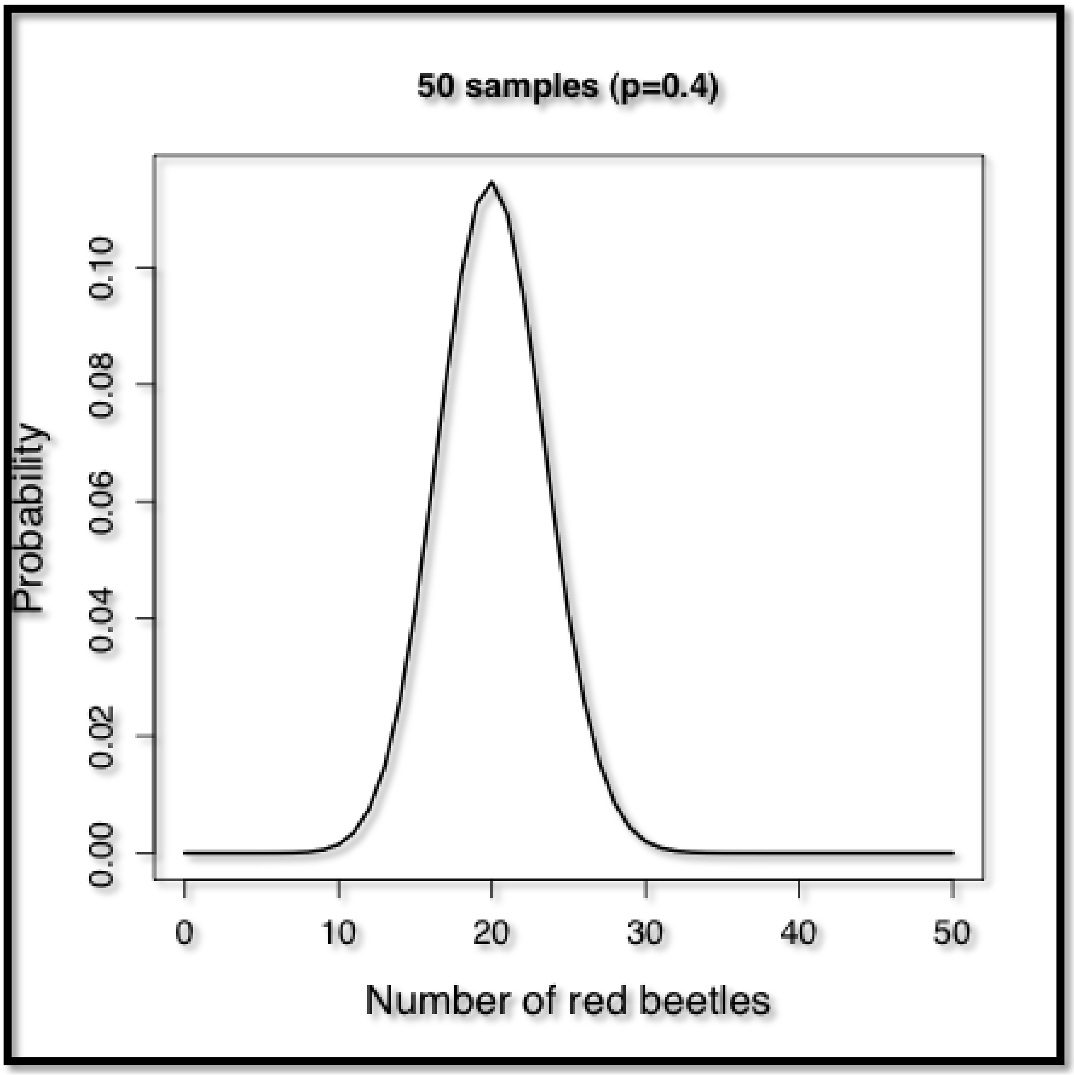
Math in Science Class: Accounting for Uncertainty When Analyzing Count Data

Teacher's Manual



Lauren Lucas, Zach Gompert, Alex Buerkle
Department of Botany, University of Wyoming

Math in Science Class: Accounting for Uncertainty When Analyzing Count Data

Teacher's Manual

Photocopy Masters

Math in Science Class: Accounting for Uncertainty When Analyzing Count Data

Overview

Using this activity students are introduced to a way in which scientists use probability to determine how confident they can be in the scientific claims they are making. Students collect data by sampling beetles (beads) from a "population" where the probability of red beetles is unknown. Students estimate the proportion of red beetles and quantify the uncertainty in their estimate using probability theory. This activity can be used as an introductory activity before starting an experiment involving count data (e.g., determining the probability of larval survival on a particular host-plant) or as a stand-alone activity.

Objectives

In this exercise students will

- collect count data.
- analyze count data and quantify uncertainty using probability theory.
- compare results with peers to recognize how sample size affects the certainty of a point estimate.
- use appropriate scientific and mathematical terminology to construct scientific explanations.
- recognize the connection between math and science.

National Science Content Standards

This activity addresses the following national standards:

Content Standard: Unifying Concepts and Processes
Systems, order, and organization
Evidence, models, and explanation

Content Standard A: Science as Inquiry, Grades 5-8
Use appropriate tools and techniques to gather, analyze, and interpret data
Develop descriptions, explanations, predictions, and models using evidence
Think critically and logically to make the relationship between evidence and explanations
Use mathematics in all aspects of scientific inquiry

Content Standard A: Science as Inquiry, Grades 9-12
Use technology and mathematics to improve investigations and communications
Formulate and revise scientific explanations and models using logic and evidence
Communicate and defend a scientific argument

Science Education Program Standard C
The science program should be coordinated with the mathematics program to enhance student use and understanding of mathematics in the study of science and to improve student understanding of mathematics

Materials

Needed for each group of two students

Zip lock bag

400 red beads and 600 white beads (do not share these numbers with your students)

Calculator

Needed for each student

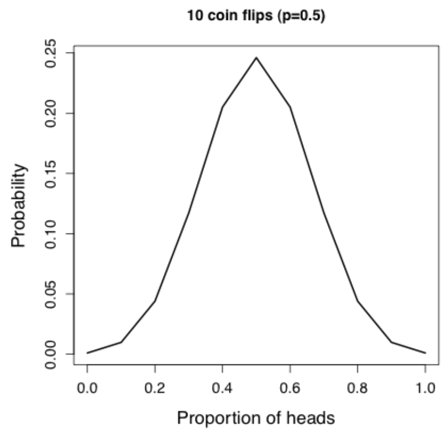
Student Handouts (Pre-Assessment and Activity)

Background

Biology textbooks tend to neglect to mention the variability inherent in the process of science, specifically in measurement and sampling. For example, it is usually impossible to study all the members of a population of a biological system; uncertainty lies in the magnitude of the discrepancy between the sample and population means. Furthermore, variability exists even in processes that are thought to be largely invariant (e.g., deviation of sex ratio in humans during wars). Some teachers delay or avoid data analysis and scientific decision-making because students have difficulty with them (math and statistics are not the favorite of the majority of students). But, we shortchange the student and undermine the goal of achieving science literacy if we defer learning about objective decision making. The trick is to include data analysis in our modeling of science processes without getting lost in the intricacies of statistical analysis. We also suggest you provide your students with ownership; have students analyze data in this way from their own hands-on inquiry project.

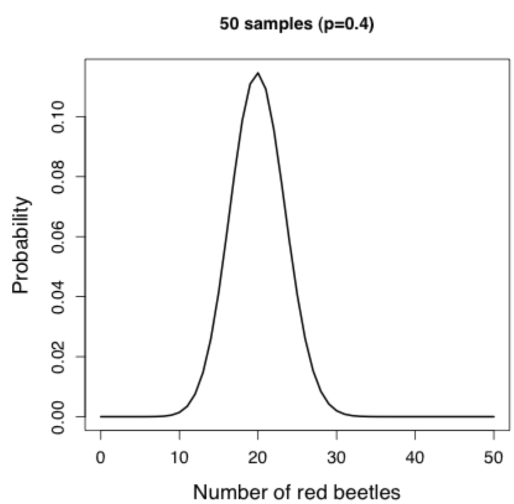
Variability is the degree to which a data set is spread out or closely clustered. **Uncertainty** can arise from our limited sampling of populations. Variability in outcomes leads us to describe processes with **probability** and corresponding models. The following activity only addresses an example of a way to describe uncertainty in discrete, or count, data (i.e., observations with only non-negative integer values such as 0, 2, 4, 3, …) where there are only two treatments (e.g., red and white). In this activity students will analyze count data using the **binomial distribution**, the discrete probability distribution of the number of successes (e.g., drawing a red beetle from a population) in a series of yes/no (i.e., this/that) trials in which the probability of success does not change throughout the experiment. Scientists and mathematicians frequently use the binomial distribution to model the number of successes in a sample (n) from a population (N).

Consider the classic example of a coin toss. We know the probability of success (heads) is 0.5. If we toss the coin 10 times, will we get 5 heads?



Notice there is only a 25% chance of tossing 5 heads, meaning there is a 75% chance of getting an outcome other than 5 heads. Even when we know the probability is 0.5, we cannot predict the exact outcomes of a coin toss.

Often, the probability of an event is not 0.5. Furthermore, often we do not know what the probability of an event is. Instead, we must estimate it. In this activity students will estimate the proportion of red beetles (beads) in a population. Your students will not know the "true" proportion of red beetles prior to the experiment, as they would in a coin toss activity. In other words, they will not know the probability of drawing a red beetle from the population. They will use their observed proportion of red beetles in a sample of 50 beetles, taken from a population of 1000 beetles, as a point estimate of the proportion of red beetles in the population. (A **point estimate** involves the use of sample data to calculate a single value to serve as a "best guess" for an unknown population parameter.) Since students will not be sampling the entire population, they will calculate a **confidence interval**, an interval estimate of the population parameter. It is used to indicate the reliability of the point estimate. As in the coin toss, sampling beetles with some probability of drawing a red beetle leads to a distribution of outcomes.



The Activity

Consider having each student answer questions about a local newspaper article (see Student Pre-Assessment Handout). Most students have misconceptions about how to accurately interpret data and how confident they can be in that interpretation. Once students express their misconceptions, it is more likely that they will challenge and modify or abandon them in the following activity. We suggest revisiting this article and the students' original answers after the activity.

<u>Question 1</u>. The article states, "Statewide, the number of women who smoke while pregnant rose from 22.6 percent to 24 percent, according to the organization's 2010 premature birth report card." Can you be certain that the proportion of smoking pregnant women in Wyoming recently increased? Explain your answer.

<u>Answer 1</u>: *No, because we do not know how much uncertainty is associated with these estimates (i.e., the confidence intervals might overlap, meaning these proportions are not different from one another).*

<u>Question 2</u>. What additional information would you want to evaluate the claim that one in four women smoke in Wyoming?

<u>Answer 2</u>*: I want to know: How was sampling conducted? What was the sample size?*

Each step of the activity is described below with examples of outcomes.

Step 1. The teacher prepares "populations" for students

Combine 400 red beads with 600 white beads in a zip lock bag for each group of students. Do not share these proportions with your students, as they are trying to estimate the proportion of red beetles (beads) in the population. We suggest you connect this abstract example to something in nature, as we have (a population of beetles with a polymorphism in carapace color).

Step 2. Students collect count data by sampling individuals from a population

Explain to students that the null hypothesis will be that the population is composed of 50% red beetles and 50% white beetles.

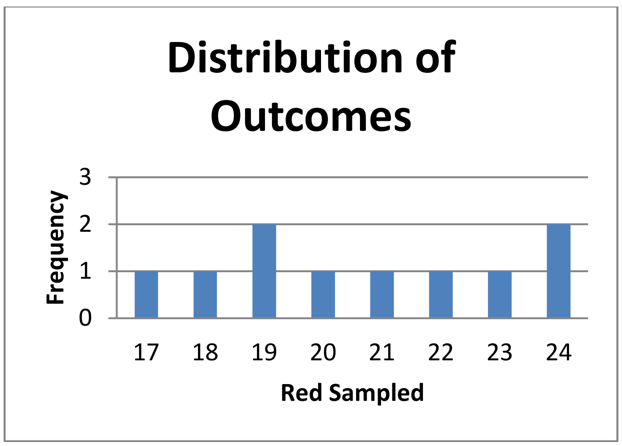
Students start by mixing the beads in the bag to ensure the samples taken from the population are random. Assign roles to students; one student will be the sampler and the other will be the recorder. The sampler will remove one bead at a time from the bag at a time while the recorder keeps track of the number of red and white beetles sampled. The sampler will take a total of 50 beetles from the population.

Step 3. The teacher graphs the class' distribution of successes

Record the number of red beetles sampled by each group. Create a histogram to demonstrate the distribution of outcomes in the class. Have the students think about the fact that given there are multiple possible outcomes for one true probability (proportion of red beetles), we cannot assume 100% certainty in a particular estimate from a single data set.

Example classroom data:

Number of Red	Number of people with red
17	1
18	1
19	2
20	1
21	1
22	1
23	1
24	2



Step 4. The teacher models how to calculate a point estimate and a 95% confidence interval (CI)

Students will use the following equation to calculate a point estimate of the proportion of red beetles in the population:

$$\hat{p} = \frac{Number\ of\ red\ beetles\ sampled}{Total\ number\ sampled}$$

Students will use the following equation to calculate a 95% CI:

Lower bounds of 95% CI: $\hat{p} - 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$

Upper bounds of 95% CI: $\hat{p} + 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$

n is 50 or the sample taken from the population
$\hat{p}$ is the point estimate
1.96 is a constant based on the standard normal distribution; in a standard normal distribution, 95% of the density is between ± 1.96 standard deviations of the mean

Example. If you removed 19 red beetles from the population out of a sample of 50, your point estimate would be 0.38, and the 95% CI would be 0.25 – 0.51. This means that 95% of the time you sample the population, this interval will include the true value (i.e., proportion of red beetles). It does not mean that "I am 95% confident the true value is within this interval."

Step 5. Students calculate a point estimate and 95% CI for their count data

Students will use the Student Handout to calculate the point estimate and 95% CI for their data. Your students can visualize the width of the 95% CI on the graph template provided on the Student Handout.

Step 6. The teacher combines the class' data to demonstrate the effect of sample size on the 95% CI

If you combine your class' data and treat it as one sample, the point estimate is more similar to the true value and the width of your CI decreases. This is a clear way to demonstrate the effect of sample size on estimating population parameters.

If you add together the example class data above, your new data would be:
Number of red beetles sampled: 207
Total beetles sampled: 500
New point estimate: 0.41
New 95% CI: 0.37-0.46

Step 7. Class discussion

Share with your students the true population proportions: 40% red and 60% white. Ask your class: how many people calculated 0.4 as their point estimate. How many people had 95% CIs that included 0.4? Did anyone get a point estimate above 0.5?

Have your students write a scientific claim based on the prompts on the Student Activity Handout.

Extension

Consider using this activity as a precursor to an inquiry project where students collect and interpret count data. See the Extension Activity for an example.

Further Reading

D. Curran-Evert. 2000. The process of scientific discovery: How certain can we be? *The American Biology Teacher* 62(4): 266-280.

W. James. 2009. The variations of human sex ratio at birth during and after wars, and their potential explanation. *Journal of Theoretical Biology* 257: 116-123.

Contact Information

If you have any questions about the content of this Teacher's Manual, please contact Lauren Lucas at laurenklucas@gmail.com

Name: ____________________

Activity: Accounting for Uncertainty When Analyzing Count Data

Student Pre-Assessment

Directions:

1. Read the following article carefully.
2. Answer questions 1 and 2 following the article.

Study: Nearly 1 in 4 Wyoming women smoke while pregnant

By JOSHUA WOLFSON

Casper Star-Tribune

| Posted: Tuesday, November 16, 2010 11:30 pm

CASPER, Wyo. — Nearly one in four Wyoming women who were pregnant last year reported smoking during their pregnancy, according to figures released Tuesday by the March of Dimes.

Statewide, the number of women who smoke while pregnant rose from 22.6 percent to 24 percent, according to the organization's 2010 premature birth report card. That runs counter to a nationwide decline that began in 2005.

Smoking can contribute to premature deliveries and low birth weights.

Despite the increase, Wyoming improved its overall rating from March of Dimes, a national organization that works to reduce premature births. The group gave the state a “C” rating — an improvement over its “D” score in 2009.

“We are moving in the right direction,” said Angela Crotsenberg, epidemiology section chief for the Wyoming Department of Health's Community and Public Health Division. “It takes time when you are looking at population health to make big changes.”

Wyoming has an overall premature birth rate — defined as births before 39 weeks — of 11.2 percent.

Depending on the circumstances, premature births can result in health problems and developmental delays. In some cases, the effects can follow children into adulthood.

Health experts say they don't know for sure why nearly a quarter of women continue to smoke while pregnant. Wyoming offers several educational programs for expectant mothers and operates a quit-tobacco phone line that provides help specifically for pregnant women.

“That is the $100 million question because we have education available,” said Debra Hamilton, a women and infant health coordinator at the health department. “We have support within Wyoming.”

Some women continue to smoke because they believe it will keep them from gaining too much weight during their pregnancy, Hamilton said. They also suspect it will result in a smaller baby, and therefore, an easier birth.

Other women might be in homes where family members continue to smoke — making quitting more difficult.

"If they are smoking and the pregnant women is trying to quit, that is going to be a barrier," she said.

Wyoming did experience a decline in the number of uninsured women last year. Access to health care before and during a pregnancy helps doctors identify conditions that contribute to premature births, according to the report card.

The health department released a maternal and child health needs assessment for Wyoming earlier this year.

It called for improving the health of women of reproductive age, Hamilton said.

"A lot of very critical development takes places in the first few weeks of pregnancy, before women know they are pregnant, perhaps," she said. "If they are smoking at the time and drinking at the time, that can affect the outcome of their pregnancy and contribute to pre-term (births)."

Wyoming also experienced a decline in late pre-term births, which have been linked to a rise in rates of early inductions and Cesarian sections, according to the March of Dimes. Late pre-term births have become enough of an issue that the organization has begun an educational campaign encouraging women to wait until at least 39 weeks to deliver.

Although people might not always be aware, the final five weeks of pregnancy are still an important part of a fetus' development, said Wyoming Medical Center Women's Services Coordinator Cheryl Graff.

"A lot of that is they are not aware of the brain development," she said.

Even with the resources available in Wyoming, the state can still improve its educational efforts, Graff said.

Women need to understand the factors that contribute to premature births and the symptoms of pre-term labor.

"It is something that they can never hear too much of," she said.

Questions:

1. The article states, "Statewide, the number of women who smoke while pregnant rose from 22.6 percent to 24 percent, according to the organization's 2010 premature birth report card."

Can you be certain that the proportion of smoking pregnant women in Wyoming recently increased? Explain your answer.

2. What additional information would you want to evaluate the claim that one in four women smoke in Wyoming?

Name: ____________________

Activity: Accounting for Uncertainty When Analyzing Count Data

Student Activity Handout

1. Record the number of red and white beetles sampled.

	Red	White
TOTAL	______	______

2. Calculate the point estimate, your observed proportion of red beetles in the population.

$$\hat{p} = \frac{Number\ of\ red\ beetles\ sampled}{Total\ number\ sampled}$$

My Point Estimate:

3. Calculate the 95% Confidence Interval (CI) of your point estimate. A 95% CI means that 95% of the time you sample the population, this interval will include the true value (i.e., proportion of red beetles).

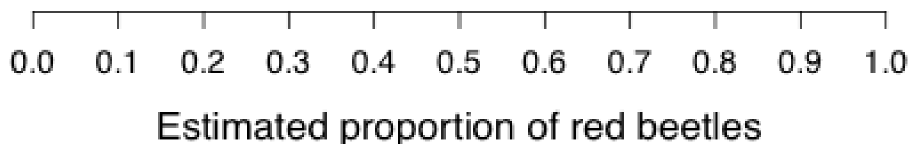
Lower bounds of 95% CI: $\hat{p} - 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$

Upper bounds of 95% CI: $\hat{p} + 1.96\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$

n is 50 or the sample taken from the population
$\hat{p}$ is the point estimate
1.96 is a constant based on the standard normal distribution; in a standard normal distribution 95% of the density is between ± 1.96 standard deviations of the mean

My 95% CI:

4. Visualize your analyzed data by drawing the point estimate and 95% CI here:

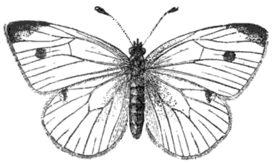


5. Write a scientific conclusion below based on your data analysis. To help, here are some examples of ways to interpret data from this activity:

“I cannot reject the hypothesis that there are 50% red and 50% white beetles in the population.”
“The point estimate is close to 0.4 but I cannot reject proportions between 25% and 51%.”

Name: ____________________

Activity: Accounting for Uncertainty When Analyzing Count Data
Extension Activity



Guided Inquiry: Butterfly larval food performance

Background Information.
The larva (caterpillar) is the main feeding stage of butterflies; the pupa eats nothing at all, and the adult simply sips fluids for energy and moisture.

Butterflies evolved along with flowering plants, and the few species that eat nonflowering plants today have evidently switched to them from flowering plants. Some tropical butterfly larvae in the family Lycaenidae eat lichens and fungi!

Most butterfly larvae feed on plants of a single family, but there are a few exceptions. Most butterfly larvae that feed on plants from more than one family eat plants from several closely related families. Again, there are some exceptions. For example, the Cabbage White Butterfly (*Pieris rapae*) feeds on plants from three closely related families, the Cruciferae, Capparidaceae, and Resedaceae, and one unrelated family, the Tropaeolaceae, because all four families contain mustard oils that attract the females and stimulate the larvae to feed.

But why do most butterfly larvae eat the plants of only a few genera or families? There is a perpetual "war" between insects and plants. Ancient insects ate ancient plants; in response, the plants evolved thick, tough, tooth-edged leaves that carry spines or hairs, hooked hairs, resins, high levels of silica, all of which physically deter animals from eating the leaves. Plants have also evolved chemical defenses against attack. Through time plants that were edible to many animals were eaten and driven to extinction, and only the plants with defenses remained. By evolving one or more of these repellents, a plant could survive all but the few insects able to withstand its defenses. The few remaining insects then evolved the ability to use the plant's own poison to find it, for the poison is generally a reliable indicator of plant identity. Furthermore, few insect species can detoxify or tolerate the numerous plant poisons residing in a great array of plants, and because most insects can detoxify only a few poisons, most eat only a few plants – generally plants that are closely related.

Some butterflies do not detoxify the plant poisons. Rather, they have evolved other chemical systems to tolerate them. The plant poisons accumulate harmlessly in their bodies, and when one is attacked by a bird or other animals that cannot tolerate the poison, the predator either releases the butterfly or vomits and learns not to hunt that species again. These butterflies then become models for mimicry.

But, what about the plant? Couldn't it still be eaten to the ground by those few species that can survive its poisons and defenses? These species are usually held in check by other predatory and parasitic insects. In this guided inquiry lab, you will test these ideas for yourself and explore how well Cabbage White Butterfly larvae survive on different plant varieties.

Research Topic.
How well are Cabbage White Butterfly larvae adapted to a particular plant?

Research Question.
Do Cabbage White Butterfly larvae have the greatest chance of survival when feeding on cabbage or brussels sprouts?

Background Research.
What other background information do you need before you form a hypothesis?

There are several useful resources for gathering scientific background information. Some suggestions are:
http://scholar.google.com/
http://www.jstor.org/

Hypothesis.
Form your hypothesis here or in your research notebook:

If____,
then____.

Materials Needed.
1. Cabbage White Butterfly eggs ordered from Carolina Biological Supply (Item #144100)
2. Laval food plants - cabbage, brussels sprouts (buy organic or wash thoroughly)
3. Clear plastic cups, plastic wrap, rubber bands – to hold the plants and larvae
4. Small paintbrushes – to carefully transfer eggs and larvae to new plants
5. Sharpies – to label the plastic cups
6. Research notebook – to record data and observations
7. Hand lenses or dissecting microscopes – to view the larvae during each instar (developmental stage)
8. Scotch tape – to tape pupae to paper during pupation
9. Adult butterfly cages (BioQuip Catalog #1466A) – to house adult butterflies for observation or further experimentation

Procedure and Tips for Caring for Larvae.
(See the care sheet by Carolina Biological Supply for further tips.)
1. In groups of 4, obtain approximately 10 eggs so that you will have 5 eggs per plant type. Make sure you have equal numbers of eggs on each plant type.
2. Set up your 2 plastic cups and label them appropriately. It is better to tear the plant leaves than to cut them with scissors.
3. With a damp small paintbrush, carefully and slowly roll an egg onto the side of the paintbrush. Carefully roll the egg back onto the plant. Do this until all eggs have been transferred to plants.
4. Record the number of eggs you started with on each plant type in your research notebook.
5. Before you see signs of plant molding, you will need to transfer your larvae to new plants. Again, use a damp small paintbrush to roll the larvae onto the side of the paintbrush and roll them back off onto new

plants. They will try to escape by hanging by threads of silk. Try to avoid transferring them during the first instar (developmental stage). When the larvae get bigger, you will have to give them new plants every other day or every day. When you notice larvae wandering off the plant, they may be molting. It is best to leave them alone, if possible, while they are molting.

6. Record the number and percentage of larvae still alive on each plant type each week. See below for a sample data table.

7. Once your larvae have pupated, you can weigh them to further answer how feeding on certain plants affects the growth and development of the Cabbage White Butterfly.

8. You can attach the end of the pupae to a piece of paper with scotch tape. Place this piece of paper in an adult butterfly cage to prepare for emergence.

9. Once your butterflies emerge as adults, you can provide them with a nectar source (e.g., sugar water sprayed on the cage). Males and females will likely mate in the cage. You can start a new experiment to ask: do females prefer to lay their eggs on the same plant larvae performed best?

Data Collection.
As a scientist, it is important to record detailed observations of your experiment during the entire process. This will help you keep track of the changes you see throughout your project. Both scientific drawings and written descriptions are recommended.

You will also need to record data throughout the course of your project. Here is an example of how you might set up a data table.

Week	Survival on cabbage: Number	Survival on cabbage: Percentage	Survival on brussels sprouts: Number	Survival on brussels sprouts: Percentage
1				
2				
3				
4				
5				
6				

Data Analysis.
What are your dependent and independent variables? Make a line graph of survival over time for each plant type. Make a bar graph of the final percentage of larvae surviving on each type of plant. Use these percentages as your point estimates. Calculate 95% confidence intervals for each point estimate to quantify the uncertainty in your estimates of survival on each type of plant.

Conclusion.
Describe what you found. Answering the following questions will help you write clear claims: How would you answer the research question? Was your hypothesis supported or refuted?

Reflection.
What did you learn during this inquiry project? The following questions will help you write a clear reflection: How have your ideas changed? What would you have done differently? Explain your possible sources of error.

Peer-Review.
All scientists must have their research papers approved by their peers before it can be published. Have another group look at your data and interpretations. Do they have similar findings? Take their comments into consideration. Furthermore, would your conclusions change if you analyzed your class' pooled data? If time permits, try this.

Reference used to develop this lab.
Scott, J. A. (1986). *The butterflies of North America: A natural history and field guide*. Stanford University Press: Stanford, CA.

Developed by Lauren Lucas and Zach Gompert. August 2010.