Winnings prediction Data Analysis

Student name

Affiliation

Question 1

In the ever-dynamic world of data, multiple regression models find themselves in a challenging situation where one has to determine the suitable variables to be used. A lot of simulation and determination of potential variables for validating models from fitted data takes care of accuracy.

In our case, there are four predictor variables namely: number of wins (wins), number of top ten (Top 10) finishes, number of top five finishes (Top 5) and the number of poles won (poles) while Winnings is the response variable. Using SPSS statistical software, the following is the output in simulation of the predictor variables in order to use the variables that best suits the winnings.

Correlations

			Winnings ($)	Poles	Wins	Top 5	Top 10
Spearman's rho	Winnings ($)	Correlation Coefficient	1.000	.342*	.774**	.806**	.869**
		Sig. (2-tailed)	.	.044	.000	.000	.000
		N	35	35	35	35	35
	Poles	Correlation Coefficient	.342*	1.000	.200	.458**	.446**
		Sig. (2-tailed)	.044	.	.250	.006	.007
		N	35	35	35	35	35
	Wins	Correlation Coefficient	.774**	.200	1.000	.844**	.794**
		Sig. (2-tailed)	.000	.250	.	.000	.000
		N	35	35	35	35	35
	Top 5	Correlation Coefficient	.806**	.458**	.844**	1.000	.932**
		Sig. (2-tailed)	.000	.006	.000	.	.000
		N	35	35	35	35	35
	Top 10	Correlation Coefficient	.869**	.446**	.794**	.932**	1.000
		Sig. (2-tailed)	.000	.007	.000	.000	.
		N	35	35	35	35	35

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 1

In search for the best and suitable single Variable for winnings, the top 10 variable forms the most suitable in relation to the winnings.

Table 1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.390	.302		-1.292	.205
	Top 10	.136	.024	.697	5.587	.000

a. Dependent Variable: Wins

$Y=0.697X_1$

a) For the multiple regression model has the number of top 10 finishes and the number of top 5 finishes as the most suitable variables as shown in the below scatter plot.

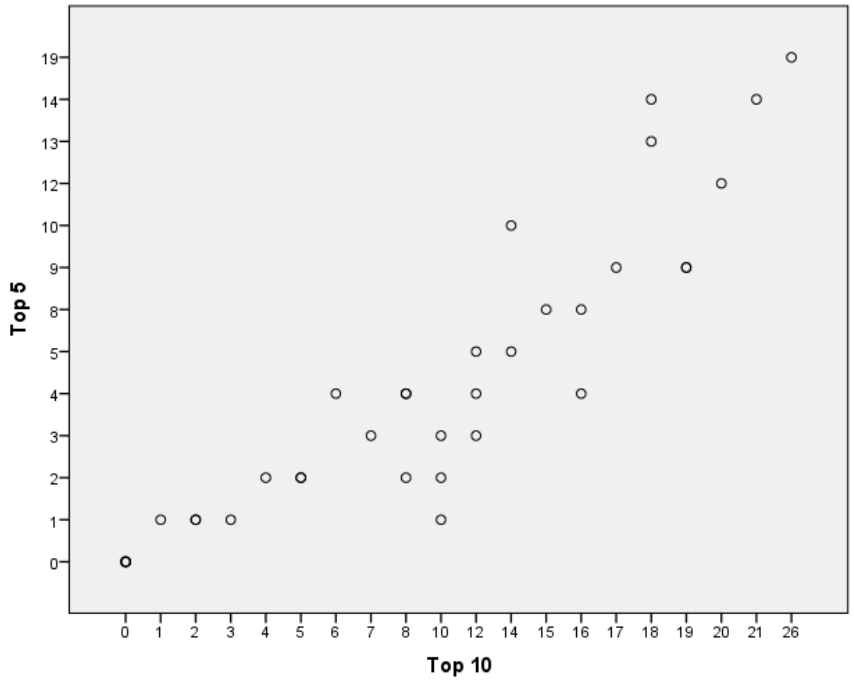

Figure 2

There is a strong positive relationship between the Top 5 finishes and Top 10 finishes and hence form the best explanatory variables for predicting winnings.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.906[a]	.820	.809	563331.584

a. Predictors: (Constant), Top 10, Top 5

Table 2

R square is 0.82 which shows that the model well fits the data provided

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3134168.251	176174.232		17.790	.000
	Top 5	73246.948	45973.342	.276	1.593	.121
	Top 10	116999.351	31276.222	.649	3.741	.001

a. Dependent Variable: Winnings ($)

Table 3

The prediction model for the winnings is as follows:

Y=$0.276X_1 + 0.649X_2 + K$; where X_1 =the number of Top 5 finishes, X_2 =the number of top 10 finishes while Y is the response variable (Winnings).

b) From the analysis above, it is evident that the model correctly fits the data. The predictive model is a productive and stable one. The analysis used an assumption that the independent variables and the dependent variable have a linear relationship. When predicting wins and according to the model fitted, the most helpful variables amongst the given ones, is the top 5 and top 10 finishes.