The dataset **ironman\_lake\_placid\_female.csv** contains data on female finishers of the Lake Placid Ironman Triathlon from 2002 to 2022. An ironman triathlon is comprised of 3 events. he motivation for this data analysis is to explore the relationship between bike and run times (in minutes) in order to gain insights into the performance patterns of the athletes. For this activity, we will specifically focus on times from 2022 Canadian Finishers.



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

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

1. What does each point in the scatterplot represent?

Each dot represents a 2022 Canadian Female Lake Placid Ironman Finisher

1. Report the least squares regression equation for predicting run time from bike time.

$$\hat{RunTime}= -72.36+0.9109(BikeTime)$$

1. 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. Sarah True had a 295.5 bike time. What is her predicted run time? Regardless of your answer to the previous question, continue to use the equation from question 3.

$$\hat{RunTime}= -72.36+0.9109\left(295.5\right)=196.8 minutes$$

1. April’s run time was 184.1 minutes. How far off was the model prediction? Explain why we might see this observation.

184.1 min – 196.8 minutes = -12.7 minutes.
Answers may vary for the explanation.
Sample: The running segment might be her best event; therefore, she is faster than expected

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 0.9109 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 -72.36 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 -72.36 minutes is possible.

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

R2 = 66.6%

$$r= \sqrt{0.666}=0.816$$

1. Based on your comprehensive analysis, describe the relationship between run times and bike times using multiple pieces of information from your findings.

Solutions may vary, but, 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. Given the structure of the Ironman race, where participants complete a 2.4-mile swim, a 112-mile bicycle ride, and a marathon 26.22-mile run in that order, explain how this information can be used to explore the relationship between the average 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.

