1) An agricultural field trial compares the yield of two varieties of tomato for commercial use. The researchers divide in half each of 10 small plots of land in different locations and plant each tomato variety on one half of each plot. After harvest, they compare the yields in pounds per plant at each location. The 10 differences (Variety A – Variety B) give $x-bar = 0.34$ and a sample standard deviation of $s = 0.83$. Is there convincing evidence of a difference in yield at the 5% significance level? Describe what a Type I and Type II error would be in this situation.

Matched Pairs t-test
Assume random, small sample size (n<15) so assume no outliers or skew.

$H_0 : \mu_d = 0$  There is no difference between the yield of the two varieties of tomatoes.

$H_a : \mu_d \neq 0$  There is a difference between the yield of the two varieties of tomatoes.

$$t = \frac{\bar{x}_d - \mu_d}{s/\sqrt{n}} = \frac{0.34 - 0}{0.83/\sqrt{10}} = 1.2954$$

$p = 2P(t > 1.2954) = 0.2274$  

$\alpha = .05$

$p > \alpha$  Fail to reject the null hypothesis.

$.2274 > .05$

There is not sufficient evidence to suggest that there is a difference between the yield of the two varieties of tomatoes.

Type 1 error: Thinking there is a difference between tomatoes when there really is not.

Type 2 error: Thinking there was not a difference in the tomatoes when there actually is a difference.
Popular wisdom is that eating presweetened cereal tends to increase the number of dental cavities in children. A sample of children was entered into a study (with parental consent) and was followed for several years. Each child was classified as a sweetened-cereal lover or a nonsweetened-cereal lover. At the end of the study, the amount of tooth damage was measured by the number of cavities. The table shows the mean number of cavities and the standard deviation for each group.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>mean</th>
<th>std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugar</td>
<td>30</td>
<td>6.41</td>
<td>1.72</td>
</tr>
<tr>
<td>No sugar</td>
<td>35</td>
<td>5.20</td>
<td>2.34</td>
</tr>
</tbody>
</table>

Use a 90% confidence interval to estimate the difference in the mean number of cavities for children who eat sweetened cereal and children who don’t.

Does the sample provide evidence that the mean number of cavities for children who eat sweetened cereal is more than the mean number of cavities for children who do not eat sweetened cereal?

**2-sample t-interval**

Assume random, medium sample sizes (15<n<40) so assume no extreme outliers or skew

1: sugar group, 2: no sugar group

\[
(x_1 - x_2) \pm t * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (6.41 - 5.20) \pm t * \sqrt{\frac{1.72^2}{30} + \frac{2.34^2}{35}} = (.3666, 2.0534)
\]

Since 0 is in the interval, there is not evidence to suggest that children who eat sugar cereal have more cavities.
A study of chromosome abnormalities and criminality examined data from 4,134 males born in Copenhagen. Each man was classified as having a criminal record or not, using the registers maintained in the local police offices. Each was also classified as having the normal male XY chromosome pair or one of the abnormalities XYY or XXY. Of the 4,096 men with normal chromosomes 381 had criminal records, while 11 of the 38 men with normal chromosomes had criminal records. Some experts believe chromosome abnormalities are associated with increased criminality. Do these data lend support to this belief? Perform a significance test to answer this question.

Construct and interpret a 95% confidence interval for the difference in proportions.

**2-proportion z-test**
Assume random
1: normal chromosomes, 2: abnormal chromosomes

\[ n_1 \hat{p}_1 \geq 10 \quad n_1 (1 - \hat{p}_1) \geq 10 \quad \text{population} > 10n_1 \]
\[ n_2 \hat{p}_2 \geq 10 \quad n_2 (1 - \hat{p}_2) \geq 10 \quad \text{population} > 10n_2 \]

\[ 381 \geq 10 \quad 3715 \geq 10 \quad \text{all men} > 40960 \]
\[ 11 \geq 10 \quad 27 \geq 10 \quad \text{all men} > 380 \]

\[ H_0 : p_1 - p_2 = 0 \quad \text{There is no difference in the proportion of criminals with normal versus abnormal chromosomes.} \]

\[ H_a : p_1 - p_2 < 0 \quad \text{The proportion of criminals with abnormal chromosomes is higher.} \]

Pool the proportions: \( \hat{p} = \frac{392}{4134} \)

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p}{1-p} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{381}{4096} - \frac{11}{38} = -4.1146 \quad p = P(z < -4.1146) = 0.00002 \]

\[ \alpha = .05 \quad p > \alpha \quad \text{Reject the null} \]

\[ .00002 > .05 \]

There is sufficient evidence to suggest that the proportion of criminals with abnormal chromosomes is higher.

**2-proportion z-interval**
Assume random
1: normal chromosomes, 2: abnormal chromosomes

\[ n_1 \hat{p}_1 \geq 10 \quad n_1 (1 - \hat{p}_1) \geq 10 \quad \text{population} > 10n_1 \]
\[ n_2 \hat{p}_2 \geq 10 \quad n_2 (1 - \hat{p}_2) \geq 10 \quad \text{population} > 10n_2 \]

\[ 381 \geq 10 \quad 3715 \geq 10 \quad \text{all men} > 40960 \]
\[ 11 \geq 10 \quad 27 \geq 10 \quad \text{all men} > 380 \]

\[ (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \left( \frac{381}{4096} - \frac{11}{38} \right) \pm 1.96 \sqrt{\frac{381}{4096} - \frac{381}{4096} + \frac{11}{38} \frac{11}{38} = (-.341, -.052) \]

We are 95% sure that the true difference in the proportion of criminals is between 34.1% and 5.2%.