Interpret the coefficients for return points won percentage and grass and clay surfaces.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** |
| Constant | -0.1898 | 0.0994 | -1.91 | 0.057 |
| ReturnPointsWonPercentage | 2.079 | 0.253 | 8.23 | 0.000 |
| EloRank | -0.001332 | 0.000120 | -11.07 | 0.000 |
| Surface |  |  |  |  |
| Grass | 0.1016 | 0.0209 | 4.85 | 0.000 |
| Hard | 0.0269 | 0.0166 | 1.62 | 0.106 |

Return Points: A 1 unit increase in the return points won percentage is associated with a 2.079 increase in win percentage holding all else constant.

Grass: Compared to playing on Clay courts, we would expect the win percentage for playing on grass courts to be 0.1016 higher, on average, holding all else constant.

1. Interpret the R-Squared value and decide if you think that this is a good model.

|  |  |  |
| --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** |
| 0.125993 | 48.20% | 47.57% |

About 47.57% of the variability of win percentage is explained by the multiple linear regression model.

This isn’t a great R-Squared value, so we should look at different models or transformations to our model that might be able to explain the data better.

1. Based off of the following plots, do you believe that this model meets the assumptions of a multiple linear regression model?

A graph with blue dots

Description automatically generatedA graph with a line

Description automatically generated

As we can see in the Q-Q plot on the left, this looks normally distributed as a vast majority of the points follow the line. On our plot of fitted values vs. residuals, the variance seems consistent with a slight narrowness towards the lower values.