Primary M&Ms Proportion Classroom Exploration
Quick Overview/Introduction

Dear Instructor,

In this exploration, we will use bags of M&Ms to explore the concept of the distribution of sample proportions. Having every student do this with their own bag will give a quick and edible representation of how we can draw many different samples from a population and see how the population of these samples can allow us to understand how a sampling distribution can offer us information about the value of a parameter.

To do this exploration, you will need the documents mentioned on step one on the page 4 as well as a bag of M&Ms for each student as well as one for the teacher. **[It is very helpful for the teacher to model the process and show students where to input values.]** As noted above, the use of Fun Size bags does lead to situations where it is not advisable to assume the population of individual bags is Normal. However, this is quickly remedied by combining bags as we go. If someone were to use traditional size bags, this issue would not exist, but I do not believe that the use of Fun Size bags would limit the learning of any student.

This exploration was designed for an introductory statistics course that heavily leverages the R software package. The only place that R is utilized here is to generate random samples to create larger “virtual” bags of M&Ms. There are other options that can do the same thing, but due to the small size, resampling is suggested if possible.

There are many places along the exploration that a teacher can choose to stop and have discussions about the statistical topics and ideas that are being worked with. The step by step directions are mainly there to ensure that the Google Sheet and its calculations are accurate.

In the end, it is your classroom. I hope that you find this exploration helpful and beneficial to your students. If you have any questions, do not hesitate to contact me at phil.huling@slu.edu.

Sincerely,

Dr. Phil Huling
Saint Louis University
Department of Mathematics and Statistics

Notes:

- **Contact the author, Dr. Phil Huling, at phil.huling@slu.edu if you have any questions, comments, or concerns about this exploration.**
- **Some screenshots may not show cells that are in the final version of the Google Sheet.**
- **When buying M&Ms, it is important to not purchase bags which may have non-traditional color choices. This is most commonly an issue around holidays.**
- **Individual Fun Size bags do not technically meet the criteria to use the Normal approximation, but the multiple bag samples should satisfy the conditions offered in most textbooks.**
- **The Step by Step Directions do get more sparse after step 20 as the process is repetitive and the individual teacher is encouraged to adapt the exploration to their own desires.**
- **A class discussion about what to do about partial M&Ms and other oddities is recommended if anyone does find something abnormal. In the author's experience, there are far fewer abnormal individuals in the milk chocolate variety of M&Ms as compared to the peanut variety.**
- **It is encouraged to use the milk chocolate variety for the above reason as well as to minimize the possibility of exposing students to potential allergens. However, it is important to note that the company's website does indicate that there is a chance that milk chocolate M&Ms may contain traces of peanuts due to being manufactured in a facility that also makes products with peanuts.**
- **The Google Sheet referenced here is meant to be "automated" and create the sample proportions as well as the numerical and graphical summaries. Please follow steps 5 through 7 carefully to ensure that this automation stays in tact. If any issues arise, please contact the author.**
- **If one does not want to use R, the website random.org is a great place to replace those steps. However, this website does not implement resampling very easily. Resampling should help in cases of smaller class sizes, but this can be overlooked for the overall benefit of such a hands-on exploration.**
- **As an example, there is a value of x = 7 and n = 16 entered as the teacher value in the red column. This can be used as it was pulled from a real fun size bag by the author, or you can delete it and enter in your own values. Note that the value of 0.4375 in cell D3 is auto computed by the Google Sheet. Do not delete that value.**
- **Starting on page 10, there are directions for exporting the class data to be used in RStudio. This is optional, but included for the teachers who want to keep the data and use it with RStudio.**

Learning Objectives

- **Understanding the following concepts:**
  - **Point estimate**
  - **Individual observations [Two different kinds here]**
    - **YES/NO to question**
    - **Each sample proportion**
  - **Population proportion**
  - **Samples of different sizes**
  - **Sampling distribution of sample proportions of a fixed size**
  - **Population of sample proportions of a fixed size**
    - **Population mean of sample proportions**
    - **Population standard deviation of sample proportions**
- **Using the formulas for the population mean and standard deviation of the distribution of sample proportions:**

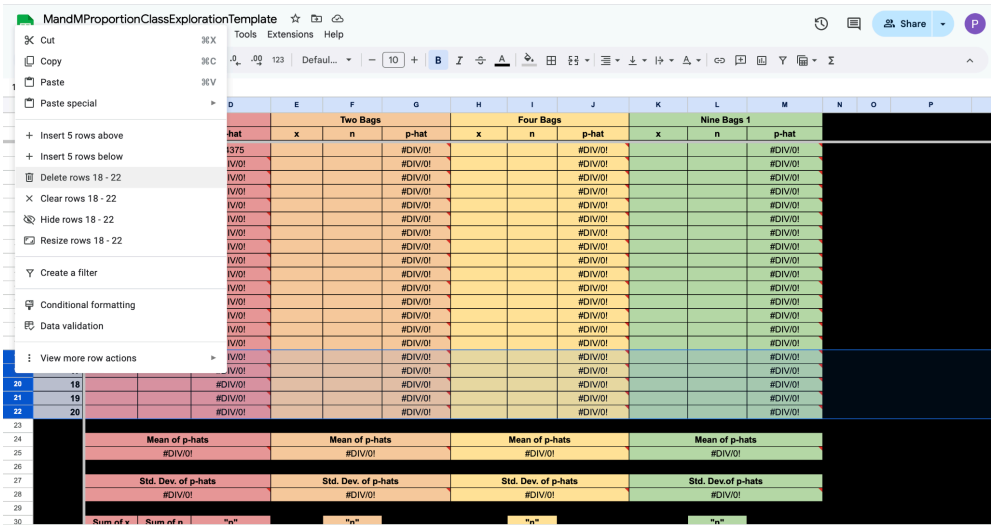
$$\mu_{\hat{p}} = p$$

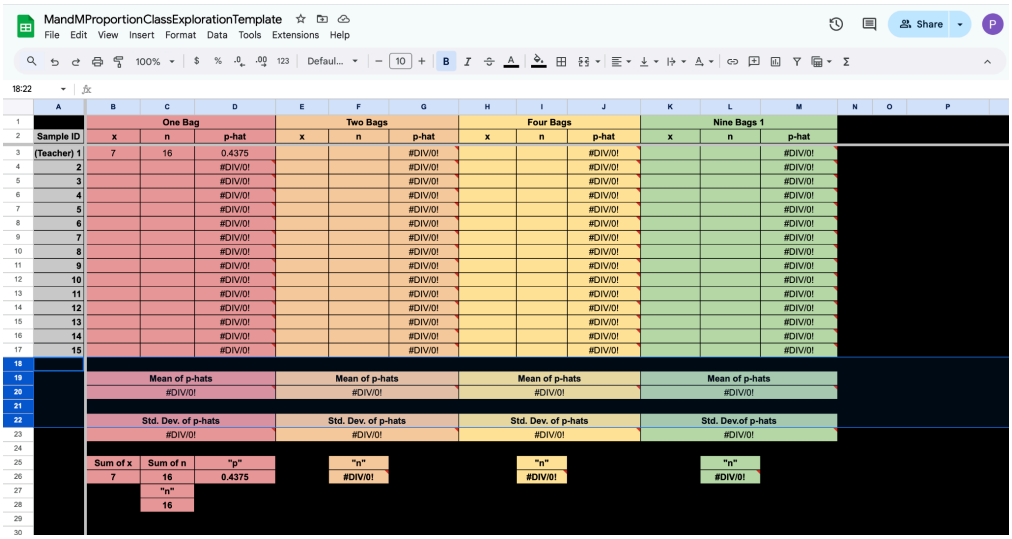
$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- **Understanding the relationship between a sample proportion and a binomial experiment**
- **Comparison of the above formulas with the parameters of the Normal approximation of a binomial random variable**
- **Difference(s) between a statistic and a parameter**
  - **Statistic**
    - **A point estimate of a parameter**
    - **Varies based on the sample**
  - **Parameter**
    - **The perfect value we want to know**
    - **Often difficult or impossible to attain**
    - **(Usually) Does not change**
- **Understanding how the sample size affects the sampling distribution**
- **Light introduction into the idea of confidence intervals based on the diminishing population standard deviation.**
- **Export a Spreadsheet and import it into RStudio**

Step by Step Directions

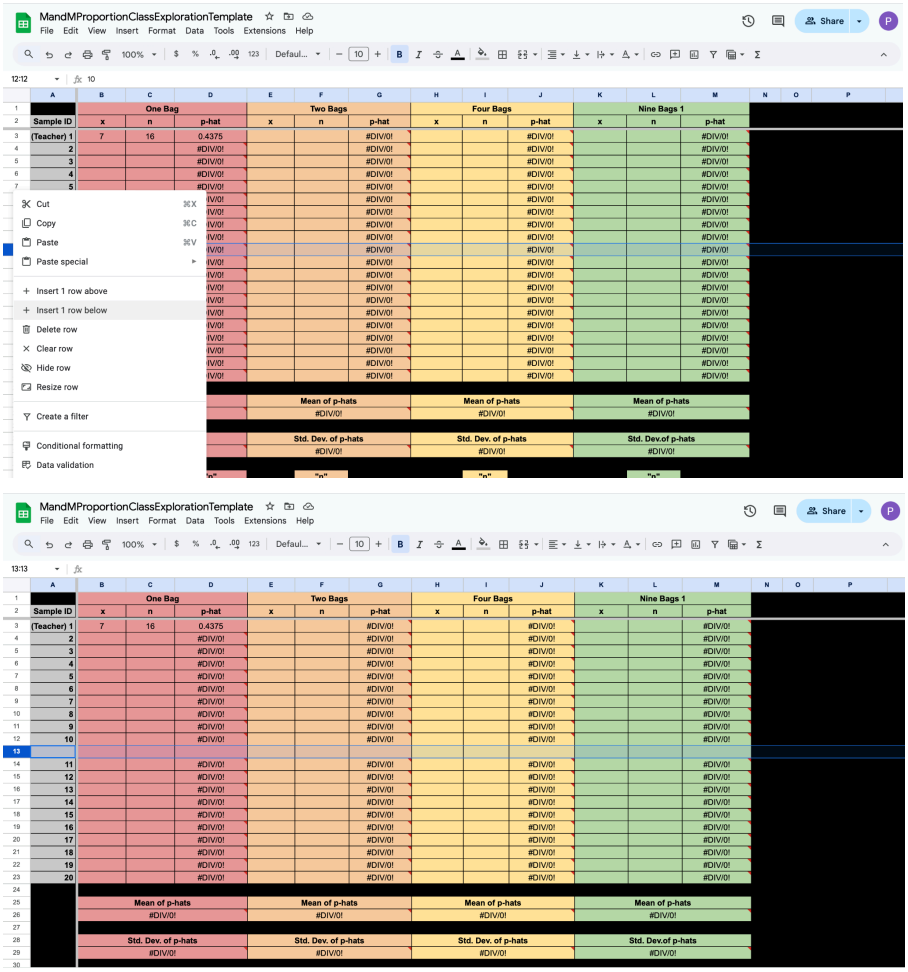
1. To do this exploration, you will need this document along with the Google Sheet entitled MandMProportionClassroomExplorationTemplate as well as the R Script called MandMProportionClassroomExploration (or an alternate RNG).
2. Share MandMProportionClassroomExploration.R in a place where your students can access it easily, such as your LMS.
3. Create a copy of the Google Sheet and remove the word Template from the file name and change it to an identifier for your particular course. For example, one could use MandMProportionClassroomExplorationSection2Fall2024.
   [Note: Please do not modify MandMProportionClassroomExplorationTemplate.]
4. You should practice completing steps 5 through 7 prior to showing up for class.
5. If necessary, adjust the number of people doing the exploration. Remember, that the number needed is the number of students plus one for the instructor.
   a. If the number needed is less than 20, select the rows that are not needed, and remove those columns. Below shows how to cut down to 15 rows.



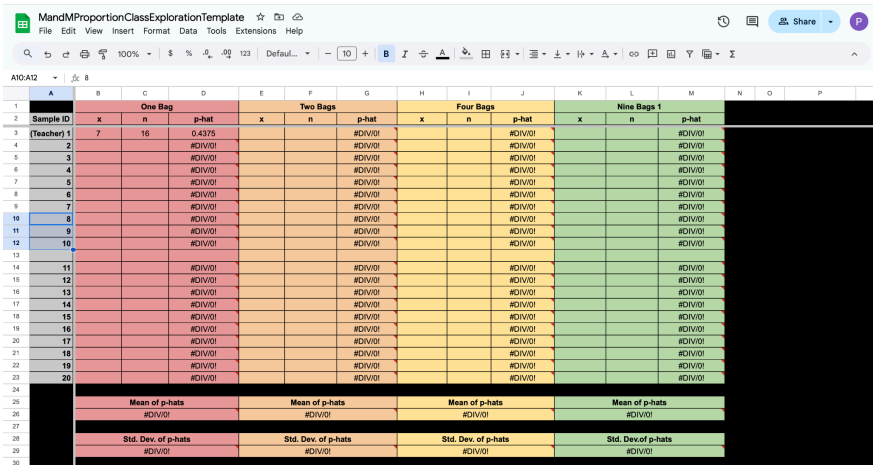


b. If the number needed is more than 20, follow the following directions.

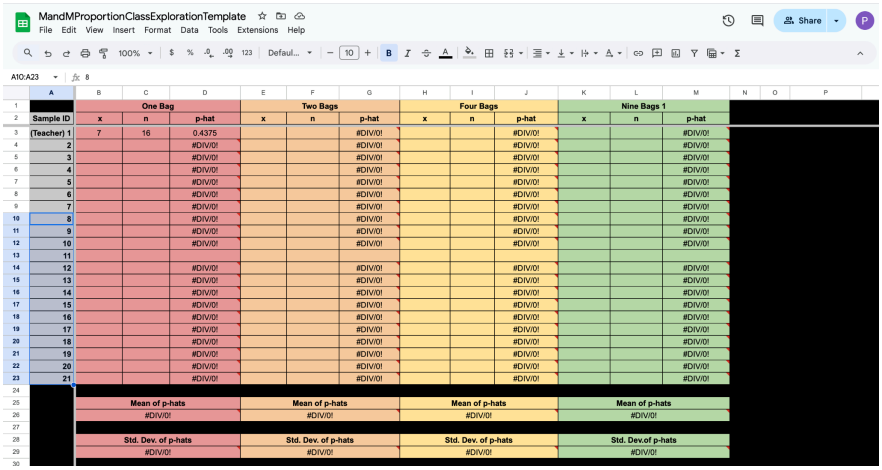
i. Right click on row 12 (which is Sample ID 10) and choose “Insert 1 row below.” Repeat this for each additional student as needed. **[Note: Only one additional row is shown being added below, but the process is identical other than having to drag further at certain points.]**



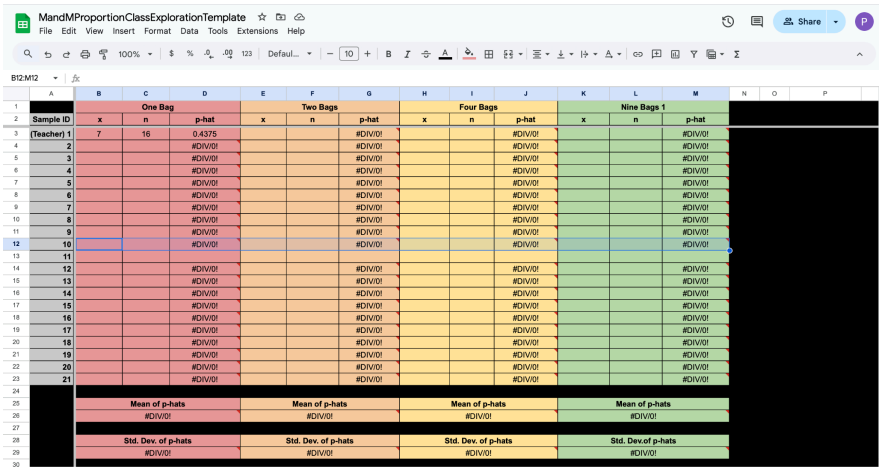
ii. In column A, select the 3 more Sample IDs above the new blank row(s). It should be cells A10 through A:12.



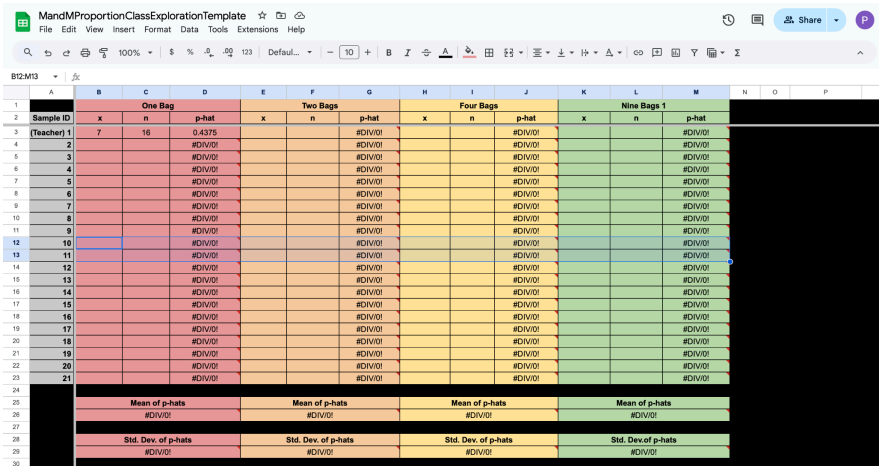
iii. Then click on and drag the lower right corner of the selected cells to the bottom of the gray area to automatically fill in the Sample IDs.



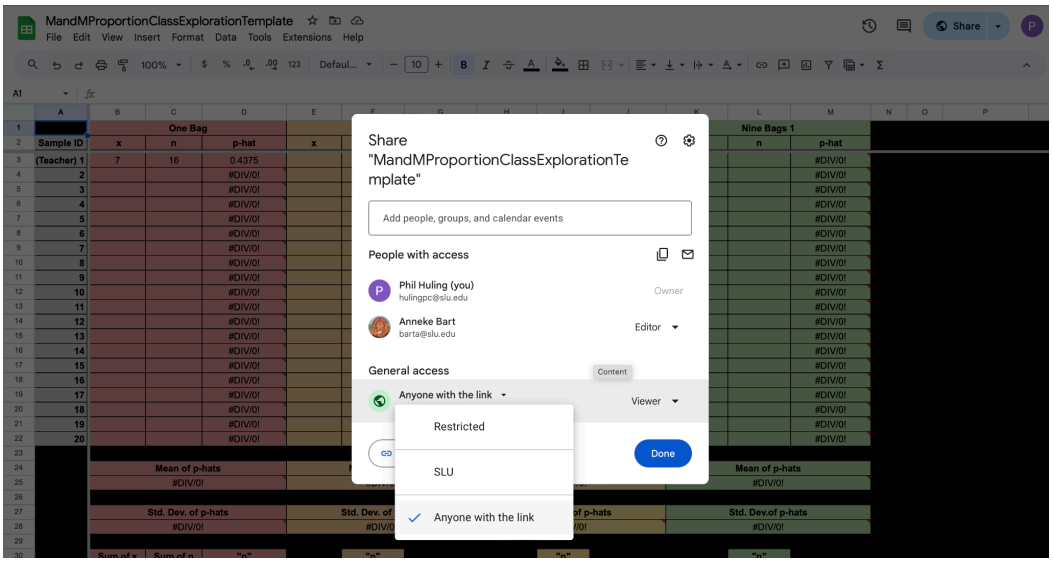
iv. Now select the colored cells in the row above the blank rows created earlier. It should be cells B12 through M12.
[Warning: Do not select the A12 cell.]



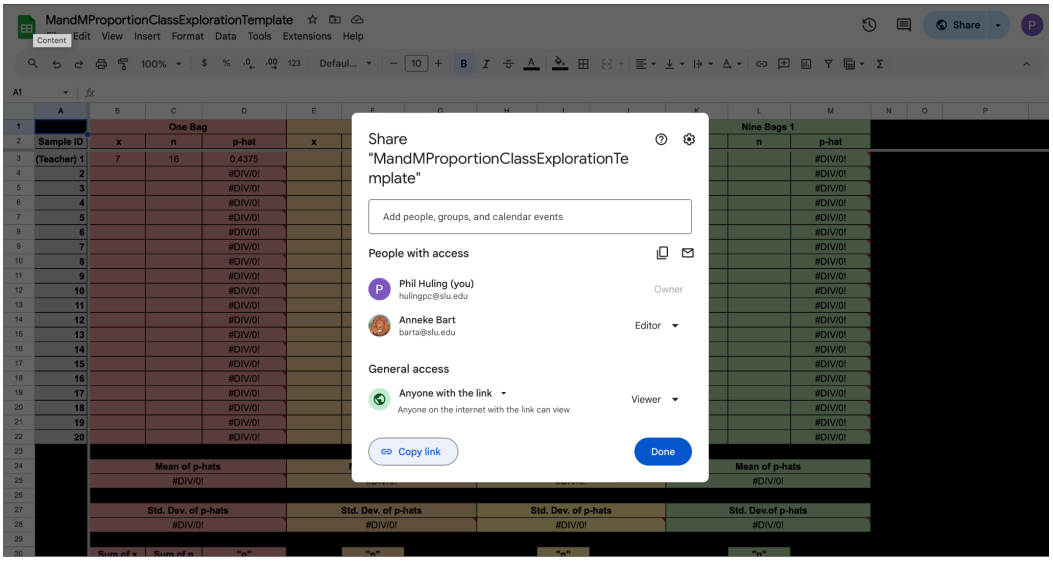
v. Then click on and drag the lower right corner of the selected cells to the bottom of the empty colored area to automatically fill in the Sample IDs.



6. This should now give you a spreadsheet which is fully automated other than the entries to be made by you and your students. Display the Google Sheet to the class.
7. At this point, you should share this Google Sheet with your class.
    a. Click on the Share button in the upper right hand corner and then under General Access, click on the drop down and choose “Anyone with the link.”



    b. You can then copy the link using the “Copy link” button and paste this link in a place your students can access it such as your LMS.



    c. Make sure each student has access to the file and has it opened on their own device.
8. You are now ready to begin the exploration in earnest. Pass out the bags of M&Ms to each student with the direction that they are not to open the bags until they are told to do so.
9. Explain what they will be doing. An example discussion: “Today we will be looking at sampling distributions of the proportions. To do this, we will be examining the proportion of primary colored, i.e. Red, Yellow, and Blue, M&Ms amongst all Milk Chocolate M&Ms.

Each of you now has a random sample of roughly the same size. Now, obviously, your one bag is not a very big sample, so we will be creating bigger and bigger samples using some random number software. Each of you will be inputting data into the Google Sheet on the screen in front of you. You will only input data into your row. I will demonstrate each step of the process and also be inputting my data into the row for Sample ID 1. **[At this point give each student a number from 2 to the end of your class.]** Now that you each have a row, please only input data into your row going forward. **[Make sure to point out the difference between the Sample ID value and the underlying row number.]** At this point, you may open your bag of M&Ms, but do not eat them until you have collected the two following pieces of information and input them into your row in the document. The value **x** represents the number of primary colored M&Ms you have and the value **n** is the total number of M&Ms in your bag. Please do not edit any cells that are not blank or that contain messages such as #Div/0!. Once you have inputted your values of **x** and **n** into the Google Sheet on your row and in the red columns, please recount your M&Ms just to verify that your values are correct. Once you have done that, you are welcome to eat your raw data."

10. Open your own bag and input your values of **x** and **n** into the Google Sheet and ask the students to do the same.
11. If needed, walk around and make sure everyone is able to input their values into the Google Sheet.
12. At this point, the values in Column D should automatically be populated with each student's value of p-hat. Explain to each student that this is their individual sample proportion for the population proportion of primary colored M&Ms. It is likely helpful to label these values on the board making sure to make distinctions between p, p-hat, and the distribution of p-hats. Having the formula for the population mean and population standard deviation for the distribution of sample proportions is also advised.
13. The cells below the main field should also display the values of the sample mean and sample standard deviation of these sample proportions. A combined estimate for **p** should also be given, labeled **"p"**, which can be used as the classroom estimate for the population proportion. You can use the rounded average labeled as **"n"** for calculation to find the projected values of the population mean of the sample proportions as well as the population standard deviation of the sample proportions.
14. You can change the Sheet of the spreadsheet by clicking on the different tabs at the bottom. Select the tab named "One Bag Histogram." This shows the distribution of sample proportions that the class has at this point. Class discussion about the shape is encouraged here.
15. Tell the students that they are going to be creating "virtual" larger bags of M&Ms using R Studio.
16. Open the R script entitled MandMProportionClassroomExploration.R in RStudio and have students do the same.

```
#This R script is to allow students to choose random samples
#from the class data.

N <- 20  #Put Number of Sample IDs Here and Execute This Line

#Two Bags
sort( sample( x = 1:N, size = 2, replace = TRUE ) )

#Four Bags
sort( sample( x = 1:N, size = 4, replace = TRUE ) )

#Nine Bags
sort( sample( x = 1:N, size = 9, replace = TRUE ) )

#Nine Bags
sort( sample( x = 1:N, size = 9, replace = TRUE ) )

```

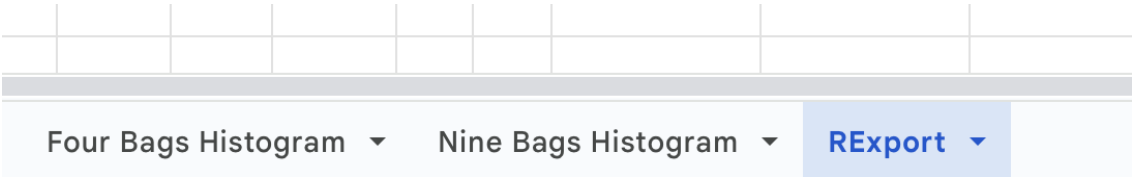
17. Enter the total number of Sample IDs (number of students plus one if the teacher is participating) into row 4 (deleting the value of 20) and execute this row and have students do the same.
18. We will now create larger “virtual” bags of M&Ms by combining random samples of the bags in the red column. Run line 7 of the R Script to generate 2 random numbers. **[You should mention it is likely that a student will not be using their own bag and that there is a chance that their sample may have a single repeated value. If this is the case, treat them as two “identical” bags. A quick discussion about “Resampling” may be necessary.]**
19. Show students how to create their random “double” sized bag by finding the **x** and **n** values in the rows with the Sample IDs found above. Add the **x** values and the **n** values and input the sums into the cells of the orange column of row 1.
20. Have students do the same using their own sample from R and put the values into the cells of the orange columns and the row of their Sample ID.
21. You can now repeat steps 12 through 14 for the orange columns making the requisite changes.
22. Repeat steps 18 through 20 to create virtual bags that are four times as large and then repeat the analysis and discussion in steps 12 through 14.
23. Repeat the above step for virtual bags that are nine times larger than our original samples.

R Data Frame Export

Notes

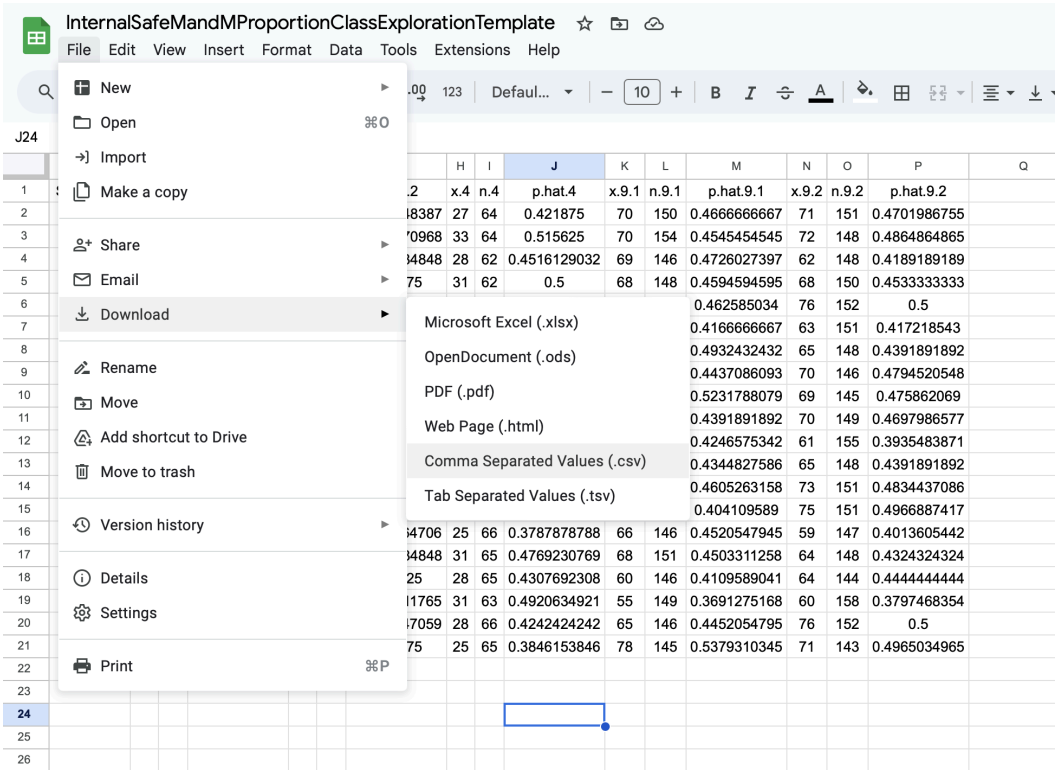
- **The following screenshots are based off random values and not real M&M data.**
- **If all of the values are not entered, this could cause issues when trying to import into RStudio. You can manually delete any cell that shows an error on the RExport sheet to remove most issues.**
- **It would be greatly appreciated if you could email your .csv file to phil.huling@slu.edu to be combined for analysis. That master file should be available for instructor use soon.**

1. Once all the values have been entered, click onto the sheet called RExport. It is the last sheet of the document.

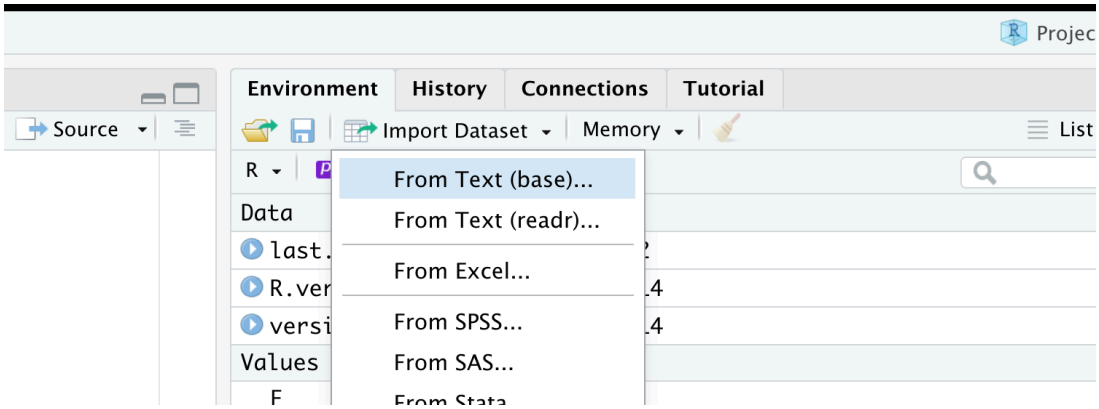


2. Once on the RExport sheet, first check that the data correctly pulled from the Data sheet.
3. Now, click on “File”, then “Download”, and finally “Comma Separated Values (.csv)” as shown below.

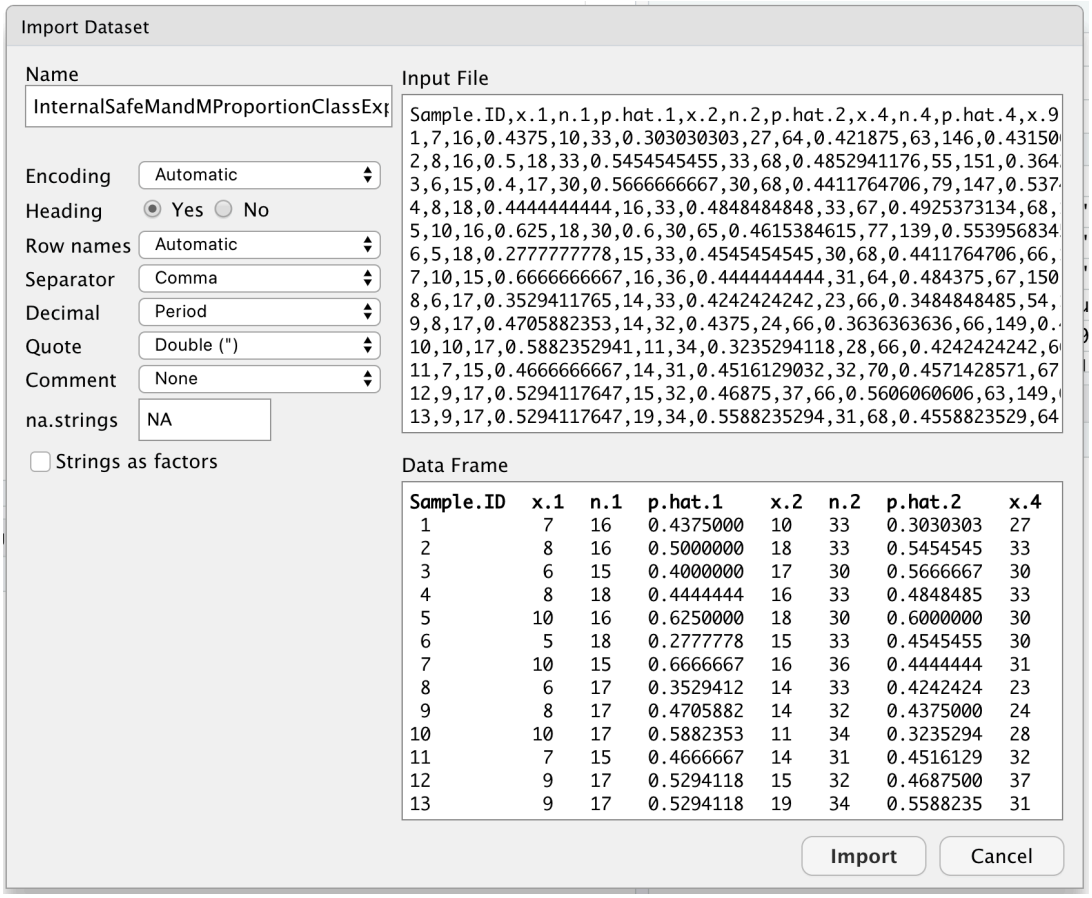
[Note: The below image is from a previous version and has additional columns no longer used.]



4. Now go to RStudio and click on “Import Dataset” and then “From Text (base)...”.

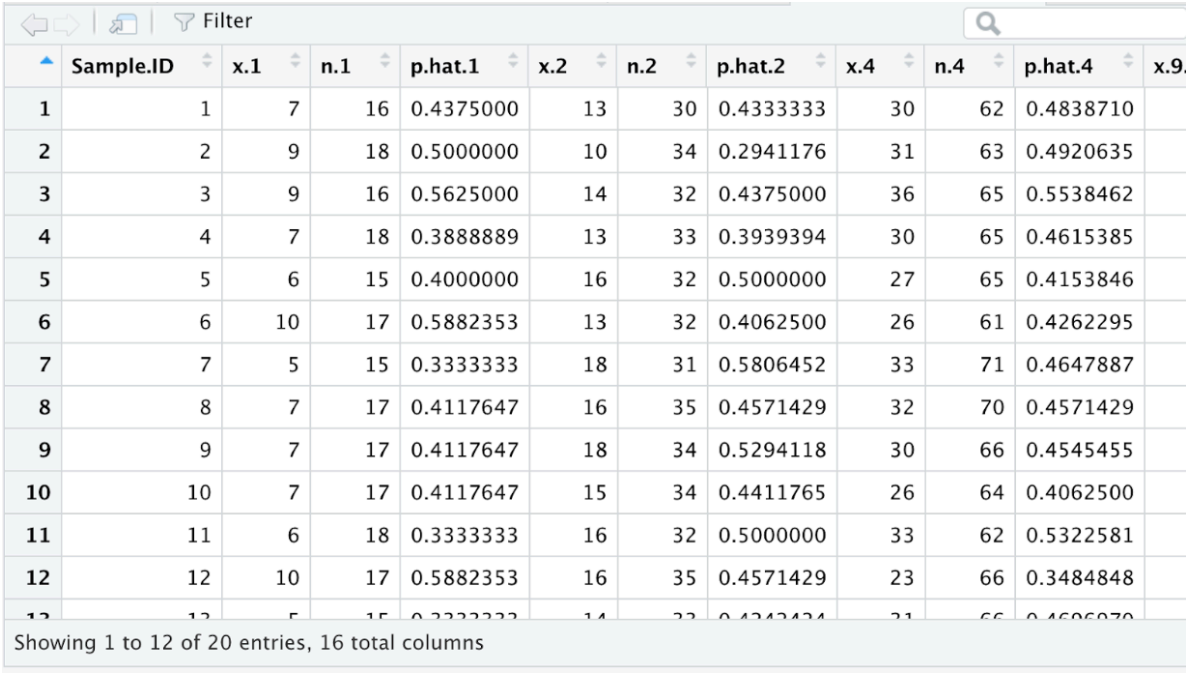


5. Then navigate to where the file was downloaded (likely your downloads folder) and select the file and click “Open.” You should see something like below.



6. Make sure to select “Yes” for “Heading”, but the other options likely will not need modification. You may want to change the “Name” field to something simpler to work with.

7. The file should now be in your Global Environment and it should display on the screen.



Filter

	Sample.ID	x.1	n.1	p.hat.1	x.2	n.2	p.hat.2	x.4	n.4	p.hat.4	x.9
1	1	7	16	0.4375000	13	30	0.4333333	30	62	0.4838710	
2	2	9	18	0.5000000	10	34	0.2941176	31	63	0.4920635	
3	3	9	16	0.5625000	14	32	0.4375000	36	65	0.5538462	
4	4	7	18	0.3888889	13	33	0.3939394	30	65	0.4615385	
5	5	6	15	0.4000000	16	32	0.5000000	27	65	0.4153846	
6	6	10	17	0.5882353	13	32	0.4062500	26	61	0.4262295	
7	7	5	15	0.3333333	18	31	0.5806452	33	71	0.4647887	
8	8	7	17	0.4117647	16	35	0.4571429	32	70	0.4571429	
9	9	7	17	0.4117647	18	34	0.5294118	30	66	0.4545455	
10	10	7	17	0.4117647	15	34	0.4411765	26	64	0.4062500	
11	11	6	18	0.3333333	16	32	0.5000000	33	62	0.5322581	
12	12	10	17	0.5882353	16	35	0.4571429	23	66	0.3484848	

Showing 1 to 12 of 20 entries, 16 total columns