

Lab 10 (M&M's) Wrap-Up

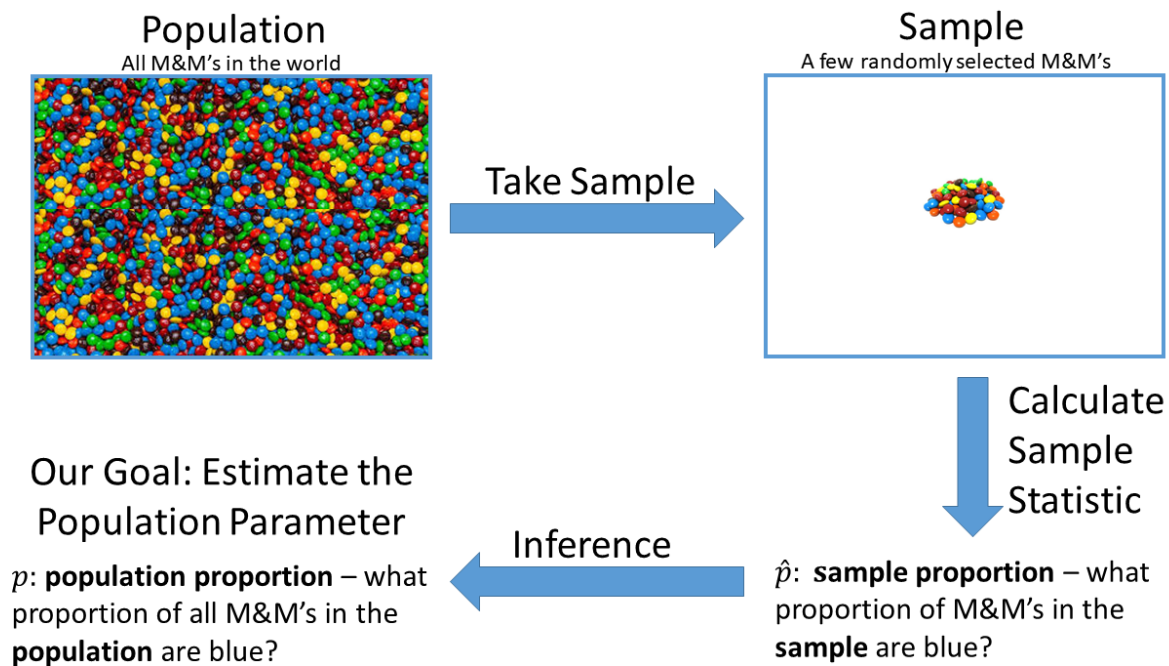
More on Hypothesis Tests and Confidence Intervals for Population Proportions

Note: There are also solutions to Lab 10 on Gryd.

What Proportion of M&M's are blue?

Mars Company (the makers of M&M's) last published the distribution of colors for M&M's in 2008, and at that time the proportion of M&M's that were blue was 0.16. Has the proportion of M&M's that are blue changed since then? Let's investigate.

1. What is the population parameter?



2. Write down suitable null and alternative hypotheses for a hypothesis test of whether the proportion of M&M's that are blue has changed since 2008.

Null hypothesis: $p = 0.16$

Alternative hypothesis: $p \neq 0.16$

3. Take a sample of roughly 10 M&M's, and a second sample of roughly 100 M&M's (the exact sample sizes are not important). In each sample, count how many M&M's were blue.

People who submitted the lab had different numbers of blue M&M's in their samples:

```
##      x  n
## 1   6 10
## 2   2 10
## 3   5 21
## 4   1 10
## 5   6 10
## 6   4 10
## 7   3 10
## 8   1 10
## 9   1 10
## 10  4 10
## 11  0 10
## 12  2 11
## 13  3 10
## 14  0 10
## 15  2 10
## 16  2 10
## 17  0 10

##      x  n
## 1  45 151
## 2  28 100
## 3  47 136
## 4  33 115
## 5  45 115
## 6  30 100
## 7  34 100
## 8  36 121
## 9  21 100
## 10 22 108
## 11 23 100
## 12 51 276
## 13 26 100
## 14 28 100
## 15 18 100
## 16 58 100
## 17 20 100
```

4. Define X to be a random variable that represents the count of how many M&M's in your larger sample were blue. What model could you use for the sampling distribution of X ?

$X \sim \text{Binomial}(100, p)$

This is shorthand for “We model X as following a binomial distribution with sample size $n = 100$ and unknown population proportion p .”

5. Are the conditions met for using the sample to learn about the population, and for using the statistical model specified in part 4?

There are three things to think about:

1. Is the sample **representative** of the population? (always required)
 - Were we careful in **randomly selecting** the people to include in the sample?
 - Are there any forms of **bias** in the sampling? In this case, 23 people did not respond (there is some non-response). Does this matter?

Answer: We took a random sample of M&M's for our sample. It's hard to say for sure if this sample was representative of the population. This was a convenience sample (we drew the M&M's from a single bag). Maybe this bag was produced in a factory that makes more or less blue M&M's than the population as a whole. For the purpose of the lab, let's go ahead with the assumption that this was a representative sample.

2. Are the people in the sample **independent** from each other? (required for all methods we will use in this class)
 - Does knowing one person's response change the information we have about another person's response?
 - Closely related to the question of randomization above.

Answer: Knowing the color of one M&M in my sample doesn't give me any information about the colors of other M&M's in the sample. The assumption of independence seems reasonable.

3. Is the variable we have recorded a **count of the number of people or items in our sample with a certain characteristic**? (required to use the Binomial model)

Answer: We counted the number of M&M's that were blue in our sample. The Binomial model is appropriate.

All of these conditions are satisfied, with the possible exception that we are using a convenience sample. Without any specific information indicating that using M&M's from a single bag is a problem, let's proceed.

6. Computations in R

Since everyone had a different number of blue M&M's in their sample, we all had different results. Here's one example, supposing you had 2 blue M&M's in a sample of size 10 and 33 blue M&M's in a sample of size 100:

For the sample of size 10:

```
binom.test(x = 2, n = 10, p = 0.16)

##
## Exact binomial test
##
## data: 2 and 10
## number of successes = 2, number of trials = 10, p-value = 0.6669
## alternative hypothesis: true probability of success is not equal to 0.16
## 95 percent confidence interval:
## 0.02521073 0.55609546
## sample estimates:
## probability of success
## 0.2
```

For the sample of size 100:

```
binom.test(x = 33, n = 100, p = 0.16)
```

```
##
## Exact binomial test
##
## data: 33 and 100
## number of successes = 33, number of trials = 100, p-value =
## 2.674e-05
## alternative hypothesis: true probability of success is not equal to 0.16
## 95 percent confidence interval:
## 0.2391985 0.4311728
## sample estimates:
## probability of success
## 0.33
```

In the screenshot below, I have annotated where the different pieces of information can be found in this R output:

```
##
## Exact binomial test
##
## data: 33 and 100
## number of successes = 33, number of trials = 100, p-value =
## 2.674e-05
## alternative hypothesis: true probability of success is not equal to 0.16
## 95 percent confidence interval:
## 0.2391985 0.4311728
## sample estimates:
## probability of success
## 0.33
```

p-value (if there's enough space, this may also show up at the end of the previous line).

confidence interval endpoints

point estimate (which is the same as the sample proportion; in this example, our sample had 33 successes in a sample of size 100, so $\hat{p} = 0.33$)

7. Based on the R output from part 6, what is your point estimate of the proportion of M&M's in the population that are blue? You can state this for just the larger sample.

If we took a sample of size $n = 100$ and observed $x = 33$ blue M&M's in the sample, the sample proportion is

$$\hat{p} = \frac{33}{100} = 0.33.$$

This is our point estimate of the population proportion.

8. Based on the R output from part 6, what is a 95% confidence interval for the proportion of M&M's in the population that are blue? Do this for both samples.

Based on the sample of size 10, I am 95% confident that the proportion of M&M's in the population that are blue is between 0.025 and 0.556.

Based on the sample of size 100, I am 95% confident that the proportion of M&M's in the population that are blue is between 0.239 and 0.431.

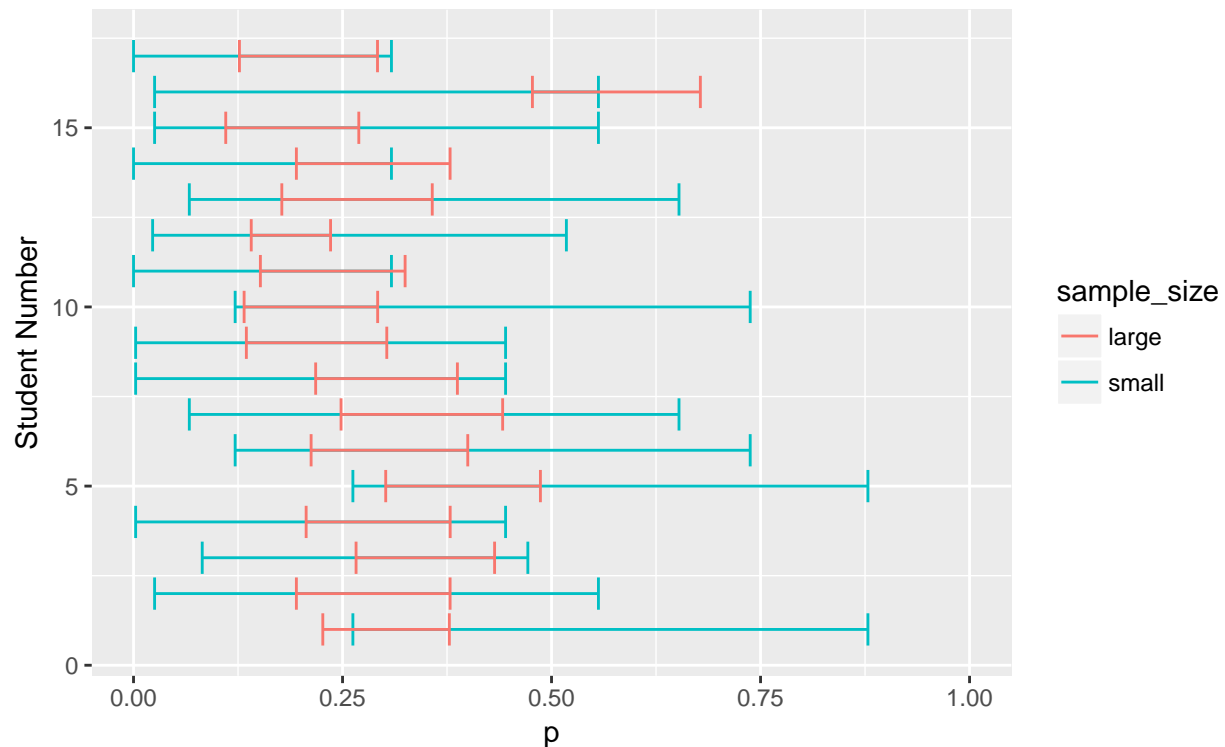
9. What is the interpretation of the confidence interval you found in part 8 based on the larger sample?

If we were to take a large number of different random samples of size 100 from the population, and compute a different 95% confidence interval for the population proportion p based on each of those samples, about 95% of those different confidence intervals would contain the true proportion of M&M's that are blue in the population.

I calculated the confidence intervals based on everyone's different samples (code not shown). Here are the first few rows of the results:

##	student_num	x	n	p_hat	ci_lb	ci_ub	sample_size
## 1	1	6	10	0.600000	0.262378077	0.8784477	small
## 2	2	2	10	0.200000	0.025210726	0.5560955	small
## 3	3	5	21	0.2380952	0.082175876	0.4716598	small
## 4	4	1	10	0.100000	0.002528579	0.4450161	small
## 5	5	6	10	0.600000	0.262378077	0.8784477	small
## 6	6	4	10	0.400000	0.121552258	0.7376219	small

Here is a plot showing the confidence intervals from the 17 different people who submitted the lab.



10. The width of a confidence interval is the difference between the upper limit and the lower limit. The width is a measure of our level of uncertainty about the value of the population parameter (a wider interval indicates more uncertainty about the population parameter value). Find the width of each confidence interval you found in part 8. Is the confidence interval based on a large sample wider or narrower than the confidence interval based on a small sample? Does this make sense?

In the plot above, the confidence intervals based on small samples were wider than the confidence intervals based on large samples.

A larger sample gives us more information about the population.

11. Based on the R output from part 6, what is the p-value for the hypothesis test you set up in part 2? What is the interpretation of that p-value, and what is your conclusion for the test?

In my output from number 6 above, the p-value for the test is $2.674\text{e-}05 = 0.00002674$.

The p-value is the probability of observing a test statistic at least as extreme as what we observed in this sample, if the null hypothesis was true.

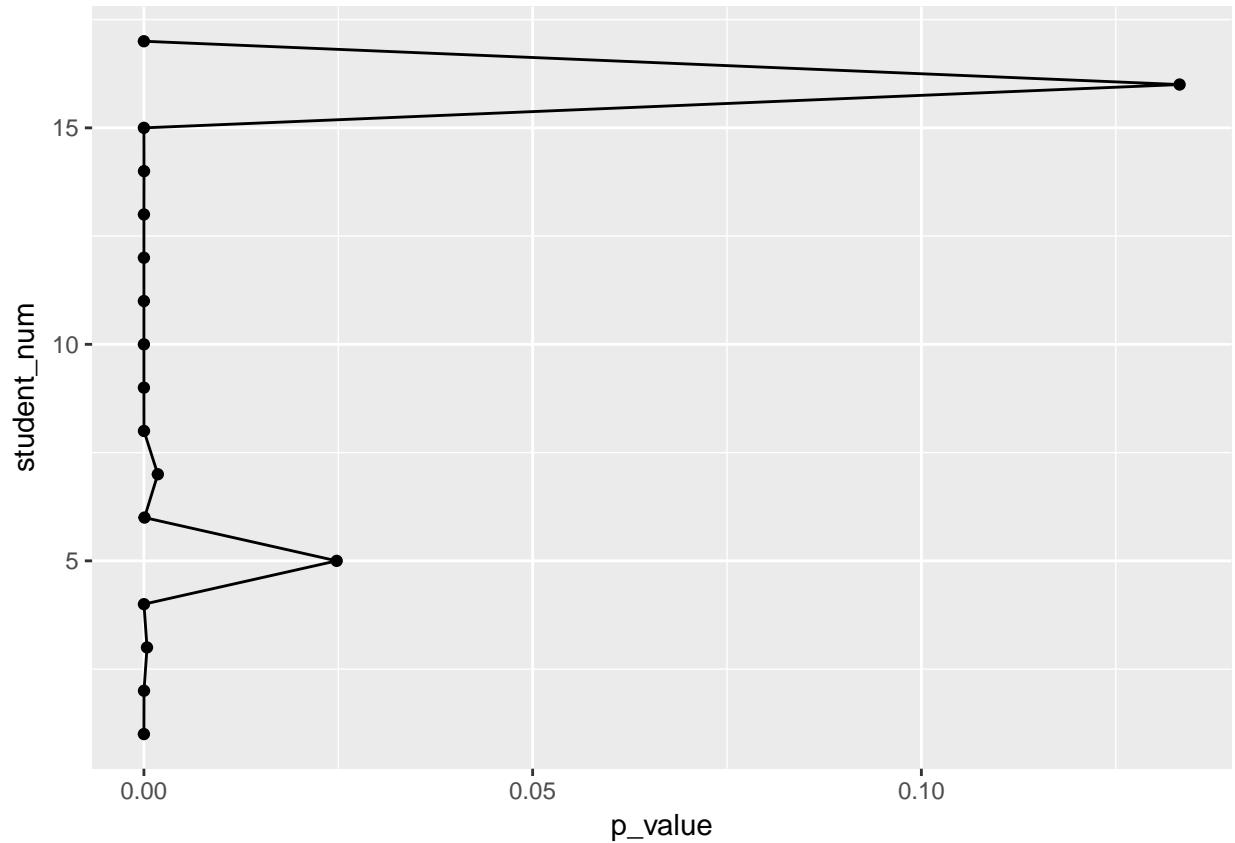
To draw a conclusion for the test, we compare the p-value to significance level cut-offs. For example, since the p-value is less than 0.001, the data provide very strong evidence that the population proportion of M&M's that are blue is not equal to 0.16.

Just as we all got different confidence intervals based on our different samples, we also all got different p-values!

Here are the p-values based on the different samples we got:

##	student_num	x	n	p_hat	p_value	sample_size
## 1	1	45	151	0.2980132	7.494000e-07	large
## 2	2	28	100	0.2800000	1.257915e-05	large
## 3	3	47	136	0.3455882	3.985665e-04	large
## 4	4	33	115	0.2869565	5.571423e-06	large
## 5	5	45	115	0.3913043	2.479354e-02	large
## 6	6	30	100	0.3000000	7.850140e-05	large
## 7	7	34	100	0.3400000	1.789930e-03	large
## 8	8	36	121	0.2975207	9.814991e-06	large
## 9	9	21	100	0.2100000	4.337367e-09	large
## 10	10	22	108	0.2037037	3.988244e-10	large
## 11	11	23	100	0.2300000	5.513581e-08	large
## 12	12	51	276	0.1847826	3.030173e-27	large
## 13	13	26	100	0.2600000	1.667363e-06	large
## 14	14	28	100	0.2800000	1.257915e-05	large
## 15	15	18	100	0.1800000	6.147807e-11	large
## 16	16	58	100	0.5800000	1.332106e-01	large
## 17	17	20	100	0.2000000	1.115909e-09	large

Here is a plot of the p-values for the test we set up in part 2, of the null hypothesis that $p = 0.16$.



12. Suppose we used a significance level of 0.05 to make our decision in the hypothesis test. Who would reject the null hypothesis that $p = 0.16$? Who would fail to reject the null hypothesis?

13. Is there any chance that everyone in the class could have made the “correct” decision in the hypothesis test based on their samples?