Recall from last unit:

Sampling Distributions for Sample Proportions

- The **mean** of the sampling distribution is the same as the proportion of the population:

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

- The **standard deviation** of the sampling distribution gets smaller according to this equation:

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

Notice how with proportions, we calculate our own SD from the info given

- We assume **Normality** through three checkpoints:

1. $np \geq 10$
2. $n(1-p) \geq 10$
3. Assuming that the sample size is less than 10% of the population

- Calculate **probabilities** using *normalcdf*

Madison's iPad statistics shows that she spends 48% of her daily time playing Farm Story while Bay uses 34% of her daily time on the game. Lola and Sam who are curious only because they want to be in another stats problem, want to see who really is more dedicated to their game so they decide to do their own sampling. Lola takes a sample of 20 days of Madison's iPad stats while Sam, preoccupied with scary TikTok's she's been watching recently only takes a sample of 10 days for Bay.

Sampling Distributions for **Differences** in Sample Proportions

- The **mean** of the sampling distribution is the same as the proportion of the population:

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

- The **standard deviation** of the sampling distribution gets smaller according to this equation:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- We assume **Normality** through the same checkpoints but for both proportions:

$n_1 p_1 \geq 10$ $\quad$ $n_2 p_2 \geq 10$

$n_1(1-p_1) \geq 10$ $\quad$ $n_2(1-p_2) \geq 10$

Assume that the samples are both independent and less than 10% of each population

- Calculate **probabilities** using *normalcdf*

Madison's iPad statistics shows that she spends 48% of her daily time playing Farm Story while Bay uses 34% of her daily time on the game. Lola and Sam who are curious only because they want to be in another stats problem, want to see who really is more dedicated to their game so they decide to do their own sampling. Lola takes a sample of 20 days of Madison's iPad stats while Sam, preoccupied with scary TikTok's she's been watching recently only takes a sample of 10 days for Bay.

Are each proportions large enough for their sample sizes to be large enough to meet sample proportion assumptions?

$$n_M p_M \geq 10$$

$$20 p_M \geq 10$$

$$p_M \geq 0.5$$

$$p_M = 0.48$$

$$n_B p_B \geq 10$$

$$10 p_B \geq 10$$

$$p_B \geq 1$$

Looks like the sample size is too small especially in the case of Bay's numbers

Madison's iPad statistics shows that she spends 48% of her daily time playing Farm Story while Bay uses 34% of her daily time on the game. Lola and Sam who are curious only because they want to be in another stats problem, want to see who really is more dedicated to their game so they decide to do their own sampling. Lola takes a sample of 20 days of Madison's iPad stats while Sam, preoccupied with scary TikTok's she's been watching recently only takes a sample of 10 days for Bay.

Since they both know that they should do a little more research to be sure now, they agree to sample 40 days for both Bay and Madison.

Find the mean and standard deviation of the difference of their sample proportions

$$40(0.48) = 19.2 \geq 10$$

$$40(.34) = 13.6 \geq 10$$

The sample size is less than 10% of the population

They've both been playing Farm Story for at least two years which is longer than 400 days

$$\mu_{\hat{p}_M - \hat{p}_B} = 0.48 - 0.34 = .14$$

$$\sigma_{\hat{p}_M - \hat{p}_B} = \sqrt{\frac{p_M(1 - p_M)}{n_M} + \frac{p_B(1 - p_B)}{n_B}}$$

$$= \sqrt{\frac{0.48(1 - 0.48)}{40} + \frac{0.34(1 - 0.34)}{40}}$$

$$= 0.1089$$

Madison's iPad statistics shows that she spends 48% of her daily time playing Farm Story while Bay uses 34% of her daily time on the game. Lola and Sam who are curious only because they want to be in another stats problem, want to see who really is more dedicated to their game so they decide to do their own sampling. Lola takes a sample of 20 days of Madison's iPad stats while Sam, preoccupied with scary TikTok's she's been watching recently only takes a sample of 10 days for Bay.

Madison, who wants Lola and Sam to understand her dedication to her game, is certain that less than 10% of the time, her Farm Story time will be less than Bay's. Is she right?

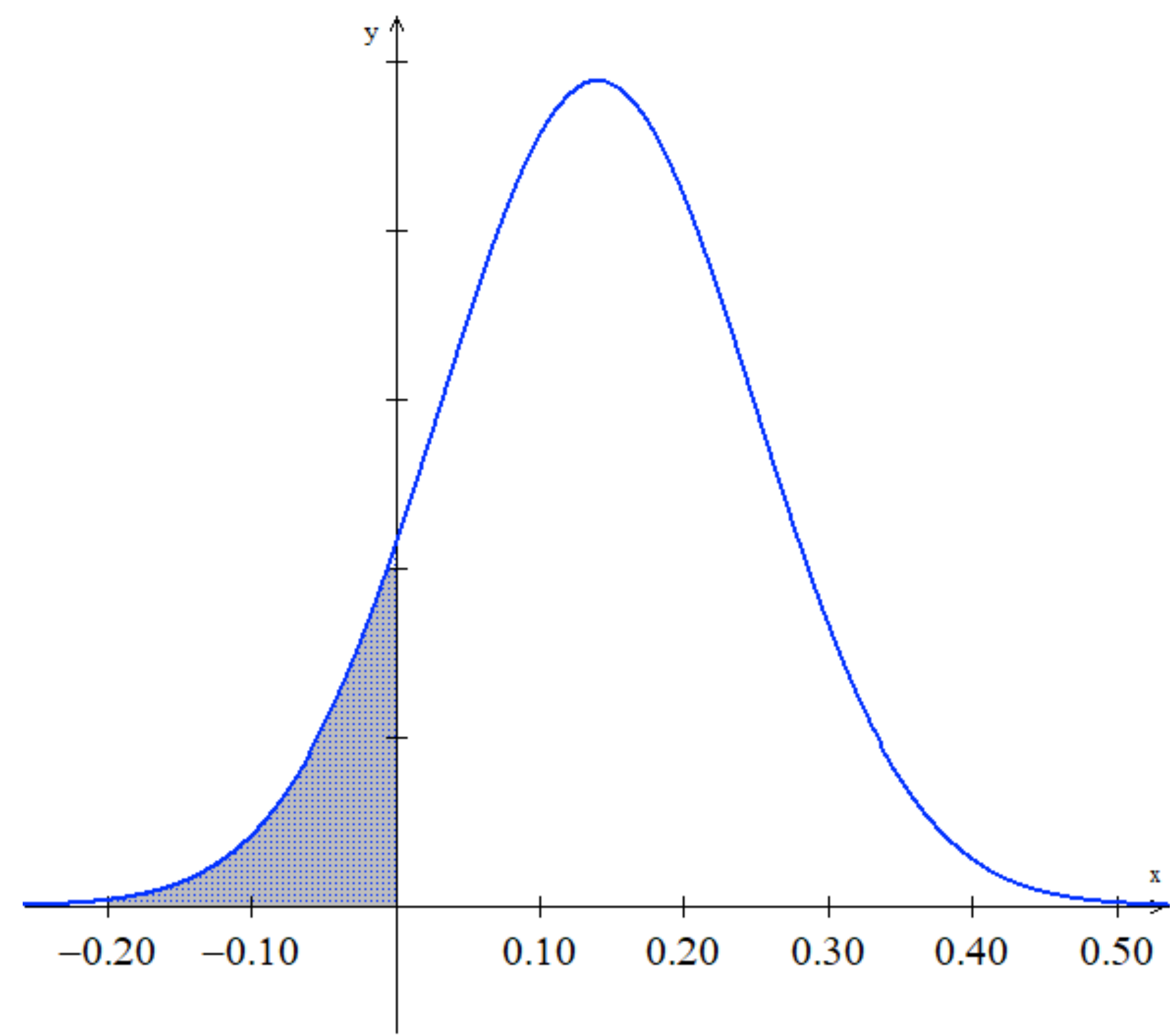
So we are looking for where the distribution is less than 0

$$P(\hat{p}_M - \hat{p}_B < 0)$$

normalcdf(–1E99, 0, 0.14, 0.1089)

0.0992 or 9.92%

Looks like Madison is right…by the tiniest of margins



Madison’s iPad statistics shows that she spends 48% of her daily time playing Farm Story while Bay uses 34% of her daily time on the game. Lola and Sam who are curious only because they want to be in another stats problem, want to see who really is more dedicated to their game so they decide to do their own sampling. Lola takes a sample of 20 days of Madison’s iPad stats while Sam, preoccupied with scary TikTok’s she’s been watching recently only takes a sample of 10 days for Bay.

Since they both know that they should do a little more research to be sure now, they agree to sample 40 days for both Bay and Madison.

You may have noticed that this doesn’t seem very practical especially if we already know the actual proportions. Most of the time in real life we won’t have much actual population data. That’s why we sample in the first place. The main purpose of these problems is to practice working with the formulas before we start working with populations with unknown parameters