A graph with a line and dots

Description automatically generated

1. In the scatterplot above what is the explanatory variable? What is the response? What are the units of both variables respectively?

Average start is the explanatory variable (position). Driver rating is the response variable (points).

2. What does each point on the scatter plot represent?

Each point represents a driver that competed in the 2022 season.

3. Assess and describe the trend displayed in the scatterplot.

The trend displays a strong, negative, and linear relationship.

4. Given the trend of the smoother, does a linear trend seem reasonable?

Yes, linearity seems reasonable to assume.

5. Write the population equation for the linear model predicting Driver Rating from Average Start.

β0

6. Fit the least squares regression equation for modeling Driver Rating from Average Start.

A screenshot of a computer error

Description automatically generated

7. Interpret the slope coefficient in the context of the data.

For a one place increase in average starting position, we expect driver rating to decrease by 2.15 points on average.

8. Provide a literal interpretation of the intercept coefficient. Does it make sense in context of the data?

When average starting position is at the place of 0, we expect the driver rating to be 109.24 points on average. This does not make sense as the best starting position would be 1. Further, it represents a case of extrapolation as the lowest seen in the data is approximately 5.

9. Although it is already known, use the model to predict the Driver Rating of Joey Logano who has an AvgStart of 10.9.

85.805

10. Given that Joey Logano's actual Driver Rating in 2022 was 89.5, calculate the residual of the model prediction. Did the model overpredict or underpredict Logano’s actual driver rating?

e = 89.805 - 89.5 = 0.305

The model underpredicted the Logano’s actual driver rating by 0.305 points.

11. What is the size of a typical error for this model?

Residual standard error: 8.384

12. What percent of variation in Driver Rating can be explained by the model using Avg Start? What is the sample correlation?

Roughly 85.4% of total variability in driver rating can be explained by the model using average start.

Sample correlation = r = = -0.924

13. What are the model assumptions for simple linear regression? Are they satisfied here?

A graph of a number of values

Description automatically generated

A graph of a number of points

Description automatically generated

A graph of a number of numbers

Description automatically generated

Model assumptions:

Linearity, Independence, Normality, Equal variance

There is some slight curvature in the relationship, but nothing too extraordinary.

Normality seems reasonably met but there may be some unusual observations.

Overall, conditions seem to be satisfied.