Biol 213 Genetics (Fall 2000)
Mendelian Genetics

You've devoted much of your education to developing mathematical skills. The reason is that mathematics provides one of the most powerful windows into reality that we have available. This lesson is often lost, however, and the various cosines and vectors may seem utterly divorced from your common experience.

Much of what we do in the next couple of weeks will make use of mathematics. This will initially dismay some, those who have associated the subject with distant abstractions. If you approach Mendelian genetics without mathematics, then you will not be a participant in it. If you use formulas blindly, then you will not participate very adeptly. ***The life scientist who does not use quantitative thinking sees the world half blind.***

The focus this week is to convince you that *simple arithmetic* and *intuition* will enable you to harness the power of mathematics to see deeply into common genetic situations. Don't let numbers push you away. If you go beyond abstraction and use numbers to make the world more concrete, you'll see that you have had for years potent tools with which to analyze your everyday world.

Outline

I. OVERVIEW

After five or six weeks into the semester, we've finally reached the beginning of the book. Why this route?

You've all heard the story: Gregor the monk, toiling in his pea patch, arriving after many years of labor at what we now recognize to be the shining truths of segregation and independent assortment of traits; rejection by his obtuse contemporaries; his final apotheosis many years after his death. Well, forget it. Those contemporaries were actually some really sharp guys, and if they thought Mendel was crazy, you can be sure they had their reasons.

What they DIDN'T have (and you do) is five weeks of Biol 213. With those lessons in hand, we can appreciate Mendel's insights in fewer than the 30 years it took history. Let's take stock.

DNA: Linear sequence of nucleotides. Purely informational. Self replicating.

Protein: Linear sequence of amino acids. Determines form / function of cells.

Gene: Region of DNA that determines a protein.

Allele: Specific version of a gene.

> A brief interlude to underscore the difference between gene and allele, a distinction that eludes many. If you and I compare our possessions, we'd find we have, to a large extent, the same things. We both have writing implements, we both have backpacks, and so forth. My possessions are homologous to yours, performing similar functions. But they are not identical. You may have a felt tip pen and I a pencil. My backpack has a hole in its lower left corner, yours may not.
>
> Similarly, it's a sure bet that we have virtually the same genes, since our human cells have the same functional requirements. But they are not identical. Your version of hexokinase might be slightly different from my own. Perhaps it works faster, or perhaps it has a different amino acid in a position that does not affect enzymatic function. You may be sure, however, that we both have hexokinase and a gene that encodes it. We have homologous genes, but we have different alleles.

SQ1. The rare degenerative condition known as Huntington's Disease is inherited as a dominant trait, and so people affected are invariably heterozygotes. You want to clone the DNA responsible for Huntington's Disease. You isolate DNA from an affected person, but you better be sure that you clone the right:

A. Chromosome
B. Gene
C. Allele

Chromosome: Highly organized form of DNA.

Homologous chromosome: Chromosome carrying the same genes but perhaps different alleles. Chromosome 7 that I obtained from my mother is homologous to Chromosome 7 that I obtained from my father, but they are not identical.

Meiosis: Cell divisions that result in gametes with half the number of chromosomes as somatic cells -- one member of each homologous pair.

With these insights in mind, let's finally turn to Mendel.

II. SIMPLE MENDELIAN CROSSES (PART I)

II.A. The cross and its terminology *(pp.16-23)*

The text takes a historical approach to Mendel's experiments. I'm going to shadow the story from a retrospective viewpoint. Consider Figure 2-4.

SQ2. At the top of Fig. 2-4, two plants are presented with a single character difference: purple vs white flowers. Supposing that this character is determined by a single gene, do the two types of plants have the same gene? Do they have the same alleles?

SQ3. Make up a one-letter name for the gene determining flower color. Like all genes, this one has a specific location on a chromosome. How many copies of this gene are in somatic pea cells? ("somatic" refers to non-gamete cells) How many copies of this gene are in pea gametes resulting from meiosis?

SQ4. Make up one-letter names for the two kinds of alleles of the gene determining white or purple flower color. Write down the genotype (with respect to flower color) of each of the two plants. Below that, write down the genotype of gametes produced by each of the two plants.

Genotype is a listing of alleles under consideration for an organism or cell. Perhaps it should be called "allelotype", since a listing of the genes would not distinguish one pea from another. In principle, one could generate the genotype directly from the organism's DNA sequence.

Phenotype is a listing of the characters under consideration for an organism. It cannot be determined from the organism's DNA sequence, since DNA determines only protein, not the effect of the protein. If you dunk a white flower in purple paint, you've changed its phenotype.

SQ5. Write down all possible combinations of the gametes you listed in SQ4. one gamete from the purple flower and one from the white. In other words, write down all possible genotypes of the progeny of a cross between the two plants. Label the progeny F_1.

According to the data on p.22, all the F_1 progeny have purple flowers ("F" stands for *filia*, "children" in Latin -- after all, Mendel was a monk). Why all purple? Suppose that the gene under consideration encodes an enzyme that catalyzes the production of purple pigment. The enzyme is normal in purple plants but doesn't work in white plants. If we symbolize the allele that encodes good enzyme as *P* and the allele that encodes bad enzyme as *p*, then the F_1 progeny has a genotype of *Pp*, having gained the *P* allele from the purple mother and the *p* allele from the white father.

SQ6. Will pigment-producing enzyme be made in a plant with a *Pp* genotype?

We've come across two kinds of genotypes with respect to a gene:

Homozygous: Both homologous chromosomes carry the same allele.

Heterozygous: Each homologous chromosome carries a different allele.

Mendel coined two terms to describe the phenotypic interaction between two alleles:

Dominant allele: The phenotype determined by this allele is evident in heterozygous progeny.

Recessive: The phenotype determined by this allele is absent in heterozygous progeny.

SQ7. If one allele encodes an active enzyme and another an inactive enzyme, which would you expect to be the dominant allele?

II.B. The monohybrid cross

Continue with the experiment shown on p.21, crossing two tall F_1 plants with each other.

SQ8. Write down the genotypes of the two F_1 parents and the gametes they produce. Does it matter which F_1 individuals you choose or even whether you cross an individual with itself (i.e., self-fertilization)? Remember this answer when you do your own F_1 cross with flies.

SQ9. Write down all possible combinations of the gametes of the F_1 parents. This is conveniently done in the form of a square, to make sure you haven't forgotten any possibilities.

Each heterozygous parent contributes one of two alleles, because meiosis separates the two homologous chromosomes and apportions them randomly into the gametes. There are thus 2x2 = 4 possible progeny:

Table 1: F_2 Progeny of a F_1xF_1 monohybrid cross

		Female	
		P	***p***
Male	***P***	*PP*	*Pp*
	p	*pP*	*pp*

SQ10. Circle the genotypes that you would expect to result in purple petals. What ratio do you expect between progeny with purple petals to those with white petals?

Mendel performed many monohybrid crosses, and the results he obtained are shown in the table on p.22. In the cross we're considering between purple and white plants, he found that 705 F_2 progeny of the F_1xF_1 cross had purple flowers and 224 had white flowers. The ratio of one to the other is 3.15 to 1. All of the ratios hovered around 3 to 1, and it is tempting to believe that the true ratio of all the crosses is 3 to 1. 3.15 to 1 is close to 3 to 1, but *is it close enough?* Hold on to that thought, for we'll spend considerable time answering that question in a moment.

Perhaps the most distinctive crosses that Mendel performed were those of the third set, in which individuals of the F_2 generation were self-crossed, producing an F3 generation (Fig. 3). If the F_2 generation was more complicated than the monocolor F_1 generation, would the F3 generation be still more complicated? No, he found that self-crossing all F_2 individuals that had a recessive phenotype gave rise to progeny that also had the recessive phenotype. In other words, the recessive phenotype bred true. Once recessive, always recessive. What about the F_2 individuals with the dominant phenotype?

SQ11. What phenotype(s) would you expect from a self-cross of a randomly chosen purple F_2 progeny shown in Fig. 1?

Sometimes F_2 individuals exhibiting the dominant phenotype bred true, meaning their progeny and grandprogeny always had purple flowers, but sometimes not. In many cases, purple F_2 individuals produced progeny that were a mixture of purple and white individuals. When this occurred, the ratio of progeny with purple flowers to those with white was always about 3 to 1. Mendel's genius was in noting the quantitative relationship between those F_2 purple plants that bred true and those that gave rise to mixed progeny (in the coming weeks you will show similar quantitative genius with flies). The ratio of hybrid F_2 dominants to pure-breeding F_2 dominants was 2:1, and so the 3:1 ratio of dominant to recessive progeny in the F_2 generation could now be seen more precisely as a 1:2:1 mixture of pure dominant : hybrid dominant : recessive progeny.

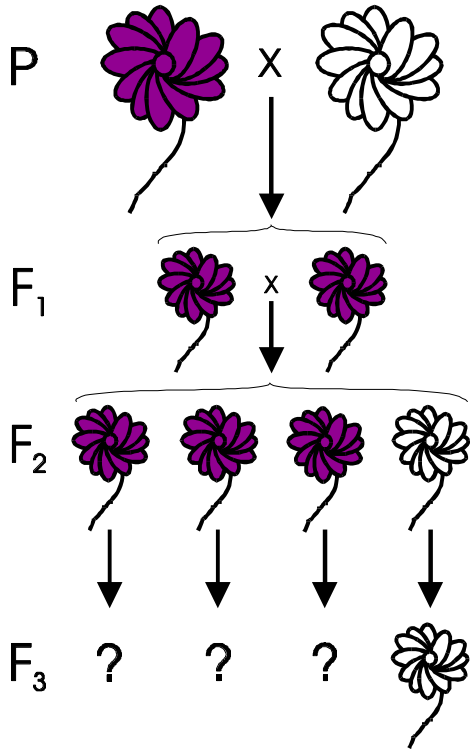

Fig. 1. F_3 generation of monohybrid cross. All crosses are self-crosses except that between the two parents.

SQ12. Return to the square you modified in SQ10. Circle <u>twice</u> F_2 progeny with the dominant phenotype that you would expect to breed true. Write down the ratio of the number of boxes twice circled to the number once circled to the number not circled at all.

The ratio you arrived at by counting boxes is valid only if each of the four boxes is equally likely to occur. This is true only if each gamete genotype is equally likely, with a probability of 50%. Mendel, totally ignorant of both chromosomes and meiosis, arrived at this conclusion in his first law, which may be stated:

> <u>Law of Segregation</u>: Heterozygotes produce equal numbers of the two alleles.

SQ13. Restate the Law of Segregation in the language of chromosomes.

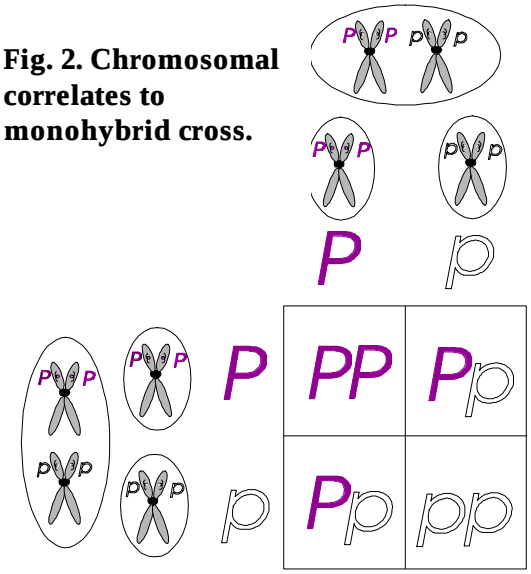

Fig. 2. Chromosomal correlates to monohybrid cross.

III. EVALUATION OF GENETIC RESULTS: CHI-SQUARED TEST *(pp.37-40)*

III.A. Rationale behind the chi-squared test

Mendel obtained ratios that were close but not identical to 3:1 and concluded that they were close enough. Was this conclusion justified? How can we decide? OK, you might be willing to cut Gregor a break this once, but time and again the similar genetic question arises, because ratios give us information about underlying mechanism. Do the numbers indicate that a certain disease is inherited? How much should my patient rely on a test result that seems to indicate that she carries a deadly genetic trait? We need a way of assessing the truth.

Back to Mendel. Was an F_2 generation of 705 purple and 224 white plants reasonably described by a 3:1 ratio? Let's do a thought experiment. Suppose we did a monohybrid cross between a purple and white plant a million times and each time noted the flower color of 929 F_2 progeny. Plotting the results, we might get something like as that shown in Fig. 3.

We get a normal curve centered about 75% of 929 (=696.75) and 25% of 929 (=232.25).

SQ14. Why is it that the number of purple plants and the number of white plants in the specific experiment considered (dotted line) are on opposite sides of the mean?

SQ15. What is the most likely experimental result? How often (out of the million trials) does that result occur?

We can now ask a simple question: How probable is the result Mendel got? What percent out of the million random trials would we expect to get precisely 705 purple and 224 white plants? If you read off the frequency and divide the number by a million, you'll get a frequency on the order of 1%: not too likely. But that procedure isn't fair. ANY specific number -- even 697, as close as we can get to exactly 3:1 -- is infrequent. It clearly is not very meaningful to ask merely how probable is a specific result.

SQ16. If you flip a coin a 1000 times, is it likely that you'll get precisely 500 heads and 500 tails? Is it likely that you'll get at least 500 heads?

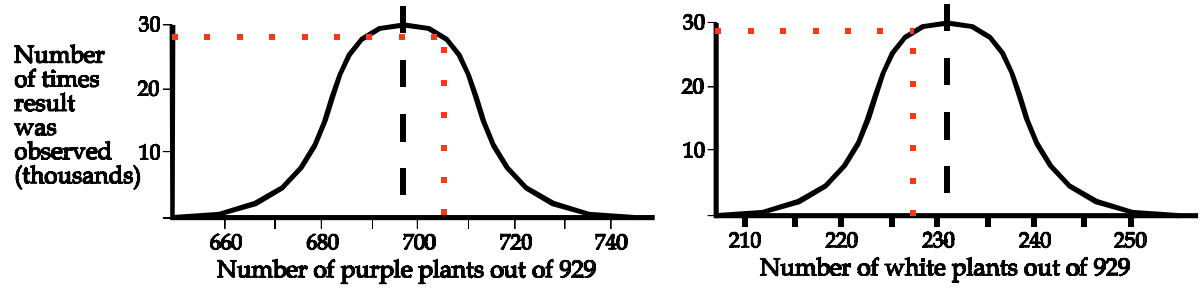

Fig. 3: Distribution of results from a million monohybrid crosses. 929 progeny were scored from a million crosses between pure breeding purple and white plants. The dashed line indicates the most common result (also the mean number averaged over all trials). The dotted line represents the result from one specific experiment.

But the task is not hopeless. The probability of a specific number of outcomes is meaningless, because it depends on how finely the possibilities have been chopped. But it is meaningful to ask how likely it is that Mendel's results deviated from the expected value (**E**) by at least the deviation he observed. The observed value (**O**) deviated by (705-696.75) = 8.75 plants. The curve in Fig. 4 (left) has shaded in those values that deviate by at least that amount.

SQ17. Why is there shading on both sides of the curve in Fig. 4?

The shaded area takes up about 60% of the area under the curve. This means that 60% of the time Mendel would have observed a number of plants at least as far from the expected value as 705. Obviously, we can't complain about a number being unreasonably far away from the number that would have produced an exact 3:1 ratio if such a number (or worse) occurs over half the time.

On the other hand, if he had obtained a more distant value of 725 purple and 204 white plants, then we would have cause for concern. The curve above (right) shows shaded those values that deviate from the expected value by at least as much as does 725. That area takes up only 3% of the curve, so in a large number of trials, only 3% of the time would Mendel have gotten values as distant from expected as he did. Mendel could have been unlucky, but it is also possible that the premise (3:1 ratio) is wrong.

If we can measure the shaded area under the curve, then we can answer the question: Is an experimental result reasonably close to that expected from theory? ...but how to measure areas? Fortunately, all has been done for us in this regard. There are voluminous tables about areas under the curve. All we have to do is tell the table how far our observed value is from that expected.

SQ18. Recall that the shaded area covers values that are as far or farther from the expected value than the observed value. The further away an observed value is from the expected value, the smaller/larger (choose one) is the shaded area under the curve?

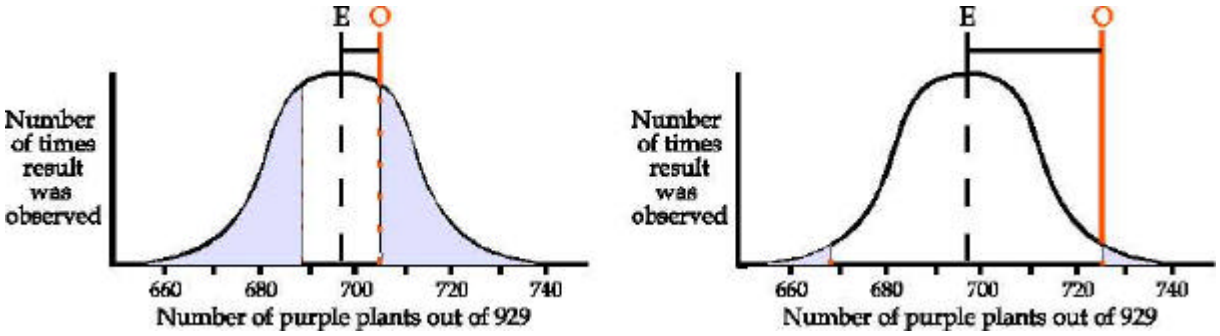

Fig. 4: Area under the curve of distribution of results. E is the expected number of purple progeny out of 929 F_2 progeny scored. **O** is the observed number of purple progeny for a specific experiment. The shaded portion of the curve represents the number of experiments that have deviations from expectation greater than that of the experiment under consideration. **(Left curve) O** = 705 purple F_2 progeny. **(Right curve) O** = 725 purple F_2 progeny.

It sounds simple -- just tell the table how far is the observed value from that expected (**O - E**) -- but it isn't quite.

SQ19. How far is 725 purple plants from the expected value, if you anticipate a 3:1 ratio of purple plants amongst 929 plants?

SQ20. Suppose Mendel had counted not 929 but 9290 plants, and the observed value was the same distance (same number of plants) from the expected values as in SQ19. Would the distribution curve and shaded values look different from that shown in Fig. 2 (right curve)?

Evidently, mere distance is not a good measure of how unlikely a value is. The problem is that a 30-plant difference is a lot if you were expecting 695 plants but not much if you were expecting 6950. We need to compare the difference to the expectation. This is done by the formula:

$$(\mathbf{O - E})^2 \ / \ \mathbf{E}$$

Note that squaring the distance removes the negative sign if the observed value is smaller than the expected value. It's the distance that counts, not the direction.

III.B. Steps in performing a chi-squared test

The χ^2 (Chi-squared) test is the embodiment of the strategy we've just employed to determine the reasonableness of a result with respect to prior expectations. In essence, it is a procedure to identify the area under a distribution of possible results representing those that are worse (compared to expectation) than that observed. The procedure:

1. Determine the expected values for the experiments. You need to have a model in hand, e.g. "The phenotypes should stand in a 3:1 ratio", and on that basis calculate the expected number for each condition.

SQ21. What are the expected values in the experiment we've been considering, where 929 plants are counted and the ratio of purple to white plants is presumed to be 3:1?

2. Calculate the squares of the deviations (related to the distance between the observed value and the expected value) for each variable.

SQ22. What is the square of the deviation calculated for 725 purple flowers observed with an expected number of 696.75?

SQ23. What is the square of the deviation calculated for 204 white flowers observed with an expected number of 232.25?

3. Add up all the normalized deviations, producing X^2:

$$X^2 = \text{Sum of } [(\mathbf{O - E})^2 \ / \ \mathbf{E}]$$

SQ24. What is X^2 for the observations of 725 purple flowers and 204 white flowers, with an expected ratio of 3:1?

4. Determine the degrees of freedom (**df**). The degrees of freedom is the number of independent variables you are considering. Be careful! There's generally one fewer than you think. For example, in the purple x white cross we've been considering, there is only one independent variable, because we set the number of plants counted (at 929 plants). Thus if you know the number of purple plants, that gives you the number of white plants. The degrees of freedom in this problem is therefore one.

SQ25. Suppose you are assessing the proposition that different majors on campus attract more females than males or vice versa. For each major, you note how many are females and how many are males. How many degrees of freedom are there?

5. Use the χ^2 value and the degrees of freedom to find the probability that a result exceeds the observed deviation from expectation. On a table of χ^2 values (e.g. Table 5-4 in the text), find the line with the appropriate degrees of freedom and on that line the χ^2 value you calculated in Step 3. Note the probability **p** that heads the column in which you find the χ^2 value. You're not likely to find a χ^2 value that exactly matches yours but rather two numbers that flank your own. This will give you a range of probabilities.

SQ26. Which best describes the significance of the probability or range of probabilities found in Step 5?

A. It is the probability that your result is correct
B. It is the probability that your hypothesis is correct
C. It is the probability that your result arose by chance
D. It is the probability that your result (or worse) could have arisen by chance if your hypothesis is correct

SQ27. Does a high p value indicate that your hypothesis is correct? Make up a situation where this would not be true.

SQ28. Does a low p value indicate that your hypothesis is incorrect? Make up a situation where this would not be true.

Note that pp.38-39 in the text provides a worked out example of a χ^2 test.

SQ29. Although dwarfism is a rare condition in humans, it is nonetheless not so uncommon to encounter a productive union between two dwarves. Suppose you consider the progeny of many such unions and find that 132 are also dwarves and 60 are of normal stature.

a. Does dwarfism appear to be dominant or recessive?

b. Make up appropriate symbols and write down the genotypes of the parents and the different types of progeny.

c. Perform a c^2 test to assess whether the ratio of progeny can reasonably be described as 3:1 and draw a conclusion.