The Relationship of Serving to Winning in Volleyball

Eric Geisler, Daniel Hrusovsky

John Carroll University

October 30, 2024

Introduction

We will be analyzing serving data from collegiate women's games (~3000) to determine the relationship between aces and errors on sets won. We hypothesize that service aces, service errors, service aces against, and service errors against will all be significant predictors of the number of sets won in a match.

Analysis and Discussion

miss

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	3084	100.0	100.0	100.0

Statistics

miss

N	Valid	3084
	Missing	0

Firstly, we establish that there is no missing data in the dataset used.

Now, we can begin by analyzing the descriptive statistics of the data.

Statistics

		SetsWon Sets Won	ServiceAcesAgainst Service Aces Against	ServiceErrorsAgainst Service Errors Against	TotalMissedServes Total Missed Serves	TotalServiceAces Total Service Aces
N	Valid	3084	3084	3084	3084	3084
	Missing	0	0	0	0	0
Mean		2.01	4.80	6.55	6.61	5.36
Skewness		-.681	.862	.715	.634	.946
Std. Error of Skewness		.044	.044	.044	.044	.044
Kurtosis		-1.262	1.281	.835	.533	1.654
Std. Error of Kurtosis		.088	.088	.088	.088	.088
Minimum		0	0	0	0	0
Maximum		4	23	23	21	26

The data for all of the selected variables is normal, as all skewness values are within the range of -3 to 3, and all kurtosis values are within the adjusted range for large sample sizes of -20 to 20.

Now, we can move on to the assumptions for regression. First, we must determine if the data has adequate variance. Specifically, we must see if there is one value or a close set of values that make up over 90% of the values in our data.

ServiceAcesAgainst Service Aces Against

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	78	2.5	2.5	2.5
	1	221	7.2	7.2	9.7
	2	377	12.2	12.2	21.9
	3	441	14.3	14.3	36.2
	4	479	15.5	15.5	51.8
	5	404	13.1	13.1	64.9
	6	313	10.1	10.1	75.0
	7	262	8.5	8.5	83.5
	8	205	6.6	6.6	90.1
	9	118	3.8	3.8	94.0
	10	58	1.9	1.9	95.8
	11	60	1.9	1.9	97.8
	12	30	1.0	1.0	98.8
	13	17	.6	.6	99.3
	14	9	.3	.3	99.6
	15	4	.1	.1	99.7
	16	2	.1	.1	99.8
	17	4	.1	.1	99.9
	19	1	.0	.0	100.0
	23	1	.0	.0	100.0
	Total	3084	100.0	100.0	

o

For *ServiceAcesAgainst*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

ServiceErrorsAgainst Service Errors Against

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	20	.6	.6	.6
	1	62	2.0	2.0	2.7
	2	152	4.9	4.9	7.6
	3	259	8.4	8.4	16.0
	4	342	11.1	11.1	27.1
	5	418	13.6	13.6	40.6
	6	413	13.4	13.4	54.0
	7	386	12.5	12.5	66.5
	8	301	9.8	9.8	76.3
	9	230	7.5	7.5	83.8
	10	144	4.7	4.7	88.4
	11	133	4.3	4.3	92.7
	12	68	2.2	2.2	94.9
	13	72	2.3	2.3	97.3
	14	40	1.3	1.3	98.6
	15	16	.5	.5	99.1
	16	10	.3	.3	99.4
	17	8	.3	.3	99.7
	18	3	.1	.1	99.8
	19	3	.1	.1	99.9
	20	1	.0	.0	99.9
	21	2	.1	.1	100.0
	23	1	.0	.0	100.0
	Total	3084	100.0	100.0	

For *ServiceErrorsAgainst*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

TotalMissedServes Total Missed Serves

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	19	.6	.6	.6
	1	64	2.1	2.1	2.7
	2	147	4.8	4.8	7.5
	3	254	8.2	8.2	15.7
	4	342	11.1	11.1	26.8
	5	409	13.3	13.3	40.0
	6	391	12.7	12.7	52.7
	7	379	12.3	12.3	65.0
	8	307	10.0	10.0	75.0
	9	245	7.9	7.9	82.9
	10	170	5.5	5.5	88.4
	11	128	4.2	4.2	92.6
	12	76	2.5	2.5	95.0
	13	69	2.2	2.2	97.3
	14	37	1.2	1.2	98.5
	15	17	.6	.6	99.0
	16	14	.5	.5	99.5
	17	9	.3	.3	99.8
	18	2	.1	.1	99.8
	19	2	.1	.1	99.9
	20	1	.0	.0	99.9
	21	2	.1	.1	100.0
	Total	3084	100.0	100.0	

For *TotalMissedServes*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

TotalServiceAces Total Service Aces

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	56	1.8	1.8	1.8
	1	179	5.8	5.8	7.6
	2	318	10.3	10.3	17.9
	3	385	12.5	12.5	30.4
	4	443	14.4	14.4	44.8
	5	397	12.9	12.9	57.7
	6	344	11.2	11.2	68.8
	7	283	9.2	9.2	78.0
	8	213	6.9	6.9	84.9
	9	160	5.2	5.2	90.1
	10	102	3.3	3.3	93.4
	11	75	2.4	2.4	95.8
	12	44	1.4	1.4	97.2
	13	33	1.1	1.1	98.3
	14	21	.7	.7	99.0
	15	15	.5	.5	99.5
	16	3	.1	.1	99.6
	17	6	.2	.2	99.8
	18	2	.1	.1	99.8
	19	2	.1	.1	99.9
	20	1	.0	.0	99.9
	24	1	.0	.0	100.0
	26	1	.0	.0	100.0
	Total	3084	100.0	100.0	

For Total*ServiceAces*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

SetsWon Sets Won

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	647	21.0	21.0	21.0
	1	415	13.5	13.5	34.4
	2	275	8.9	8.9	43.4
	3	1746	56.6	56.6	100.0
	4	1	.0	.0	100.0
	Total	3084	100.0	100.0	

For *SetsWon*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

We have found there to be adequate variance in all variables. Moving on, we must analyze any influential cases.

Outlier Statistics[a]

		Case Number	Statistic	Sig. F
Stud. Deleted Residual	1	2922	-3.512	
	2	2941	-2.919	
	3	2935	-2.848	
	4	2920	-2.784	
	5	890	-2.616	
	6	2262	-2.537	
	7	2927	-2.525	
	8	522	-2.487	
	9	2579	-2.460	
	10	1244	-2.439	
Mahal. Distance	1	2536	50.935	
	2	3061	50.460	
	3	2696	47.550	
	4	675	32.195	
	5	1616	29.442	
	6	3037	28.713	
	7	1598	28.130	
	8	1718	28.051	
	9	2874	27.005	
	10	696	26.298	
Cook's Distance	1	3061	.017	1.000
	2	2922	.010	1.000
	3	2519	.005	1.000
	4	2930	.005	1.000
	5	3079	.003	1.000
	6	2924	.003	1.000
	7	589	.003	1.000
	8	2941	.003	1.000
	9	2536	.003	1.000
	10	1112	.003	1.000

a. Dependent Variable: SetsWon Sets Won

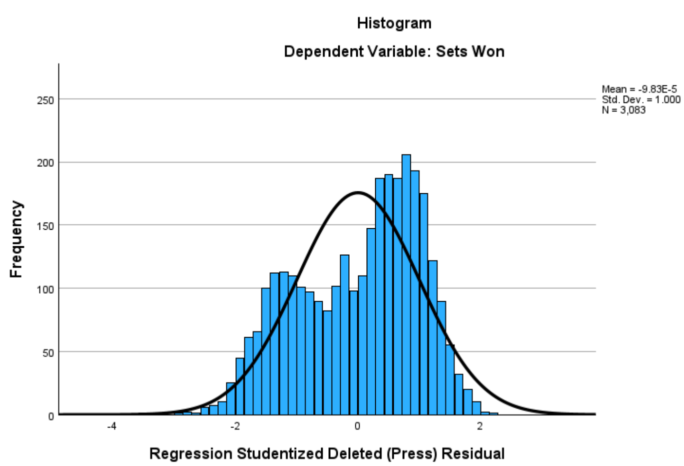
The Cook's Distance value examines both the predictor and dependent values and measures the amount by which a regression model changes by removing a single point. We can see that the case number (3061) with the maximum Cook's Distance value in the data appears 2nd highest for Mahal. Distance (measures Independent Variables) and does not appear for Stud. Deleted Residuals (measures Dependent Variables). Since the maximum of Cook's Distance is less than 1 (.017) however, we have no influential cases.

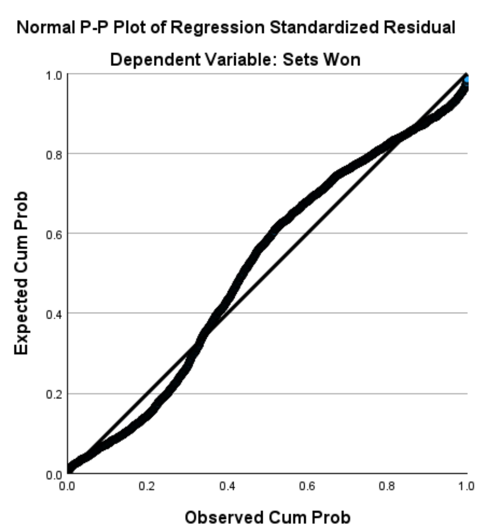
For the last primary assumption, we must analyze the linearity of the data. Plots were excluded for simplicity. While, for some variables, the differences between the R^2 cubic value and the R^2 linear value as well as the R^2 quadratic value and the R^2 linear value exceeded 2%, there were no major violations of linearity between the independent variables and the dependent variable.

Moving onto our secondary assumptions, we now must examine if there is sufficiently constant error variance (homoscedasticity).

Plots again were excluded for simplicity. The ratio for the difference in the greatest part of the scatter and the smallest part of the scatter for each variable and the standardized predicted values was less than 3:1, and thus we can assume that there is sufficiently constant error variance.

We have sufficiently constant error variance for all variables. For the next secondary assumption, we must examine if there is sufficient normality of the error residuals.





From the P-P Plot and the Histogram for the studentized deleted residuals, it appears as though there is sufficient normality of the error residuals. However, the data does deviate slightly in the

P-P Plot and the Histogram data appears to be leptokurtic/bimodal, so we need to check the underlying statistics to be sure.

Statistics

SDR_6 Studentized Deleted Residual

N	Valid	3084
	Missing	0
Mean		-.0000982
Skewness		-.384
Std. Error of Skewness		.044
Kurtosis		-.841
Std. Error of Kurtosis		.088
Minimum		-3.51181
Maximum		2.18252

We can assume normality for the error residuals, as the skewness of -.384 is between -3 and 3, and the kurtosis of -.841 is within the adjusted range of -20 to 20 for large sample sizes.

We have now met all the primary and secondary assumptions, and can continue with our regression analysis.

Correlations

		SetsWon Sets Won	TotalMissedServes Total Missed Serves	TotalServiceAces Total Service Aces	ServiceAcesAgainst Service Aces Against	ServiceErrorsAgainst Service Errors Against
Pearson Correlation	SetsWon Sets Won	1.000	.184	.419	-.272	-.010
	TotalMissedServes Total Missed Serves	.184	1.000	.217	.060	.200
	TotalServiceAces Total Service Aces	.419	.217	1.000	.016	.000
	ServiceAcesAgainst Service Aces Against	-.272	.060	.016	1.000	.247
	ServiceErrorsAgainst Service Errors Against	-.010	.200	.000	.247	1.000
Sig. (1-tailed)	SetsWon Sets Won	.	<.001	<.001	<.001	.285
	TotalMissedServes Total Missed Serves	.000	.	.000	.000	.000
	TotalServiceAces Total Service Aces	.000	.000	.	.195	.493
	ServiceAcesAgainst Service Aces Against	.000	.000	.195	.	.000
	ServiceErrorsAgainst Service Errors Against	.285	.000	.493	.000	.
N	SetsWon Sets Won	3083	3083	3083	3083	3083
	TotalMissedServes Total Missed Serves	3083	3083	3083	3083	3083
	TotalServiceAces Total Service Aces	3083	3083	3083	3083	3083
	ServiceAcesAgainst Service Aces Against	3083	3083	3083	3083	3083
	ServiceErrorsAgainst Service Errors Against	3083	3083	3083	3083	3083

From the correlation table, we can see that:

- The correlation relationship between *TotalMissedServes* and *SetsWon* is significant, with **r(3083) = .184** and **p < .001**. The coefficient of determination of the association is .184, meaning there is a positive, low correlation between total missed serves and sets won.
- The correlation relationship between *TotalServiceAces* and *SetsWon* is significant, with **r(3083) = .419** and **p < .001**. The coefficient of determination of the association is .419, meaning there is a positive, moderate correlation between total service aces and sets won.
- The correlation relationship between *ServiceAcesAgainst* and *SetsWon* is significant, with **r(3083) = -.272** and **p < .001**. The coefficient of determination of the association is -.272, meaning there is a positive, low correlation between total service aces against and sets won.
- The correlation relationship between *ServiceErrorsAgainst* and *SetsWon* is not significant, with **r(3083) = -.010** and **p = .285**.

Moving onto our ANOVA analysis within the overall regression analysis, our null hypothesis, H_O, is that there is no independent predictor variable that helps predict our dependent observation variable, where **α** = .05. If the F-statistic is significant (p < .05), we have sufficient evidence to reject the null hypothesis.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1267.689	4	316.922	280.251	<.001[b]
	Residual	3480.766	3078	1.131		
	Total	4748.455	3082			

a. Dependent Variable: SetsWon Sets Won

b. Predictors: (Constant), ServiceErrorsAgainst Service Errors Against, TotalServiceAces Total Service Aces, ServiceAcesAgainst Service Aces Against, TotalMissedServes Total Missed Serves

Based on the ANOVA table, there is an independent predictor variable that helps predict our dependent observation variable, as **F(4,3078) = 280.251** and **p < .001**. Thus, there is sufficient statistical evidence that at least one IPV has significant explanatory power.

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.517[a]	.267	.266	1.063	.267	280.251	4	3078	<.001

a. Predictors: (Constant), ServiceErrorsAgainst Service Errors Against, TotalServiceAces Total Service Aces, ServiceAcesAgainst Service Aces Against, TotalMissedServes Total Missed Serves

b. Dependent Variable: SetsWon Sets Won

Looking at the adjusted R-Square statistic **(R^2 = .266)**, 26.6% of the variance in *SetsWon* is explained by *TotalServiceAces, TotalMissedServes, ServiceAcesAgainst,* and *ServiceErrorsAgainst.*

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	1.403	.065		21.616	<.001	1.276	1.530					
	TotalMissedServes Total Missed Serves	.042	.006	.107	6.614	<.001	.029	.054	.184	.118	.102	.913	1.095
	TotalServiceAces Total Service Aces	.158	.006	.400	25.280	<.001	.146	.171	.419	.415	.390	.951	1.052
	ServiceAcesAgainst Service Aces Against	-.129	.007	-.295	-18.517	<.001	-.143	-.116	-.272	-.317	-.286	.939	1.065
	ServiceErrorsAgainst Service Errors Against	.016	.006	.041	2.524	.012	.004	.029	-.010	.045	.039	.903	1.108

a. Dependent Variable: SetsWon Sets Won

Based on the coefficients table:

- For every increase of one missed serve, we predict an increase of .042 in *SetsWon* (b = .042, t = 6.614, p < .001, 95% CI [.029, .054]). Because the p-value is below .05, the predictor is statistically significant. Additionally, B = .107, meaning for every one standard deviation increase in *TotalMissedServes*, we predict a .107 standard deviation increase in *SetsWon*. Without controlling for the influence of any other Independent

Predictor Variables, the associative relationship between *TotalMissedServes* and *SetsWon* is .184. When controlling for all other Independent Predictor Variables, the associative relationship between *TotalMissedServes* and *SetsWon* decreases to .102.

- For every increase of one service ace, we predict an increase of .158 in *SetsWon* ($b = .158$, $t = 25.280$, $p < .001$, 95% CI [.146, .171]). Because the p-value is below .05, the predictor is statistically significant. Additionally, $B = .400$, meaning for every one standard deviation increase in *TotalServiceAces*, we predict a .400 standard deviation increase in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *TotalServiceAces* and *SetsWon* is .419. When controlling for all other Independent Predictor Variables, the associative relationship between *TotalServiceAces* and *SetsWon* decreases to .390.
- For every increase of one service ace against, we predict a decrease of -.129 in *SetsWon* ($b = -.129$, $t = -18.517$, $p < .001$, 95% CI [-.143, -.116]). Because the p-value is below .05, the predictor is statistically significant. Additionally, $B = -.295$, meaning for every one standard deviation increase in *ServiceAcesAgainst*, we predict a -.295 standard deviation decrease in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *ServiceAcesAgainst* and *SetsWon* is -.272. When controlling for all other Independent Predictor Variables, the associative relationship between *ServiceAcesAgainst* and *SetsWon* decreases to -.286.
- For every increase of one service error against, we predict an increase of .016 in *SetsWon* ($b = .016$, $t = 2.524$, $p = .012$, 95% CI [.004, .029]). Because the p-value is below .05, the predictor is statistically significant. Additionally, $B = .041$, meaning for every one standard deviation increase in *ServiceAcesAgainst*, we predict a .041 standard deviation decrease in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *ServiceErrorsAgainst* and *SetsWon* is -.010. When controlling for all other Independent Predictor Variables, the associative relationship between *ServiceErrorsAgainst* and *SetsWon* increases to .039.

Conclusion

As expected, these findings show that service aces, service errors, service aces against, and service errors against are all significant predictors of the number of sets a team wins in a given game. Most surprisingly, there is a positive relationship between service errors and sets won. We theorize that the reason for this positive relationship could be:

1) serving aggressively equates to winning (high numbers of service errors are equated with high numbers of service aces)

2) the data contains copious amounts of lopsided matchups where high numbers of service errors are overcome by an overall gap in skill (possible future research can explore this)

3) there are confounding factors not accounted for in our data

Overall, these results suggest that serving aggressively has strategic value when compared to the risk of missing serves. For every one standard deviation increase in service aces, we predict a .400 standard deviation increase in sets won in a match, compared to just a .107 standard deviation increase in sets won for a one standard deviation increase in service errors. This relationship is more clearly displayed by service and service errors against: for every one standard deviation increase in service aces against, we predict a -.295 standard deviation decrease in sets won, compared to just a .041 standard deviation increase in sets won for a one standard deviation increase in service errors against.

Now, we will conduct another regression analysis just on data from Division III matches.

Statistics

		SetsWon Sets Won	TotalMissedServes Total Missed Serves	TotalServiceAces Total Service Aces	ServiceAcesAgainst Service Aces Against	ServiceErrorsAgainst Service Errors Against
N	Valid	798	798	798	798	798
	Missing	0	0	0	0	0
Mean		2.12	6.20	6.57	5.51	6.44
Skewness		-.858	.664	.936	.865	.740
Std. Error of Skewness		.087	.087	.087	.087	.087
Kurtosis		-1.019	.398	1.666	1.214	.538
Std. Error of Kurtosis		.173	.173	.173	.173	.173
Minimum		0	0	0	0	0
Maximum		3	18	26	23	19

The data for all of the selected variables is normal, as all skewness values are within the range of -3 to 3, and the kurtosis values are all within the adjusted range for large sample sizes of -20 to 20.

Now, we can move on to the assumptions for regression. First, we must determine if the data has adequate variance. Specifically, we must see if there is one value or a close set of values that make up over 90% of the values in our data.

ServiceAcesAgainst Service Aces Against

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	20	2.5	2.5	2.5
	1	47	5.9	5.9	8.4
	2	74	9.3	9.3	17.7
	3	102	12.8	12.8	30.5
	4	103	12.9	12.9	43.4
	5	100	12.5	12.5	55.9
	6	76	9.5	9.5	65.4
	7	80	10.0	10.0	75.4
	8	68	8.5	8.5	84.0
	9	38	4.8	4.8	88.7
	10	18	2.3	2.3	91.0
	11	32	4.0	4.0	95.0
	12	17	2.1	2.1	97.1
	13	8	1.0	1.0	98.1
	14	5	.6	.6	98.7
	15	3	.4	.4	99.1
	16	2	.3	.3	99.4
	17	4	.5	.5	99.9
	23	1	.1	.1	100.0
	Total	798	100.0	100.0	

For *ServiceAcesAgainst*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

ServiceErrorsAgainst Service Errors Against

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	8	1.0	1.0	1.0
	1	20	2.5	2.5	3.5
	2	42	5.3	5.3	8.8
	3	81	10.2	10.2	18.9
	4	93	11.7	11.7	30.6
	5	100	12.5	12.5	43.1
	6	101	12.7	12.7	55.8
	7	96	12.0	12.0	67.8
	8	75	9.4	9.4	77.2
	9	50	6.3	6.3	83.5
	10	37	4.6	4.6	88.1
	11	33	4.1	4.1	92.2
	12	9	1.1	1.1	93.4
	13	25	3.1	3.1	96.5
	14	11	1.4	1.4	97.9
	15	6	.8	.8	98.6
	16	5	.6	.6	99.2
	17	4	.5	.5	99.7
	18	1	.1	.1	99.9
	19	1	.1	.1	100.0
	Total	798	100.0	100.0	

For *ServiceErrorsAgainst*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

TotalMissedServes Total Missed Serves

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	19	.6	.6	.6
	1	64	2.1	2.1	2.7
	2	147	4.8	4.8	7.5
	3	254	8.2	8.2	15.7
	4	342	11.1	11.1	26.8
	5	409	13.3	13.3	40.0
	6	391	12.7	12.7	52.7
	7	379	12.3	12.3	65.0
	8	307	10.0	10.0	75.0
	9	245	7.9	7.9	82.9
	10	170	5.5	5.5	88.4
	11	128	4.2	4.2	92.6
	12	76	2.5	2.5	95.0
	13	69	2.2	2.2	97.3
	14	37	1.2	1.2	98.5
	15	17	.6	.6	99.0
	16	14	.5	.5	99.5
	17	9	.3	.3	99.8
	18	2	.1	.1	99.8
	19	2	.1	.1	99.9
	20	1	.0	.0	99.9
	21	2	.1	.1	100.0
	Total	3084	100.0	100.0	

For *TotalMissedServes*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

TotalServiceAces Total Service Aces

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	9	1.1	1.1	1.1
	1	26	3.3	3.3	4.4
	2	51	6.4	6.4	10.8
	3	74	9.3	9.3	20.1
	4	91	11.4	11.4	31.5
	5	96	12.0	12.0	43.5
	6	93	11.7	11.7	55.1
	7	83	10.4	10.4	65.5
	8	66	8.3	8.3	73.8
	9	53	6.6	6.6	80.5
	10	45	5.6	5.6	86.1
	11	38	4.8	4.8	90.9
	12	23	2.9	2.9	93.7
	13	15	1.9	1.9	95.6
	14	10	1.3	1.3	96.9
	15	10	1.3	1.3	98.1
	16	3	.4	.4	98.5
	17	6	.8	.8	99.2
	18	2	.3	.3	99.5
	19	1	.1	.1	99.6
	20	1	.1	.1	99.7
	24	1	.1	.1	99.9
	26	1	.1	.1	100.0
	Total	798	100.0	100.0	

For *TotalServiceAces*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

SetsWon Sets Won

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	153	19.2	19.2	19.2
	1	96	12.0	12.0	31.2
	2	52	6.5	6.5	37.7
	3	497	62.3	62.3	100.0
	Total	798	100.0	100.0	

For *SetsWon*, there is no violation of adequate variance, as no single response or close group of responses constitutes over 90% of the values.

We have found there to be adequate variance in all variables. Moving on, we must analyze any influential cases.

Outlier Statistics[a]

		Case Number	Statistic	Sig. F
Stud. Deleted Residual	1	2922	-3.286	
	2	2941	-2.823	
	3	2935	-2.802	
	4	2920	-2.703	
	5	2927	-2.562	
	6	2932	-2.545	
	7	2579	-2.473	
	8	2772	-2.418	
	9	2302	-2.335	
	10	2854	-2.329	
Mahal. Distance	1	2536	35.259	
	2	3061	34.463	
	3	2696	32.861	
	4	2287	20.338	
	5	3079	19.144	
	6	3037	18.116	
	7	2874	17.433	
	8	2839	17.284	
	9	2321	16.865	
	10	3004	16.527	
Cook's Distance	1	3061	.028	1.000
	2	2922	.021	1.000
	3	2536	.015	1.000
	4	2519	.014	1.000
	5	3079	.012	1.000
	6	2746	.011	1.000
	7	2930	.010	1.000
	8	3058	.009	1.000
	9	2939	.009	1.000
	10	2924	.009	1.000

a. Dependent Variable: SetsWon Sets Won

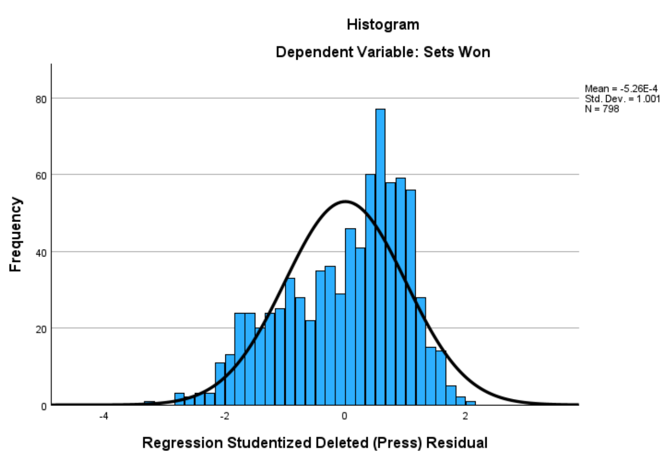
The Cook's Distance value examines both the predictor and dependent values and measures the amount by which a regression model changes by removing a single point. We can see that the case number (3061) with the maximum Cook's Distance value in the data appears 2nd highest for Mahal. Distance (measures Independent Variables) and does not appear for Stud. Deleted Residuals (measures Dependent Variables). Since the maximum of Cook's Distance is less than 1 (.028) however, we have no influential cases.

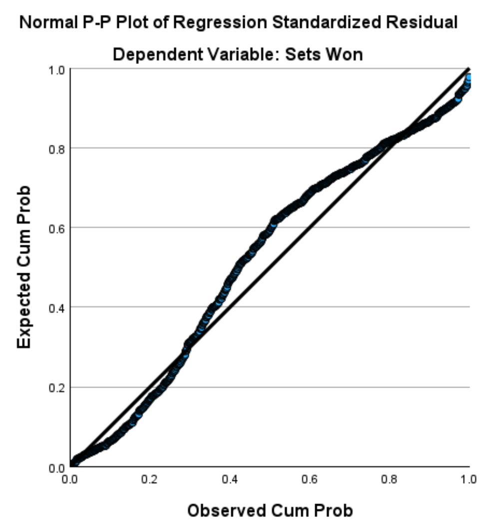
For the last primary assumption, we must analyze the linearity of the data. Plots were excluded for simplicity. While, for some variables, the differences between the R^2 cubic value and the R^2 linear value as well as the R^2 quadratic value and the R^2 linear value exceeded 2%, there were no major violations of linearity between the independent variables and the dependent variable.

Moving onto our secondary assumptions, we now must examine if there is sufficiently constant error variance (homoscedasticity).

Plots again were excluded for simplicity. The ratio for the difference in the greatest part of the scatter and the smallest part of the scatter for each variable and the standardized predicted values was less than 3:1, and thus we can assume that there is sufficiently constant error variance.

We have sufficiently constant error variance for all variables. For the next secondary assumption, we must examine if there is sufficient normality of the error residuals.





From the P-P Plot and the Histogram for the studentized deleted residuals, it appears as though there is sufficient normality of the error residuals. However, the data does deviate slightly in the P-P Plot and the Histogram data appears to be leptokurtic, so we need to check the underlying statistics to be sure.

Statistics

SDR_6 Studentized Deleted Residual

N	Valid	798
	Missing	0
Mean		.0232618
Skewness		-.464
Std. Error of Skewness		.087
Kurtosis		-.473
Std. Error of Kurtosis		.173
Minimum		-3.51181
Maximum		1.96831

We can assume normality for the error residuals, as the skewness of -.464 is between -3 and 3, and the kurtosis of -.473 is within the adjusted range of -20 to 20 for large sample sizes.

Correlations

		SetsWon Sets Won	TotalMissedServes Total Missed Serves	TotalServiceAces Total Service Aces	ServiceAcesAgainst Service Aces Against	ServiceErrorsAgainst Service Errors Against
Pearson Correlation	SetsWon Sets Won	1.000	.174	.389	-.313	.007
	TotalMissedServes Total Missed Serves	.174	1.000	.245	.142	.262
	TotalServiceAces Total Service Aces	.389	.245	1.000	.010	.029
	ServiceAcesAgainst Service Aces Against	-.313	.142	.010	1.000	.294
	ServiceErrorsAgainst Service Errors Against	.007	.262	.029	.294	1.000
Sig. (1-tailed)	SetsWon Sets Won	.	<.001	<.001	<.001	.425
	TotalMissedServes Total Missed Serves	.000	.	.000	.000	.000
	TotalServiceAces Total Service Aces	.000	.000	.	.384	.208
	ServiceAcesAgainst Service Aces Against	.000	.000	.384	.	.000
	ServiceErrorsAgainst Service Errors Against	.425	.000	.208	.000	.
N	SetsWon Sets Won	798	798	798	798	798
	TotalMissedServes Total Missed Serves	798	798	798	798	798
	TotalServiceAces Total Service Aces	798	798	798	798	798
	ServiceAcesAgainst Service Aces Against	798	798	798	798	798
	ServiceErrorsAgainst Service Errors Against	798	798	798	798	798

From the correlation table, we can see that:

- The correlation relationship between *TotalMissedServes* and *SetsWon* is significant, with **r(798) = .174** and **p < .001**. The coefficient of determination of the association is .174, meaning there is a positive, low correlation between total missed serves and sets won.
- The correlation relationship between *TotalServiceAces* and *SetsWon* is significant, with **r(798) = .389** and **p < .001**. The coefficient of determination of the association is .389, meaning there is a positive, moderate correlation between total service aces and sets won.
- The correlation relationship between *ServiceAcesAgainst* and *SetsWon* is significant, with **r(798) = -.313** and **p < .001**. The coefficient of determination of the association is -.313, meaning there is a positive, low correlation between total service aces against and sets won.
- The correlation relationship between *ServiceErrorsAgainst* and *SetsWon* is not significant, with **r(798) = .007** and **p = .425**.

Moving onto our ANOVA analysis within the overall regression analysis, our null hypothesis, H_O, is that there is no independent predictor variable that helps predict our dependent observation variable, where **α** = .05. If the F-statistic is significant (p < .05), we have sufficient evidence to reject the null hypothesis.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	325.085	4	81.271	74.197	<.001[b]
	Residual	868.605	793	1.095		
	Total	1193.690	797			

a. Dependent Variable: SetsWon Sets Won

b. Predictors: (Constant), ServiceErrorsAgainst Service Errors Against, TotalServiceAces Total Service Aces, ServiceAcesAgainst Service Aces Against, TotalMissedServes Total Missed Serves

Based on the ANOVA table, there is an independent predictor variable that helps predict our dependent observation variable, as **F(4,793) = 74.197** and **p < .001**. Thus, there is sufficient statistical evidence that at least one IPV has significant explanatory power.

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.522[a]	.272	.269	1.047	.272	74.197	4	793	<.001

a. Predictors: (Constant), ServiceErrorsAgainst Service Errors Against, TotalServiceAces Total Service Aces, ServiceAcesAgainst Service Aces Against, TotalMissedServes Total Missed Serves

b. Dependent Variable: SetsWon Sets Won

Looking at the adjusted R-Square statistic **(R^2 = .269)**, 26.9% of the variance in *SetsWon* is explained by *TotalServiceAces, TotalMissedServes, ServiceAcesAgainst,* and *ServiceErrorsAgainst.*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	1.603	.118		13.610	<.001	1.372	1.835					
	TotalMissedServes Total Missed Serves	.044	.012	.117	3.614	<.001	.020	.068	.174	.127	.109	.871	1.148
	TotalServiceAces Total Service Aces	.122	.011	.362	11.580	<.001	.101	.142	.389	.380	.351	.939	1.065
	ServiceAcesAgainst Service Aces Against	-.131	.012	-.354	-11.128	<.001	-.155	-.108	-.313	-.368	-.337	.909	1.100
	ServiceErrorsAgainst Service Errors Against	.026	.012	.070	2.136	.033	.002	.049	.007	.076	.065	.863	1.158

a. Dependent Variable: SetsWon Sets Won

Based on the coefficients table:

- For every increase of one missed serve, we predict an increase of .044 in *SetsWon* (b = .044, t = 3.614, p < .001, 95% CI [.020, .068]). Because the p-value is below .05, the predictor is statistically significant. Additionally, Β = .117, meaning for every one standard deviation increase in *TotalMissedServes*, we predict a .117 standard deviation increase in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *TotalMissedServes* and *SetsWon* is .174. When controlling for all other Independent Predictor Variables, the associative relationship between *TotalMissedServes* and *SetsWon* decreases to .109.
- For every increase of one service ace, we predict an increase of .158 in *SetsWon* (b = .122, t = 11.580, p < .001, 95% CI [.101, .142]). Because the p-value is below .05, the predictor is statistically significant. Additionally, Β = .362, meaning for every one standard deviation increase in *TotalServiceAces*, we predict a .362 standard deviation increase in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *TotalServiceAces* and *SetsWon* is .389. When controlling for all other Independent Predictor Variables, the associative relationship between *TotalServiceAces* and *SetsWon* decreases to .351.
- For every increase of one service ace against, we predict a decrease of -.131 in *SetsWon* (b = -.131, t = -11.128, p < .001, 95% CI [-.155, -.108]). Because the p-value is below .05, the predictor is statistically significant. Additionally, Β = -.354, meaning for every one standard deviation increase in *ServiceAcesAgainst*, we predict a -.354 standard deviation decrease in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *ServiceAcesAgainst* and *SetsWon* is -.313. When controlling for all other Independent Predictor Variables, the associative relationship between *ServiceAcesAgainst* and *SetsWon* decreases to -.337.
- For every increase of one service error against, we predict an increase of .026 in *SetsWon* (b = .026, t = 2.136, p = .033, 95% CI [.002, .049]). Because the p-value is below .05, the predictor is statistically significant. Additionally, Β = .070, meaning for every one

standard deviation increase in *ServiceAcesAgainst*, we predict a .070 standard deviation decrease in *SetsWon*. Without controlling for the influence of any other Independent Predictor Variables, the associative relationship between *ServiceErrorsAgainst* and *SetsWon* is .007. When controlling for all other Independent Predictor Variables, the associative relationship between *ServiceErrorsAgainst* and *SetsWon* increases to .065.

Conclusion

For only Division III matches, the outputs largely remain the same. All variables are significant predictors. Again, the value of accruing (and preventing) aces is notable.