TASK #2

A multiple linear regression model is built with Wins (W) as the dependent variable and other variables (except team names) as predictors. We have a total of 330 observations. The results of the model are discussed below:

a. Is the model statistically significant at the 0.05 significance level? State the hypotheses and include supporting figures that provide the decision for the test.

To evaluate whether the multiple linear regression model has statistical significance at the 0.05 significance level, we need to perform a hypothesis test. The null hypothesis is that all regression coefficients in the model are equal to zero, and the alternative hypothesis is that at least one regression coefficient is not equal to zero. From the analysis of variance (ANOVA) table presented below, we can view the results of the hypothesis test. The F-value for the model is 174.31, and the corresponding p-value is less than 0.0001, which is much lower than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that the model has statistical significance.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	25	44634	1785.36525	174.31	<.0001
Error	304	3113.67475	10.24235		
Corrected Total	329	47748			

b. Discuss the model's goodness of fit.

Root MSE	3.20037
Dependent Mean	40.02424
R-Square	0.9348
Adj R-Sq	0.9294
AIC	1124.67386
AICC	1129.68048
SBC	891.45027

The multiple linear regression model has a very good fit. From the table above, we can see that the R-squared value for the model is 0.9348, which means that 93.48% of the variability in Wins can be explained by the independent variables in the model. This indicates that the model can capture a significant portion of the variability in the data and has a strong ability to predict Wins. Additionally, the adjusted R-squared value for the model is 0.9294. Adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in the model. In this case, the adjusted R-squared is only slightly lower than the R-squared value, which indicates that the model is not overfitting the data by including too many predictors. This is a good sign, as overfitting can lead to poor predictive performance on new data.

c. Which predictors, if any, are significant and which are not? Include the supporting figures in your answer.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-91.969348	34.737229	-2.65	0.0085
Age	1	0.414717	0.134010	3.09	0.0022
SOS	1	0.099511	0.742871	0.13	0.8935
ORtg	1	1.770542	0.850755	2.08	0.0383
Pace	1	-0.081053	0.100404	-0.81	0.4201
FTr	1	-5.286883	40.948065	-0.13	0.8974
3Par	1	-4.355997	6.164141	-0.71	0.4803
eFG%	1	90.256938	129.281402	0.70	0.4856
TOV%	1	-0.638439	1.217515	-0.52	0.6004
ORB%	1	0.215151	0.472210	0.46	0.6490
FT/FGA	1	10.805947	73.980192	0.15	0.8840
OppeFG%	1	-371.374727	15.211693	-24.41	<.0001
OppTOV%	1	3.030607	0.206133	14.70	<.0001
DRB%	1	0.839850	0.138628	6.06	<.0001
OppFT/FGA	1	-85.190106	10.758710	-7.92	<.0001
Season 12-13	1	-0.937199	1.660444	-0.56	0.5729
Season 13-14	1	-0.109650	1.469984	-0.07	0.9406
Season 14-15	1	-0.240973	1.514062	-0.16	0.8737
Season 15-16	1	0.288284	1.330701	0.22	0.8286
Season 16-17	1	0.005855	1.117565	0.01	0.9958
Season 17-18	1	0.213964	1.061727	0.20	0.8404
Season 18-19	1	0.492399	0.970050	0.51	0.6121
Season 19-20	1	-4.577570	0.953579	-4.80	<.0001

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Season 20-21	1	-4.833047	0.903102	-5.35	<.0001
Season 21-22	1	-0.252743	0.913035	-0.28	0.7821
Season 22-23	0	0	.	.	.
Conference Eastern	1	-0.122055	0.509360	-0.24	0.8108
Conference Western	0	0	.	.	.

From the parameter estimate table above, we can determine the significance of predictors by looking at the p-values of each parameter (predictor). A p-value less than the significance level (0.05) indicates that the predictor is significant, while a p-value greater than the significance level indicates that the predictor is not significant. Based on the specified criteria, significant parameters have been color-coded "Green" and parameters that do not significantly contribute to the model are color-coded "Red".

Note: From the above table, we can also see that the predictors "Season 22-23" and "Conference Western" have zero estimates (Highlighted in Yellow). This means that these two predictors have no effect on the Wins and therefore, are not included in the model. The reason for this could be because they were either not significant predictors or they were highly correlated with other predictor variables in the model.

A summary of the Significant and Not Significant Predictors with their corresponding p-values rounded off to the nearest 2 decimals is given below:

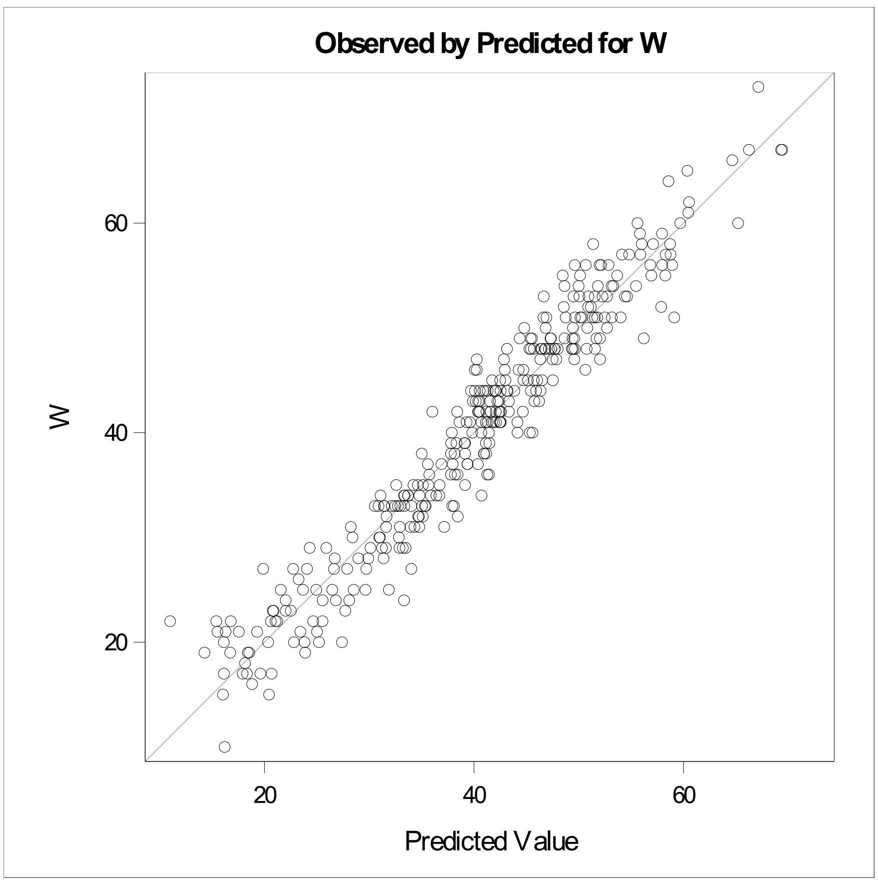
SIGNIFICANT PREDICTORS		NOT SIGNIFICANT PREDICTORS	
Parameter Name	**p-value**	**Parameter Name**	**p-value**
Intercept	0.01	SOS	0.89
Age	0.00	Pace	0.42
ORtg	0.04	FTr	0.90
OppeFG%	<.0001	3PAr	0.48
OppTOV%	<.0001	eFG%	0.49
DRB%	<.0001	TOV%	0.60
OppFT/FGA	<.0001	ORB%	0.65
Season 19-20	<.0001	FT/FGA	0.88
Season 20-21	<.0001	Season 12-13	0.57
		Season 13-14	0.94
		Season 14-15	0.87

		Season 15-16	0.83
		Season 16-17	1.00
		Season 17-18	0.84
		Season 18-19	0.61
		Season 21-22	0.78

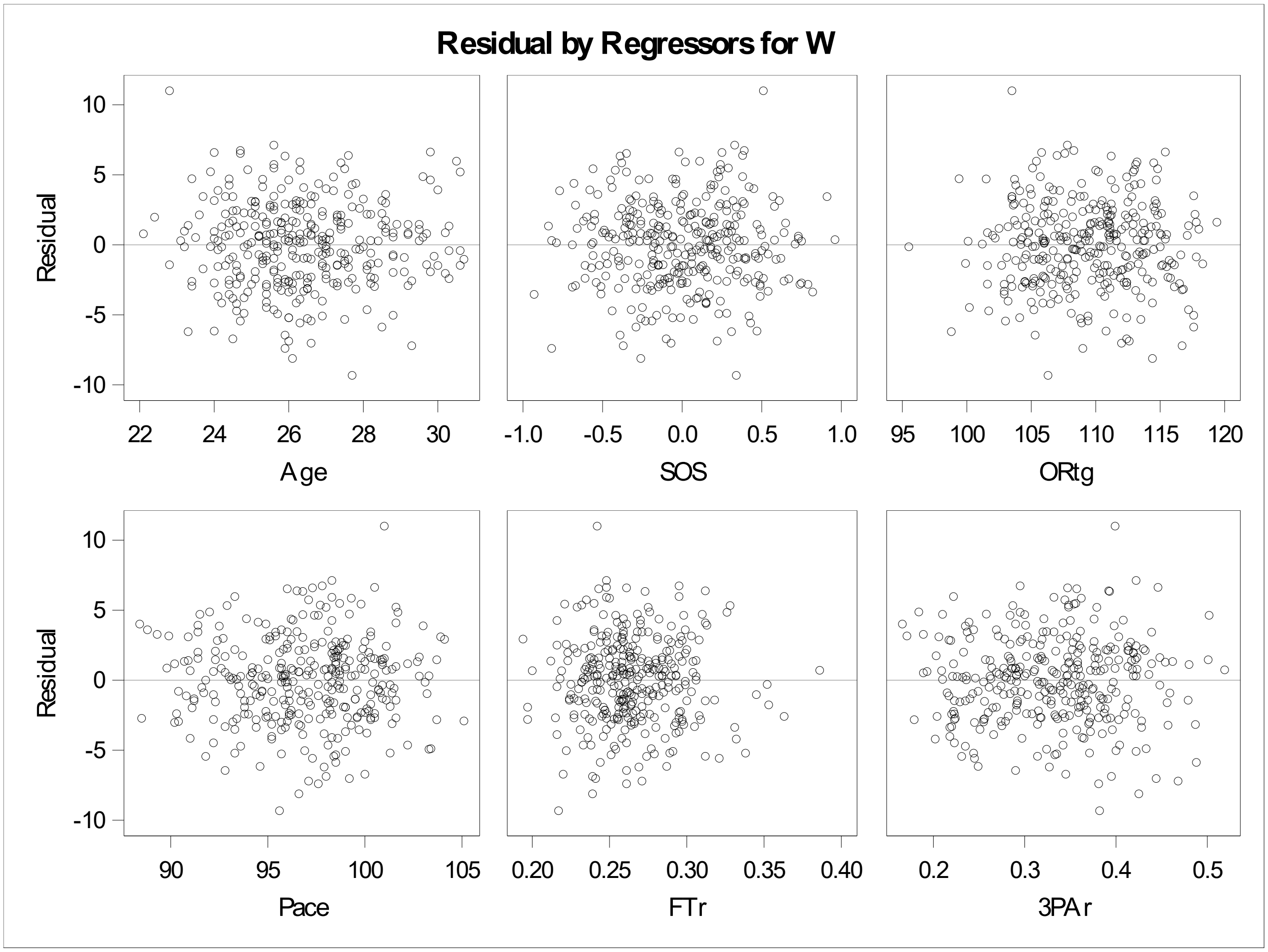
d. Examine the residuals. Do regression assumptions hold?

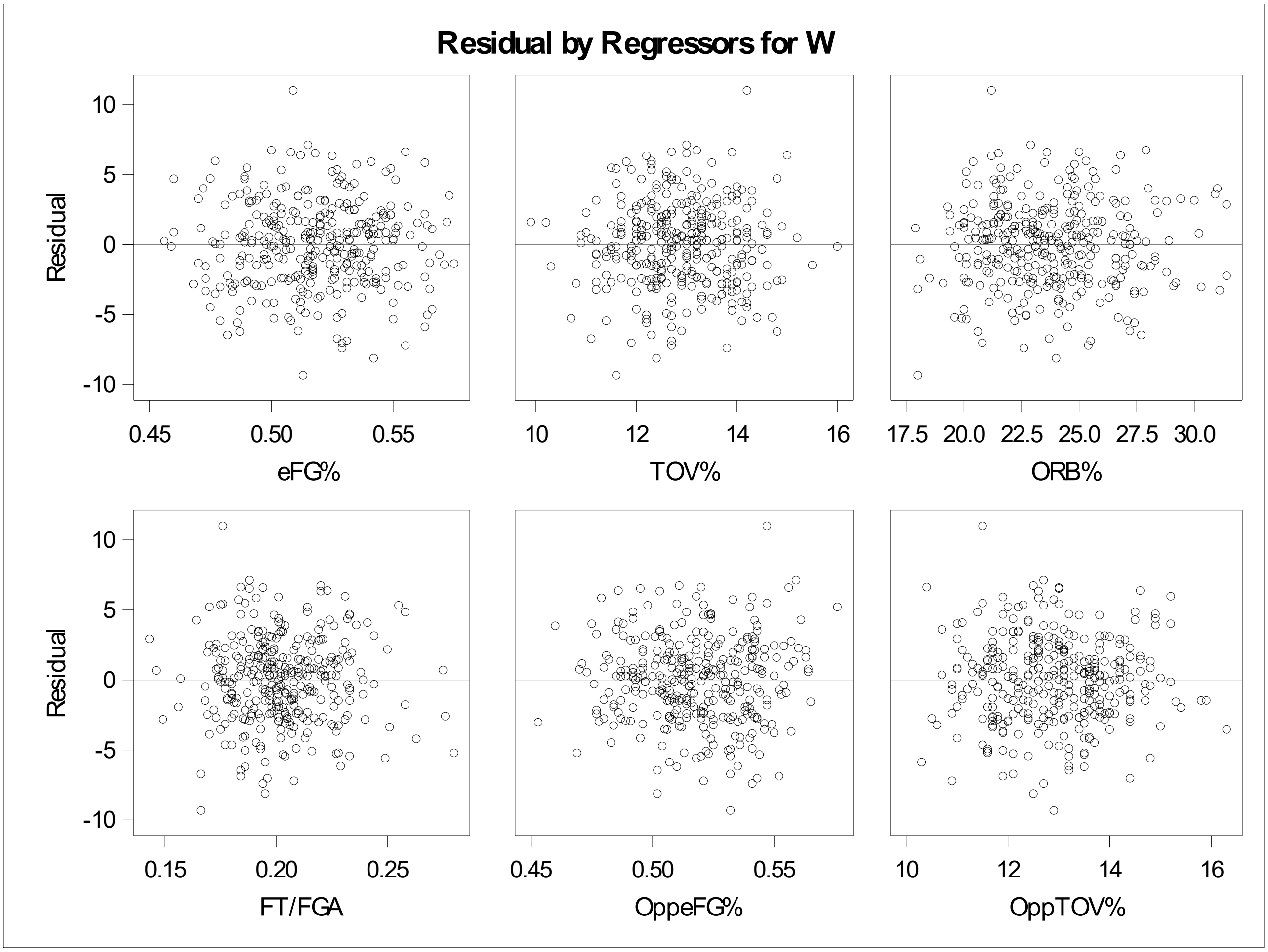
A Linear regression model has the following assumptions:
Linearity: The mean of the Ys is accurately modeled by a linear function of the Xs.
Normality: The random error term is assumed to have a normal distribution with a mean of zero.
Homoscedasticity: The random error term is assumed to have a constant variance.
Independence: The errors are independent.

<u>Checking Linearity</u>



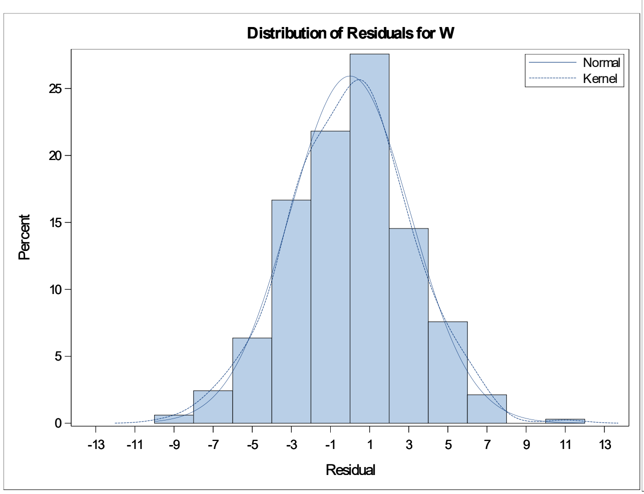
The observed vs. fitted plot suggests that there is a linear relationship between the predictor variables and the response variable. The points appear to be randomly scattered around the fitted line, indicating that the model is adequately capturing the relationship between the predictors and the response. Therefore, we can assume that the linearity assumption is met.

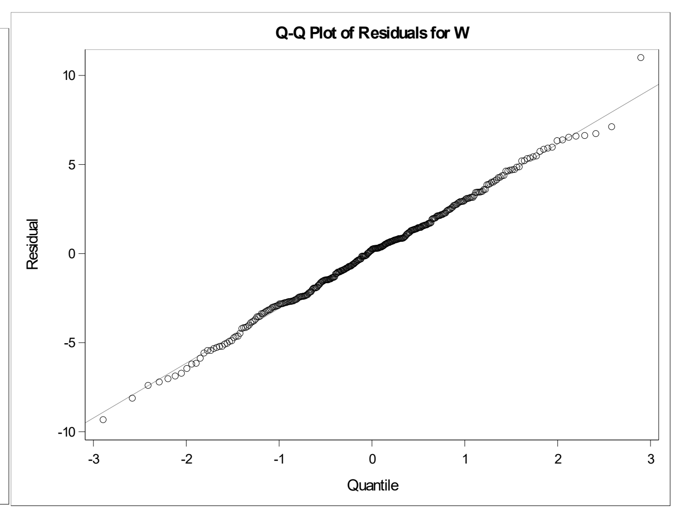




From the above plot, we can see that each plot displays a random scatter of points around the zero line, indicating that the residuals are distributed evenly and not following any discernible pattern. This suggests that the relationship between the predictor variable and the response variable is linear and can be accurately modeled using linear regression.

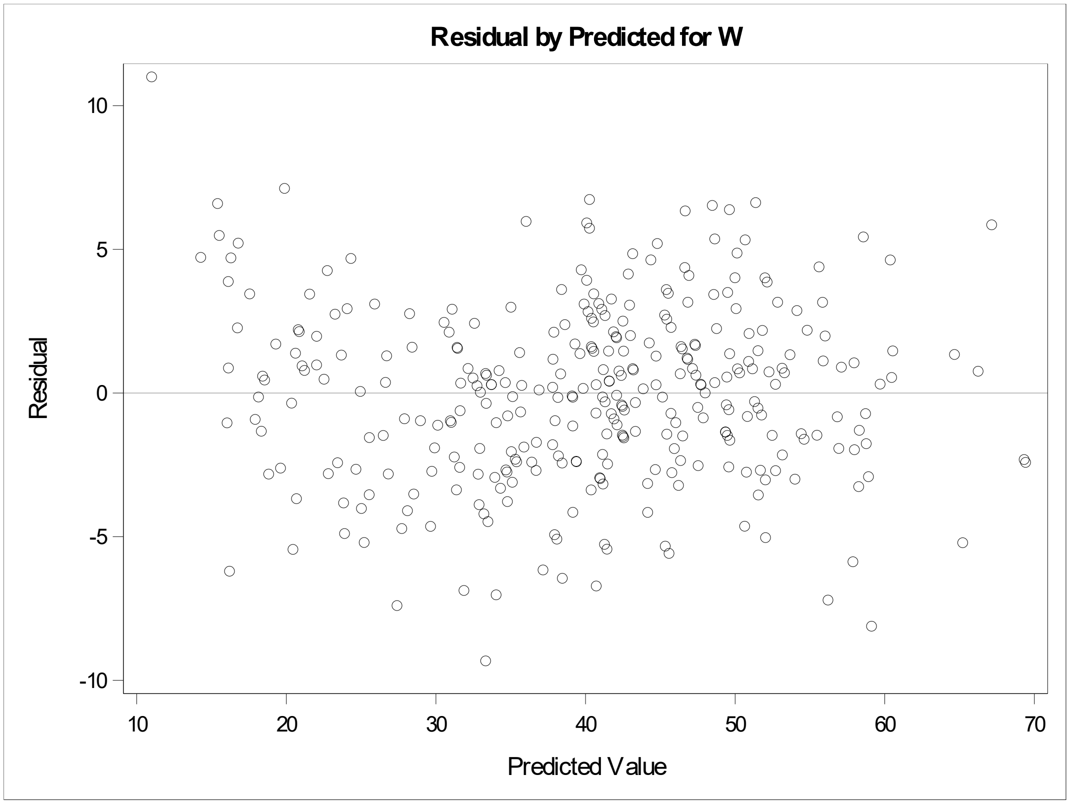
Checking Normality





Based on the histogram, the residuals appear to follow a bell curve and are centered around zero, suggesting that the normality assumption holds true. From the Q-Q Plot, we can see that the points on the plot fall close to the straight diagonal line, indicating that the residuals follow a normal distribution. Therefore, we can conclude that the normality assumption is met, and the model is valid for further analysis.

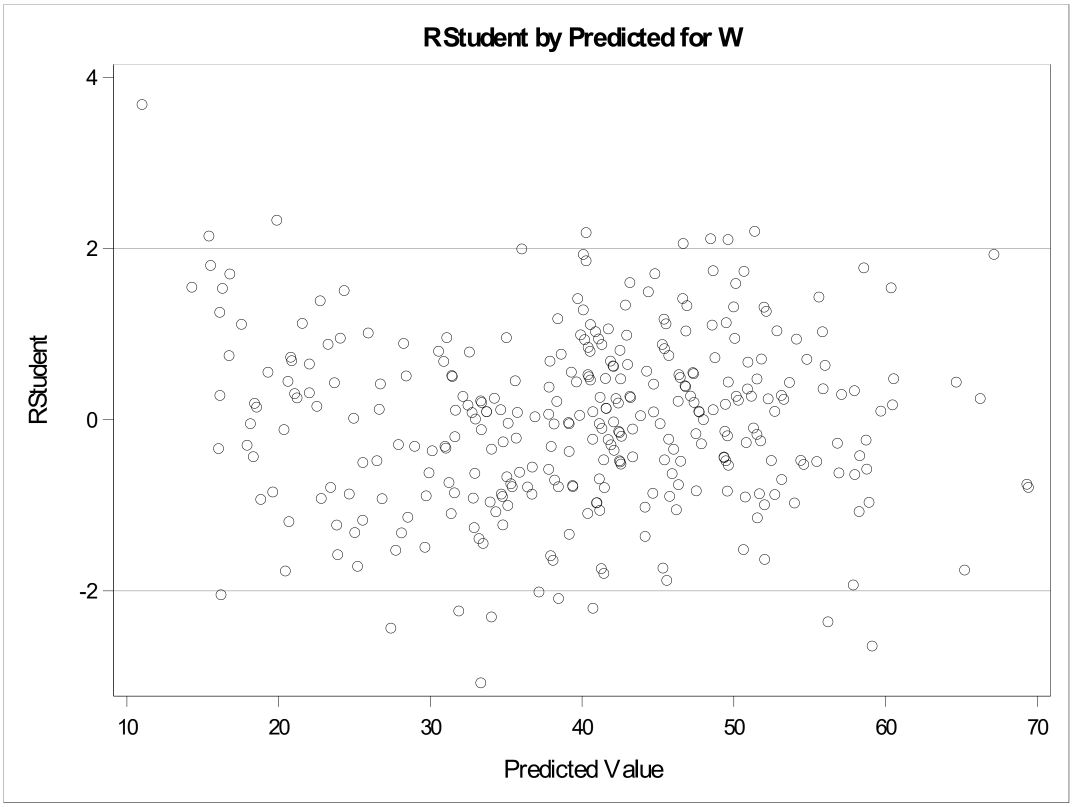
Checking Homoscedasticity



After examining the "Residuals by Predicted" plot, we can conclude that the homoscedasticity assumption holds true for our regression model. The plot shows a random scatter of the residuals around zero for all values of predicted Y, without any clear pattern of funnel shape or changing variance. This suggests that the variance of the errors is constant across all levels of the predicted values and that the assumption of homoscedasticity is not violated.

Checking Independence

The "RStudent by Predicted" plot is used to check the assumption of independence in the linear regression model. After examining the plot (given below), we can conclude that the assumption of independence holds true. The plot shows no clear pattern, with the residuals appearing to be randomly scattered around the horizontal line at zero. Therefore, we can assume that the errors in our model are independent of each other.



e. Is there any evidence of multicollinearity?

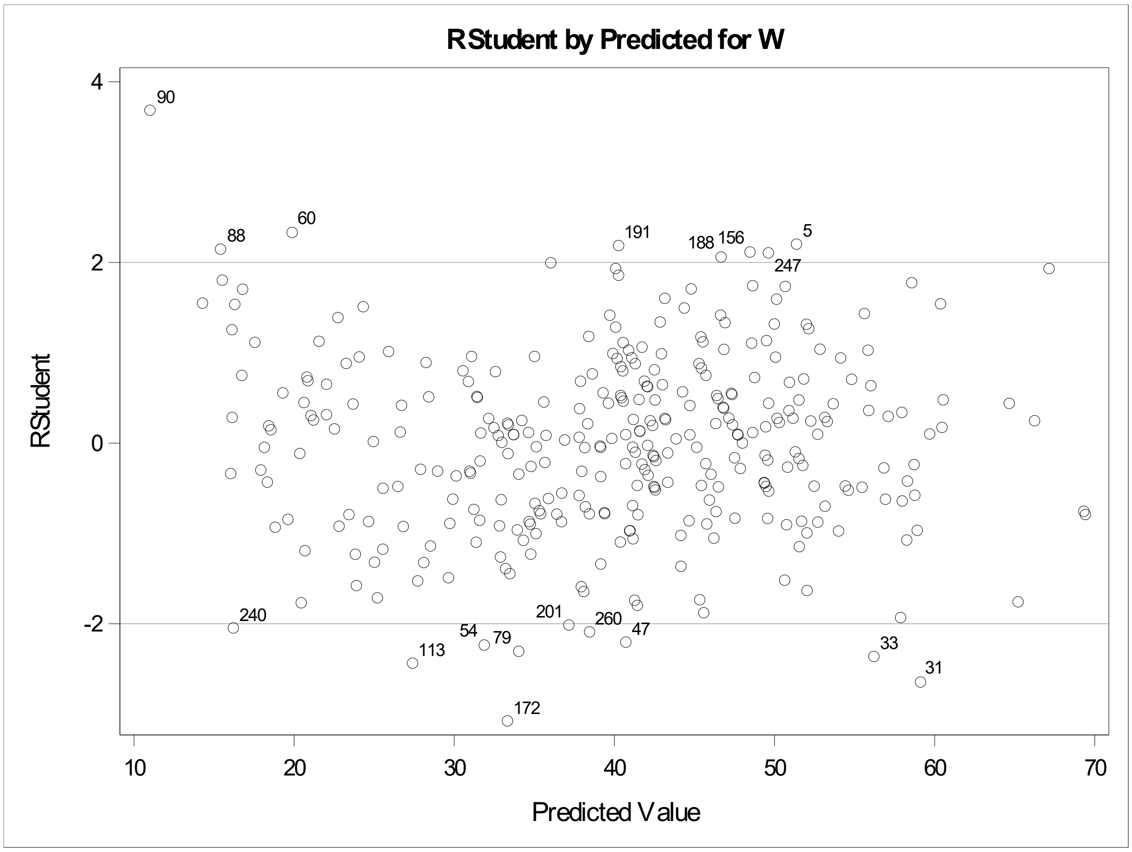
Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated with each other, making it difficult to determine the individual effect of each variable on the response variable. This can lead to unstable and unreliable estimates of the regression coefficients, which can affect the accuracy and interpretability of the model. Variance Inflation Factor (VIF) has been used to diagnose the collinearity problem. From the below table, we can see that some predictors have VIFs that are much larger than 10. These VIFs are color-coded in “Red”. Predictors with VIF > 10 in descending order are as follows– ORtg (432.19), eFG% (322.17), FT/FGA (80.08), ORB% (49.81), TOV% (46.14), FTr (43.68).

Hence, we can conclude that there is evidence of a multi-collinearity problem. Generally, in the absence of a subject-matter expert, we systematically remove variables starting with the highest VIF and re-run the analysis. Much like p-values, the VIF values will need to be updated with each successive variable removal.

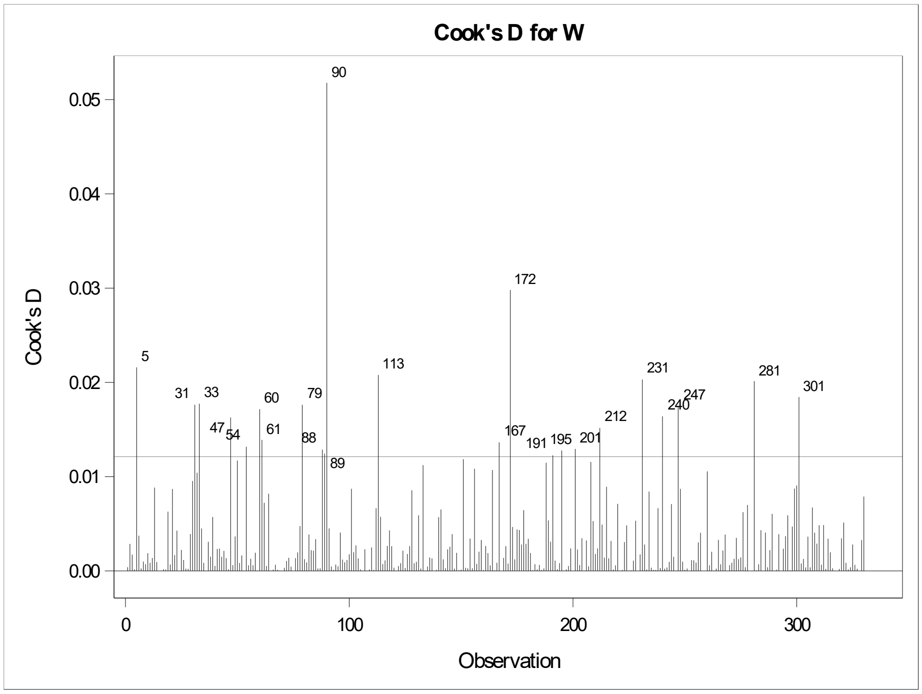
Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	B	-91.96935	34.73723	-2.65	0.0085	0
Age	Age	1	0.41472	0.13401	3.09	0.0022	1.78023
SOS	SOS	1	0.09951	0.74287	0.13	0.8935	2.33971
ORtg	ORtg	B	1.77054	0.85076	2.08	0.0383	**432.18949**
Pace	Pace	1	-0.08105	0.10040	-0.81	0.4201	3.57928
FTr	FTr	1	-5.28688	40.94807	-0.13	0.8974	**43.67675**
3PAr	3PAr	1	-4.35600	6.16414	-0.71	0.4803	6.15530
eFG%	eFG%	B	90.25694	129.28140	0.70	0.4856	**322.16988**
TOV%	TOV%	1	-0.63844	1.21751	-0.52	0.6004	**46.14036**
ORB%	ORB%	1	0.21515	0.47221	0.46	0.6490	**49.80780**
FT/FGA	FT/FGA	B	10.80595	73.98019	0.15	0.8840	**80.07778**
OppeFG%	OppeFG%	1	-371.37473	15.21169	-24.41	<.0001	3.63073
OppTOV%	OppTOV%	1	3.03061	0.20613	14.70	<.0001	1.67235
DRB%	DRB%	1	0.83985	0.13863	6.06	<.0001	2.59362
OppFT/FGA	OppFT/FGA	1	-85.19011	10.75871	-7.92	<.0001	1.50386
Season 12-13	Season 12-13	B	-0.93720	1.66044	-0.56	0.5729	7.34137
Season 13-14	Season 13-14	B	-0.10965	1.46998	-0.07	0.9406	5.75379
Season 14-15	Season 14-15	B	-0.24097	1.51406	-0.16	0.8737	6.10402
Season 15-16	Season 15-16	B	0.28828	1.33070	0.22	0.8286	4.71509
Season 16-17	Season 16-17	B	0.00585	1.11756	0.01	0.9958	3.32563
Season 17-18	Season 17-18	B	0.21396	1.06173	0.20	0.8404	3.00161
Season 18-19	Season 18-19	B	0.49240	0.97005	0.51	0.6121	2.50563
Season 19-20	Season 19-20	B	-4.57757	0.95358	-4.80	<.0001	2.42127
Season 20-21	Season 20-21	B	-4.83305	0.90310	-5.35	<.0001	2.17171
Season 21-22	Season 21-22	B	-0.25274	0.91303	-0.28	0.7821	2.21975
Season 22-23	Season 22-23	0	0	.	.	.	.
Conference Eastern	Conference Eastern	B	-0.12206	0.50936	-0.24	0.8108	2.08980
Conference Western	Conference Western	0	0	.	.	.	.

f. Are there any influential observations present?

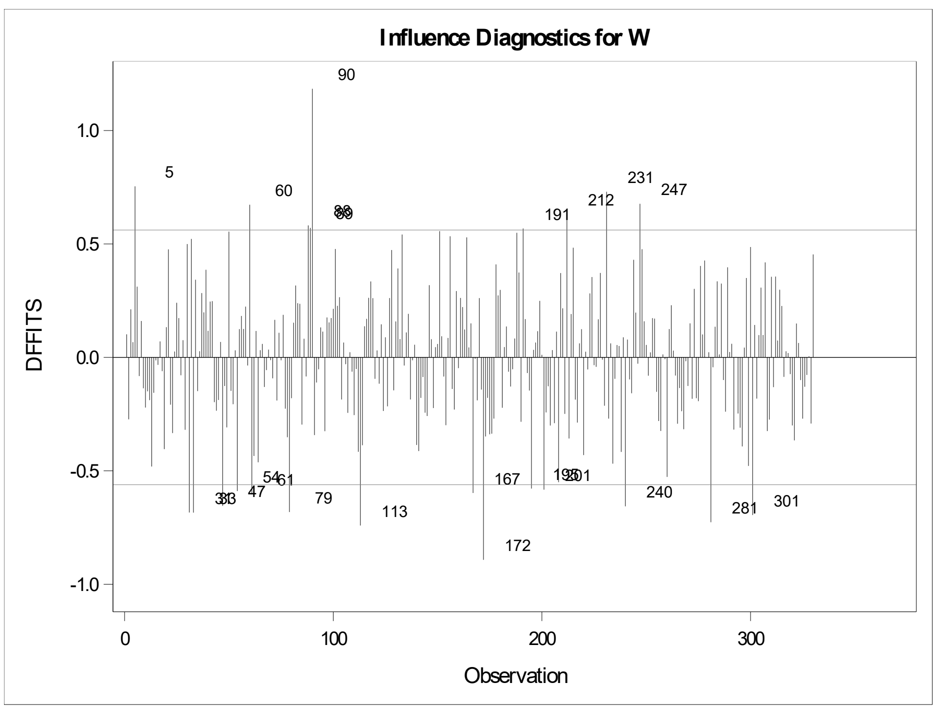
Presence of Influential observations is assessed using RStudentByPredicted, Cooks distance, DFFITS, DFBETA Plots. An influential observation is one that has a significant impact on the regression line, meaning that if it were removed from the dataset, the regression line would change substantially. This can happen when an observation has an extreme value for one or more of the predictors, or when an observation is an outlier compared to the rest of the data.



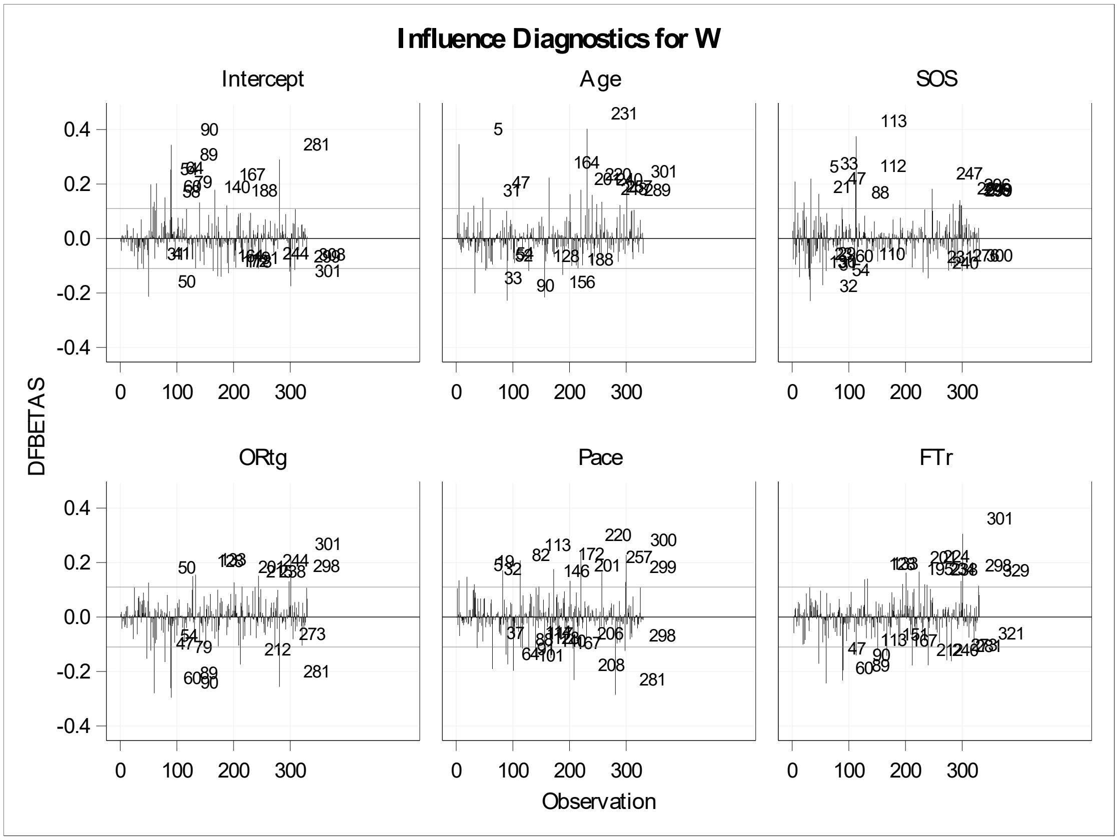
The above RStudent plot shows 18 observations (5.45% of the total number of observations) beyond two standard errors from the mean of 0. These are identified with their observation numbers. These can be considered as potential influential observations.

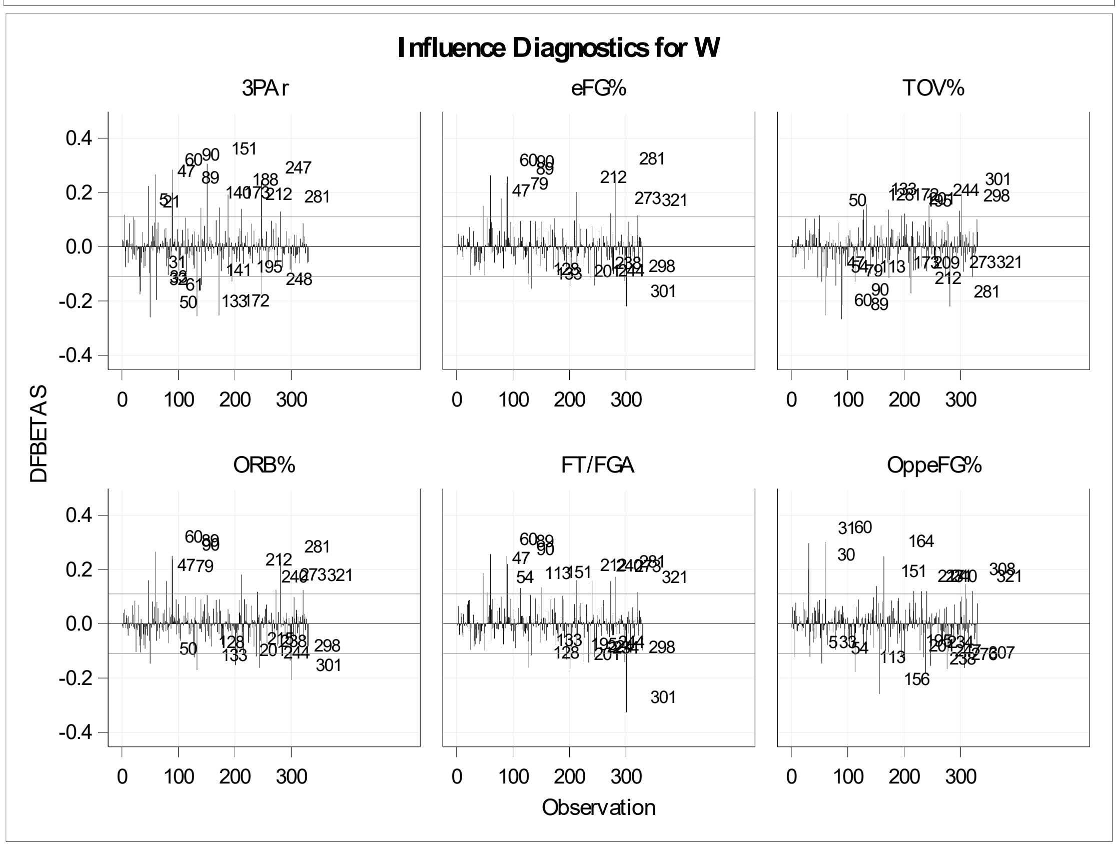


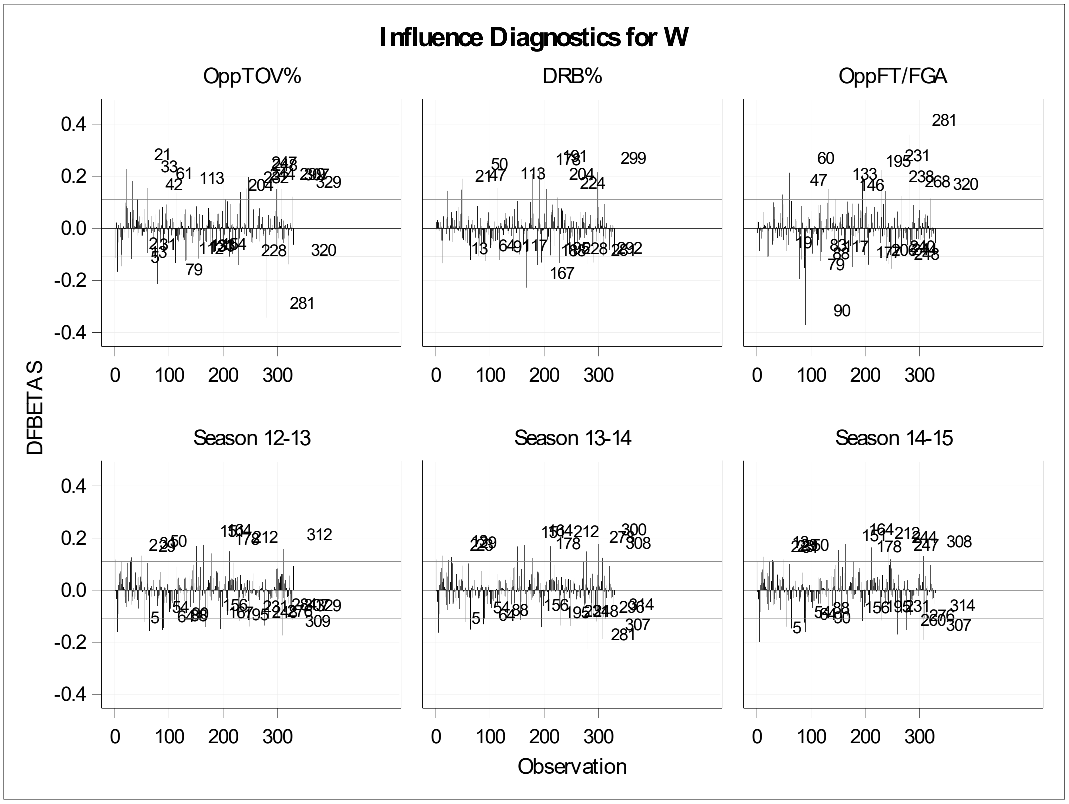
We can observe from the above Cook's D plot that the potential influential observations are those that are greater than the cut-off value (represented by the horizontal line above the x-axis). There are 23 observations (approximately 7% of the total number of observations) here as well that are falling outside the cut-off line. These can be considered potential influential observations. We can observe the same observations in the below DFFITS Plot. These observations go over and beyond the cut-off line indicating that they are potentially influential observations.

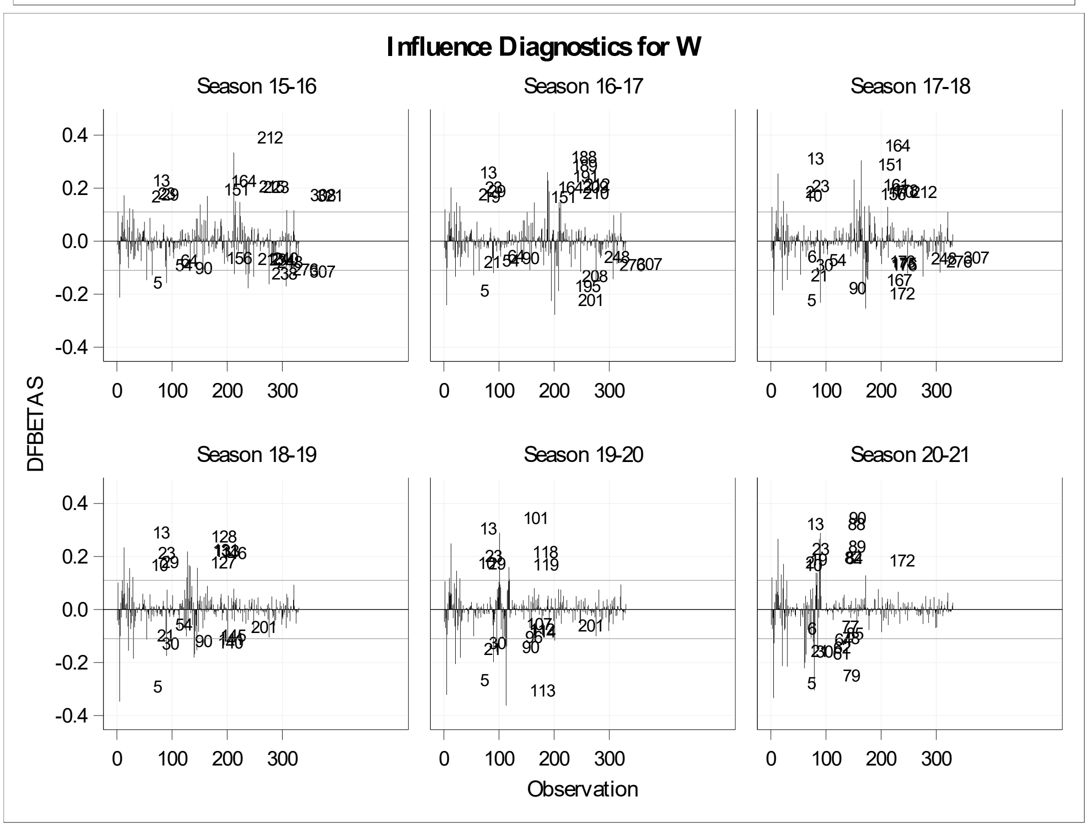


We can notice potentially influential observations from the below DFBETA plots as well. The number of these outliers varies depending on the independent variable.


Influence Diagnostics for W
Intercept
Age
SOS
ORtg
Pace
FTr
DFBETAS
Observation


Influence Diagnostics for W
3PAr
eFG%
TOV%
ORB%
FT/FGA
OppeFG%
DFBETAS
Observation


Influence Diagnostics for W
OppTOV%
DRB%
OppFT/FGA
Season 12-13
Season 13-14
Season 14-15
DFBETAS
Observation


Influence Diagnostics for W
Season 15-16
Season 16-17
Season 17-18
Season 18-19
Season 19-20
Season 20-21
DFBETAS
Observation

It is important to address these influential observations in the data, as they can have a significant impact on the results and interpretation of the regression analysis. One way to address influential observations is to remove them from the dataset and rerun the analysis without them. However, it is important to exercise caution when removing influential observations, as they may contain important information or reflect real-world conditions.

TASK #3

a. Is the model significant?

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	44581	2786.29480	275.37	<.0001
Error	313	3167.08925	10.11850		
Corrected Total	329	47748			

Based on the ANOVA table above, we can say that the model is statistically significant with an F-value of 275.37 and a corresponding p-value of less than 0.0001, indicating strong evidence against the null hypothesis that the model is not significant.

b. Discuss the model's goodness of fit.

Root MSE	3.18096
Dependent Mean	40.02424
R-Square	0.9337
Adj R-Sq	0.9303
AIC	1112.28694
AICC	1114.48630
BIC	784.56333
C(p)	13.21506
SBC	844.87151

From the above table, we can see that the R-squared value of the model is 0.9337, which means that approximately 93.37% of the variation in the dependent variable can be explained by the independent variables included in the model. The Adjusted R-squared value is 0.9303, which is slightly lower than the

R-squared value, suggesting that some of the independent variables in the model may not be contributing significantly to the model's predictive power.

The Root Mean Square Error (RMSE) of the model is 3.18096, which indicates that the average prediction error of the model is relatively small.

Overall, the model appears to have a good fit for the data.

c. Discuss the significance of predictors.

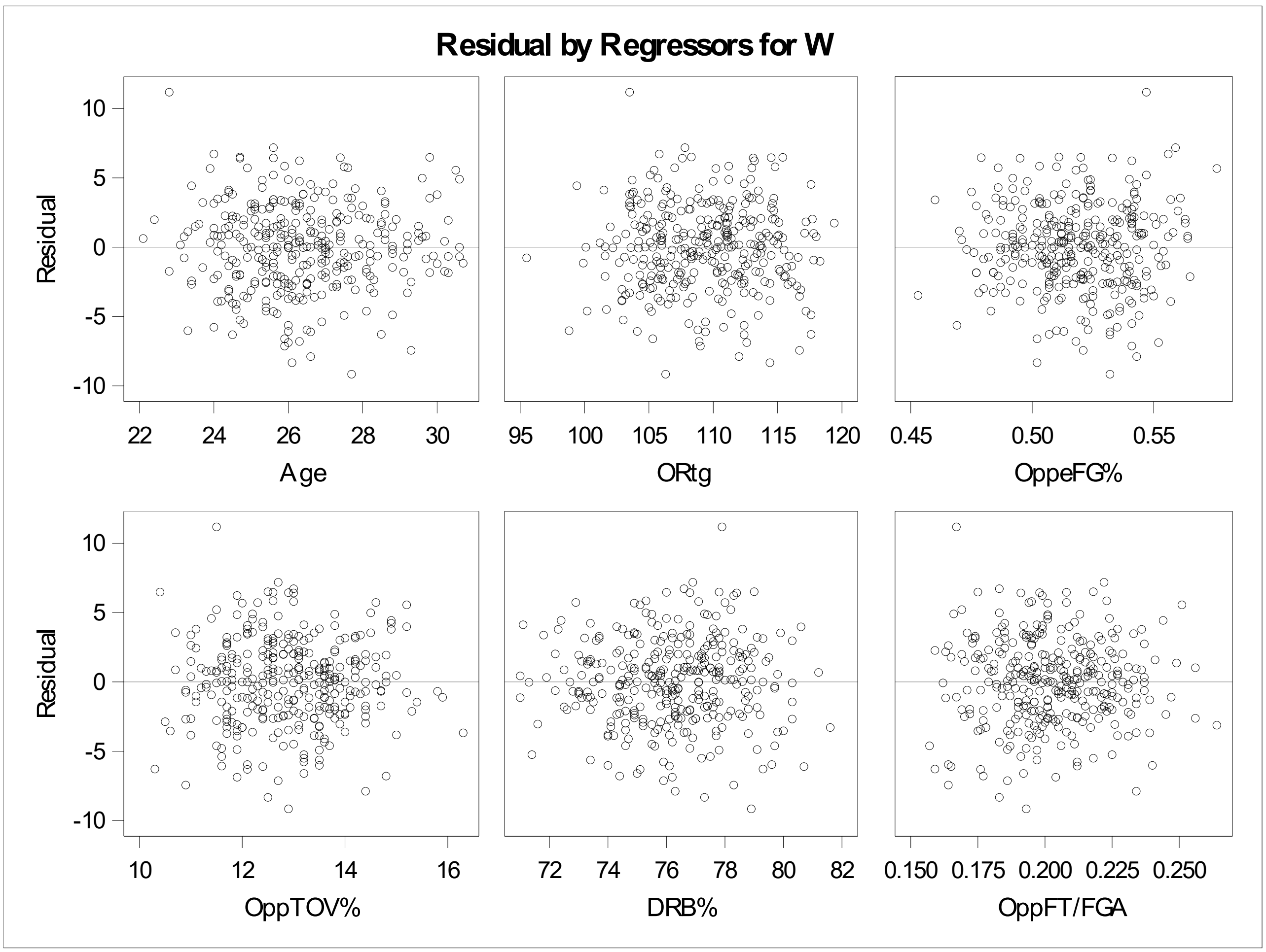
Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-112.822186	16.661679	-6.77	<.0001
Age	1	0.457767	0.119164	3.84	0.0001
ORtg	1	2.247151	0.062995	35.67	<.0001
OppeFG%	1	-373.884441	13.553302	-27.59	<.0001
OppTOV%	1	3.102006	0.193128	16.06	<.0001
DRB%	1	0.884609	0.125661	7.04	<.0001
OppFT/FGA	1	-87.333260	10.442581	-8.36	<.0001
Season 12-13	1	-0.276540	1.355808	-0.20	**0.8385**
Season 13-14	1	0.269802	1.241260	0.22	**0.8281**
Season 14-15	1	0.111704	1.309271	0.09	**0.9321**
Season 15-16	1	0.463687	1.197605	0.39	**0.6989**
Season 16-17	1	0.079864	1.032217	0.08	**0.9384**
Season 17-18	1	0.379793	1.019612	0.37	**0.7098**
Season 18-19	1	0.360526	0.932596	0.39	**0.6993**
Season 19-20	1	-4.867460	0.918298	-5.30	<.0001
Season 20-21	1	-4.856984	0.876472	-5.54	<.0001
Season 21-22	1	-0.352924	0.880092	-0.40	**0.6887**
Season 22-23	0	0	.	.	.

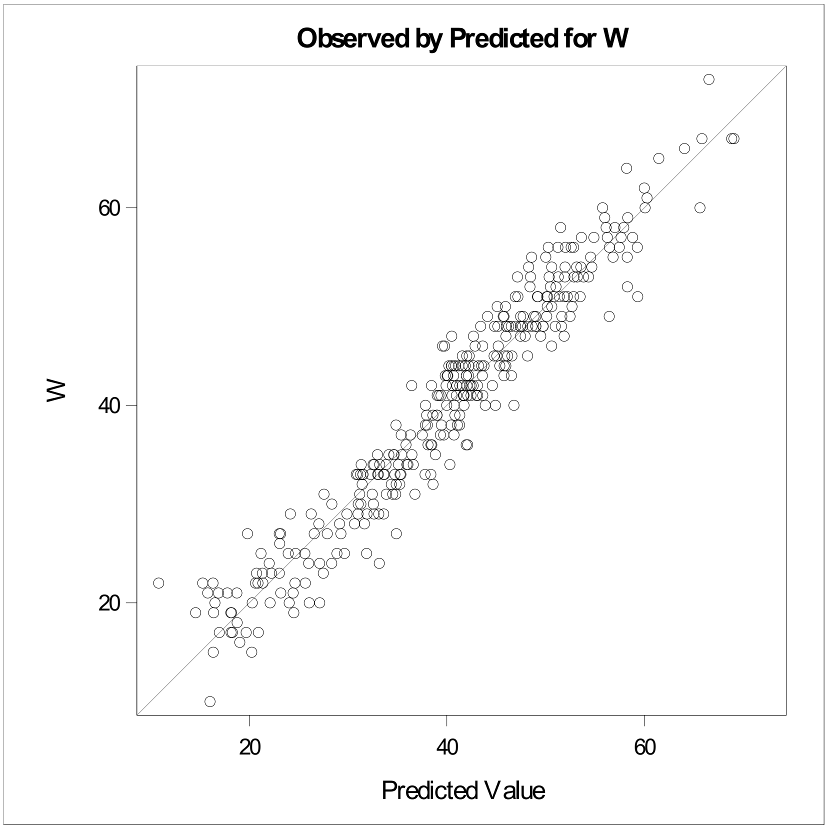
We can assess the significance of the predictors from the above table. The predictors with a p-value of less than 0.05 are statistically significant, which means that there is strong evidence to suggest that the variable is contributing to the model's ability to explain the variation in the dependent variable. If a predictor has a p-value of greater than 0.05, it means that there is not enough evidence to reject the null hypothesis that the coefficient for that variable is equal to zero. This suggests that the variable may not be important in predicting the Wins (W). Variables that are not significant to the model turned out to be all seasons except 19-20 & 20-21. These are color-coded in "Red" in the above table. Usually, predictors that are not significant are removed from the model. However, the step-wise selection has not removed any of those predictors here. The reason is that even if most of the categories have p-values greater than 0.05, the categorical variable (Season) as a whole may still be significant. Moreover, removing a categorical variable completely can result in the loss of important information and can lead to a less accurate model.

d. Examine regression assumptions.

Checking Linearity

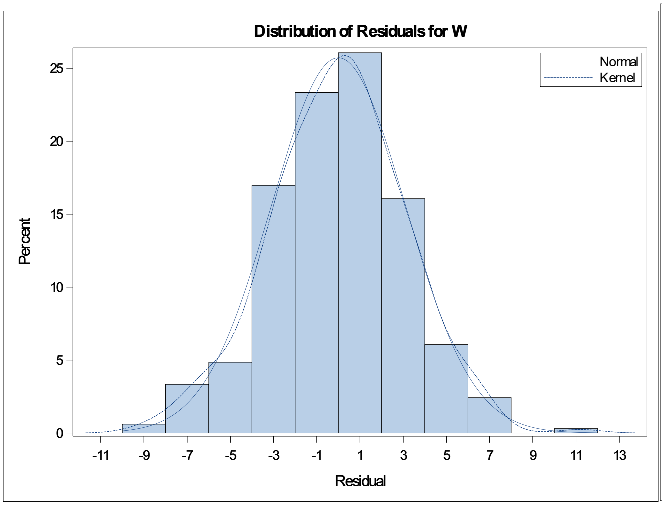


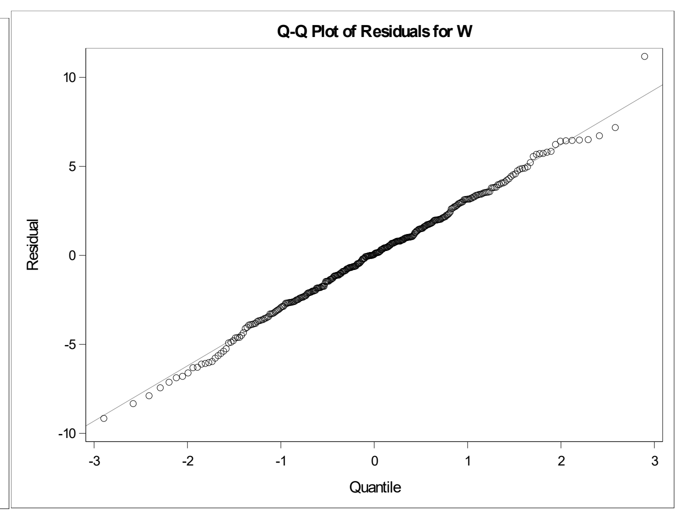
The above plot shows that there is a random distribution of points around the zero line, indicating that there is no specific pattern observed in the residuals. This suggests that the relationship between the predictor and response variables is linear, and it can be well modeled by using a linear regression model.



According to the observed vs. fitted plot, there seems to be a linear correlation between the predictor variables and the response variable. The points are dispersed randomly around the fitted line, indicating that the model has appropriately captured the association between the predictors and the response. Hence, we can infer that the assumption of linearity is fulfilled.

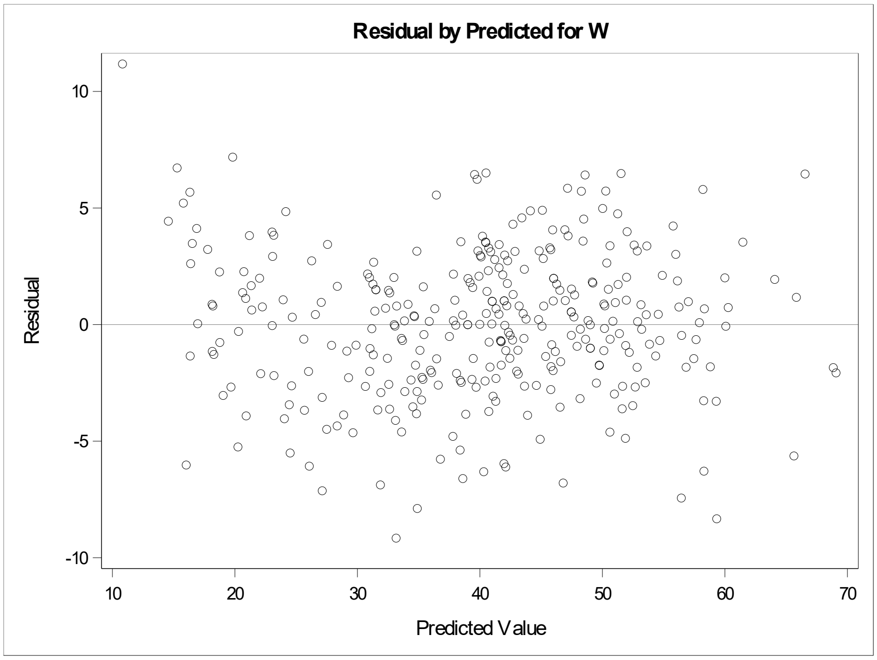
Checking Normality





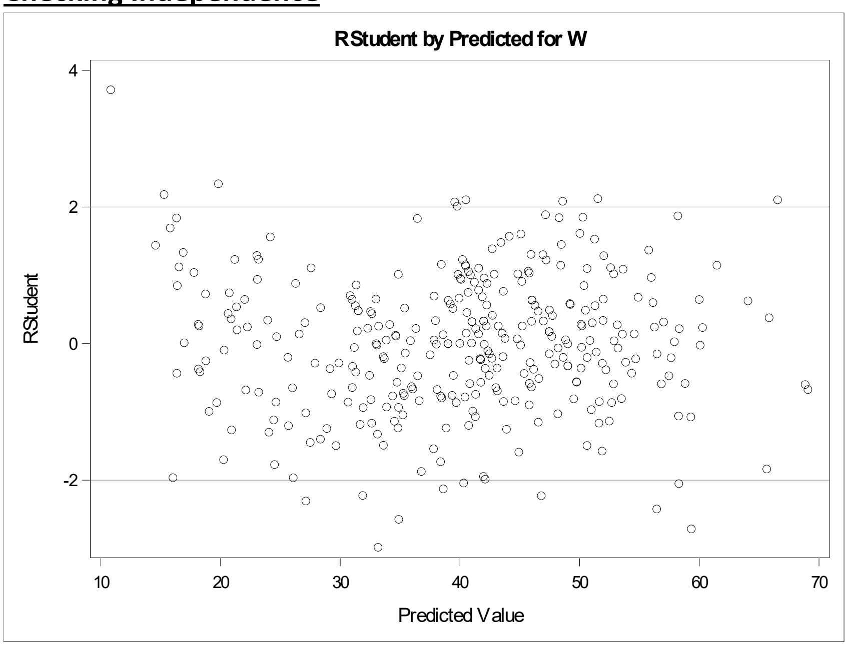
The histogram of residuals displays a bell-shaped curve, and the center is around zero, which indicates that the normality assumption is valid. In addition, the Q-Q Plot displays the residuals' points fall close to the straight diagonal line, indicating that the residuals are normally distributed. As a result, we can conclude that the normality assumption is fulfilled.

Checking Homoscedasticity



From the "Residuals by Predicted" plot, we can see that the residuals are scattered randomly around the horizontal line at zero, indicating that there is no pattern or trend in the distribution of residuals. Moreover, there is no visible cone-shaped or fan-shaped pattern in the plot, which indicates that the variance of the residuals is constant across the range of predicted values. Therefore, we can conclude that the homoscedasticity assumption holds true for our regression model.

Checking Independence



The "RStudent by Predicted" plot depicts a random scattering of residuals around the zero horizontal line, without any clear pattern. Hence, it can be deduced that the errors in the model are not associated with each other, and the independence assumption can be assumed to hold true.

e. Is multicollinearity a problem?

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	B	-112.82219	16.66168	-6.77	<.0001	0
Age	Age	1	0.45777	0.11916	3.84	0.0001	1.42487
ORtg	ORtg	1	2.24715	0.06300	35.67	<.0001	2.39864
OppeFG%	OppeFG%	1	-373.88444	13.55330	-27.59	<.0001	2.91751
OppTOV%	OppTOV%	1	3.10201	0.19313	16.06	<.0001	1.48595
DRB%	DRB%	1	0.88461	0.12566	7.04	<.0001	2.15719
OppFT/FGA	OppFT/FGA	1	-87.33326	10.44258	-8.36	<.0001	1.43413
Season 12-13	Season 12-13	B	-0.27654	1.35581	-0.20	0.8385	4.95460
Season 13-14	Season 13-14	B	0.26980	1.24126	0.22	0.8281	4.15277
Season 14-15	Season 14-15	B	0.11170	1.30927	0.09	0.9321	4.62032
Season 15-16	Season 15-16	B	0.46369	1.19760	0.39	0.6989	3.86580
Season 16-17	Season 16-17	B	0.07986	1.03222	0.08	0.9384	2.87181
Season 17-18	Season 17-18	B	0.37979	1.01961	0.37	0.7098	2.80209
Season 18-19	Season 18-19	B	0.36053	0.93260	0.39	0.6993	2.34423
Season 19-20	Season 19-20	B	-4.86746	0.91830	-5.30	<.0001	2.27290
Season 20-21	Season 20-21	B	-4.85698	0.87647	-5.54	<.0001	2.07056
Season 21-22	Season 21-22	B	-0.35292	0.88009	-0.40	0.6887	2.08770
Season 22-23	Season 22-23	0	0	.	.	.	.

Variance Inflation Factors (VIFs) have been used to diagnose the multi-collinearity problem. We can see from the above table that there are no predictors with a VIF > 10. The stepwise selection model has automatically taken care of not including a predictor with a high VIF. So, we can conclude that there is no evidence of the multi-collinearity problem in the stepwise model.

TASK #4

Are models from parts 2 & 3 different? If so, discuss the differences.

Objective of the Analysis

The main difference between the models from Part 2 and Part 3 is the objective of the analysis. Part 2 aimed to create a multiple linear regression model that predicts the number of wins (W) in NBA games using ALL available predictor variables. Part 3, on the other hand, focused on identifying the best subset of predictor variables using stepwise selection with Mallow's C(p) as the selection criterion. The analysis involved the same tasks as Part 2 but with the primary objective of selecting the most relevant predictors. As a result, Part 3 included only one categorical variable "Season" whereas Part 2 involved both categories.

Selection Criteria

Another major difference between these two models is the selection criteria. In the model from Part 2, all available variables were included without any selection criteria. As a result, the model included some predictors that do not significantly contribute to explaining the variation in the wins (W). This means that these variables do not have a statistically significant relationship with the outcome variable, and their inclusion in the model does not improve the model's ability to predict the outcome. This can lead to issues such as overfitting, where the model becomes too complex and fits the training data too closely, leading to poor performance on new data. Including irrelevant predictors can also decrease the interpretability of the model, as it becomes more difficult to identify which variables are truly driving the relationship with the outcome. This model also posed a ***multi-collinearity problem*** by including some predictors with VIF > 10. This means that the model from part 2 cannot be used to analyze which factors contribute more to winning an NBA game since it did not address these major issues that adversely affect the model's ability to predict.

On the other hand, the model from part 3 addresses these issues by using a stepwise selection approach with Mallow's C(p) as the selection criteria to add or remove effects. The model first included the Intercept, even though it did not explain anything in the variation of the Wins (W), because it was statistically significant to the model with a p-value < 0.05. The model then added Age, which explained 30.25% of the variation in Wins (W) along with the intercept. Similarly, other variables were added one by one in such a way that they contributed to explaining a much higher proportion of the variation in Wins (W) along with other predictors already added to the model. This included Age, ORtg, OppeFG%, OppTOV%, DRB%, OppFT/FGA, and the categorical variable Season. By doing a stepwise selection, the model automatically resolved the problem of multi-collinearity.

Overall, the model from part 3 is a better choice for analyzing which factors contribute more to winning an NBA game since it addressed the major issues that the model from part 2 failed to address. It used a stepwise selection approach with Mallow's C(p) as the selection criteria to add or remove effects, which resulted in a more accurate and reliable model.

TASK #5

Interpret the results from part 3. Note that reporting the figures from the regression result set in SAS is not equivalent to interpreting the results. If in doubt, think of a scenario where you act as a consultant to the general manager of a basketball team who is trying to assess what factors the team should focus on for it to have a successful season.

The following table is used to interpret the results of the stepwise regression model.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-112.822186	16.661679	-6.77	<.0001
Age	1	0.457767	0.119164	3.84	0.0001
ORtg	1	2.247151	0.062995	35.67	<.0001
OppeFG%	1	-373.884441	13.553302	-27.59	<.0001
OppTOV%	1	3.102006	0.193128	16.06	<.0001
DRB%	1	0.884609	0.125661	7.04	<.0001
OppFT/FGA	1	-87.333260	10.442581	-8.36	<.0001
Season 12-13	1	-0.276540	1.355808	-0.20	0.8385
Season 13-14	1	0.269802	1.241260	0.22	0.8281
Season 14-15	1	0.111704	1.309271	0.09	0.9321
Season 15-16	1	0.463687	1.197605	0.39	0.6989
Season 16-17	1	0.079864	1.032217	0.08	0.9384
Season 17-18	1	0.379793	1.019612	0.37	0.7098
Season 18-19	1	0.360526	0.932596	0.39	0.6993
Season 19-20	1	-4.867460	0.918298	-5.30	<.0001
Season 20-21	1	-4.856984	0.876472	-5.54	<.0001
Season 21-22	1	-0.352924	0.880092	-0.40	0.6887
Season 22-23	0	0	.	.	.

From the above table, we can see that the important factors that contribute to winning an NBA game are: Age, ORtg, OppeFG%, OppTOV%, DRB% and OppFT/FGA. From their p-values, we can say that they are statistically significant. Though we can observe that Seasons 19-20, and 20-21 are also statistically significant, they are categorical variables only. In general, the year in which a game takes place may not impact the results of the game. The results for Season 19-20 and 20-21 might have been skewed due to the Covid-19 pandemic. So, they are not used for interpretation of the results.

Age

The coefficient estimate for Age is 0.4578, indicating that the increase in Wins for every one-year increase in the average age of the team is approximately 0.46, holding all other variables constant. The p-value suggests that the Age coefficient is statistically significant at the 95% confidence level, meaning that we can be fairly confident that the relationship between Age and Wins is not due to chance.

As a consultant to the general manager of a basketball team, I would recommend that the general manager prioritize recruiting or retaining players with a higher level of experience to increase the team's chances of having a successful season.

OFFENSIVE FACTORS

Offensive Rating (ORtg)

The ORtg variable has a strong positive effect on Wins, meaning that for every increase in ORtg by one point, we can expect a 2.247 increase in Wins, all other factors remaining constant. The ORtg variable is also highly statistically significant, indicating that this relationship is unlikely to be due to chance. Therefore, I would suggest the manager to focus on improving the offensive rating (ORtg) of his team in order to increase its chances of having a successful season.

Some potential ways to improve an NBA team's offensive rating are:

- Improving the team's shooting accuracy by practicing more shooting and improving shot selection.
- Increasing the pace of the game which can lead to more possessions and more scoring opportunities. This can be achieved through various means, such as encouraging players to push the ball up the court quickly after a rebound or turnover.
- Optimizing player rotations and lineups by carefully managing player minutes to ensure that key contributors are rested and performing at their best when on the court.

DEFENSIVE FACTORS

Opponent Effective Field Goal Percentage (OppeFG%)

The OppeFG% coefficient is -373.88. This result implies that an increase in the opponent's Effective Field Goal Percentage is associated with a decrease in Wins, holding all other variables constant. The OppeFG% variable is highly statistically significant as its p-value is much less than 0.05.

To improve the team's Wins, it is important to focus on reducing the Opponent's Effective Field Goal Percentage. This means that the team should aim to make it harder for the opposing team to make field goals by improving their defense. This could be achieved by :

- Practicing defensive drills and strategies during team practice sessions.
- Identifying the opponent's strengths and weaknesses and adjusting the defensive game plan accordingly.
- Recruiting or trading for players with strong defensive skills and incorporating them into the team's roster.

Opponent Turnover Percentage (OppTOV%)

The OppTOV% coefficient is 3.102. This implies that an increase in the team's opponent Turnover Percentage is associated with an increase in Wins, holding all other variables constant. The OppTOV% is highly statistically significant as its p-value is much less than 0.05.

To improve a team's opponent turnover percentage, the team can focus on :

- Strategies that force turnovers such as pressing, trapping, and aggressive man-to-man defense.
- Improving individual player skills such as footwork and other defensive instincts.
- Studying opponents' offensive strategies and weaknesses to better anticipate and disrupt their plays.

Ultimately, a strong defensive effort and smart defensive tactics are key to forcing turnovers and improving OppTOV%, which can lead to more wins.

Defensive Rebound Percentage (DRB%)

The DRB% coefficient is 0.885. This implies that an increase in the team's Defensive Rebound Percentage is associated with an increase in Wins, holding all other variables constant. The DRB% variable is highly statistically significant as its p-value is much less than 0.05.

It appears that improving the team's defensive rebound percentage could be a key factor in increasing their number of wins. Defensive rebounding is an important aspect of basketball because it allows a team to prevent their opponents from getting second-chance opportunities and can lead to more possessions for the team on offense.

Defensive Rebound Percentage can be improved by:

- Improving the players rebounding skills and positioning
- Adding players to the roster who excel in rebounding.
- Reviewing game footage and statistics to identify areas of weakness and develop strategies.

Opponent Free Throws Per Field Goal Attempt (OppFT/FGA)

The OppFT/FGA coefficient is -87.33. This implies that an increase in the team's opponent Free Throws per Field Goal Attempt (OppFT/FGA) is associated with a decrease in Wins, holding all other variables constant. The OppFT/FGA variable is highly statistically significant as its p-value is much less than 0.05

It seems that minimizing the opponent's Free Throws per Field Goal Attempt is a key factor to focus on for the team to have a successful season. This suggests that the team should focus on improving its defense, particularly in terms of avoiding fouls that could result in free throws for the opponent.

Some strategies that could help in minimizing OppFT/FGA include:

- Encouraging players to focus on good defensive positioning to reduce the likelihood of committing fouls.
- Ensuring that players maintain good physical conditioning and stamina, as fatigue can often lead to defensive breakdowns and fouls.

By focusing on these strategies, the team can improve its defensive performance, limit the number of free throws its opponents get, and ultimately increase its chances of winning more games.

---------- X ----------