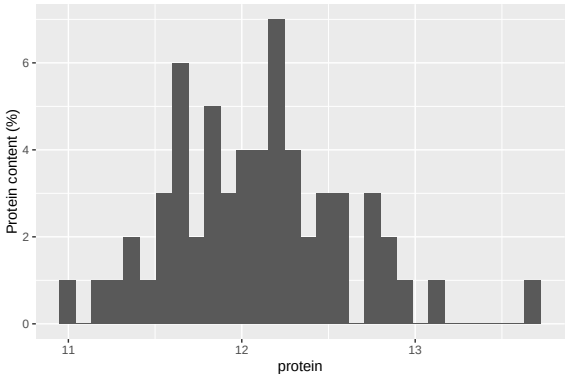
Submit one .pdf file that includes the answers to both sets of questions. Also submit your .R script that you used to answer the questions in the R part.

R Questions

1. (5 points) This question examines the bootstrap approach to sampling distributions, confidence intervals, and hypothesis testing.

   Suppose you are embarking on a baking business. A key ingredient for you is flour, and of most importance is that the protein content has minimal variation. That is, for consistent baked products you need the protein levels to be consistent across bags. You sample and test 60 bags from your current favorite flour brand for their protein content (measured in %), a histogram of the results is shown below.



   (a) Get this protein flour data into R with the following code:

```
protein <- c(12.06, 11.16, 11.35, 11.89, 12.49, 12.19, 11.89, 12.47, 12.42,
11.57, 12.2, 11.04, 12.17, 12.82, 11.81, 11.86, 11.75, 11.82,
12.17, 11.63, 11.54, 12.76, 12.2, 12.13, 12.08, 12.56, 12.77,
13.12, 12.15, 12.07, 11.48, 11.61, 12.28, 12.38, 11.67, 11.67,
11.55, 12.16, 12.92, 11.85, 12.53, 12.29, 12.06, 12.06, 12.01,
12.81, 11.78, 11.66, 11.4, 12.33, 12.21, 11.93, 12.71, 11.65,
12.32, 12.52, 11.84, 12.56, 13.72, 11.29)
```

   (b) What is the point estimate for the standard deviation of protein content for the population of bags of this brand of flour?

   (c) Write a function called `boots()` that takes a bootstrap sample of size 60 from the flour protein data; and calculates and returns the standard deviation of the bootstrap sample.

   (d) Now use your function and `replicate()` to take and calculate the standard deviation of 10,000 bootstrap samples. Make a histogram of the results with `qplot()`.

   (e) Find the 95% bootstrap confidence interval using the percentile method and `quantile()`. One of the arguments for `quantile()` is `probs`, which allows you to specify the empirical quantiles you want returned. See Lecture 6 (Module 8) to review the percentile method.

   (f) Suppose we want to use bootstrapping to test the following null and alternative hypotheses at the 5% level.

$$H_0 : \sigma = 0.5$$

$$H_A : \sigma \neq 0.5$$

In two sentences at most, what do we conclude and why?

2. (3 points) Not surprisingly, there is an R package called `bootstrap`, with a function `bootstrap()` that can expedite the bootstrap procedure.
   (a) Load the `bootstrap` package (you may need to install it first). Take a look at the help and then use `bootstrap()` to replicate your work in 1(d), i.e. take and calculate the standard deviation of 10,000 bootstrap samples. (Hint: the argument `theta` should take the value `sd`) Store the result in `strap`
   (b) Examine the contents `strap` using `str()`. Can you see where the 10,000 bootstrap values are stored? Pull them out and draw a histgram of them.
   (c) Use `quantile()` to find the 95% bootstrap confidence interval from `bootstrap()`, and then repeat Problem 1, part (f).

Conceptual Questions

Answer **two** of the following questions.

3. (1 point) A psychologist is researching the difference between people in their twenties and people in their fourties in how long they sleep. He asks a random sample of 9 people in their twenties and 11 people in their fourties, how much they sleep (in hours) during a typical night. The responses are:

Twenties	4	6	5	9	8
	9	3	5	7	
Fourties	9	5	8	8	7
	9	8	8	9	6
	4				

   (a) Calculate a rank-sum statistic for the twenties group by hand.
   (b) What is a potential problem with using the rank-sum statistic to analyze this data?

4. (1 point) A pathologist wishes to study the length of sinus inflamation due to rhinosinusitis. She asks a random sample of people who have suffered from rhinosinusitis how long they had the inflamation (in days). Their responses are as follows:

Length of time in days of inflamation							
7	14	21	20	19	30	22	18

   She wishes to test the hypothesis that the center of the distribution of inflamation times is 21 days. Calculate a signed-rank statistic for this hypothesis by hand.

5. (1 point) A state park ranger would like to test the hypothesis that visitors spend about a typical time of 3 hours at the state park at which he works. He records the time at which cars arrive and depart, and calculates the time spent in the park in hours. He obtains the following information:

Planned camping times (in hours)									
3.6	3.3	1.2	3.6	2.3	5.6	3.4	3.5	3.1	2.6

Use the signed-rank test to test the hypothesis using R and state a conclusion in the context of the problem.

6. (1 point) A musician would like to make inference about the median number of times per year people go to musical concerts.
   (a) How might bootstrapping help him?
   (b) How would he go about performing the bootstrap?

7. (1 point) A potato farmer wants to estimate the population standard deviation of the weights of potatoes at his farm, which consists of 20 rows. He selects a random sample of potatoes from the two closest rows to a stream and weighs them. He then proceeds to use bootstrapping to obtain an estimate of the desired standard deviation. Explain what is wrong with this scenario.

8. (1 point) A representative of a company at a career fair has ten different stacks of fliers. She counts the number of fliers each interested person takes. Her company wants her to estimate the mean number of fliers people will take with this particular set of ten fliers. However, only three people are interested in this company so she has only three observations. She heard that bootstrapping is a way to obtain an estimate of the mean without knowing anything about the population. Should she proceed with bootstrapping to obtain her estimate? Why or why not?