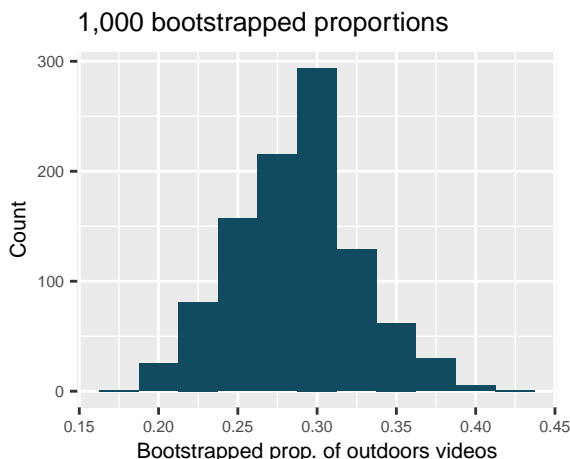


Problem Set 9

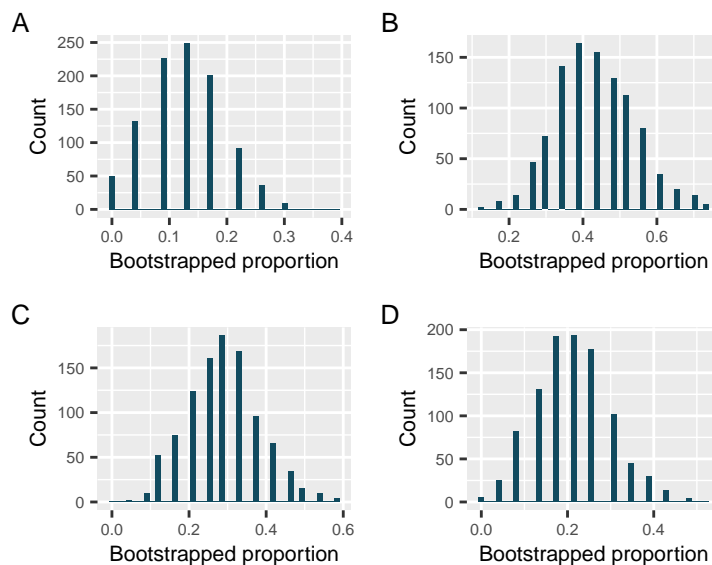
Confidence Intervals with the Bootstrap

1. **Outside YouTube videos.** Let's say that you want to estimate the proportion of YouTube videos which take place outside (define "outside" to be if any part of the video takes place outdoors). You take a random sample of 128 YouTube videos¹ and determine that 37 of them take place outside. You'd like to estimate the proportion of all YouTube videos which take place outside, so you decide to create a bootstrap interval from the original sample of 128 videos.



- Describe in words the relevant statistic and parameter for this problem. If you know the numerical value for either one, provide it. If you don't know the numerical value, explain why the value is unknown.
 - What notation is used to describe, respectively, the statistic and the parameter?
 - If using software to bootstrap the original dataset, what is the statistic calculated on each bootstrap sample?
 - When creating a bootstrap sampling distribution (histogram) of the bootstrapped sample proportions, where should the center of the histogram lie?
 - The histogram provides a bootstrap sampling distribution for the sample proportion (with 1000 bootstrap repetitions). Using the histogram, estimate a 90% confidence interval for the proportion of YouTube videos which take place outdoors.
 - In words of the problem, interpret the confidence interval which was estimated in the previous part.
2. **Bootstrap distributions of \hat{p}** Each of the following four distributions was created using a different dataset. Each dataset was based on $n = 23$ observations.

¹There are many choices for implementing a random selection of YouTube videos, but it isn't clear how "random" they are.



Consider each of the following values for the true population p (proportion of success). Datasets A, B, C, D were bootstrapped 1000 times, with bootstrap proportions as given in the histograms provided. For each parameter value, list the datasets which could plausibly have come from that population. (Hint: there may be more than one dataset for each parameter value.)

- $p = 0.05$
 - $p = 0.25$
 - $p = 0.45$
 - $p = 0.55$
 - $p = 0.75$
3. Work through the tutorial on bootstrapping found at <https://openintro.shinyapps.io/ims-04-foundations-04/>. This tutorial mentions hypothesis testing, a topic that we will be covering next. Skip over those sections for now to focus on forming confidence intervals using the bootstrap. It demonstrates two different ways to use bootstrapping to form a confidence interval: the empirical rule and the percentile bootstrap. We primarily focus on the percentile bootstrap in this class.

Please record in your problem at least one question that you have regarding this tutorial about forming confidence intervals using the bootstrap.