Introduction to Statistics
Activity 7.4a **SOLUTIONS**

Reese's Candies Colors (cont.)

Our class results suggest that sample proportions vary depending on which sample you happen to get, but there is a *pattern* to this variation. To investigate this pattern more thoroughly, however, you need more samples. Because it is time-consuming to literally sample candies, you will use technology to ***simulate*** the sampling process.

l) To perform these simulations, you need to assume you know the actual value of the parameter p. In lecture, we used a sample size of 75 candies with $p = 0.45$. Now consider drawing samples of 75 candies from a different population where 25% of the candies are orange ($p = 0.25$). What does CLT predict the shape, center, and spread will be, compare this to the sampling distribution found when simulating this in lecture (when $p = 0.45$, we saw normal shape, se = 0.05, and mean = 0.45)?
Same shape (~ normal) and spread (se ~ 0.05), but center will move to $\mu_{\hat{p}} = 0.25$.

l) Use our class GoogleDoc applet (Sampling Distr for Proportions) to draw 1000 samples of 75 candies from a population with $p = 0.25$. Record the mean and standard deviation of these values and comment on the shape. How does this distribution compare to the previous one? Does this match your predictions?

Mean $\mu_{\hat{p}}$ of p̂ values: ~ 0.25 Standard error of p̂ values: ~ 0.05

Shape: ~ normal

m) If you go back to assuming the population proportion of orange candies is $p = 0.45$, what does CLT predict for the shape, the mean, and standard error of the distribution of sample proportions when the sample consists of $n = 25$ candies? Do these values come close to the applet's simulated results seen in lecture (Normal, se = 0.1, mean = 0.45)?

Shape: ~ normal

Center, theoretical mean of p̂ values: $\mu_{\hat{p}} = 0.45$

Spread, theoretical se of p̂ values: $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.45 \times 0.55}{25}} \approx 0.09950.$

Come close? YES!

n) Repeat the previous question for a sample size of n = 75. Compare your theoretical answers to the applet results seen in lecture (Normal, se = 0.06, mean = 0.45)?
Shape: ~ normal
Center, theoretical mean of p̂ values: $\mu_{\hat{p}} = 0.45$
Spread, theoretical standard deviation of p̂ values: $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.45\times0.55}{75}} \approx 0.05745$
Come close? YES!

o) Write a one-sentence interpretation of the se value that you calculated in part n.
[*Hint:* Think about what std dev implies in general and apply that interpretation to this context.]

On avg, samples of 75 candies will have their proportion of orange candies be about 5.745% away from 45%.

p) Now suppose only 10% of Reese's Pieces are orange. Use the applet to draw 1000 samples of size n = 25 in this case. Reproduce a sketch of this sampling distribution.

q) Explain why it makes sense that this sampling distribution is not symmetric.
It is scrunched up near 0%. There are many more ways to be above 10% (3,4,5,6,7,etc. candies) than there are to be below 10% (0,1,2 candies).

r) Explain why the observation that this sampling distribution does not follow a normal distribution is *not* a contradiction to the Central Limit Theorem.
[*Hint:* Consider the conditions needed for the normal approximation to be valid.]
Does not satisfy the technical conditions: $np = 25 \times 0.1 = 2.5 < 10$.

s) Now change sample size to n = 250 and continue to suppose that 10% of the candies are orange. Generate 1000 random samples. Now, does the sampling distribution appear to follow a normal distribution? Explain why the Central Limit Theorem predicts this.
Yes, because now the technical conditions are met:

SRS **and** $np = 250 \times 0.1 = 25 > 10$ **and** $n(1-p) = 250 \times 0.9 = 225 > 10$.