The George Washington University

Department of Statistics

Professor Kim

A Study of NFL Kickers

5/2/2019

By Eloka Obi

Abstract

The object of our study is to determine whether the experience level and performance of 209 National Football League (NFL) kickers plays a role in their salary. This is a topic that has been heavily debated in the National Football League. There is a lot of speculation about what factors go into determining different NFL players' salaries. A particular position, the kicker, has a very different job from the rest of the team. A kicker's responsibility is to score a field goal. For context, a field goal is an event in American Football where the offensive team, or the team in possession of the ball, attempts to kick the ball through a set of tall upright poles. If the team is successful in doing this, the offensive team is then awarded three points. Because a kicker's ability and position is very distinct, the measurements of factors that contribute to their salary cannot be determined as easily as other players, such as linebackers or a quarterbacks. It is for this reason that our study helps explain what goes into determining if a kicker's salary is higher or lower. To determine this, we tested two factors: the kicker's performance level (measured by their field goal completion percentage) and their experience level in the National Football League (measured by years active). We know that these are two big factors in a kicker's career, so we assume that they play a role in determining the kickers' salary. The study is conducted and tested using statistics that are provided by the NFL from their statistics page. We hypothesized that there would be a correlation between more years in the league along with better performance and a higher salary. The final determination after the evaluation is that while there is no interaction between the two factors, they are independently significant towards a kicker's determined salary, in particular the experience having more weight than the success rate of the field goal. It would be helpful, for a more precise response towards what determines a kicker's salary, to test additional factors, such as ball kicking distance or financial state of the team as a whole.

Introduction

Our group decided to conduct a statistical study analyzing kickers in the National Football League and the different variables that affect their individual salaries. With public information available from the National Football League's website, we collected the following data on two hundred nine active and inactive professional kickers; salary, years of professional experience, and the kickers' field goal completion percentage. To calculate a kickers completion percentage, you take their successful field goals divided by their attempted field goals. The data was presented in a more convoluted way, but we organized the data in sas for our study. Our study is based on a combination of an experiment and an observation and is a fixed factors model. Moving forward, we will have two factors, A and B, with factor A being the kickers' performance and B being the number of years experience. Both factors were divided into three

levels, using a 95% confidence surrounding the mean of each factor to determine what was high, average and low. For experience level, we used basic football terminology: Rookie, Intermediate, and Veteran.

Study Overview

Our study is to determine how two fixed factors (A) kicker performance level and (B) experience level correlate with the salary of the sample of NFL kickers. Additionally, we will be testing for interaction between our factors. Breaking down our factors further, we segmented years experience into three groups, Rookie, Intermediate, and Veteran. Rookies are identified as any kickers with zero to five years experience, Intermediates are individuals with six to ten years experience, and Veterans are players with eleven or more years of experience. Field goal performance is quantified as a percentage, a continuous value between zero and one hundred, meaning that experience is a qualitative variable which was broken down into levels of High, Average, and Low based on a 95% confidence interval of the mean percentage values.

The model we have chosen to utilize for this study is a Factor Affects ANOVA model with both effects fixed. The model is shown below with the corresponding hypotheses.

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + ijk$$

Main effects:	General Model:
$H_0: \alpha_j = 0$ $H_a: \alpha_j \neq 0$	$H_0: \alpha_j = \beta_k = \alpha_j\beta_k = 0$
$H_0: \beta_k = 0$ $H_a: \beta_k \neq 0$	H_a: at least one factor does not equal zero

In our study we tested two separate dependent variables corresponding with salary: salary quartiles and salaries as explicit numerical values in dollars. We chose to analyze the salary quartiles in addition to the numerical values as quartiles allow us to further segment salaries in levels, painting a more clear picture of kickers' salary ranges.

Our group's hypothesis is that that National Football League kickers' salaries will increase in correlation with more years spent in the National Football League as well as better performance levels. In short, the longer and more accomplished a National Football League kickers' career is, the higher their yearly salary will be.

Methods

Full Model

As was said in the overview, a fixed factors effect with both effects fixed is the model used for our data. **μ**.. is used as our constant. ***α***i are constants subject to the restrictions$\Sigma\boldsymbol{\alpha}i = 0$. **β**j are constants subject to the restriction $\Sigma\boldsymbol{\beta}j = 0$. (***αβ***)ij are constants subject to the restrictions:

$\Sigma i(\boldsymbol{\alpha\beta})ij = 0$ j= 1,...b
$\Sigma j(\boldsymbol{\alpha\beta})ij = 0$ i= 1,....a

In testing our hypothesis a F-test was used to see if we should reject or fail to reject our null hypothesis of whether salary will increase in correlation with more years spent in National Football League as well as higher field goal completion percentages. We used our p-value to make sure whether we should reject or fail to reject a null hypothesis as well. We used a significance level of ***α***=0.05 for both the F-statistic and our P-value. The same test was done when testing if our interaction term was significant.

Our categorical dependent variable quartile shows the levels of salary from 1-4. 1 is the lowest while 4 is the highest. We delineated this model again based on experience and performance. More experience and better performance will lead to a higher quartile but this may not exactly be the case with both of our models showing a low R^2 for both of the quartiles. A F-test was conducted for both quartiles in regards to salary and experience with a significance level of ***α***=0.05. Our p-value also was used to confirm our null hypothesis. Whether we should reject the null hypothesis or fail to reject the null hypothesis at the ***α***=0.05 level. From doing this test we can better interpret our how are dependent variable interact with our independent variables. The dependent variable quartile, had a higher R^2 value and thus was a better fit for our predictions.

In terms of the collinearity of the variables, it is clear that this would be something to take into consideration; checking for collinearity essentially is checking for the correlation between the variables in question, and if this were to be high then this would tend to inflate the variance of at least one of the estimated regression coefficients. Therefore, we used our SAS output to quantify the level of collinearity that exists; the SAS output is shown below both for salary quartiles salary in dollars in Table 1 and Table 2, respectively.

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	188157	218288	0.86	0.3897	.	0
Completion_Level	1	126162	72656	1.74	0.0840	0.97917	1.02127
Experience_Level	1	521785	80189	6.51	<.0001	0.97917	1.02127

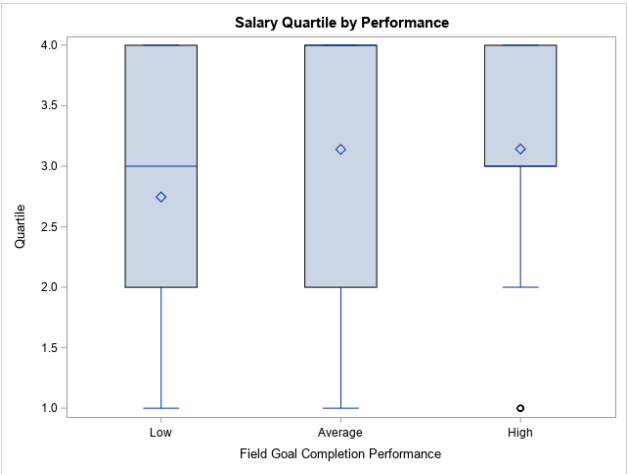
Table 1: Salary Quartile VIF values

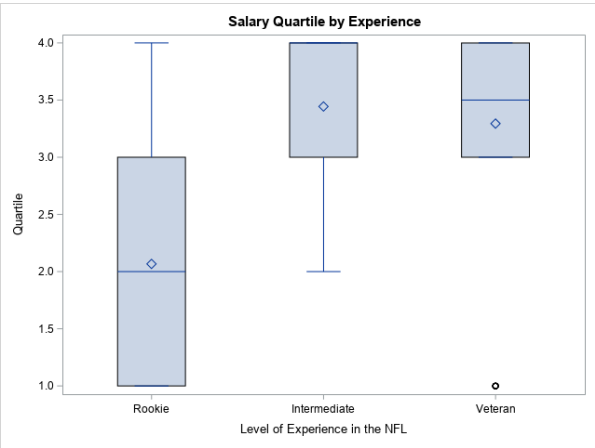
Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1.58280	0.21401	7.40	<.0001	.	0
Completion_Level	1	0.12134	0.07123	1.70	0.0900	0.97917	1.02127
Experience_Level	1	0.55729	0.07862	7.09	<.0001	0.97917	1.02127

Table 2: Salary (in dollars) VIF values

As is clear from the outputs above there there is no significant evidence for the presence of collinearity as none of the 'Variance Inflation' values are above 10. Therefore the use of both predictor variables in the same model is warranted.

The box-and-whisker plots are also crucial in analyzing the main effect of performance on salary as well as the main effect of experience on salary. This can be seen below in Graph 1 and Graph 2. Both show that they contribute heavily to salary.





Graph 1: Box & Whisker Plot Performance Graph 2: Box & Whisker Experience

Reduced Model - Performance

Our reduced model for performance excluded experience out of the model to test how salary would just be affected by performance independently. To test the model we used an F-test along with a p-value to test how significant performance would be in a reduced model. Our F-test was used with a significance level of 0.05 and a p-value that was tested against a significance level of 0.05. To better understand if the reduced model with the variable of performance holds a more significant correlation by itself we can check the adjusted R^2. This can be seen below in Table 3.

The GLM Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.2263674E12	6.2263674E12	6.09	0.0144
Error	207	2.1151666E14	1.0218196E12		
Corrected Total	208	2.1774302E14			

R-Square	Coeff Var	Root MSE	Salary Mean
0.028595	65.59094	1010851	1541144

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Completion_Level	1	6.2263674E12	6.2263674E12	6.09	0.0144

Table 3: GLM Reduced Model Procedure for Salary according to Performance only

Table 3 shows that the R^2 is at 0.029, which, out of 1, is not particularly high but it is not extremely low either. This goes to show that the reduced model is not a great model to show the correlation of performance towards the salary. The P-value is 0.0144, which is lower than the significance level of 0.05. This goes to show that the the performance, however is significant towards the salary.

Reduced Model - Experience

A reduced model for experience excluded the variable of performance to test whether the model was a better fit and how salary would be affected by experience in a model that was reduced. To test the effectiveness of this model an F-test along with a P-value with a significance level of 0.05 was used. From this we gain an even greater understanding for which variable holds more significance with our full regression model with all variables accounted for. To see how correlated this line is with the model we can look at the adjusted R^2. This can be seen below in Table 4.

The GLM Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.9720572E13	3.9720572E13	46.19	<.0001
Error	207	1.7802245E14	860011844638		
Corrected Total	208	2.1774302E14			

R-Square	Coeff Var	Root MSE	Salary Mean
0.182419	60.17401	927368.2	1541144

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Experience_Level	1	3.9720572E13	3.9720572E13	46.19	<.0001

Table 4: GLM Reduced Model Procedure for Salary according to Experience only.

Table 4 shows that the R^2 is 0.182, which is higher than the R^2 was for performance. While this number is still low and is still representative of why this model is not an accurate depiction of the amount that either experience or performance plays in salary, it still is accurate in its depiction of how experience is more correlated to salary than performance. The P-value is <.0001 which is smaller than 0.05, the alpha and significance level. Again, this shows that experience is significant towards the salary.

Results and Conclusions

Conclusions of Interaction Model

After analyzing the results of our study, we determined that the two factors are significant, however they have no interaction with each other. This means that neither of our main effects are influence by a third variable, although we see that if a kicker is a rookie, there can be interaction present. This has to do with what we have previously discussed, about a kicker's performance at the time of contract signing. Illustrated in Figures 1 and 2 are the interactions plots for our two dependent models.

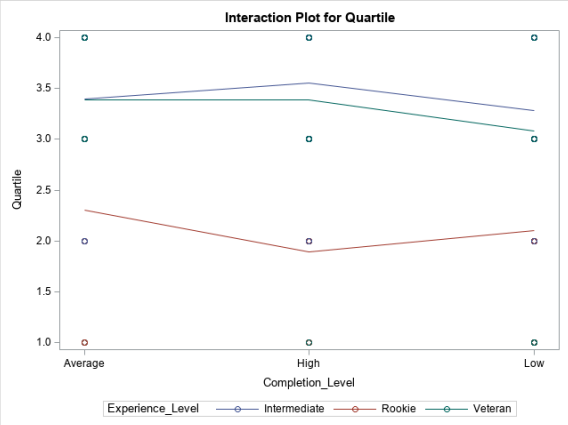


Figure 1: Interaction Plot for Quartile

Figure 2: Interaction Plot for Salary($)

Salary Dollar Amount Model

The SAS System

The GLM Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	5.7376847E13	7.1721059E12	8.94	<.0001
Error	200	1.6036618E14	801830884840		
Corrected Total	208	2.1774302E14			

R-Square	Coeff Var	Root MSE	Salary Mean
0.263507	58.10295	895450.1	1541144

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Completion_Level	2	7.8650665E12	3.9325333E12	4.90	0.0083
Experience_Level	2	4.6927085E13	2.3463543E13	29.26	<.0001
Completio*Experience	4	2.5846955E12	646173876158	0.81	0.5227

Table 3: ANOVA output for Salary($)

Analyzing our complete model above, we can determine that our two factors of completion percentage and experience level are both significant based on their P-values. .0083 for performance and <.0001 for experience; they are both below our alpha of .05. However, evaluating the P-value of our interaction term, the interaction between completion and

experience, we have a considerably higher P-value of 0.5227. This allows us to determine that, according to our model, there is no significant interaction when using a NFL kicker's completion percentage and experience level to predict salary. Furthermore, our R^2 value of .263 means that our model does not account for a significant amount of the variability of these kickers' salaries, so it is not a very strong model.

Salary Quartile Model

The SAS System

The GLM Procedure

Dependent Variable: Quartile

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	75.9692565	9.4961571	13.56	<.0001
Error	200	140.0307435	0.7001537		
Corrected Total	208	216.0000000			

R-Square	Coeff Var	Root MSE	Quartile Mean
0.351710	27.89173	0.836752	3.000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Completion_Level	2	7.50768549	3.75384275	5.36	0.0054
Experience_Level	2	65.99615102	32.99807551	47.13	<.0001
Completio*Experience	4	2.46541996	0.61635499	0.88	0.4767

Table 4: ANOVA output for Quartile

As you can see above, the P-value level of completion percentage and experience are both below our selected alpha of 0.05, being .0054 and <.0001 respectively, meaning that again, they are both to be considered significant in our model. However, when we once again look to the P-value of our interaction term, we see a level significantly higher than our alpha at 0.4767. This means that we have determined there in so significant interaction between the two. Furthermore, our R^2 value for this model 0.35 which means that it accounts for more of the variability of salary quartiles than the previous model for dollar values of salary, however, at under 0.5, it is not a great model in accounting for these kickers' salaries.

Conclusions of Reduced Models

Looking at the results of the reduced models we can see from the SAS outputs that although both the reduced models (performance and experience) were significant - and therefore qualify for variables that do indeed relate to players salaries, they account for very little variation amongst these kickers' salaries and are not reliable models in determining salary for NFL kickers.

Performance:

The GLM Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.2263674E12	6.2263674E12	6.09	0.0144
Error	207	2.1151666E14	1.0218196E12		
Corrected Total	208	2.1774302E14			

R-Square	Coeff Var	Root MSE	Salary Mean
0.028595	65.59094	1010851	1541144

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Completion_Level	1	6.2263674E12	6.2263674E12	6.09	0.0144

Table 5: ANOVA output for Salary reduced to Performance

As we can see, the P-value for the reduced model regarding performance is 0.0144 which is indeed smaller than 0.05 (our alpha value). However the R^2 of this model 0.02, meaning that our performance levels do not account for a large amount of the variability of kickers' salaries.

Experience:

The GLM Procedure

Dependent Variable: Salary

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.9720572E13	3.9720572E13	46.19	<.0001
Error	207	1.7802245E14	860011844638		
Corrected Total	208	2.1774302E14			

R-Square	Coeff Var	Root MSE	Salary Mean
0.182419	60.17401	927368.2	1541144

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Experience_Level	1	3.9720572E13	3.9720572E13	46.19	<.0001

Table 6: ANOVA output for Salary reduced to Experience

As we can see, the P-value for the reduced model regarding experience is <0.0001 which is indeed smaller than 0.05 (our alpha value).

However, even though these reduced models show that the variables are indeed legitimate in helping predict player's salaries, the reduced models themselves are not good models in their entirety. This can be seen above when looking at the 'R-square' values for each reduced model. The R-square values for performance and experience are 0.0285 and 0.182 respectively which shows that only 2.9% and 18.2% of the variance in the data can be explained by the models. This

does not suggest that either of these two reduced models should be used alone to predict players salaries.

Limitations of the Study

After analyzing and interpreting our data, and forming conclusions, the group has also formulated certain limitations we believe our study faced. To begin, we have determined that a kicker's performance (completion percentage) and their years of experience may be dependent on each other. Players that are in the league for longer may have a better possibility of having higher completion percentage. While rookies may have a great season, and as such, positively skewed completion percentages, the converse can be true as well. Additionally, the salary that a kicker earns, or any player for that matter, may be dependent on the financial situation of the team they are on. There are teams in bigger markets that have more money to spend on players, or more simply, there are rich teams and there are poor teams. Finally, kickers' salaries are very often dependent on their performance at the time of their contract signing more than their performance at any other point in time. The kicker could have a fantastic year, sign a contract with a team for a large salary, and then perform poorly thereafter.

References

National Football League. (n.d.). Official 2018 National Football League Record & FactBook. Retrieved from http://www.nfl.com/static/content/public/photo/2018/08/09/0ap3000000946101.pdf

Appendix I: SAS Code

```
*create table with experience*;
proc sql;
create table Players as
select Player, Experience
from tmp1.Nfl_rosters Where (Position= 'PK');
quit;
*create completion percent variable*;
proc sql;
create table Kickers_stats as
select Player, FGA_20_29, FGM_20_29,FGA_30_39, FGM_30_39,FGA_40_49,
FGM_40_49,FGA_50_plus, FGM_50_plus
from tmp1.Nfl_kicker_stats;
quit;
data Kickers_stats;
set Kickers_stats;
TFGA=(FGA_20_29+FGA_30_39+FGA_40_49+FGA_50_plus);
run;
data Kickers_stats;
set Kickers_stats;
TFGM=(FGM_20_29+FGM_30_39+FGM_40_49+FGM_50_plus);
run;
data Kickers_stats;
set kickers_stats;
Completion_Percent=(TFGM/TFGA);
run;
*gather salaries*;
proc sql;
create table kicker_sal as
select Player, Salary
from tmp1.Nfl_salaries where (Position = 'K');
quit;
*qualitate salaries*;
data kicker_sal;
set kicker_sal;
if salary<465000 then Quartile=1;
if 465000<=salary<732499 then Quartile=2;
```

```
if 732500<salary<1574999 then Quartile=3;
if salary>1575000 then Quartile=4;
run;
proc sql;
create table kickers2 as
select a.Player, a.salary, a.Quartile, (completion_percent*100) as Completion
from Kicker_sal a inner JOIN Kickers_stats b
on upcase(a.Player)= upcase(b.Player);
quit;
*combine player, experience, completion and salary in one table*;
proc sql;
create table kickers as
select a.Player, a.Salary, a.Quartile, a.Completion, b.Experience
from kickers2 a inner join players b
on upcase(a.Player)= upcase(b.Player);
quit;
*qualitate years of experience*;
data kickers;
set kickers;
if experience<=5 then Experience_Level=1;
if 5<experience<=10 then Experience_Level=2;
if 10<experience then Experience_Level=3;
run;
proc format;
value experiencefmt
1= "Rookie"
2= "Intermediate"
3= "Veteran";
run;
*to classify completion into categories*;
proc means data=kickers1 max min mean median CLM alpha=.01;
var completion;
run;
data kickers;
set kickers;
if completion<82 then Completion_Level=1;
if 82<=completion<86 then Completion_Level=2;
if completion=>86 then Completion_Level=3;
run;
```

```
proc format;
value completionfmt
1= "Low"
2="Average"
3= "High";
run;
data kickers;
set kickers;
label completion_level="Field Goal Completion Performance"
experience_level="Level of Experience in the NFL";
run;
*Run qualitative model Salary Quartile = Completion Percent + Experience Level + Interaction
Term*;
proc glm data=kickers;
class completion_level experience_level;
model quartile= completion_level experience_level completion_level*experience_level;
format completion_level completionfmt. experience_level experiencefmt.;
output p=pred r=resid rstudent=stresid;
run;
*quantitative model for salary actual value*;
proc glm data=kickers;
class completion_level experience_level;
model salary= completion_level experience_level completion_level*experience_level;
format completion_level completionfmt. experience_level experiencefmt.;
output p=pred r=resid rstudent=stresid;
run;
*obtain VIF values*;
proc reg data=kickers;
model salary= completion_level experience_level /tol vif collin ;
format completion_level completionfmt. experience_level experiencefmt.;
output p=pred r=resid rstudent=stresid;
run;
proc reg data=kickers;
model quartile= completion_level experience_level/ tol vif collin ;
format completion_level completionfmt. experience_level experiencefmt.;
output p=pred r=resid rstudent=stresid;
run;

*create graphs to illustrate main effects*;
```

```
proc sgplot data=kickers;
vbox salary/category=completion_level;
format completion_level completionfmt.;
run;
title "Salary($) by Experience";
proc sgplot data=kickers;
vbox salary/category=experience_level;
format experience_level experiencefmt.;
run;
title "Salary Quartile by Performance";
proc sgplot data=kickers;
vbox quartile/category=completion_level;
format completion_level completionfmt.;
run;
end;
title "Salary Quartile by Experience";
proc sgplot data=kickers;
vbox quartile/category=experience_level;
format experience_level experiencefmt.;
run;
*testing for interaction between completion levels and experience levels, there's none*;
title "Performance by Experience";
proc sgplot data=kickers;
vbox completion_level/ category=experience_level;
format completion_level completionfmt. experience_level experiencefmt.;
run;
proc univariate;
var resid;
qqplot resid/normal;
Run;
*reduced models regression*;
proc glm data=kickers;
model salary=completion_level;
format completion_level completionfmt.;
run;
proc glm data=kickers;
model salary=experience_level;
format experience_level experiencefmt.;
run;
```

```
proc glm data=kickers;
model quartile=completion_level;
format completion_level completionfmt.;
run;
proc glm data=kickers;
model quartile=experience_level;
format experience_level experiencefmt.;
run;
ODS RTF CLOSE;
```