The Scientific Method and Data Analysis

The eminent German philosopher Karl Popper wrote in his famous essay, *Science as Falsification*, that it is vulnerability to falsification--not constant verification--that is the mark of truly powerful scientific idea.

I. The Scientific Method

Scientists have many ways to approach the testing of scientific ideas, which generally follow the basic sequence you no doubt have seen before.

observation - The investigator notes a phenomenon that poses a problem/question.

hypothesis formulation - The investigator poses the question in such a way that it can be tested by rigorously designed experiments or field observations.

null hypothesis - stated in terms of "no difference between observed results and expected results" of an experiment

alternative hypothesis - the opposite of the null, and actually the statement of interest

prediction - The investigator makes a statement about what s/he believes is true about the hypothesis.

experimental design - The investigator designs an experiment which will yield data to either support or refute the hypothesis.

* data collection - The experiments are run, and data are collected.

* data analysis - The data are subjected to rigorous analysis via quantification and statistical tests to determine whether any deviation from the expected result is truly meaningful, or merely due to chance.

conclusion - The investigator accepts or rejects the null hypothesis, and offers logical explanations for the observed results.

A. Deductive and Inductive Reasoning

Science proceeds via several trains of reasoning and thought, each appropriate to a given set of circumstances. Deduction is defined as reasoning from a general observation to a specific conclusion. For example:

1. The sparrow is a bird.
2. All birds have feathers. (premise)
3. Sparrows must have feathers.

The investigator engaging in deductive reasoning finds the conclusion by analyzing previously known facts (the premise). If the premise is true, the conclusion must be true. A false conclusion can be reached if a premise is not true, so one must be sure that the premise is true before one makes a deductive conclusion.

Induction is defined as reasoning from a specific case to the general. For example:

1. The sparrow is a bird.
2. The sparrow can fly. (premise)
3. All birds can fly.

In inductive reasoning, the premise may predict a high probability of the conclusion, but does not ensure that the conclusion is true. One must observe every possible case to know for certain that an inductive conclusion is correct. Hence, one must be cautious before making the “inductive leap” to a potentially false conclusion.
Scientific questions are answered via hypothetico-deductive reasoning: The formulation of a hypothesis (a tentative answer to a question) regarding a specific phenomenon, and the execution of rigorous, repeatable experiments from which one may deduce a general answer to the hypothesis.

B. Hypothesis, Theory and Law
Keep these definitions distinct in your mind for the rest of your life.

Hypothesis
As we've stated before, the hypothesis is an educated guess based upon observation. It is a rational explanation of a single event or phenomenon based upon what is observed, but which has not been tested or proven.

Theory
In science, the term theory has a very precise meaning that is quite different from the colloquial use of the term.

A theory is an hypothesis whose predictions have stood up to thorough and rigorous testing with experiments and observations performed by many different, objective scientists. A single scientist cannot create a theory. An hypothesis may become a theory if it is tested rigorously many, many times and is never proven incorrect.

A theory is generally accepted by the scientific community to be true, and it can be used to make predictions for the further elucidation of the phenomenon in question.

Note that neither a hypothesis nor a theory can be "proven" with the scientific method. An objective scientist refuses to let even the most powerful theory become dogma (something held as an established, unquestioned opinion). Even a very powerful scientific theory is subject to revision in light of new information.

Law
A scientific law is a statement of fact meant to explain, in concise terms, an action or set of actions. It is generally accepted by the scientific community to be universally true. A law is often expressed in terms of a single mathematical equation.

A scientific law does not provide an explanation for a phenomenon. It is accepted at face value because it has always been observed to hold true. Thus, a scientific law is simple, true, universal, and absolute. As such, it forms a cornerstone of investigation and discovery in its area, and is the basis of hypothesis formulation.

Examples of scientific laws include the Law of Gravity, the Newtonian Laws of Motion, the Laws of Thermodynamics and Boyle’s Law of Gases

C. One-and Two-tailed Hypotheses
A hypothesis is nothing more than a possible explanation of a particular phenomenon. It is based on past experience about the phenomenon under study, and can be considered an "educated guess." (emphasis on educated)

Multiple hypotheses make good science. (If you have only one possible answer, you may bias your experiment and your analysis.) Hypotheses should be testable via experimental procedures or field studies based on the hypothetico-deductive approach.

Hypotheses can be refuted (proven wrong, or falsified), but they cannot be proven to be true. It is impossible to perform enough experiments to be certain that the answer will always be the same in every case, and that the same explanation will hold true for every case of a particular phenomenon. On the other hand, if a hypothesis is shown to
be incorrect in a single case, then the hypothesis must be rejected, as it does not hold true in every case.

By stating the question as null and alternative hypotheses, the investigator allows less ambiguity in accepting or rejecting one or the other. Once the null hypothesis is rejected, the alternative hypothesis is accepted by default, although it then becomes subject to greater scrutiny and further testing. For example:

"There is no difference in the rate of germination in lettuce seeds exposed to red light and far-red light."

A second hypothesis, the alternative hypothesis ($H_a$), states the exact opposite of the null hypothesis. $H_a$ is, of course, the hypothesis of interest:

"There is a difference in the rate of germination in lettuce seeds exposed to red light and far-red light."

Note that the null hypothesis does not necessarily state that seeds exposed to red or far red light germinate at a faster or slower rate. It states only that there is no difference between the groups being compared. Such a hypothesis, which does not predict a direction in which the data might deviate from the expected (e.g., "higher" or "lower") is called a two-tailed hypothesis (it can “go either way”). Similarly, the alternate hypothesis does not state whether lettuce seeds are more or less likely to germinate under red or far red light. It simply says there is a difference between the two groups. The analyzed data will suggest the direction (i.e. "higher," "lower," "less likely," "more likely") of the alternate hypothesis, but the investigator is not making a direct prediction of this direction.

In some situations, it is of great interest to determine the direction in which observed results deviate from the expected. In this case, one should devise a one-tailed hypothesis. It is more difficult to reject a one-tailed hypothesis than a two-tailed hypothesis, as you will learn when we analyze probabilities. Statistical formulas specially designed to test one- and two-tailed hypotheses do exist, but they are beyond the scope of this appendix.

Before you begin any experiment, you must formulate a null hypothesis pertaining to your experimental groups. The hypotheses should be stated in the introduction of your scientific paper or presentation.

II. Data, parameters and statistics

Most investigations in the biological sciences today are quantitative. The investigator's goal is to collect biological observations which can be tabulated as numerical facts, also known as data (singular = datum).

A. Types of Data

Biological research can yield several different types of data:

1. Attribute data. This simplest type consists of descriptive, "either-or" measurements, and usually describe the presence or absence of a particular attribute. The presence or absence of a genetic trait ("pubescence" or "no pubescence") or the type of genetic trait (different forms of a protein receptor, for example) are examples. Because this type of data has no specific sequence, it is considered unordered data.
2. **Discrete numerical data.** These data correspond to biological observations that are counted, and are integers (whole numbers). The number of leaves on each member of a group of plants, the number of breaths per minute in a group of newborns or the number of beetles per square meter of forest floor are all examples of numerical discrete data. Although these data are ordered, they do not describe physical attributes of the things being counted.

3. **Continuous numerical data.** The most quantitative data fall along a numerical continuum. The limit of resolution of such data is the accuracy of the methods and instruments used to collect them. Examples of continuous numerical data are tail length, brain volume, percent body fat...anything that varies on a continuous scale. Rates (such as rate of transpiration over the course of an hour) are also numerical continuous data.

   The statistical test appropriate in any given situation depends upon the type of data, and this should be chosen in advance, before the experiment is performed. Be sure of the type of data you will be collecting before you begin.

   When an investigator collects *numerical* data from a group of subjects, s/he must determine how and with what frequency the data vary. For example, if one wished to study the distribution of leaf sizes of a particular species of plants, one might measure the leaf size of a sample of the plants (say, 50 individuals) and graph the numbers with "average leaf size" on the x-axis and "number of individuals" on the y-axis. The resulting figure shows the frequency distribution of the data, a representation of how often a particular data point occurs at a given measurement.

   Usually, data measurements are distributed over a range of values. Measures of the tendency of measurements to occur near the center of the range include the population mean (the average measurement), the median (the measurement located at the exact center of the range) and the mode (the most common measurement in the range).

   It is also important to understand how much variation a group of subjects exhibits around the mean. For example, if the average leaf size is 10 cm² we must determine whether leaf size forms a very wide distribution (with a relatively small number of individuals having the smallest leaves or largest leaves) or one which hovers near the mean (with a relatively large number of individuals having leaves close to the average size). Measurements of dispersion around the mean include the range, variance and standard deviation.

**B. Parameters and Statistics**

If you were able to measure the height of every adult male *Homo sapiens* who ever existed, and then calculate a mean, median, mode, range, variance and standard deviation from your measurements, those values would be known as *parameters*. They represent the actual values as calculated from measuring every member of a population of interest. Obviously, it is very difficult to obtain data from every member of a population of interest, and impossible of that population is theoretically infinite in size. However, one can estimate parameters by randomly sampling members of the population. Such an estimate, calculated from measurements of a subset of the entire population, is known as a *statistic*. 
Parameters are written as Greek symbols equivalent to the Roman symbols used to represent statistics. For example, the standard deviation for a subset of an entire population is written as "s", whereas the true population parameter is written as $\sigma$.

Statistics and statistical tests are used to test whether the results of an experiment are significantly different from what is expected. What is meant by "significant" or "expected" results? The answers lie in understanding probability.

### III. Statistical tests

Let's return to our germinating seeds. After the data have been collected (Table 1), the investigators can determine whether any difference in germination rate between the two groups is significant or simply due to random chance. To do so, the investigators must perform a statistical test on the data collected. The results of this test will enable them to either ACCEPT or REJECT the null hypothesis.

#### A. Descriptive Statistics: mean, variance and standard deviation

You probably will be dealing most often with numerical continuous data, and so should be familiar with the definitions and abbreviations of several important quantities:

- $x = \text{data point}$ the individual values of a measured parameter ($=x_i$)
- $\overline{x} = \text{mean}$ the average value of a measured parameter
- $n = \text{sample size}$ the number of individuals in a particular test group
- $df = \text{degrees of freedom}$ the number of independent quantities in a system
- $s^2 = \text{variance}$ a measure of individual data points' variability from the mean
- $s = \text{standard deviation}$ the positive square root of the variance

**Table 1. Germination in lettuce seeds exposed to either red or far red light. Each sample consists of a dish with 100 seeds, and data are tabulated as % germination.**

<table>
<thead>
<tr>
<th>Sample ID#</th>
<th>% germination</th>
<th>Sample ID#</th>
<th>% germination</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-1</td>
<td>80</td>
<td>FR-1</td>
<td>21</td>
</tr>
<tr>
<td>R-2</td>
<td>78</td>
<td>FR-2</td>
<td>15</td>
</tr>
<tr>
<td>R-3</td>
<td>76</td>
<td>FR-3</td>
<td>10</td>
</tr>
<tr>
<td>R-4</td>
<td>90</td>
<td>FR-4</td>
<td>11</td>
</tr>
<tr>
<td>R-5</td>
<td>88</td>
<td>FR-5</td>
<td>3</td>
</tr>
<tr>
<td>R-6</td>
<td>92</td>
<td>FR-6</td>
<td>9</td>
</tr>
<tr>
<td>R-7</td>
<td>74</td>
<td>FR-7</td>
<td>15</td>
</tr>
<tr>
<td>R-8</td>
<td>85</td>
<td>FR-8</td>
<td>15</td>
</tr>
<tr>
<td>R-9</td>
<td>90</td>
<td>FR-9</td>
<td>18</td>
</tr>
<tr>
<td>R-10</td>
<td>77</td>
<td>FR-10</td>
<td>22</td>
</tr>
<tr>
<td>N</td>
<td>10</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>MEAN</td>
<td>83</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Variance</td>
<td>45.33</td>
<td></td>
<td>33.6</td>
</tr>
<tr>
<td>Standard dev.</td>
<td>6.7</td>
<td></td>
<td>5.8</td>
</tr>
</tbody>
</table>

To calculate the mean germination rate of either group, the investigators count the number of germinated seeds and un-germinated seeds in each sample dish, and then determine the mean rate of germination (in this case, a percentage) for each treatment. 9 25 25 49 25 81
Thus calculated, the mean germination rate of the red light exposed seeds is 83, and of the treatment group, 13 (13%) (Table 1).

To determine the degree of the individual samples’ deviation from the mean germination rate, the investigators calculate several quantities. The first is the sum of squares (SS) of the deviations from the mean, defined as:

\[ SS = \sum (x - \bar{x})^2 \]

Whenever there is more than one test group, statistics referring to each test group are given a subscript as a label. In our example, we will designate any statistic from the Red Light group with a subscript "R" and any statistic from the Far Red Light group with a subscript "FR." Thus, sum of squares of our control group (SS_R) is equal to 69298 and SS_FR is equal to 2235 (Table 2).

The variance \( (s^2) \) of the data, the mean SS of each test group, is defined as:

\[ s^2 = \frac{\sum (\bar{x} - x)^2}{n-1} \]

Calculate the variance for both the Red and Far Red groups. Check your answers against the correct ones listed Table 2.

**Standard deviation (s), the square root of the variance:**

\[ s = \sqrt{\frac{\sum (\bar{x} - x)^2}{n-1}} \]

Calculate the standard deviation for the Red and Far Red groups. Check your answers against the correct ones listed in Table 2.


Inferential statistics are calculated to make logical conclusions about the numerical characteristics of population under study, including accounting for random variation. In the natural sciences, such calculations often are done with algorithms (statistical tests) that generate a statistic representing the population.

**1. Parametric tests**

A parametric test is used to test the significance of continuous numerical data (e.g. - lizard tail length, change in weight, reaction rate, etc.). Examples of commonly used parametric tests are the Student's t-test (which generates a statistic that compares
the difference between two means) and the ANOVA (which generates a statistic that compares the differences between multiple means.

**The Student's t test – A parametric test**

The Student t-test can be used to determine whether a difference between two means is significant. These means may calculated from observations that are either **paired** (as when individuals in a single group are subjected to "before and after" measurements, and data points are paired for each tested individual) or **independent** (as when individuals in two similar sample populations are measured, but each individual in each sample population is measured only once). Slightly different calculations of the t-statistic must be used in each case.

**a. Independent Sample t-test**

In the case of our germination experiment, a rate (number of seeds germinated per 100) was generated for each of the two groups being compared. But because each seed was exposed only to red light or far red light—not to sequential stimuli—the members of the two experimental groups are different in possibly subtle ways. Their means must be analyzed with the independent sample. None of the individual seeds were subjected to paired measurements, which would mean that the same seed was used in both Red and Far Red treatments. Since each seed germinated in your experiment, it was not possible to subject them to a paired test. Therefore, we will use a statistical test designed to show whether there is a significant difference between the means of two **independent** samples, those of your (1) Red and (2) Far Red groups.

Use the following equation to calculate a t-statistic for your two means:

\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{s_p^2(1/n_1 + 1/n_2)}}
\]

...where \(x_1\) and \(x_2\) are the means of your two groups, \(n_1\) and \(n_2\) are the numbers of seeds you used in each group, and \(s_p^2\) is the **pooled variance**, calculated as:

\[
s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}
\]

...where \(s_1^2\) is the variance of group 1, \(s_2^2\) is the variance of group 1, \(df_1\) is the degrees of freedom for group 1 \((df_1 = n_1 - 1)\) and \(df_2\) is the degrees of freedom for group 2 \((df_2 = n_1 - 1)\)

The **variance** \((s^2)\) for a mean is calculated as
\[ s^2 = \frac{\sum (\bar{x} - x)^2}{n-1} \]

In which \( \bar{x} \) is the mean, \( x \) is each individual value, and \( n \) is the sample size.

Finally, determine the number of independent quantities in your system, or **degrees of freedom**. The degrees of freedom determine the significance level tied to every possible value of a statistic (such as the t-statistic). The degrees of freedom is the number of data points that are free to vary without changing the test statistic, and this changes depending on the type of statistic you are calculating. The degrees of freedom for a two-sample t-test with independent means is calculated as the sum of the degrees of freedom of each test group:

\[ df = (n_1 - 1) + (n_2 - 1) \]

b. **Paired Sample t-test**

In some cases, an experiment will subject the same sets of individuals to sequential variables. For example, if you were to grow a set of 30 bean seedlings under blue light, measure their growth rate, and then switch them to red light and again measure their growth rate, you would essentially be performing "before" and "after" measurements of a single plant responding to two different variables. This means that the two values you attain will not be independent of one another, and are said to be **paired**. (As a note, the trial described here would have to be reciprocated by first growing seedlings under red and then blue light, to make certain that any differences in growth were not simply due to the life stage of the plants.)

In a paired sample t-test, the separate means of two sample populations is not measured. Instead, the difference between the first measurement and the second measurement on the same individual is determined and used to calculate the t-statistic. The difference between the first and second measurements is represented as \( d \), and the mean of all your individual differences as \( \bar{d} \).

The standard deviation and variance are calculated as before, but substituting \( d \) and \( \bar{d} \) for \( x \) and \( \bar{x} \) in the equations used in the independent sample test. Thus, to calculate the **variance** of paired samples, use the following equation:

\[ s^2_{d} = \frac{\sum (\bar{d} - d)^2}{n-1} \]

To calculate the **standard deviation**, take the square root of the variance.

Recall that degrees of freedom is the number of data points that are free to vary without changing the test statistic, and this changes depending on the type of statistic you are calculating. For the paired sample t-test,

\[ df = (n - 1) \]
...where \( n \) is the number of independent quantities in your system. For paired samples, \( n = \) the number of pairs in your system. In this case, \( n = 30 \).

Use the following equation for the paired sample t-test:

\[
t = \frac{( \bar{d} - 0 ) \sqrt{n}}{s_d}
\]

2. Non-parametric tests

A non-parametric test is used to test the significance of qualitative data (e.g. numbers of purple versus yellow corn kernels, presence or absence of freckles in members of a population etc.). Both attribute data and discrete numerical data can be analyzed with non-parametric tests such as the Chi-square and Mann-Whitney U test. Although these tests are often simpler to perform, they are not as powerful as parametric tests. In other words, non-parametric tests less able than parametric tests to accurately predict whether unexpected results are due to random chance.

Chi-square: A non-parametric test

A commonly used non-parametric inferential statistic is the Chi square \( (\chi^2) \). Although this test has several complex permutations, we will use only the simplest formula, but you can also use it to test a wide variety of attribute or discrete data. The formula for calculating the Chi square statistic is as follows:

\[
\chi^2 = \sum \frac{(O - E)^2}{E}
\]

In which:
- \( O \) = the observed results
- \( E \) = the expected results
- \( \Sigma \) means the summation of \( \chi^2 \) values over every phenotypic category

In the Chi square test, \( n \) has a slightly different meaning than it has in parametric tests. In this case, \( n \) is the total number of categories possible. For example, if you are counting purple and yellow corn kernels, \( n = 2 \) (purple and yellow). If you are counting expression of two phenotypes, such as red versus yellow flower petals and curly versus straight stamens, \( n = 4 \) (red curly, red straight, yellow curly and yellow straight).

The degrees of freedom (df) in this Chi square test is equal to \( n-1 \).

V. Probability and significance

The term "significant" is often used in everyday conversation, yet few people know the statistical meaning of the word. In scientific endeavors, significance has a highly specific and important definition. The difference between an observed and expected result is said to be statistically significant if and only if:

Under the assumption that there is no true difference, the probability that the observed difference would be at least as large as that actually seen is less than or equal to 5% (0.05).
Conversely, under the assumption that there is no true difference, the probability that the observed difference would be smaller than that actually seen is greater than 95% (0.95).

Once an investigator has calculated a Chi-square or t-statistic, s/he must be able to draw conclusions from it. How does one determine whether deviations from the expected (null hypothesis) are significant?

As mentioned previously, depending upon the degrees of freedom, there is a specific probability value linked to every possible value of any statistic.

A. Determining the significance level of a parametric statistic

In a sample where two different groups of individuals are subjected to different variables, one should use an independent sample t-test. (An experiment in which the same group of individuals is subjected to two different treatments should be analyzed with a paired t-test.) Values for the red and far red lettuce seeds appear in Table 2.

The next step is to interpret what our t-statistic tells us about the difference in mean germination rate between the groups. Is the difference significant? Or is it simply due to random chance? The answer lies in the table of critical values for the t-statistic, part of which is illustrated in Table 3.

Table 2. Statistics for Lettuce germination under red and far red light.

<table>
<thead>
<tr>
<th>statistic</th>
<th>red light</th>
<th>far red light</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean (x)</td>
<td>83</td>
<td>13.9</td>
</tr>
<tr>
<td>variance (s^2)</td>
<td>45.3333</td>
<td>33.6556</td>
</tr>
<tr>
<td>standard deviation (s)</td>
<td>6.7330</td>
<td>5.8013</td>
</tr>
<tr>
<td>overall statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>24.58</td>
<td></td>
</tr>
<tr>
<td>degrees of freedom (df)</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>P value (significance)</td>
<td>&lt;0.0001</td>
<td>HIGHLY SIGNIFICANT</td>
</tr>
</tbody>
</table>

To determine whether the t-statistic indicates rejection of the null hypothesis, do the following:

1. Locate the appropriate degrees of freedom in the far left (yellow) column of Table 3..
2. Look across the df row (df=18) to find a t value closest to the one we obtained.
3. If the exact value does not appear on the table, note the two t values which most closely border our value.
4. Find the P values corresponding to the two bordering values. Our P value lies between them.

____ > P > _____

At 18 degrees of freedom, our seed germination t statistic (24.58) lies far to the right of the last value on the table on that row (3.922) on the table of critical t-statistic values. Thus, the probability that the difference in germination rate between the Red and Far Red groups is due to chance is much less than 0.0001 (1/1000%). This is highly significant, meaning that there is a vanishingly small chance that the difference is due to random chance. Hence, the only variable (the difference in light wavelength) apparently affects germination rate. We can reject our original two-tailed hypothesis and accept the alternate hypothesis:
"There is a significant difference in the rate of germination between seeds exposed to red versus far red light."

Table 3. Partial table of critical values for the independent-sample t-test (from Zar, 1974). The first row of P values is used for a two-tailed alternative hypothesis (i.e., one which does not specify the direction (lower or higher germination rate) of the alternative hypothesis). The second row of P values should be used for a one-tailed hypothesis (i.e., one which does specify the direction of the alternative hypothesis). P values lower than 0.05 (shaded in blue for each hypothesis type) indicate 5% or lower probability that deviation from the null predictions are due to chance, and the null hypothesis should be rejected. Degrees of freedom are shaded in yellow.

<table>
<thead>
<tr>
<th>P (2) =</th>
<th>0.50</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (1) =</td>
<td>0.25</td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
<td>0.005</td>
<td>0.0025</td>
<td>0.001</td>
<td>0.0005</td>
</tr>
<tr>
<td>df</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.765</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
<td>10.215</td>
<td>12.924</td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
<td>7.173</td>
<td>8.610</td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>1.478</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>6.869</td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
<td>5.208</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>1.415</td>
<td>1.989</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.408</td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>1.397</td>
<td>1.895</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.781</td>
</tr>
<tr>
<td>10</td>
<td>0.703</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
<td>4.144</td>
<td>4.587</td>
</tr>
<tr>
<td>11</td>
<td>0.690</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
<td>4.025</td>
<td>4.437</td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
<td>3.930</td>
<td>4.318</td>
</tr>
<tr>
<td>13</td>
<td>0.694</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.372</td>
<td>3.852</td>
<td>4.221</td>
</tr>
<tr>
<td>14</td>
<td>0.692</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
<td>3.787</td>
<td>4.140</td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
<td>3.733</td>
<td>4.073</td>
</tr>
<tr>
<td>16</td>
<td>0.680</td>
<td>1.337</td>
<td>1.746</td>
<td>2.120</td>
<td>2.583</td>
<td>2.821</td>
<td>3.252</td>
<td>3.686</td>
<td>4.015</td>
</tr>
<tr>
<td>17</td>
<td>0.689</td>
<td>1.333</td>
<td>1.740</td>
<td>2.110</td>
<td>2.567</td>
<td>2.898</td>
<td>3.222</td>
<td>3.646</td>
<td>3.965</td>
</tr>
<tr>
<td>18</td>
<td>0.688</td>
<td>1.330</td>
<td>1.734</td>
<td>2.101</td>
<td>2.552</td>
<td>2.878</td>
<td>3.197</td>
<td>3.610</td>
<td>3.922</td>
</tr>
<tr>
<td>19</td>
<td>0.688</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.539</td>
<td>2.861</td>
<td>3.174</td>
<td>3.579</td>
<td>3.883</td>
</tr>
<tr>
<td>20</td>
<td>0.687</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.528</td>
<td>2.845</td>
<td>3.153</td>
<td>3.552</td>
<td>3.850</td>
</tr>
<tr>
<td>21</td>
<td>0.686</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.518</td>
<td>2.831</td>
<td>3.135</td>
<td>3.527</td>
<td>3.819</td>
</tr>
<tr>
<td>22</td>
<td>0.686</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td>2.819</td>
<td>3.119</td>
<td>3.505</td>
<td>3.792</td>
</tr>
<tr>
<td>23</td>
<td>0.685</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.500</td>
<td>2.707</td>
<td>3.104</td>
<td>3.485</td>
<td>3.768</td>
</tr>
<tr>
<td>24</td>
<td>0.685</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
<td>3.091</td>
<td>3.467</td>
<td>3.745</td>
</tr>
<tr>
<td>25</td>
<td>0.684</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td>2.787</td>
<td>3.078</td>
<td>3.450</td>
<td>3.725</td>
</tr>
<tr>
<td>26</td>
<td>0.684</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
<td>3.067</td>
<td>3.435</td>
<td>3.707</td>
</tr>
<tr>
<td>27</td>
<td>0.684</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.057</td>
<td>3.421</td>
<td>3.690</td>
</tr>
<tr>
<td>28</td>
<td>0.683</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.047</td>
<td>3.408</td>
<td>3.674</td>
</tr>
<tr>
<td>29</td>
<td>0.683</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.452</td>
<td>2.756</td>
<td>3.038</td>
<td>3.396</td>
<td>3.659</td>
</tr>
<tr>
<td>30</td>
<td>0.683</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.445</td>
<td>2.750</td>
<td>3.030</td>
<td>3.385</td>
<td>3.646</td>
</tr>
</tbody>
</table>

Notice that the t-value calculated for our seed germination data indicates rejection of even one-tailed hypothesis. However, because all honest researchers state their hypotheses before they see their results, Team Germination should stick by their original hypothesis and let the direction of the data speak for itself.

Remember that you must have a representative sample of the population—not a single experimental run—in order to perform the t-test (A single experiment cannot have a mean, variance or standard deviation.). And the larger your sample size, the closer your P value will be to that of the actual, ideal (entire) population.
B. Determining significance level of a non-parametric statistic

First let us determine the probability value for our non parametric test, the Chi square. In this semester’s laboratory on Mendelian Genetics, you will use the Chi Square to determine whether the proportion of physical types of offspring (purple or yellow corn kernels) in a single cohort is different from the expected. In the example presented in the chapter, data yield a $\chi^2$ value equal to 1.333. Because there are two independent categories (purple and yellow), $df = 2-1 = 1$.

1. In the far left column of Table A1-4, locate the appropriate df.
2. Go across the appropriate df row, and locate the Chi square value closest to the one we obtained with the example data. As you can see, 1.333 is not listed on the table. Rather, it lies between two values listed on the table, 1.323 and 2.706.
3. Go to the top row above each of the Chi square values bordering our example value. Above each is listed a corresponding probability (P) value.
4. The P value corresponding to 1.323 is 0.25; this means that a Chi square value of 1.323 indicates a 25% possibility that the deviation from the expected is due to chance. Thus, there is only a 75% chance that these deviations are due to some factor other than chance.
5. The P value corresponding to 2.706 is 0.10; this means that a Chi square value of 2.706 indicates a 10% probability that the deviation from the expected is due to chance, and a 90% probability that the deviation is due to some factor other than chance.
6. The probability value of our example Chi square lies between 0.25 and 0.10. This is most often expressed as

$$0.25 > P > 0.10$$

This P value is outside the accepted standards for statistical significance. The null hypothesis (the observed ratio of purple to yellow corn kernels will not differ from those predicted by Mendel’s Laws) cannot be rejected.

| Table A1-4. A partial table of the probability values for the Chi square statistic. |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| P = 0.999 0.995 0.990 0.975 0.950 0.900 0.750 0.50 0.25 0.10 | 0.05 0.02 0.01 0.005 0.001 |
| df | 1 | 0.000 0.000 0.000 0.001 0.004 0.016 0.102 0.455 1.323 2.706 | 3.841 5.024 6.635 7.879 10.82 |
| 2 | 0.002 0.010 0.020 0.051 0.103 0.211 0.575 1.386 2.773 4.605 | 5.991 7.378 9.210 10.59 13.82 |
| 3 | 0.024 0.072 0.115 0.216 0.352 0.584 1.213 2.366 4.108 6.251 | 7.815 9.348 11.35 12.84 16.27 |
| 4 | 0.091 0.207 0.297 0.484 0.711 1.064 1.923 3.357 5.385 7.779 | 9.488 11.14 13.27 14.86 18.47 |
| 5 | 0.210 0.412 0.554 0.831 1.145 1.610 2.675 4.351 6.626 9.236 | 11.07 12.83 15.09 16.75 20.52 |

VI. Sample Size

To test the null hypothesis, the investigators design an experiment. For the seed hypotheses above, two groups, Red Light and Far Red Light, are created.

In a properly designed experiment, the two groups being compared must be subjected to exactly the same physical conditions with the exception of a single variable. The only difference between the groups is a single variable, in this case, the different wavelengths of light. Such rigor reduces the influence of confounding effects, uncontrolled differences between the two groups that could affect the results.

Because Team Germination cannot control for the possible confounding effect of genetic differences in the seeds, they must try to reduce the influence of that effect by
using a **large sample size**—as many experimental subjects as possible—so that the influence of genetic differences will be minimized in a largely genetically similar sample. **It is a general rule that the larger the sample size, the closer the approximation of the statistic to the actual parameter.** Even so, it is never wise to completely ignore the possibility of confounding effects. Honest investigators should mention them when reporting their findings.

**Literature Cited**