1. C
2. D
3. B
4. A
5. B
6. C
7. D

8. The scatterplot shows a moderately strong, positive linear association between water temperature in degrees Fahrenheit and fish activity.

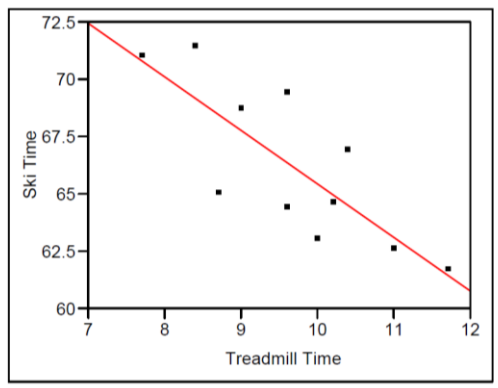
9. The correlation would not change!

10. The point would reduce the correlation coefficient. It is a strong outlier in the y-direction and would reduce the strength of the relationship.

12.

a. Treadmill time is the explanatory variable since we are using treadmill time to predict the results of the ski race.

b. The scatterplot shows that there is a moderate, negative, linear association between the treadmill time and ski time.



c. Predicted Ski Time = 88.795649 - 2.3335102 (Treadmill Time)

d. The LSRL predicts that for each additional minute an athlete is able to run on the treadmill, the time it will take to complete the ski race will decrease by 2.33 minutes.

e. Y: The LSRL predicts that athletes that can last 0 minutes on the treadmill will have a ski time of 88.80 minutes. This doesn't make sense. It isn't reasonable that an athlete couldn't run for any amount of time on the treadmill. Also, 0 is way below the other x-values in the data set, so we would be extrapolating. X: The LSRL predicts that athletes who last 38.11 minutes on the treadmill will take 0 minutes to finish the ski race. This is humanly impossible. Again, 38 minutes is way beyond the x-values in our data set, so we were extrapolating.

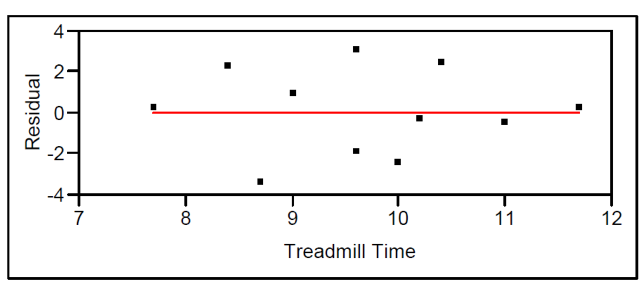
f. r = -0.796. The correlation coefficient tell us that there is a moderately strong linear relationship between treadmill time and ski time.

g. It wouldn't change. The correlation coefficient is not dependent on units.

h. No, both variable probably depend on general fitness. Besides, we would need an experiment to conclude cause and effect. CORRELATION DOES NOT IMPLY CAUSATION

i. $y - \hat{y} = 71 - 70.859 = 0.141$. The model's prediction was only 0.141 minutes too low for the first athlete (very close!).

j. Since there is no pattern in the residual plot, this means that the linear model is appropriate for this data.



k. r^2: 63% of the variability in ski time can be explained by variability in treadmill time (63% of the variability in ski time can be explained by the LSRL relating ski time to treadmill time). s: When using treadmill time to predict ski time, the predictions will be about 2.19 minutes away from the actual times, on average.

l. r^2 wouldn't change at all and s would be 60 times smaller (.0365 hours).

m. The LSRL predicts that athletes who last 8 minutes on a treadmill will complete the ski race in 70.16 minutes.
n. This point will be influential since it's x-value is bigger than the others. However, since it falls very close to the LSRL above, the LSRL will change very little (slope and y-intercept will be similar). However, the other measures will change. For example, s will go down, since the residual for this point will be smaller than average. The value of r will be closer to -1 and r^2 will be closer to 1 since this point gives extra evidence of a negative association (it follows the pattern of the scatterplot).
o. This point will also be influential since it's x-value is bigger than the others. However, since it falls very far from the LSRL above, the LSRL will be affected (the y-intercept will be lower and the slope closer to 0). This also means that s will go up, since moving the line will make the typical residual get bigger. The value of r will be closer to 0 and r^2 will be closer to 0 since this point makes the data more scattered. (The data point falls away from the pattern of the scatterplot)