The following questions are based on a dataset from a Nordic Ski race in Trondheim, Norway on the 17th December 2023. The race is one of many World Cup events held by the International Ski Federation throughout the months of December through March. This race in particular was a 10-kilometer women’s race in the classic technique, with 61 women enrolled. The course for this race consisted of two 5-kilometer loops. In the original data set, there are 4 different distance intervals taken at 1.3, 4.3, 7.5, and 10k respectively.

For these questions, there will be two similar but different data sets. The first is one called ski\_speeds.csv, which I have mutated from the original dataset to include a speed variable for each distance interval which I calculated using the time and distance variables. For the questions dealing with this data set the Rank4 and Speed1 variables will be relevant, since Rank4 refers to finishing place of the skier and Speed1 variable refers to the speed of the skier after the 1.3-kilometer mark.

The second dataset, ski\_speeds\_anova.csv, has been pivoted so that there is a single column for speed and a separate column for the speed interval. The questions with this dataset will work with Rank4 again and Bib number which refers to the starting order of the racers. As well as the Speed and Name variables in which the name variable is the name of the skier.

**Questions 1-2 will be using the ski\_speeds csv:**

1. Create a simple linear regression model of Rank4 predicted by Speed1. Check the conditions for the model, assess its variance, linearity, and residual normality.
   1. Variance:
   2. Linearity:
   3. Residual normality:
2. Find the correlation between Speed1 and Rank4. Is there a strong correlation and if so, is it positive or negative?
   1. Correlation = \_\_\_\_\_\_\_\_\_\_\_\_
   2. Interpretation:

**Use the ski\_speeds\_anova csv for questions 3-4:**

1. Make a model of Rank4 predicted by Bib and Speed. Interpret the coefficient of Bib.
2. Use Randomized Block Design to decide if there is a considerable difference in the mean speed for each time interval and the mean speed for each skier. First, perform the test by hand, writing out hypotheses and fill in the ANOVA table below. Then, check it by running it through R.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **D.F.** | **S.S.** | **M.S.** | **F-Stat** | **P-value** |
| **Name** |  |  |  |  |  |
| **Speed Interval** |  |  |  |  |  |
| **Residuals** |  |  |  |
| **Total** |  |  |

**Factor A:**

**Conclusion:**

**Factor B:**

**Conclusion:**