## Gummy Bear Lab

**Gummy Bear Descriptive Statistics – Teacher prep notes**

Prep: For each lab station (I have 7, so adjust the handout appropriately for your number)

* 1 larger beaker/cup (~200 ml) filled with any solution (DI water or any reasonable molarity of sucrose or NaCl) containing 5 bears of the same color. All stations need to have the same solution and same quantity, but with different color Bears, so that one station has red, one station has clear, one has yellow, and so on. The bears I buy come in 5 different colors, so some bears are double represented. This, I think, is a good thing so kids have to work with different sample sizes.
* 5 smaller (~50 ml) beakers/cups filled with equal quantity of solutions: DI Water, 0.25 M sucrose, 0.5 M sucrose, 0.75 M sucrose, 1 M sucrose. Place one bear of any color in each cup.

Allow these to soak for at least an hour before class. You CAN do it the night before for a first period class, but the bears get pretty gross and you probably can’t re-use them.

Also needed: balances, paper towels, plastic spoons, and TI-84 (or similar) calculators.

\*Note: I do realize that samples this small are really not sufficient to do a t-test, but since I’m trying to teach the process, I’m not worrying about it.



**The Gummy Bear Lab**

Experimental Design, Descriptive Statistics, Graphing at the AP Level,

and an Introduction to Statistical Analysis

*How does the mass of gummy bears change when soaked in different solutions?*

Obtain a lab tray. In it you will find one large cup containing 1M NaCl and several gummy bears of the same color and several smaller cups containing a single gummy bear each. We are going to work as a class to collect and analyze two sets of data with these bears.

First, for consistency, all bears started with a mass of \_\_\_\_\_\_ grams. Anything different from that was not used in this experiment.

**Part A: Do all colors of bears gain/lose the same amount of mass in the same solution?**

Experimental Design:

* What’s the independent variable? (What could we fill into the table prior to data collection?) Is this variable categorical or continuous?
* What’s the dependent variable? (What will we measure?) Is this variable categorical or continuous?
* What’s an appropriate control? (How could we show that changes in our dependent variable were due to changes in our independent variable?) Is this a positive (looking for a known response) or a negative (looking for no response) control?
* What should be held constant to ensure validity of our results? What aspects of this experiment, if changed, could affect the data collected? List three.

Once your teacher has approved your experimental design, collect your data: Color: \_\_\_\_\_\_\_\_\_

|  |
| --- |
| Mass (g) |
|  |
|  |
|  |
|  |
|  |

1. Using only the large cup containing five bears of the same color, use a spoon to fish out the bears. Place them on a paper towel. Gently pat them dry.
2. Weigh them one at a time on the electronic balance, and keep track of your data in the table at the right.
3. Return them to the cup to continue soaking. Don’t throw them away; we’re using the same bears all day.
4. Bring your data to the front of the class to contribute to the class spreadsheet.

While we wait for everyone to finish, consider this situation:

Your teacher is shopping on Amazon for new pH meters. Two meters are comparably priced and have good reviews. Here are the “stars” with the number of reviewers:



Which “star” count is more reliable? Why? From which seller am I more likely to get a disappointing product?

**Table 1: Class Data - Mass of Different Color Gummy Bears Soaked in a solution**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color:** |  |  |  |  |  |
| Trial 1 Mass (g) |  |  |  |  |  |
| Trial 2 Mass (g) |  |  |  |  |  |
| Trial 3 Mass (g) |  |  |  |  |  |
| Trial 4 Mass (g) |  |  |  |  |  |
| Trial 5 Mass (g) |  |  |  |  |  |
| Trial 6 Mass (g) |  |  |  |  |  |
| Trial 7 Mass (g) |  |  |  |  |  |
| Trial 8 Mass (g) |  |  |  |  |  |
| Trial 9 Mass (g) |  |  |  |  |  |
| Trial 10 Mass (g) |  |  |  |  |  |

This is our *raw data*. Unlike most of your prior classes, raw data is almost never graphed in this class. Instead, we want to do some analysis first and represent the data using descriptive statistics, or numbers that summarize the data. As we work through the descriptive statistics you’ll need for this course, you’ll record your answers in table 2 below.

First, enter your data in lists 1-5 in your calculator, exactly as shown in Table 1. We will use your calculator heavily to both calculate descriptive statistics as well as check your work. To do this, hit “STAT” and then “EDIT.” Clear the lists if necessary and enter your Gummy Bear data. Don’t worry that some lists are longer than others are.

**Table 2: Descriptive Statistics of the mass of different colors of gummy bears soaked in a solution**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color:** |  |  |  |  |  |
| Sample Size (n) |  |  |  |  |  |
| Mean ($\overbar{x})$ |  |  |  |  |  |
| Standard Deviation (s) |  |  |  |  |  |
| Standard Error of the Mean (SEM) |  |  |  |  |  |
| 2\*SEM |  |  |  |  |  |
| Mean + 2SEM |  |  |  |  |  |
| Mean - 2SEM |  |  |  |  |  |

Determine your **sample size**, or ‘n’, for each column. This is simple – how many trials do you have? Notice that because there were 7 lab groups and 5 colors, some colors have a larger sample size than others. That’s okay! Just record each color’s sample size in each column.

Most of you have used the “**mean**” or “average” before; it’s the just value of central tendency, or the middle of a data set. The formula on your sheet looks scary, but all it says is to find the mean, add up all of your values for that condition (so all the data values for one color) and divide by your sample size, or n.

Show your work for calculating the mean of the data in column 1 in the space below, and then just report the mean for the other colors.

The mean is a useful representation of data, but it has limitations. For example, a small dataset consisting of numbers 19, 20, and 21 has a mean of 20, but so do the numbers 10, 20, and 30. The two datasets are drastically different, but the mean value doesn’t represent that. We use a value called “**standard deviation**” to represent how spread out the data is. Calculating standard deviation isn’t hard, but is very tedious, so I will never ask you to do it by hand. Go to your calculator.

Starting with the first color, go to STAT🡪CALC🡪1:1-Var Stats and hit enter. Select a list by hitting 2nd 🡪 L1, and hitting Enter. You’re going to see a bunch of stuff pop up that looks like this:

Most of this information we have no use for, but knowing how to interpret this information will save you a ton of time. The $\overbar{x}$ is the mean of the dataset; compare it to the one you calculated by hand. The n at the bottom is your sample size; again compare it to the one you counted. We do these to make sure you didn’t typo the data into the calculator. The Sx value is the standard deviation of the dataset. Round it to hundredths and write it in table 2. Repeat this process for the other Lists/colors.

Ultimately, in science, our goal is to learn about the entire world, not just what’s happening in our classroom. That is, if we are looking at the change in mass of 5 red gummy bears, we’re actually interested in what happens to all the red gummy bears on the planet in this same situation, not just our 5. In order for us to make the leap from our small sample of red gummy bears in a cup to the entire population of red gummy bears on the planet, we calculate a value called the **standard error of the mean**, or SEM, or $SE\_{\overbar{x}}$. The SEM represents how accurately the sample mean represents the population mean. The larger the sample size, the more accurately the mean represents the population, just as you probably concluded with the Amazon example at the beginning of the lab. You will need to calculate SEM by hand, but it’s quite easy: Just divide the standard deviation by the square root of the sample size. The standard error of the mean has many applications in statistical analysis, but we will use it primarily for graphical purposes. In general, when we graph data, we graph the mean±2\*SEM. We do this because it represents something called a **95% confidence interval**. That is, we are 95% sure that the true population mean falls somewhere in that range, based on the data collected. Reasonably, the larger a confidence interval is, the less confident we are in that mean being accurate to the population as a whole.

So, let’s make that graph. As a reminder, the independent variable goes on the x-axis and the dependent variable goes on the y-axis. Since the independent variable is categorical, create a BAR graph. For each color bear, graph the mean±2\*SEM so it looks similar to the one on the left, but with one bar for each color bear. The top of the bar is the mean. The I-shaped thing at the top of each bar represents the top and bottom of the 2SEM, so that mean+2SEM is the top of the I, and mean-2SEM is the bottom of the I.

Create a graph of your data:



Now, let’s return to our original question: **Do all colors of bears gain/lose the same amount of mass when soaked in the same solution?** We probably have some differences in the calculated means of each color, but how different is different enough to claim that the colors don’t all respond equitably? Remember, our goal is to make an inference about the entire world’s population of gummy bears, not just the ones in this classroom. So, rather than just looking at the calculated means, we look at the error bars you drew. If two colors’ error bars ***do not*** overlap, we can say that they are different in terms of the dependent variable. If two colors’ error bars ***do*** overlap, we say they are the same in terms of the dependent variable.

Write a CLAIM to answer this question. You may not start with the word ‘yes’ or ‘no.’

Describe the EVIDENCE you have to support your claim.

**Part B: How do different concentrations of sucrose influence the mass of gummy bears?**

Now, we’ll use the cups that contain the individual bears. Look at the cup. You’ll see that each is marked with a particular molarity of solution.

Experimental Design:

* What’s the independent variable? (What could we fill into the table prior to data collection?) Is this variable categorical/discrete or continuous?
* What’s the dependent variable? (What will we measure?) Is this variable categorical/discrete or continuous?
* The control and constants are the same as Part A, so you don’t need to write them again. When we graph this data, will we use a line graph or a bar graph?

Now, collect your data:

1. Using the five smaller cups that contain a single bear, use a spoon to fish out the bear. Place it on a paper towel. Gently pat it dry.
2. Weigh it on the electronic balance, and keep track of your data in the table below.
3. Return it to the cup to continue soaking. DO NOT CONFUSE YOUR BEARS. Make sure they get back into the appropriate cup.
4. Bring your data to the front of the class to contribute to the class spreadsheet.

**Table 3: Mass of bears in different concentrations of sucrose**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Solution** | **0 M (DI Water)** | **0.25 M** | **0.5M** | **.75 M** | **1 M** |
| Group 1 Mass (g) |  |  |  |  |  |
| Group 2 Mass (g) |  |  |  |  |  |
| Group 3 Mass (g) |  |  |  |  |  |
| Group 4 Mass (g) |  |  |  |  |  |
| Group 5 Mass (g) |  |  |  |  |  |
| Group 6 Mass (g) |  |  |  |  |  |
| Group 7 Mass (g) |  |  |  |  |  |

Now, complete the descriptive statistics for this dataset using the same procedures as used in Part A:

**Table 4: Descriptive Statistics of the mass of bears in different concentrations of sucrose**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Solution** | **0 M (DI Water)** | **0.25 M** | **0.5M** | **.75 M** | **1 M** |
| Sample Size (n) |  |  |  |  |  |
| Mean ($\overbar{x})$ |  |  |  |  |  |
| Standard Deviation (s) |  |  |  |  |  |
| Standard Error of the Mean (SEM) |  |  |  |  |  |
| 2\*SEM |  |  |  |  |  |
| Mean + 2SEM |  |  |  |  |  |
| Mean - 2SEM |  |  |  |  |  |

Here are the formulas again, and some room to show work if you’d like:

Once again, create a graph of your data. This time, the independent variable is continuous, so you’ll make a line graph, however, don’t draw the line yet. For now just make the scatterplot. It should look something like this but with labels and units on your axes:





When taking exams, drawing/sketching a line of best fit that approximates the data is sufficient. However, when completing a lab, I’ll often ask you to find the best-fit line using the regression feature of your calculator. To do that, go back to the lists in your calculator (STAT 🡪EDIT) and clear the raw data you used to find your descriptive statistics. Enter the (X,Y) coordinates of the means you plotted in lists 1 and 2, respectively. Then go to STAT🡪CALC🡪4:LinReg(ax+b) and hit enter. The calculator will give you the formula for the best-fit line (called a Linear Regression) of your data. Write it down here, and add it to your graph:

BIG DEAL: When graphing a line, do not extend the line beyond the last collected data point. If you are going to extrapolate your trend beyond the last data point, use a dashed line by actually lifting up your pencil at the last data point and using a - - - - to indicate that you’ve extrapolated.

Once again, let’s go back to our original question: **How do different concentrations of sucrose influence the mass of gummy bears?**

Write a CLAIM to answer this question.

Describe the EVIDENCE you have to support your claim.

**Day 2: Introduction to Statistical Tests**

Last class we worked on describing data with statistical measures such as standard deviation, SEM, and error bars. You learned (or were reminded how to) find a best-fit line on a scatterplot and that (ahem) bar graphs are a thing. We learned how to interpret error bars to state a claim about the difference (or lack thereof) of two different groups of data.

So, how certain are we that any two groups of bears are different? The SEM bars allow us to state a claim of similarity or different with 95% confidence, but can we be more specific? The purpose of statistical tests is to generate a value that describes how sure we are of the similarity or difference of populations. Today we’ll learn about statistical hypothesis testing to generate something called a p-value that allows us to be more specific.

To begin with, statistical hypotheses are different from scientific hypotheses. Don’t get them confused. All hypotheses are based on the premise that it’s impossible to prove something *right* but it is possible to prove something *wrong*. This is why scientific hypotheses have to be falsifiable. There are two statistical hypothesis, the **null hypothesis** and the **alternate hypothesis**. The word “null” means “zero,” so the null hypothesis simply states that there is no difference between two datasets. Generally speaking, you’re looking to show that you have sufficient data to prove the null hypothesis false. Then there is the “alternate” hypothesis. The alternate hypothesis is the converse of the null hypothesis. By showing the null hypothesis is false, you are supporting (but not proving) the opposite, the alternate hypothesis.

For Part A, where we looked at how different colors of gummy bears mass changed, our hypotheses would be:

* Null hypothesis: There is no difference in mass between the red and clear gummy bears.
* Alternate hypothesis: There is a difference in mass between the red and clear gummy bears.

While there is a statistical test that will compare more than two datasets, (honestly there are dozens of different statistical tests that serve a variety of purposes) we are only going to use two this year: one called a t-test and another called a Chi-square test. Today we’re using a t-test. All statistical tests are based on a bunch of calculus and areas under a curve that you don’t need to understand. What this calculus generates is a p-value. If you think ‘p’ means probability, the question becomes, “What is the probability that the populations represented by these two datasets are the same?” The larger the p-value, the more similar the datasets (and, thus, the populations) are, supporting the null hypothesis. The smaller the p-value, the less similar the datasets (and, thus, the populations) are, refuting the null hypothesis. Ultimately, the cut-off p-value in Biology 5%, or a p-value of 0.05. We reject our null hypothesis only when there’s less than a 5% chance the two populations are actually the same.

**If your p < 0.05, reject your null hypothesis and accept your alternate hypothesis.**

**If your p > 0.05, fail to reject your null hypothesis.**

So how do you get a p-value? Conveniently, your calculator will find one for you once you enter your raw data. Go to STAT 🡪 EDIT and enter the red gummy bear masses in List 1 and the clear gummy bear masses in List 2. Then back to STAT 🡪 TESTS 🡪 4:2-SampTTest and ensure you have selected Data, L1, L2, 1, 1, ≠µ2, No, Calculate and hit Enter.

Your screen will appear similar to the one at the right. First, CONFIRM that it says µ1≠µ2. Then verify your means by comparing $\overbar{x}1$ and $\overbar{x}2$ to the means you calculated in your table of descriptive statistics. If there’s a discrepancy, double-check your data entry in your lists; you probably have a typo somewhere.

You only need two pieces of information: the t-statistic and the p-value. The t-statistic just tells you how much evidence you have against the null hypothesis. The number can be positive or negative. Round it to hundredths and write it down. The p-value tells you whether you can reject your null hypothesis. Since it’s a probability, it will ALWAYS be between 0 and 1. BE CAREFUL, as your calculator will use scientific notation, so pay attention. Round that value to thousandths, and write it down. If it’s miniscule, you can write p < 0.001. Last, interpret the p-value. Are you rejecting or failing to reject the null hypothesis? Then, interpret it: Are the two populations different? Either write, “The red and clear gummy bears have statistically different masses when soaked in the same solution” or “The mass of the red and clear gummy bears are statistically indistinguishable when soaked in the same sucrose solution.”

So, put it all together! Just copy the hypotheses from the previous page.

Null Hypothesis:

Alternate Hypothesis:

t= p=

Do you reject or fail to reject the null hypothesis?

What does this mean?

Now try one of your own from Part B of the lab, where we looked at the masses of bears that soaked in different concentrations of sucrose. Compare the masses of the bears soaked in 0M sucrose (Water) and 1M sucrose.

Null Hypothesis:

Alternate Hypothesis:

t= p=

Do you reject or fail to reject the null hypothesis?

What does this mean?

To recap, a Two-sample t-test is used to determine if two datasets (samples) are sufficiently different to state with less than 5% doubt that the populations from which they came are also different. This test is typically used to compare two different levels of independent variable (such as what we did in this lab), or to compare the independent variable to the control, or some other baseline measurement, all depending on the experimental design.

A statistical test is additional *evidence* that can be used when writing a CER. Next class we’ll explore the use of Chi-square analysis, a test where we compare observed results to expected values. The hypothesis writing, calculator use, and p-value interpretation are very similar.