Data
For this exercise we will use measurements of snow water equivalence (SWE) collected by the State of Colorado using a snow course and snow water equivalence measured by the Natural Resources Conservation Service using a SnoTel system. How well do SnoTel sites work for measuring SWE? Here we assume that SWE measured at snow courses is the "true" SWE measurement. At Wolf Creek pass we will compare annual SWE measurements made in April from snow course measurements to SnoTel measurements.

Question 1
For both sites, calculate the following parameters (10 points):
1. mean
   \[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]
2. standard deviation
   \[ s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \]
3. coefficient of variation
   \[ CV = \frac{s}{\bar{x}} \times 100\% \]
4. standard error
   \[ SE = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}} \]

Using the formulas above, the statistics were computed for both course measurements and snotel measurements.
<table>
<thead>
<tr>
<th>Year</th>
<th>Course [inches]</th>
<th>SnoTel [inches]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>52.3</td>
<td>57.5</td>
</tr>
<tr>
<td>1980</td>
<td>45.6</td>
<td>50.1</td>
</tr>
<tr>
<td>1981</td>
<td>19.2</td>
<td>20.6</td>
</tr>
<tr>
<td>1982</td>
<td>41.6</td>
<td>45.6</td>
</tr>
<tr>
<td>1983</td>
<td>34.7</td>
<td>38.5</td>
</tr>
<tr>
<td>1984</td>
<td>29</td>
<td>31.6</td>
</tr>
<tr>
<td>1985</td>
<td>44</td>
<td>48.3</td>
</tr>
<tr>
<td>1986</td>
<td>37</td>
<td>40.5</td>
</tr>
<tr>
<td>1987</td>
<td>36.1</td>
<td>37.9</td>
</tr>
<tr>
<td>1988</td>
<td>21.9</td>
<td>23.6</td>
</tr>
<tr>
<td>1989</td>
<td>33.2</td>
<td>37.9</td>
</tr>
</tbody>
</table>

**Question 2**

Compare SnoTel and snow course annual SWE values using a paired-difference t-test (10 points).

1. What is the mean of the difference between the two methods?

\[
\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = 3.41 \text{ in}
\]

where \( d \) is the difference between the course and snotel measurements and \( n \) is the number of measurements.

2. What is the standard deviation of the difference between the two methods?

\[
s = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}} = 1.33 \text{ in}
\]

where \( s \) is the standard deviation.

3. What is the t-statistic?

\[
t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.41 \text{ in} - 0 \text{ in}}{1.33 \text{ in}/\sqrt{11}} = 8.530
\]

where \( t \) is the t-statistic and \( \mu_0 \) is the null value. For a paired difference test, we are testing if the difference is different than zero \( (\mu_0 = 0) \). The null hypothesis for this test is

\[H_0: \text{the mean of the difference is zero}\]

4. What is the probability or p-value?

The number of degrees of freedom is 1 less than then number of measurement pairs:

\[df = n - 1 = 10\]

The paired difference test is a two-tailed test, because the mean can either be significantly greater than zero or less than zero (and we don’t care which one it is). We will test at the 95% confidence level \( (\alpha = 0.05) \). Using Excel, the probability is determined to be:
5. Is the SnoTel measurement of SWE significantly different than the snow course measurement of SWE at the 0.05 probability level?

For a two-tailed test at the 95% confidence level ($\alpha = 0.05$), the critical $t$-statistic is 2.228.

$$ t > t_{\alpha/2=0.05/2,df=10} $$

Because the calculated $t$-statistic is greater than the critical $t$-statistic, we reject the null hypothesis. Therefore we conclude that snow course and SnoTel measurements are significantly different.

Additional Excel work:

<table>
<thead>
<tr>
<th>Year</th>
<th>Course [inches]</th>
<th>SnoTel [inches]</th>
<th>difference [inches]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>52.3</td>
<td>57.5</td>
<td>5.2</td>
</tr>
<tr>
<td>1980</td>
<td>45.6</td>
<td>50.1</td>
<td>4.5</td>
</tr>
<tr>
<td>1981</td>
<td>19.2</td>
<td>20.6</td>
<td>1.4</td>
</tr>
<tr>
<td>1982</td>
<td>41.6</td>
<td>45.6</td>
<td>4</td>
</tr>
<tr>
<td>1983</td>
<td>34.7</td>
<td>38.5</td>
<td>3.8</td>
</tr>
<tr>
<td>1984</td>
<td>29</td>
<td>31.6</td>
<td>2.6</td>
</tr>
<tr>
<td>1985</td>
<td>44</td>
<td>48.3</td>
<td>4.3</td>
</tr>
<tr>
<td>1986</td>
<td>37</td>
<td>40.5</td>
<td>3.5</td>
</tr>
<tr>
<td>1987</td>
<td>36.1</td>
<td>37.9</td>
<td>1.8</td>
</tr>
<tr>
<td>1988</td>
<td>21.9</td>
<td>23.6</td>
<td>1.7</td>
</tr>
<tr>
<td>1989</td>
<td>33.2</td>
<td>37.9</td>
<td>4.7</td>
</tr>
</tbody>
</table>

| count: | **11** |
| mean:  | **3.41 inches** |
| std deviation: | **1.33 inches** |
| coeff of variation: | **39%** |
| std error: | **0.40 inches** |

**Question 3**

Make the following separate graphs (10 points):

1. Time series of annual SWE for the snow course data.
2. Time series of annual SWE for the SnoTel data.
3. XY scatterplot of SWE measured at the snow course (x-axis) versus SWE measured from the SnoTel site (y-axis).

Question 4
Regress SnoTel annual SWE values against snow course data (20 points).
1. make the same XY scatterplot of SWE measured at the snow course (x-axis) versus SWE measured from the SnoTel site (y-axis).

2. Regress SWE from the SnoTel site (dependent variable) against SWE from the snow course and place the $r^2$ value and the equation for the line on the graph.

In Excel, a linear regression line can be added to the graph by selecting the data and choosing the menu item 'Chart->Add Trendline...'. You can cause the equation and $R^2$ value to appear on the graph by checking the appropriate checkboxes in the Options tab.

A more detailed regression analysis can be run in Excel by choosing the Tools->Data Analysis... menu item. The 'Regression' analysis yields the following results:

```
SUMMARY OUTPUT

Regression Statistics
Multiple R 0.997754089
R Square 0.995513223
Adjusted R Square 0.995014692
Standard Error 0.783077927
Observations 11

ANOMA

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>1224.517464</td>
<td>1224.517464</td>
<td>1996.894032</td>
<td>7.03532E-12</td>
</tr>
<tr>
<td>Residual</td>
<td>9</td>
<td>5.518899353</td>
<td>0.613211039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>1230.036364</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0%
Intercept -0.541397678 0.921913263 -0.587254463 0.571469187 -2.626911959 1.544116602 -2.626911959 1.544116602
X Variable 1 1.110125125 0.02484245 44.6862028 7.03532E-12 1.053927557 1.166322694 1.053927557 1.166322694
```

3. what is the slope of the regression line?

From the graph or the regression output, \( \text{slope} = 1.11 \). \( \text{Snotel SWE [in]} \) = \( \frac{\text{snowcourse SWE [in]}}{\text{Snotel SWE [in]}} \)

4. what is the y-intercept of the regression line?

From the graph or the regression output, \( \text{y-intercept} = -0.54 \) Snotel SWE [in]
5. **what is the t-statistic for the slope of the line?**
   From the regression output, $t_{stat} = 44.69$. This large t-statistic value occurs because the slope is different than 0.

6. **what is the probability for the slope of the line?**
   From the regression output, $P = 7.03 \times 10^{-12}$. This small p-value also occurs because the slope is different than 0.

7. **is this a significant relationship?**
   With our current set of tools, we cannot determine whether or not there is a significant statistical relationship. There are statistical tests to determine whether or not a relationship is statistically significant, but we have not studied these statistical tests.
   The high $R^2$ value of 0.9955 indicates that there is a high linear correlation between the Snow Course and Snotel measurements. And, because the 95% confidence range for the slope (1.054 to 1.166) does not contain a slope of 1 (unity), the Snow Course and Snotel measurements do not give us the same measurements.