Outline

1. Situation: a categorical variable with two levels and a quantitative variable
   (a) Simple random samples from each of two populations
   (b) Simple random sample from one population divided into two groups based on a
        categorical variable
   (c) Randomized comparative experiment with two treatments

2. Notation:

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>St. Dev.</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>St. Dev.</td>
</tr>
<tr>
<td>1</td>
<td>$x_1$</td>
<td>$\mu_1$</td>
<td>$n_1$</td>
</tr>
<tr>
<td>2</td>
<td>$x_2$</td>
<td>$\mu_2$</td>
<td>$n_2$</td>
</tr>
</tbody>
</table>

3. Goals of inference: to write confidence intervals for $\mu_1 - \mu_2$ and to test the hypothesis
   $H_0: \mu_1 - \mu_2 = 0$.

4. The normality assumption: each population is normally distributed

5. The fundamental distribution fact:

   \[
   \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
   \]

   has a distribution which is approximately $t$ with degrees of freedom at least the smaller of
   $n_1 - 1$ and $n_2 - 1$ and at most $n_1 + n_2 - 2$.

6. Confidence interval for $\mu_1 - \mu_2$:

   \[
   (\bar{x}_1 - \bar{x}_2) \pm t^* \text{SE} \quad \text{where} \quad SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
   \]

   $t^*$ is the critical value for the confidence level and the right number of degrees of freedom
   (use software or choose the smaller of $n_1 - 1$ and $n_2 - 1$).