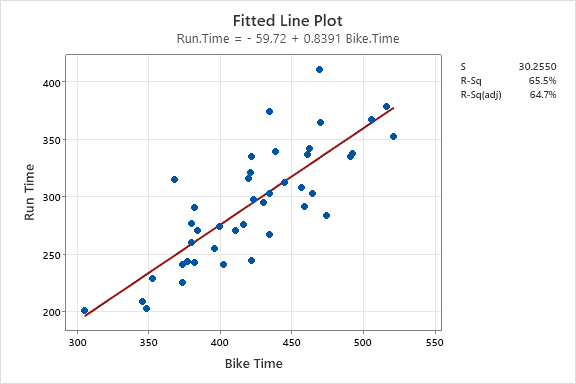
Data: triDataLakePlacidFinal.json

Description: The motivation for this data analysis is to explore the relationships between bike times and run times (in minutes) in order to gain insights into the performance patterns of triathletes. By analyzing this relationship, we can understand the interplay between different segments of the race and potentially identify areas of improvement for athletes. For this activity, we will specifically focus on times from Canadian finishers in the 2018 Lake Placid Ironman.



1. What is the explanatory variable in this situation? What is the response variable? What type are both variables?

Explanatory Variable: Bike Time  
Response: Run Time  
\*\*Both variables are Numerical\*\*

1. What does each point in the scatterplot represent?

Each dot represents a 2018 Canadian Female Lake Placid Ironman Finisher

1. Report the least squares regression equation for predicting price from points.
2. Suppose that you were interested in using this regression model to predict the run time for Sarah True, the USA Olympic Triathlete. What assumption about your data would you need to make?

You need to assume that the relationship between run time and bike time is the same for United States women (like Sarah True) as it is for Canadian women (which is what the regression model is based on).

1. April Clausen had a 470 minute bike time. What is her predicted run time? Regardless of your answer to the previous question, continue to use the equation from question 3.

According to the model, April Clausen

1. April’s run time was 411 minutes. How far off was the model prediction? Explain why we might see this observation.

411 min – 334.3 minutes = 76.7 minutes.  
Answers may vary for the explanation.   
Sample: Since the run is the last leg of the race, her run time may have been slower due to burnout.

1. Interpret the slope of the model in the context of the application. Be sure to be mindful of the units.

For every 1 minute in Bike Time, the Run Time is expected to increase by 0.8391 minutes

1. Interpret the intercept of the model in the context of the application.

When a Bike Time is zero, the expected Run time is -59.72 minutes

1. Is the intercept interpretation meaningful? Explain.

No, this is not a meaningful interpretation.   
  
Neither a Bike Time of 0 minutes or a Run Time of -59.72 minutes is possible.

1. What percent of variation in Run Times is explained by the model using Bike Time?

R2 = 65.5%

1. What is the sample correlation between run times and bike times?
2. Write a few sentences to completely describing the relationship between run times and bike times. Try to incorporate multiple pieces of information from your analysis.

Solutions may vary, but basically, a solution should incorporate a description of the scatterplot supplemented with the results from the regression model. For example,

Lake Placid Canadian Women’s Ironman Triathlon run and bike times have a clear positive relationship with no extreme outliers. This relationship seems linear and fairly strong (resulting in an R-Squared of 65.5%).

1. Recall that the Ironman consists of a 2.4-mile swim, a 112-mile bicycle ride, and a marathon 26.22-mile run (in that order). How might you use this information to investigate the relationship between speed at which a triathlete completes the bike and run portions of the race?

Answers can vary. One possible solution is to convert bike and run events into speeds (e.g., miles per hour). Other solutions could be based on pace (e.g., “X minute mile”)

1. If you have the technology to do so, fit the least squares regression model predicting running speed from biking speed. Record the equation and interpret the slope coefficient in the context of the application.

Answers are dependent on the choice made in the previous question. Here is sample output when converting to speed in miles per hour.

