Genetics: Human Genetic and Environmental Variation

Introduction

Variation is a ubiquitous feature of natural populations. All organisms exhibit variation for a significant number of morphological, biochemical, and behavioral characteristics. Genetic variation is essential for the process of natural selection to produce evolutionary change. This is a teaching cliché, but do it anyway-look around the classroom and you will immediately notice a great deal of variation among members of this class. Some of this variation is morphological: hair color, height, eye color, etc.. Some is behavioral: preference for certain foods, knowledge of languages, choice of clothing, etc.. For centuries, biologists have sought an explanation for the existence of variation. Much of it has a basis in our genes, a fact that is of tremendous evolutionary significance. Other variation is primarily due to environmental influences on our development. For nearly every trait, however, both genes and environment interact to some extent to produce the organism's phenotype.

Phenotype: the observable properties of an organism.

Types of Variation

Every observable trait can, in theory, be measured or scored (to score a trait simply means to record what category that observation falls in to). There are different types of variables:

Attributes, or qualitative variables, can be scored, but not fall into a continuum. Examples include human eye color, political party, blood type, and gender

Quantitative, or **measurable, variables** fall along an axis, and can be measured to observe their place relative to others.

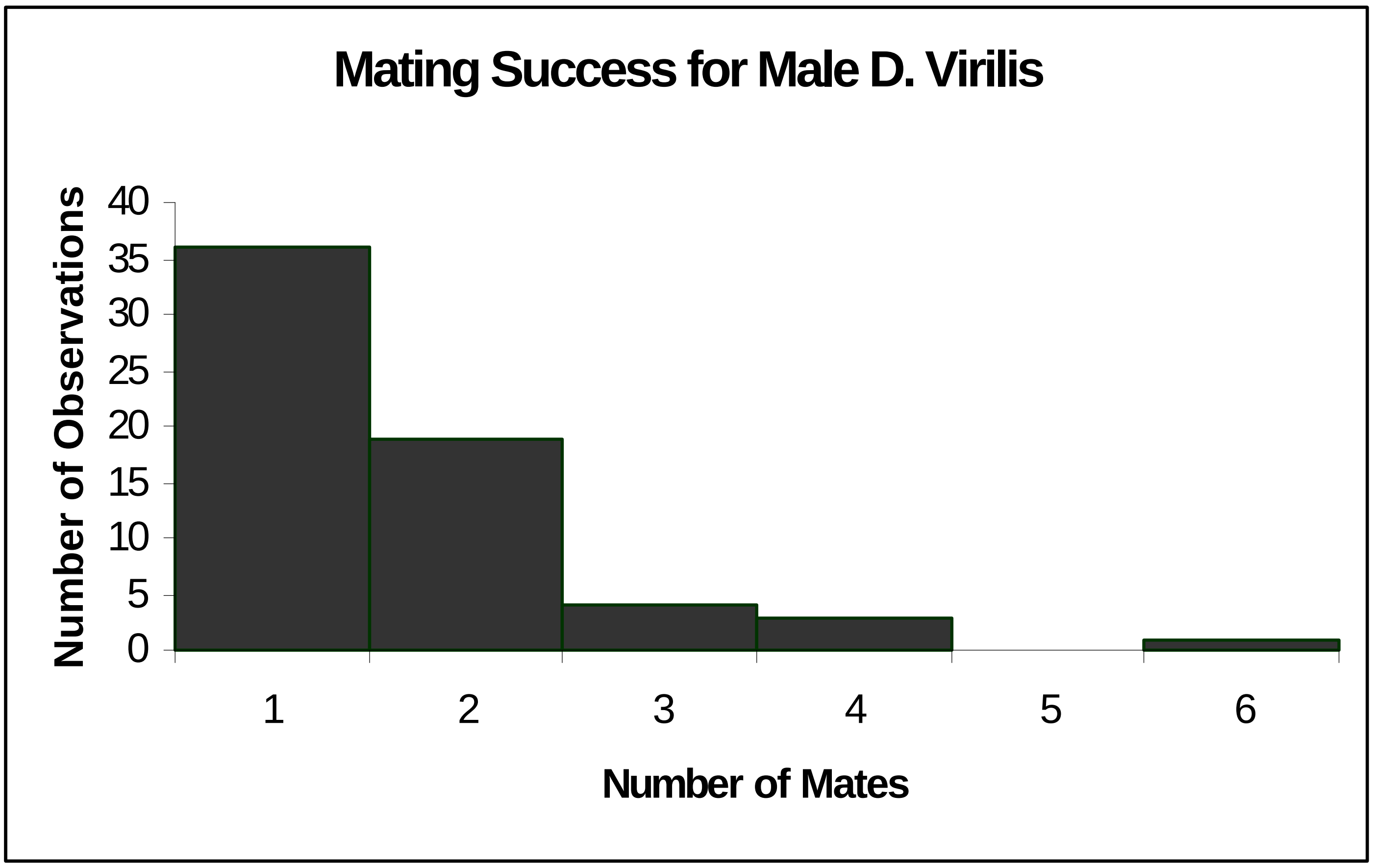
Discontinuous measurable variables: fall into discrete intervals. Examples include shoe size, number of mates, number of scales along a certain stretch of a snake's face.

Continuous measurable variables do not fall into discrete intervals, they exist along a continuum. Examples include height, weight, age, length of an insect's thorax.

Distributions of Values

Scientists frequently study populations of organisms, both natural and artificial. By measuring a group of organisms, they are implicitly studying a **statistical population**. A statistical population is the group of organisms in which a scientist is interested and is attempting to characterize. A scientist can measure every member of their statistical population, or they can attempt to characterize it by measuring a random sample of individuals. It is important that this random sample actually be a random, representative sample, or the statistical population under study will not reflect the real population that the scientist may think they are studying. Every population has a **distribution of values** for every quantitative variable. This reflects the number of individuals possessing each value for the trait. These distributions are frequently expressed as a **histogram.** In drawing a histogram, the range of values is broken into intervals (the x axis), and the number of individuals within that interval is expressed as the height of a bar (y axis).

A Histogram



Mean, Median, Variance, etc.

The distribution of numerical values can be described by several statistics. Some of the most helpful and basic stats are the **mean**, the **median**, the **variance**, the **standard deviation,** and the **range**.

The arithmetic mean (this is usually what we mean when we say "mean", but there are others as well, such as the harmonic mean) is simply the average.

Thus $\mathbf{x} = {}^{\Sigma x}/_{N}$. Where **x** is the average, **N** is the number of observations, and Σ**x** is the sum of the series of observations**.**

The median is the value with the same number of values preceding it, and following it in a series ordered from smallest to largest. To find the median, order the list of measurements from smallest to largest. Find the central value (if there are an odd number of values), or the midpoint between the two central value (for an even number of values).

The **mode** is the most common value in the data set. There can be several modes.

The **variance** describes how variable the data set is; it is an index of how likely each value is to depart from the mean**.** This is a very important aspect of nature that we tend not to think about in our everyday lives. For instance, the mean ovarall temperature in Chicago is about the same as the mean overall temperature in Bermuda, but the variance in temperature (at least the monthly variance and day to day variance) is much higher in Chicago. This makes the climates of the two cities very different (because variation is so important to climate, you hardly ever see overall means in weather forecasting- since most people don't understand variance, so monthly means are listed for high and low temperatures.)

Variance, $\sigma^2 = {}^{(\Sigma(x\text{-}x)^2}/_{N\text{-}1}$.

The **standard deviation** is the square root of the variance. In many ways, it is more useful as a summary statistic, because it is in the same units as the mean, and should be about the same order of magnitude.

The range is another good summary statistic to describe variation. It is sensitive to the size of the data set, because the more values that accumulate, the more likely an extremely high or an extremely low value are to crop up. The best way to give the range is to simply list the high and low values.

Types of Distributions

Populations of actual organisms exhibit a great variety of distributions for different measurable variables. Some common distributions are: **normal, bimodal, and multimodal**. Distributions may also be **skewed**, or exhibit **kurtosis**.

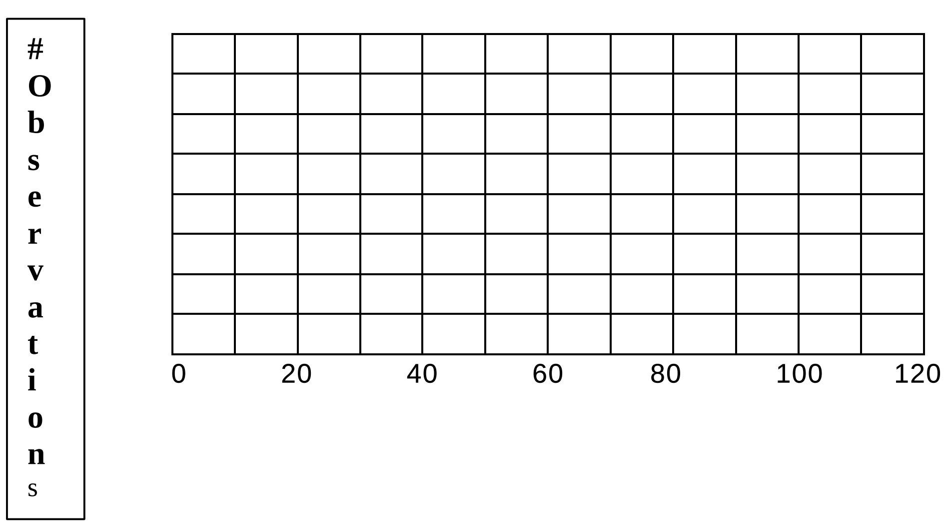
Pre-Lab Discussion

Before coming to class, answer the following questions on a separate sheet of paper and turn them in to your laboratory instructor at the beginning of class.

1) A biologist hikes up a mountain and measures a population of Pondersosa pine. The following series of observations are trunk diameters, in centimeters, measured at a height of 1m from the ground, and taken within 200m of a central point.

12	33	77	45	21	11	9	6	8	9	81	103	14	11	3
8	31	74	12	11	4	6	1	7	2	12	4	21	21	66
81	77	75	13	10	107	14	26	9	9	4	1	1	67	88
5	4	5	9	21	7	84	11	8	70	12	9	13	4	16

2) Calculate the mean, variance, and standard deviation for the above data set. Find the variance as well.
3) How are the data distributed? What might this pattern say about this population of pines?
4) Construct a histogram of your data in the space below.



Questions for Group Discussion

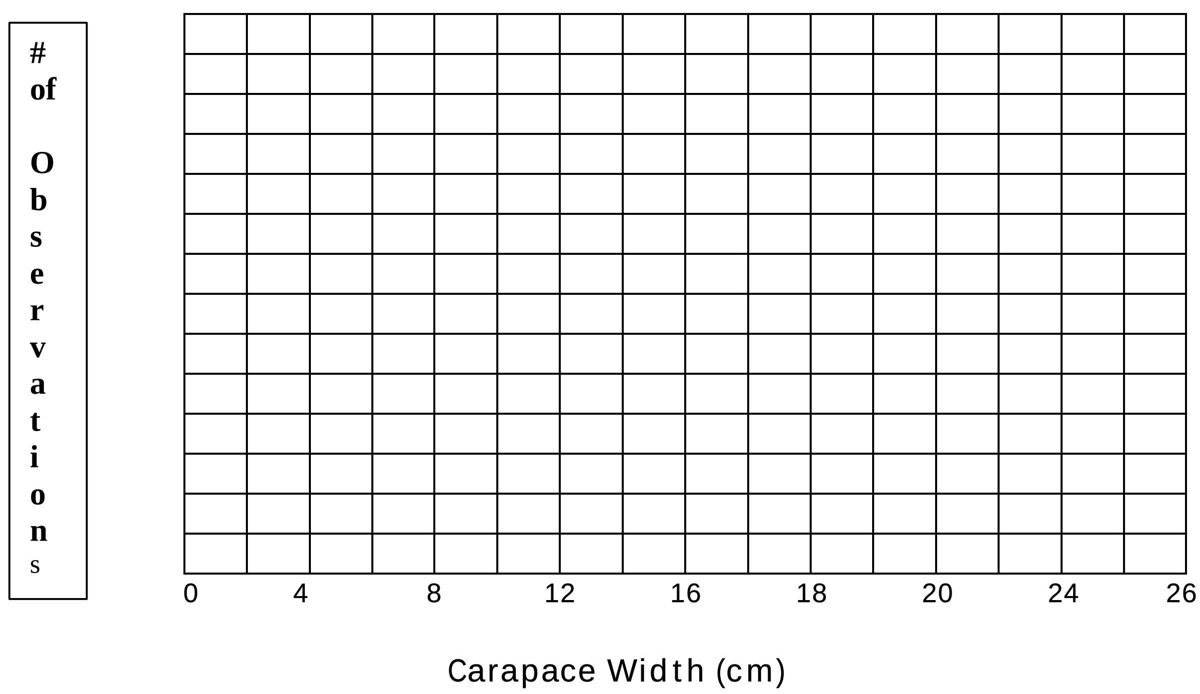
A) A biologist measures a population of Blue Crabs. The following series of observations are carapace widths, in centimeters, measured at the widest point of the animal.

12.0	3.1	7.7	7.5	15.0	11.0	16.8	16.6	8.4
8.3	19.6	17.4	12.0	14.4	14.4	6.1	21.1	12.5
18.1	17.7	4.4	13.6	16.5	17.7	14.4	26.3	12.1
19.1	16.9	15.4	19.0	16.0	16.4	13.4	11.0	8.0

B) Calculate the mean, variance, and standard deviation for the above data set. Find the variance as well.

C) How are the data distributed? What might this pattern say about this population of blue crabs?

C) Construct a histogram of your data in the space below.



Laboratory

In laboratory today, you will be measuring a small sample of human beings; this sample will be yourself and the other members of your class, including your laboratory instructor. You will select one attribute, or categorical variable, and one measurable variable. Some of these variables have a well-established genetic basis, some of them are mostly due to the environment, and some have notable contributions from both genes and environment.

For each variable, you interview each member of the class, and score them for gender, as well as another attribute, and measure them (or interview them) for two measurable variables. To do this, give everybody in class a number (probably the number in which they are interviewed). For each number, score the attribute and measure the measurable variable. Keep the data in pairs, so each attribute score has a measurable variable associated with it. See the sample data table.

The list below contains some good attributes. You may select others, but do NOT pick anything touchy such as political party or sexual orientation.

Attributes-Score your subjects for GENDER, CILANTRO, and for ONE OTHER of the following:

Gender is the sex of the person as recognized by society. A person's chromosomal sex is controlled by genetics (XX vs. XY chromosomes), but a person's gender may be influenced by developmental and cultural factors. For instance, some people born XY appear to be women because of a defect in the enzyme that converts testosterone to dihydrotestosterone.
Score this by asking the subject: Male vs. Female, (or transexual, if they choose to identify themselves as such).

Hair Color is controlled by multiple genetic and environmental factors as well. A single, dominant gene controls whether the person has black hair, (BB or Bb) or other-colored hair (bb). A single, recessive allele, controls whether a person has red hair(RR or Rr), or other colored hair(rr). In the absence of these two hair colors, (ie., bb and Rr), the interaction of several loci cause a range of colors between blonde and dark brown. In addition to genes, a person's environment affects their hair. Sun bleaches brown hair to blonde, if it is light enough. Age causes hair to become gray, as cells that produce the pigment loose their ability to function. Most people have also considered coloring their hair at one point or another.
Score your subject: Black, Red, Brown, Blonde, or Other.

Taste PTC-Phenylthiocarbimine is a chemical that some people can taste and other people cannot. The ability to taste PTC is caused by a dominant allele that codes for a receptor for the chemical (PP and Pp). Homozygous recessive individuals (pp) cannot taste PTC. DO NOT TASTE THE STRIP if you are pregnant/might be pregnant (If this is a sensitive issue to you, tell the interviewer you already know you can taste it and you hate the taste. Likewise, if you really already know and hate the taste, you do not have to taste the strip either).
Score each subject Taster vs. Nontaster.

Hitchiker's Thumb-The ability to bend the tip of one's thumb back past the 90 degree angle with the rest of the digit, is due to a recessive allele. Homozygous recessive individuals (tt) have hitchiker's thumb. Tt and TT individuals have normal thumbs. Score each subject Hitchiker vs. Not Hitchiker

Taste of Cilantro-Some people like the taste of cilantro. Other people hate it. There is actually a genetic polymorphism for a receptor that influences the way a person senses the taste of cilantro. Nobody knows for sure, but it is probably a codominant trait. Individuals with one or more copies of one variant tend to taste it as "soapy" or unpleasant. Ask your subject if they like cilantro. Force them to answer as yes or no.

Measurable variables. Pick two of the following and measure or interview everyone in the class. You may pick others, but DO NOT pick weight, people are touchy about that. People can be a little touchy about age and shoe size, too, so do not balk or raise an eyebrow if you do not believe a person's reported measurement-simply record it.

Discontinuous Measurable Variables

Shoe Size is clearly a cultural construct. Shoes come in sizes because it would be impractical to make shoes in a spectrum of sizes. Shoe size is a discontinuous measurable variable because there are a discrete number of shoe sizes (half sizes are discrete categories just like full sizes, there is no 9.357 shoe size, for example). It reflects foot size, however, which is controlled by a variety of genetic and environmental factors affecting overall body size and growth of different parts of the body.

Interview each subject and ask their shoe size.

Number of languages a person speaks. This is largely cultural, though it is quite possible that certain genetic factors influence a person's ability to learn language in a given environmental context.

Interview each subject by asking them how many languages they are able to speak well enough order take-out food over the phone.

Continuous Measurable Variables

Middle finger length-This is a continuous measurable variable determined by a variety of genetic and environmental factors determining overall body size and the relative proportions of different parts of the body.

Height-measure your subject in cm. There should be a series of height marks on the wall of the lab room. This is a continuous measurable variable determined by a variety of genetic and environmental factors as well.

Age-if you think about it, age is entirely environmental. Since this is a laboratory for an introductory biology course, the ages of your classroom will not be a random sample of the population at large. A variety of genetic factors influence a person's ability to live long enough to live a given age, however, so there is a genetic influence.

The attributes you studied were:

The measurable variables you studied were:

Enter your data into the table below:

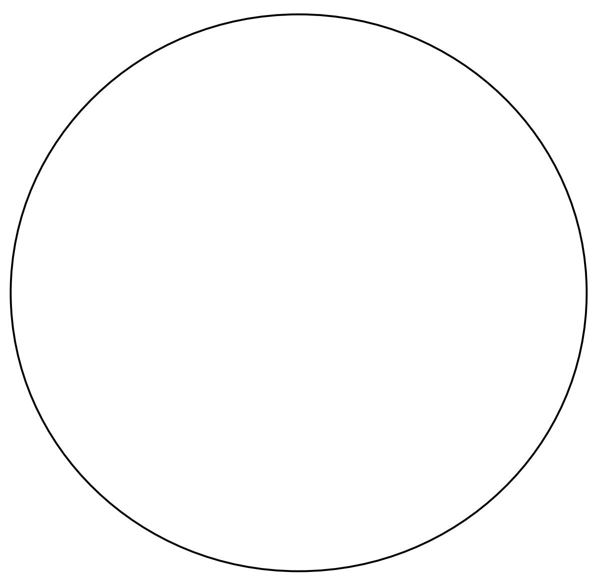
Subject #	Gender	Cilantro	Attribute Value	Measurement 1	Measurement 2

Find the **Mean, Median, Variance**, and **Standard Deviation** for one of your measurable variables. Also list any modal values.

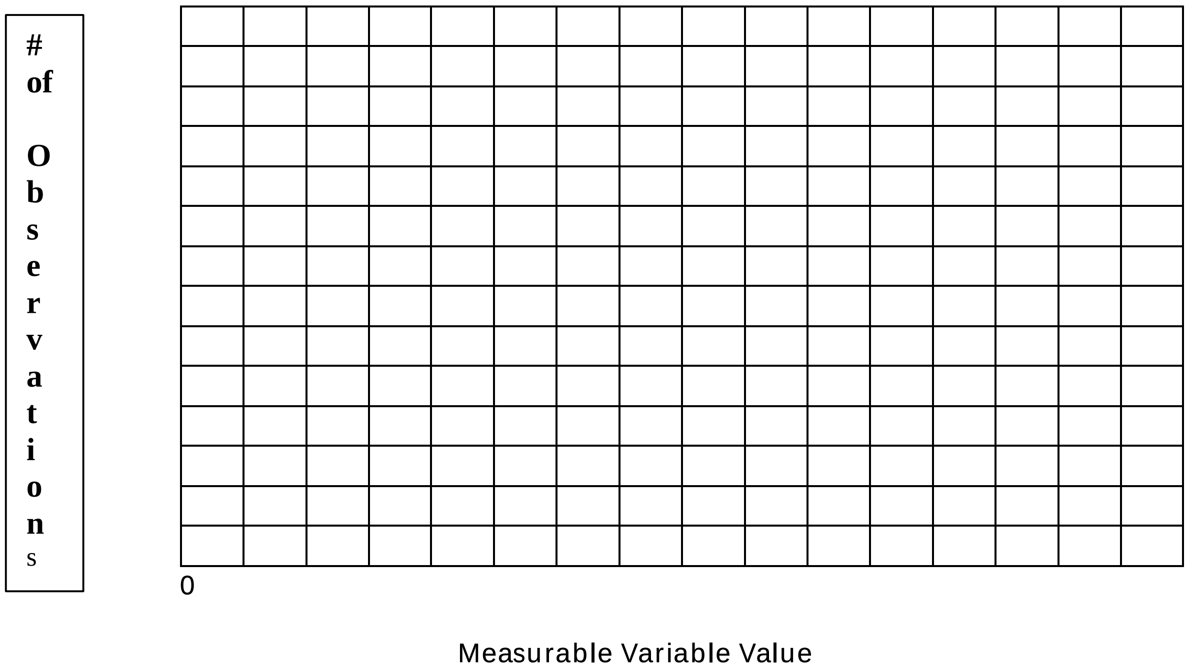
For one the attributes, give the frequencies of each of your phenotypes:

Attribute Value	Number of Observations	Phenotype Frequency

Make a pie graph of the distribution of one of the attributes you studied for your population. Since the figure is hand-drawn, an approximate figure is fine:



Make a histogram for the distribution of one of your measurable variables.



Do a Chi-Square Test to determine whether GENDER influenced the other attribute you studies.

i) Your null hypothesis will be that there is NO DIFFERENCE.

ii) Thus, to generate your expected values, find the relative proportions of each gender in the class (divide the number of males divided by the total, and the number of females by the total).

iii) Take the total number of people who liked cilantro and multiply it by the relative numbers of males and females, respectively These are your expected numbers of males and females that like cilantro.

iv) Next, multiply the relative numbers of males and females by the total number of people that dislike cilantro. These are your expected numbers of males and females that dislike cilantro.

v) Find the Chi-square statistic. Is there a significant difference?

vi) If any of your expected values is under 5, this result might not be meaningful, because of limitations of the Chi-square test.

	Male		Female	
	Likes Cilantro	Dislikes Cilantro	Likes Cilantro	Dislikes Cilantro
OBSERVED	______________	_______________	_____________	____________________
EXPECTED	______________	________________	_____________	____________________

More Statistics

Your laboratory instructor will show you how to do an analysis of variance using the computer spreadsheet program MS Excel. After class, go down to a computer lab or find a computer at home and do these tests. Print the results and bring them to discussion next week.

Perform a T-Test. Use gender attribute as the independent variable and one of your measurable variable as the dependent variable. Is there a significant effect gender on your measurable variable? Why might this be the case?

Perform a linear regression analysis between your two measurable variables. Is there a significant correlation between the two? Why might this be the case?