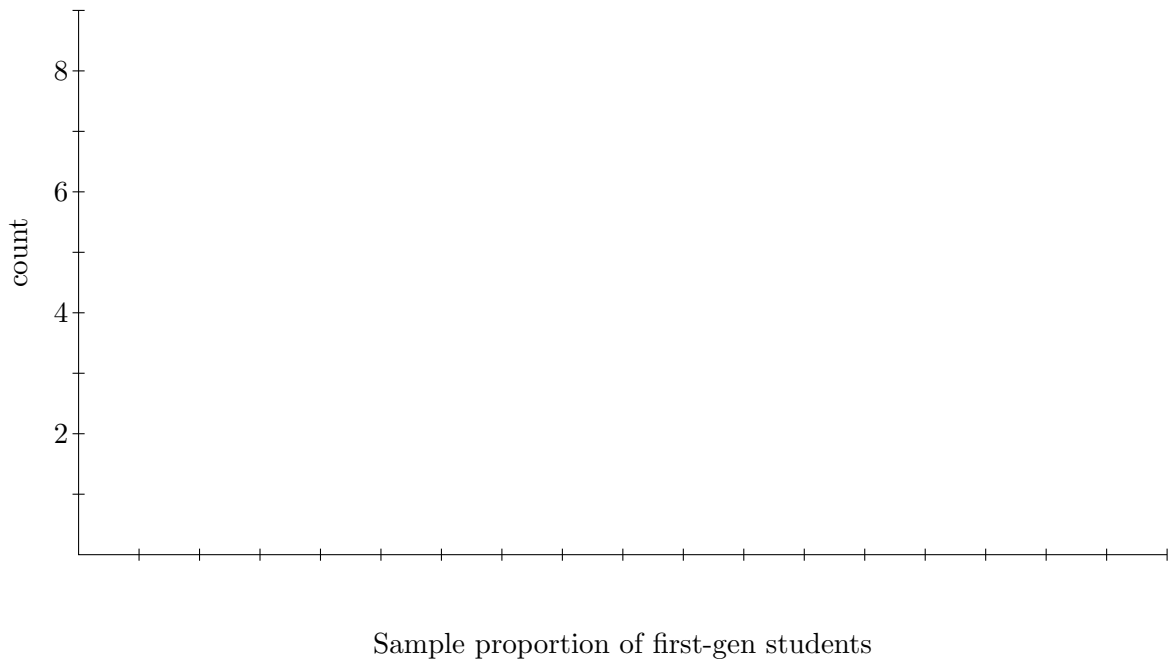


R Activity: Sampling Distributions

Scene:

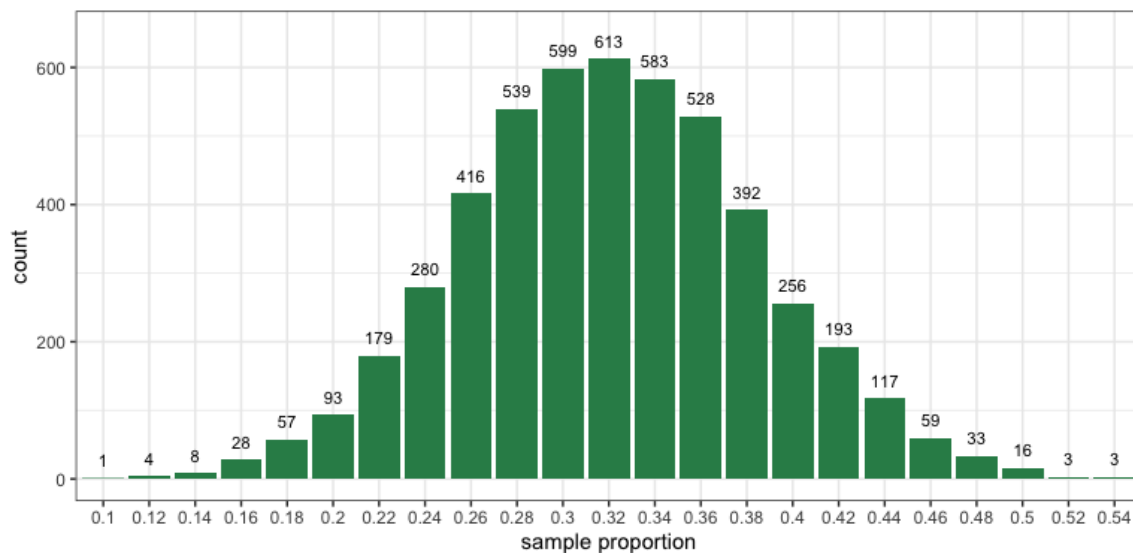
According to the Linfield Fact book for 20-21, updated in April 2021, 32% of the 1,911 Linfield students are first-generation students. (<https://inside.linfield.edu/institutional-research/factbook.html>)

1. If you select a simple random sample of  $n = 50$  students, about how many of them would you expect will be first-generation students? (Give a single number.)
2. Referring to your answer to Q1, how confident are you that you would have exactly that many first-gen students in your sample?
3. Which do you think is more likely if you gather a simple random sample of  $n = 50$  students: that your sample has 6 first-gen students, or that your sample has 16 first-gen students?
4. Now, with RStudio let's each of us gather a simple random sample of 50 Linfield students, by running the code provided.  
Record here the number  $x$  of first-gen students in your sample:  $x = \underline{\hspace{2cm}}$   
Record your sample proportion  $\hat{p}$  of first-gen students:  $\hat{p} = \underline{\hspace{2cm}}$
5. Now we compile the class data. Make a bar plot of the results below (label the tick marks appropriately)



6. Based on the bar plot, about how often should we expect a sample of 50 students to have exactly 16 first-gen students? Express your answer as a percentage. Knowing this, would you change your answer to Q2 above, assuming you gave 16 as your answer to Q1?
  
7. Based on the bar plot, about how likely is it that the sample proportion of first-gen students in a random sample of  $n = 50$  Linfield students is less than 0.16?
  
8. According to our bar plot, how likely would it be to have 25 or more first-gen students in a simple random sample of 50 Linfield students?

Whereas the bar plot on the first page shows the results (i.e., sample proportions) of about 25 to 30 simulations (one for each of you), the bar plot below shows the results of **5,000 simulations** such simulations made in RStudio.



9. Using more simulations, we can get better approximations to the answers to Q6, Q7, and Q8. Using this new bar plot, record below new answers to these three questions.
  - Q6: About \_\_\_\_\_ percent of the samples have  $\hat{p} = .32$ .
  - Q7: About \_\_\_\_\_ percent of the samples fall in this range.
  - Q8: About \_\_\_\_\_ percent of the samples fall in this range.

### Moral

Although the sample proportion  $\hat{p}$  of first-gen students can vary from sample to sample, there is long term distribution for possible values of  $\hat{p}$  from this sampling process, and this distribution is called the **sampling distribution** for  $\hat{p}$ . And hey! This distribution looks sort of normal!