Section 1: Introduction

The purpose of this paper is to investigate whether or not full-time employment status has an impact on the success in a game titled Warzone. My interest in this peaked as a way to win an argument with a friend who claims that you can't be that good at the game and work full-time. I am also interested in the ways you could predict wins or hours to be played in a given week by a player. The variables are **work**, **hours**, **KD**, and **wins**.

Work is a binary categorical variable that indicates full-time employment or not.

Hours represents the number of hours played per week.

KD is a popular in-game metric that calculates the ratio of number of enemies eliminated to number of times killed. It is often used as a measure of overall skill in the community.

Wins represents the number of games won. A game can only be won by outlasting all other opponents. **Good** is a judging criteria. If an observation **KD** and **wins** are greater than the product of the median of our sample data and median of the population (retrieved from CODwarzonestats.com) then 'good' was assigned, otherwise, 'bad' was assigned.

The data was collected by asking for responses for a questionnaire on different social platforms that host the Warzone community. First it was reddit but it was going slow so I moved to their Discord server. After I was approaching the end, I noticed that I had too few non-full-time observations. So I went into the game, found random players, and the first question I asked was whether they worked full-time. If they didn't then I'd move to the next player. After finding a player it was easy to get the stats I needed because of the resources the games developers provide. I could always ask them how many wins they had and what their KD was but to get how many hours they played, I was able to use a third party website, cod.tracker.gg, to view how many minutes they've played in the past 7 days. My prompts and screenshots of the messages I sent out are available in the appendix.

Section 2: Univariate Summary Statistics



Figure 2.1.1

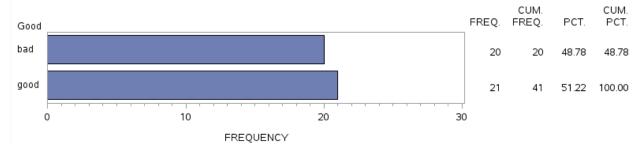


Figure 2.1.2

The summary stats for the only generated variable, **good**, shows a very even division which was exactly the goal.

The FREQ Procedure									
full time employment status									
Work	Frequency	Percent	Cumulative Frequency	Cumulative Percent					
no	15	36.59	15	36.59					
yes	26	63.41	41	100.00					

Figure 2.1.1

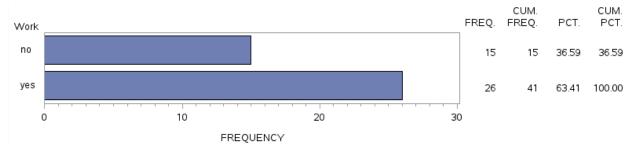


Figure 2.1.2

For **work**, it shows that the majority of our sample data was received from players who are full-time workers.

Quantiles (Definition 5)

										& crea
							tile	Quan		Level
							515	í	Max	100%
							515	í		99%
							165	1		95%
							134	1		90%
							104	1	2 3	75%
							55		Median	50%
		IATE Procedure					20		Q1	25%
		ole: Wins	Variab	,			12			10%
		Moments					9			5%
41		Sum Weights			N		0			1%
3286		Sum Observations		Mean 80.1463415			0		in	0% M
11173.778		Std Deviation 105.706093 Variance								
	11.297	Kurtosis		3.21055	Skewness		ns	servatio	me Ob	Extre
	446951				est	High	est	Low		
85182	16.508	Std Error Mean	1351	131.891	ff Variatior	Coe	Obs	Value	Obs	Value
		tical Measures	Statis	Basic			37	134	1	0
	y	Variability		ation	Loc		38	135	2	5
9	105.70609	Deviation	Std	80.14634	Mean		39	165	3	9
1	11174	ance	Vari	55.00000	Median		40	483	4	11
	515.00000	_	Ran	14.00000	Mode		41	515	5	12
)	84.00000	Interquartile Range							-	

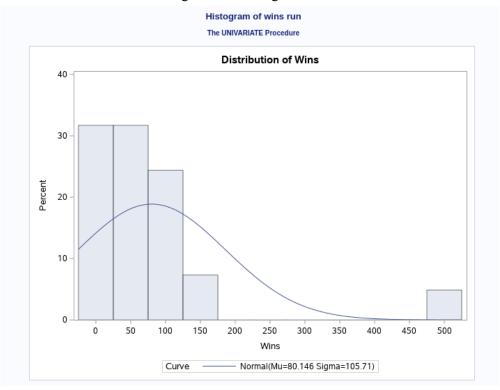
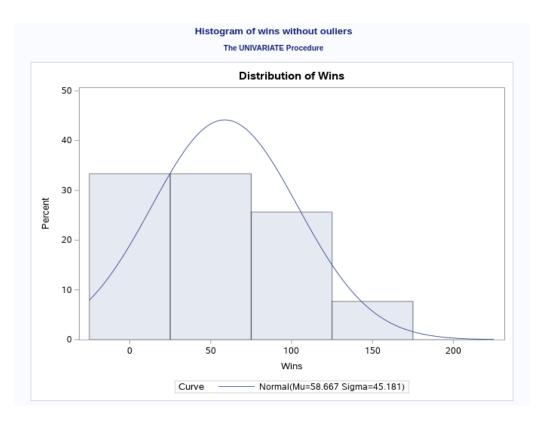


Figure 2.2.1 & Figure 2.2.2

Figure 2.2.3

The **wins** variable has two clear outliers with two observations greater than 200. Because of this it looks skewed to the right. In the next figure, a more unimodal, normal and less skewed distribution can be seen.



				IATE Procedure le: hours			
			Мо	ments			
N			41	Sum Weights			41
Mean		19.0300	244	Sum Observation	ons	780	.231
Std De	eviation	13.6746	919	Variance		186.9	9972
Skewness 1.25		1.25460	5460871 Kurtosis			1.1650386	
Uncorrected SS		S 22327.7	7029	Corrected SS		7479.88798	
Coeff Variation		71.8585	098	Std Error Mean		2.13562	2808
	Loc	,					
1	Mean	19.03002	Std	Deviation		3.67469	
1	Median	16.00000	Vari	ance	186.99720		
1	Mode	6.00000	Ran	ge	56	00000	
			Inte	rquartile Range	15	00000	

Figure 2.3.1

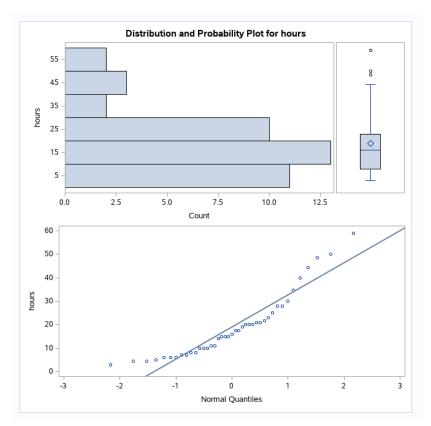


Figure 2.3.2

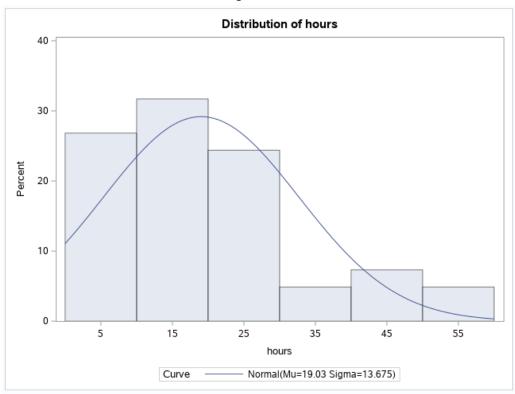


Figure 2.3.3

The amount of time is unimodally and normally distributed about the mean of 19 hours. It is left skewed with some players wanting to play for much longer than the majority. There are three outliers that play for much longer.

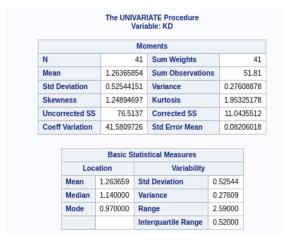


Figure 2.4.1

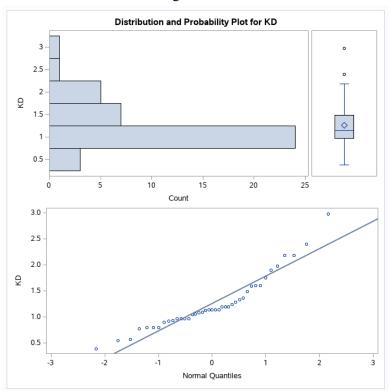


Figure 2.4.2

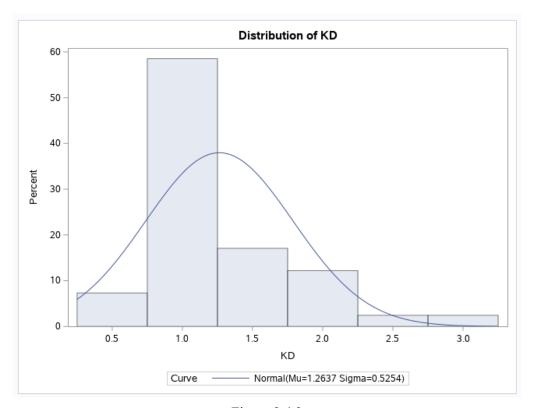


Figure 2.4.3

There are two outliers for the \mathbf{KD} variable that both go significantly above the average. It is unimodally and normally distributed about the mean of 1.2637

Section 3: Finding relationships

Summary stats grouped by employment status								
The MEANS Procedure								
			Work=no					
Variable	N	Mean	Std Dev	Minimum	Maximum			
Wins hours KD	15 15 15	93.2666667 28.1154000 1.3186667	115.4361089 17.2229440 0.5954694	12.0000000 4.5000000 0.3900000	483.0000000 59.0000000 2.9800000			
			Work=yes					
Variable	N	Mean	Std Dev	Minimum	Maximum			
Wins	26	72.5769231	101.2565743	0	515.0000000			

Figure 3.1.1

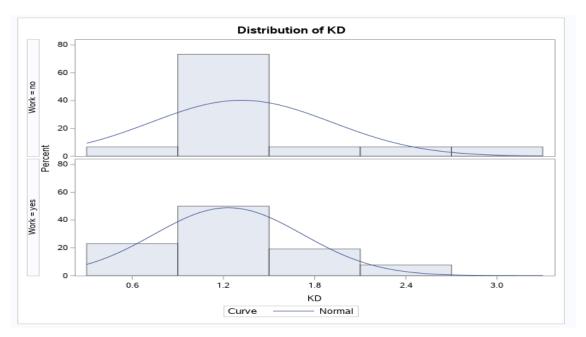


Figure 3.1.2

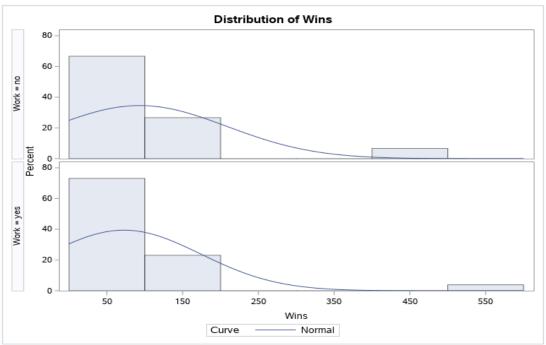


Figure 3.1.3

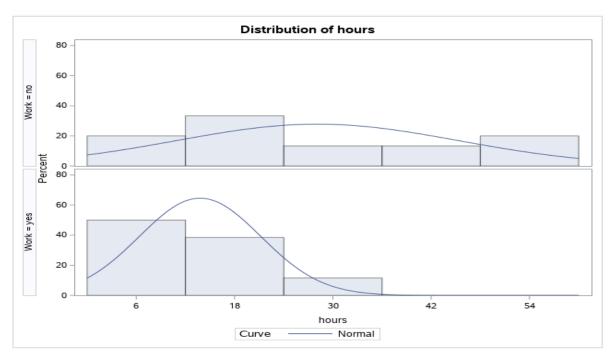


Figure 3.1.4

One observation to make when the distributions are split by employment status, is that the number of hours spent on the game are more normally distributed for the data from those who aren't full time employees.

	Summary stats grouped by player quality									
	The MEANS Procedure									
			Good=bad							
Variable	. N	l Mean	Std Dev	Minimum	Maximum					
Wins	20	27.8000000	26.7711549	0	110.0000000					
hours	20	12.7250000	9.2344820	3.0000000	40.0000000					
KD	20	1.0255000	0.5285877	0.3900000	2.9800000					
			Good=good							
Variable	N	Mean	Std Dev	Minimum	Maximum					
Wins	21	130.0000000	127.7376217	40.0000000	515.0000000					
	21	25.0348095	14.6754514	6.0000000	59.0000000					
hours										

Figure 3.2.1

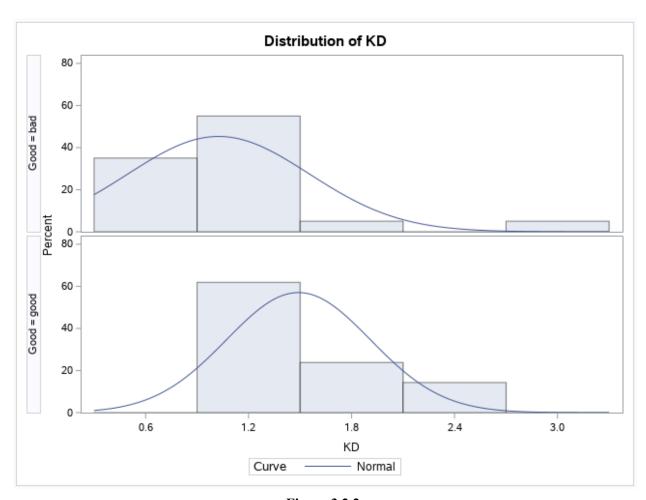


Figure 3.2.2

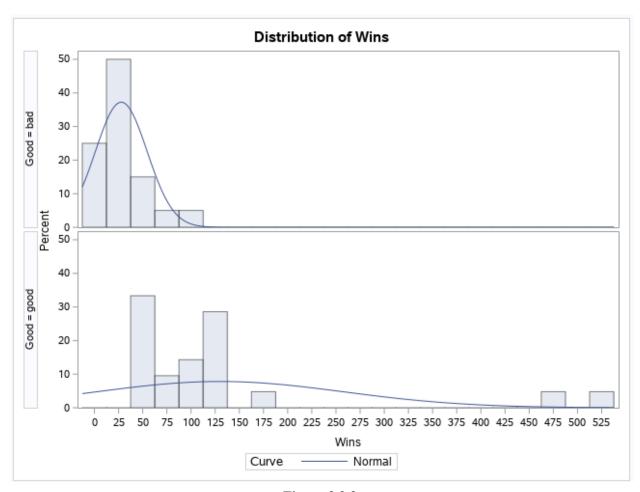


Figure 3.2.3

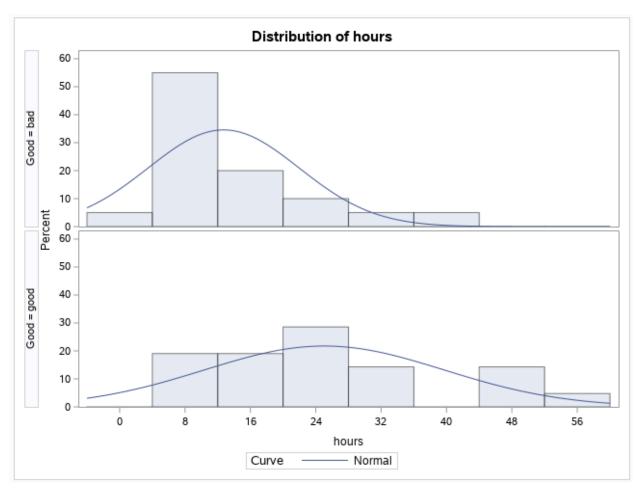


Figure 3.2.4

An interesting note is that of our subjectively "good" players, their hours played are more uniformly distributed.

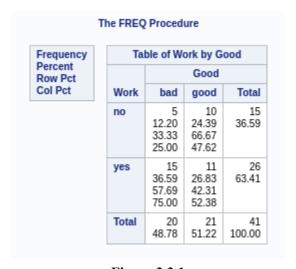


Figure 3.3.1

Statistic	DF	Value	Prob
Chi-Square	1	2.2590	0.1328
Likelihood Ratio Chi-Square	1	2.2924	0.1300
Continuity Adj. Chi-Square	1	1.3893	0.2385
Mantel-Haenszel Chi-Square	1	2.2039	0.1377
Phi Coefficient		-0.2347	
Contingency Coefficient		0.2285	
Cramer's V		-0.2347	

Figure 3.3.2

This is a test of independence conducted on the variables **good** and **work.** In the test we assume the null hypothesis that they are independent and choose an alpha value of 0.05. Since our p-value is greater than that, we can say that there is not sufficient enough evidence to say that they are not independent. Thus disproving my friend's argument, but **good** is a very subjective variable and there's more that can be explored, such as a model for wins.

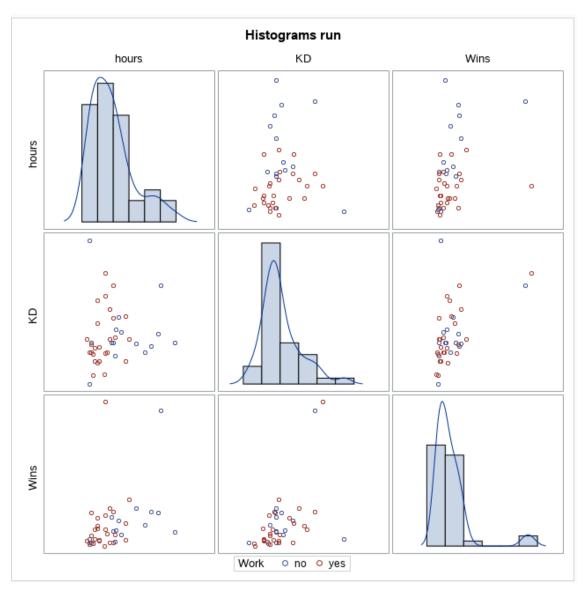


Figure 3.4.1

Most of the scatter plots look random and show little to know relationship. However, looking at Wins vs Hours without the two obvious outliers seems as though there is a positive association. IT will be shown without the outliers below.

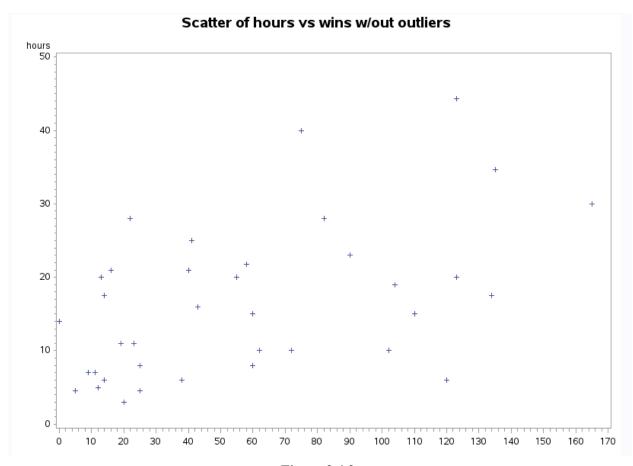


Figure 3.4.2

It's weak but linear and positive.



Figure 3.5.1

The correlation coefficient between **hours** and **wins** is 0.36467 with an r^2 of 0.133. This indicates that there is a weak positive relationship with only 13% of the variation in **hours** can be explained by the variation in **wins**. The p-value of 0.0191 suggests that it is significantly significant as well.

The correlation coefficient between **hours** and **KD** is 0.11021 with an r^2 of 0.0121. This indicates that there is a weak positive relationship with only 1.21% of the variation in **hours** can be explained by the variation in **KD**. The p-value of 0.4927 suggests that it is not significantly significant.

The correlation coefficient between **wins** and **KD** is 0.51614 with an r^2 of 0.2664. This indicates that there is a weak positive relationship with only 26.64% of the variation in **wins** can be explained by the variation in **KD**. The p-value of 0.4927 suggests that it is not significantly significant as well.

Section 4: Modelling

Of the most interesting things to model from the data, **wins** would be an obvious first choice. However, Time played or time that will be played would be a good thing to be able to predict from the business perspective. If you can manipulate one of the dependent variables like in our case, **wins** or **KD**, by selectively choosing who players fight against (i.e. putting good players against bad players and vice versa), then you could theoretically get the user to play longer.

Since **good** is a generated variable dependent on variables I already know, I'll just stick with the raw data and leave that out of the models I wish to produce. Also I'll be using a dataset that is modified to be free outliers to ensure normality of variables. The correlation matrix only had one significant statistic and it

suggests a weak relationship so collinearity is not a concern. It should be noted that observations are independent from each other in addition to all being normally distributed.

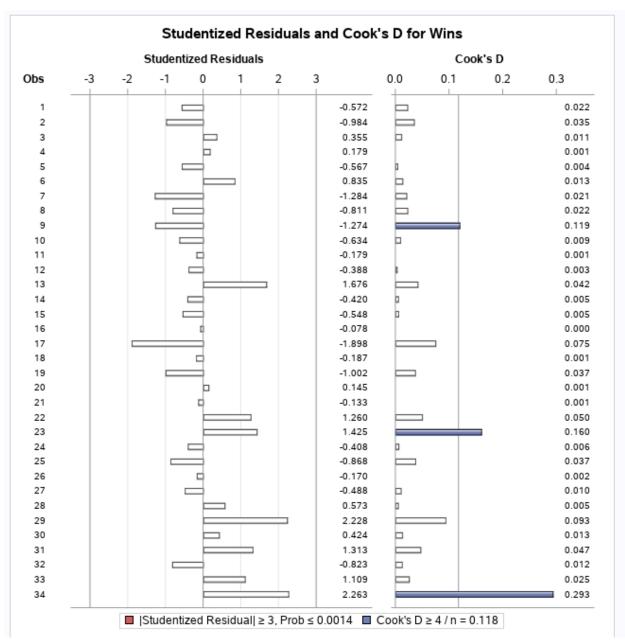


Figure 4.1.1

I will remove the outliers at observations 9, 23, and 34.

Lets use the formula $\hat{Y} = \Box_0 + x_1 \Box_1 + x_2 \Box_2 + x_3 \Box_3$, where \Box_i are the constants and x_i is **KD**, hours, and work(binary dummy version) respectively.

			Α	nalysis o	of Var	iance				
Sourc	Source		F	Sum of Squares		Mea Squa		F Val	ue	Pr > F
Model			3	15851	52	83.5107	72	4.	45	0.0115
Error		1	27	32052	11	87.1224	16			
Corre	cted Tota	al :	30	47903						
	Root MSE 34.45464 R-Squa		nare 0.3309							
	Dependent		Mear	50.80	0645 Adj R-Sq		-Sq	0.2565		
	Coeff Var			67.81	549					
			P	arameter	Estir	nates				
Va	riable	DF		ameter stimate	Sta	Standard Error t\		alue	Pr	> t
Int	ercept	1	-15	5.26355	6355 22.81785		-(0.67	0.5	092
KD)	1	51	1.16283	17.5	53026	2	2.92	0.0	070
ho	urs	1	(0.36330	1.:	13438	(0.32	0.7	512
w1		1		0.81346	15.63376		-	0.05	0.0	589

Figure 4.1.2

With an adjusted R² of 25%, using all the variables only explains 25% of the the **wins** variation. The p-value for the F-test shows that at least one of the variables is useful for predicting **wins**. Looking at the variable's individual T-test however we see that only **KD** has a p value that allows me to reject the null hypothesis with the others being wildly larger than alpha. I then used a partial F-test on the less significant variables and see if the p-value is large enough to accept the null hypothesis that they are insignificant predictors.

	M	odel: MODEL1							
Test test1 Results for Dependent Variable Wins									
Source	DF	Mean Square	F Value	Pr > F					
Numerator	2	65.79474	0.06	0.9462					
Denominator	27	1187.12246							

Figure 4.1.3

The p-value suggests that there is no evidence to reject my null hypothesis. So I will now make a model for wins that relies solely on KD. The equation is $\hat{Y} = \Box_0 + x_1 \Box_1$.

The REG Procedure Model: MODEL1 Dependent Variable: Wins Number of Observations Read 31 Number of Observations Used 31 **Analysis of Variance** Sum of Mean Source DF Squares Square F Value Pr > F Model 1 15719 15719 14.16 0.0008 Error 29 32184 1109.78952 **Corrected Total** 30 47903 Root MSE R-Square 0.3281 33.31350 0.3050 Dependent Mean 50.80645 Adj R-Sq Coeff Var 65.56944 Parameter Estimates

Parameter Standard Variable DF Estimate Error t Value Pr > |t| Intercept 1 -13.03998 17.98887 -0.72 0.4743 KD 1 54.01854 14.35329 3.76 8000.0

Figure 4.1.4

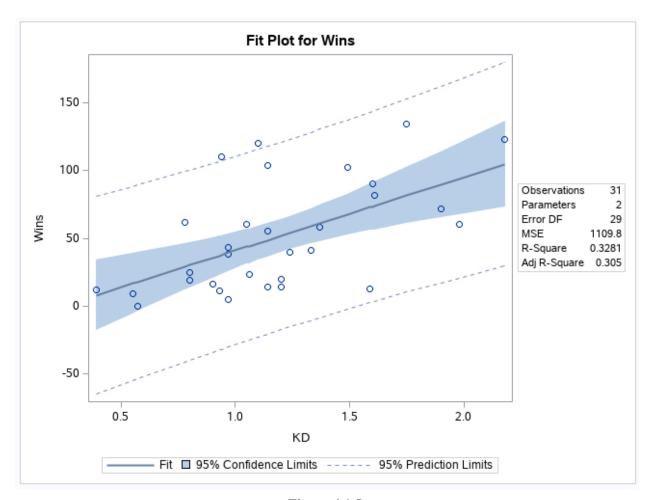


Figure 4.1.5

Final equation: wins = -13.0398887 + 54.01854(KD)

For a model to model hours.

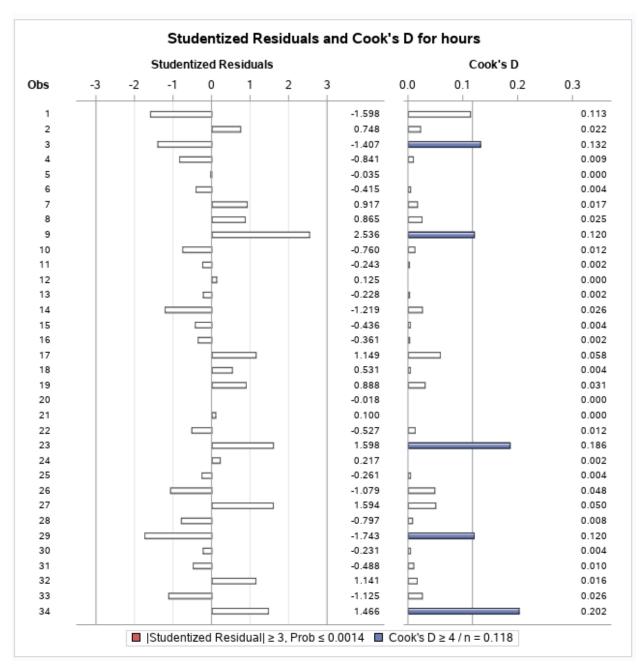


Figure 4.2.1

I will remove the outliers at observations 3,9, 23, 29, and 34.

Lets use the formula $\hat{Y} = \Box_0 + x_1 \Box_1 + x_2 \Box_2 + x_3 \Box_3$, where \Box_i are the constants and x_i is **KD**, wins, and work(binary dummy version) respectively.

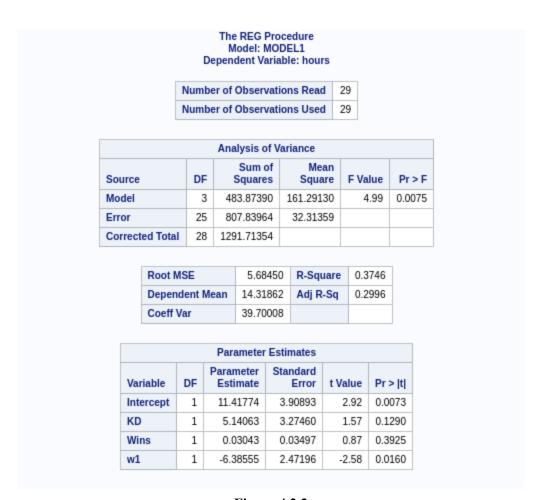


Figure 4.2.2

With an adjusted R² of 29%, using all the variables only explains 29% of the variable **hours** variation. The p-value for the F-test shows that at least one of the variables is useful for predicting **hours**. Looking at the variable's individual T-test however we see that only **work(binary dummy)** has a p value that allows me to reject the null hypothesis with the others being wildly larger than alpha. I then used a partial F-test on the less significant variables and see if the p-value is large enough to accept the null hypothesis that they are insignificant predictors.

		REG Procedu odel: MODEL:					
Test test1 Results for Dependent Variable hours							
Source	DF	Mean Square	F Value	Pr > F			
Numerator	2	123.21249	3.81	0.0359			
Denominator	25	32.31359					

Figure 4.2.3

This figure shows that the p-value is small enough to reject the null hypothesis. So I will test the hypothesis that only **wins** are insignificant

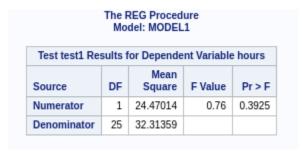


Figure 4.2.4

Sure enough the p-value is not small enough to reject the null hypothesis that **wins** are insignificant, so I will make a model that uses **KD** and **work(binary double).**

It will instead use the formula $\hat{Y} = \Box_0 + x_1 \Box_1 + x_2 \Box_2$.

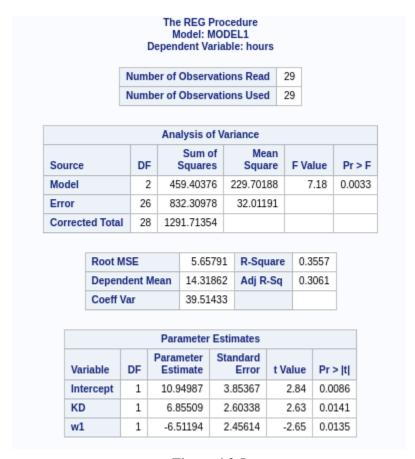


Figure 4.2.5

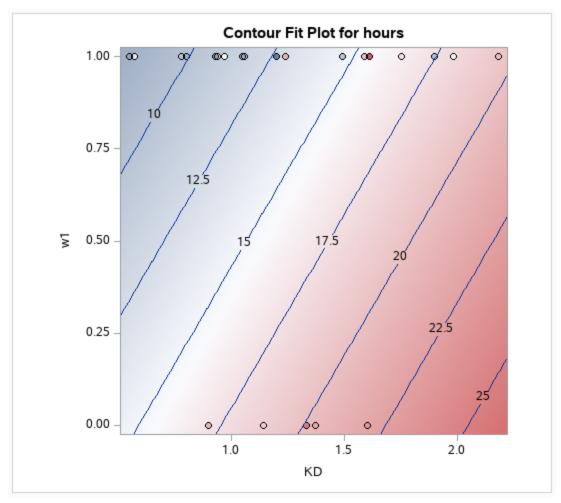


Figure 4.2.6

The final equation: hours = 10.94987 + 6.85509(KD) - 6.51194(work)

Section 5: Limitations

Certain factors just could not be accounted for with the lack of data. One interesting point made by a redditor was that most people play at night so the games during the night are almost always harder, therefore it would make sense to account for the time in which they most frequently play. This is data on games played so surely players will have their outlier games. There is no way to remove that from our data. Also a kill-death ratio and number of wins is not a full proof way to measure skill. Things such as shot accuracy and time spent alive are also good indicators. Speaking of wins, some people are newer to the game than others so they will surely have less wins. It doesn't necessarily follow that their low games won total is an indicator of a "bad" player. Certain choices were made to make this an achievable project but it is still on 1 a small project with only 40 observations.

Section 6: Conclusion

Given the sample it cannot be concluded that full-time employment status has an effect on wins which will be exciting news to report to others in the community as well as my friend. It was disappointing that a model for wins that used all other variables could not be made. Mostly because it was my guess that it

relied on these selected variables, especially **KD** and **hours.** Perhaps with more variables and a larger sample size, more could be discovered. However, I am very satisfied with the information I learned from this. I found a website that I could scrape to do a larger scale dive into this topic so maybe I will learn more.

Section 7: Appendix

```
Code:
*Author: Chibuikem Oparaoji
FILENAME REFFILE '/home/u57042395/warzoneData.csv';
PROC IMPORT DATAFILE=REFFILE
       DBMS=CSV
       OUT=Wardata2;
       GETNAMES=YES;
RUN;
PROC CONTENTS DATA=Wardata2; RUN;
*summary for good variable;
proc freq data=Wardata2;
label good='player quality';
table good;
run;
*bar chart for good varible;
proc gchart data=wardata2;
hbar good;
run;
*summary for work variable;
proc freq data=Wardata2;
label work='full time employment status';
table work;
run;
*bar chart for work varible;
proc gchart data=wardata2;
hbar work;
run;
*historgram for wins;
*also creates summary stats for all numeric varibles;
proc univariate data = wardata2;
histogram wins / midpoints=0 to 500 by 50 normal;
```

```
title "Histogram of wins"
run;
*stem and leaf plot for hours;
proc univariate data=wardata2 plot;
var hours;
run;
*stem and leaf plot for KD;
proc univariate data=wardata2 plot;
var KD;
run;
*sorting for summary stats;
proc sort data=wardata2;
by work;
run;
*Summary stats based on employment status;
proc means data=wardata2;
var wins hours KD;
by work;
title "Summary stats grouped by employment status";
*Making the histograms by the work varibale;
proc univariate data = wardata2 noprint;
histogram KD wins hours /normal;
class work;
title "Histograms"
run;
*sorting for summary stats but for good;
proc sort data=wardata2;
by good;
run;
*Summary stats based on player quality;
proc means data=wardata2;
var wins hours KD;
by good;
title "Summary stats grouped by player quality";
*Making the histograms by the player quality;
proc univariate data = wardata2 noprint;
histogram KD wins hours /normal;
class good;
title "Histograms"
```

```
run;
*Creating data set without wins outliers so I
can use it when I need to identify relationships using the wins variable;
Data wardata1;
set wardata2;
if wins < 300;
if hours < 40;
if KD < 2.2;
run;
*Scatter plots of numeric data;
proc sgscatter data=wardata2;
matrix hours KD wins / group=work diagonal=(histogram kernel);
run:
proc univariate data = wardata2 noprint;
histogram hours KD /normal;
title "Histogram";
run;
*Data step to creat a dummy variable for work;
data wardata1;
set wardata1;
if work='yes' then do;
w1 = 1;
end;
else if work ='no' then do;
w1 = 0;
end;
run;
*Veiwing the proc reg results;
proc reg data=wardata1;
model wins = KD hours w1 /R;
run;
*removing outliers;
Data warreg;
set wardata1;
if wins = 22 then delete;
if hours = 34.66 then delete;
if wins = 165 then delete;
run;
```

```
*crearting model;
proc reg data=warreg;
model wins = KD hours w1;
run;
*Testing the significance of hours and w1;
proc reg data=warreg;
model wins = KD hours w1;
test1: test hours=0,w1=0;
run;
*final model;
proc reg data=warreg;
model wins = KD;
run;
*Veiwing the proc reg results;
proc reg data=wardata1;
model hours = KD wins w1 /R;
run;
*removing outliers;
Data warreg1;
set wardata1;
if wins = 22 then delete;
if hours = 34.66 then delete;
if wins = 165 then delete;
if KD = 0.39 then delete;
if KD = 1.1 then delete;
run;
*crearting model;
proc reg data=warreg1;
model hours = KD wins w1;
run;
*Testing the significance of KD and wins;
proc reg data=warreg1;
model hours = KD wins w1;
test1: test wins=0,KD=0;
run;
```

```
*Testing the significance of wins;
proc reg data=warreg1;
model hours = KD wins w1;
test1: test wins=0;
run;
*final model;
proc glm data=warreg1;
model hours = KD w1;
run;
```

Data:

Work	KD	Wins	hours	Good
no	1.14	14	6	bad
yes	0.97	38	6	bad
no	0.9	16	21	bad
yes	1.06	23	11	bad
yes	0.78	62	10	bad
yes	1.98	60	15	good
no	2.98	25	4.5	bad
yes	1.2	14	17.5	bad
yes	0.57	0	14	bad
yes	0.8	22	28	bad
yes	0.97	5	4.5	bad
yes	1.9	72	10	good
yes	1.61	82	28	good
yes	0.55	9	7	bad
yes	1.05	60	8	good
no	0.39	12	5	bad
yes	1.1	120	6	good
no	1.33	41	25	good
yes	2.18	123	20	good
yes	1.75	134	17.5	good
yes	0.8	19	11	bad
yes	0.94	110	15	bad
yes	1.24	40	21	good

yes	1.49	102	10	good
yes	2.4	515	15	good
yes	1.2	165	30	good
yes	1.2	20	3	bad
yes	0.93	11	7	bad
yes	0.8	25	8	bad
yes	1.59	13	20	bad
yes	0.97	43	16	bad
no	1.6	90	23	good
no	1.14	55	20	good
no	0.97	75	40	bad
no	1.14	104	19	good
no	1.14	49	59	good
no	1.08	123	44.3	good
no	1.13	135	34.66	good
no	2.18	483	50.004	good
no	1.29	119	48.527	good
no	1.37	58	21.74	good

Pictures of sampling method:

Promts

Hi, that was a good post and I see you're an active member in the community, I was hoping you could help me answer a few questions for my school project on warzone players. I don't need much I just need your answer to these questions

Do you work full time?

What's your K/d?

How many wins do you have? (not including plunders and LTMs)

How many hours a week do you think you play?

If you can help that's fine, otherwise sorry for the spam message.

Hi, I get what you, I was hoping you could help me answer a few questions for my school project on warzone players. I don't need much I just need your answer to these questions

Do you work full time?

What's your K/d?

How many wins do you have? (not including plunders and LTMs)

How many hours a week do you think you play?

If you can help that's fine, otherwise sorry for the spam message.

LOL that was a funny post. Sorry for the long message but I was hoping you could help me answer a few questions for my school project on warzone players. I don't need much I just need your answer to these questions

Do you work full time?

What's your K/d?

How many wins do you have? (not including plunders and LTMs)

How many hours a week do you think you play?

If you can help that's fine, otherwise sorry for the spam message.

Hey nice clutch and sniper shots. Sorry for the long message but I was hoping you could help me answer a few questions for my school project on warzone players. I don't need much I just need your answer to these questions

Do you work full time?

What's your K/d?

How many wins do you have? (not including plunders and LTMs)

How many hours a week do you think you play?

If you can help that's fine, otherwise sorry for the spam message.

Hey, sorry to interrupt the flow. Could you help me answer a few questions for my school project on



GreenPhantom161

Hi, that was a funny post lol. Sorry ...



ShrekSeager123

Hi, that was a funny post lol. Sorry ...



Tyber_Roman

Hi, that was a good post and I see ...



Scobes96

DAMN that was a nice shot! I see y...



DontSmokeMate

Hey man, great content and I see y...



WatchmakerJJ

Hi, nice Sykov play! Sorry for the lo...



Paddy560

Hi, that was a funny post lol. Sorry ...



Datsgustrong

Hi, that was a funny post. I was ho...



WarzoneTerritory

Hi, that was a good compilation lol ...



VeteranWookie

Hi, that was some good content an...



DPRBassman

Thanks, No problem! Yes I work full...



BruhBrehBroh

Thanks a ton!



Kindly_Fancy

Good luck!



briklot

Thank you so much for the info!



ContentConch

Hi, that was a good post and I see ...



Wattsonmaster667

Hi, no clue if thats the case but suc...



SorryIdontknowmyName

Hi, that was a good post and I see ...





So I decided to do a Warzone project (2nd attempt for responses)

r/CODWarzone · Posted by u/IKrocking 9 days ago

So I decided to do a Warzone project for school Question

Need some help to get back on Warzone

Not a lot of help but I do need to know a little bit about you. Before I hop back on Warzone I need to complete a project for school and I've chosen to do it on Warzone players because... I wanted to play warzone when I started the project.

Here's what I need from you:

Are you a full-time worker?

Whats your k/d ratio?

How many hours a week do you think you play?

How many wins do you have? (does not include plunder, rebirth, and LTMs)

You can comment, send, a message, or a chat. I'll record all of it.

I don't need a lot of data. (Only 40 entries and I asked elsewhere too)

Thank you! Here's my answer in case you're wondering No 1.18 6 hours a week 14 wins

2 points · 17 comments