PART I: INTRODUCTION TO STATISTICS

Purpose of Statistics

What is the meaning of variation in nature? Plato argued that every entity in nature has its ideal form, and that variation we observe is an illusion that distracts us from what is really significant about the natural world. Today we know otherwise. Biologists are as interested in variation as they are in the average.

Consider this problem: We want to measure the body size of some kind of lizard. We collect a sample of that species from a local area and measure the length from tip of nose to the anal opening. We are careful to collect only adult males of the same age. If we measure the same lizard ten times it is likely that we will end up with some variation in our data simply because of measurement error. Then, if we measure ten individuals from the same population we will certainly also have variation in the data set, perhaps a result of differing growth rates of the animals. Last, we might gather another sample from a site 100 km away and note that there seems to be variation between the two samples (perhaps the average growth rate at one population is greater than the other because food for the lizards is more common there).

Thus, we have three kinds of variation to consider:

1– variation due to measurement error. This is a problem of technique and usually not of much biological interest.

2 – we have variation *within* a population. We want to *describe* this variation and eventually account for it.

3 – we have variation between the two populations. We would want to *describe* this variation and somehow determine if the populations are really different, and if so, why?

The study of statistics is an area of applied mathematics that allows us to *describe* variation in populations and to make some *inferences* about that population based on a *sample* drawn from the population. (The dictionary definition of "to infer" is "to derive as a consequence, conclusion, or probability.") Also, we can *compare* two or more populations. Our goal, then, is *to understand the origin of variation in nature.*

Sampling From a Large Population

Today most natural and social scientists use statistical methods to help in testing hypotheses. Statistics, however, was invented by biologists (especially geneticists) who discovered that to understand nature requires an understanding of variation in nature. Today it is not possible to study ecology, evolution, or animal behavior without recourse to statistical methods. Indeed, throughout Biology 102 you will be examining data and doing statistical tests. Suppose wildlife biologists wanted to implement a management scheme for alligators in Florida, but if they captured every one of the creatures for a census (even if this were possible), there would be no wild alligators to manage. Ecologists must rely heavily on statistical methods because often the population of interest is ether large or widespread, making sampling necessary.

Whenever we sample from a large population we must face a very real problem: samples may not represent the real nature of the population. For example, consider this thought experiment: you flip a coin (assume it is a "fair" coin that delivers half heads and half tails when it is flipped) ten times. Suppose you get 7 heads and 3 tails (we will call this 7H:3T). If you flipped this fair coin an infinite number of times you should get 50% heads and 50% tails. Why did you get 7H:3T? There are two possible reasons (remember that we have eliminated in our thought experiment the possibility that the coin is not fair). First, random sampling error was the cause; just by chance you got 7 heads. The issue of random sampling error is central to statistical testing, and is incorporated into every statistical test you will perform in this course. Second, there was biased sampling. Perhaps you tossed the coin in such a way to favor heads. There is no way to incorporate sampling bias into statistical testing, so you must be very careful when gathering data not to introduce any bias (even unconscious) into the sampling!

Now that you have your data on a sample drawn from the large population you can calculate several descriptive statistics. The first type are *sample parameters* which are simply *descriptions* of the sample. It is very important to distinguish *sample parameters* from *population parameters* which are the real “states of nature.” For example, your population measures for the coin toss experiment are 70% heads and 30% tails, whereas the population measures would be 50% and 50%.

By convention, population parameters are designated by Greek letters and sample parameters by Roman letters. For example, the arithmetic mean is indicated with a μ for a population and $\bar{x}$ for a sample, and the standard deviation (a measure of degree of variation in the population or sample) is indicated by σ and the sample by s.

Why do we need to calculate sample parameters? We hope that the sample parameters reflect the population measurements. That is, if we calculate the mean (a measure of “average”) and the standard deviation (a measure of amount of variation) for the sample, we hope that we can infer the mean and standard deviation of the entire population

Probability Theory and Statistics

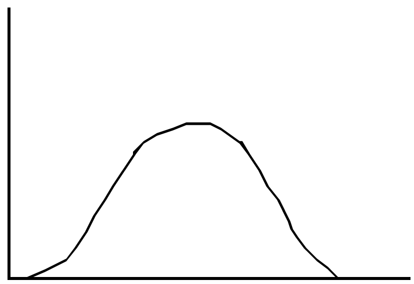
The laws of probability were worked out in the 18th and 19th century by European gamblers who were interested in understanding how the “odds” worked in games of chance. We assumed earlier that a “fair” coin when flipped will come up 50% heads if we flipped the coin an infinite number of times. This assumption comes very easily to most people; it seems obvious. But, why *should* this happen? The laws of probability describe some of the most profound laws of nature, laws that have intrigued philosophers for generations. For our purposes it is important to remember that statistical analysis and tests depend completely on knowledge of the laws of probability.

For example, suppose we flip a coin ten times. There are 11 possible outcomes. The *binomial distribution* describes the probability of each outcome:

Outcome	Probability
10H:0T	0.00098
9H:1T	0.00977
8H:2T	0.04395
7H:3T	0.11719
6H:4T	0.20508
5H:5T	0.24609
4H:6T	________
3H:7T	________
2H:8T	________
1H:9T	________
0H:10T	________

Fill in the rest of this table. Note that the probability of having 5H:5T is only about 1/4, lower than many people might guess at first. The probability of getting a head *each time* the coin is tossed is 50%, but the probability of getting *exactly* 5H:5T is only .246.

There are many other kinds of distributions besides the binomial. A very important distribution, and one that is probably familiar to you, is the *normal distribution*, or so-called *bell-shaped distribution:*



The normal distribution fits many patterns in nature: human heights, human IQ scores, lengths of wings of birds or insects, amount of hemoglobin in the blood of snakes, running speed of lizards, etc. Many such measures fit the normal distribution because they are controlled by many independent forces (consider all the factors that determine the IQ score of a 13 year old child). Some of the other distributions we will use this semester are: the *Poisson, t, F,* and *Chi Square*. Some of these have characteristic shapes much like the normal and binomial. Others depend on *degrees of freedom* of the sample. The concept of degrees of freedom is well beyond the scope of this course; we must simply recognize that for certain distributions d.f. is important, such as the *t*, *F*, and *Chi-Square*

Hypothesis Testing

A second purpose of statistics, other than simply describing a particular population, is in the evaluation of hypotheses. Consider again our coin tossing experiment. Suppose your hypothesis is that the coin is “fair.” You flip it ten times and get 10H:0T. Can you reject your hypothesis? Look on the table of probabilities and note that there is a .00098 chance of getting 10H:0T. That is a very low probability. Are you willing to state that you reject the “fair coin” hypothesis since there is only a 1 in a thousand chance that you could have flipped a 10H:0T? Here you see that making such decisions is a “judgment call” on the part of any experimenter. *Any time* we test a hypothesis using statistical procedures we must make such a judgment call.

Normally the hypothesis to be tested would be stated as: This coin is not fair. The alternative, termed the *null hypothesis*, states that the coin is fair. That is, the null hypothesis is one of “no difference.” In the case of our coin tossing experiment, we are interested in “no difference” from the expected outcome under random conditions. Statistical procedures allow you to accept or reject a hypothesis by either accepting or rejecting the *null hypothesis.* Note again: <u>we test the null hypothesis</u>.

OK, suppose we make the following judgment call in advance. We will reject the null hypothesis if the probability of obtaining the result of an experiment or observation is less than 5% under chance conditions (stated as $P < .05$). For example, we might reject the null hypothesis that a coin is fair if the probability of obtaining a result from a tossing experiment is less than five percent.

But, there is an important point to keep in mind here. Suppose we decide to reject the null hypothesis if there is < 5% chance of obtaining our experimental result under random conditions.

That means there is still a certain probability that we have *rejected the null hypothesis when it is actually true.* If we flip the coin 1000 times and get 1000 heads, there is still a tiny probability that the coin was in fact a fair coin...if we reject the null hypothesis and declare that the coin is unfair, we have a tiny chance of being in error.

If we reject the null hypothesis when it is actually correct, we have made what is called a Type I error. There is another kind of error, a Type II error, in which we accept the null hypothesis when in fact it is false.

How can we be sure we haven't made an error when testing a hypothesis? Answer: we can't! This is why we must never apply statistical methods in our research without great caution and common sense. We must never bow before the god of statistics. The procedures are only tools to help us understand nature. We should no more fear statistical procedures than we would our 10x lens or spectrophotometer.

Summary

In using statistical methods, we first gather data, ask relevant biological questions, and then decide if statistical analysis is helpful. If so, we select the appropriate procedure, then decide on the *significance level* suitable for the study (the significance level is just deciding on the tolerable chance of making a Type I error...this may be very low if the experiment is a life-or-death matter [perhaps $P < .001$ might be used] or somewhat larger for other experiments [typically $P < .05$ in most research]). We run the test, check the tables in a statistics book to determine the probability of having the result we obtain, then decide if we will reject or accept the null hypothesis (the null hypothesis is usually designated as H_o). Note that statistical procedures can never be used to "*prove*" a hypothesis...we can only speak of probabilities that a hypothesis is true.

Thought question: Suppose we conduct an experiment to determine if fruit flies lay more eggs at warmer temperatures. We run a batch of flies at 20° and another batch at 30°. We then run the experiment over and over. After 100 experiments, we test each experiment to determine if the number of eggs produced by the flies differs in the two temperature treatments. We use significance level of .05 as our cut-off. Result: 96 experiments showed no difference, but 4 experiments showed a significant difference. We conclude that the flies in those experiments must have been genetically different from the others, that is why they were the only ones to show a difference in number of eggs produced. Is this a correct conclusion to draw?

PART II:
DERIVING THE CHI-SQUARE DISTRIBUTION

We will now do an experiment with coin tossing to derive empirically one of the most important distributions used in statistics, the *Chi-Square Distribution*. The formula for the value called Chi-Square (abbreviated as χ^2) is:

$$\chi^2 = \frac{(\text{observed number - expected number})}{\text{expected number}}$$

Suppose we flip a coin ten times. If we wish to test the hypothesis that the coin is fair vs. unfair, we propose that the expected number of heads in the ten tosses is 5 *if the coin is fair*. We can then apply the formula above to get the value of χ^2, then look this value up in a text to find out the probability of having that outcome.

Now derive the distribution of χ^2. Each member of the class will flip a coin ten times and record the outcome of number of heads and tails. Repeat this exercise ten times (so you have ten sets of ten flips). Then, for each set of flips, determine the χ^2 value. We assume in these experiments that our coins are fair and we flip them in an unbiased way. So, be sure to give your coin a good spin each time you flip. No backsliding here on your 89th flip! We don't want to see any half hearted flipping in this course!

Now, enter your data on the following table:

Trial	# Heads	# Tails	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1	________	________	________	________
2	________	________	________	________
3	________	________	________	________
4	________	________	________	________
5	________	________	________	________
6	________	________	________	________
7	________	________	________	________
8	________	________	________	________
9	________	________	________	________
10	________	________	________	________

Next, we want to plot the distribution of χ^2. We will have only about 160 values. The Laboratory Instructor will be in charge of the group histogram. When everyone gets all the points plotted, then you can made a summary graph for your own use of the whole class's results. In this case you will want to plot the value of χ^2 on the horizontal axis and the percent of times that value came up on the vertical axis, then draw a "smoothed" curve to describe this result. Your Laboratory Instructor will show you how.

What is the form of the distribution of χ^2? Your distribution will look a bit choppy because you only have 160 points. As you might guess, the distribution can be derived theoretically from the laws of probability. But here you see how *any* statistical distribution can be derived empirically. In some cases we cannot derive a particular distribution theoretically and must use empirical means, usually using a computer to "flip coins."

Now, let's put your curve to use. Suppose you step into the Sit-and-Bit Cafe one lonely evening for a quiet snack, but are soon accosted by a very large gentleman in a leather jacket painted with nasty words. This fellow tells you that he just lost $20 on a coin toss and he suspects that the coin is biased. He demands that you test the coin. A group of other large gentlemen in leather coats watch your actions. You flip the coin ten times and get 8H:2T. You calculate the value of χ^2. Now you dig into your pack and pull out the graph you have made of the distribution of χ^2. From the graph you determine the chance that the coin would turn up this value of χ^2 *or greater* (why do you want to know this value?). Now you can make some statement to the nice gentleman about the coin. What cut-off point for P do you choose? Why?

The Vermont Megabucks

Each week thousands of people in Maine, New Hampshire, and Vermont look forward to the weekly drawing of the tri-state lottery. A set of balls, each with a number ranging from 1 to 42 is tumbled, then they fall into a series of slots so that six of the balls are taken “at random.” These six numbers are the winners in that week’s lottery.

Some people become possessed with lotteries, feeling there is some pattern to the numbers being drawn. If the balls fall randomly, then there can be no pattern except for randomness. Every set of six numbers is equally likely when compared to every other possible set. But, is the contraption with all those balls really fair? Let’s use our new skill with the χ^2 test to determine if the lottery is fair.

Attached is the complete list of winning numbers for the tri-state megabucks lottery for the past year. We have a record of the winning numbers for 102 games. We can divide up the list and record the number of times each of the 42 numbers has been drawn. These will be our observed numbers. We expect that each of the 42 numbers would have an equal chance of being drawn. So, our expected numbers would be the total number of balls drawn during the year (= number of games or 102 times the number of balls drawn, or 7) divided by the number of balls available, or 42:

$$\text{expected}=\frac{102*7}{42}=17$$

Next, make a table showing the observed and expected numbers for each number possible in the lottery (the first three numbers have been done for you):

Previous Winning numbers in the TRI-STATE MEGABUCKS are:
(Source: http://www.mainelottery.com)

June 17, 2000	01-02-10-18-26-28	Bonus - 39
June 14, 2000	06-14-21-24-25-42	Bonus - 37
June 10, 2000	01-05-24-32-37-42	Bonus - 20
June 7, 2000	03-07-12-23-25-26	Bonus - 31
June 3, 2000	01-17-23-36-37-38	Bonus - 05
May 31, 2000	01-20-21-24-25-41	Bonus - 15
May 27, 2000	06-09-12-21-22-30	Bonus - 17
May 24, 2000	13-20-21-22-27-42	Bonus - 04
May 20, 2000	21-22-25-28-32-39	Bonus - 12
May 17, 2000	04-08-09-10-30-41	Bonus - 31
May 13, 2000	16-19-23-33-35-37	Bonus - 29
May 10, 2000	05-09-18-32-34-42	Bonus - 22
May 6, 2000	02-12-21-22-32-42	Bonus - 17
May 3, 2000	07-13-14-18-32-36	Bonus - 37
April 29, 2000	03-09-17-19-22-26	Bonus - 42
April 26, 2000	14-25-26-27-38-40	Bonus - 36
April 22, 2000	01-08-12-25-27-41	Bonus - 05
April 19, 2000	02-04-13-17-38-39	Bonus - 23
April 15, 2000	10-25-26-28-38-42	Bonus - 09
April 12, 2000	02-05-07-17-23-40	Bonus - 32
April 8, 2000	01-07-08-15-17-29	Bonus - 24
April 5, 2000	01-03-05-25-26-34	Bonus - 36
April 1, 2000	10-12-17-36-39-42	Bonus - 07

March 29, 2000	03-08-09-15-38-39	Bonus - 26
March 25, 2000	01-10-21-31-32-40	Bonus - 28
March 22, 2000	04-09-11-19-24-34	Bonus - 21
March 18, 2000	04-06-19-24-33-38	Bonus - 28
March 15, 2000	06-13-21-27-31-37	Bonus - 39
March 11, 2000	03-04-05-29-37-41	Bonus - 06
March 8, 2000	07-08-13-17-23-37	Bonus - 33
March 4, 2000	04-24-35-39-41-42	Bonus - 37
March 1, 2000	02-10-13-29-34-40	Bonus - 07
February 26, 2000	07-10-11-22-32-34	Bonus - 33
February 23, 2000	09-24-32-38-39-42	Bonus - 17
February 19, 2000	18-19-22-23-39-40	Bonus - 42
February 16, 2000	03-04-05-10-37-40	Bonus - 16
February 12, 2000	04-05-08-09-11-25	Bonus - 15
February 9, 2000	03-05-17-20-22-41	Bonus - 10
February 5, 2000	07-18-29-31-39-42	Bonus - 32
February 2, 2000	03-04-08-13-18-37	Bonus - 12
January 29, 2000	05-10-15-23-30-32	Bonus - 18
January 26, 2000	08-18-23-24-27-36	Bonus - 03
January 22, 2000	07-12-16-26-28-33	Bonus - 18
January 19, 2000	05-12-17-23-26-40	Bonus - 24
January 15, 2000	02-13-25-26-27-36	Bonus - 16
January 12, 2000	07-14-16-31-34-39	Bonus - 15
January 5, 2000	04-13-25-32-35-37	Bonus - 10
January 1, 2000	04-07-10-18-27-40	Bonus - 34
December 29, 1999	09-19-21-33-34-35	Bonus - 20
December 25, 1999	02-03-04-11-12-30	Bonus - 39
December 22, 1999	13-17-25-26-30-34	Bonus - 21
December 18, 1999	01-08-16-17-31-32	Bonus - 33
December 15, 1999	01-09-18-23-33-34	Bonus - 31
December 11, 1999	08-14-16-19-32-38	Bonus - 29
December 8, 1999	11-22-24-35-37-41	Bonus - 19
December 4, 1999	11-15-18-20-22-27	Bonus - 38
December 1, 1999	17-20-24-33-38-42	Bonus - 11
November 27, 1999	07-10-11-26-36-42	Bonus - 17
November 24, 1999	11-14-22-24-38-42	Bonus - 04
November 20, 1999	06-09-15-35-37-40	Bonus - 28
November 17, 1999	05-13-21-25-40-41	Bonus - 27
November 13, 1999	02-08-09-12-19-23	Bonus - 36
November 10, 1999	14-19-20-28-38-39	Bonus - 24
November 6, 1999	06-08-15-17-23-29	Bonus - 21
November 3, 1999	11-18-24-31-36-42	Bonus - 39
October 30, 1999	05-15-17-18-38-41	Bonus - 20
October 27, 1999	03-04-26-29-31-42	Bonus - 36
October 23, 1999	02-11-13-15-27-34	Bonus - 31
October 20, 1999	01-09-13-21-31-39	Bonus - 07
October 16, 1999	07-19-24-25-29-39	Bonus - 33
October 13, 1999	14-15-21-22-28-32	Bonus - 12
October 9, 1999	09-11-12-14-33-39	Bonus - 17
October 6, 1999	04-13-15-18-21-29	Bonus - 20
October 2, 1999	08-09-10-22-26-39	Bonus - 20
September 29, 1999	07-21-23-24-32-42	Bonus - 25
September 25, 1999	12-15-23-25-32-38	Bonus - 31
September 22, 1999	08-16-19-31-33-35	Bonus - 17
September 18, 1999	02-11-26-31-40-41	Bonus - 15
September 15, 1999	01-17-30-34-36-39	Bonus - 09
September 11, 1999	04-07-19-20-28-29	Bonus - 21

September 8, 1999	07-12-13-19-20-24	Bonus - 02
September 4, 1999	08-27-28-32-34-39	Bonus - 41
September 1, 1999	08-14-25-34-35-42	Bonus - 30
August 28, 1999	03-05-09-11-24-28	Bonus - 12
August 25, 1999	08-16-20-24-26-36	Bonus - 31
August 21, 1999	02-10-12-17-35-42	Bonus - 30
August 18, 1999	27-28-31-37-39-41	Bonus - 17
August 14, 1999	10-11-14-30-33-41	Bonus - 09
August 11, 1999	08-17-27-28-31-40	Bonus - 36
August 7, 1999	04-10-19-22-27-38	Bonus - 35
August 4, 1999	02-16-21-22-26-40	Bonus - 25
July 31, 1999	01-11-29-34-37-39	Bonus - 26
July 28, 1999	04-08-14-35-38-40	Bonus - 05
July 24, 1999	03-04-13-29-38-41	Bonus - 24
July 21, 1999	02-05-16-25-35-39	Bonus - 37
July 17, 1999	09-14-15-19-23-34	Bonus - 42
July 14, 1999	14-16-21-25-35-40	Bonus - 28
July 10, 1999	11-12-14-23-37-40	Bonus - 02
July 7, 1999	14-15-28-30-33-41	Bonus - 35
July 3, 1999	06-08-09-10-19-41	Bonus - 13
June 30, 1999	09-14-16-21-35-38	Bonus - 37
June 26, 1999	07-14-18-31-34-40	Bonus - 10

	Observed	Expected
1	______	17
2	______	17
3	______	17
4	______	______
5	______	______
6	______	______
7	______	______
8	______	______
9	______	______
10	______	______
11	______	______
12	______	______
13	______	______
14	______	______
15	______	______
16	______	______
17	______	______
18	______	______
19	______	______
20	______	______
21	______	______
22	______	______
23	______	______
24	______	______
25	______	______
26	______	______
27	______	______
28	______	______
29	______	______
30	______	______
31	______	______
32	______	______
33	______	______
34	______	______
35	______	______
36	______	______
37	______	______
38	______	______
39	______	______
40	______	______
41	______	______
42	______	______

For each of these pairs of numbers, apply the equation for χ^2 and sum all the results. This will give you the final value of χ^2 with 41 d.f.. Looking in a statistics book we can find that the critical value of χ^2 for 41 d.f. and $P < .05$ is 55.76.

PART III:
COMPARING TWO SAMPLES:
AN EXAMPLE OF ONE STATISTICAL TEST

We will now consider the problem of comparing two samples. Suppose we wish to determine if two species of flies differ in their reproductive output. Let's call them Species A and B. We collect ten of each species and allow them to lay eggs in the lab so we can count them. We do this trial twice; that is, we have two sets of ten egg counts for each species. The data are plotted up on Figure 1.

Look at the figure and see that there is variation in number of eggs produced by each species. Note that the two sample of Species A are not identical (variation between samples) and there is certainly variation between the results for Species A and Species B. There are several ways we could describe the variation in each sample. We could simply state the RANGE (the lowest and highest value) and perhaps could use that information to argue that Species 2 has a greater variation among individuals in the number of eggs produced. Another measure is the STANDARD DEVIATION which is the most common way information on variation is presented in the biological literature.

Standard deviation (SD) is simply an average of the deviations of each point from the mean for the sample. The reason SD is used more often than the range is that a single point could be an outlier – far from the mean – and other the others could be close to the mean, yet the range would be great. That is, the single outlier point would be given too much "weight" in the analysis. By calculating the average deviation from the mean (SD), we get a better picture of the variation.

SD is useful in another way IF the data are NORMALLY DISTRIBUTED (when plotted, look like the normal or bell-shaped curve). We can add 1 SD to the mean, and subtract 1 SD from the mean; the resulting range will include 67% of the points, and if we use 2 SD above and below the mean, then we have 99% of the points. Thus, if we use 1SD above and below, we can say, "Most of the flies produce within this many eggs." Or, if we use 2SD we can say, "Almost all of the flies produce within this many eggs." Pretty neat, huh?

We can now calculate the mean and standard deviation for each sample. Most good scientific calculators can calculate means and standard deviations, saving you the tedium of looking up the formula (for standard deviation) and doing all the calculations. The means and standard deviations for all four samples are indicated on the figure. (In a later exercise you will calculate the SD using the proper equation.)

We propose that the two species differ in their reproductive output...one species produces more eggs than the other. But, we also propose that the first sample drawn from each species is representative of the population, as is the second sample. That is, *within* species, there is no statistical difference between the samples. The variation we see between samples is a result of random sampling errors.

We can now “eyeball” the data on the figure. Ha, the samples within species don’t look very different, but the two species do seem to differ in reproductive output. Eyeballs can fail; we need rigorous testing of our hypotheses. First, let’s test if the two samples of Species A differ, then if the two samples of Species B differ, then finally we can test if A and B differ.

There are two common tests to determine if the average values of two samples differ: the Students t-test and the Mann-Whitney U test. The formula for calculation of the value of t is given in the chapter on variation in natural populations. The t-test allows us to determine if there is a significant difference between the means of two samples. Applying the t-test to the means of the two samples of Species A gives us t = .647. Looking up this value of t in the table in a statistics book gives us a P = .2623 , which is P > .05. Likewise, for a comparison of species B, we get t = .805 and P = .428 and thus P > .05. We cannot, therefore, reject the null hypothesis; that is, the means of the two populations of A and for B do not differ. We can thus combine the data for our two samples for each species.

Next, let’s compare the two species. The values of the means look different (36.3 vs. 46.5 for combined data). But, we might notice something else important, the variation seen in Species A looks less than the variation seen in Species B. Just by looking at the plotted data hints at this, but also the values of the standard deviations also look different (4.4 vs. 7.2). If the within sample variation differs between the two species, we *cannot use the t-test*. Why is this so? Every statistical test has some built-in assumptions. For example, when we did the coin flipping experiment we assumed that the result from flip number 1 did not influence the outcome of flip number 2 (this very reasonable assumption, of independence of the flips, is often missed by people when they think of probabilities...if the coin comes up heads the first 7 times they believe it has a greater than 50% chance of coming up tails on the 8th flip). Some statistical tests require some important assumptions about the distribution of the data being examined. The t-test, for example, requires that the variation in the two samples be *homogeneous*. Our eyeballing of the data suggests the variation actually differs between the two species.

One test of homogeneity of variation is the F test which is described also in the chapter on variation in natural populations. When we apply the F test to these data we find that the variation seen within each species differs significantly (F = 2.75, P < .05). Therefore, we cannot use the t-test. Statistical tests that have assumptions about the distribution of the data are called *parametric tests* and those that have no such assumptions are called *nonparametric tests*. We must therefore use a nonparametric test on these data. Nonparametric tests are often more difficult to calculate (not important if you are using computer statistical packages), and are less powerful, that is, we are more likely to make a Type II error. But, often they are the best ones to use when working with biological data.

The Mann-Whitney U test is a very useful nonparametric test to determine if the distributions of two populations differ in their average value (not the mean in this case, but the median). To calculate the value of U we first mix the data for both species (or, in general, the two samples), then order the data from lowest to highest value. We would expect, if the two samples have the same distribution, to have the values for Species A and Species B to be more-or-less evenly mixed: ABAABABBAAB, rather than one species occurring predominantly at the high or low end: AAABAABABBBB.

Now give a rank of 1 to the lowest score, a 2 to the next lowest, etc. Sum up the rank scores for Species 1 and the sum for Species 2. Pick the species that has the largest sample size (if the two samples have equal number of data values, pick either one), and call this sample 2. Then, apply the following formula:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

OR

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

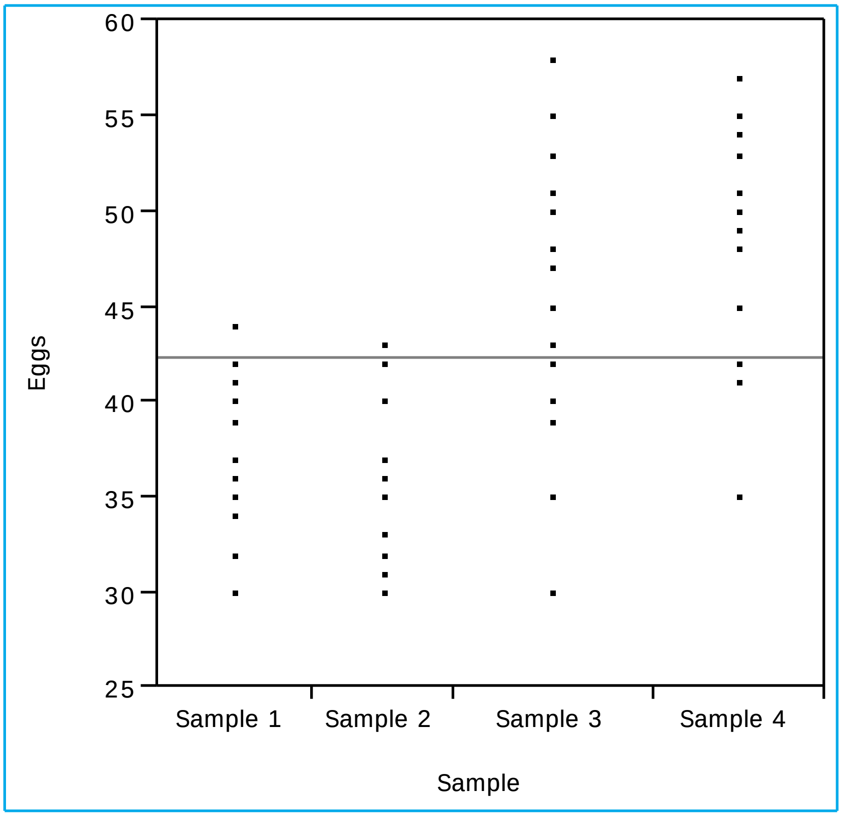
If we plug in our numbers (n_2 = the sample size of the species with the larger sample size, R_2 is the sum of the ranks of that species, etc.), we get two different values of U. We want the *smaller* value. The smaller U value is called U and the larger is called U'. They are related by $U = n_1 n_2 - U'$. We can now look up this smaller value of U in a table showing the distribution of U to determine its significance. When this is done we find that the U for these data is 79.5 and P = .0001 which is $P < .05$. So, we can declare that the number of eggs produced by Species 1 differs from the number produced by species 2.

If our sample sizes are large, the distribution of U approaches the normal distribution. Your Laboratory Instructor will show you how to use a table of normal values. You calculate the value of *z* used in such a table as:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{(n_1)(n_2)(n_1+n_2+1)}{12}}}$$

As we go through the laboratory exercises we will work with a number of statistical tests. They will be described as we meet them, but these basic ideas on statistics will be very useful to you. You might now and then reread this chapter to clear up any problems you might have.

Eggs By Sample



Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
Sample 1	11	37.2727	4.36098	1.3149
Sample 2	10	35.9000	4.58136	1.4488
Sample 3	14	45.4286	7.86144	2.1011
Sample 4	12	48.3333	6.51339	1.8803

Figure 1. Data on clutch size (number of eggs produced) for two species of flies. Samples 1 and two come from species A and samples 3 and 4 come from species B. Means and standard deviations are given for each sample .

Note that the samples differ within each species. We need to test if this difference between samples is significantly different from what we might expect if we drew two samples from the **SAME** population.

Also note that the two species differ in both means and standard deviations. We need to test if these two species differ significantly.

Statistics Summary

Data Distribution

Is there a difference in the mean of my samples?

Are my variances homogeneous?

F – test

H_o: $\mu_1 = \mu_2$ - $df_{num}=n_{(larger\ variance)}-1$

H_1: $\mu_1 \neq \mu_2$ - $df_{denom}=n_{(smaller\ variance)}-1$

$$F_{stat} = \frac{\text{Larger Variance}}{\text{Smaller Variance}}$$

$F_{stat} > F_{crit}$ → Reject H_o → Mann-Whitney U-test

$F_{stat} < F_{crit}$ → Accept H_o → t - test

Mann-Whitney U-test

H_o: $Median_1 = Median_2$

H_1: $Median_1 \neq Median_2$

See manual for procedures

t - test

H_o: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$ – $df = n_1 + n_2 - 2$

$$t_{stat} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right|$$

$t_{sat} > t_{crit}$ → Reject H_o

$t_{stat} < t_{crit}$ → Accept H_o

Is my observed data distribution Different from my expected data distribution?

χ^2 test:

H_o: Expected = Observed

H_1: Expected ≠ Observed

df = n-1

$$\chi^2_{stat} = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

$\chi^2_{stat} > \chi^2_{crit}$ → Reject H_o

$\chi^2_{stat} < \chi^2_{crit}$ → Accept H_o