T-tests for 2 Dependent Means

February 24, 2020

Contents
- t-test For Two Dependent Means Tutorial
- Example 1: Two-tailed t-test for dependent means
- Effect size (d)
- Power
- 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:

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>173</td>
<td>153</td>
</tr>
<tr>
<td>187</td>
<td>167</td>
</tr>
<tr>
<td>121</td>
<td>101</td>
</tr>
<tr>
<td>159</td>
<td>139</td>
</tr>
<tr>
<td>128</td>
<td>108</td>
</tr>
<tr>
<td>162</td>
<td>142</td>
</tr>
<tr>
<td>189</td>
<td>169</td>
</tr>
<tr>
<td>180</td>
<td>160</td>
</tr>
<tr>
<td>213</td>
<td>193</td>
</tr>
<tr>
<td>205</td>
<td>185</td>
</tr>
</tbody>
</table>

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 27 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:

<table>
<thead>
<tr>
<th>High School</th>
<th>College</th>
<th>difference (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>3.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3.47</td>
<td>2.86</td>
<td>0.61</td>
</tr>
<tr>
<td>3.5</td>
<td>3.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>3.96</td>
<td>3.4</td>
<td>0.56</td>
</tr>
<tr>
<td>3.75</td>
<td>3.61</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>3.3</td>
<td>0.7</td>
</tr>
<tr>
<td>4.4</td>
<td>2.98</td>
<td>1.42</td>
</tr>
<tr>
<td>3.95</td>
<td>3.3</td>
<td>0.65</td>
</tr>
<tr>
<td>3.8</td>
<td>3.4</td>
<td>0.4</td>
</tr>
<tr>
<td>3.44</td>
<td>3.9</td>
<td>-0.46</td>
</tr>
<tr>
<td>4</td>
<td>3.69</td>
<td>0.31</td>
</tr>
<tr>
<td>2.7</td>
<td>2.73</td>
<td>-0.03</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>3.6</td>
<td>3.2</td>
<td>0.4</td>
</tr>
<tr>
<td>3.5</td>
<td>2.8</td>
<td>0.7</td>
</tr>
<tr>
<td>3.9</td>
<td>3.99</td>
<td>-0.09</td>
</tr>
<tr>
<td>3.8</td>
<td>3.26</td>
<td>0.54</td>
</tr>
<tr>
<td>3.3</td>
<td>3.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>3.3</td>
<td>3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>3.9</td>
<td>3.8</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>3.52</td>
<td>0.48</td>
</tr>
<tr>
<td>4.3</td>
<td>3.9</td>
<td>0.4</td>
</tr>
<tr>
<td>3.88</td>
<td>3.75</td>
<td>0.13</td>
</tr>
<tr>
<td>3.7</td>
<td>3.7</td>
<td>0</td>
</tr>
<tr>
<td>3.26</td>
<td>3.6</td>
<td>-0.34</td>
</tr>
<tr>
<td>3.93</td>
<td>3.4</td>
<td>0.53</td>
</tr>
</tbody>
</table>

An independent measures t-test is done by simply running a t-test on that third column of differences. The mean of differences is $\bar{D} = -0.2$. The standard deviation of the differences is $S_D = 0.4948$.

You can verify that this mean of differences is the same as the difference of the means: the mean of High School GPAs is 3.63 and the mean of the College GPAs is 3.43. The difference of these means is $3.63 - 3.43 = 0.2$.

The standard error of the mean is:

$$s_D = \frac{s_D}{\sqrt{n}} = \frac{0.4948}{\sqrt{27}} = 0.1$$
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.

$$t = \frac{\bar{D} - \mu_{hyp}}{s_{\bar{D}}} = \frac{-0.2}{0.1} = -2$$

Finally, we use the t-table to see if this is a statistically significant t-statistic. We’ll be using the row for $df = 26$ since we have 27 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.

<table>
<thead>
<tr>
<th>df, one tail</th>
<th>0.25</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.684</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
<td>3.707</td>
</tr>
<tr>
<td>27</td>
<td>0.684</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.690</td>
</tr>
<tr>
<td>28</td>
<td>0.683</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.674</td>
</tr>
</tbody>
</table>

The critical value of $t$ is 2.0555:

Our observed value of $t$ is -2 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 0.056:

<table>
<thead>
<tr>
<th>t</th>
<th>df</th>
<th>$\alpha$ (one tail)</th>
<th>$\alpha$ (two tail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>26</td>
<td>0.028</td>
<td>0.056</td>
</tr>
</tbody>
</table>

The critical value of $t$ is 2.0555:

Our observed value of $t$ is -2 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 0.056:
Convert α to t

<table>
<thead>
<tr>
<th>α</th>
<th>df</th>
<th>t (one tail)</th>
<th>t (two tail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>26</td>
<td>1.7056</td>
<td>2.0555</td>
</tr>
</tbody>
</table>

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

The GPA of High School male students (M = 3.63, SD = 0.4967) is not significantly different than the GPA of College male students (M=3.43, SD = 0.3481), t(26) = -2, p = 0.056.

**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:

\[ d = \frac{|D - \mu_{hyp}|}{s_D} \]

For this example on GPAs:

\[ d = \frac{|\bar{D} - \mu_{hyp}|}{s_D} = \frac{|-0.2 - 0|}{0.4948} = 0.4 \]

This is considered to be a medium 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 0.4, sample size of 27 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.

<table>
<thead>
<tr>
<th>effect size (d)</th>
<th>n</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>27</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( t_{crit} )</th>
<th>( t_{crit} - t_{obs} )</th>
<th>area</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7056</td>
<td>-0.3728</td>
<td>0.6439</td>
<td>0.6439</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( t_{crit} )</th>
<th>( t_{crit} - t_{obs} )</th>
<th>area</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.0555</td>
<td>-4.134</td>
<td>0.0002</td>
<td>0.5092</td>
</tr>
<tr>
<td>2.0555</td>
<td>-0.0229</td>
<td>0.5091</td>
<td></td>
</tr>
</tbody>
</table>
So our observed power is 0.5092.

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:

<table>
<thead>
<tr>
<th>fathers</th>
<th>students</th>
<th>difference (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>71</td>
<td>6</td>
</tr>
<tr>
<td>72</td>
<td>67</td>
<td>5</td>
</tr>
<tr>
<td>72</td>
<td>72</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>62</td>
<td>-2</td>
</tr>
<tr>
<td>70</td>
<td>72</td>
<td>-2</td>
</tr>
<tr>
<td>68</td>
<td>72</td>
<td>-4</td>
</tr>
<tr>
<td>68</td>
<td>69</td>
<td>-1</td>
</tr>
<tr>
<td>67</td>
<td>70</td>
<td>-3</td>
</tr>
<tr>
<td>67</td>
<td>71</td>
<td>-4</td>
</tr>
<tr>
<td>68</td>
<td>72</td>
<td>-4</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>69</td>
<td>-1</td>
</tr>
<tr>
<td>71</td>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td>73</td>
<td>72</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>74</td>
<td>-6</td>
</tr>
<tr>
<td>56</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>67</td>
<td>70</td>
<td>-3</td>
</tr>
<tr>
<td>74</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>70</td>
<td>-3</td>
</tr>
<tr>
<td>67</td>
<td>73</td>
<td>-6</td>
</tr>
<tr>
<td>73</td>
<td>74</td>
<td>-1</td>
</tr>
<tr>
<td>76</td>
<td>72</td>
<td>4</td>
</tr>
<tr>
<td>67</td>
<td>69</td>
<td>-2</td>
</tr>
<tr>
<td>66</td>
<td>67</td>
<td>-1</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>74</td>
<td>-1</td>
</tr>
<tr>
<td>72</td>
<td>72</td>
<td>0</td>
</tr>
</tbody>
</table>

The mean of differences is $\bar{D} = 0.96$. The standard deviation of the differences is $S_D = 2.902$.

The standard error of the mean is:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{2.902}{\sqrt{27}} = 0.56$$

Our t-statistic is:

$$t = \frac{\bar{D} - \mu_{hyp}}{s_{\bar{D}}} = \frac{0.96 - 0}{0.56} = 1.7143$$

Finally, we use the t-table to see if this is a statistically significant t-statistic. We’ll be using
the row for df = 26 since we have 27 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.

<table>
<thead>
<tr>
<th>df, one tail</th>
<th>0.25</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.685</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
<td>3.745</td>
</tr>
<tr>
<td>25</td>
<td>0.684</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td>2.787</td>
<td>3.725</td>
</tr>
<tr>
<td><strong>26</strong></td>
<td><strong>0.684</strong></td>
<td><strong>1.315</strong></td>
<td><strong>1.706</strong></td>
<td><strong>2.056</strong></td>
<td><strong>2.479</strong></td>
<td><strong>2.779</strong></td>
<td><strong>3.707</strong></td>
</tr>
<tr>
<td>27</td>
<td>0.684</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.690</td>
</tr>
<tr>
<td>28</td>
<td>0.683</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.674</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td></td>
</tr>
</tbody>
</table>

The critical value of t is 2.0555:

Our observed value of t is 1.7143 which is not in rejection region.

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

<table>
<thead>
<tr>
<th>Convert t to α</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1.7143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Convert α to t</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>0.05</td>
</tr>
</tbody>
</table>

We therefore fail to reject \( H_0 \) and conclude that, using APA format: "The height of fathers of male students (M = 69.15, SD = 4.4177) is not significantly different than the height of male students (M = 70.11, SD = 3.9839), t(26) = 1.7143, p = 0.0984."
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

```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/Psych315W20survey.csv")
# First find the UW GPA's for the male students
x <- survey$GPA_UW[survey$gender == "Male"]
# 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]
y <- y[goodvals]
# run the t-test. Use 'paired = TRUE' because x and y are dependent
out <- t.test(x,y,
               paired = TRUE,
               alternative = "two.sided",
               var.equal = TRUE)
# The p-pvalue is:
out$p.value
[1] 0.04745029

# Displaying the result in APA format:
sprintf("'t('%.g') = %4.2f, p = %5.4f',out$parameter,out$statistic,out$p.value)
[1] "'t(26) = -2.08, p = 0.0475"
mx <- mean(x)
my <- mean(y)
s <- sd(x-y)
n <- length(x)
#effect size
d <- abs(mx-my)/s
d
[1] 0.4004389

# Find observed power from d, alpha and n
out <- power.t.test(n=n,
                    d = d,
                    sig.level = .05,
                    power = NULL,
                    alternative = "two.sided",
```
# 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"]
# Then find the heights of their fathers
y <- survey$pheight[survey$gender == "Male"]
# Remove the pairs that have a NA in either x or y:
goodvals = !is.na(x) & !is.na(y)
x <- x[goodvals]
y <- y[goodvals]
# run the t-test. Use ‘paired = TRUE’ because x and y are dependent
out <- t.test(x,y,
  paired = TRUE,
  alternative = "two.sided",
  var.equal = TRUE)

# The p-value is:
out$p.value
[1] 0.09654006

# Displaying the result in APA format:
sprintf('t(%g) = %4.2f, p = %5.4f',out$parameter,out$statistic,out$p.value)
[1] "t(26) = 1.72, p = 0.0965"
Questions
Here are 10 practice t-test questions followed by their answers.

1) The conduct of furry and parched brothers

Without anything better to do, you measure the conduct of 33 brothers under two conditions: 'furry' and 'parched'. You then subtract the conduct of the 'furry' from the 'parched' conditions for each brothers and obtain a mean pair-wise difference of 4 with a standard deviation is 11.7306.

Using an alpha value of 0.05, is the conduct from the 'furry' condition significantly less than from the 'parched' condition?
What is the effect size?
What is the observed power of this test?

2) The damage of wealthy and meaty grandmothers

You measure the damage of 67 grandmothers under two conditions: 'wealthy' and 'meaty'. You then subtract the damage of the 'wealthy' from the 'meaty' conditions for each grandmothers and obtain a mean pair-wise difference of 0.66 with a standard deviation is 2.4124.

Using an alpha value of 0.01, is the damage from the 'wealthy' condition significantly less than from the 'meaty' condition?
What is the effect size?
What is the observed power of this test?

3) The evil of uttermost and nutritious cartoon characters

You measure the evil of 13 cartoon characters under two conditions: 'uttermost' and 'nutritious'. You then subtract the evil of the 'uttermost' from the 'nutritious' conditions for each cartoon characters and obtain a mean pair-wise difference of 2.91 with a standard deviation is 7.5312.

Using an alpha value of 0.01, is the evil from the 'uttermost' condition significantly less than from the 'nutritious' condition?
What is the effect size?
What is the observed power of this test?

4) The density of minor and mundane potatoes

Your boss makes you measure the density of 79 potatoes under two conditions: 'minor' and 'mundane'. You then subtract the density of the 'minor' from the 'mundane' conditions for each potatoes and obtain a mean pair-wise difference of 0.14 with a standard deviation is 14.4081.

Using an alpha value of 0.05, is the density from the 'minor' condition significantly less than from the 'mundane' condition?
What is the effect size?
What is the observed power of this test?

5) The advice of direful and chilly psych 315 students
You want to measure the advice of 66 psych 315 students under two conditions: 'direful' and 'chilly'. You then subtract the advice of the 'direful' from the 'chilly' conditions for each psych 315 students and obtain a mean pair-wise difference of -1.38 with a standard deviation is 13.8975.

Using an alpha value of 0.05, is the advice from the 'direful' condition significantly greater than from the 'chilly' condition?

What is the effect size?

What is the observed power of this test?

6) The news of small and large fingers

Your stats professor asks you to measure the news of 64 fingers under two conditions: 'small' and 'large'. You then subtract the news of the 'small' from the 'large' conditions for each fingers and obtain a mean pair-wise difference of 0.16 with a standard deviation is 11.5763.

Using an alpha value of 0.05, is the news from the 'small' condition significantly less than from the 'large' condition?

What is the effect size?

What is the observed power of this test?

7) The jewelry of smoggy and obedient musical groups

On a dare, you measure the jewelry of 75 musical groups under two conditions: 'smoggy' and 'obedient'. You then subtract the jewelry of the 'smoggy' from the 'obedient' conditions for each musical groups and obtain a mean pair-wise difference of 3.25 with a standard deviation is 14.2087.

Using an alpha value of 0.01, is the jewelry from the 'smoggy' condition significantly different than from the 'obedient' condition?

What is the effect size?

What is the observed power of this test?

8) The frequency of foregoing and abundant democrats

We decide to measure the frequency of 28 democrats under two conditions: 'foregoing' and 'abundant'. You then subtract the frequency of the 'foregoing' from the 'abundant' conditions for each democrats and obtain a mean pair-wise difference of 1.1 with a standard deviation is 15.7174.

Using an alpha value of 0.05, is the frequency from the 'foregoing' condition significantly less than from the 'abundant' condition?

What is the effect size?

What is the observed power of this test?

9) The shopping of kaput and smart fingers

You decide to measure the shopping of 13 fingers under two conditions: 'kaput' and 'smart'. You then subtract the shopping of the 'kaput' from the 'smart' conditions for each fingers and obtain a mean pair-wise difference of 2.25 with a standard deviation is 9.1147.

Using an alpha value of 0.05, is the shopping from the 'kaput' condition significantly less
than from the 'smart' condition?
What is the effect size?
What is the observed power of this test?

10) The work of mere and nosy Europeans

You measure the work of 86 Europeans under two conditions: 'mere' and 'nosy'. You then subtract the work of the 'mere' from the 'nosy' conditions for each Europeans and obtain a mean pair-wise difference of 0.88 with a standard deviation is 11.8601. Using an alpha value of 0.01, is the work from the 'mere' condition significantly less than from the 'nosy' condition?
What is the effect size?
What is the observed power of this test?
Answers

1) The conduct of furry and parched brothers

\[ \bar{D} = 4, s_D = 11.7306, n = 33 \]
\[ s_D = \frac{11.7306}{\sqrt{33}} = 2.04 \]
\[ df = 33-1 = 32 \]
\[ t = \frac{4}{2.04} = 1.9608 \]
\[ t_{crit} = 1.69 \]

We reject \( H_0 \).

The conduct of furry brothers (M = 31.33, SD = 6.5949) is significantly less than the conduct of parched brothers (M=35.33, SD = 8.4972), \( t(32) = 1.9608, p = 0.0293 \).

Effect size: \( d = \frac{|\bar{D}|}{s_D} = \frac{4}{11.7306} = 0.34 \) This is a small effect size.

The observed power for one tailed test with an effect size of \( d = 0.34 \), \( n = 33 \) and \( \alpha = 0.05 \) is 0.6015.

# Using R:
sem <- 11.7306/sqrt(33)
t <- (31.3252-35.3276)/2.04
t
[1] -1.961961
p <- pt(t,32,lower.tail = TRUE)
# APA format:
sprintf('t(32) = %4.2f, p = %5.4f',t,p)
[1] "t(32) = -1.96, p = 0.0293"
# Effect size:
d <- abs(4 - 0)/11.7306
d
[1] 0.3409885
# power:
out <- power.t.test(n = 33,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.6073074
2) The damage of wealthy and meaty grandmothers

\[ \bar{D} = 0.66, s_D = 2.4124, n = 67 \]
\[ s \bar{D} = \frac{2.4124}{\sqrt{67}} = 0.29 \]
\[ df = 67-1 = 66 \]
\[ t = \frac{0.66}{0.29} = 2.2759 \]
\[ t_{crit} = 2.38 \]
We fail to reject \( H_0 \).

The damage of wealthy grandmothers (M = 94.57, SD = 1.5943) is not significantly less than the damage of meaty grandmothers (M = 95.23, SD = 1.736), \( t(66) = 2.2759, p = 0.0131 \).

Effect size: 
\[ d = \frac{|\bar{D}|}{s_D} = \frac{0.66}{2.4124} = 0.27 \text{ This is a small effect size.} \]

The observed power for one tailed test with an effect size of \( d = 0.27 \), \( n = 67 \) and \( \alpha = 0.01 \) is 0.4311.

# Using R:
\[
\text{sem} \leftarrow 2.4124/\text{sqrt}(67) \\
t \leftarrow (94.5684-95.2257)/0.29 \\
t
\]
\[ [1] \ -2.266552 \]
\[
p \leftarrow \text{pt}(t,66,\text{lower.tail} = \text{TRUE}) \\
# APA format: \\
\text{sprintf('t(66) = \%4.2f, p = \%5.4f',t,p)} \\
\[ [1] \ "t(66) = -2.27, p = 0.0134" \]
# Effect size: \\
\[
d \leftarrow \text{abs}(0.66 - 0)/2.4124 \\
d
\]
\[ [1] \ 0.2735865 \]
# power: \\
\[
\text{out} \leftarrow \text{power.t.test}(n = 67,d= d,\text{sig.level} = 0.01,\text{power} = \text{NULL}, \\
\text{type} = "\text{one.sample}"\text{,alternative} = "\text{one.sided}"
\]
\[ \text{out$power} \\
\[ [1] \ 0.4471645 \]
3) The evil of uttermost and nutritious cartoon characters

\[
\bar{D} = 2.91, s_D = 7.5312, n = 13
\]

\[
s_D = \frac{7.5312}{\sqrt{13}} = 2.09
\]

df = 13-1 = 12

\[
t = \frac{2.91}{2.09} = 1.3923
\]

\[
t_{crit} = 2.68
\]

We fail to reject \( H_0 \).

The evil of uttermost cartoon characters (\( M = 33.47, \ SD = 5.698 \)) is not significantly less than the evil of nutritious cartoon characters (\( M=36.38, \ SD = 5.5003 \)), \( t(12) = 1.3923, p = 0.0945 \).

Effect size: \( d = \frac{\bar{D}}{s_D} = \frac{2.91}{7.5312} = 0.39 \) This is a small effect size.

The observed power for one tailed test with an effect size of \( d = 0.39, n = 13 \) and \( \alpha = 0.01 \) is 0.1132.

# Using R:

```r
sem <- 7.5312/sqrt(13)
t <- (33.4681-36.3785)/2.09
t
[1] -1.392536
p <- pt(t,12,lower.tail = TRUE)
# APA format:
sprintf('t(12) = %.4f, p = %.5f',t,p)
[1] "t(12) = -1.39, p = 0.0945"
# Effect size:
d <- abs(2.91 - 0)/7.5312
d
[1] 0.3863926
# power:
out <- power.t.test(n = 13,d= d, sig.level = 0.01, power = NULL, type = "one.sample", alternative = "one.sided")
out$power
[1] 0.1390311
```
4) The density of minor and mundane potatoes

\[ \bar{D} = 0.14, s_D = 14.4081, n = 79 \]

\[ s_D = \frac{14.4081}{\sqrt{79}} = 1.62 \]

\[ df = 79-1 = 78 \]

\[ t = \frac{0.14}{1.62} = 0.0864 \]

\[ t_{crit} = 1.66 \]

We fail to reject \( H_0 \).

The density of minor potatoes (M = 43.08, SD = 8.4052) is not significantly less than the density of mundane potatoes (M=43.22, SD = 9.9153), t(78) = 0.0864, p = 0.4657.

Effect size: 
\[ d = \frac{|\bar{D}|}{s_D} = \frac{0.14}{14.4081} = 0.01 \] This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.01, n = 79 and \( \alpha = 0.05 \) is 0.0596.

# Using R:
```r
sem <- 14.4081/sqrt(79)
t <- (43.0761-43.216)/1.62 
t [1] -0.08635802
p <- pt(t,78,lower.tail = TRUE)
# APA format:
sprintf('t(78) = %4.2f, p = %5.4f',t,p)
[1] "t(78) = -0.09, p = 0.4657"
# Effect size:
d <- abs(0.14 - 0)/14.4081
d [1] 0.009716757
# power:
out <- power.t.test(n = 79,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.05947035
```
5) The advice of direful and chilly psych 315 students

\[ \bar{D} = -1.38, s_D = 13.8975, n = 66 \]

\[ s\bar{D} = \frac{13.8975}{\sqrt{66}} = 1.71 \]

\[ df = 66-1 = 65 \]

\[ t = \frac{-1.38}{1.71} = -0.807 \]

\[ t_{crit} = -1.67 \]

We fail to reject \( H_0 \).

The advice of direful psych 315 students (M = 11.94, SD = 11.0697) is not significantly greater than the advice of chilly psych 315 students (M=10.56, SD = 9.0879), t(65) = -0.807, p = 0.2113.

Effect size: \( d = \frac{\bar{D}}{s_D} = \frac{-1.38}{13.8975} = 0.1 \) This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.1, n = 66 and \( \alpha = 0.05 \) is 0.1975.

# Using R:

```r
sem <- 13.8975/sqrt(66)
t <- (11.9394-10.5639)/1.71
t
[1] 0.804386
p <- pt(t,65,lower.tail = FALSE)
# APA format:
sprintf('t(65) = %4.2f, p = %5.4f',t,p)
[1] "t(65) = 0.80, p = 0.2121"
# Effect size:
d <- abs(-1.38 - 0)/13.8975
d
[1] 0.09929843
# power:
out <- power.t.test(n = 66,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.1986283
```
6) The news of small and large fingers

$$\bar{D} = 0.16, s_D = 11.5763, n = 64$$

$$s_D = \frac{11.5763}{\sqrt{64}} = 1.45$$

df = 64-1 = 63

$$t = \frac{0.16}{1.45} = 0.1103$$

$$t_{crit} = 1.67$$

We fail to reject $$H_0$$.

The news of small fingers ($M = 88.37, SD = 7.3783$) is not significantly less than the news of large fingers ($M=88.53, SD = 7.6427$), $$t(63) = 0.1103, p = 0.4563$$.

Effect size: $d = \frac{|\bar{D}|}{s_D} = \frac{0.16}{11.5763} = 0.01$ This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.01, n = 64$ and $\alpha = 0.05$ is 0.0585.

# Using R:

```R
sem <- 11.5763/sqrt(64)
t <- (88.3696-88.5265)/1.45
t
[1] -0.1082069
p <- pt(t,63,lower.tail = TRUE)
# APA format:
sprintf('t(63) = %4.2f, p = %5.4f',t,p)
[1] "t(63) = -0.11, p = 0.4571"
# Effect size:
d <- abs(0.16 - 0)/11.5763
d
[1] 0.01382134
# power:
out <- power.t.test(n = 64,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.06233488
```
7) The jewelry of smoggy and obedient musical groups

\[ \bar{D} = 3.25, s_D = 14.2087, n = 75 \]

\[ s_D = \frac{14.2087}{\sqrt{75}} = 1.64 \]

df = 75-1 = 74

\[ t = \frac{3.25}{1.64} = 1.9817 \]

\[ t_{crit} = \pm 2.64 \]

We fail to reject \( H_0 \).

The jewelry of smoggy musical groups (\( M = 85.49, SD = 9.7644 \)) is not significantly different than the jewelry of obedient musical groups (\( M=88.74, SD = 8.9408 \)), \( t(74) = 1.9817, p = 0.0512 \).

Effect size: \( d = \frac{\bar{D}}{s_D} = \frac{3.25}{14.2087} = 0.23 \) This is a small effect size.

The observed power for two tailed test with an effect size of \( d = 0.23 \), \( n = 75 \) and \( \alpha = 0.01 \) is 0.2582.

# Using R:

```r
sem <- 14.2087/sqrt(75)
t <- (85.4859-88.7397)/1.64
t
[1] -1.984024
# Since this is a two-tailed test, use abs(t) and lower.tail = FALSE
p <- 2*pt(abs(t),74,lower.tail = FALSE)
# APA format:
sprintf('t(74) = %4.2f, p = %5.4f',t,p)
[1] "t(74) = -1.98, p = 0.0510"
# Effect size:
d <- abs(3.25 - 0)/14.2087
d
[1] 0.2287331
# power:
out <- power.t.test(n = 75,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "two.sided")
out$power
[1] 0.2613595
```
8) The frequency of foregoing and abundant democrats

\[ \bar{D} = 1.1, s_D = 15.7174, n = 28 \]

\[ s_D = \frac{15.7174}{\sqrt{28}} = 2.97 \]

df = 28-1 = 27

\[ t = \frac{1.1}{2.97} = 0.3704 \]

\[ t_{crit} = 1.70 \]

We fail to reject \( H_0 \).

The frequency of foregoing democrats (M = 54.68, SD = 10.1185) is not significantly less than the frequency of abundant democrats (M=55.78, SD = 8.4382), \( t(27) = 0.3704, p = 0.357 \).

Effect size: \( d = \frac{|\bar{D}|}{s_D} = \frac{1.1}{15.7174} = 0.07 \) This is a small effect size.

The observed power for one tailed test with an effect size of \( d = 0.07, n = 28 \) and \( \alpha = 0.05 \) is 0.0969.

# Using R:

```r
sem <- 15.7174/sqrt(28)
t <- (54.6771-55.782)/2.97
t
[1] -0.3720202
p <- pt(t,27,lower.tail = TRUE)
# APA format:
sprintf('t(27) = %4.2f, p = %5.4f',t,p)
[1] "t(27) = -0.37, p = 0.3564"
# Effect size:
d <- abs(1.1 - 0)/15.7174
d
[1] 0.06998613
# power:
out <- power.t.test(n = 28,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.09961985
```
9) The shopping of kaput and smart fingers

\[ \bar{D} = 2.25, s_D = 9.1147, n = 13 \]

\[ \bar{D} = \frac{9.1147}{\sqrt{13}} = 2.53 \]

\[ df = 13-1 = 12 \]

\[ t = \frac{2.25}{2.53} = 0.8893 \]

\[ t_{crit} = 1.78 \]

We fail to reject \( H_0 \).

The shopping of kaput fingers (M = 69.48, SD = 6.8591) is not significantly less than the shopping of smart fingers (M=71.74, SD = 5.1979), t(12) = 0.8893, p = 0.1957.

Effect size: \( d = \frac{|\bar{D}|}{s_D} = \frac{2.25}{9.1147} = 0.25 \) This is a small effect size.

The observed power for one tailed test with an effect size of d = 0.25, n = 13 and \( \alpha = 0.05 \) is 0.1978.

# Using R:

```r
sem <- 9.1147/sqrt(13)
t <- (69.4837-71.7372)/2.53
t
[1] -0.8907115
p <- pt(t,12,lower.tail = TRUE)
# APA format:
sprintf('t(12) = %.4f, p = %.5f',t,p)
[1] "t(12) = -0.89, p = 0.1953"
# Effect size:
d <- abs(2.25 - 0)/9.1147
d
[1] 0.246854
# power:
out <- power.t.test(n = 13,d= d,sig.level = 0.05,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.21055
```
**10)** The work of mere and nosy Europeans

\[ \bar{D} = 0.88, s_D = 11.8601, n = 86 \]

\[ s_D = \frac{11.8601}{\sqrt{86}} = 1.28 \]

\[ df = 86-1 = 85 \]

\[ t = \frac{0.88}{1.28} = 0.6875 \]

\[ t_{crit} = 2.37 \]

We fail to reject \( H_0 \).

The work of mere Europeans (M = 17.68, SD = 8.1884) is not significantly less than the work of nosy Europeans (M=18.56, SD = 7.8397), \( t(85) = 0.6875, p = 0.2468 \).

Effect size: \( d = \frac{|\bar{D}|}{s_D} = \frac{0.88}{11.8601} = 0.07 \) This is a small effect size.

The observed power for one tailed test with an effect size of \( d = 0.07 \), \( n = 86 \) and \( \alpha = 0.01 \) is 0.0444.

```r
# Using R:
sem <- 11.8601/sqrt(86)
t <- (17.6811-18.5622)/1.28
t
[1] -0.6883594
p <- pt(t,85,lower.tail = TRUE)
# APA format:
sprintf('t(85) = %.2f, p = %.5f,t,p)
[1] "t(85) = -0.69, p = 0.2466"
# Effect size:
d <- abs(0.88 - 0)/11.8601
d
[1] 0.07419836
# power:
out <- power.t.test(n = 86,d= d,sig.level = 0.01,power = NULL,
type = "one.sample",alternative = "one.sided")
out$power
[1] 0.04956016
```