

Final Report

Spenser White

May 4, 2017

Abstract: This research investigates if opponents and quarter of the basketball game are significant factors in determining the number of fouls called against the Denver Nuggets during a National Basketball Association game. Poisson processes are identified as reasonable models for sequences of fouls called against the Nuggets, so we use a Poisson regression model to identify differences in foul intensity with respect to our factors of interest. We find that both factors have significant effects. For comparison, we also fit a similar model for each of the Northwest Division teams. This information is useful in helping teams assess their play style and improve on their fouling tendencies.

1 Introduction

This project seeks to analyze the effect differences in fouling tendencies among NBA teams, specifically the Denver Nuggets. The question asked in this study is ‘What is the foul intensity of the Nuggets and the teams they play?’ with the hypothesized answer that they are all equal. We examine the change of the response factor, number of fouls, by looking at how it is effected by the factors of interest, team and quarter of play.

By analyzing the fouling tendencies of the Denver Nuggets, when matched against teams in the Northwest Division, we can provide the team with valuable insight into their own behavior. This can lead to a competitive advantage against the teams studied, as they will know when it is safe to be more aggressive or when it is wise to play closer to the chest.

For a brief view of the complete dataset, see A1a. For the dataset used on the Denver Nuggets division, see A1b.

2 Definitions

For this project we modelled time sequences of fouls using the Poisson process. This was decided on after exploratory analysis of the model revealed a uniform distribution of fouls across the time span 9.5 minutes remaining to 1 minute remaining within every quarter of play. The term ‘foul intensity’ is used here in reference to the value of λ , which is the mean value of a Poisson distribution. Specifically:

2. DEFINITIONS

$$\hat{\lambda}_n = \frac{1}{n} \sum_{j=1}^n x_j$$

To use a Poisson we first collected the data in such a fashion that it fulfilled the requirements of a counting process. The underlying assumptions of a counting process are:

- i) $N(t) \geq 0$.
- ii) $N(t)$ is integer valued.
- iii) if $s < t$, then $N(s) \leq N(t)$.
- iv) For $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval (s, t) .

These assumptions are intrinsically satisfied for our data; there can never be less than 0 fouls in a quarter, the number of fouls can only increase by 1 at a time, and the number of fouls in a quarter increase in a positive additive fashion. Once the counting process assumptions have been met, we can use the Poisson model to examine our data, if the assumptions of the Poisson model are met.

The assumptions of the Poisson model are:

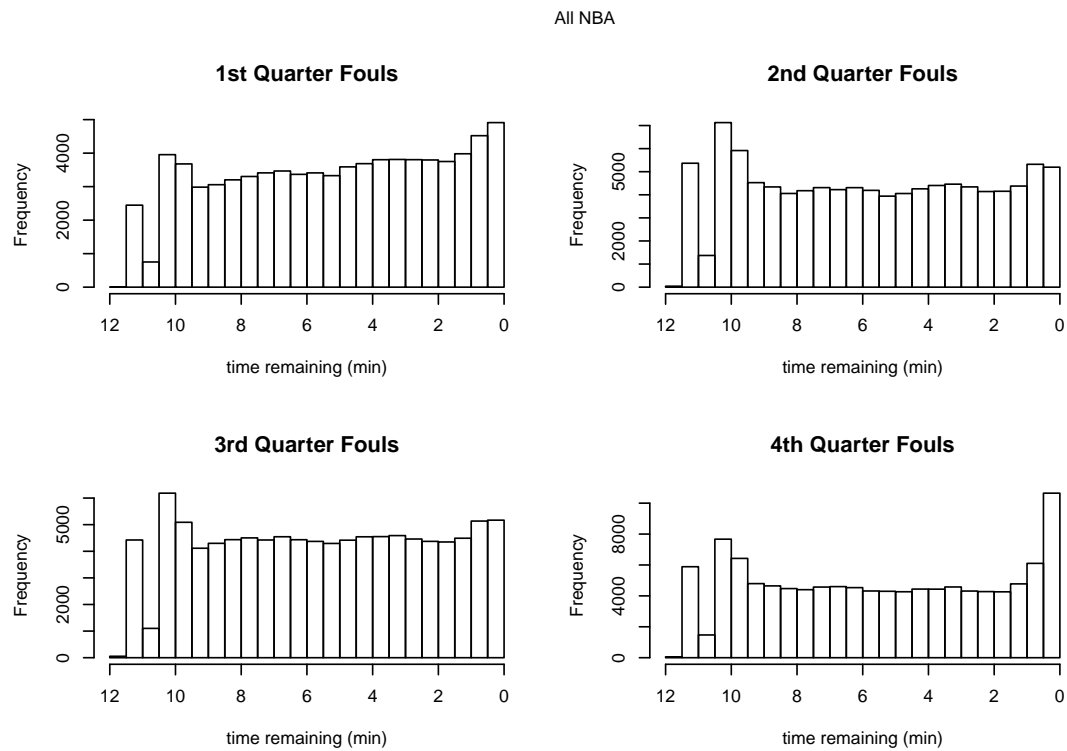
- i) $N(0) = 0$.
- ii) The process has independent increments.
- iii) The number of events in any interval of length t is Poisson distributed with mean λt . That is, for all $s, t \geq 0$

$$PN(t+s) - N(s) = n = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \text{ for } n = 0, 1, \dots$$

Again, the data we collected fits these assumptions well. The number of fouls at time 0 is 0, where we define time 0 to begin at sub 2.5 minutes from the beginning of the quarter (the reasoning for which is explained below). There is a possibility, however, that the second assumption is not met in that while any foul is ideally not related to the occurrence of any past foul, in reality many choices regarding future foul-drawing behavior are made on the past amount of fouls called. For example, a player may become upset at a past call and play more aggressively out of frustration, making his future fouling rate dependent on his past fouling rate. Another example is that coaches often remove players approaching a fouling limit, which alters the amount of future fouls likely to be committed by said team.

We have pressed forward with using a Poisson model, despite these possible objections, as the exploratory analysis of the data revealed a uniform distribution consistently present within the data for consistent time periods across quarters. A result of the Poisson model is that for any time interval $N(t) = n$, the distribution of the n random variables should be uniform on the interval $(0, t)$. This was observed for the interval $(9.5, 1)$ in minutes remaining for every quarter of play, and points to a Poisson process existing within that time frame. That is the reason we have set 9.5 equal to 0 in the counting of fouls in our data. See graph 1 for a visual representation.

Graph 1



Having observed this data and concluded that a Poisson process was appropriate for explaining what was seen, a test was carried out to determine if the value of λ differed for the Denver Nuggets depending on the team they played against.

For clarity: Our response variable is the amount of fouls called on the Denver Nuggets. The factors of interest are opposing team (with a possible 4 levels being the Utah Jazz, the Oklahoma City Thunder, the Minnesota Timberwolves, and the Portland Trail Blazers) and quarter of play (with 4 possible levels, being 1st quarter, 2nd quarter, 3rd quarter, and 4th quarter).

See A2a for code. See A2b for generic function.

$$H_0 : \lambda_{q1} = \lambda_{q2} = \lambda_{q3} = \lambda_{q4}$$

$H_a : H_0$ is not true.

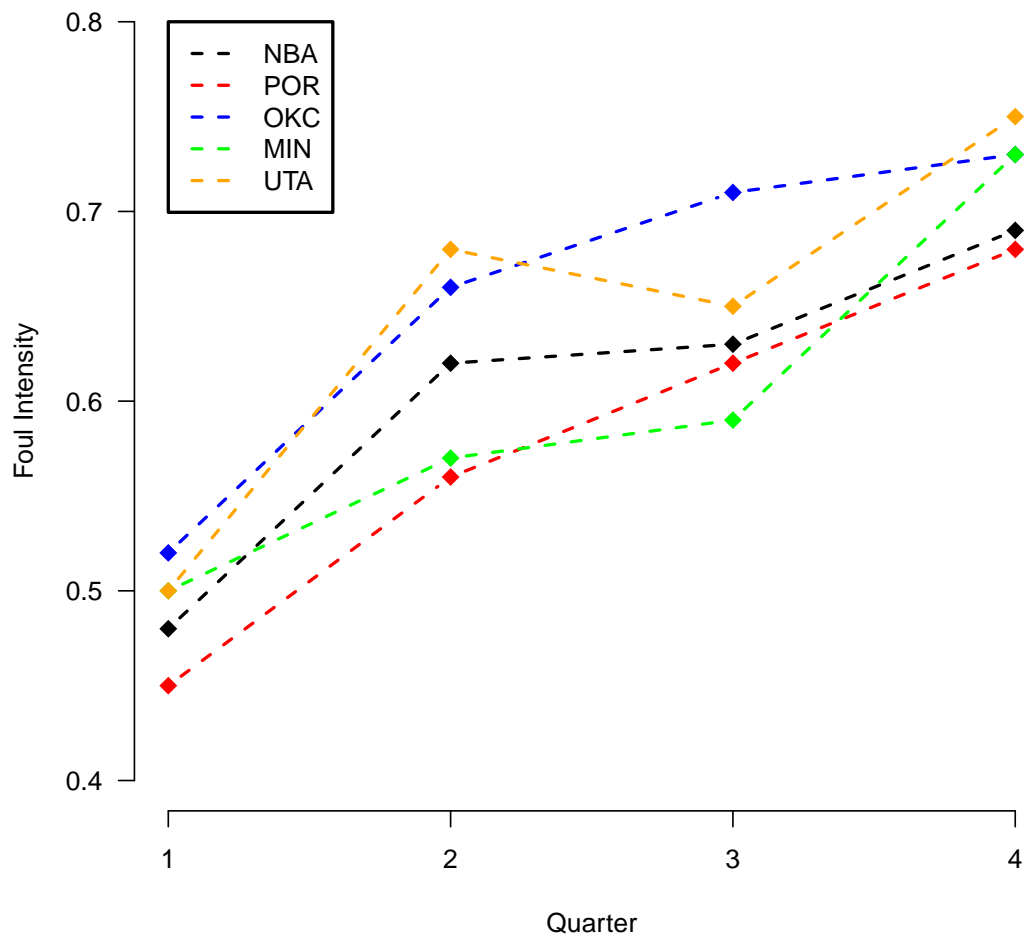
testing at $\alpha = .10$

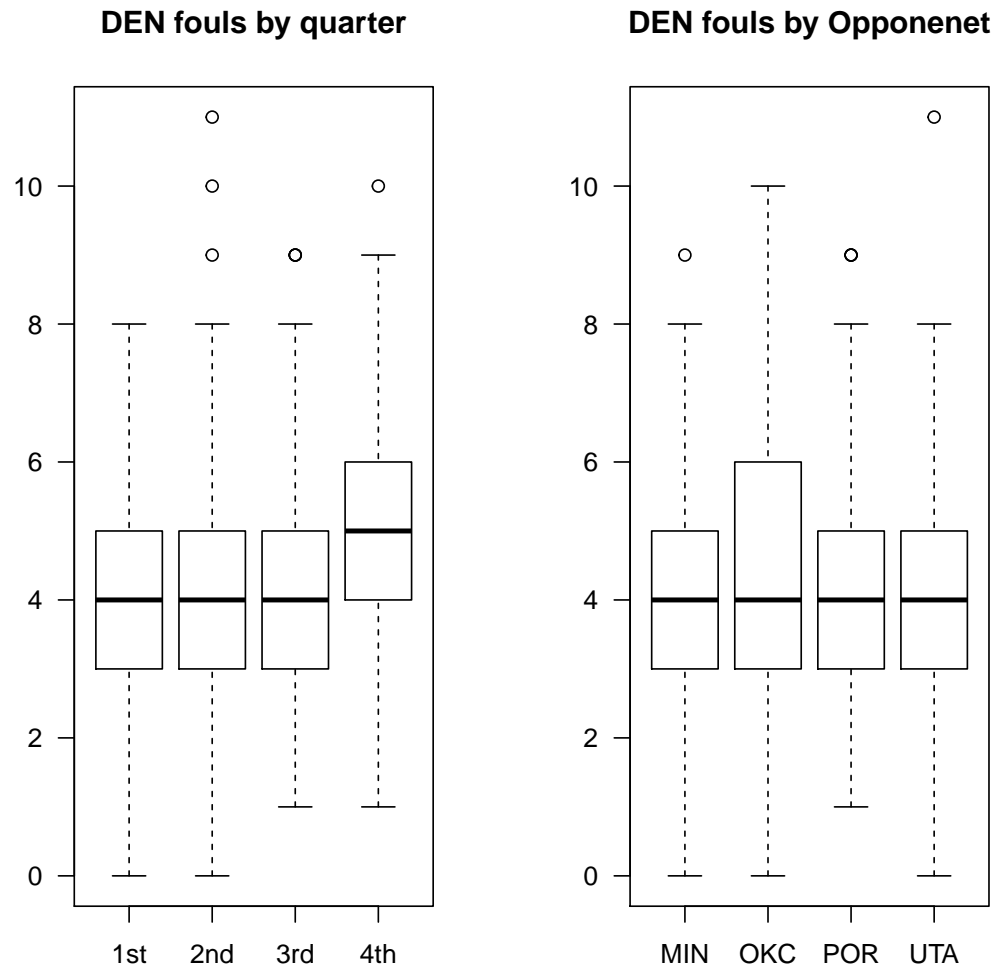
Table 1:

	NBA	POR	OKC	MIN	UTA
Full Game	0.49	0.48	0.54	0.50	0.51
Quarter 1	0.41	0.42	0.44	0.45	0.43
Quarter 2	0.50	0.48	0.57	0.47	0.52
Quarter 3	0.51	0.48	0.55	0.49	0.50
Quarter 4	0.54	0.52	0.59	0.58	0.59

See A3 for code.

Denver Nuggets Foul Intensity by Quarter by Team





Tukey's test was ran on the data set, concerning fouls committed as predicted by quarter and fouls committed as predicted by opponent. The ANOVA test for fouls predicted by team came back insignificant, but the ANOVA for fouls committed as predicted by quarter was significant. The Tukey's table for that ANOVA can be seen below.

1st	2nd	3rd	4th

The results of the test show that 2nd and 3rd quarters aren't very different, and neither are the 3rd and 4th quarters. We see strong differences between every other quarter, however, so there does in fact appear to be a difference in fouls as sorted by quarter.

Assuming, then, that the assumptions of the Poisson process are met we built a Poisson

Table 2:

	term	comparison	estimate	conf.low	conf.high	adj.p.value
1	quarter	2nd-1st	0.63	0.04	1.23	0.03
2	quarter	3rd-1st	0.61	0.01	1.20	0.04
3	quarter	4th-1st	1.18	0.59	1.78	0.00
4	quarter	3rd-2nd	-0.02	-0.62	0.57	1.00
5	quarter	4th-2nd	0.55	-0.04	1.15	0.08
6	quarter	4th-3rd	0.58	-0.02	1.17	0.06

regression model to describe the events seen in the data. The model uses binary qualitative variables to predict the foul count for a combination of the factors of interest. Our factors of interest for this model, after exploratory analysis showed them to be of potential significance, are quarter of play and team played against. Specifically, we are interested in the North-West division of the NBA, which the Denver Nuggets are a part of, for teams played against. The teams within the North-West Division are the Portland Trailblazers (POR), the Minnesota Timberwolves (MIN), the Utah Jazz (UTA), and the Oklahoma City Thunder (OKC).

The final model we used can be written as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

where β_0 represents the intercept, which further represents the base model of DEN against MIN in the first quarter. All other values within the model are then compared to this. Furthermore, $\beta_1 - \beta_6$ represent the regression coefficients of our model, which show by how much the foul count, Y , can be expected to change if that β 's corresponding X value is included in the model.

The X values are:

$$X_1 = \begin{cases} 1, & \text{if the opponent is OKC} \\ 0, & \text{otherwise} \end{cases}$$

$$X_4 = \begin{cases} 1, & \text{if it is the first quarter} \\ 0, & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if the opponent is POR} \\ 0, & \text{otherwise} \end{cases}$$

$$X_5 = \begin{cases} 1, & \text{if it is the second quarter} \\ 0, & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1, & \text{if the opponent is UTA} \\ 0, & \text{otherwise} \end{cases}$$

$$X_6 = \begin{cases} 1, & \text{if it is the third quarter} \\ 0, & \text{otherwise} \end{cases}$$

An example of interpreting the output of this model can be found in the next section.

3 Design and Justification

As seen in the section above, the final model to explain the number of fouls the Denver Nugget accrue over the course of a quarter was an additive Poisson model. An additive model was preferred over an interaction model, as the inclusion of interaction terms resulted in null significance for nearly every coefficient term. See Table 3 below for a coefficients summary of the interactive model.

Table 3:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3383	0.0905	14.78	0.0000
Oklahoma City Thunder	-0.0273	0.1300	-0.21	0.8334
Portland Trail Blazers	-0.0718	0.1315	-0.55	0.5851
Utah Jazz	-0.0449	0.1306	-0.34	0.7310
quarter2nd	0.0480	0.1265	0.38	0.7044
quarter3rd	0.0863	0.1254	0.69	0.4911
quarter4th	0.2649	0.1204	2.20	0.0277
Oklahoma City Thunder:quarter2nd	0.2110	0.1772	1.19	0.2339
Portland Trail Blazers:quarter2nd	0.0878	0.1818	0.48	0.6291
Utah Jazz:quarter2nd	0.1373	0.1795	0.76	0.4444
Oklahoma City Thunder:quarter3rd	0.1455	0.1769	0.82	0.4110
Portland Trail Blazers:quarter3rd	0.0574	0.1808	0.32	0.7509
Utah Jazz:quarter3rd	0.0615	0.1794	0.34	0.7319
Oklahoma City Thunder:quarter4th	0.0400	0.1720	0.23	0.8159
Portland Trail Blazers:quarter4th	-0.0381	0.1756	-0.22	0.8282
Utah Jazz:quarter4th	0.0512	0.1726	0.30	0.7669

As such, an additive model was used to explain the data instead. See Table 4 below for the summary output.

Table 4:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.2908	0.0596	21.66	0.0000
Oklahoma City Thunder	0.0733	0.0601	1.22	0.2227
Portland Trail Blazers	-0.0469	0.0620	-0.76	0.4491
Utah Jazz	0.0187	0.0609	0.31	0.7583
2nd Quarter	0.1585	0.0635	2.50	0.0125
3rd Quarter	0.1529	0.0636	2.41	0.0161
4th Quarter	0.2789	0.0618	4.51	0.0000

Although not significant in the model, the team factor remains in this final model in an effort to detract from any over signifying quarter may have on the response. It is apparent from Table 4 above that while there may not be a significant change in fouls as according to team there is some level of variance caused by this factor. A more clear example can be seen in Graph 1, with the Utah Jazz. While the Jazz are following the overall trend of the data, they deviate in being

the only team to cause a decrease in foul intensity in the third quarter. So, while their effects may not be pronounced enough to find significant difference between the teams, this factor's inclusion in the model will help give better results than a model without.

To read this output, first we must note that model is a logmatic regression, or that the response variable is modeled on the log of the predictors. Then is is clear that to realize the true coefficients of the model we must raise e to the regression results, which will then tell us in the expected amount of fouls per quarter we expect to see given the factors included.

Furthermore, to read the output of this table we must understand the implications of the base model. Being a purely qualitative model, our regression summary output shows us the coefficients of the factors of interest if that level of that factor is to be examined. If we chose to look at non of the factors, and only consider the intercept, then we are left with a model which represents the base to which the rest of the coefficients are comparative to.

An example:

Our β_0 here represents the expected number of fouls in a certain quarter against a certain team. We will receive only β_0 , and therefore that value, when all other factors in the model are equal to zero. As a result, the intercept represents Denver versus Minnesota in the first quarter. This number is read off the summary as 1.29079. To interpret this model in terms of fouls expected, we raise e to this value, resulting in 3.635658. That is, we expect the Nuggets to commit 3.64 fouls, on average, in the first quarter of play against the Minnesota Timberwolves.

If we were interested in examining the expected fouls the Nuggets commit against OKC in the third quarter, we could take the sum of β_0 and the coefficients that OKC and the third quarter represent, in this case 0.07331 and 0.15291 respectively. Then we take e raised to this sum to find the expected number of fouls per minute, which results in 4.558575 See Table 5 for a list of the transformed coefficient estimates.

Table 5:

	Fouls
(Intercept)	3.34
4th Quarter	1.32
3rd Quarter	1.17
2nd Quarter	1.17
Oklahoma City Thunder	1.08
Utah Jazz	1.02
Portland Trail Blazers	0.95

Note that above in Table 5 that all estimates for each factor are positive, meaning that the amount of fouls in the base model, represented by β_0 , are the lowest amount we expect the Denver Nuggets to commit in any quarter against any North-West Division team. This was incidental, as the levels of factors included in the base model are decided alphabetically.

To understand the effectiveness of the Poisson regression model we created, two pseudo- R^2 values were used - McFadden's and Cox-Snell's. See A5 for an explanation of these values. The results of the pseudo- R^2 values are 0.0101 and 0.0394 respectively. These values inspire little confidence in our model and, as a result, it must be said that model appears ineffective in

predicting the foul counts of the Denver Nuggets.

4 Model Testing

While no model testing was done due to time constraints, we were able to build models for each of the other North-West Division teams in regards to fouls they commit. See tables FILL THIS IN to see the coefficients summary for MIN, POR, UTA, and OKC.

Table 6: MIN

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3775	0.0587	23.48	0.0000
Oklahoma City Thunder	-0.1326	0.0616	-2.15	0.0313
Portland Trail Blazers	-0.0787	0.0607	-1.30	0.1952
Utah Jazz	-0.1123	0.0613	-1.83	0.0669
2nd Quarter	0.1384	0.0639	2.17	0.0303
3rd Quarter	0.1666	0.0635	2.62	0.0087
4th Quarter	0.1776	0.0633	2.80	0.0051

Table 7: POR

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3144	0.0601	21.87	0.0000
Minnesota Timberwolves	-0.0694	0.0613	-1.13	0.2574
Oklahoma City Thunder	-0.0556	0.0606	-0.92	0.3587
Utah Jazz	-0.1428	0.0619	-2.31	0.0212
2nd Quarter	0.1732	0.0648	2.67	0.0075
3rd Quarter	0.2089	0.0642	3.25	0.0011
4th Quarter	0.3091	0.0628	4.92	0.0000

Table 8: UTA

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3546	0.0593	22.85	0.0000
Minnesota Timberwolves	-0.1767	0.0605	-2.92	0.0035
Oklahoma City Thunder	-0.0638	0.0587	-1.09	0.2776
Portland Trail Blazers	-0.2204	0.0607	-3.63	0.0003
2nd Quarter	0.2428	0.0642	3.78	0.0002
3rd Quarter	0.2784	0.0637	4.37	0.0000
4th Quarter	0.3295	0.0630	5.23	0.0000

See A7 for these models.

For all these models the base model was picked to be the team observed against Denver in the first. These additional models show us that while the results might be non-significant,

Table 9: OKC

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3202	0.0606	21.78	0.0000
Minnesota Timberwolves	-0.0620	0.0613	-1.01	0.3120
Portland Trail Blazers	-0.2086	0.0632	-3.30	0.0010
Utah Jazz	-0.1137	0.0621	-1.83	0.0673
2nd Quarter	0.1152	0.0666	1.73	0.0839
3rd Quarter	0.1839	0.0656	2.80	0.0050
4th Quarter	0.3408	0.0634	5.38	0.0000

they are sound in their methodology.

Studying these results reveals several interesting facts, namely the implication that DEN has the highest foul drawing tendency in the North-West Division and that all teams are progressively committing more fouls as time goes on. Taking tally of the results, every team shows DEN with at least a slight advantage in the foul drawing department, with MIN displaying the highest amount, being significantly more than every other North-West Division team, and POR with only one team slightly crossing the significance threshold. To asses the validity of this, a model was built comparing the number of fouls Denver draws against opponents, that is fouls that opponents commit against them, and then performing an analysis on that. The results can be seen below.

Table 10: DEN

	POR	OKC	MIN	UTA
1	0.45	0.45	0.44	0.47
2	0.52	0.48	0.58	0.56
3	0.55	0.60	0.57	0.61
4	0.58	0.56	0.53	0.63

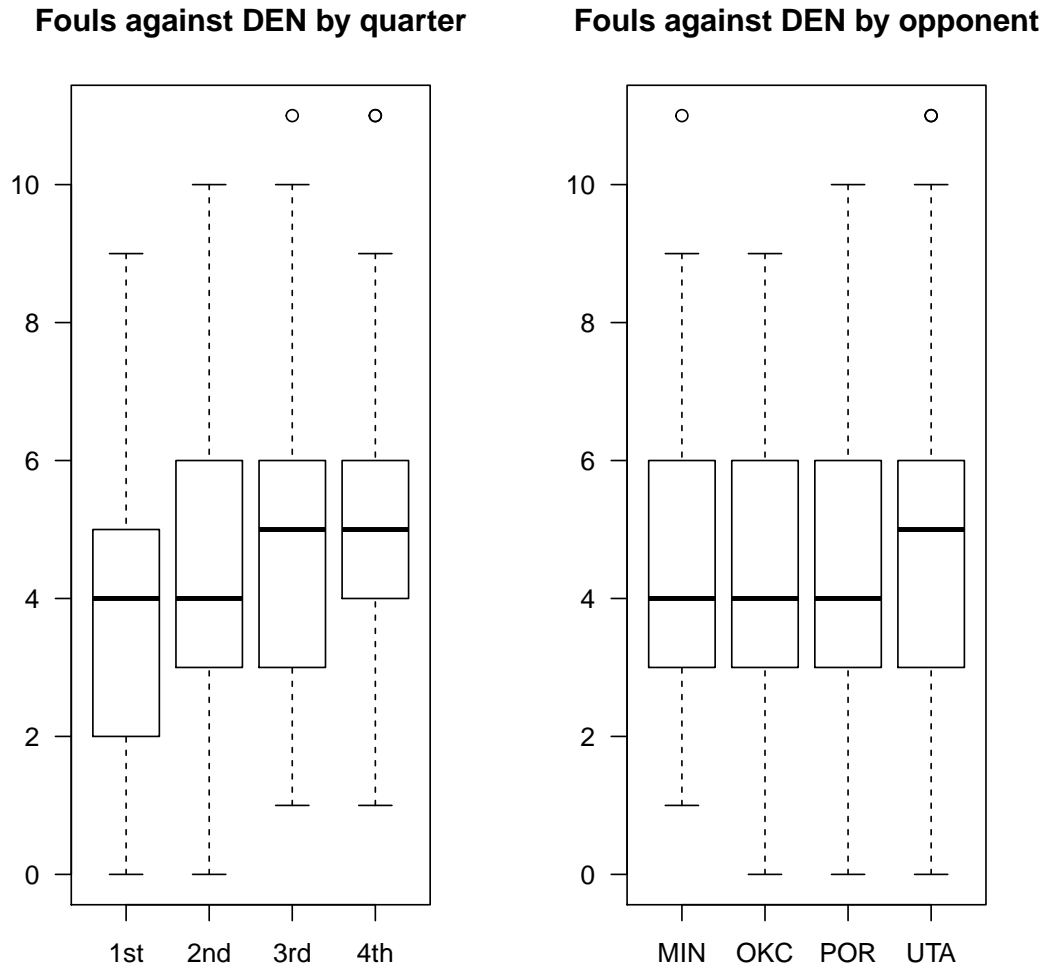
Running two ANOVA tests on this new data has interesting results. As with last time, we run one ANOVA examining the fouls as explained by quarter, and one ANOVA examining the fouls as explained by opposing team. Unlike last time, we can see, in Table 11 and Table 12, that there is a significant difference for both. Analyzing using Tukey's we get the results seen below.

Table 11:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
opp_name	3	13.38	4.46	1.20	0.3095
Residuals	496	1844.47	3.72		

Table 12:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
quarter	3	90.95	30.32	8.51	0.0000
Residuals	496	1766.90	3.56		



Running Tukey's on this yields the following results.

3rd 2nd 4th 1st

Table 13:

	term	comparison	estimate	conf.low	conf.high	adj.p.value
1	quarter	2nd-1st	0.62	0.09	1.14	0.01
2	quarter	3rd-1st	0.98	0.46	1.51	0.00
3	quarter	4th-1st	0.88	0.35	1.40	0.00
4	quarter	3rd-2nd	0.36	-0.16	0.89	0.28
5	quarter	4th-2nd	0.26	-0.26	0.79	0.58
6	quarter	4th-3rd	-0.10	-0.63	0.42	0.96

Table 14:

	term	comparison	estimate	conf.low	conf.high	adj.p.value
1	opp_name	OKC-MIN	-0.04	-0.61	0.53	1.00
2	opp_name	POR-MIN	-0.06	-0.57	0.45	0.99
3	opp_name	UTA-MIN	0.52	0.01	1.03	0.04
4	opp_name	POR-OKC	-0.02	-0.59	0.55	1.00
5	opp_name	UTA-OKC	0.56	-0.01	1.13	0.06
6	opp_name	UTA-POR	0.58	0.07	1.09	0.02

OKC POR MIN UTA

These results appear to mirror the results from the earlier model. For quarters, the first quarter is significantly lower than the others. This means that the Denver Nuggets are fouled about the same amount for quarters 2, 3, and 4, but fouled significantly less during quarter 1. And for opponent there appears to be no significant difference among teams.

Having this knowledge it became apparent that a model was needed to explain this phenomena. Since we already have a Poisson regression model used for fouls DEN commits, it only makes sense to use a Poisson model for fouls DEN draws. The model meets all the assumptions of a Poisson distribution in the same way that the first model did. The summary output of the model can be read below in Table 15.

Table 15:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3355	0.0581	22.98	0.0000
quarter2nd	0.1624	0.0620	2.62	0.0088
quarter3rd	0.2437	0.0608	4.01	0.0001
quarter4th	0.2355	0.0609	3.86	0.0001
opp_nameOklahoma City Thunder	-0.0128	0.0597	-0.21	0.8305
opp_namePortland Trail Blazers	-0.0073	0.0596	-0.12	0.9021
opp_nameUtah Jazz	0.0744	0.0584	1.27	0.2029

As expected from the finding using the combination of Tukey's and the prior model, quarter

was a significant factor at every level, and opponent was only significant for the Utah Jazz (outside of the base model). The model suggests that the Denver Nuggets draw more fouls as the game goes on, something which was also suggested by the model analyzing their foul committing tendency. Deviating from prior expectations, the model suggests that the Nuggets draw the most fouls against UTA, followed by MIN, then POR, and lastly OKC. This is interesting in that it appears the Nuggets both draw and commit the a high amount of fouls against the Jazz while drawing and committing a low amount of fouls against the Trail Blazers. See Table 16 below for a comparison of the transformed coefficients.

It should be noted that for the Poisson regression model for fouls DEN draws, the McFadden and Cox-Snell Pseudo- R^2 values are 0.0118 and 0.0478 respectively. These values are very low, but still better than the values received for the previous model examining fouls committed.

Table 16:

	Committed	Drawn
(Intercept)	3.6356743	3.8020558
2nd Quarter	1.1717391	1.1763485
3rd Quarter	1.1652174	1.2759336
4th Quarter	1.3217391	1.2655602
Oklahoma City Thunder	1.0760624	0.9872991
Portland Trail Blazers	0.9541721	0.9926942
Utah Jazz	1.0189263	1.0772170

5 Strengths and Weaknesses of my Approach

My project, ultimately, failed to reach many of the goals I set out to accomplish. The final models used to explain the response variable of interest does a very poor job of it. As a result the conclusions I come to in my paper are not very applicable to helping the Denver Nuggets gain any advantage over their local competition.

However, what my approach did do right is find solid theoretical grounding for the models used, and used sounds analysis to come to sound, if insignificant, conclusions. My process has a lot of mathematical footing to stand on should it be scrutinized.

The only thing I really think this project needs to have fixed is the final model chosen to explain the data. While a Poisson process is observable from the data, a straight Poisson regression does not seem to be the most appropriate for this data. A Cart model would have been a more likely fit for the data and would do a, hopefully, better job of explaining the phenomenon we looked at here. Additionally, the sheer amount of time data collection took, including compiling time for large set after collection, created a lot of tedious and frustrating setbacks. A more complete and easily accessible data set would have been preferable.

6 Further Work

Given the time, this would be only the first step of a much longer project. I would like to see my work carried on wards to discover what percentage of responsibility fouls have on the outcomes of games.

Additionally, further work can be done applying the mode built here to different datasets. To create a robust model, we used a massive amount of data that spans several years, which gives conclusions that may not be applicable to real world issues. Taking this model here and applying it to much more focused and smaller samples would provide a lot of utility for teams assesing anything from staff changes to roster line ups to the influence of referee teams.

To create a truly useful study that could review the tendencies of this team, it would be necessary to review tapes of games and count the number of fouls called versus the number of fouls committed but not called. While the Nuggets have an average foul intensity, it may be possible that they are playing abnormally in regards to percentage of fouls caught. I believe this would be a much more fruitful area of focus in determining which alterations to make to team behavior, but the data needed is very vast and beyond my grasp at this point.

7 Appendix

```
##          date          v_name v_points
## 1 Tue Oct 28 2008 Cleveland Cavaliers      85
## 2 Tue Oct 28 2008 Cleveland Cavaliers      85
## 3 Tue Oct 28 2008 Cleveland Cavaliers      85
## 4 Tue Oct 28 2008 Cleveland Cavaliers      85
## 5 Tue Oct 28 2008 Cleveland Cavaliers      85
## 6 Tue Oct 28 2008 Cleveland Cavaliers      85
##          h_name h_points  ovt          foul_team
## 1 Boston Celtics      90 <NA> Boston Celtics
## 2 Boston Celtics      90 <NA> Boston Celtics
## 3 Boston Celtics      90 <NA> Cleveland Cavaliers
## 4 Boston Celtics      90 <NA> Boston Celtics
## 5 Boston Celtics      90 <NA> Cleveland Cavaliers
## 6 Boston Celtics      90 <NA> Cleveland Cavaliers
##          foul_type          foul_by          foul_drawn
## 1      shooting      Rajon Rondo      Delonte West
## 2      shooting      Kendrick Perkins Zydrunas Ilgauskas
## 3      offensive      Rajon Rondo      Mo Williams
## 4      shooting      Eddie House Zydrunas Ilgauskas
## 5 personal foul      LeBron James      Paul Pierce
## 6      shooting Zydrunas Ilgauskas      Paul Pierce
##          foul_quarter foul_score foul_time foul_score_diff
## 1 1st quarter score: 0-0      11:40      0
## 2 1st quarter score: 11-4      07:39      7
## 3 1st quarter score: 12-4      06:38      8
```

7. APPENDIX

```
## 4 1st quarter score: 14-6      05:45      8
## 5 1st quarter score: 16-6      05:32     10
## 6 1st quarter score: 16-8      04:57      8
##      mean      sd foul_time_dec v_fc v_first_q v_second_q
## 1 1.160714 5.739106    11.066667  30      7      4
## 2 1.160714 5.739106     7.650000  30      7      4
## 3 1.160714 5.739106     6.633333  30      7      4
## 4 1.160714 5.739106     5.750000  30      7      4
## 5 1.160714 5.739106     5.533333  30      7      4
## 6 1.160714 5.739106     4.950000  30      7      4
## v_third_q v_fourth_q v_2m h_fc h_first_q h_second_q
## 1      7      6      2    26      3      5
## 2      7      6      2    26      3      5
## 3      7      6      2    26      3      5
## 4      7      6      2    26      3      5
## 5      7      6      2    26      3      5
## 6      7      6      2    26      3      5
## h_third_q h_fourth_q h_2m winner_hv v_addF h_addF v_addP
## 1      2      6      3      h      0      1      0
## 2      2      6      3      h     NA      2     NA
## 3      2      6      3      h      1      0      1
## 4      2      6      3      h     NA      3     NA
## 5      2      6      3      h      2     NA      2
## 6      2      6      3      h      3     NA      3
## h_addP foul_time_dec_diff
## 1      1      3.4166667
## 2      2      1.0166667
## 3      0      0.8833333
## 4      3      0.2166667
## 5     NA      0.5833333
## 6     NA      0.2666667
```

A1a

```
##      date      v_name v_points
## 595 Wed Oct 29 2008   Denver Nuggets      94
## 882 Fri Oct 31 2008   Denver Nuggets     113
## 1271 Sat Nov 1 2008 Los Angeles Lakers     104
## 2252 Wed Nov 5 2008   Denver Nuggets     101
## 3100 Fri Nov 7 2008   Dallas Mavericks     105
## 3774 Sun Nov 9 2008   Memphis Grizzlies      90
##      h_name h_points ovt      mean
## 595      Utah Jazz     98 <NA> -3.4893617
## 882 Los Angeles Clippers    103 OT -5.0925926
## 1271      Denver Nuggets     97 <NA>  0.2352941
## 2252 Golden State Warriors    111 <NA>  1.8400000
## 3100      Denver Nuggets    108 <NA> -1.0434783
```

7. APPENDIX

```
## 3774      Denver Nuggets      100 <NA> -2.3437500
##          sd winner_hv v_fc v_first_q v_second_q v_third_q
## 595  3.628824      h   25      6      5      2
## 882  6.760787      v   25      4      6      3
## 1271 3.507639      v   22      5      6      1
## 2252 6.421774      h   27      4      6      5
## 3100 3.279272      h   24      5      3      4
## 3774 5.626521      h   31      5      5      9
##          v_fourth_q h_fc h_first_q h_second_q h_third_q
## 595      6   22      7      2      6
## 882      7   29      3      5      5
## 1271     4   29      6      5      5
## 2252     4   23      3      7      5
## 3100     6   22      8      4      3
## 3774     8   33      5      6      7
##          h_fourth_q nugs_fc opp_fc      opp_name
## 595      4      25   22      Utah Jazz
## 882      2      25   29   Los Angeles Clippers
## 1271     4      29   22   Los Angeles Lakers
## 2252     3      27   23   Golden State Warriors
## 3100     5      22   24   Dallas Mavericks
## 3774     9      33   31   Memphis Grizzlies
##          nugs_first_q nugs_second_q nugs_third_q nugs_fourth_q
## 595      6      5      2      6
## 882      4      6      3      7
## 1271     6      5      5      4
## 2252     4      6      5      4
## 3100     8      4      3      5
## 3774     5      6      7      9
##          opp_first_q opp_second_q opp_third_q opp_fourth_q
## 595      7      2      6      4
## 882      3      5      5      2
## 1271     5      6      1      4
## 2252     3      7      5      3
## 3100     5      3      4      6
## 3774     5      5      9      8
```

A1b

```
# df of lambdas for nugs vs nw div teams NBA

nba_f <- sum(c(df_nugs$nugs_first_q, df_nugs$nugs_second_q,
  df_nugs$nugs_third_q,
  df_nugs$nugs_fourth_q))/(length(df_nugs$nugs_fc) * 8.5 *
  4)
```



```
nba_1 <- sum(df_nugs$nugs_first_q)/(length(df_nugs$nugs_first_q)
*
8.5)

nba_2 <- sum(df_nugs$nugs_second_q)/(length(df_nugs$nugs_second_q
) *
8.5)

nba_3 <- sum(df_nugs$nugs_third_q)/(length(df_nugs$nugs_third_q)
*
8.5)

nba_4 <- sum(df_nugs$nugs_fourth_q)/(length(df_nugs$nugs_fourth_q
) *
8.5)

# POR

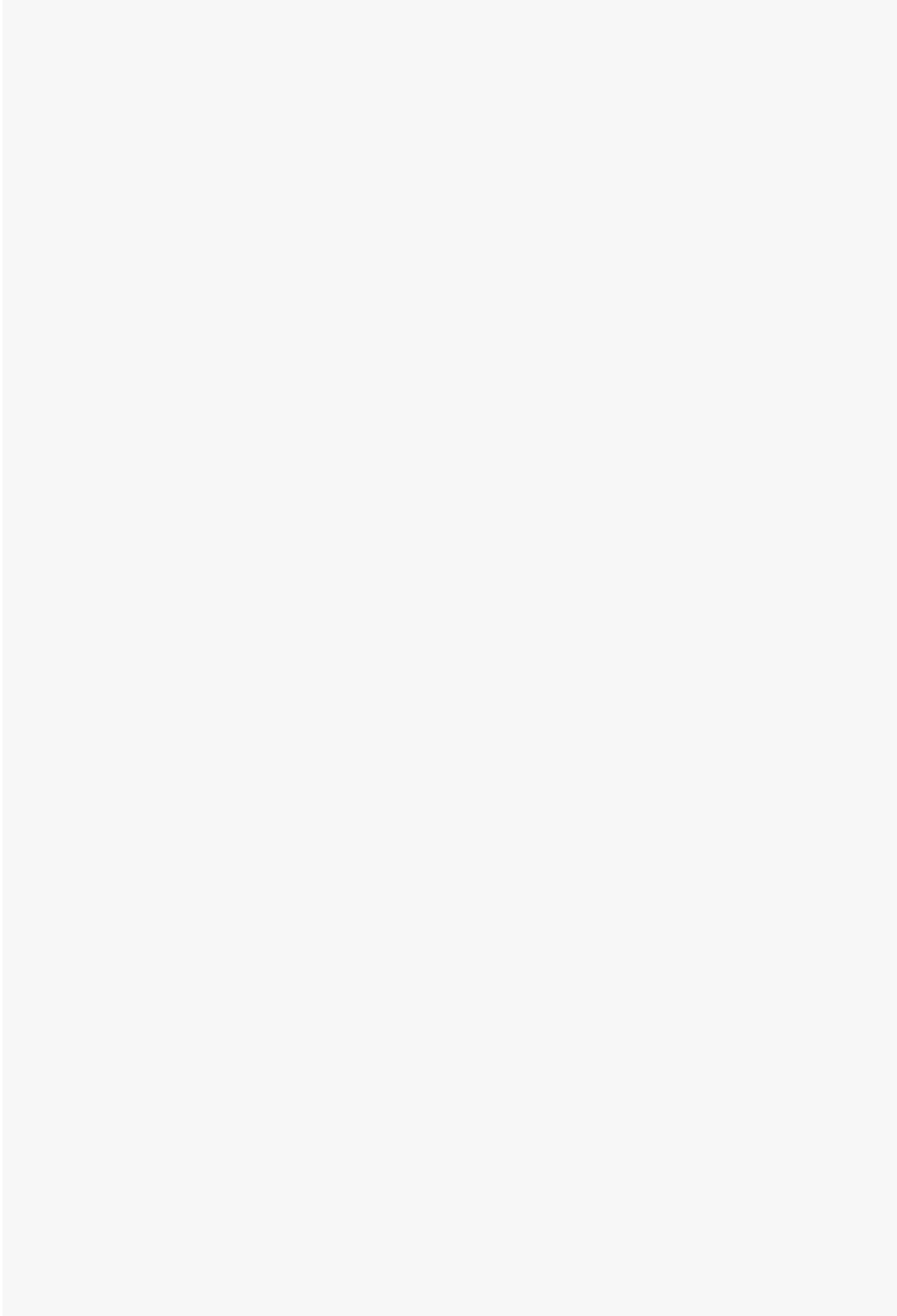
por_f <- sum(c(df_nugs$nugs_first_q[df_nugs$opp_name == nw_div
[1]],
df_nugs$nugs_second_q[df_nugs$opp_name == nw_div[1]],
df_nugs$nugs_third_q[df_nugs$opp_name ==
nw_div[1]], df_nugs$nugs_fourth_q[df_nugs$opp_name ==
nw_div[1]])/(length((df_nugs$nugs_fc[df_nugs$opp_name ==
nw_div[1]])) * 8.5 * 4)

por_1 <- sum(df_nugs$nugs_first_q[df_nugs$opp_name == nw_div[1]])
/(length((df_nugs$nugs_first_q[df_nugs$opp_name ==
nw_div[1]])) * 8.5)

por_2 <- sum(df_nugs$nugs_second_q[df_nugs$opp_name == nw_div
[1]]/(length((df_nugs$nugs_second_q[df_nugs$opp_name ==
nw_div[1]])) * 8.5)

por_3 <- sum(df_nugs$nugs_third_q[df_nugs$opp_name == nw_div[1]])
/(length((df_nugs$nugs_third_q[df_nugs$opp_name ==
nw_div[1]])) * 8.5)

por_4 <- sum(df_nugs$nugs_fourth_q[df_nugs$opp_name == nw_div
[1]]/(length((df_nugs$nugs_fourth_q[df_nugs$opp_name ==
nw_div[1]])) * 8.5)
```



```
min_2 <- sum(df_nugs$nugs_second_q[df_nugs$opp_name == nw_div
[3]])/(length((df_nugs$nugs_second_q[df_nugs$opp_name ==
nw_div[3]])) * 8.5)

min_3 <- sum(df_nugs$nugs_third_q[df_nugs$opp_name == nw_div[3]])
/(length((df_nugs$nugs_third_q[df_nugs$opp_name ==
nw_div[3]])) * 8.5)

min_4 <- sum(df_nugs$nugs_fourth_q[df_nugs$opp_name == nw_div
[3]])/(length((df_nugs$nugs_fourth_q[df_nugs$opp_name ==
nw_div[3]])) * 8.5)

# UTA

uta_f <- sum(c(df_nugs$nugs_first_q[df_nugs$opp_name == nw_div
[4]],
df_nugs$nugs_second_q[df_nugs$opp_name == nw_div[4]],
df_nugs$nugs_third_q[df_nugs$opp_name ==
nw_div[4]], df_nugs$nugs_fourth_q[df_nugs$opp_name ==
nw_div[4]])/(length((df_nugs$nugs_fc[df_nugs$opp_name ==
nw_div[4]])) * 8.5 * 4)

uta_1 <- sum(df_nugs$nugs_first_q[df_nugs$opp_name == nw_div[4]])
/(length((df_nugs$nugs_first_q[df_nugs$opp_name ==
nw_div[4]])) * 8.5)

uta_2 <- sum(df_nugs$nugs_second_q[df_nugs$opp_name == nw_div
[4]])/(length((df_nugs$nugs_second_q[df_nugs$opp_name ==
nw_div[4]])) * 8.5)

uta_3 <- sum(df_nugs$nugs_third_q[df_nugs$opp_name == nw_div[4]])
/(length((df_nugs$nugs_third_q[df_nugs$opp_name ==
nw_div[4]])) * 8.5)

uta_4 <- sum(df_nugs$nugs_fourth_q[df_nugs$opp_name == nw_div
[4]])/(length((df_nugs$nugs_fourth_q[df_nugs$opp_name ==
nw_div[4]])) * 8.5)

df_lam_nugs <- data.frame(NBA = c(nba_f, nba_1, nba_2, nba_3,
nba_4), POR = c(por_f, por_1, por_2, por_3, por_4), OKC = c(
okc_f,
okc_1, okc_2, okc_3, okc_4), MIN = c(min_f, min_1, min_2,
min_3, min_4), UTA = c(uta_f, uta_1, uta_2, uta_3, uta_4))
```

A2a

```
get_lam <- function(team) {
  df_temp <- unique(df[c("date", "v_name", "v_points", "h_name",
    "h_points", "ovt", "mean", "sd", "winner_hv", "v_fc",
    "v_first_q", "v_second_q", "v_third_q", "v_fourth_q",
    "h_fc", "h_first_q", "h_second_q", "h_third_q", "h_fourth_q")])

  for (i in 1:1) {

    df_temp$team_fc <- NA
    df_temp$opp_fc <- NA
    df_temp$opp_name <- NA
    df_temp$team_first_q <- NA
    df_temp$team_second_q <- NA
    df_temp$team_third_q <- NA
    df_temp$team_fourth_q <- NA
    df_temp$opp_first_q <- NA
    df_temp$opp_second_q <- NA
    df_temp$opp_third_q <- NA
    df_temp$opp_fourth_q <- NA

    df_temp$team_fc[which(df_temp$h_name == team)] <-
      df_temp$h_fc[df_temp$h_name ==
        team]
    df_temp$team_fc[which(df_temp$v_name == team)] <-
      df_temp$v_fc[df_temp$v_name ==
        team]
    df_temp$opp_fc[which(df_temp$h_name == team)] <-
      df_temp$v_fc[df_temp$h_name ==
        team]
    df_temp$opp_fc[which(df_temp$v_name == team)] <-
      df_temp$h_fc[df_temp$v_name ==
        team]
    df_temp$opp_name[which(df_temp$h_name == team)] <-
      df_temp$v_name[df_temp$h_name ==
        team]
    df_temp$opp_name[which(df_temp$v_name == team)] <-
      df_temp$h_name[df_temp$v_name ==
        team]

    df_temp$team_first_q[which(df_temp$h_name == team)] <-
      df_temp$h_first_q[df_temp$h_name ==
        team]
    df_temp$team_first_q[which(df_temp$v_name == team)] <-
      df_temp$v_first_q[df_temp$v_name ==
```

```
      team]
df_temp$team_second_q[which(df_temp$h_name == team)] <-
  df_temp$h_second_q[df_temp$h_name ==
    team]
df_temp$team_second_q[which(df_temp$v_name == team)] <-
  df_temp$v_second_q[df_temp$v_name ==
    team]
df_temp$team_third_q[which(df_temp$h_name == team)] <-
  df_temp$h_third_q[df_temp$h_name ==
    team]
df_temp$team_third_q[which(df_temp$v_name == team)] <-
  df_temp$v_third_q[df_temp$v_name ==
    team]
df_temp$team_fourth_q[which(df_temp$h_name == team)] <-
  df_temp$h_fourth_q[df_temp$h_name ==
    team]
df_temp$team_fourth_q[which(df_temp$v_name == team)] <-
  df_temp$v_fourth_q[df_temp$v_name ==
    team]

df_temp$opp_first_q[which(df_temp$h_name == team)] <-
  df_temp$v_first_q[df_temp$h_name ==
    team]
df_temp$opp_first_q[which(df_temp$v_name == team)] <-
  df_temp$h_first_q[df_temp$v_name ==
    team]
df_temp$opp_second_q[which(df_temp$h_name == team)] <-
  df_temp$v_second_q[df_temp$h_name ==
    team]
df_temp$opp_second_q[which(df_temp$v_name == team)] <-
  df_temp$h_second_q[df_temp$v_name ==
    team]
df_temp$opp_third_q[which(df_temp$h_name == team)] <-
  df_temp$v_third_q[df_temp$h_name ==
    team]
df_temp$opp_third_q[which(df_temp$v_name == team)] <-
  df_temp$h_third_q[df_temp$v_name ==
    team]
df_temp$opp_fourth_q[which(df_temp$h_name == team)] <-
  df_temp$v_fourth_q[df_temp$h_name ==
    team]
df_temp$opp_fourth_q[which(df_temp$v_name == team)] <-
  df_temp$h_fourth_q[df_temp$v_name ==
    team]

}

df_temp <- df_temp[which(df_temp$opp_name == nw_div[1] |
```

```
df_temp$opp_name == nw_div[2] | df_temp$opp_name ==
  nw_div[3] |
df_temp$opp_name == nw_div[4] | df_temp$opp_name ==
  nw_div2[5]),
]

# NBA
t_nba_f <- sum(c(df_temp$team_first_q, df_temp$team_second_q,
  df_temp$team_third_q, df_temp$team_fourth_q))/(length(
  df_temp$team_fc) *
  8.5 * 4)
t_nba_1 <- sum(df_temp$team_first_q)/(length(
  df_temp$team_first_q) *
  8.5)
t_nba_2 <- sum(df_temp$team_second_q)/(length(
  df_temp$team_second_q) *
  8.5)
t_nba_3 <- sum(df_temp$team_third_q)/(length(
  df_temp$team_third_q) *
  8.5)
t_nba_4 <- sum(df_temp$team_fourth_q)/(length(
  df_temp$team_fourth_q) *
  8.5)

# POR
t_por_f <- sum(c(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[1]], df_temp$team_second_q[df_temp$opp_name ==
  nw_div[1]], df_temp$team_third_q[df_temp$opp_name ==
  nw_div[1]], df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[1]]))/(length((df_temp$team_fc[df_temp$opp_name ==
  nw_div[1]])) * 8.5 * 4)
t_por_1 <- sum(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[1]])/(length((df_temp$team_first_q[df_temp$opp_name
  ==
  nw_div[1]])) * 8.5)
t_por_2 <- sum(df_temp$team_second_q[df_temp$opp_name ==
  nw_div[1]])/(length((df_temp$team_second_q[
  df_temp$opp_name ==
  nw_div[1]])) * 8.5)
t_por_3 <- sum(df_temp$team_third_q[df_temp$opp_name ==
  nw_div[1]])/(length((df_temp$team_third_q[df_temp$opp_name
  ==
  nw_div[1]])) * 8.5)
t_por_4 <- sum(df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[1]])/(length((df_temp$team_fourth_q[
  df_temp$opp_name ==
  nw_div[1]])) * 8.5)
```

```
# OKC
t_okc_f <- sum(c(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[2]], df_temp$team_second_q[df_temp$opp_name ==
  nw_div[2]], df_temp$team_third_q[df_temp$opp_name ==
  nw_div[2]], df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[2]]))/(length((df_temp$team_fc[df_temp$opp_name ==
  nw_div[2]])) * 8.5 * 4)
t_okc_1 <- sum(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[2]])/(length((df_temp$team_first_q[df_temp$opp_name ==
  ==
  nw_div[2]])) * 8.5)
t_okc_2 <- sum(df_temp$team_second_q[df_temp$opp_name ==
  nw_div[2]])/(length((df_temp$team_second_q[
  df_temp$opp_name ==
  nw_div[2]])) * 8.5)
t_okc_3 <- sum(df_temp$team_third_q[df_temp$opp_name ==
  nw_div[2]])/(length((df_temp$team_third_q[df_temp$opp_name ==
  ==
  nw_div[2]])) * 8.5)
t_okc_4 <- sum(df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[2]])/(length((df_temp$team_fourth_q[
  df_temp$opp_name ==
  nw_div[2]])) * 8.5)

# MIN
t_min_f <- sum(c(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[3]], df_temp$team_second_q[df_temp$opp_name ==
  nw_div[3]], df_temp$team_third_q[df_temp$opp_name ==
  nw_div[3]], df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[3]]))/(length((df_temp$team_fc[df_temp$opp_name ==
  nw_div[3]])) * 8.5 * 4)
t_min_1 <- sum(df_temp$team_first_q[df_temp$opp_name ==
  nw_div[3]])/(length((df_temp$team_first_q[df_temp$opp_name ==
  ==
  nw_div[3]])) * 8.5)
t_min_2 <- sum(df_temp$team_second_q[df_temp$opp_name ==
  nw_div[3]])/(length((df_temp$team_second_q[
  df_temp$opp_name ==
  nw_div[3]])) * 8.5)
t_min_3 <- sum(df_temp$team_third_q[df_temp$opp_name ==
  nw_div[3]])/(length((df_temp$team_third_q[df_temp$opp_name ==
  ==
  nw_div[3]])) * 8.5)
t_min_4 <- sum(df_temp$team_fourth_q[df_temp$opp_name ==
  nw_div[3]])/(length((df_temp$team_fourth_q[
  df_temp$opp_name ==
```

```
nw_div[3])) * 8.5)

# UTA
t_uta_f <- sum(c(df_temp$team_first_q[df_temp$opp_name ==
nw_div[4]], df_temp$team_second_q[df_temp$opp_name ==
nw_div[4]], df_temp$team_third_q[df_temp$opp_name ==
nw_div[4]], df_temp$team_fourth_q[df_temp$opp_name ==
nw_div[4]]))/(length((df_temp$team_fc[df_temp$opp_name ==
nw_div[4])) * 8.5 * 4)
t_uta_1 <- sum(df_temp$team_first_q[df_temp$opp_name ==
nw_div[4]]/(length((df_temp$team_first_q[df_temp$opp_name
==
nw_div[4])) * 8.5)
t_uta_2 <- sum(df_temp$team_second_q[df_temp$opp_name ==
nw_div[4]]/(length((df_temp$team_second_q[
df_temp$opp_name ==
nw_div[4])) * 8.5)
t_uta_3 <- sum(df_temp$team_third_q[df_temp$opp_name ==
nw_div[4]]/(length((df_temp$team_third_q[df_temp$opp_name
==
nw_div[4])) * 8.5)
t_uta_4 <- sum(df_temp$team_fourth_q[df_temp$opp_name ==
nw_div[4]]/(length((df_temp$team_fourth_q[
df_temp$opp_name ==
nw_div[4])) * 8.5)

# DEN
t_den_f <- sum(c(df_temp$team_first_q[df_temp$opp_name ==
nw_div[5]], df_temp$team_second_q[df_temp$opp_name ==
nw_div[5]], df_temp$team_third_q[df_temp$opp_name ==
nw_div[5]], df_temp$team_fourth_q[df_temp$opp_name ==
nw_div[5]]))/(length((df_temp$team_fc[df_temp$opp_name ==
nw_div2[5])) * 8.5 * 4)
t_den_1 <- sum(df_temp$team_first_q[df_temp$opp_name ==
nw_div2[5]]/(length((df_temp$team_first_q[
df_temp$opp_name ==
nw_div2[5])) * 8.5)
t_den_2 <- sum(df_temp$team_second_q[df_temp$opp_name ==
nw_div2[5]]/(length((df_temp$team_second_q[
df_temp$opp_name ==
nw_div2[5])) * 8.5)
t_den_3 <- sum(df_temp$team_third_q[df_temp$opp_name ==
nw_div2[5]]/(length((df_temp$team_third_q[
df_temp$opp_name ==
nw_div2[5])) * 8.5)
t_den_4 <- sum(df_temp$team_fourth_q[df_temp$opp_name ==
nw_div2[5]]/(length((df_temp$team_fourth_q[
```



```
df_temp$opp_name ==
nw_div2[5])) * 8.5)

df_temp <- data.frame(NBA = c(t_nba_f, t_nba_1, t_nba_2,
t_nba_3, t_nba_4), POR = c(t_por_f, t_por_1, t_por_2,
t_por_3, t_por_4), OKC = c(t_okc_f, t_okc_1, t_okc_2,
t_okc_3, t_okc_4), MIN = c(t_min_f, min_1, min_2, min_3,
min_4), UTA = c(t_uta_f, t_uta_1, t_uta_2, t_uta_3,
t_uta_4),
DEN = c(t_den_f, t_den_1, t_den_2, t_den_3, t_den_4))

assign(paste0("df_lam_", tolower(substr(team, 1, 3))),
df_temp,
envir = .GlobalEnv)
}
```

A2b

```
anova(lm(fouls ~ quarter, data = df_model))

## Analysis of Variance Table
##
## Response: fouls
##          Df Sum Sq Mean Sq F value    Pr(>F)
## quarter    3   87.75  29.2500    8.7274 1.19e-05 ***
## Residuals 496 1662.35   3.3515
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(aov(lm(fouls ~ quarter, data = df_model)))

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm(fouls ~ quarter, data = df_model))
##
## $quarter
##          diff          lwr          upr      p adj
## 2nd-1st  0.632  0.03507007  1.2289299  0.0331791
## 3rd-1st  0.608  0.01107007  1.2049299  0.0440462
## 4th-1st  1.184  0.58707007  1.7809299  0.0000027
## 3rd-2nd -0.024 -0.62092993  0.5729299  0.9996012
## 4th-2nd  0.552 -0.04492993  1.1489299  0.0815287
## 4th-3rd  0.576 -0.02092993  1.1729299  0.0631135
```

7. APPENDIX

A3

```
glm(fouls ~ opp_name + quarter, data = df_model, family = poisson
)

##
## Call:  glm(formula = fouls ~ opp_name + quarter, family =
##        poisson, data = df_model)
##
## Coefficients:
##              (Intercept)
##                1.29079
##   opp_nameOklahoma City Thunder
##                0.07331
## opp_namePortland Trail Blazers
##               -0.04691
##           opp_nameUtah Jazz
##                0.01875
##               quarter2nd
##                0.15849
##               quarter3rd
##                0.15291
##               quarter4th
##                0.27895
##
## Degrees of Freedom: 499 Total (i.e. Null);  493 Residual
## Null Deviance:                436.6
## Residual Deviance: 412.1      AIC: 2033
```

A4a

```
glm(fouls ~ opp_name * quarter, data = df_model, family = poisson
)

##
## Call:  glm(formula = fouls ~ opp_name * quarter, family =
##        poisson, data = df_model)
##
## Coefficients:
##              (Intercept)
##                1.33829
##           opp_nameOklahoma City Thunder
##               -0.02734
##   opp_namePortland Trail Blazers
##               -0.07179
##           opp_nameUtah Jazz
```

```
## -0.04488
## quarter2nd
## 0.04801
## quarter3rd
## 0.08633
## quarter4th
## 0.26488
## opp_nameOklahoma City Thunder:quarter2nd
## 0.21100
## opp_namePortland Trail Blazers:quarter2nd
## 0.08779
## opp_nameUtah Jazz:quarter2nd
## 0.13726
## opp_nameOklahoma City Thunder:quarter3rd
## 0.14547
## opp_namePortland Trail Blazers:quarter3rd
## 0.05738
## opp_nameUtah Jazz:quarter3rd
## 0.06148
## opp_nameOklahoma City Thunder:quarter4th
## 0.04004
## opp_namePortland Trail Blazers:quarter4th
## -0.03811
## opp_nameUtah Jazz:quarter4th
## 0.05115
##
## Degrees of Freedom: 499 Total (i.e. Null); 484 Residual
## Null Deviance: 436.6
## Residual Deviance: 409.8 AIC: 2049
```

A4b

A5 McFadden:

$$R^2 = 1 - \frac{\ln - \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})}$$

Where M_{Full} is the model with predictors and $M_{Intercept}$ is the model without.

\hat{L} is the estimated likelihood.

$$\text{Cox-Snell: } R^2 = 1 - \left(\frac{\ln - \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} \right)^{2/N}$$

A6 The poisson regression model is defined as follows:

$$\ln(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1}$$

By raising e to both sides of the equation we are left with:

$$\lambda = e^{\beta_0} + e^{\beta_1 X_1} + \dots + e^{\beta_{p-1} X_{p-1}}$$

which gives us a λ that is representative of the number of fouls recived under the conditions set by the model.

Because this is a qualitative model, all β values are set to either 0 or 1, depending on the factors in play, with only one of each category (team, quarter) being able to hold a value

7. APPENDIX

of 1 at any time. This implies that when all are set to 0 that the model is measuring the not-included factors, which is represented by the intercept.

```
#POR

glm(fouls ~ opp_name + quarter, data = POR_model, family =
  poisson)

##
## Call:  glm(formula = fouls ~ opp_name + quarter, family =
  poisson, data = POR_model)
##
## Coefficients:
##              (Intercept)
##                  1.31437
## opp_nameMinnesota Timberwolves
##                  -0.06938
##  opp_nameOklahoma City Thunder
##                  -0.05558
##           opp_nameUtah Jazz
##                  -0.14276
##               quarter2nd
##                   0.17317
##               quarter3rd
##                   0.20892
##               quarter4th
##                   0.30909
##
## Degrees of Freedom: 503 Total (i.e. Null);  497 Residual
## Null Deviance:          470.1
## Residual Deviance: 439.5      AIC: 2055

#OKC

glm(fouls ~ opp_name + quarter, data = OKL_model, family =
  poisson)

##
## Call:  glm(formula = fouls ~ opp_name + quarter, family =
  poisson, data = OKL_model)
##
## Coefficients:
##              (Intercept)
##                  1.32018
## opp_nameMinnesota Timberwolves
##                  -0.06199
## opp_namePortland Trail Blazers
##                  -0.20862
```

```
##           opp_nameUtah Jazz
##           -0.11369
##           quarter2nd
##           0.11517
##           quarter3rd
##           0.18389
##           quarter4th
##           0.34082
##
## Degrees of Freedom: 499 Total (i.e. Null);  493 Residual
## Null Deviance:          468.4
## Residual Deviance: 425.6      AIC: 2010

#MIN

glm(fouls ~ opp_name + quarter, data = MIN_model, family =
  poisson)

##
## Call:  glm(formula = fouls ~ opp_name + quarter, family =
  poisson, data = MIN_model)
##
## Coefficients:
##           (Intercept)
##           1.37749
##   opp_nameOklahoma City Thunder
##           -0.13262
##   opp_namePortland Trail Blazers
##           -0.07866
##           opp_nameUtah Jazz
##           -0.11229
##           quarter2nd
##           0.13843
##           quarter3rd
##           0.16655
##           quarter4th
##           0.17758
##
## Degrees of Freedom: 499 Total (i.e. Null);  493 Residual
## Null Deviance:          502.5
## Residual Deviance: 487.1      AIC: 2078

#UTA

glm(fouls ~ opp_name + quarter, data = UTA_model, family =
  poisson)

##
```

```
## Call: glm(formula = fouls ~ opp_name + quarter, family =
  poisson, data = UTA_model)
##
## Coefficients:
##                (Intercept)
##                1.35461
## opp_nameMinnesota Timberwolves
##                -0.17666
## opp_nameOklahoma City Thunder
##                -0.06376
## opp_namePortland Trail Blazers
##                -0.22043
##                quarter2nd
##                0.24281
##                quarter3rd
##                0.27840
##                quarter4th
##                0.32952
##
## Degrees of Freedom: 499 Total (i.e. Null);  493 Residual
## Null Deviance:          507.5
## Residual Deviance: 459      AIC: 2073
```

A7

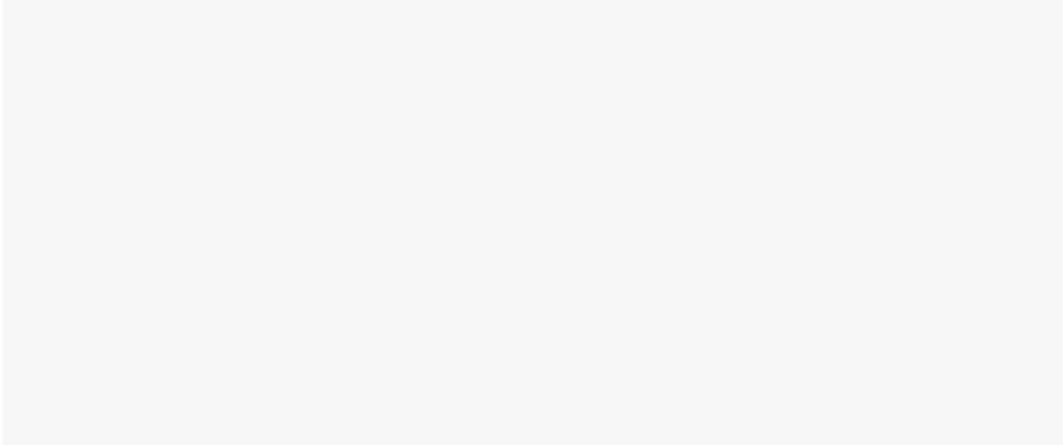
```
anova(lm(fouls ~ quarter, data = df_model))

## Analysis of Variance Table
##
## Response: fouls
##          Df Sum Sq Mean Sq F value    Pr(>F)
## quarter    3   87.75  29.2500   8.7274 1.19e-05 ***
## Residuals 496 1662.35   3.3515
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(lm(fouls ~ opp_name, data = df_model))

## Analysis of Variance Table
##
## Response: fouls
##          Df Sum Sq Mean Sq F value    Pr(>F)
## opp_name    3   17.0   5.6654   1.6214 0.1835
## Residuals 496 1733.1   3.4942

TukeyHSD(aov(lm(fouls ~ quarter, data = df_model)))
```



A8 8 References

Poisson distribution - Maximum Likelihood Estimation. Normal distribution - Maximum likelihood estimation. <https://www.statlect.com/fundamentals-of-statistics/Poisson-distribution-maximum-likelihood>. Accessed 3 March 2018

Ross, S. M. (1993). Introduction to Probability Models. San Diego, US.: Harcourt/Academic Press.

Weisberg, S. (2013). Applied linear regression. Retrieved from <https://ebookcentral.proquest.com>

9 Additional Links

<https://www.basketball-reference.com>

<https://github.com/92swhite/NBA-Project>