LAB #3: Statistical descriptions of data; comparing two samples

---

BEFORE LAB

- **Read the Introduction and skim the lab exercises below.**
- **Review microscope use and safety and take the Blackboard quiz**

OBJECTIVES

1. Understand the primary functions of statistics; describing populations and testing alternative hypotheses.
2. Familiarity and understanding of statistical terms; Population vs sample, parameter, mean, median, mode, continuous vs categorical data, alternative vs null hypothesis, test statistic including the example of *t*, alpha and *P*-value, population and sample variance, population and sample standard deviation, standard error.
3. Understand the meaning of *P*-value in the context of hypothesis testing.
4. Practice skills in data management, graphing, and statistical analysis (excel).

INTRODUCTION

Populations and samples

Populations in statistics are the complete groups of interest; the full set of individuals we are interested in making inferences about. It could be all men or women, or all voters in an election, or all members of a single lizard species. There are two sampling problems statistics tries to solve. First, we almost never can collect data on an entire population. Instead, we have to make inferences about the **parameters**, variables that describe a population [like mean, median, or variance], from smaller **samples***. We measure these parameters in the samples, then use the variability in those samples to figure out how close we thing our sample parameter estimates are to the true population parameters. Second, we often don't even know if we are dealing with two or more separate populations for a given trait of interest, like leg length, or if we are in fact looking at a single population. Note that two different species might well be good biological species, but if they don't differ in some specific trait, like leg length, we consider them statistically to be a single population (for that trait). In this lab we will measure a couple different kinds of samples; *Paramecium bursaria* vs *ceratium*, and the morphology of Galapagos finches before and after a drought in the Galapagos. In both cases we will use statistics to ask whether the two samples come from the same or different populations.

---

* Proper sampling is a subject of its own, but any factor that makes a sample less than a completely random selection from the population of interest can bias results, and almost all actual samples are called samples of convenience, meaning they are what we could get, not what is actually out there. One of many good discussions is here; https://blogs.scientificamerican.com/guest-blog/where-are-the-real-errors-in-political-polls/

Before we get to statistical parameters we should define variable types. Data can take at least three forms; continuous, categorical, and ordinal. **Continuous** data varies as it sounds, quantitatively in either integer or decimal units. Think length, width, running speed, height, weight, age at death, development time. **Categorical** data refers to discrete, discontinuous states that variables can take. Think sex, species, phylum, state or country of origin, color, or genotype. Ordinal data are beyond the scope of this course, but are used frequently in behavioral studies, they are categories with ranks, such that the order is important but the scale is not regular or linear. In this lab we will describe continuous variables using the sample parameters of mean, variance and standard deviation. The parameters we will deal with today are measures of central tendency and spread.

Measures of central tendency

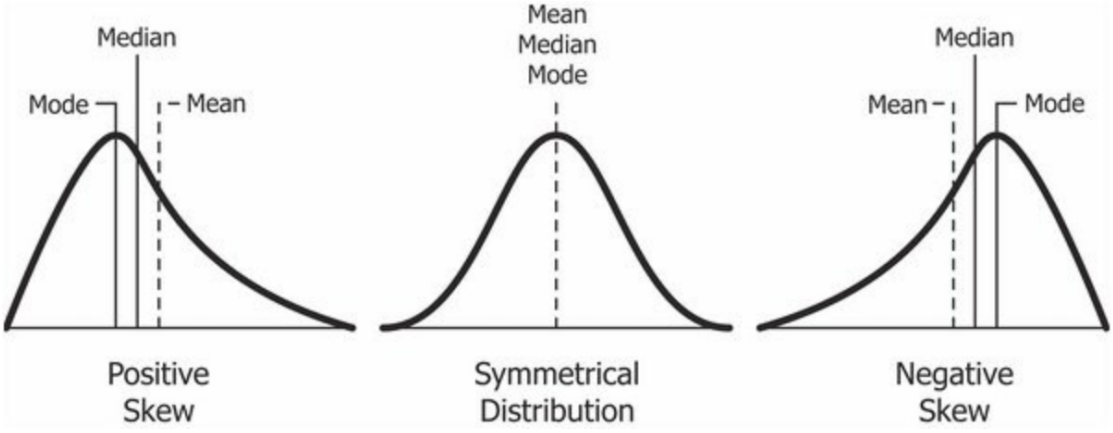
Central tendency can be described by the **mean**, or average of all values, the **median**, or the value separating the larger half of the values from the smaller half, and the **mode**, the most common value. The mean is the parameter we are probably most comfortable with. The mean score a class achieves on an exam is often what we use to measure our own success. We use a bar above a variable to show that it is a mean ($\bar{X}$). The entire population has a mean μ, and an estimate of the population mean from a sample is $\bar{X}$.

$$\textbf{Population mean} = \mu = \Sigma X_i / N$$

$$\textbf{Sample mean} = \overline{X} = \Sigma X_i / n$$

In the above formulas the Σ is the symbol for "sum of", so ΣX_i means sum of the ***X's*** - that is summing the values from each individual in the sample and dividing by ***n,*** which is the number of individuals in the sample

A **normally distributed population** is one that is fully described by the mean and the variance. It is symmetrical around the mean, giving it the familiar 'bell' shape. In theory the tails extend to infinity in each direction. The mean, median and mode are the same. In the right (positive)-skewed distribution the mean is larger than the mode because it is affected more by extremely large values. An example of a right-skewed population is household income in the US.



Author: Diva Jain. Source Wikipedia

The median individual income in the US was under $33,000 in 2018. However, the extremely high annual incomes of the wealthiest 1% (average income in this percentile is over $1,300,000) raises the mean individual income to $50,000 (US Soc Sec Admin). The mean is more vulnerable to outliers than the median. The left (negative) skewed population has a mean smaller than the median or mode because it is more affected by extreme negative outliers.

Measures of spread

Measures of variability in data are just as important as measures of central tendency. This is because the more variable a population, the more often you might collect two samples with means pretty far from each other by chance. The less variable a population, the closer we would expect the means of multiple samples to be to each other. This means that the variability (sometimes call spread or dispersion) of our data is critical to determining the statistical significance of any difference in the means of different samples. One measure of spread is pretty crude, the **range** of values from low to high. The ones we will focus on are based on the differences of each point from the mean.

Variance: If we were to just use the average difference of each point from the mean as a measure of spread, we would always end up with zero. This is because by definition some points will be above and some below the mean, and the sum of those differences must cancel out. One way (not the only way, but that's history) to make the differences positive is to square them. The fundamental measure of spread in statistics is the variance, the average of the squares of the differences of each point from the mean. The population variance, denoted σ^2, is then:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Where *i* is each invidual data point, *N* is the population size, X_i is the value of an individual, and μ is the population mean. Now remember we almost never get to measure the population variance or mean (μ). Instead we collect data from samples, that have sample mean $\bar{X}$ and the sample variance is slightly different:

$$s^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{n - 1}$$

This is very similar, but not identical to the population variance. Some things are just symbols. $\bar{X}$ really is the same calculation as a population mean, just for the sample collected, similarly *n* is just the sample size rather than the population size. However, why *n*-1 and not just *n*? Stats math geniuses over a century ago realized that small samples tend to under-estimate variance, because they are less likely to pick up rare extreme values, so they put in a correction; using *n*-1 in the denominator will make a fraction bigger than using just *n.* The smaller the sample, the more of an enlarging effect subtracting one from the denominator will increase the sample variance estimate. Large samples, that are likely to catch some rare extreme values won't have sample variances increased much at all. The sample **standard deviation**, which is sort of like (not exactly, don't worry about it until you're in a stats class) the average difference from the mean, is the square root of the variance, so $s = \sqrt{s^2}$. Variance and standard deviation (SD) alone are measures of how variable your sample data are, but to put a range around a mean estimate or test the difference between two means, we need to measure the uncertainty of the mean estimate itself, not the variability in our data. The term that describes variability around the sample mean estimate is the **standard error, SE**, sometimes written $SE_{\bar{X}}$, to show that it is describing

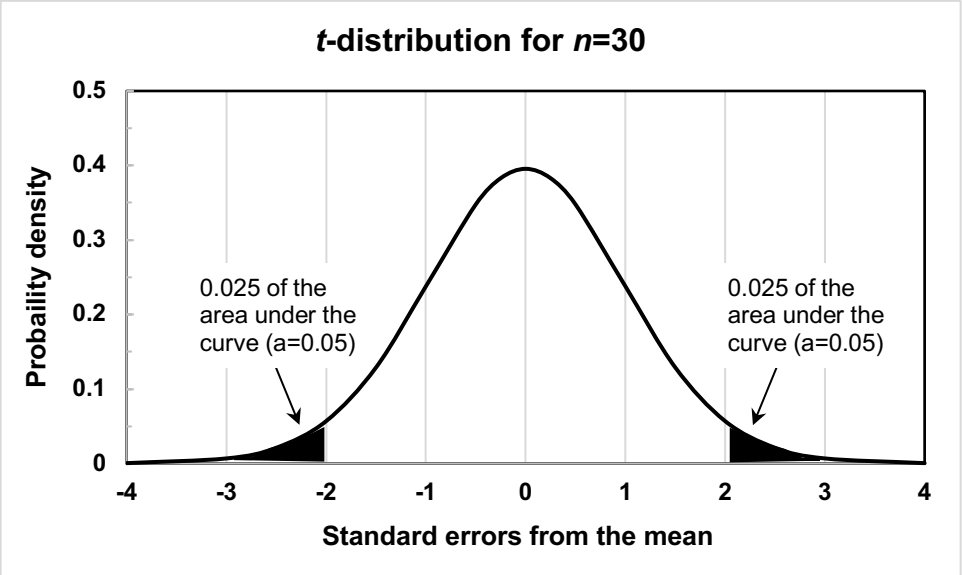
variability of the mean estimate, not variability of the individual data points. It is calculated by taking the sample standard deviation and dividing it by the square root of the sample size,

$$\mathrm{SE}_{\bar{X}} = \frac{s}{\sqrt{n}}$$

In this way SE is influenced by variability among data points, but there is also a large effect of sample size; larger samples will have larger denominators, and smaller standard errors. This makes a lot of sense. If individual data points are less variable (low *s*) then there is less variability around the estimate of the mean (any sample will probably be close to the true mean). However, even if individual data points are highly variable, it you sample a lot of them the sample mean will also be pretty close to the true mean; there will be less uncertainty around your mean estimate.

The range within which the true mean is found

The standard error ($\mathrm{SE}_{\bar{X}}$) is used to define the range within which we have a certain level of confidence (90%, 95%, 99%) the true mean lies. We determine that **confidence interval** by multiplying the SE by a new variable, *t*. ***t*** a continuous probability distribution (a curve that defines the probabilities of different values of a variable) that describes how many standard errors away from the true mean small samples are expected to be. It was developed by a statistician working to improve quality controls at Guinness brewing company from small samples at the start of the 20th century. For a given value of degrees of freedom, which for a single mean is just one less than the sample size (*n*-1), and a given level of confidence in your mean estimate, *t* will tell you the range where that proportion of sample means will fall. In the figure below there is a probability density function for sample means of size 30. The filled areas each cover 0.025 (2.5%) of the area of the curve, showing that 5% of the time sample means will be greater than 2.042 standard errors away from the true mean.



So, to generate a 95% confidence interval around a mean estimate from a sample, you add and subtract *t* times the standard error from the sample mean:

$$\mathrm{CI} = \bar{X} \pm t_{\frac{\alpha}{2},df} * SE_{\bar{X}}$$

Where $t_{\frac{\alpha}{2},df}$ can be looked up in a table like the one below, for a 95% confidence interval $\alpha/2$ is 0.025, *df* is *n*-1, and $SE_{\bar{X}} = \frac{s}{\sqrt{n}}$. Note from the table below that for large sample size a good rule of thumb is that 95% of the sample means, like 95% of the data in a normally distributed population, lie within 2 standard errors of the sample mean.

Table 1: Values of *t* for various *df* and α-levels.

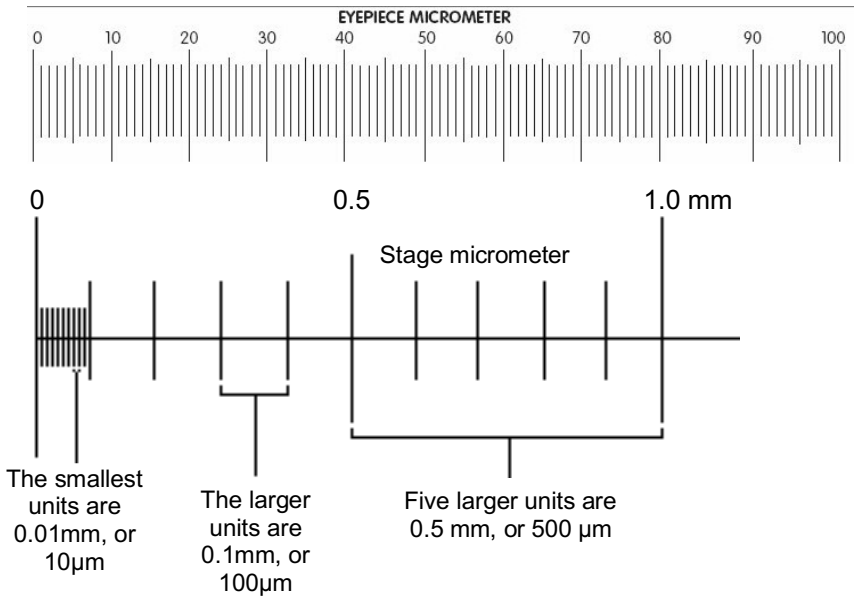
			CI *P*=0.95		**CI *P*=0.99**
df	**α/2=0.1**	**α /2=0.05**	**α /2=0.025**	**α /2=0.01**	**α /2=0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
infinity	1.282	1.645	1.96	2.327	1.576

Exercise 1: Comparing the size of two protists.

Hopefully we learned how to use the stage micrometer in Bio 105 or at least in the chapter to prepare for the microscope quiz. We will quickly review. The first rule of QC microscopy is … Never get involved in a land war in Asia! No, that's the Princess Bride. Here's the real one:

Never use coarse focus above 4X. That's because the stage micrometers we use cost $100-200, and even a paramecium slide is $35. Oh and we don't want you to get cut by broken glass. Basic safety is that you use course focus only in 4x, after you've centered your subject, then increase the magnification, always by turning the ring and not yanking on the objectives, looking from the side to make sure you're not going to grind the objective into the slide, and then use only fine focus at the higher magnification. My favorite tips for basic microscopy are to not be lazy eyed and use both eyes, adjusting the spread of the oculars so you see a single image circle, and don't set the lamp to 11 all the time. Lower light, and even narrower aperture will often give much better contrast and detail.

The first step will be to work with your partner to each measure 10 protists, either *Paramecium caudatum*, a population ecology classic, or *Ceratium*, an armored dinoflagellate. Take a look at your protist slide. Can you see the ocular micrometer, the scale that rotates when you rotate the eyepiece? Great. While you're at it, rotate the ocular micrometer so it is horizontal. Now, are you going to use 4x, 10x (most likely), or 40x to measure your protist? The next step is to calibrate your ocular micrometer at the magnification you have chosen, to determine how many actual µm each unit on the ocular micrometer is. Remember the stage micrometer is 1 or 2 mm long, divided by larger divisions every 0.1mm (100µm) and the smallest divisions are 0.01mm (10µm). Align the ocular and stage micrometers so you can get a good count of how many ocular units are in a known number of stage units.



In the figure to the left, there are 80 units on the ocular micrometer spanning a range of 1 mm (1000 µm) on the stage micrometer. 1000µm / 80 units is .. 100µm/8 units, or … 25µm/2 units, or … 12.5µm per ocular unit. You're ready now! You can remove the stage micrometer, put it away carefully, and now you can count the ocular units for each protist and multiply it by the factor for your microscope for each sample.

Ocular units: _____ Stage micrometer µm: _____
µm / ocular unit_____

Fill in the table below with your 10 measurements in ocular units. For fun let's use excel to calculate the actual length in µm.

Species:		
Observation	Ocular units	µm
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

There is a companion excel file for this lab, called "**Lab 3 T-testinator**." Since the whole point of this and any lab is to mess things up, download the excel file, and rename it, best to include your name and lab. Use the first sheet in the workbook, "**Single sample worksheet**", enter the species, the ocular units and the stage micrometer µm. In cell C4, to the right of the cell that says "µm per unit" type "=C3/C2" without quotation marks. The '=' sign tells Excel that you are typing a formula, and either clicking on cells or typing the column letter and row number will tell excel to use the contents of those cells in any formulae. There is a huge number of functions excel can calculate for you, each with its own non-intuitive syntax. After you have entered the 10 measurements in ocular units, can you enter a formula in cell C8 to calculate the first sample's length in µm? When you have done so, perhaps with help from your partner or lab instructor, copy and paste that formula to calculate the lengths of all the samples. You'll see that the average of the sample is calculated in cell C18, but also that this workbook used the Insert\Name\Define Name menu command to assign the name 'average' to that cell, so now other formulas can refer to that fixed cell value in their own formulas. The first cell in the colum that calculates the difference between each datapoint and the sample mean use that variable name. Take a look at those formulas and paste them down, so now the sheet will have the sum of squares and the variance calculated. Now you enter the formulas to calculate the SD and SE. In excel the square root of a value x would be written (x)^0.5. Now you're ready to calculate the confidence interval:
What is the value of t for n=10, $\alpha/2 = 0.025$? _____
Use that value of t to calculate the CI_{95} as described on pages 4-5 _____

We are now ready to compare the sample means of the two species and ask if they are significantly different.

Compare the means of two samples:

To statistically test whether two means are different we need a new measure of uncertainty around our sample estimates. Instead of the $SE_{\bar{X}}$, the standard error of the sample mean estimate, we need the $SE_{\overline{X_1}-\overline{X_2}}$, which is to say how variable is the estimate of the difference between the two sample means.

$$SE_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{s_1}{N_1} + \frac{s_2}{N_2}}$$

Where s_1 and s_2 are the standard deviations of each sample, and N_1 and N_2 are the sample sizes. There are different ways to calculate the test statistic, *t*, depending on assumptions of equal variances, but let's use the simplest, not because the assumptions are met, but because it shows what shapes the test statistic. The simplest test statistic to compare two means is

$$t = \frac{X_1 - X_2}{SE_{\overline{X_1}-\overline{X_2}}}$$

How will *t* vary with an increasing difference between the sample means?

_____

How will *t* vary with increasing variability of the data within each sample?

_____

How will *t* vary with an increasing sample size (N_1 and N_2)?

_____

Degrees of freedom in a two-sample *t*-test is easy to calculate if you know or can assume that the variances are equal in the two samples. In that case it's N_1+N_2-2. It has to be approximated if that assumption can't me made, and that's beyond this lab.

Work with your partner to complete the following table:

Parameter	**Length of *Paramecium caudatum* (in *μ*m)**	**Length of *Paramecium caudatum* (in *μ*m)**
Sample size N		
Mean $\overline{X}$		
Stnd Deviation *s*		
Variance s^2		
95% Confidence Interval CI		
$SE_{\overline{X_1}-\overline{X_2}}$		
$t_{\overline{X_1}-\overline{X_2}}$		

Now, look up in the *t*-table above for 18 *df* (N_1+N_2-2) whether the *t* calculated from your samples is greater than the **critical value** of *t* in the table for α=0.05 (α/2=0.025). The critical value of a test statistic is the value of the test statistic above which we reject the null hypothesis. Let's unpack that definition. A null hypothesis is the hypothesis of no difference, relationship or effect. For the two-sample mean it is that the two means are equal. The alternative hypothesis is that the means are different. The critical value depends on ***p***, which is a measure of how unlikely our data are if the null hypothesis was in fact true. The *p*-value is the probability of obtaining a result as far from the null hypothesis prediction as our data were, or farther, if the null hypothesis were true. In this case it is asking how often we would observe means as far from each other or farther if the true means were the same. A low *p*-value suggests that it would be very unlikely that two samples from the same population would have means as different as we observed, and provides evidence to reject the null hypothesis. When the *p*-values is less than the alpha-level you select, the result is considered statistically significant. Typically, 0.05 is used as the alpha-level.

What's the critical value of *t* for 18 *df* and *a*=0.05? _____

How does the test statistic *t* compare with the critical value?

_____

What's your conclusion?

_____

The next part of the lab is adapted from HHMI Bio interactive

INTRODUCTION

In 1973, Princeton University evolutionary biologists Peter and Rosemary Grant began studying the finches of the Galápagos archipelago, a group of islands about 600 miles off the coast of Ecuador. They collected thousands of measurements every year to track changes in the physical characteristics of finch populations over time. One of their major goals was to collect enough data to identify associations between environmental and evolutionary changes in finch populations.

For their study, the Grants focused on the medium ground finch (*Geospiza fortis*), a seed-eating species of finch on the island of Daphne Major. Every year, the Grants measured physical characteristics like wing length, body mass, tarsus length (the section of leg between the ankle and knee), and beak depth for hundreds of individual medium ground finches. Small changes in these structures can be important for survival in different environments. In addition, these traits tend to vary widely within populations.

In early 1977, a drought began on Daphne Major. The drought lasted for 18 months and caused the type and abundance of food available to the finches to change rapidly. Medium ground finches prefer to eat the small, soft seeds of the bushy plant chamaesyce (*Chamaesyce amplexicaulis*), but the supply of chamaesyce seeds was extremely limited as a result of the drought. As the drought progressed and the hungry finches quickly ate the small, soft chamaesyce seeds, one of the only remaining food sources for the medium ground finch became the seeds of a plant called caltrop (*Tribulus cistoides*). Caltrop seeds are much larger and harder than those of the chamaesyce and are covered with pointy spines. Fewer than 20% of the 1,200 medium ground finches on the island survived the drought of 1977.

The Grants were interested in determining whether there were any differences between the finches that survived the drought and the finches that did not—and in particular, whether any physical characteristics were key to survival. To answer this question, they compared the average value of different characteristics in the finches that survived the drought to the average values of the same characteristics in those that did not survive. They then applied statistical methods to determine whether the differences they found between the two groups were likely to be real or merely occurred by chance.

You now have the opportunity to statistically analyze data collected by the Grants.

The final sheet in the workbook contains actual data from 100 medium ground finches living on Daphne Major in 1976. Fifty of those birds did not survive the 1977 drought (nonsurvivors) and 50 did (survivors). Use the T-testinator spreadsheet to complete the table below comparing the samples of survivors and non-survivors. To the right is an example of a good **bar-graph,** showing how one continuous trait, dorsal fin height varies as a function of the explanatory categorical variable sex. The bars above the means are 95% confidence intervals. In the box below the table, draw a bar graph of how one variable varies between survivors and non-survivors.

Mean Dorsal Fin Height Among Male and Female Orca Whales

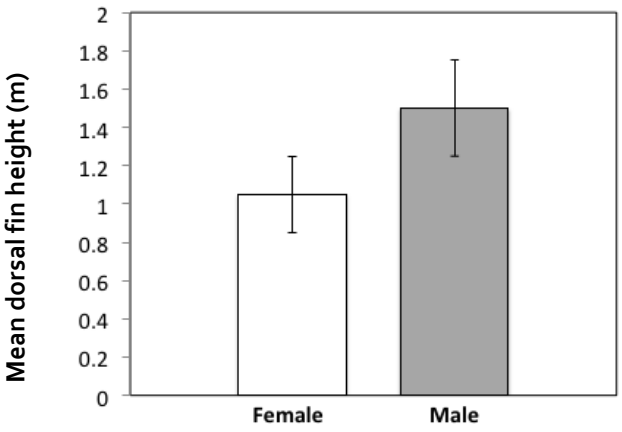


Table: Morphological variation in surviving and nonsurviving Darwin's finches.

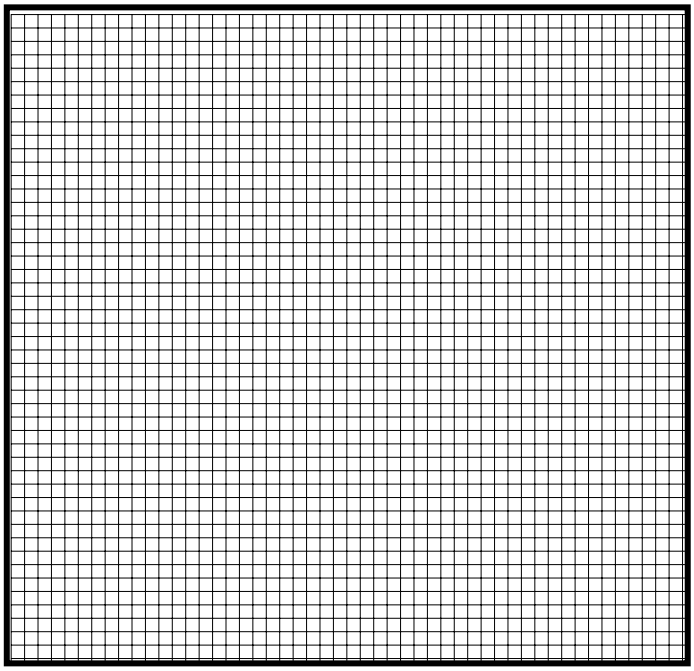
Descriptive Statistics	Nonsurvivors				Survivors			
	Body Mass (g)	Wing Length (mm)	Tarsus Length (mm)	Beak Depth (mm)	Body Mass (g)	Wing Length (mm)	Tarsus Length (mm)	Beak Depth (mm)
Mean								
Variance (s^2)	1.842	5.181	0.701	0.775	3.087	5.448	0.735	0.709
Standard Deviation								
Standard Error of the Mean								
95% Confidence Interval								

Figure caption: _____

_____

_____

_____



From our discussion of this example in lecture, and from the video ‘Beak of the finch’, which variable would you expect to change the most – what was selection supposed to be acting on?

Of the four variables, which seemed to change the most over the course of the drought. Why do you say that?

What do you think is a better metric to measure amount of change; SD standard deviation or SE standard error, and why? Is it more meaningful to say a mean shifted 1.5 SD’s or 1.5 SE’s?

Using the answer above, what trait do you now think changed the most, and what measure are you using to measure change?

Statistical terms worksheet – make sure you understand these!

Population vs Sample	
Mean	
Median	
Mode	
Sum of squares	
Variance	
Standard deviation (of the sample or population)	
Standard error (of the mean) **(How are SD and SE different?)**	
Confidence interval	
***t*-test**	
***t*-test is used to:**	
***t*-value is an example of a test statistic**	What's a test statistic? It's a measure of how different our observations are from the null hypothesis. Examples include *t, F,* and χ^2
Null hypothesis	
Alternative hypothesis	
***P*-value**	