Reese's Pieces Activity: Sampling Distribution of $\hat{p}$

Inspired by Activity 13-1 in *Workshop Statistics: Discovery with Data*
Allan J. Rossman and Beth L. Chance

1. Have you noticed that orange is a prominent color with the Reese's brand of candy? What percent of Reese's Pieces so you suppose are orange? Go ahead, guess: ________

2. I will now take a random sample of 25 Reese's Pieces.

   My sample proportion is ________. We have a special symbol for this: $\hat{p}$ ("p-hat")

   So now we know, right? Well, maybe. How likely is it that MY ONE SAMPLE is a PERFECT representation of the population of ALL Reese's Pieces?

3. We should also examine my method. If a sample is to represent the population, it must be

   _____________________. And one of the best ways to guarantee that is with a sample

   that is ___________________. How did I do?

4. Another condition that must be true (remember binomial/Bernoulli conditions?):

   I__________________ How did I do on this one?

5. One more question: what if my sample contained only three pieces? There's no way I could have detected an accurate estimate of the percent orange. So my sample must also be ____________ enough. Was it?

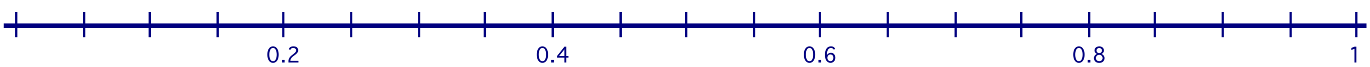
6. OK. I think we're ready to believe that I did in fact have a good representative sample that also fits a binomial/Bernoulli model. But even if my sample is a good one, what do you think will happen if I take another sample of 25 and calculate the percent orange?

7. And better yet, what would happen if we took LOTS of samples of size 25 and calculated the percent orange?

8. So let’s find out how LOTS of sample proportions behave. Take as many samples of 25 as you can, and calculate the percent orange.

Sample #1: _______% Sample #2: ________% Sample #3: _________%

Make a dot plot of the ***class*** $\hat{p}$’s below. Use $\hat{p}$’s instead of dots.



REMEMBER: This is a distribution of st__________________, NOT d_________.

Therefore, it is a very different distribution called a s____________________ distribution.

Describe our class *sampling distribution* below: (remember shape, center, spread?)

Notice that you now have some idea of the *sampling variability:* values of statistics ($\hat{p}$’s) from many different samples will vary somewhat from sample to sample.

ANSWERS to Reese's Pieces Activity: Sampling Distribution of $\hat{p}$

Inspired by Activity 13-1 in *Workshop Statistics: Discovery with Data*
Allan J. Rossman and Beth L. Chance

1. Have you noticed that orange is a prominent color with the Reese's brand of candy? What percent of Reese's Pieces so you suppose are orange? Go ahead, guess: ________

2. I will now take a random sample of 25 Reese's Pieces.

   My sample proportion is ________. We have a special symbol for this: $\hat{p}$ ("p-hat")

   So now we know, right? Well, maybe. How likely is it that MY ONE SAMPLE is a PERFECT representation of the population of ALL Reese's Pieces?
   Probably not very likely...there will be some sampling variability from sample to sample…

3. We should also examine my method. If a sample is to represent the population, it must be representative. And one of the best ways to guarantee that is with a sample that is random How did I do?

4. Another condition that must be true (remember binomial/Bernoulli conditions?): independence between Reese's Pieces How did I do on this one? From previous chapters on Binomial probabilities, we learned that as long as your sample is not to big, we can assume we are "close enough" to independence to not worry. As long a n < 10% of the population, the independence condition is "met."

5. One more question: what if my sample contained only three pieces? There's no way I could have detected an accurate estimate of the percent orange. So my sample must also be large enough. Was it? Again, from previous chapters, we learned that as long as np and n(1 – p) are both greater than 10, we can assume our sample is large enough (Actually, we learned that it is large enough to use a Normal Model instead of the Binomial Model, which is what this sampling distribution will involve…later…).

6. OK. I think we're ready to believe that I did in fact have a good representative sample that also fits a binomial/Bernoulli model. But even if my sample is a good one, what do you think will happen if I take another sample of 25 and calculate the percent orange? Other samples could produce different sample proportions, even if they are random.

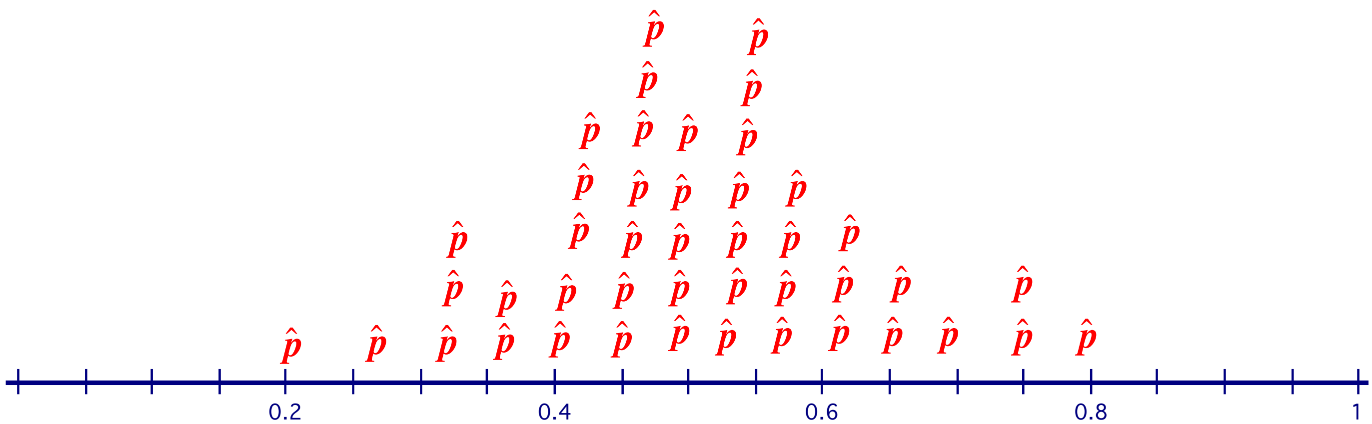
7. And better yet, what would happen if we took LOTS of samples of size 25 and calculated the percent orange? We would begin to get an idea of how *sample* *proportions* behave. We would see how much variation there is from sample to sample. (If we know how sample proportions behave, we can then begin to understand how confident we can be that our ONE sample taken ONE TIME is near the true proportion of orange pieces.)

8. So let's find out how LOTS of sample proportions behave. Take as many samples of 25 as you can, and calculate the percent orange.

Sample #1: _______% Sample #2: ________% Sample #3: _________%

Make a dot plot of the ***class*** $\hat{p}$'s below. Use $\hat{p}$'s instead of dots.

Answers may vary, but it might look something like this:



REMEMBER: This is a distribution of statistics, NOT data.

Therefore, it is a very different distribution called a sampling distribution.

Describe our class *sampling distribution* below: (remember shape, center, spread?)

Answers may vary, but might be something like this: "The distribution is approximately normal, the mean is ≈ 0.47 and the standard deviation is ≈ 0.10"

Notice that you now have some idea of the *sampling variability:* values of statistics ($\hat{p}$'s) from many different samples will vary somewhat from sample to sample.