# WORKING WITH STATISTICS

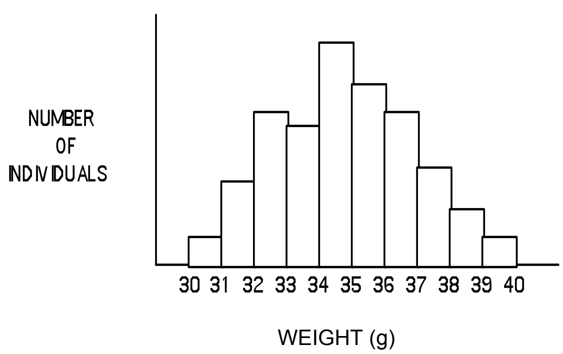
(Reviewed Spring 2010)

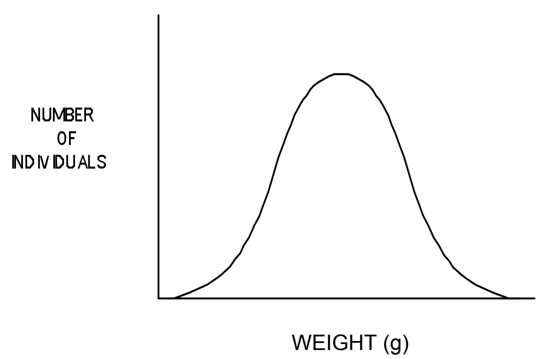
## BIOLOGICAL VARIATION

**Variation** is a noticeable characteristic of living organisms. This is true even among individuals of a single species or other closely related biological group. For example, all humans share a very similar general morphology, yet still exhibit obvious differences in height, pigmentation, voice, and numerous other traits.

As such, a scientist collecting information about such a group is faced with the challenge of making appropriate general statements about a group or population consisting of individuals that vary from one another. Suppose you wanted to learn something about the weight of a population of adult male robins. Since measuring all the individuals in a population is usually impractical, you will probably capture and weigh an unbiased **representative sample** of, say, 1000 birds. Having done so, you should have a good *general* idea about the weight of the robins in the population, but will almost certainly find that there is some degree of variation among the individuals in the sample.

Wanting to convey what you have learned, you might be inclined to simply list the weights of all 1000 birds. Such an approach, however, would be cumbersome and would fail to convey much useful information. A more concise and meaningful description can be achieved by plotting your sample data in a **histogram** or **frequency distribution.** (Note that if the size of the categories was made small enough, our frequency distribution would approximate a smooth curve!)




A frequency distribution provides a pictorial description of our sample population. However, since we probably want to make comparisons of frequency distributions from different populations, having a numerical description of the major features of the distribution would be more useful. We accomplish this by calculating and reporting **descriptive statistics**.

## DESCRIPTIVE STATISTICS

There are two general categories of descriptive statistics. **Measures of location** reflect the "central tendency" of the population, and often represent the number around which the majority of individuals are found. **Measures of variability** reflect the "spread" of the measurements of the sample population.

**Common measures of location include:**

- **Mean**. The mean (or average) is calculated by dividing the sum of all sample measurements by the sample size (n).

- **Median**. The median is the sample measurement within the distribution that has an equal number of measurements on either side of it.

Although the mean is most commonly utilized, it becomes less meaningful when describing a population that does not approximate a **normal distribution** (see Appendix A). When sample measurements are heavily skewed to one side of a distribution, the median often becomes a more useful descriptor. (The **mode**—the sample measurement that occurs most frequently in the distribution—is another measure of location that is infrequently used.)

**Common measures of variability include:**

- **Range**. The range is the difference between the largest and smallest measurements in a sample. When presenting the range, the largest and smallest values are usually indicated.

- **Variance**. Variance ($s^2$) is a measure of the difference between each individual in the sample and the mean of the sample. It is calculated as:

$$\text{variance = the sum of } \frac{(\text{each item - mean})^2}{\text{number of items minus one}} \quad \text{--OR--} \quad s^2 = \sum_{i=1}^{n} \frac{(y_i - y)^2}{n - 1}$$

- **Standard Deviation**. Standard deviation (s or SD) represents the “average” deviation from the mean. It is calculated by taking the square root of the variance.

- **Standard Error**. Standard error (SE) represents the “uncertainty” in the mean of a sample population. It is calculated by dividing the standard deviation (s) by the square root of the sample size (n).

$$\text{standard deviation (s)} = \sqrt{\text{variance}} \qquad \text{standard error (SE)} = \frac{s}{\sqrt{n}}$$

Standard deviation and standard error are most commonly reported as values extending on either side of the mean. Standard deviation is usually used to present the variability around the mean when the *entire* population is known, while standard error is employed when a *sample* of the entire population is being analyzed. (As you will see later, standard error is also a useful tool for making quick statistical comparisons among populations.)

**Reporting Descriptive Statistics**

Depending upon what is appropriate for your particular study, descriptive statistics may reported in the text, in a table, or in a figure. In all cases, it is important to clearly indicate *which* descriptive statistics you are presenting. Consider the following statement; it is short and to the point, but makes it clear to the reader that the values presented represent the mean and standard error, as well as indicating the sample size.

> “The mean (+/- SE) weight of male robins in the sample population was 81 +/- 5.1 g (n=1000).”

In *figures,* measures of variability are usually expressed as *error bars* extending above and below the measure of location. In the following example, the weights of male robins in four different habitats are compared. The large bars represent the means while the smaller error bars represent the standard errors. Note that the figure legend clearly indicates which statistics are being presented, as well as the sample size for each population (Figure 1).

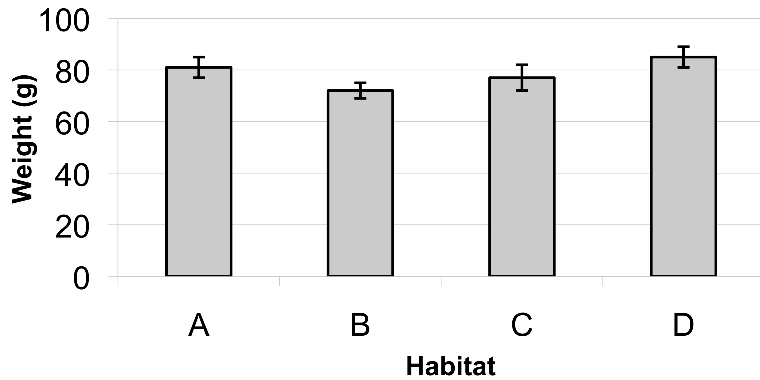



**Figure 1. Mean (+/- SE) weight of male robins in four different habitats. ($n_A$=823; $n_B$=903; $n_C$=877; $n_D$=789)**

## COMPARATIVE STATISTICS

Having described our population numerically, it is likely that we will want to make objective, quantitative comparisons with other populations. For example, we may want to compare the weight of our male robins to that of female robins (or to other male robins in a different region or year). Here is where things get a bit more challenging.

Suppose all our male robins weighed 85g, and all our female robins weighed 75g. In such a circumstance our task would be easy—we would conclude that males weigh more than females. However, since natural variation is evident in all populations, such unambiguous results are seldom seen. A more realistic result would be that each individual bird had a different weight, and that the range of male robin weights overlapped the range of female robin weights. The *average* (mean) weight of the two groups might be different, but, because the ranges overlap, we may not be able to decide objectively if the weights of male and females are *really* different, or if the *observed* differences (in the means) are simply the result of *natural variation*.

Formulating conclusions based solely on comparing the means of two samples is a mistake. Although the means may differ, the observed differences may be the result of natural chance variation. In other words, the samples are *statistically* the same--the observed differences are said to be *not significant.*

One way to *estimate* whether the differences between sample groups are significant is to compare their standard errors. Recall that standard error is based on both variability and sample size--it is expressed as a value bracketing either side of the mean. Examine Figure 2. Note that the *lowest* value indicated by the error bars in the ‘male’ sample overlaps the *highest* value indicated by the error bars in the ‘female’ sample. If the standard errors of two samples overlap, it is generally assumed that the two samples are *not significantly different*. Now examine Figure 3. Note that in this case, the standard errors of the ‘male’ and ‘female’ samples DO NOT overlap. If the standard errors of two samples do not overlap, the two samples are assumed to be *significantly different.*

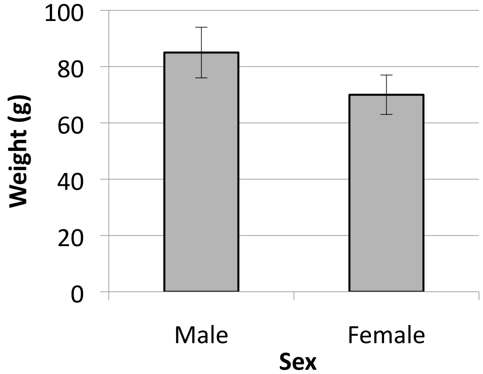



**Figure 2. Mean (+/- SE) weight of male vs. female robins. ($n_m$=719; $n_f$=739)**

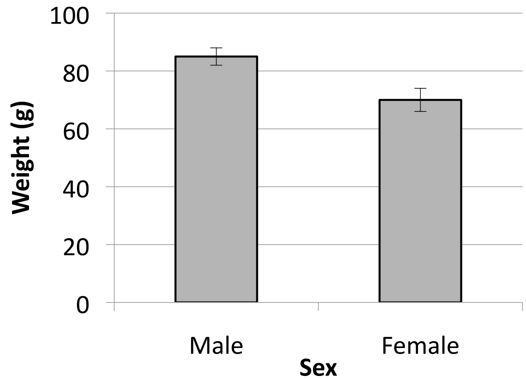



**Figure 3. Mean (+/- SE) weight of male vs. female robins. ($n_m$=719; $n_f$=739)**

Note that the *means* reported in Figure 2 vs. Figure 3 are the same; only the *standard errors* differ. However, you would likely draw a very different conclusion from the information shown in one figure vs. that shown in the other!

Comparative (or differential) statistics provide a more reliable method of determining whether observed differences are significant. These mathematical tests make comparisons among sample groups by taking into account both the mean (a measure of central tendency) AND the variation of each group. Comparative statistical tests are used to calculate a *p-value*. Without belaboring the point, the p-value represents the *probability* that differences observed among groups are the result of chance alone. As a convention, 0.05 is considered the "cutoff" for determining whether two compared groups are significantly different or not. If $p>0.05$, the observed differences are said to be *not significant* (assumed to be from chance alone). Conversely, if $p<0.05$ (or =0.05), the observed differences are said to be *significant* (assumed to be from a treatment or some other experimental condition).

The *Student's t-test* is one statistical test commonly used to compare two sample groups. (Microsoft Excel is capable of performing t-tests quickly and easily!) One powerful feature of the t-test is that it may be used to compare *unpaired* data (e.g., the resting heart rates of two different groups of people such as athletes and non-athletes), or *paired* data (e.g., the heart rate of a single group of people before and after some experimental treatment such as vigorous exercise). Applying the appropriate version of the test is critically important! (Numerous other statistical tests exist, but a comprehensive accounting of them is outside the scope of this document. A course devoted to statistics is recommended for those wishing to understand both the mechanics and the proper use of such tests.)

It should be evident by now that the term "significant"—when used in the context of a statistical comparison—has a very specific meaning. As scientists, avoid using this term in a casual manner! Also, keep in mind that saying that two samples are not *significantly* different is essentially the same as saying they are not different. You MUST take the results of your statistical comparisons into account when formulating your conclusions!

**Reporting Comparative Statistics**

The results of comparative statistics are usually reported with a *statement* indicating whether the compared groups are or are not significantly different. Although formats for specific statistical tests vary, it is always important to indicate (If applicable) the test applied and the p-value obtained. For example:

> "Male robins in habitat A were found to be significantly larger than those in habitat B (t-test, $p<0.05$)."
>
> "There was no significant difference in the weight of male robins as compared to female robins (t-test, $p=0.45$)."

You will note that it is NOT necessary to explain *how* you interpreted the p-value. This something you may assume your reader already understands.

## SIGNIFICANT FIGURES

(Not to be confused with *significant differences* described above!)

When making quantitative measurements and performing statistical analysis (particularly with a computer or calculator), careful attention to the number of **significant figures** reported is critical to ensure that you do not imply a degree of precision greater than that which you are able to observe. Significant figures are defined as *the necessary number of figures required to express the result of a measurement so that only the last digit in the number is in doubt*.

**Significant Figures in Measurements**

Suppose you are out measuring pine needles with a ruler that is calibrated only in centimeters. If you find that a pine needle is between 9 and 10 cm long, it would be appropriate for you to *estimate* the additional fraction of a centimeter (in tenths of a centimeter by convention) and add it to 9 cm. Reporting your measurement as, for example, 9.3 cm indicates that the final digit is in doubt. However, be aware that reporting your measurement as 9.30 cm would be in error--you would be implying a level of precision that is beyond that of your measuring device.

**Doing Arithmetic with Significant Figures**

Careful attention to significant figures in your measurements *must* be carried through to your statistical calculations. Failure to do so will make your original measurements appear more precise than they are. Keep these few rules in mind when working with data:

- When performing *multiplication* or *division*, the final answer should be expressed with the precision of the number in the calculation that shows the *least number of significant figures*. For example, if you wish to calculate the weight of 10.1 ml of water and you are told the density of water is 0.9976 g/ml, you would multiply the density times the volume to obtain the weight. However, the correct answer would be 10.1 g, not 10.07576 g. Any answer with more three significant figures conveys a precision that is not justified given the uncertainty of the water volume measurement.

- When performing *addition* or *subtraction*, the answer should contain no more decimal places than the number with the *least number of digits following the decimal place*. For example, 7.2$^{o}$ C subtracted from 7.663$^{o}$ C yields a correct answer of 0.5$^{o}$ C, not 0.463$^{o}$ C. (If the first number had been known with a precision of 7.200$^{o}$ C, then the latter answer would have been correct.)

- When converting measurements from one set of metric units to another, take care to avoid introducing additional significant figures. 4.3 cm is equal to 43 mm, not 43.0 mm. Similarly, 1.63 m could be expressed as $1.63 \times 10^3$ mm (three significant figures), but not 1630 mm (four significant figures).

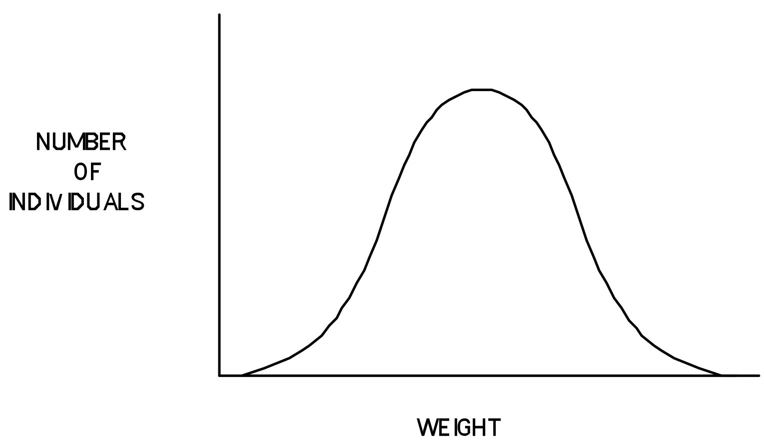
- When performing calculations, only the *final* answer should be reduced to the appropriate number of significant figures. When doing so, be sure to *round* the final digit properly!

**APPENDIX A: THE NORMAL DISTRIBUTION**

Frequency distributions can, of course, have many different shapes. However, many variable biological characteristics are commonly described by a **normal distribution**. Normal distributions exhibit three general properties:

1. Most observations occur near the mean (at the center of the distribution).
2. Fewer observations occur as you get further from the mean (toward the "tails" of the distribution).
3. Observations have an equal probability of being larger or smaller than the mean (the distribution is symmetrical around the mean).

For obvious reason, the normal distribution is sometimes referred to as a "bell curve:"




Theoretical reasons exist to explain the normal distribution of many biological characteristics. Simply put, the value of any variable depends on many different independent factors. These factors may have a large or small effect and may increase or decrease the size of the variable. These factors tend to cancel each other out so that most variables are near the mean. Therefore, the probability of a value being very different from the mean becomes very small, and the probability of the combined factors having a positive effect (causing the variable to have a value larger than the mean) is equal to the probability of the sum of factors having a negative effect (producing a variable smaller than the mean).