One-Sample and Two-Sample Means Tests
1 Sample t Test

The 1 sample t test allows us to determine whether the mean of a sample data set is different than a known value.

• Used when the population variance is not known.
• Can be used when the sample size is small.
• Use n-1 degrees of freedom.
For example, we are interested in determining if the mean per capita income of West Virginia counties is different than the national average, and we suspect based on *a priori* (before hand) knowledge that it is lower.

Our hypotheses are:

\( H_0 \): The per capita income of West Virginia counties is not significantly less than the national average.

\( H_a \): The per capita income of West Virginia counties is significantly less than the national average.
Procedure Steps

1. Determine if West Virginia county per capita income is normally distributed.

2. Proceed with the 1 sample t test.

We do not need to know the normality of US per capita income. All we need to know is the mean.
<table>
<thead>
<tr>
<th></th>
<th>Per Capita Income</th>
<th>Median Household Income</th>
<th>Median Family Income</th>
<th>Population</th>
<th>Households</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States</strong></td>
<td>$27,334</td>
<td>$51,914</td>
<td>$62,982</td>
<td>308,745,538</td>
<td>116,716,292</td>
</tr>
<tr>
<td><strong>West Virginia</strong></td>
<td>$19,443</td>
<td>$38,380</td>
<td>$48,896</td>
<td>1,852,994</td>
<td>763,831</td>
</tr>
</tbody>
</table>

This is why we suspect WV is lower. But is this difference statistically significant?

Important: remember that we develop null hypotheses based on theory and/or what we see in the data.
### Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>PcapInc</td>
<td>.089</td>
<td>54</td>
</tr>
</tbody>
</table>

<sup>a</sup> Lilliefors Significance Correction

<sup>*</sup> This is a lower bound of the true significance.

![Normal Q-Q Plot of PcapInc](image)
W/S Test for Normality

\[ sd = \$3,075.28 \]
\[ range = \$16,778 \]

\[ q = \frac{w}{s} \]

\[ q = \frac{\$16,778}{\$3,075} \]

\[ q = 5.46 \]

\[ q_{\text{critical}} = 3.90, 5.46 \]

Since \(3.90 < q = 5.46 < 5.46\), accept Ho, the data are not significantly different than normal \((w/s_{5.46}, p > 0.05)\). Just barely normal.
The ‘barely normal’ result of the w/s test is due to its testing of $k_3$, so it is more sensitive to leptokurtic data.
The $t$ statistic is calculated as:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Where $\bar{x}$ is the sample mean, $\mu$ is the true mean, $s$ is the sample standard deviation, and $n$ is the sample size.
The effect of sample size on t test results:

\[
t(n=7) = \frac{15 - 13}{2 / \sqrt{7}} = \frac{2}{2 / 2.65} = 2 \times \frac{2}{0.76} = 2.63
\]

\[
t(n=14) = \frac{15 - 13}{2 / \sqrt{14}} = \frac{2}{2 / 3.74} = \frac{2}{0.53} = 3.77
\]

Small sample sizes make tests MORE conservative (e.g. harder to gain significance).

- Small but important differences may not be detected.

Increasing the sample size allows us to detect smaller, but statistically significant, differences.

- Larger sample sizes make statistical test more powerful... but remember there is a point of diminishing returns.
We take the calculated t value and the df to the t table to determine the probability.

Although we ignore the sign on the t table, we can tell from the sign that WV per capita income is less than the national average.
\[ t = -9.93 \]
\[ df = 53 \]

Since our df is not listed, use the next lower df.

Where does the calculated \( t \) (-9.93) fall?

<table>
<thead>
<tr>
<th>df</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
<th>.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
<td>6.965</td>
<td>9.925</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
</tr>
<tr>
<td>13</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
<td>2.120</td>
<td>2.583</td>
<td>2.921</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.740</td>
<td>2.110</td>
<td>2.567</td>
<td>2.898</td>
</tr>
<tr>
<td>18</td>
<td>1.330</td>
<td>1.734</td>
<td>2.101</td>
<td>2.552</td>
<td>2.878</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.539</td>
<td>2.861</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.528</td>
<td>2.845</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.518</td>
<td>2.831</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td>2.819</td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.500</td>
<td>2.807</td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td>2.787</td>
</tr>
<tr>
<td>26</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
</tr>
<tr>
<td>27</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
</tr>
<tr>
<td>28</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.462</td>
<td>2.756</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td><strong>1.684</strong></td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
</tr>
<tr>
<td>120</td>
<td>1.289</td>
<td>1.658</td>
<td>1.980</td>
<td>2.358</td>
<td>2.617</td>
</tr>
<tr>
<td>∞</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
</tr>
</tbody>
</table>

Adapted from Table III of Fisher and Yates (1974).
Since $-9.93 < 1.684$, reject $H_0$.

_The per capita income of West Virginia is significantly less than the national average ($t_{-9.93}, p < 0.005$)._

West Virginia’s per capita income is very low.
Note that the confidence intervals for the mean are calculated using the critical $t$ value from the table. Use the 2-tailed value or 0.025 if it is a 1-tailed table.

\[
pr \left[ \left( \bar{x} - t_{\sigma, df} \frac{s}{\sqrt{n}} \right) \leq \mu \leq \left( \bar{x} + t_{\sigma, df} \frac{s}{\sqrt{n}} \right) \right] = 1 - \alpha
\]

If the df you need is not in the table, use the next LOWER value.

\[
pr \left[ \left( 19553.2 - 2.021 \frac{2757.04}{\sqrt{54}} \right) \leq \mu \leq \left( 19553.2 + 2.021 \frac{2757.04}{\sqrt{54}} \right) \right] =
\]

\[18794.9 \leq \mu \leq 20311.5 \quad \leftarrow \text{We are 95\% confident that the mean income for West Virginia falls within this range.}\]
We’re using a 1-tailed table, so we need to split 0.05 between the tails.

Therefore, use the 0.025 column.
Two-Sample T-Test for Means

• Used to compare one sample mean to another.

• Two different test:
  
  − Equal variances
  
  − Unequal variances

• *Homoscedasticity* – the assumption of equal variances.
When data meet the assumption of normality we can use confidence intervals to quantify the area of uncertainty (in this case 5%).

We are 95% confident that the mean from this group…

... will not fall in this range.

Here we are 95% confident that the means do not overlap and that these two groups are significantly different.
In this example the means could occur in an overlapping region and we are less than 95% certain that they are significantly different.
When the variances are not equal, the region of overlap is asymmetrical and the probabilities associated with the location of the mean are not the same. To avoid a Type 1 error we use a more conservative approach.
Test 1: Equal Variances

The test statistic is:

\[ t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \]

\[ df = n_1+n_2-2 \]

\( s^2_1 \) and \( s^2_2 \) are the variances (or squared standard deviations) for each sample, \( n_1 \) and \( n_2 \) are the group samples sizes.

Therefore \( t \) is the distance between the two means, in standard deviation units, taking into consideration sample size and spread.

For the 2-sample t test we know 2 means, therefore the degrees of freedom would be: \( df = n_1+n_2-2 \).
When comparing two or more groups, each group MUST tested for normality individually.

If we pool the data and test for normality, then we are assuming that the data are from the same population... which is what we are trying to determine with the t test.
To determine if the variances are equal, use the equation:

$$F = \frac{s_1^2}{s_2^2}$$

$$df = (n_1 - 1, n_2 - 1)$$

Which is just the ratio of one variance to the other. The $F$ results are then compared to the $F$ table.
### Critical Values of the F-Distribution

*Taken from Rohlf and Sokal, 1981 Table 16*

<table>
<thead>
<tr>
<th>Numerator Degrees of Freedom (V1)</th>
<th>Denominator Degrees of Freedom (V1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>0.01</td>
<td>1.72</td>
</tr>
<tr>
<td>0.05</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>161</td>
</tr>
<tr>
<td>0.025</td>
<td>648</td>
</tr>
<tr>
<td>.005</td>
<td>40500</td>
</tr>
<tr>
<td>0.01</td>
<td>1620000</td>
</tr>
<tr>
<td>0.005</td>
<td>4350000</td>
</tr>
</tbody>
</table>

Note that there are several additional pages not shown...
Example: 2-Sample T Test with Equal Variances

Research question: Is there a difference in the sitting height between the Arctic and Great Plains native Americans?

\[ H_0 : \text{There is no significant difference in the sitting heights of Arctic and Great Plains native Americans.} \]

\[ H_a : \text{There is a significant difference in the sitting heights of Arctic and Great Plains native Americans.} \]

<table>
<thead>
<tr>
<th></th>
<th>Arctic</th>
<th>Plains</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Mean</td>
<td>864.9</td>
<td>882.7</td>
</tr>
<tr>
<td>( s^2 )</td>
<td>488.2</td>
<td>432.6</td>
</tr>
<tr>
<td>Range</td>
<td>71.2</td>
<td>84.22</td>
</tr>
</tbody>
</table>
The order of operations for conducting the test is:

1. Test each group for normality. Proceed with t test if both groups are normal.

2. Conduct an F test.
   A. If $H_o$ is accepted then proceed with equal variances t test.
   B. If $H_0$ is rejected then proceed with Welch’s approximate t test.

3. Conduct appropriate t test.
Test each group for normality:

**Arctic**

\[ n = 8 \]

\[ \text{Range}_{\text{Arctic}} = 71.2 \]

\[ \text{Variance}_{\text{Arctic}} = 488.2 \]

\[ q_{\text{Arctic}} = \frac{71.2}{\sqrt{488.2}} = 3.22 \]

\[ q_{\text{Critical}} = 2.50, 3.399 \]

Since \[ 2.50 < q_{3.22} < 3.399 \], accept \( H_0 \).

Arctic sitting height is not different than normal \( (q_{3.22}, p>0.05) \).

**Plains**

\[ n = 16 \]

\[ \text{Range}_{\text{Plains}} = 84.22 \]

\[ \text{Variance}_{\text{Plains}} = 432.6 \]

\[ q_{\text{Arctic}} = \frac{84.22}{\sqrt{432.6}} = 4.049 \]

\[ q_{\text{Critical}} = 3.01, 4.24 \]

Since \[ 3.01 < q_{4.049} < 4.24 \], accept \( H_0 \).

Plains sitting height is not different than normal \( (q_{4.049}, p>0.05) \).

Since both groups are normal, proceed to the variance (F) test.
Variance or F Test:

$H_0$ : The variances are not significantly different.

$H_a$ : The variances are significantly different.

$\alpha = 0.05$

$$F = \frac{488.2}{432.6} = 1.13 \quad \text{df} = (16 - 1, 8 - 1) = (15, 7)$$

From the F table the critical value for 15,7 df is 2.71.
Note that on the F table the probabilities are read in the columns rather than the rows.

The calculated F statistic falls here between 0.25 and 0.50.
Since our calculated F statistic is less than the critical value from the table (1.13 < 2.71) we assume that the variances are equal.

*The variances for two sitting heights are not significantly different (F\textsubscript{1.13}, 0.50 > p > 0.25).*

Now we can proceed with the two-sample t test...
Since we do not care about direction (larger or smaller) the sign does not matter. If we do care about direction then the sign is very important.
For example:

No direction implied (doesn’t matter):

\[
t = \frac{864.9 - 882.7}{\sqrt{\frac{488.2}{8} + \frac{432.6}{16}}} = \frac{-17.8}{9.38} = -1.9
\]
Sign has no meaning.

Arctic tribe shorter than plains tribe:

\[
t = \frac{864.9 - 882.7}{\sqrt{\frac{488.2}{8} + \frac{432.6}{16}}} = \frac{-17.8}{9.38} = -1.9
\]
Sign means the arctic tribe’s mean is less than the plains tribe’s mean.

Plains tribe taller than arctic tribe:

\[
t = \frac{882.7 - 864.9}{\sqrt{\frac{432.6}{16} + \frac{488.2}{8}}} = \frac{-17.8}{9.38} = 1.9
\]
Sign means plains the tribe’s mean is greater than the arctic tribe’s mean.
Remember that the positioning of each group in the numerator of the $t$ equation is related to the sign of the results.

Be very careful how you construct the equation’s numerator, AND the way you interpret the results.
Table from Zar, 1984 Table B.3

Critical Values of the $t$ Distribution

Note: use a 2-tailed test since we are just interested in whether there is a difference, not if one is greater or less than the other.
In Table A.3 the critical value for $df=22$ is 2.074.

Since $1.95 < 2.074$ we therefore accept $H_0$:

There is no significant difference in the sitting heights of Arctic and Great Plains native Americans ($t_{-1.95}, 0.10 > p > 0.05$).

The probability range (e.g. $0.10 > p > 0.05$) are from the t-table.
Critical Values of the t-Distribution
Taken from Zar, 1984 Table B.3

Note: 2-tailed t = 1.95 for a df(v) of 22 falls between 0.10 and 0.05.
Test 2: Unequal Variances (Welch’s approximate t)

The test statistic is:

\[
t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}
\quad \text{and} \quad
\text{df} = \frac{\left(\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}\right)^2}{\frac{\left(\frac{s^2_1}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s^2_2}{n_2}\right)^2}{n_2 - 1}}
\]

\(s^2_1\) and \(s^2_2\) are the variances (or squared standard deviations) for each sample, \(n_1\) and \(n_2\) are the group samples sizes.

The df may be non-integer, in which case the next smaller integer should be used. Take this number to the t table.
Is the vegetation index for wetland 35 less than the index for 23?

<table>
<thead>
<tr>
<th>ID</th>
<th>Kolmogorov-Smirnov(^a) Statistic</th>
<th>Shapiro-Wilk Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID23</td>
<td>.054</td>
<td>.991</td>
</tr>
<tr>
<td>ID35</td>
<td>.066</td>
<td>.989</td>
</tr>
</tbody>
</table>

\(F_{\text{Critical}}(144,144) = 1.35\)

From F table for df 120,120 (next lower on the table)

The wetland variances are significantly different \((F_{11.8}, p < 0.001)\).
The vegetation index for wetland 35 is significantly less than the index for wetland 23 \((t_{-103.5}, p < 0.001)\).
Variable Distributions and Test Statistic Distributions

The variable may have a negative exponential distribution (e.g. frequency of a Ebola cases per outbreak).

- Most locations have very few people with Ebola.
• Take 100 samples from the previous distribution.

• Calculate a mean for each sample.

• The central limit theorem states that the sample means will be normally distributed.

24 sample means

48 sample means

72 sample means
• Some sample means will be greater than the true mean, some will be less.

• However, if the number of sample means is large they will take on the properties of a normal distribution.

• This is true even if the underlying population has a different distribution.
Robustness in Statistics

*Robust* – when a statistical technique remains useful even when one or more of its assumptions are violated.

In general, if the sample size is large moderate deviations from the statistical technique’s assumptions will not invalidate the results.

The F and t tests are considered to be fairly robust... other tests are not, so simply increasing the sample size may not be the best approach if the data violate assumptions.