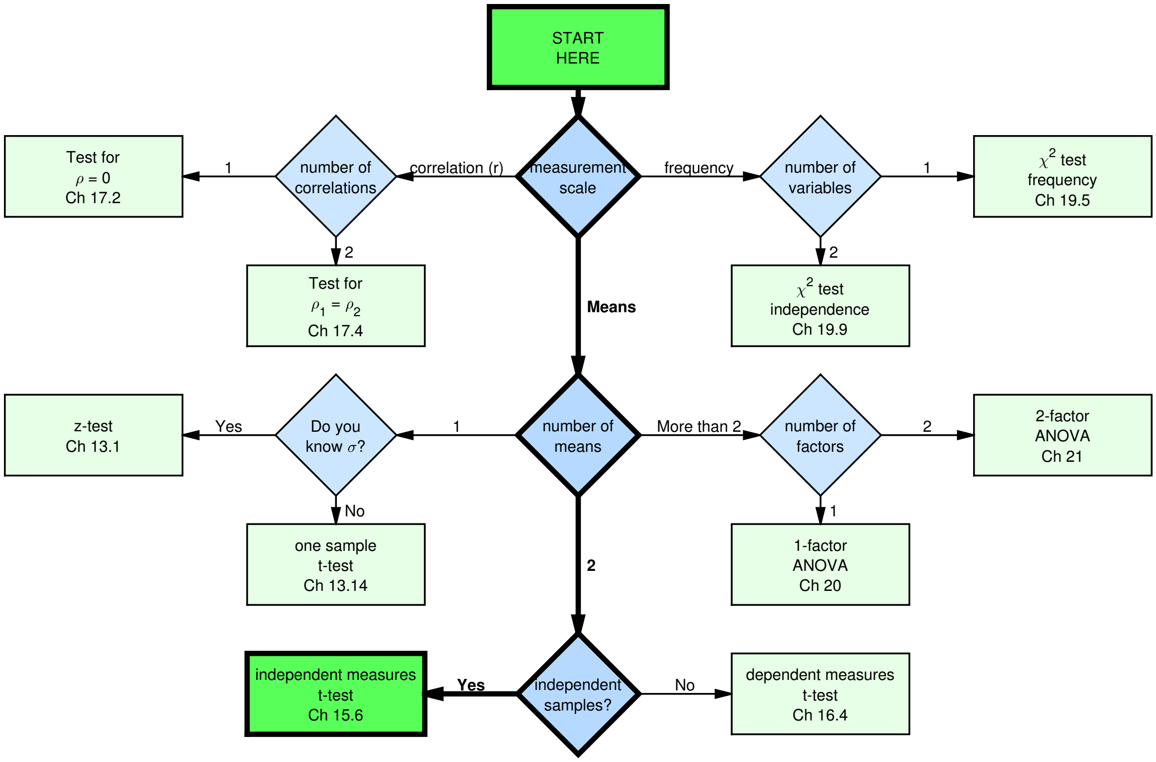
T-tests for 2 Independent Means

February 17, 2021

Contents

t-test for 2 independent means Tutorial

This t-test compares two means that are drawn independently. More specifically, the null hypothesis is that the two means are drawn from populations with the same mean (or means that differ by some fixed amount). Here's how to get to the independent means t-test in the flow chart:



The test estimates the probability of obtaining the observed means (or more extreme) if the null hypothesis is true. If this probability is small (less than alpha), then we reject the null hypothesis in support of the alternative hypothesis that the population means for the two samples are not the same (or differ by more than the fixed difference). Both the null and alternative hypotheses assume that the two population standard deviations are the same.

To conduct the test, we convert the two means and standard deviations into a test statistic which is drawn from a t-distribution under the null hypothesis:

$$t = \frac{\bar{x}-\bar{y}-(\mu_x-\mu_y)_{hyp}}{s_{\bar{x}-\bar{y}}}$$

The term in the numerator, $(\mu_x - \mu_y)_{hyp}$, is the expected difference between means under the null hypothesis. Usually this number is zero, which is when we are simply testing if the two means are significantly different from each other. So the simple case when we are comparing means to each other the calculation for t simplifies to:

$$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}}$$

The denominator $s_{\bar{x}-\bar{y}}$ is called the **pooled standard error of the mean**. It is calculated by first calculating the **pooled standard deviation**:

$$s_p = \sqrt{\frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{(n_x-1)+(n_y-1)}}$$

If you look at the formula for the pooled standard deviation, s_p, you should see that it's sort of a complicated average of the two standard deviations. Technically, the calculation inside the square root is really an average of the sample variances, weighted by their degrees of freedom. Importantly, s_p should always fall somewhere between the two sample standard deviations.

The pooled standard error is calculated from s_p by:

$$s_{\bar{x}-\bar{y}} = s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

This is sort of like how we calculated the standard error for the single sample t-test by diving the standard deviation by $\sqrt{n}$.

You can go from standard deviations and sample sizes to the pooled standard error of the mean in one step if you prefer:

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{(n_x-1)+(n_y-1)}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$$

Note: if the two sample sizes are the same ($s_x = s_y = n$), then the pooled standard error of the mean simplifies to:

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s_x^2+s_y^2}{n}}$$

The degrees of freedom of the independent means t-test is the sum of the degrees of freedom for each mean:

$$df = (n_x - 1) + (n_y - 1) = n_x + n_y - 2$$

Example 1: one-tailed test for independent means, equal sample sizes

Suppose you're a 315 Stats professor who is intersted in the impact on remote education on learning. You do this by comparing Exam 2 scores from course taught in 2020 before the pandemic to the Exam 2 scores in the course taught in 2021 during the pandemic.

In 2020, the 81 Exam 2 scores had a mean of 81.72 and a standard deviation of 28.2834. The 81 Exam 2 scores in 2021 had a mean of 71.68 and a standard deviation of 33.3654. Let's run a hypothesis test to determine of the mean Exam 2 scores from 2020 is significantly greater than from 2021. Use $\alpha = 0.05$.

First we calculate the pooled standard error of the mean. Since the sample sizes are the same ($n_x = n_y = 81$):

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s_x^2+s_y^2}{n}} = \sqrt{\frac{28.2834^2+33.3654^2}{81}} = 4.86$$

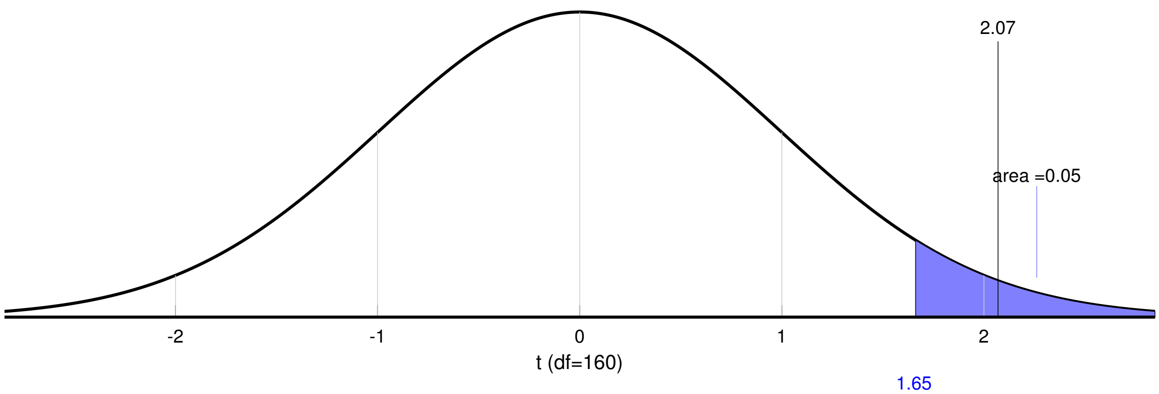
Our t-statistic is therefore:

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{81.72-71.68}{4.86} = 2.07$

This is a one-tailed t-test with df = 81 + 81 - 2 = 160 and α = 0.05. We can find our critical value of t from the t-table:

df, one tail	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
158	0.676	1.287	1.655	1.975	2.350	2.607	3.353
159	0.676	1.287	1.654	1.975	2.350	2.607	3.353
160	**0.676**	**1.287**	**1.654**	**1.975**	**2.350**	**2.607**	**3.352**
161	0.676	1.287	1.654	1.975	2.350	2.607	3.352
162	0.676	1.287	1.654	1.975	2.350	2.607	3.352

Our critical value of t is 1.654. Here's where our observed value of t (2.07) sits on the t-distribution compared to the critical value (1.654):



Our observed value of t is 2.07 which is greater than the critical value of 1.654. We therefore reject H_0 and conclude that the mean Exam 2 scores from 2020 is significantly greater than from 2021.

We can use the t-calculator to calculate that the p-value is 0.02:

Convert t to α			
t	df	α (one tail)	α (two tail)
2.07	160	0.02	0.0401

Convert α to t			
α	df	t (one tail)	t (two tail)
0.05	160	1.6544	1.9749

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

The Exam 2 scores of 2020 Psych 315 students (M = 81.72, SD = 28.2834) is significantly greater than the Exam 2 scores of 2021 Psych 315 students (M = 71.68, SD = 33.3654) t(160) = 2.07, p = 0.02.

Error Bars

For tests of independent means it's useful to plot our means as bars on a bar graph with error bars representing the standard errors of the mean. We calculate each standard error for each mean the usual way by dividing the standard deviation by the square root of each sample size:

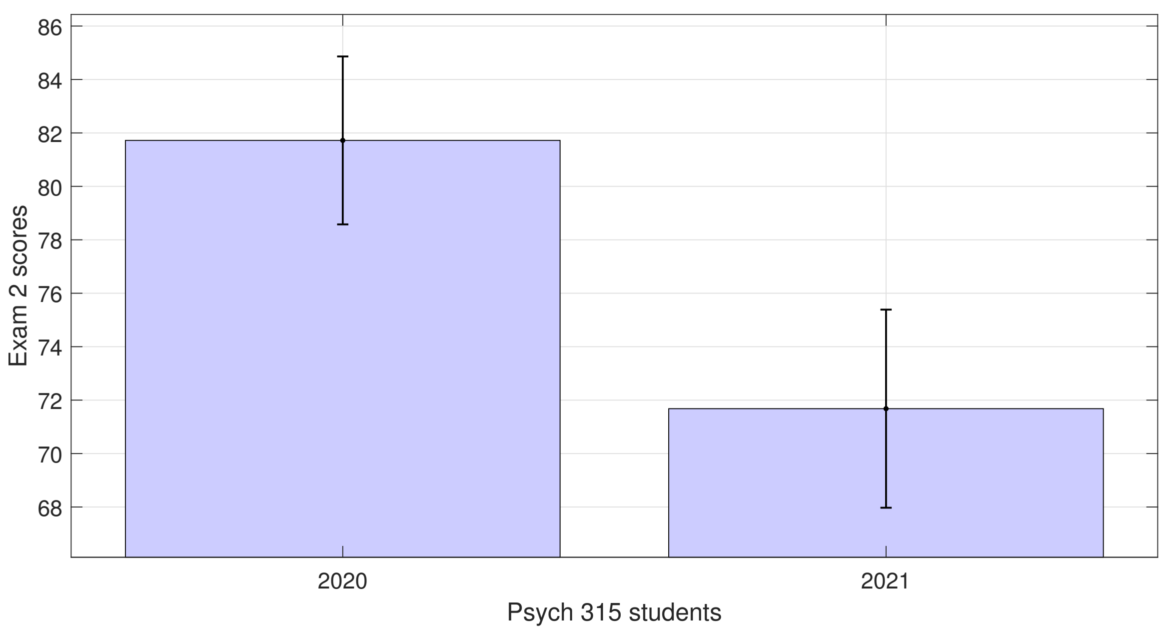
For 2020,

$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{28.2834}{\sqrt{81}} = 3.14$

and for 2021,

$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{33.3654}{\sqrt{81}} = 3.71$

The error bars are drawn by moving up and down one standard error of the mean ($s_{\bar{x}}$) for each mean ($\bar{x}$):



Bar graphs with error bars are useful for visualizing the significance of the difference between means.

Remember this rule of thumb: If the error bars overlap, then a one-tailed t-test will probably fail to reject H_0 with alpha = .05.

Keep in mind that you need a bigger gap between the error bars to reach significance for a two-tailed test, and/or for smaller values of alpha, like .01.

Example 2: two-tailed test for independent means, unequal sample sizes

Suppose you want to test the hypothesis that women with tall mothers are taller than women with less tall mothers. We'll use our class for our sample and divide the students into women with mothers that are taller and shorter than the median of of 64 inches (5 feet 4 inches). For our class, the heights of the 55 women with tall mothers has a mean of 65.9 inches and a standard deviation of 2.6 inches. The heights of the 63 women with less tall mothers has a mean of 63.6 inches and a standard deviation of 2.55 inches. Are these heights significantly different? Use alpha = 0.01.

Since our sample sizes are different, we have to use the more complicated formula for the pooled standard error of the mean:

$$s_p = \sqrt{\frac{(55-1)2.6^2+(63-1)2.55^2}{(55-1)+(63-1)}} = 2.57$$

$$s_{\bar{x}-\bar{y}} = 2.57\sqrt{\frac{1}{55}+\frac{1}{63}} = 0.47$$

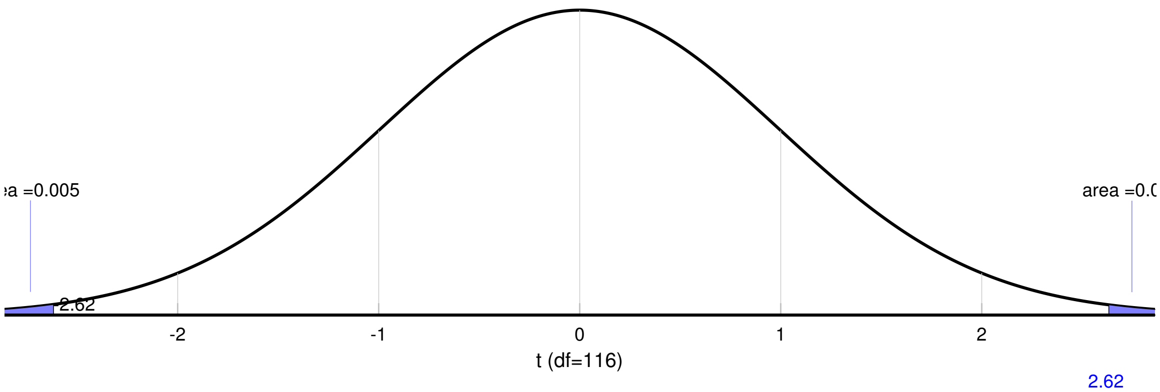
Our t-statistic is

$$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{65.9-63.6}{0.47} = 4.89$$

And use the t-table to see if this is statistically significant for df = 55 + 63 - 2 = 116:

df, one tail	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
114	0.677	1.289	1.658	1.981	2.360	2.620	3.378
115	0.677	1.289	1.658	1.981	2.359	2.619	3.377
116	**0.677**	**1.289**	**1.658**	**1.981**	**2.359**	**2.619**	**3.376**
117	0.677	1.289	1.658	1.980	2.359	2.619	3.376
118	0.677	1.289	1.658	1.980	2.358	2.618	3.375

For a two-tailed test, we find the critical value of t in the table by looking under the column for alpha/2 = 0.005. This is because the total area under both tails to add up to α, so the area in each of the two tails is $\frac{\alpha}{2}$. Our critical value is $\pm$ 2.619:



Our observed value of t is 4.89 which is greater than the critical value of 2.619. We therefore reject H_0 and conclude that the mean height of the women with tall mothers is statistically different from the height of the women with less tall mothers .

We can get a p-value from our calculator on the Excel spreadsheet. This gives us p $<$ 0.0001:

Convert t to α			
t	df	α (one tail)	α (two tail)
4.89	116	0	0

Convert α to t			
α	df	t (one tail)	t (two tail)
0.01	116	2.3589	2.6189

To conclude using APA format we'd state:

The height of women with tall mothers (M = 65.9, SD = 2.6) is significantly different than the height of women with less tall mothers (M = 63.6, SD = 2.55) t(116) = 4.89, p = $<$ 0.0001.

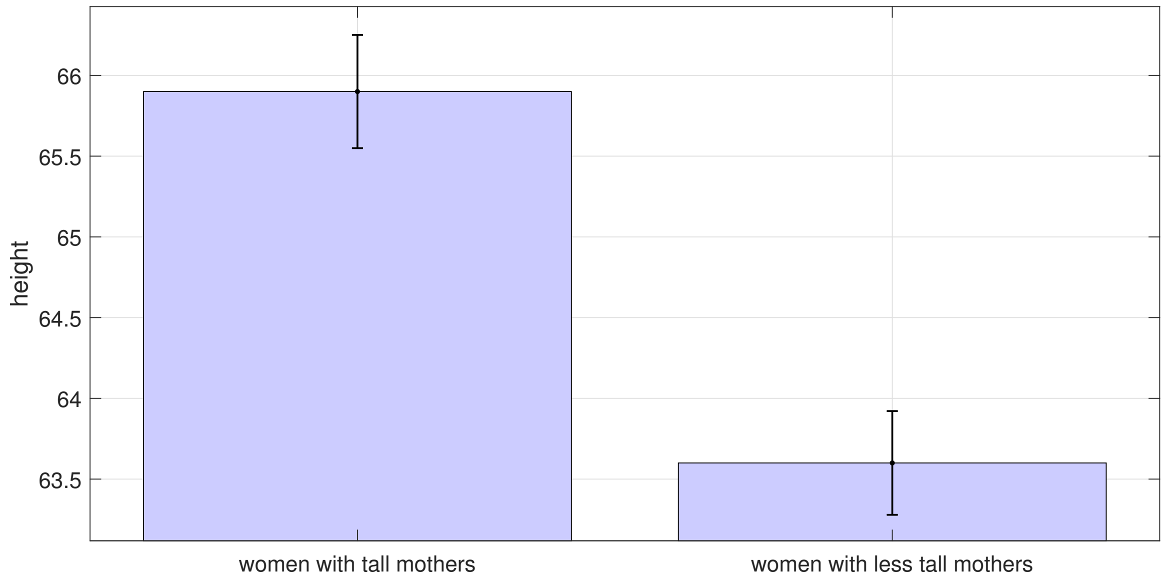
To show these means with error bars, for women with tall mothers,

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{2.6}{\sqrt{55}} = 0.35$$

and for women with less tall mothers,

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{2.55}{\sqrt{63}} = 0.32$$

The error bars are drawn by moving up and down one standard error of the mean ($s_{\bar{x}}$) for each mean ($\bar{x}$):



Effect size

The effect size for an independent measures t-test is Cohen's d again. This time it's measured as:

$$d = \frac{|\bar{x} - \bar{y} - (\mu_x - \mu_y)_{hyp}|}{s_p}$$

Or more commonly when $(\mu_x - \mu_y)_{hyp} = 0$:

$$d = \frac{|\bar{x} - \bar{y}|}{s_p}$$

where, again, the denominator is the pooled standard deviation:

$$s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{(n_x - 1) + (n_y - 1)}}$$

Recall that equal sample sizes the formula for s_p simplifies to:

$$s_p = \sqrt{s_x^2 + s_y^2}$$

For the first example,

$$s_p = \sqrt{28.2834^2 + 33.3654^2} = 30.929$$

so

$$d = \frac{|81.72 - 71.68|}{30.929} = 0.32$$

Which is considered to be a small effect size.

For the second example,

$$s_p = \sqrt{\frac{(55-1)2.6^2+(63-1)2.55^2}{(55-1)+(63-1)}} = 2.57$$

so

$$d = \frac{|65.9-63.6|}{2.57} = 0.89$$

Which is considered to be a large effect size.

Power

Power calculations for the independent mean t-test are conceptually the same as for the t-test for one mean. It's still the probability of correctly rejecting H_0.

I've provided a power calculator in the Excel spreadsheet. To use the power calculator **enter the average of the two sample sizes** for your value of n. For the first example, it's n=0. Here's what the results look like for our first example:

effect size (d)	n	α
0.32	81	0.05

One tailed test two means			
t_{crit}	$t_{crit} - t_{obs}$	area	power
161.6449	159.6084	0	0

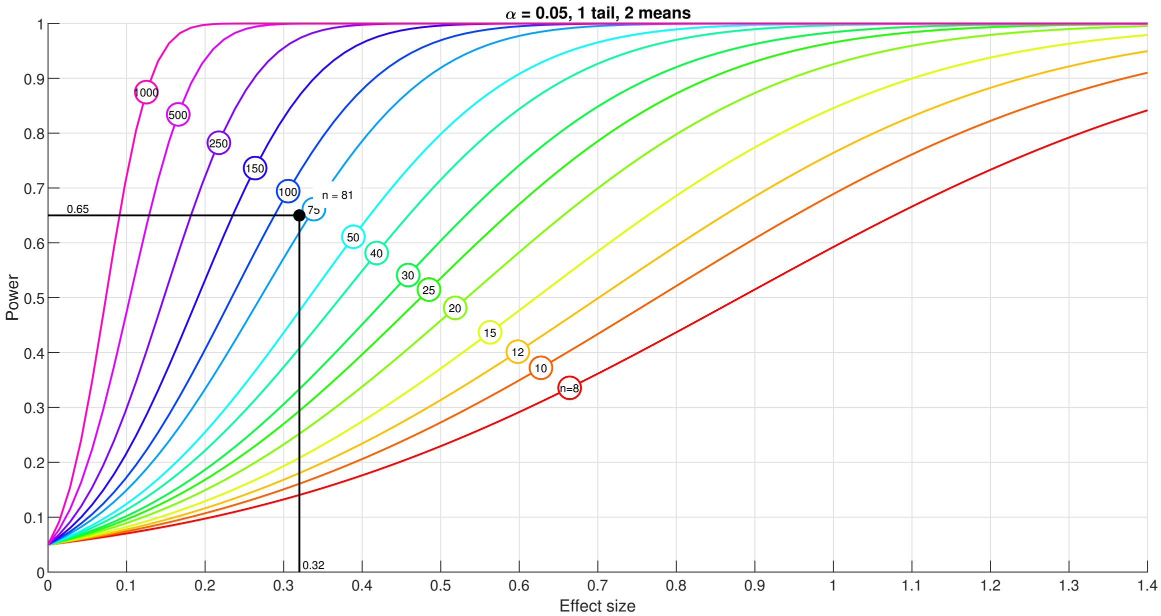
Two tailed test two means			
t_{crit}	$t_{crit} - t_{obs}$	area	power
1.9749	-0.0616	0.5245	0.5246
-1.9749	-4.0114	0	

This was a one-tailed test, the power is 0.65.

I've also provided you power curves for the test for independent means at:

http://courses.washington.edu/psy315/pdf/PowerCurves.pdf

Here is the family of power curves for a one-tailed test with $\alpha = 0.05$ and two means. To estimate power, use the **mean of the two sample sizes** again.



Can you estimate what sample size you'd need to have a power of 0.8? Answer: You'd need around 120 Psych 315 students.

For the second example, we had an effect size of 0.89 and an average sample size of 59:

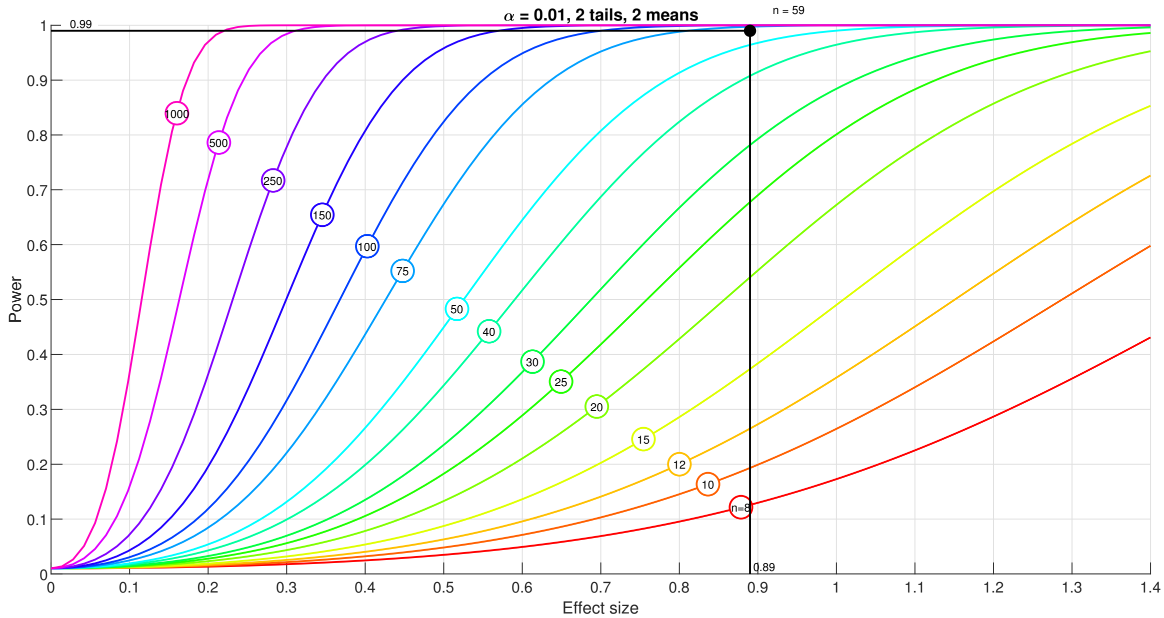
effect size (d)	n	α
0.89	59	0.01

One tailed test two means			
t_{crit}	$t_{crit} - t_{obs}$	area	power
118.3263	113.4924	0	0

Two tailed test two means			
t_{crit}	$t_{crit} - t_{obs}$	area	power
2.6189	-2.2151	0.9856	0.9856
-2.6189	-7.4528	0	

This was a two-tailed test, the power is 0.99.

Here's the family of power curves for a two-tailed test with $\alpha = 0.01$ and two means. To estimate power, use the **mean of the two sample sizes** again.



Can you estimate what sample size you'd need to have a power of 0.8? Answer: You'd need around 30 .

Using R to run a t-test for independent means

The following R script is a bit long, but it covers how to run a t-test for independent means, calculate power, and plot a bar graph with error bars

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

```
# TwoSampleIndependentTTest.R
#
# The following the example is the t-test for independent means, where we compared
# heights of female students who's mothers were taller or shorter than the median.

# Load in the survey data
survey <-read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W21survey.csv")
# First find the heights of the mothers of female students, removing NA's
mheight <- survey$mheight[!is.na(survey$mheight) & survey$gender=="Female" ]
# This is the median of the mother's heights:
median(mheight)
[1] 64

# Find the heights of the female students who's mother's aren't NA's:
height <- survey$height[!is.na(survey$mheight) & survey$gender=="Female"]
# Find the heights of female students who's mothers are taller than the median.  Call them 'x'
x <- height[mheight>median(mheight)]
```

```
# Find the heights of female students who's mother's heights are lessthan or equal to the median
y <- height[mheight<=median(mheight)]
# Run the two-tailed t-test. If you send in both x and y, t.test
# assumes it's a two-sample independent measures t-test.  The 'var.equal = TRUE'
# tells R to use the pooled standard deviation to combine the two measures
# of standard deviation.
out <- t.test(x,y,
       alternative = "two.sided",
       var.equal = TRUE)
# The p-pvalue is:
out$p.value
[1] 4.268402e-06

# Displaying the result in APA format:
sprintf('t(%g) = %4.2f, p = %5.1f',out$parameter,out$statistic,out$p.value)
[1] "t(116) = 4.83, p =   0.0"
mx <- mean(x)
my <- mean(y)
nx <- length(x)
ny <- length(y)
sx <- sd(x)
sy <- sd(y)
# pooled sd
sp <- sqrt( ((nx-1)*sx^2 + (ny-1)*sy^2)/(nx-1+ny-1))
sp
[1] 2.573593

#effect size
d <- abs(mx-my)/sp
d
[1] 0.8907205

# Find observed power from d, alpha and n
out <- power.t.test(n = (nx+ny)/2,
                    d = d,
                    sig.level = .05,
                    power = NULL,
                    alternative = "two.sided",
                    type = "two.sample")

out$power
[1] 0.9977265

# Find desired n from d, alpha and power =0.8
out <- power.t.test(n = NULL,
                    d = d,
                    sig.level = .05,
                    power = 0.8,
```

```
                    alternative = "two.sided",
                    type = "two.sample")
out$n
[1] 20.79135

# Making a bar graph with error bars.  Adding error bars is a bit more
# complicated than the basic bar graph.  First, it requires loading in
# a new 'library' called 'ggplot2' which can be done with the 'install.library'
# function:

# install.library(ggplot2)

# I've commented it out here because I've already done this.  Once you've
# installed a library once, you can tell R that you want to use it with
# the 'library' function:
library(ggplot2)
Warning message:
package 'ggplot2' was built under R version 4.0.3

# Now we'll generate a 'data frame' containing statistics for x and y:
summary <- data.frame(
  mean <- c(mean(x),mean(y)),
  n <- c(length(x),length(y)),
  sd <- c(sd(x),sd(y)))
summary$sem <- summary$sd/sqrt(summary$n)
colnames(summary) = c("mean","n","sd","sem")
row.names(summary) = c("Tall Mothers","Less Tall Mothers")
# This was a bit of work, but it creates a nice table:
summary
                      mean  n       sd       sem
Tall Mothers      65.92727 55 2.595256 0.3499442
Less Tall Mothers 63.63492 63 2.554576 0.3218463

# Once you have this summary table, the rest will give you a nice looking
# bar plot with error bars:

# Define y limits for the bar graph based on means and sem's
ylimit <-  c(min(summary$mean-1.5*summary$sem),
             max(summary$mean+1.5*summary$sem))
# Plot bar graph with error bar as one standard error (standard error of the mean/SEM)
ggplot(summary, aes(x = row.names(summary), y = mean)) +
  xlab("Students") +
  geom_bar(position = position_dodge(), stat="identity", fill="blue") +
  geom_errorbar(aes(ymin=mean-sem, ymax=mean+sem),width = .5) +
  theme_bw() +
  theme(panel.grid.major = element_blank()) +
  scale_y_continuous(name = "Height (in)") +
  coord_cartesian(ylim=ylimit)
```

Questions

You knew it was coming. Here are 5 random practice questions followed by their answers.

1) You go out and measure the smell of 34 drab and 84 breezy movies and obtain for drab movies a mean smell of 68.01 and a standard deviation of 4.4309, and for breezy movies a mean of 69.63 and a standard deviation of 4.3934.
Make a bar graph of the means with error bars representing the standard error of the means
Using an alpha value of 0.01, is the mean smell of drab movies significantly less than for the breezy movies?
What is the effect size?
What is the observed power of this test?

2) You get a grant to measure the baggage of 55 overrated and 18 curved underwear and obtain for overrated underwear a mean baggage of 13.4 and a standard deviation of 9.9228, and for curved underwear a mean of 20.8 and a standard deviation of 9.8979.
Make a bar graph of the means with error bars representing the standard error of the means
Using an alpha value of 0.01, is the mean baggage of overrated underwear significantly different than for the curved underwear?
What is the effect size?
What is the observed power of this test?

3) Because you don't have anything better to do you measure the speed of 105 boorish and 28 invincible iPhones and obtain for boorish iPhones a mean speed of 42.3 and a standard deviation of 7.2692, and for invincible iPhones a mean of 40.59 and a standard deviation of 6.3024.
Make a bar graph of the means with error bars representing the standard error of the means
Using an alpha value of 0.01, is the mean speed of boorish iPhones significantly different than for the invincible iPhones?
What is the effect size?
What is the observed power of this test?

4) Your stats professor asks you to measure the happiness of 98 proud and 31 infamous brothers and obtain for proud brothers a mean happiness of 33.24 and a standard deviation of 5.4555, and for infamous brothers a mean of 34.76 and a standard deviation of 6.0071.
Make a bar graph of the means with error bars representing the standard error of the means
Using an alpha value of 0.01, is the mean happiness of proud brothers significantly different than for the infamous brothers?
What is the effect size?
What is the observed power of this test?

5) Let's measure the homework of 102 spiky and 55 old ice dancers and obtain for

spiky ice dancers a mean homework of 40.98 and a standard deviation of 7.8785, and for old ice dancers a mean of 44.76 and a standard deviation of 8.546.

Make a bar graph of the means with error bars representing the standard error of the means

Using an alpha value of 0.01, is the mean homework of spiky ice dancers significantly less than for the old ice dancers?

What is the effect size?

What is the observed power of this test?

Answers

1) The smell of drab and breezy movies

$\bar{x} - \bar{y} = 68.01 - 69.63 = -1.62$

$s_p = \sqrt{\frac{(34-1)4.4309^2 + (84-1)4.3934^2}{(34-1)+(84-1)}} = 4.4041$

$s_{\bar{x}-\bar{y}} = 4.4041\sqrt{\frac{1}{34} + \frac{1}{84}} = 0.8952$

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{68.01-69.63}{0.8952} = -1.81$

$t_{crit} = -2.36(df = 116)$

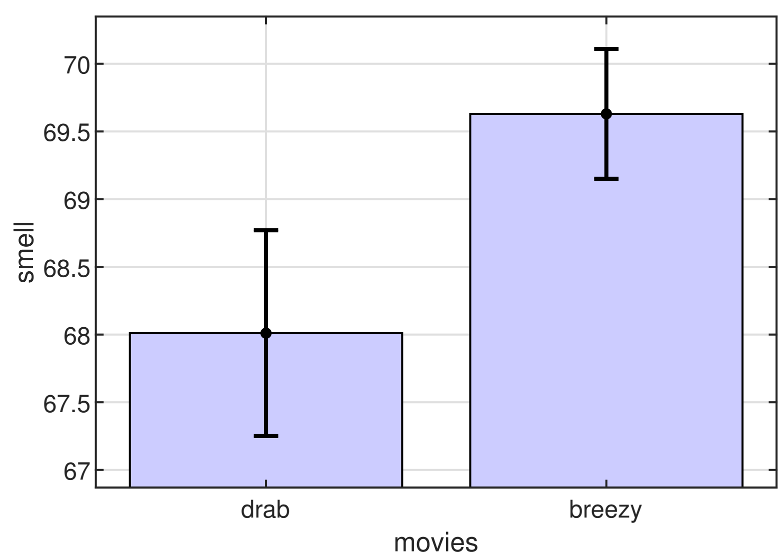
We fail to reject H_0.

The smell of drab movies (M = 68.01, SD = 4.4309) is not significantly less than the smell of breezy movies (M = 69.63, SD = 4.3934) t(116) = -1.81, p = 0.0364.

The effect size is $d = \frac{|\bar{x}-\bar{y}|}{s_p} = \frac{|68.01-69.63|}{4.4041} = 0.37$
This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.37, n = $\frac{(34+84)}{2}$ = 59 and $\alpha = 0.01$ is 0.3600.
$s_{\bar{x}} = = \frac{4.4309}{\sqrt{34}} = 0.7599$

$s_{\bar{y}} = \frac{4.3934}{\sqrt{84}} = 0.4794$



```
# Using R:
m1 <- 68.01
m2 <- 69.63
s1 <- 4.4309
s2 <- 4.3934
n1 <- 34
n2 <- 84
df <- 34 + 84 - 2
# Calculate pooled standard deviation
sp <- sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1-1+n2-1))
sp
[1] 4.404101
# pooled standard error of the mean
sxy <- sp*sqrt(1/n1+1/n2)
sxy
[1] 0.8951981
t = (m1-m2)/sxy
t
[1] -1.809655
p = pt(t,df)
p
[1] 0.03646928
# APA format:
sprintf('t(116) = %4.2f, p = %5.4f',t,p)
[1] "t(116) = -1.81, p = 0.0365"
# effect size
d <- abs(m1-m2)/sp
d
[1] 0.367839
# power:
out <- power.t.test(n = (n1+n2)/2,d= d,sig.level = 0.01,power = NULL,
type = "two.sample",alternative = "one.sided")
out$power
[1] 0.362513
```

2) The baggage of overrated and curved underwear

$\bar{x} - \bar{y} = 13.4 - 20.8 = -7.4$

$s_p = \sqrt{\frac{(55-1)9.9228^2+(18-1)9.8979^2}{(55-1)+(18-1)}} = 9.9168$

$s_{\bar{x}-\bar{y}} = 9.9168\sqrt{\frac{1}{55} + \frac{1}{18}} = 2.6929$

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{13.4-20.8}{2.6929} = -2.75$

$t_{crit} = \pm 2.65 (df = 71)$

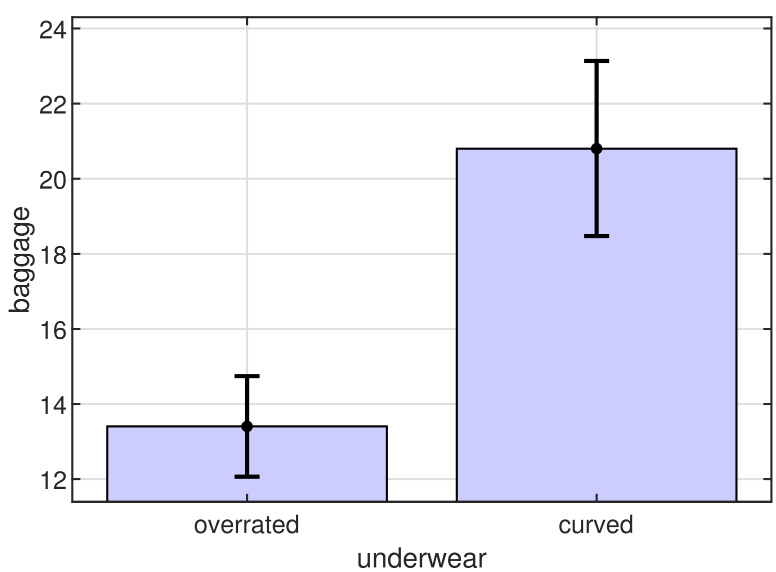
We reject H_0.

The baggage of overrated underwear (M = 13.4, SD = 9.9228) is significantly different than the baggage of curved underwear (M = 20.8, SD = 9.8979) t(71) = -2.75, p = 0.0076.

The effect size is $d = \frac{|\bar{x}-\bar{y}|}{s_p} = \frac{|13.4-20.8|}{9.9168} = 0.75$
This is a small effect size.

The observed power for two tailed test with an effect size of d = 0.75, n = $\frac{(55+18)}{2}$ = 37 and $\alpha = 0.01$ is 0.7200.
$s_{\bar{x}} = = \frac{9.9228}{\sqrt{55}} = 1.338$

$s_{\bar{y}} = \frac{9.8979}{\sqrt{18}} = 2.333$



```
# Using R:
m1 <- 13.4
m2 <- 20.8
s1 <- 9.9228
s2 <- 9.8979
n1 <- 55
n2 <- 18
df <- 55 + 18 - 2
# Calculate pooled standard deviation
sp <- sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1-1+n2-1))
sp
[1] 9.916844
# pooled standard error of the mean
sxy <- sp*sqrt(1/n1+1/n2)
sxy
[1] 2.692882
t = (m1-m2)/sxy
t
[1] -2.747985
p = 2*pt(abs(t),df,lower.tail = FALSE)
p
[1] 0.007596156
# APA format:
sprintf('t(71) = %4.2f, p = %5.4f',t,p)
[1] "t(71) = -2.75, p = 0.0076"
# effect size
d <- abs(m1-m2)/sp
d
[1] 0.7462052
# power:
out <- power.t.test(n = (n1+n2)/2,d= d,sig.level = 0.01,power = NULL,
type = "two.sample",alternative = "two.sided")
out$power
[1] 0.7044647
```

3) The speed of boorish and invincible iPhones

$\bar{x} - \bar{y} = 42.3 - 40.59 = 1.71$

$s_p = \sqrt{\frac{(105-1)7.2692^2+(28-1)6.3024^2}{(105-1)+(28-1)}} = 7.0807$

$s_{\bar{x}-\bar{y}} = 7.0807\sqrt{\frac{1}{105} + \frac{1}{28}} = 1.506$

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{42.3-40.59}{1.506} = 1.14$

$t_{crit} = \pm 2.61(df = 131)$

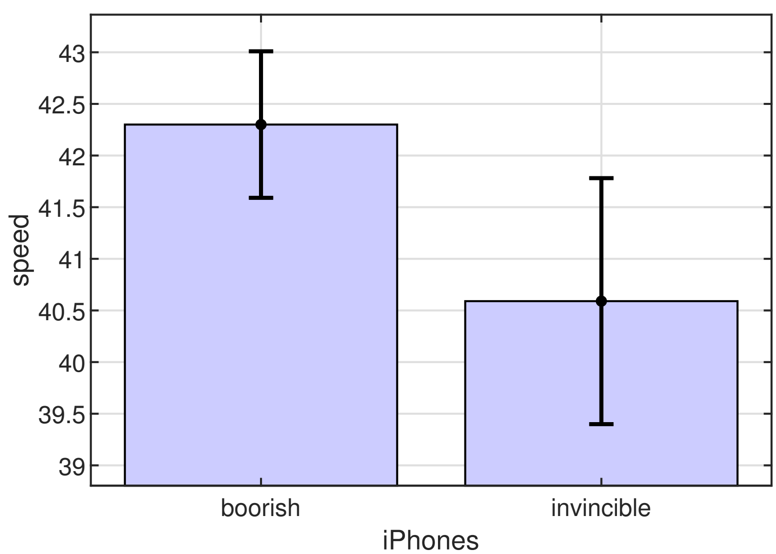
We fail to reject H_0.

The speed of boorish iPhones (M = 42.3, SD = 7.2692) is not significantly different than the speed of invincible iPhones (M = 40.59, SD = 6.3024) t(131) = 1.14, p = 0.2564.

The effect size is $d = \frac{|\bar{x}-\bar{y}|}{s_p} = \frac{|42.3-40.59|}{7.0807} = 0.24$
This is a small effect size.

The observed power for two tailed test with an effect size of d = 0.24, n = $\frac{(105+28)}{2}$ = 67 and $\alpha = 0.01$ is 0.1100.
$s_{\bar{x}} = = \frac{7.2692}{\sqrt{105}} = 0.7094$

$s_{\bar{y}} = \frac{6.3024}{\sqrt{28}} = 1.191$



```
# Using R:
m1 <- 42.3
m2 <- 40.59
s1 <- 7.2692
s2 <- 6.3024
n1 <- 105
n2 <- 28
df <- 105 + 28 - 2
# Calculate pooled standard deviation
sp <- sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1-1+n2-1))
sp
[1] 7.080744
# pooled standard error of the mean
sxy <- sp*sqrt(1/n1+1/n2)
sxy
[1] 1.506021
t = (m1-m2)/sxy
t
[1] 1.135442
p = 2*pt(abs(t),df,lower.tail = FALSE)
p
[1] 0.2582632
# APA format:
sprintf('t(131) = %4.2f, p = %5.4f',t,p)
[1] "t(131) = 1.14, p = 0.2583"
# effect size
d <- abs(m1-m2)/sp
d
[1] 0.2415
# power:
out <- power.t.test(n = (n1+n2)/2,d= d,sig.level = 0.01,power = NULL,
type = "two.sample",alternative = "two.sided")
out$power
[1] 0.1149083
```

4) The happiness of proud and infamous brothers

$\bar{x} - \bar{y} = 33.24 - 34.76 = -1.52$

$s_p = \sqrt{\frac{(98-1)5.4555^2 + (31-1)6.0071^2}{(98-1)+(31-1)}} = 5.5907$

$s_{\bar{x}-\bar{y}} = 5.5907\sqrt{\frac{1}{98} + \frac{1}{31}} = 1.152$

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{33.24-34.76}{1.152} = -1.32$

$t_{crit} = \pm 2.62 (df = 127)$

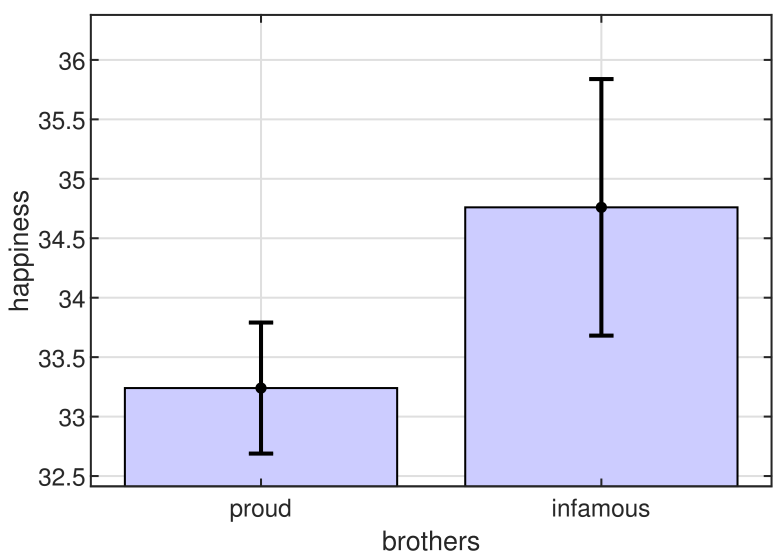
We fail to reject H_0.

The happiness of proud brothers (M = 33.24, SD = 5.4555) is not significantly different than the happiness of infamous brothers (M = 34.76, SD = 6.0071) t(127) = -1.32, p = 0.1892.

The effect size is $d = \frac{|\bar{x}-\bar{y}|}{s_p} = \frac{|33.24-34.76|}{5.5907} = 0.27$
This is a small effect size.

The observed power for two tailed test with an effect size of d = 0.27, n = $\frac{(98+31)}{2}$ = 65 and $\alpha = 0.01$ is 0.1400.
$s_{\bar{x}} = = \frac{5.4555}{\sqrt{98}} = 0.5511$

$s_{\bar{y}} = \frac{6.0071}{\sqrt{31}} = 1.0789$



```
# Using R:
m1 <- 33.24
m2 <- 34.76
s1 <- 5.4555
s2 <- 6.0071
n1 <- 98
n2 <- 31
df <- 98 + 31 - 2
# Calculate pooled standard deviation
sp <- sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1-1+n2-1))
sp
[1] 5.590711
# pooled standard error of the mean
sxy <- sp*sqrt(1/n1+1/n2)
sxy
[1] 1.152041
t = (m1-m2)/sxy
t
[1] -1.319397
p = 2*pt(abs(t),df,lower.tail = FALSE)
p
[1] 0.18941
# APA format:
sprintf('t(127) = %4.2f, p = %5.4f',t,p)
[1] "t(127) = -1.32, p = 0.1894"
# effect size
d <- abs(m1-m2)/sp
d
[1] 0.2718796
# power:
out <- power.t.test(n = (n1+n2)/2,d= d,sig.level = 0.01,power = NULL,
type = "two.sample",alternative = "two.sided")
out$power
[1] 0.1464137
```

5) The homework of spiky and old ice dancers

$\bar{x} - \bar{y} = 40.98 - 44.76 = -3.78$

$s_p = \sqrt{\frac{(102-1)7.8785^2 + (55-1)8.546^2}{(102-1)+(55-1)}} = 8.1173$

$s_{\bar{x}-\bar{y}} = 8.1173\sqrt{\frac{1}{102} + \frac{1}{55}} = 1.3579$

$t = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{40.98-44.76}{1.3579} = -2.78$

$t_{crit} = -2.35(df = 155)$

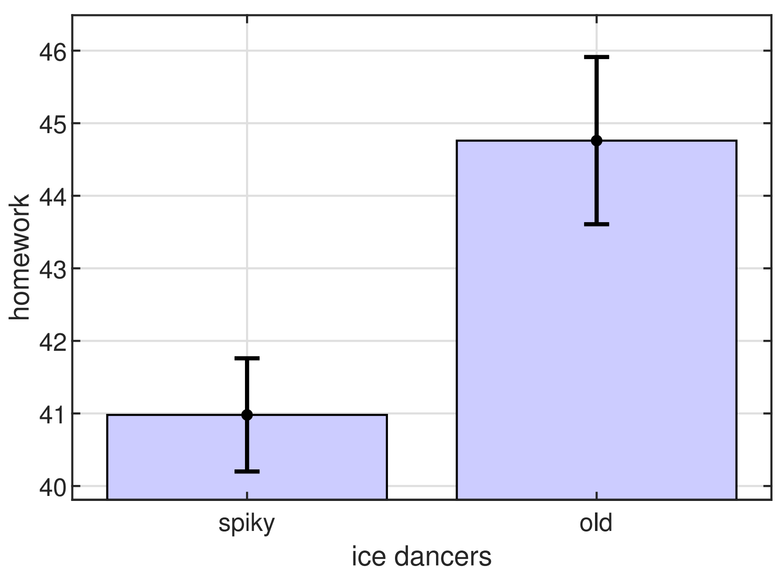
We reject H_0.

The homework of spiky ice dancers (M = 40.98, SD = 7.8785) is significantly less than the homework of old ice dancers (M = 44.76, SD = 8.546) t(155) = -2.78, p = 0.0031.

The effect size is $d = \frac{|\bar{x}-\bar{y}|}{s_p} = \frac{|40.98-44.76|}{8.1173} = 0.47$
This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.47, n = $\frac{(102+55)}{2} = 79$ and $\alpha = 0.01$ is 0.7300.
$s_{\bar{x}} = = \frac{7.8785}{\sqrt{102}} = 0.7801$

$s_{\bar{y}} = \frac{8.546}{\sqrt{55}} = 1.1523$



```r
# Using R:
m1 <- 40.98
m2 <- 44.76
s1 <- 7.8785
s2 <- 8.546
n1 <- 102
n2 <- 55
df <- 102 + 55 - 2
# Calculate pooled standard deviation
sp <- sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1-1+n2-1))
sp
[1] 8.117281
# pooled standard error of the mean
sxy <- sp*sqrt(1/n1+1/n2)
sxy
[1] 1.357935
t = (m1-m2)/sxy
t
[1] -2.783638
p = pt(t,df)
p
[1] 0.003022347
# APA format:
sprintf('t(155) = %4.2f, p = %5.4f',t,p)
[1] "t(155) = -2.78, p = 0.0030"
# effect size
d <- abs(m1-m2)/sp
d
[1] 0.4656732
# power:
out <- power.t.test(n = (n1+n2)/2,d= d,sig.level = 0.01,power = NULL,
type = "two.sample",alternative = "one.sided")
out$power
[1] 0.7141606
```