Module III An Introduction to Hypothesis Testing/ Chi Square Test

Upon completion of this lab, you will be able to:

- Understand Mendelian genetics and segregation
- Be able to discern phenotypes and, using this knowledge, hypothesize parental genotype and phenotype.
- Deduce genotypic and phenotypic ratio of offspring
- Use Chi square statistical analysis to confirm or reject hypotheses

Background: When we observe a ratio of different phenotypes, we can try to explain these with some hypothesis. For example, in humans, the sex chromosomes segregate for sex determination, offspring which are XX are female, and XY individuals are male. The sex ratio in humans should thus be 1:1. Although you know that theoretical expectations are 1:1, you could not reasonably reject this hypothesis if a sample of 100 contained 48 females and 52 males. What about 60:40, or 75:25? The problem then becomes how to test this hypothesis.

The idea behind hypothesis testing is that the observed deviation could come from two possible sources:

1. Chance or random variation

2.Real phenomena not considered (that is, perhaps the hypothesis is wrong).

To separate these two possibilities, **we will determine the probability that the deviation <u>could</u> be due to chance using the chi square test** (= X^2), The chi square test is one of the most widely used statistical tests. We will use the chi square test to evaluate our hypotheses.

In genetics, and in much of biology, the arbitrary <u>convention</u> is that, if an observed an outcome <u>has less than a 5% (= 1/20) probability of occurring by chance alone, that event is described as "unexpected" or as "a SIGNIFICANT deviation</u> from expectations". Nothing more is implied. Keep in mind that when an event is "unexpected", this is not to imply that it is "impossible", just improbable.

In this lab, we will be dealing with the first law of Mendelian genetics, the law of segregation. **This indicates that, after coming together in a hybrid, two forms of a gene (alleles) can separate (segregate) with no change in their original quality.**

We will be plating and observing tobacco seedlings that are the progeny of a self-fertilized parent. The progeny potentially, will have three different phenotypes of color on their cotyledons (first leaves): yellow/ white, light green, and dark green.

Procedure:

Day1: Plate Nicotiana Tabacum (tobacco) seedlings according to instructions given by the instructor.

Day 14: Examine the tobacco seedlings.

Randomly count approximately 100 seedlings for color by pulling them out using tweezers and organizing them on a sheet of paper. If you have less than 100 seedlings on your plate, count all.

Answer the following questions in your lab notebook; include your complete chi square analysis!

1. What is the observed **phenotypic** ratio (setting the lowest number to 1)?

____ DARK green:_____PALE GREEN:_____WHITE.

2. What are probable **genotypes** of DARK green, PALE−green and WHITE seedlings?

3. What are the probable **genotypes of the parent** of the seedlings?

4. What is the probable **phenotype of the parent plant**?

5. Write a hypothesis, including the hypothesized genotypes of the parent and the expected ratio of the seedlings that you counted, then use the table on the next page to test you hypothesis using the chi-square statistical method.

Answer the above questions in your lab notebook.

Use the chi square method to determine if the observed phenotypic ratios deviate significantly from expectations (aka your hypothesis) and if so, with what level of significance.

As an example of an appropriate answer for the latter part of the question, a chi sq. of 3.00, with 1 degree of freedom (df). would be described as, "a probability between 5 and 20%" or "0.20 > P > 0.05" or "Not Significant". A chi sq. of 3.90, with 1 df. would be "P < 0.05" or "Significant at the 5% level".

Start your analysis by filling in the table:

Make sure to put this table in your notebook as your results.

Phenotype	Observed	Expected	Deviation	$(\text{Deviation})^2$ /Expected
Dark Green				
Pale Green				
White				
Totals				

Chi sq. = d.f. = p=

Do you reject or accept your hypothesis? Explain!

This should be your conclusion for this lab exercise.

SELECTED CRITICAL VALUES IN THE DISTRIBUTION OF CHI– SQ.

Degrees of freedom	PROBABILITY (P)						
	0.95	0.80	0.50	0.20	0.05	0.01	0.005
1	0.004	0.064	0.455	1.642	3.841	6.635	7.879
2	0.103	0.446	1.386	3.219	5.991	9.210	10.579
3	0.352	1.005	2.366	4.642	7.815	11.345	12.838
4	0.711	1.649	3.357	5.989	9.488	13.277	14.860
5	1.145	2.343	4.351	7.289	11.070	15.086	16.750
6	1.635	3.070	5.248	8.558	12.592	16.812	18.548

Example of how to work through the math of the Chi Square Test:

Example calculation and evaluation of a different experiment with the hypothesis that two leaf colors will be observed at a 3:1 ratio.

Calculation of X^2 for a monohybrid ratio (3:1)

Phenotype (class)	Observed	Expected	Deviation	$\frac{(\text{Deviation})^2}{\text{Expected}}$
Wild type	99	108	−9	0.75
Mutant	45	36	+9	2.25
Totals	144	144		**3.00 = X^2**

The expected frequencies were calculated with the 3:1 ratio as the expected ratio. Since there were 144 individuals observed (99 + 45), three fourths of the 144 (= 108) should be wild type, etc. $(-9)^2/108 = 0.75$, *etc. THE SUM (3.00) OF the two* D^2/E *values IS THE CHI SQUARE VALUE*

To calculate the degree of freedom, subtract one from the number of classes of observation. In the example above, there are 2 classes of phenotypes, wild type and mutant, thus there is 1 "degree of freedom" (d.f. = number of classes − 1). Mathematicians have kindly calculated (See the table above) that, with 1 degree of freedom, a chi square value of 1.642 will occur, BY CHANCE, with a probability of 20%.

A chi square value of 3.841 will occur, 5% of the time, etc. A chi sq. of 3.00 will occur less than 20% of the time, but more than 5%. Thus the observed deviation from expectations (3:1) is not a "significant" one. Thus we have no reason to reject the hypothesis that the ratio is indeed 3:1. Only if the chi sq. had been greater than 3.841, could we have said that the deviations from expectations were significant.