# T-tests for 2 Dependent Means

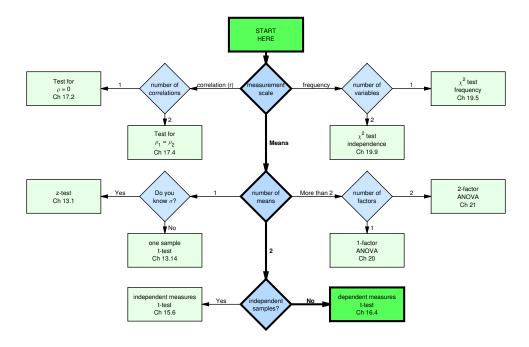
# January 10, 2021

### Contents

- t-test For Two Dependent Means Tutorial
- Example 1: Two-tailed t-test for dependent means
- Effect size (d)
- Power
- $\bullet$  Example 2
- Using R to run a t-test for independent means
- Questions
- Answers

# t-test For Two Dependent Means Tutorial

This test is used to compare two means for two samples for which we have reason to believe are dependent or correlated. The most common example is a repeated-measure design where each subject is sampled twice- that's why this test is sometimes called a 'repeated measures t-test'. Here's how to get to the dependent measures t-test on the flow chart:



Consider a weight-loss program where everyone lost exactly 20 pounds. Here's an example of weights before and after the program (in pounds) for 10 subjects:

Before	After
173	153
187	167
121	101
159	139
128	108
162	142
189	169
180	160
213	193
205	185

If you were to run an independent measures t-test on these two samples, you'd find that you'd fail to reject the hypothesis that the program changed the subject's weights with t(18) = 1.49, p = 0.1535.

But everyone lost 20 pounds! How could we not conclude that the weight loss program was effective? The problem is that there is a lot of variability in the weights across subjects. This variability ends up in the pooled standard deviation for the t-test.

But we don't care about the overall variability of the weights across subjects. We only care about the change due to the weight-loss program.

Experimental designs like this where we expect a correlation between measures are called 'dependent measures' designs. Most often they involve repeated measurements of the same subjects across conditions, so these designs are often called 'repeated measures' designs.

If you know how to run a t-test for one mean, then you know how to run a t-test for two dependent means. It's easy.

The trick is to create a third variable, D, which is the pair-wise differences between corresponding scores in the two groups. You then simply run a t-test on the mean of these differences - usually to test if the mean of the differences, D, is different from zero.

# Example 1: Two-tailed t-test for dependent means

Suppose you want to see if GPAs from High School are significantly different than College for male students. You use the 28 male students from our class as a sample. We'll use an alpha value of 0.05.

Here's the table of GPAs, along with the column of differences:

High School	College	difference (D)
3.6	2	1.6
3.83	3.35	0.48
3.89	3.84	0.05
4	3.91	0.09
2.18	2.89	-0.71
4.6	2.6	2
3.95	3.66	0.29
2.9	3.83	-0.93
3.4	3.23	0.17
3	3	0
4	3.65	0.35
3.5	3.51	-0.01
3.2	3.3	-0.1
3.2	3.65	-0.45
2.2	4	-1.8
3.7	3.07	0.63
3.8	3.31	0.49
3.95	3.8	0.15
3.92	3.95	-0.03
3.7	3.2	0.5
3.85	3.66	0.19
3.8	3.85	-0.05
3.65	3.3	0.35
3.88	3.53	0.35
3.87	3.68	0.19
3.2	3.9	-0.7
3.7	3.8	-0.1
3.7	3.43	0.27

An dependent measures t-test is done by simply running a t-test on that third column of differences. The mean of differences is  $\bar{D}=$  -0.12. The standard deviation of the differences is  $S_D=0.7013$ .

You can verify that this mean of differences is the same as the difference of the means: the mean of High School GPAs is 3.58 and the mean of the College GPAs is 3.46. The difference of these means is 3.58 - 3.46 = 0.12.

The standard error of the mean is:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{0.7013}{\sqrt{28}} = 0.13$$

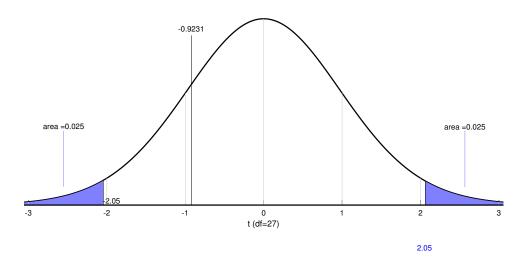
Just like for a t-test for a single mean, we calculate our t-statistic by subtracting the mean for the null hypothesis and divide by the estimated standard error of the mean. In this example, the mean for the null hypothesis,  $\mu_{hyp}$ , is zero.

$$t = \frac{\bar{D} - \mu_{hyp}}{s_{\bar{D}}} = \frac{-0.12}{0.13} = -0.9231$$

Finally, we use the t-table to see if this is a statistically significant t-statistic. We'll be using the row for df = 27 since we have 28 **pairs** of GPAs. This is a two-tailed test, so we need to divide alpha by 2:  $\frac{0.05}{2} = 0.025$ . Here's a sample section from the t-table.

df, one tail	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
:	:	:	:	:	:	:	:
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
:	:	:	:	:	:	:	:

The critical value of t is  $\pm 2.0518$ :



Our observed value of t is -0.9231 which is not in rejection region. We therefore fail to reject  $H_0$  and conclude that GPAs from High School are not significantly different than GPAs from College.

We can use the t-calculator to find that the p-value is 0.3641:

Convert t to $\alpha$					
t df $\alpha$ (one tail) $\alpha$ (two tail)					
0.9231 27 0.1821 0.3641					

Convert $\alpha$ to t					
$\alpha$ df t (one tail) t (two tail)					
0.05 27 1.7033 2.0518					

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

The GPA of High School infants (M = 3.58, SD = 0.5264) is not significantly different than the GPA of College infants (M = 3.46, SD = 0.4541), t(27) = -0.9231, p = 0.3641.

# Effect size (d)

The effect size for the dependent measures t-test is just like that for the t-test for a single mean, except that it's done on the differences, D. Cohen's d is:

$$d = \frac{|\bar{D} - \mu_{hyp}|}{s_D}$$

For this example on GPAs:

$$d = \frac{|\bar{D} - \mu_{hyp}|}{s_D} = \frac{|-0.12 - 0|}{0.7013} = 0.17$$

This is considered to be a small effect size.

### Power

Calculating power for the t-test with dependent means is just like calculating power for the single-sample t-test. For the power calculator, we just plug in our effect size, our sample size (size of each sample, or number of pairs), and alpha. For our example of an effect size of 0.17, sample size of 28 and  $\alpha = 0.05$ , we get:

The thing to remember is that although the data has two means, the hypothesis test is really a test of a single mean  $(H_0: \bar{D}=0)$ . So we use the power value from the single mean.

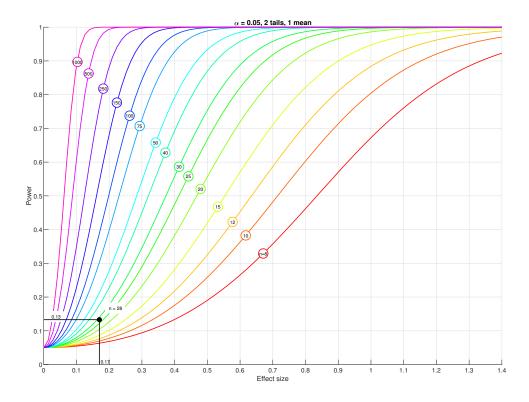
effect size (d)	n	α
0.17	28	0.05

One tailed test one mean					
$t_{crit}$ $t_{crit} - t_{obs}$ area power					
1.7033 0.8037 0.2143 0.2143					

Two tailed test one mean					
$t_{crit}$ $t_{crit} - t_{obs}$ area power					
-2.0518	-2.9514	0.0032	0.1329		
2.0518 1.1523 0.1297					

So our observed power is 0.1329.

Similarly, if there is a power outage (pun sort of intended) and you have to use the power curve, use the power curve for **one mean**:



# Example 2

Let's see if there is a significant difference between student's heights and their father's heights for male students in our class. We'll use an alpha value of 0.05.

Here's the table of heights, along with the column of differences:

fathers	students	difference (D)
72	72	0
71	71	0
69	66	3
70	75	-5
70	70	0
66	68	-2
75	71	4
72	72	0
67	68	-1
65	62	3
70	70	0
68	66	2
70	73	-3
72	74	-2
69	72	-3
64	66	-2
71	74	-3
71	72	-1
68	72	-4
71	70	1
70	70	0
64	66	-2
66	68	-2
67	67	0
72	74	-2
73	72	1

The mean of differences is  $\bar{D}=0.69.$  The standard deviation of the differences is  $S_D=2.2049.$ 

The standard error of the mean is:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{2.2049}{\sqrt{26}} = 0.43$$

Our t-statistic is:

$$t = \frac{\bar{D} - \mu_{hyp}}{s_{\bar{D}}} = \frac{0.69}{0.43} = 1.6047$$

Finally, we use the t-table to see if this is a statistically significant t-statistic. We'll be using the row for df = 25 since we have 26 pairs of heights. This is a two-tailed test, so we need

to divide alpha by 2:  $\frac{0.05}{2} = 0.025$ . Here's a sample section from the t-table.

df, one tail	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
:	:	:	:	:	:	:	:
23	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
:	:	:	:	:	:	:	:

The critical value of t is 2.0595:

Our observed value of t is 1.6047 which is not in rejection region.

We can use the Excel stats calculator to find the exact p-value:

Convert t to $\alpha$				
t df $\alpha$ (one tail) $\alpha$ (two tail)				
1.6047 25 0.0606 0.1211				

Convert $\alpha$ to t				
$\alpha$ df t (one tail) t (two tail)				
0.05 25 1.7081 2.0595				

We therefore fail to reject  $H_0$  and conclude that, using APA format: "The height of fathers of sororities (M = 69.35, SD = 2.8276) is not significantly different than the height of sororities (M=70.04, SD = 3.206), t(25) = 1.6047, p = 0.1211."

# Using R to run a t-test for independent means

The following R script shows how to run t-tests for the two dependent measures t-test examples in this tutorial.

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

```
# The following the example is the t-test for dependent means, where we compared
# GPA's from high school to GPA's from UW
# Load in the survey data
survey <-read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W21survey.csv")</pre>
# First find the UW GPA's for the male students
x <- survey$GPA_UW[survey$gender == "Male"]</pre>
# Then find the high school GPA's for the male students
y <- survey$GPA_HS[survey$gender == "Male"]
# Remove the pairs that have a NA in either x or y:
goodvals = !is.na(x) & !is.na(y)
x <- x[goodvals]</pre>
y <- y[goodvals]
# run the t-test. Use 'paired = TRUE' because x and y are dependent
out <- t.test(x,y,</pre>
       paired = TRUE,
       alternative = "two.sided",
       var.equal = TRUE)
# The p-pvalue is:
out$p.value
[1] 0.3859977
# Displaying the result in APA format:
sprintf('t(%g) = %4.2f, p = %5.4f',out$parameter,out$statistic,out$p.value)
[1] "t(27) = -0.88, p = 0.3860"
mx \leftarrow mean(x)
my <- mean(y)</pre>
s = sd(x-y)
n <- length(x)
#effect size
d <- abs(mx-my)/s
[1] 0.1665273
# Find observed power from d, alpha and n
out <- power.t.test(n =n,</pre>
                     d = d,
                     sig.level = .05,
                     power = NULL,
                     alternative = "two.sided",
```

```
type = "one.sample")
out$power
[1] 0.1335437
# Example 2: Is there a significant difference between male student's heights and their
# father's heights?
# First find the heights of the male students
x <- survey$height[survey$gender == "Male"]</pre>
# Then find the heights of their fathers
y <- survey$pheight[survey$gender == "Male"]</pre>
# Remove the pairs that have a NA in either x or y:
goodvals = !is.na(x) & !is.na(y)
x <- x[goodvals]</pre>
y <- y[goodvals]</pre>
# run the t-test. Use 'paired = TRUE' because x and y are dependent
out <- t.test(x,y,</pre>
              paired = TRUE,
              alternative = "two.sided",
              var.equal = TRUE)
# The p-pvalue is:
out$p.value
[1] 0.1219329
# Displaying the result in APA format:
sprintf('t(%g) = %4.2f, p = %5.4f',out$parameter,out$statistic,out$p.value)
[1] "t(25) = 1.60, p = 0.1219"
```

## Questions

Here are 10 practice t-test questions followed by their answers.

#### 1) The scenery of colossal and worthless colors

For a 499 project you measure the scenery of 96 colors under two conditions: 'colossal' and 'worthless'. You then subtract the scenery of the 'colossal' from the 'worthless' conditions for each colors and obtain a mean pair-wise difference of 1.27 with a standard deviation is 6.4377.

Using an alpha value of 0.05, is the scenery from the 'colossal' condition significantly less than from the 'worthless' condition?

What is the effect size?

What is the observed power of this test?

#### 2) The importance of pointless and nonstop nerds

You get a grant to measure the importance of 78 nerds under two conditions: 'pointless' and 'nonstop'. You then subtract the importance of the 'pointless' from the 'nonstop' conditions for each nerds and obtain a mean pair-wise difference of 0.82 with a standard deviation is 5.2376.

Using an alpha value of 0.05, is the importance from the 'pointless' condition significantly different than from the 'nonstop' condition?

What is the effect size?

What is the observed power of this test?

#### 3) The health of smelly and chivalrous candy bars

You ask a friend to measure the health of 82 candy bars under two conditions: 'smelly' and 'chivalrous'. You then subtract the health of the 'smelly' from the 'chivalrous' conditions for each candy bars and obtain a mean pair-wise difference of 3.31 with a standard deviation is 12 7073

Using an alpha value of 0.01, is the health from the 'smelly' condition significantly less than from the 'chivalrous' condition?

What is the effect size?

What is the observed power of this test?

### 4) The visual acuity of left and supreme bananas

Your boss makes you measure the visual acuity of 31 bananas under two conditions: 'left' and 'supreme'. You then subtract the visual acuity of the 'left' from the 'supreme' conditions for each bananas and obtain a mean pair-wise difference of 0.97 with a standard deviation is 7.2732.

Using an alpha value of 0.05, is the visual acuity from the 'left' condition significantly less than from the 'supreme' condition?

What is the effect size?

What is the observed power of this test?

#### 5) The friendship of depressed and nice skin color

We measure the friendship of 71 skin color under two conditions: 'depressed' and 'nice'. You then subtract the friendship of the 'depressed' from the 'nice' conditions for each skin color and obtain a mean pair-wise difference of 3.5 with a standard deviation is 11.0489. Using an alpha value of 0.01, is the friendship from the 'depressed' condition significantly less than from the 'nice' condition?

What is the effect size?

What is the observed power of this test?

#### 6) The gravity of lamentable and male elements

I'd like you to measure the gravity of 49 elements under two conditions: 'lamentable' and 'male'. You then subtract the gravity of the 'lamentable' from the 'male' conditions for each elements and obtain a mean pair-wise difference of 0.25 with a standard deviation is 4.0537.

Using an alpha value of 0.01, is the gravity from the 'lamentable' condition significantly less than from the 'male' condition?

What is the effect size?

What is the observed power of this test?

#### 7) The happiness of gentle and voiceless dinosaurs

On a dare, you measure the happiness of 103 dinosaurs under two conditions: 'gentle' and 'voiceless'. You then subtract the happiness of the 'gentle' from the 'voiceless' conditions for each dinosaurs and obtain a mean pair-wise difference of 1.46 with a standard deviation is 9.5993.

Using an alpha value of 0.05, is the happiness from the 'gentle' condition significantly less than from the 'voiceless' condition?

What is the effect size?

What is the observed power of this test?

#### 8) The happiness of understood and barbarous skittles

In your spare time you measure the happiness of 62 skittles under two conditions: 'understood' and 'barbarous'. You then subtract the happiness of the 'understood' from the 'barbarous' conditions for each skittles and obtain a mean pair-wise difference of -1.37 with a standard deviation is 11.4024.

Using an alpha value of 0.05, is the happiness from the 'understood' condition significantly greater than from the 'barbarous' condition?

What is the effect size?

What is the observed power of this test?

#### 9) The information of zany and bad chickens

I measure the information of 84 chickens under two conditions: 'zany' and 'bad'. You then subtract the information of the 'zany' from the 'bad' conditions for each chickens and obtain a mean pair-wise difference of -0.71 with a standard deviation is 8.322.

Using an alpha value of 0.01, is the information from the 'zany' condition significantly different than from the 'bad' condition?

What is the effect size?

What is the observed power of this test?

### 10) The violance of gusty and cold computers

You go out and measure the violance of 30 computers under two conditions: 'gusty' and 'cold'. You then subtract the violance of the 'gusty' from the 'cold' conditions for each computers and obtain a mean pair-wise difference of 0.8 with a standard deviation is 6.0687. Using an alpha value of 0.05, is the violance from the 'gusty' condition significantly less than from the 'cold' condition?

What is the effect size?

What is the observed power of this test?

### Answers

1) The scenery of colossal and worthless colors

$$\begin{split} \bar{D} &= 1.27, s_D = 6.4377, n = 96 \\ s_{\bar{D}} &= \frac{6.4377}{\sqrt{96}} = 0.66 \\ \mathrm{df} &= 96\text{-}1 = 95 \\ t &= \frac{1.27}{0.66} = 1.9242 \\ t_{crit} &= 1.66 \end{split}$$

We reject  $H_0$ .

The scenery of colossal colors (M = 64.31, SD = 4.4321) is significantly less than the scenery of worthless colors (M=65.58, SD = 4.3533), t(95) = 1.9242, p = 0.0287.

Effect size:  $d = \frac{|\bar{D}|}{{}^sD} = \frac{1.27}{6.4377} = 0.2$  This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.2, n = 96 and  $\alpha$  = 0.05 is 0.6170.

```
# Using R:
sem <- 6.4377/sqrt(96)
t <- (64.3113-65.584)/0.66
[1] -1.928333
p <- pt(t,95,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(95) = %4.2f, p = %5.4f',t,p)
[1] "t(95) = -1.93, p = 0.0284"
# Effect size:
d \leftarrow abs(1.27 - 0)/6.4377
[1] 0.1972754
# power:
out <- power.t.test(n = 96,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.608054
```

2) The importance of pointless and nonstop nerds

$$\begin{split} \bar{D} &= 0.82, s_D = 5.2376, n = 78 \\ s_{\bar{D}} &= \frac{5.2376}{\sqrt{78}} = 0.59 \\ \mathrm{df} &= 78\text{-}1 = 77 \\ t &= \frac{0.82}{0.59} = 1.3898 \\ t_{crit} &= \pm 1.99 \end{split}$$

We fail to reject  $H_0$ .

The importance of pointless nerds (M = 83.82, SD = 3.4055) is not significantly different than the importance of nonstop nerds (M=84.64, SD = 3.755), t(77) = 1.3898, p = 0.1686.

Effect size: d =  $\frac{|\bar{D}|}{^sD}$  =  $\frac{0.82}{5.2376}$  = 0.16 This is a small effect size.

The observed power for two tailed test with an effect size of d = 0.16, n = 78 and  $\alpha = 0.05$  is 0.2829.

```
# Using R:
sem < -5.2376/sqrt(78)
t <- (83.8197-84.636)/0.59
t
[1] -1.383559
# Since this is a two-tailed test, use abs(t) and lower.tail = FALSE
p <- 2*pt(abs(t),77,lower.tail = FALSE)</pre>
# APA format:
sprintf('t(77) = \%4.2f, p = \%5.4f',t,p)
[1] "t(77) = -1.38, p = 0.1705"
# Effect size:
d \leftarrow abs(0.82 - 0)/5.2376
[1] 0.1565603
# power:
out <- power.t.test(n = 78,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "two.sided")
out$power
[1] 0.2760948
```

3) The health of smelly and chivalrous candy bars

$$\begin{split} \bar{D} &= 3.31, s_D = 12.7973, n = 82 \\ s_{\bar{D}} &= \frac{12.7973}{\sqrt{82}} = 1.41 \\ \mathrm{df} &= 82\text{-}1 = 81 \\ t &= \frac{3.31}{1.41} = 2.3475 \\ t_{crit} &= 2.37 \text{ (using df} = 80) \end{split}$$

We fail to reject  $H_0$ .

The health of smelly candy bars (M = 18.91, SD = 8.777) is not significantly less than the health of chivalrous candy bars (M=22.22, SD = 9.213), t(81) = 2.3475, p = 0.0107.

Effect size:  $d = \frac{|\overline{D}|}{s_D} = \frac{3.31}{12.7973} = 0.26$  This is a small effect size.

The observed power for one tailed test with an effect size of d=0.26, n=82 and  $\alpha=0.01$  is 0.4925.

```
# Using R:
sem <- 12.7973/sqrt(82)
t <- (18.9089-22.2219)/1.41
[1] -2.349645
p <- pt(t,81,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(81) = %4.2f, p = %5.4f',t,p)
[1] "t(81) = -2.35, p = 0.0106"
# Effect size:
d \leftarrow abs(3.31 - 0)/12.7973
[1] 0.2586483
# power:
out <- power.t.test(n = 82,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.4906999
```

4) The visual acuity of left and supreme bananas

$$\begin{split} \bar{D} &= 0.97, s_D = 7.2732, n = 31 \\ s_{\bar{D}} &= \frac{7.2732}{\sqrt{31}} = 1.31 \\ \mathrm{df} &= 31\text{-}1 = 30 \\ t &= \frac{0.97}{1.31} = 0.7405 \\ t_{crit} &= 1.70 \end{split}$$

We fail to reject  $H_0$ .

The visual acuity of left bananas (M = 94.4, SD = 5.1754) is not significantly less than the visual acuity of supreme bananas (M=95.38, SD = 6.2365), t(30) = 0.7405, p = 0.2324.

Effect size:  $d = \frac{|\bar{D}|}{s_D} = \frac{0.97}{7.2732} = 0.13$  This is a small effect size.

The observed power for one tailed test with an effect size of d=0.13, n=31 and  $\alpha=0.05$  is 0.1691.

```
# Using R:
sem < -7.2732/sqrt(31)
t <- (94.4037-95.3785)/1.31
[1] -0.7441221
p <- pt(t,30,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(30) = %4.2f, p = %5.4f',t,p)
[1] "t(30) = -0.74, p = 0.2313"
# Effect size:
d \leftarrow abs(0.97 - 0)/7.2732
[1] 0.1333663
# power:
out <- power.t.test(n = 31,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.1790549
```

5) The friendship of depressed and nice skin color

$$\begin{split} \bar{D} &= 3.5, s_D = 11.0489, n = 71\\ s_{\bar{D}} &= \frac{11.0489}{\sqrt{71}} = 1.31\\ \mathrm{df} &= 71\text{-}1 = 70\\ t &= \frac{3.5}{1.31} = 2.6718\\ t_{crit} &= 2.38 \end{split}$$

We reject  $H_0$ .

The friendship of depressed skin color (M = 11.43, SD = 8.3102) is significantly less than the friendship of nice skin color (M=14.93, SD = 7.2131), t(70) = 2.6718, p = 0.0047.

Effect size:  $d = \frac{|\bar{D}|}{{}^sD} = \frac{3.5}{11.0489} = 0.32$  This is a small effect size.

The observed power for one tailed test with an effect size of d=0.32, n=71 and  $\alpha=0.01$  is 0.6234.

```
# Using R:
sem <- 11.0489/sqrt(71)
t <- (11.4324-14.932)/1.31
[1] -2.67145
p <- pt(t,70,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(70) = %4.2f, p = %5.4f',t,p)
[1] "t(70) = -2.67, p = 0.0047"
# Effect size:
d \leftarrow abs(3.5 - 0)/11.0489
[1] 0.3167736
# power:
out <- power.t.test(n = 71,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.6145311
```

6) The gravity of lamentable and male elements

$$\begin{split} \bar{D} &= 0.25, s_D = 4.0537, n = 49 \\ s_{\bar{D}} &= \frac{4.0537}{\sqrt{49}} = 0.58 \\ \mathrm{df} &= 49\text{-}1 = 48 \\ t &= \frac{0.25}{0.58} = 0.431 \\ t_{crit} &= 2.41 \end{split}$$

We fail to reject  $H_0$ .

The gravity of lamentable elements (M = 73.73, SD = 2.8222) is not significantly less than the gravity of male elements (M = 73.98, SD = 2.8455), t(48) = 0.431, p = 0.3342.

Effect size:  $d = \frac{|\bar{D}|}{s_D} = \frac{0.25}{4.0537} = 0.06$  This is a small effect size.

The observed power for one tailed test with an effect size of d=0.06, n=49 and  $\alpha=0.01$  is 0.0263.

```
# Using R:
sem < -4.0537/sqrt(49)
t <- (73.7292-73.9759)/0.58
[1] -0.4253448
p <- pt(t,48,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(48) = \%4.2f, p = \%5.4f',t,p)
[1] "t(48) = -0.43, p = 0.3362"
# Effect size:
d \leftarrow abs(0.25 - 0)/4.0537
[1] 0.06167205
# power:
out <- power.t.test(n = 49,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.0282843
```

7) The happiness of gentle and voiceless dinosaurs

$$\begin{split} \bar{D} &= 1.46, s_D = 9.5993, n = 103 \\ s_{\bar{D}} &= \frac{9.5993}{\sqrt{103}} = 0.95 \\ \mathrm{df} &= 103\text{-}1 = 102 \\ t &= \frac{1.46}{0.95} = 1.5368 \\ t_{crit} &= 1.66 \text{ (using df} = 100) \end{split}$$

We fail to reject  $H_0$ .

The happiness of gentle dinosaurs (M = 82.87, SD = 6.8821) is not significantly less than the happiness of voiceless dinosaurs (M = 84.33, SD = 6.9609), t(102) = 1.5368, p = 0.0637.

Effect size: d =  $\frac{|\bar{D}|}{^sD}$  =  $\frac{1.46}{9.5993}$  = 0.15 This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.15, n = 103 and  $\alpha = 0.05$  is 0.4454.

```
# Using R:
sem < -9.5993/sqrt(103)
t <- (82.8687-84.333)/0.95
t
[1] -1.541368
p <- pt(t,102,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(102) = %4.2f, p = %5.4f',t,p)
[1] "t(102) = -1.54, p = 0.0632"
# Effect size:
d \leftarrow abs(1.46 - 0)/9.5993
d
[1] 0.1520944
# power:
out <- power.t.test(n = 103,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.4556059
```

8) The happiness of understood and barbarous skittles

$$\begin{split} \bar{D} &= -1.37, s_D = 11.4024, n = 62 \\ s_{\bar{D}} &= \frac{11.4024}{\sqrt{62}} = 1.45 \\ \mathrm{df} &= 62\text{-}1 = 61 \\ t &= \frac{-1.37}{1.45} = -0.9448 \\ t_{crit} &= -1.67 \end{split}$$

We fail to reject  $H_0$ .

The happiness of understood skittles (M = 30.35, SD = 8.5396) is not significantly greater than the happiness of barbarous skittles (M=28.98, SD = 9.2862), t(61) = -0.9448, p = 0.1742.

Effect size: d =  $\frac{|\bar{D}|}{s_D}$  =  $\frac{-1.37}{11.4024}$  = 0.12 This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.12, n = 62 and  $\alpha = 0.05$  is 0.2355.

```
# Using R:
sem <- 11.4024/sqrt(62)
t <- (30.3529-28.9822)/1.45
t
[1] 0.9453103
p <- pt(t,61,lower.tail = FALSE)</pre>
# APA format:
sprintf('t(61) = %4.2f, p = %5.4f',t,p)
[1] "t(61) = 0.95, p = 0.1741"
# Effect size:
d \leftarrow abs(-1.37 - 0)/11.4024
d
[1] 0.1201501
# power:
out <- power.t.test(n = 62,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.2390785
```

9) The information of zany and bad chickens

$$\begin{split} \bar{D} &= -0.71, s_D = 8.322, n = 84 \\ s_{\bar{D}} &= \frac{8.322}{\sqrt{84}} = 0.91 \\ \mathrm{df} &= 84\text{-}1 = 83 \\ t &= \frac{-0.71}{0.91} = -0.7802 \\ t_{crit} &= \pm 2.64 \text{ (using df} = 80) \end{split}$$

We fail to reject  $H_0$ .

The information of zany chickens (M = 81.58, SD = 5.982) is not significantly different than the information of bad chickens (M=80.87, SD = 5.2741), t(83) = -0.7802, p = 0.4375.

Effect size:  $d = \frac{|\bar{D}|}{s_D} = \frac{-0.71}{8.322} = 0.09$  This is a small effect size.

The observed power for two tailed test with an effect size of d=0.09, n=84 and  $\alpha=0.01$  is 0.0373.

```
# Using R:
sem <- 8.322/sqrt(84)
t <- (81.5754-80.8704)/0.91
[1] 0.7747253
# Since this is a two-tailed test, use abs(t) and lower.tail = FALSE
p <- 2*pt(abs(t),83,lower.tail = FALSE)</pre>
# APA format:
sprintf('t(83) = %4.2f, p = %5.4f',t,p)
[1] "t(83) = 0.77, p = 0.4407"
# Effect size:
d \leftarrow abs(-0.71 - 0)/8.322
d
[1] 0.08531603
# power:
out <- power.t.test(n = 84,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "two.sided")
out$power
[1] 0.03519677
```

10) The violance of gusty and cold computers

$$\begin{split} \bar{D} &= 0.8, s_D = 6.0687, n = 30 \\ s_{\bar{D}} &= \frac{6.0687}{\sqrt{30}} = 1.11 \\ \mathrm{df} &= 30\text{-}1 = 29 \\ t &= \frac{0.8}{1.11} = 0.7207 \\ t_{crit} &= 1.70 \end{split}$$

We fail to reject  $H_0$ .

The violance of gusty computers (M = 16.28, SD = 3.7731) is not significantly less than the violance of cold computers (M=17.08, SD = 4.6179), t(29) = 0.7207, t(29) = 0.7

Effect size:  $d = \frac{|\bar{D}|}{{}^sD} = \frac{0.8}{6.0687} = 0.13$  This is a small effect size.

The observed power for one tailed test with an effect size of d=0.13, n=30 and  $\alpha=0.05$  is 0.1659.

```
# Using R:
sem < -6.0687/sqrt(30)
t <- (16.279-17.0769)/1.11
[1] -0.7188288
p <- pt(t,29,lower.tail = TRUE)</pre>
# APA format:
sprintf('t(29) = %4.2f, p = %5.4f',t,p)
[1] "t(29) = -0.72, p = 0.2390"
# Effect size:
d \leftarrow abs(0.8 - 0)/6.0687
[1] 0.1318239
# power:
out <- power.t.test(n = 30,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.1737145
```