The following questions are based on a dataset from a 10 kilometer Nordic Ski race in Trondheim, Norway. The International Ski and Snowboard Federation (FIS), organizes and records a series of events such as the World Cup which exists for all disciplines of FIS such as Nordic Skiing. World Cup races occur every weekend between January and March, with events hosted in different countries. These events consist of multiple different races, held in different disciplines and distances. Nordic skiing is split up into two different techniques, classic and skate, all race weekends have a mixture of skate and classic technique races. The race in this dataset is a womens' 10k classic race set in Trondheim, Norway. In long distance pieces like 10ks there is often a lot of discussion about pacing. Racers' times are recorded at different distance intervals throughout the course. I think it is interesting to look at how well racers' times at different distances predict their rank at the end of the race. I think especially the first time is most interesting to see with regard to the final rank, to see if the same racers who have top times in the first distance have good ranks at the end of the race. In this csv in particular, I created a variable for the speed for each time interval using mutate.

1. Create a simple linear regression model of Time4 predicted by Time1. Check the conditions for the model, to assess its variance, linearity, and normality.

2. Find the correlation between Time1 and Rank4. Is there a strong correlation and if so, is it positive or negative?

2. Make a model of Rank4 predicted by Bib. Interpret the coefficient of Bib.

3. Use an ANOVA for means test to decide if there is a considerable difference in the mean speed for each time interval. First, perform the test by constructing a one-way ANOVA model for means by hand. Then, check it by running it through R.

Fill the below ANOVA table with your results: