Hypothesis tests are used to test certain statements about population parameters. The general method to conduct a hypothesis test is:

1. **Formulate the hypotheses.** Before looking at the data (usually before collecting the data), decide on a null hypothesis $H_0$ and an alternative hypothesis $H_a$. The null hypothesis should be of the form: “some (unknown) population parameter equals some number”. The alternative hypothesis should encompass all the possibilities that the null hypothesis does not. Usually, this will take the form of “the (unknown) population parameter does not equal the chosen number”. However, in certain cases where physical or other circumstances rule out some of the possibilities, the alternative hypothesis may take the form “the (unknown) population parameter is greater than the chosen number” or “the (unknown) population parameter is less than the chosen number”.

2. **Compute the appropriate test statistic.** The appropriate test statistic depends on what parameter is being tested for and is indicated in the table below.

The Big Mathematical Fact underlying hypothesis tests is that: If $H_0$ is true, the test statistic

$$\text{test statistic} = \frac{(\text{estimate of parameter}) - (\text{parameter value under } H_0)}{(\text{standard error of estimate of parameter})}$$

has (or in some cases is approximated by) a known distribution. That is, if we were to take simple random samples of the same size again and again, the distribution of the test statistic above would follow the known distribution.

The following table indicates which parameters can be tested for with which test statistics, and what the associated known distributions are.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Test statistic</th>
<th>Standard error</th>
<th>Distribution</th>
<th>Ch.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$z = \frac{\bar{x} - \mu_0}{\text{SE}_x}$</td>
<td>$\frac{\sigma}{\sqrt{n}}$</td>
<td>$N(0, 1)$</td>
<td>6</td>
</tr>
<tr>
<td>$\Delta \mu = \mu_2 - \mu_1$</td>
<td>$z = \frac{(\bar{x}_2 - \bar{x}<em>1) - \Delta \mu_0}{\text{SE}</em>{\bar{x}_2 - \bar{x}_1}}$</td>
<td>$\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$</td>
<td>$N(0, 1)$</td>
<td>6</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$t = \frac{\bar{x} - \mu_0}{\text{SE}_x}$</td>
<td>$\frac{s}{\sqrt{n}}$</td>
<td>$t(n-1)$</td>
<td>7</td>
</tr>
<tr>
<td>$\Delta \mu = \mu_2 - \mu_1$</td>
<td>$t = \frac{(\bar{x}_2 - \bar{x}<em>1) - \Delta \mu_0}{\text{SE}</em>{\bar{x}_2 - \bar{x}_1}}$</td>
<td>$\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$</td>
<td>$t(\min(n_1, n_2) - 1)$</td>
<td>7</td>
</tr>
<tr>
<td>$p$</td>
<td>$z = \frac{\hat{p} - p_0}{\text{SE}_{\hat{p}}}$</td>
<td>$\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$</td>
<td>$N(0, 1)$</td>
<td>8</td>
</tr>
<tr>
<td>$\Delta p = p_2 - p_1$</td>
<td>$z = \frac{(\hat{p}_2 - \hat{p}<em>1) - \Delta p_0}{\text{SE}</em>{\hat{p}_2 - \hat{p}_1}}$</td>
<td>$\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$</td>
<td>$N(0, 1)$</td>
<td>8</td>
</tr>
</tbody>
</table>

In the last row, if $\Delta p_0 = 0$ (as is usually the case), then the pooled standard error $\text{SE}_{\hat{p}}$ should be used instead of the one given above. To compute the pooled standard error, first compute the pooled estimate of the proportion:

$$\hat{p} = \frac{\text{total success count in both populations}}{\text{total size of both populations}} = \frac{X_1 + X_2}{n_1 + n_2}.$$ 

Then the pooled standard error is

$$\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

3. **Compute the P-value.** Using the Big Mathematical Fact along with the appropriate test statistics, standard errors, and distributions from the table above, we can answer the question: If $H_0$ is true, what is the probability that a simple random sample (or pair of samples) of the given size(s) would produce a test statistic as extreme as or more extreme than the one our sample produced? The answer to this question is called the P-value resulting from the test. It is given by one of the three shaded areas below, depending on what $H_a$ is.

$H_a$: parameter $\neq$ number  
$H_a$: parameter $<$ number  
$H_a$: parameter $>$ number
The distributions depicted above are normal distributions, but the same analysis applies to t-distributions as well when appropriate.

4. Interpreting the $P$-value. When communicating with people who are familiar with statistical inference, a write-up of a hypothesis test may actually conclude simply with a statement of the test conducted and the $P$-value obtained. However, many times you will want to convey the results of the hypothesis test to people unfamiliar with statistical inference.

If you know (or think) this will be the case, before analyzing the data, you should decide upon a significance level $\alpha$ that you will use in assessing the hypothesis test. Traditionally, $\alpha$ is taken to be .05, but this is fairly arbitrary. You may want a smaller or larger significance level, depending on what you are testing.

A small $\alpha$ will mean that if $H_0$ really is true, there is a higher probability that you will (correctly) not reject it. On the other hand, if $H_0$ really is false, there is a higher probability that you will not detect that it is false with the hypothesis test.

A large $\alpha$ will mean that if $H_0$ really is false, there is a higher probability that you will (correctly) reject it. On the other hand, if $H_0$ really is true, there is a higher probability that you will reject it anyway.

Whatever your choice of $\alpha$, to interpret the hypothesis test in terms of statistical significance, the rule is:

\[ \text{If the } P\text{-value is less than or equal to } \alpha, \text{ there is statistically significant evidence to reject the null hypothesis. If the } P\text{-value is greater than } \alpha, \text{ there is not statistically significant evidence to reject the null hypothesis.} \]

Notice that, no matter what the $P$-value is, you never accept the null hypothesis. Not having statistically significant evidence to reject the null hypothesis is not the same as accepting the null hypothesis. In fact, with another test (or maybe just with a larger sample size), you might be able to gather statistically significant evidence to reject the null hypothesis, no matter what the $P$-value for this hypothesis test was.

I would also discourage you from stating that the statistical inference tells you to reject the null hypothesis. Instead, think of a hypothesis test as evidence-gathering. If the $P$-value is below the significance level $\alpha$, then you have evidence to support rejecting the null hypothesis. However, it’s really up to a human being assessing the evidence to decide whether or not that evidence is convincing enough to warrant rejection of the null hypothesis. (Likewise, since the statistics never explicitly implies acceptance of the null hypothesis itself, it is up to a human being interpreting the statistics to decide whether or not to accept the null hypothesis.)

Also, when you interpret your results, state them in terms of what the actual null hypothesis is. For example, do not simply state that you have statistically significant evidence to reject the null hypothesis, but rather that you have statistically significant evidence that the mean of the heights of squirrels on campus does not equal 10 cm. This way, the reader (or listener) does not have to search back through the statistical analysis to understand what your conclusions are.