LABORATORY 6. THE CHI-SQUARE (X^2) TEST: A STATISTICAL TEST FOR EXPERIMENTS

LEARNING OUTCOMES

When students have completed this laboratory practice, they should be able to:

- Explain how a dihybrid cross can be used to test Mendel's law of independent assortment.
- Propose a scientific hypothesis (i.e., null hypothesis) for a genetic cross.
- Generate simulated F_2 plant phenotype data using R.
- Calculate expected values from simulated observed data.
- Use a chi-square test to determine whether simulated F_2 data fit Mendelian expectations.
- Interpret chi-square results and decide whether to reject or fail to reject the null hypothesis.

INTRODUCTION

Observed outcomes from genetic crosses rarely match Mendelian predictions exactly. These differences can arise for two main reasons. **First**, deviations may occur due to random sampling variation (chance alone). **Second**, the assumptions used to calculate the expected values may be incorrect. If the assumptions are invalid, then the predicted ratios must be revised to better reflect the underlying genetic model.

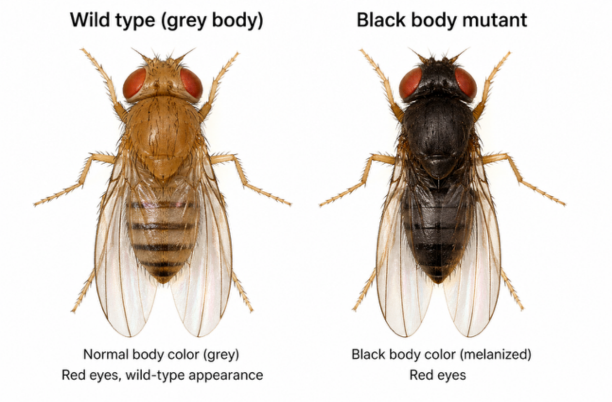
For example, in several previous genetics exercises in this lab (e.g., sex-linkage), we predicted the outcome of a particular cross without having observed data. In most cases, these predictions were based on the **Mendelian principles** of segregation and independent assortment.

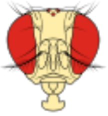
After obtaining and phenotyping the offspring, the key question becomes: Do the observed results fit the predicted outcomes? To answer this, we compare the expected and observed values to evaluate our predictions.

In this context, the **null hypothesis (H_0)** states that there is no difference between the observed and expected data. In contrast, the **alternative hypothesis (H_1)** states that there is a significant difference between them.

To determine whether any observed differences are meaningful, we use **a chi-square (χ^2) statistical test**. This test allows us to assess how well the observed data fit the expected ratios and to decide whether to reject or fail to reject the null hypothesis.

Imagine that you would like to test if the black body color in *Drosophila* follows a Mendelian inheritance pattern. The following offspring was produced:





P_1= males, **normal** body X **P_2**= females, **black** body

F_1= All **normal** body

F_1 X F_1

F_2 progeny:

Normal body	**102**
Black body	**44**

1) THE NULL HYPOTHESIS (H_0)

In this example, if the segregation of alleles responsible for the black body color follows **Mendel's first law of segregation (F_2 3:1)**, there should be **no** statistical **difference** between the observed and the expected phenotypic ratios. That is, the 102 normal: 44 black body should be very close to a 3-normal to 1-black body ratio.

THE ALTERNATIVE HYPOTHESIS (H_1)

As simple as indicating the opposite of the H_0, the H_1 would state that the **F_2** progeny (102 grey: 44 black body) **does not fit** a ratio of 3-grey to 1-black body. As mentioned previously, the **X^2** statistic will allows us to determine whether the null hypothesis (H_0) is valid or appropriate.

2) THE CHI-SQUARE FORMULA

The following formula is used to calculate the **X^2** statistic

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

The **"observed" data** represents the number of individuals found in a phenotypic class, while the **"expected" data** represents the number of individuals that the null hypothesis **predicts** for that phenotypic class. Notice that the equation is based on the difference between the observed and the expected data, and therefore, the greater the difference, the higher the value of the statistic. For the calculations, please consider that it is necessary to calculate the **X^2** from **actual numbers (amount) rather than percentages or fractions**. A table may facilitate the calculation of the X^2 value:

Phenotype	Observed number	Expected number	O-E	$(O-E)^2 / E$
Grey body (3/4)	102	(146 x ¾)= **109.5**	-7.5	0.5137
Black body (1/4)	44	(146 x ¼)= **36.5**	7.5	1.5411
Total	146			
X^2 value				2.055

3) X^2 INTERPRETATION AND CONCLUSION

Once the overall **X^2** value has been calculated (**X^2=2.055** in our example), it can be used to determine if there is a significant difference between the observed and expected results, under a certain **significance level (α)** (the probability of rejecting a H_0 when it is true, type I error). Researchers choose relatively low significance values (0.01, **0.05** or 0.10), being **0.05** the most used.

As the X^2 distribution changes depending on the **degrees of freedom (df)**, it is essential to estimate this value before drafting conclusions. This value depends on the number of phenotypic classes, and in this example, it is determined as follows:

df = number of phenotypic classes – 1.

There are **two ways** to determine if the outcome of the experiment is statistically significant. **The first method** involves comparing "calculated X^2 values" vs. "distribution X^2 values". These distribution X^2 values corresponding to the selected **significance levels** and the **degrees of freedom** are available from many sources, usually in the form of a table, like the one presented below:

Df/α	**0.1**	**0.05**	**0.01**
1	2.71	**3.84**	6.64
2	4.61	5.99	9.21
3	6.25	7.82	11.35
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.65	12.59	16.81

The numbers on each column are the **X^2** values that represent the **cutoff** point to reject or fail to reject the null hypothesis. To select the X^2 value from the table (critical value), first, we must find the appropriate **df row** and the **significance value** column.

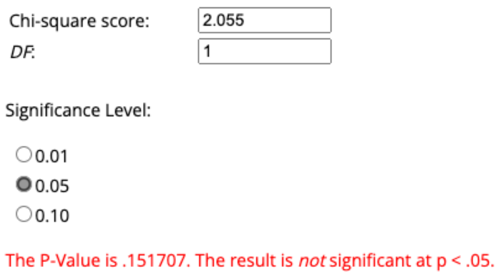
In this example, the df value is **2-1=1**. The X^2 cutoff for **α=0.05** (**critical value**) is **3.84**.

As the **calculated X^2** value= 2.055 **does not surpass** the **critical value from** the table (**2.055**<**3.84**), we conclude that the difference between the observed and expected results is negligible so, **we fail to reject the null hypothesis**. Then, the conclusion is that the observed data seems to fit the expectation of a 3:1 segregation as proposed by Mendel's laws.

The second method to determine if the experiment was statistically significant involves the estimation of a **p-value**. This value is the probability that the observed statistic occurred by chance alone, assuming that the null hypothesis is true. In other words, **the smaller the p-value, the more unlikely the observed results.** An easy way to estimate the p-value for the X^2 distribution and a given degree of freedom is to use online tools, as the one presented in the following link:

https://www.socscistatistics.com/pvalues/chidistribution.aspx

Here, you can type your calculated X^2 value (i.e., 2.055), the significance level **(α=0.05)** and press the "calculate button". The results will be presented as in the picture:



Chi-square score: 2.055
DF: 1

Significance Level:

○ 0.01
● 0.05
○ 0.10

The P-Value is .151707. The result is *not* significant at $p < .05$.

Generally, a study is statistically significant if the **p-value is less than the pre-specified α value** (0.05). A **p-value greater than or equal to α** is not a statistically significant result. In our case, p=0.1517 is greater than 0.05 and hence, it is not statistically significant. Then, the conclusion is that the observed data seems to fit the null hypothesis of a 3:1 segregation as proposed by Mendel's laws.

MAIN MATERIALS

- TLC result from last week
- Laptop computer with access to a browser.
- Access to the **Genetics Virtual Lab** platform here: www.ampossot.com/virtual_lab
- Notebook and pen (or pencil)

PROCEDURE FOR LABORATORY WORK

1) Collect your TLC plate from the previous week and observe the chromatography results under UV light.

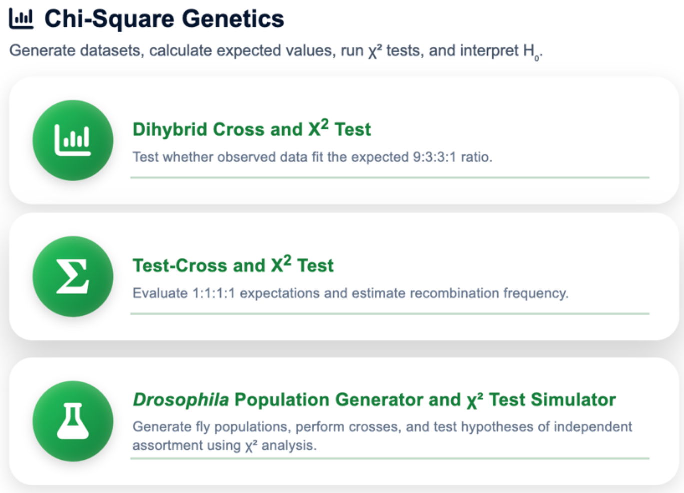
2) Analyze your results. You will need this information to complete your assignment. Take a picture with your mobile device if needed.

3) Collect the phenotypic information from the F_2 generation (posted in front of the lab).

PROCEDURE FOR VIRTUAL DATA GENERATION AND TESTS

1) **Launch the Genetics Virtual Lab**. Access here: www.ampossot.com/virtual_lab

2) You will use **three** virtual tools in this lab:



Chi-Square Genetics

Generate datasets, calculate expected values, run χ^2 tests, and interpret H_0.

Dihybrid Cross and X^2 Test

Test whether observed data fit the expected 9:3:3:1 ratio.

Test-Cross and X^2 Test

Evaluate 1:1:1:1 expectations and estimate recombination frequency.

***Drosophila* Population Generator and χ^2 Test Simulator**

Generate fly populations, perform crosses, and test hypotheses of independent assortment using χ^2 analysis.

3) Follow the instructions on each of the questions on your assignment. In some of the tools, you can **sort the flies** (by sex, phenotypes or genotypes) in the **wild population**.

4) **Perform the number of crosses described in your assignment**. Record the number of offspring phenotypes and **ratios** in the provided templates. An easy way to obtain phenotypic ratios is dividing the **higher-class number** by the **lowest class number**. Example: If you get 60 wild-type flies and 19 Sepia flies, the ratio will be 60/19 wild-type : 19/19 sepia flies. That is 3.15 wild-type: 1 sepia flies.

5) **Complete and submit your assignment by the deadline**. Please **carefully read** the questions in your assignment.