**Scenario A: 5k vs 10k Splits**

1.a.

The x variable is the time of the 5K split in the marathon, and the y variable is the time of the 10K split.

2.a. There is a strong, positive, linear association between the time of the 5K split and the 10K split. The runner at (46.45, 27.43) is an influential point, pulling the line down towards it. The runner at (50.08, 41.33) is an outlier because the runner’s 10K time was faster than expected based on his 5K time.

3.a. The explanatory variable is the 5K split, and the response variable is the 10K split. 5K split time is being used as the explanatory variable because it is being used to predict the 10K split time, which is the response variable. The y-intercept is -.592, meaning that if a runner takes 0 minutes to complete the 5K split, the 10K split would take -.592 minutes to complete. The y-intercept is not useful in this scenario because it is not feasible for a person to complete a 5K in 0 minutes, and it is even less possible for a person to run a 10K in negative minutes. The slope is 1.0366. For every 1 minute increase in the 5K split time, there is a predicted 1.0366 minute increase in the person’s 10K split time.

The LSRL equation is: ŷ = -.592 + 1.0366x or (predicted 10K split time) = .592 + 1.0366(5K split time)

4.a. The correlation coefficient is r = .9714. There is a strong, positive, linear association between the 5K split and 10K split times. The correlation coefficient matches the description of the scatterplot. The coefficient of determination is r² = .943564. Approximately 94.36% of the variation in the predicted 10K split times can be explained by the LSRL of 5K split times.

5.a.

### Based on the residual plot, the LSRL is a good model for the data given, as there is no leftover pattern. The equation for calculating residuals is: y – ŷ or actual 10K split time minus predicted 10K split time. For example, the runner that took 17.15 minutes to run the 5K took 17.50 minutes to run the 10K. Plugging the point 17.15 into the LSRL equation gives us 17.19 minutes. Actual y-value minus predicted y-value gives us the residual .31, meaning the runner took .31 minutes longer to complete the 10K than expected based on the LSRL line. The residual is graphed in relation to the x-value, and a residual for every y value is calculated and plotted.

**Scenario B: 10k vs Half Marathon Splits**

1.b.

The x variable is the 5K split time, and the y variable is the half marathon time.

2.b. There is a strong, positive, linear association between 5K split times and half marathon times. The runners at (46.45, 128.00), (50.08, 162.09), and (22.10, 172.55) are outliers, with the one at (46.45, 128.00) likely being an influential point as it pulls down the equation a bit.

3.b. The explanatory variable is the 5K split time, and the response variable is the half marathon time. The slope is 4.23. For every one minute increase in the 5K split time, there is a predicted 4.23 minute increase in the half marathon time. The y-intercept is -4.0568. When 5K split time is 0, the half marathon time is -4.0568. This number is not realistically feasible because time taken to run a 5K can’t be 0, and time can’t be negative.

The LSRL equation is: ŷ = -4.0568 + 4.23x or (predicted half marathon time) = -4.0568 + 4.23(5K split time)

4.b. The correlation coefficient is r = .9741. There is a strong, positive, linear association between 5K split times and half marathon times. The correlation coefficient matches the description of the scatterplot. The coefficient of determination is r² = .948874. Approximately 94.89% of the variation in the predicted half marathon times can be explained by the LSRL of 5K split times.

5.b.

### Based on the residual plot, the LSRL is a good model for the data given, as there is no leftover pattern. The equation for calculating residuals is: y – ŷ or actual half marathon time minus predicted half marathon time. For example, the runner that took 14.90 minutes to run the 5K took 60.20 minutes to run the half marathon. Plugging the point 14.90 into the LSRL equation gives us 58.97 minutes. Actual y-value minus predicted y-value gives us the residual 1.23, meaning the runner took 1.23 minutes longer to complete the half marathon than expected based on the LSRL line. The residual is graphed in relation to the x-value, and a residual for every y value is calculated and plotted.

**Scenario C: Half Marathon vs Total Marathon Time Splits**

1.c.

The x variable is half marathon time, and the y variable is final marathon time.

2.c. There is a strong, positive, linear association between half marathon time and final marathon time. A possible outlier lies at (76.75, 310.37), but there are no influential points.

3.c. The explanatory variable is half marathon time, and the response variable is final marathon time. The slope is 2.2532. For every one minute increase in half marathon time, there is a predicted 2.2532 minute increase in final marathon time. The y-intercept is -14.882. When half marathon time is 0 minutes, final marathon time is -14.882 minutes. This is not realistically feasible because time taken to run a half marathon can’t be 0, and time can’t be negative.

The LSRL equation is ŷ = -14.882 + 2.2532x or (predicted full marathon time) = -14.882 + 2.2532(half marathon time)

4.c. The correlation coefficient is r = .9582. There is a strong, positive, linear association between half marathon times and final marathon times. The correlation coefficient matches the description of the scatterplot. The coefficient of determination is r² = .918178. Approximately 91.82% of the variation in the predicted final marathon times can be explained by the LSRL of half marathon times.

5.c.

### Based on the residual plot, the LSRL is a good model for the data given, as there is no leftover pattern. The equation for calculating residuals is: y – ŷ or actual final marathon time minus predicted final marathon time. For example, the runner that took 117.14 minutes to run the half marathon took 263.09 minutes to run the final marathon. Plugging the point 117.14 into the LSRL equation gives us 249.06 minutes. Actual y-value minus predicted y-value gives us the residual 14.03, meaning the runner took 14.03 minutes longer to complete the final marathon than expected based on the LSRL line. The residual is graphed in relation to the x-value, and a residual for every y value is calculated and plotted.