FIFA Player Ratings: An ANOVA Analysis of Wage, Age, and Foot Preference

STAT 3220 Section 001

Team Name: stat group

Xinyue Qiu, Emily Puleo, Nancy Wang, Siri Gundapaneni

December 6, 2019

Fall, 2019

Introduction

We intend to look at this data and analyze what attributes make or break a soccer player in terms of rating score. From analyzing past data on soccer players, we will be able to predict if a player's rating will be higher or not on a scale from 1-100. This study can impact people's everyday decisions and lifestyles whether its what team to root for in a soccer match or what player to choose for your fantasy league.

Though this study can be used for recreational purposes, it can also serve a very realistic purpose for coaches and recruiters for soccer. This was sample only consists of FIFA players, but we can apply it for high school players looking to move towards college soccer and college soccer players looking to move forward and play professionally in the FIFA league. Approximately, only 5.5% of high school soccer players move on to play NCAA's men's soccer.[5] NCAA is the National Collegiate Athletic Association that centralizes and regulates college sports. After analyzing this data we can look at soccer players' attributes from an early stage and see which players have the characteristics to have a high rating in FIFA and be successful.

We chose to analyze this data using Three-Factor ANOVA with our factors being wage group, age group, and preferred foot. Since our analysis method requires qualitative independent variables we separated the wage group into low, medium, and high earnings and the age group into young and old. Preferred foot corresponds with left and right foot. We began our analysis by looking at each of the twelve possible three-way interactions between the groups and then looked at the two-way interactions. At each analysis step, we test the hypothesis that all of the treatment groups have equal means of overall rating score. Through our analysis, we found none of the three way interactions to be significant. We found our two way interactions between age group and wage group to be significant and by analyzing the two way interactions through extensive post-hoc testing we found five significantly different treatment groups: Young/Low Wage (1), Young/High Wage (2), Old/Low Wage (3), Old/High Wage (4), Young/Medium Wage and Old/High Wage (5). Each of three post-hoc tests (with Bonferroni's test being the preferred post-hoc test) reach the same conclusion regarding the significantly different groups causing us to be confident in our results.

Data Summary

Data Description

FIFA 19 is a football simulation video game, developed by EA Vancouver with several versions as an electric game series.[3] The dataset, obtained from Kaggle.com, contains information of FIFA 19 players.[3] The data was scraped from a website containing statistical information of all players in FIFA 19.[4] Having information like Nationality, Potential, Club, Value, Wage, Preferred Foot, International Reputation, this dataset covers describes about 19000 players using 90 detailed attributes that allow us to perform analysis on the general performance of the players. Each observation represents 1 single player. The data set is restricted only to the version released in June 2018. To further analyze the data, we first filtered the data to remove any observations with missing data points and selected a random, balanced sample of 600 observations. Given the abundance of qualitative and quantitative variables contained in our data set, we then selected three variables: wage, preference foot and age, for analysis of general performance. The selection is based on the fact that these variables can be coded into categories and intuitively it appears that these variables may affect overall rating. For the two variables--wage and age-- that are quantitative, we recode them each into groups for further exploration using a three-way ANOVA test.

Data Dictionary

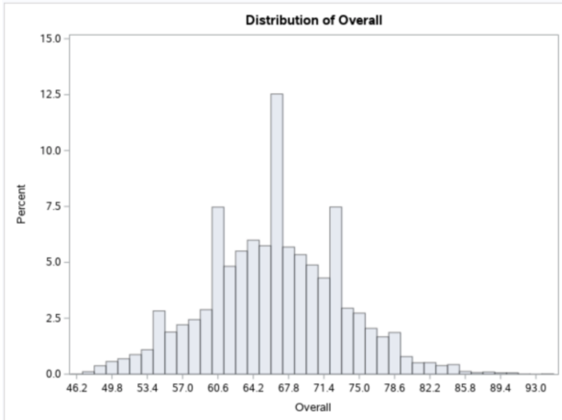
Reference Variable Name	**Variable name**	**Variable description**	**Variable type**	**Dummy variable**
y	Overall Rating	The rating of player, ranging from 0 to 100	Quantitative	-
x1	Preferred foot	The preferred foot of the player	Qualitative	Dummy=0 when foot= 'left' Dummy=1 when foot= 'right'
x2	Age	The age of the player	Qualitative	Dummy=0 when age= 'young' (≤ 30 years) Dummy=1 when age= 'old' (> 30 years)
x3	Wage	The average wage of the player	Qualitative	x1=0 when wage= 'low'(≤ 30,000€) x2=1 when wage= 'medium'(30,000€ < x2 < 60,000€) x3=2 when wage= 'high'(≥ 60,000€)

Why our data is appropriate for Three-Factor ANOVA

We are going to use three-factor ANOVA in this project. The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups. We will be able to conduct a high dimensional analysis by looking at how means vary across groups that are in multiple categories. As mentioned before, the three factors we choose are income level, age level and preferred foot, and each of them has 2 or 3 levels. And the response variable we choose is a quantitative variable: overall rating, therefore ANOVA is appropriate for our data in analyzing whether there are statistically significant differences between the means of two or three independent factors.

Continuity of Response Variable

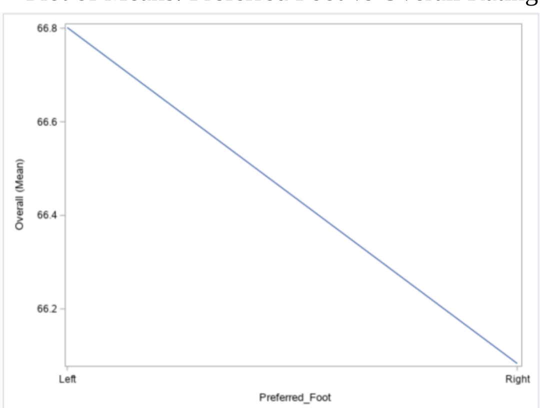
Since we designated the quantitative variable overall rating (y) to be the response variable we needed to verify that the data was continuous. We checked the histogram of the distribution of overall rating (y) to visualize continuity. As shown below, the continuity of the data allows us to select overall rating (y) to be our response variable.



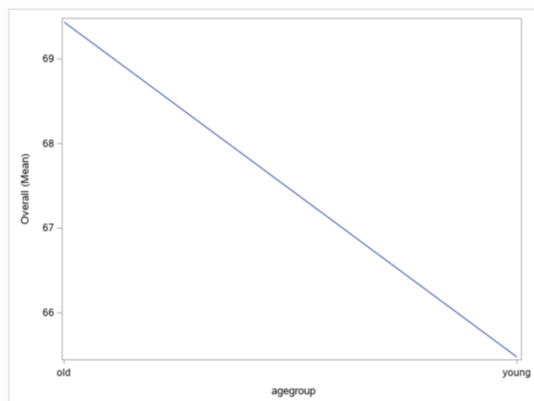
Qualitative EDA

For the three factors, we graphed the means of the qualitative variables using the vline statement in SAS. The following graphs show the qualitative variable means for income level (x1), age level (x2) and preferred foot (x3). Therefore, the ANOVA is appropriate for our analysis.

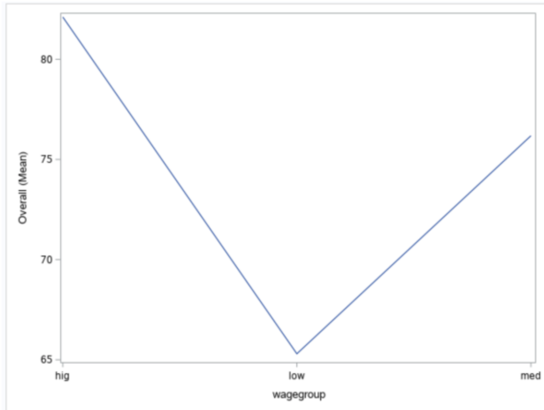
Plot of Means: Preferred Foot vs Overall Rating



Plot of Means: Age vs Overall Rating



Plot of Means: Wage vs Overall Rating



The plot of means shows us that our response variable varies across th chosen independent variables so we can conclude that our data set will be a good match for our analysis technique.

EDA Conclusions

From observing the above scatterplots and plot of means, we believe that the independent variables, age, wage, and foot preference, will interact to be useful for estimating a soccer player's overall rating.

Methodo

The Three-Factor ANOVA statistical method is used to estimate a quantitative variable based on qualitative data. The first step in any statistical analysis process is to check the assumptions of the method. When using ANOVA, we assume that each of the treatment groups are normally distributed and that the variance within the treatment groups are equal. ANOVA is robust against these assumptions, but it is good practice to check and correct them if necessary. The procedure of the Three-Factor ANOVA method is very systematic and simply builds on its related statistical methods: One-Factor and Two-Factor ANOVA. This method begins by considering including all possible three way interaction terms and if necessary moving on to analyzing the two-way interactions and main effects.

We complete a nested F-test using the model with all possible interaction terms (one way, two way, and three way interactions) as the complete model and using the model with all terms except for the three-way interaction terms as the reduced model.[8] The sum squared errors of the two models are compared and we can determine if the three-way interactions are significant (they significantly affect the model). If the terms are not significant, we remove them from the model and then look at the significance of all of the two-way interaction terms. Similar to testing the three-way interactions, we perform a nested F-test that has a complete model (all main effects and all two-way interactions) and a reduced model (main effects only). After comparing the sum squared errors of the models we can determine if the interactions are significant. If they are not significant, then we move on to perform an ANOVA comparing only the main effects.

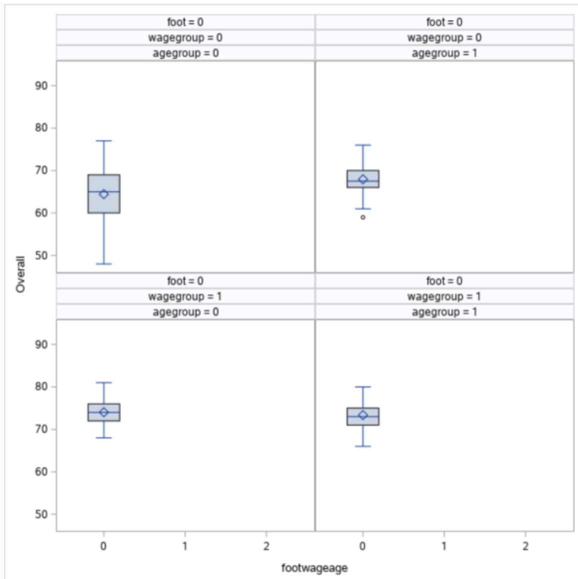
If at any point during the analysis the interaction terms are found to be significant, then it is appropriate to perform post hoc analysis on the terms. In this report, we are using Scheffe's test, Bonferroni's test, and Tukey's test for post hoc analysis. In addition to analysis, we will compare the properties of these tests to see which is most applicable to our data. Scheffe's test is a general procedure for testing all pairwise combinations of the treatment means that is usually used for exploratory data analysis. This test is not useful for large samples.[6] Bonferroni's method is useful for either balanced or unbalanced data sets because the method produces exact results for either case. The method also makes multiple comparisons between the treatment means, such as general contrasts, pairwise comparisons, and combinations of pairwise and complex contrasts.[6] Furthermore, this test does not lose its power when used with large samples. The other post hoc test we are using in this analysis is Tukey's test. Tukey's test is best when used with small sample sizes and balanced data.[6] It is the most restrictive of the tests we are using. Each of these tests work by calculating a specific critical distance and then designating which groups are significantly different based on whether the treatment group means are within the critical distance. If treatment group means are within the critical distance, then the are not significantly different from each other.

The three-factor ANOVA method is most useful when estimating a quantitative response variable from qualitative independent variables.[1] This method is also useful for reducing the risk of alpha inflation (or Type I Error) that arises from multiple comparisons.[2] ANOVA is able to reduce this risk by reducing the number of individual t-tests that are conducted. A higher number of t-test inherently increases the risk of making a Type I Error. ANOVA controls for these errors by comparing all means simultaneously.[7] Additionally, this method is beneficial because this analysis method yields the same information as multiple linear regression, but is simpler to understand and gives a direct comparison.[2] However, because of the way the hypothesis for this method is set up, we are unable to determine which mean is different.[1] We can only verify if a mean is different and therefore, we should be wary of making specific decisions based on this method. This method is insensitive to the ordering of the groups. This means that we can obtain the same results from ANOVA analysis if the categories are reordered, therefore we cannot draw conclusions about the relationship between groups that have multiple ways of being ordered.[2]

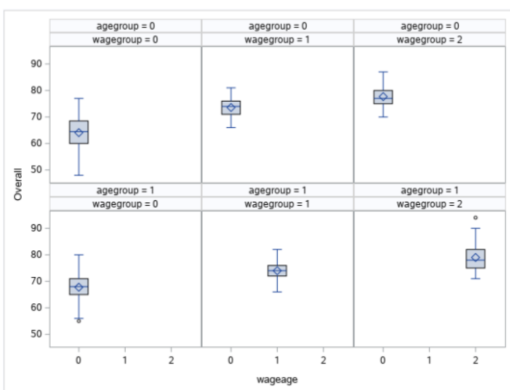
We will be using SAS for our analysis. This software has ample computing power to run this analysis, but generally this method would need computing power similar to that of multiple linear regression because these analysis techniques produce similar results and manipulate the data in similar ways. Since this method allows us to be able to make clear comparisons between variables we are able to pick a 'best' linear regression model. We can easily interpret this model as we would any other linear regression model. This method will give us a model that can be easily and practically interpreted.

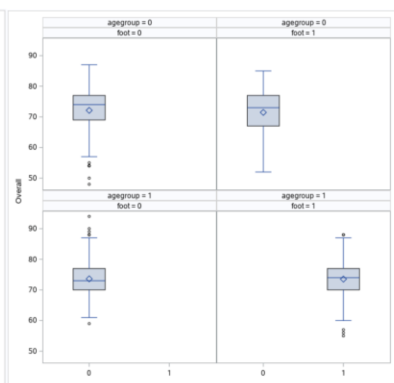
Above and Beyond: Influential Point Analysis

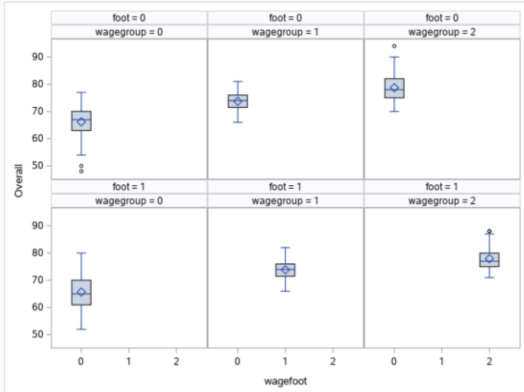
In an effort to further clean our data before completing ANOVA analysis, we investigated potential influential points that may skew our results. To identify these points we found the outlying points in each of the potential treatment groups. This meant that we first created boxplots for each of the possible three way interactions, a total of 12 different treatment groups. As shown below, only the Left Foot/Low Wage/Old group had a single outlier. The other groups did not have outlying points (see Appendix).



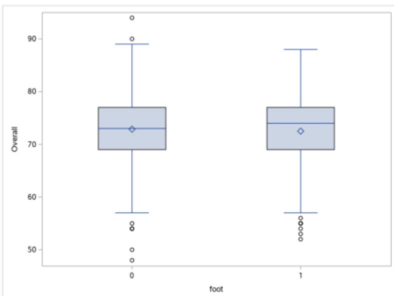
Next, we examined all of the observations that would make up the potential two way interaction groups. Out of the 18 possible treatment groups, eight of the groups had outlying points: Old/Low Wage, Old/ High Wage, Left Foot/Young, Left Foot/Old, Right Foot/Old, Left Foot/Low Wage, Left Foot/High Wage, Right Foot/High Wage. The outlying points are visible in the boxplots below.

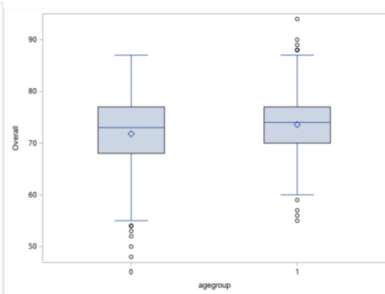


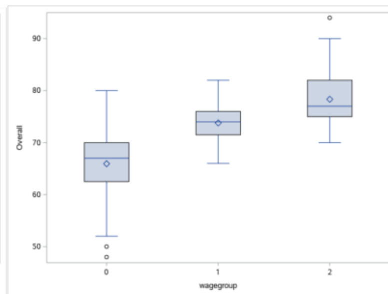




Finally, we examined all of the observations in the main effect groups. All of the treatment groups except for the Medium Wage group have outlying points, as seen below:







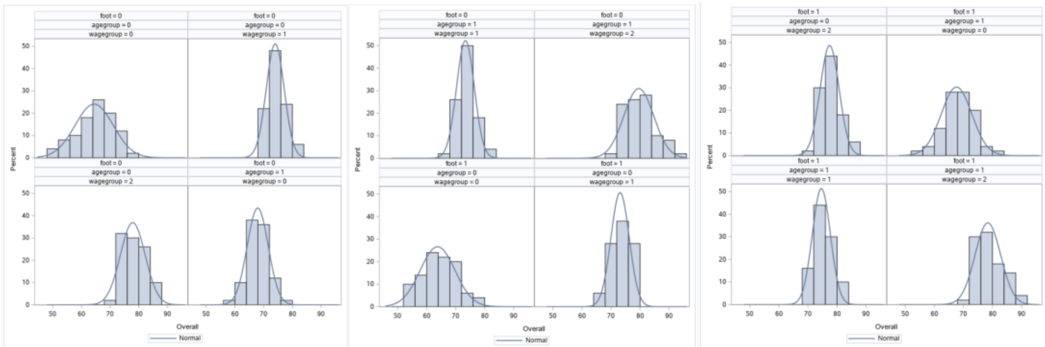
Influential Point Analysis Conclusion

We were able to successfully identify potential influential points in fourteen of the potentially useful treatment groups. However, this analysis does not give us enough information about the point so we cannot be confident that the points are influential to the analysis rather than simply outliers in the data set. Therefore, because we can confirm the extent to which these points will influence the data we leave the points in and continue our analysis as planned.

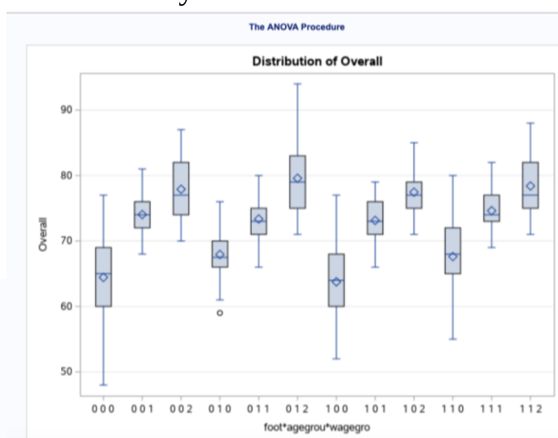
Three-Factor ANOVA Analysis

Checking Model Assumptions

The first assumption for ANOVA is that the population of each of our twelve treatment groups are normally distributed. To check this assumption we plotted histograms of the observations of each treatment group against the response variable. As seen in the plots below, each of the populations are approximately normal, and since we know that ANOVA testing is robust against normality we can conclude that we do not violate this assumption.



The second assumption for ANOVA is that each of the treatment populations has equal variances. We tested this assumption using Bartelet's Test (H0: $\boldsymbol{\sigma}^2 1 = \boldsymbol{\sigma}^2 2 = \ldots = \boldsymbol{\sigma}^2 12$, H1: at least two of them are unequal). The plot below shows how the spread of the data between the treatment groups is not approximately equal. The Bartlett's Test chi square statistic and the associated p value of <0.001 indicate that we reject the null hypothesis that all of the treatment populations have equal variance and therefore violate the assumption of equal variance. However, since ANOVA is robust against this assumption violation, we can proceed with analysis.



The ANOVA Procedure

Bartlett's Test for Homogeneity of Overall Variance			
Source	DF	Chi-Square	Pr > ChiSq
foot*agegrou*wagegro	11	87.3844	<.0001

Three-Way Interactions

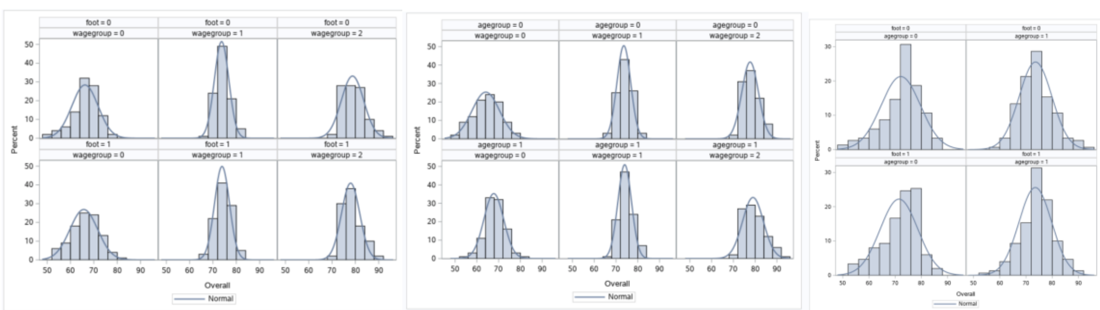
We begin ANOVA analysis by looking at the three way interactions between our three factors, age, wage, and foot preference. Our hypothesis is that the mean overall rating score for all of the three way interactions are equal. The alternate hypothesis is that at least one of the three way interactions are not equal. After conducting a Nested F-Test which tested a model with all one way, two way, and three way interactions against a model with only one way and two way interactions. We determined that the three way interaction term was not significant and had a p-value of 0.2565. We can conclude that the three-way interaction terms are not significant, so we remove them from the model and analyze the two-way interactions.

Source	DF	Anova SS	Mean Square	F Value	Pr > F
foot	1	21.28167	21.28167	1.08	0.2984
agegroup	1	484.20167	484.20167	24.65	<.0001
wagegroup	2	15727.70333	7863.85167	400.33	<.0001
foot*agegroup	1	13.20167	13.20167	0.67	0.4127
foot*wagegroup	2	25.64333	12.82167	0.65	0.5210
agegroup*wagegroup	2	286.72333	143.36167	7.30	0.0007
foot*agegrou*wagegro	2	53.58333	26.79167	1.36	0.2565

Recheck Model Assumptions

When we move on with analysis and test new interaction terms we need to recheck the ANOVA regression assumptions since our treatment groups change to be all two-way interactions groups rather than three-way interaction groups.

We check the first assumption for ANOVA to see if the two way interaction groups have normal distributions. The histograms of the observations of each treatment group below show that each of the populations are approximately normal, and since we know that ANOVA testing is robust against normality we can conclude that we do not violate this assumption.



We recheck the second assumption for ANOVA regarding assumed constant variance of the treatment groups and conclude that we violate this assumption. We reach this conclusion because the each of the interactions (Wage/Age, Wage/Foot, Foot/Age) have a Bartlett's Test chi square statistic and associated p value of <0.05 indicate that we reject the null hypothesis that all of the treatment populations have equal variances. We know that ANOVA is robust against this assumption, so we continue with our analysis.

The ANOVA Procedure

Bartlett's Test for Homogeneity of Overall Variance

Source	DF	Chi-Square	Pr > ChiSq
agegroup*wagegroup	5	74.1897	<.0001

The ANOVA Procedure

Bartlett's Test for Homogeneity of Overall Variance

Source	DF	Chi-Square	Pr > ChiSq
foot*wagegroup	5	74.5669	<.0001

The ANOVA Procedure

Bartlett's Test for Homogeneity of Overall Variance

Source	DF	Chi-Square	Pr > ChiSq
foot*agegroup	3	8.0517	0.0450

Two-Way Interactions

Our next step in ANOVA analysis is to look at the possible two way interactions between our three factors, age, wage, and foot preference. Our hypothesis is that the mean overall rating score for all of the two way interactions are equal. The alternate hypothesis is that at least one of the three way interactions are not equal. After conducting a Nested F-Test which tested a model with all one way and two way interactions against a model with only one way interactions. We determined that one of the two way interactions was significant based on the p-values of the interactions. The interaction between foot preference and age and the interaction between wage and foot preference yielded p-values of 0.4130 and 0.5214 respectively. Because of the high p-values we removed these terms. Since foot preference does not interact significantly with the other independent variables we remove it from analysis completely. The interaction between age and wage was statistically significant with a p-value of 0.0007. To further analyze the relationship between the wage variable and the age variable we will conduct post-hoc analysis.

Source	DF	Anova SS	Mean Square	F Value	Pr > F
foot	1	21.28167	21.28167	1.08	0.2987
agegroup	1	484.20167	484.20167	24.62	<.0001
wagegroup	2	15727.70333	7863.85167	399.84	<.0001
foot*agegroup	1	13.20167	13.20167	0.67	0.4130
foot*wagegroup	2	25.64333	12.82167	0.65	0.5214
agegroup*wagegroup	2	286.72333	143.36167	7.29	0.0007

Post Hoc Test 1: Bonferroni's Test

Unfortunately, it is not possible to conduct Bonferroni's Test in SAS when looking at interactions between treatment groups. In lieu of a SAS command, we calculated the critical distance, B_{ij}, for our data using the following formula:

$$B_{ij} = t_{\alpha/(2g)} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Where

g = Number of pairwise comparisons=(6)

s =√MSE=4.43131

ni = Number of observations in sample for treatment i=300

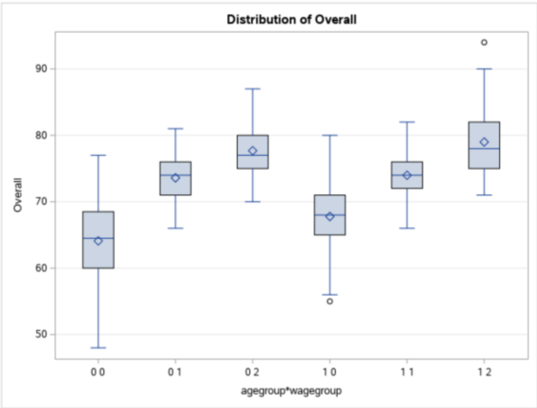
nj = Number of observations in sample for treatment j=200

B_{ij} = 1.17

Using the critical distance from Bonferroni's Test, we can use this to determine if the tested treatment groups are significantly different. By adding and subtracting the critical distance from the mean of each of the six different treatment groups and checking for overlap, we can verify whether or not the treatment groups have significantly different means. We use the table of means below to calculate the overlap regions:

Level of agegroup	Level of wagegroup	N	Overall	
			Mean	Std Dev
0	0	100	64.1000000	6.30295746
0	1	100	73.5900000	3.15618681
0	2	100	77.6800000	3.82939191
1	0	100	67.7800000	4.52508162
1	1	100	73.9900000	3.13176944
1	2	100	78.9900000	4.81473957

After applying the critical distance to each of the treatment groups we discovered that we have a total of five significantly different treatment groups: Young/Low Wage (1), Young/High Wage (2), Old/Low Wage (3), Old/High Wage (4), Young/Medium Wage and Old/High Wage (5). From the results of Bonferroni's test, we can conclude that the Young/Low Wage group has the lowest mean overall rating and the Old/High Wage has the highest mean overall rating.



Post Hoc Test 2: Scheffe's Test

Unfortunately, it is not possible to conduct Scheffe's Test in SAS when looking at interactions between treatment groups. Instead of using SAS for calculations, we calculated the critical distance, S_{ij}, for our data using the following formula:

$$S_{ij} = \sqrt{(p-1)(\mathrm{F}_\alpha)(\mathrm{MSE})\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Where

p = Number of sample (treatment) means= 6

s =$\sqrt{\mathrm{MSE}}$=4.43131

ni = Number of observations in sample for treatment i=300

nj = Number of observations in sample for treatment j=200

S_{ij}= 0.64

We apply the critical distance from Scheffe's Test in the same way at Bonferroni's test by looking at treatment group means that overlap within the critical distance range. Even with a smaller critical distance, Scheffe's test yields the same result as Bonferroni's test. We obtain the same five significantly different treatment groups: Young/Low Wage (1), Young/High Wage (2), Old/Low Wage (3), Old/High Wage (4), Young/Medium Wage and Old/High Wage (5). Because the means of the treatment groups do not change when we change post-hoc tests, the high and low mean treatment groups remain the same, the Young/Low Wage group and the Old/High Wage group respectively.

Post Hoc Test 3: Tukey's Test

As with the previous two post-hoc tests, it is not possible to conduct Tukey's Test in SAS when looking at interactions between treatment groups. To make sure we could still use this post-hoc test, we calculated the critical distance, w_{ij}, for our data using the following formula:

$$\omega_{ij} = q_\alpha(p, \nu)\frac{s}{\sqrt{2}}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Where

p = Number of sample (treatment) means= 6
ν = Number of degrees of freedom associated with MSE=594
s =√MSE=4.43131
ni = Number of observations in sample for treatment i =300
nj = Number of observations in sample for treatment j =200

w_{ij}= 1.05

We use the critical distance from Tukey's Test in the exact same way as the critical distances in Bonferroni's and Scheffe's test were used. Again, we arrive at the same conclusion and further confirm that there are five significantly different treatment groups (Young/Low Wage (1), Young/High Wage (2), Old/Low Wage (3), Old/High Wage (4), Young/Medium Wage and Old/High Wage (5)).

Comparison of Post-Hoc Tests

It is important to note again that we conduct post-hoc analysis tests in order to correct for Type I error and verify that the conclusions drawn from ANOVA testing are correct. One way to compare post-hoc tests is to compare the critical distances that each test yields. The tests have different methods for conducting their respective critical differences, so it is important to decide which method calculates the best critical distance for our data. First, we compare relative critical distance size and observe that Scheffe's test has the smallest critical distance, Tukey's test has the next smallest, and Bonferroni's test has the largest critical distance. Solely based on critical distance size, we can say the Scheffe's test has the least strict guidelines when deciding whether two groups are significantly different or not. The smaller the critical distance, the less like an overlap will be and more likely two groups are to be significantly different. Similarly, Bonferroni's test has the largest critical distance which increases the likelihood that a neighboring mean will overlap with it rendering the groups not significantly different. Simply by analyzing the relative sizes of the critical means of each of the post hoc tests, we can conclude the Bonferroni's has the highest standards for two treatment groups to be significantly different.

We can also evaluate post-hoc tests based on the comparisons that they make. All of the tests make pairwise comparisons, but Bonferroni's test has the added capability of making compound comparisons. However, in our case, the compound comparison does not apply. So, when judging the post-hoc tests based on the comparisons they make all of the tests are equal. Additionally, it is important to consider the purpose for which the test is meant to be used. Bonferroni's test is meant to be used as a correction for Type I error that occurs in ANOVA testing as well as for pairwise comparisons. Scheffe's test is meant to be used in exploratory data analysis or when looking generally at the data, but not with a targeted goal. Tukey's test is meant to be used for pairwise comparisons after ANOVA analysis. Given this information, we can confirm that we are not using the post-hoc tests in the context in which Scheffe's test is meant to be used.

Furthermore, we can make evaluations of how well each of the post-hoc tests fits our data based on the experimental design requirements of the test. Bonferroni's and Scheffe's test do not require anything specific from the data, but Tukey's test requires the data set to be balanced.

Taking all of these factors into consideration, we conclude that Bonferroni's test is the best fit for our data. We arrive at this conclusion because of the large critical distance which makes it the most difficult for treatment groups to be significantly different with this test. Additionally, this test has strong power for large sample sizes, while the other tests do not have the same power. Given that our sample encompasses 600 observations, we feel that it is important to use a post-hoc test that works well with large sample sizes. Further, we feel most comfortable using Bonferroni's test given the context in which we are conducting the test. Bonferroni's test is meant to be used to correct for Type I error and make pairwise comparisons, all of which we are interested in doing by using this post-hoc test.

Summary Table

Test Name	Distribution	Experimental Design Requirement	Comparison Type	Power of Test	Purpose of Test
Bonferroni's test	t-distribution	None	Compound comparison/ Pairwise comparison	Strong for both large and small sample size	Different combinations of comparison
Scheffe's test	F-distribution	None	Pairwise comparison	Weak	Exploratory data analysis
Tukey	t-distribution	Balanced	Pairwise comparison	Strong for only large sample size	Pairwise comparison

Conclusion

Analysis Interpretation

A three-way ANOVA was conducted on overall rating scores of soccer players considering age, wage, and foot preference as factors. We found that foot preference is not a significant estimator of a player's overall rating since it does not interact significantly with age or wage. We found that wage and age interact to estimate a player's overall rating. Post-hoc testing indicates that there are five significantly different groups from this interaction. The groups are: Young/Low Wage (1), Young/High Wage (2), Old/Low Wage (3), Old/High Wage (4), Young/Medium Wage and Old/High Wage (5). Of these groups, the Young/Low Wage group has the lowest mean overall rating and the Old/High Wage has the highest mean overall rating.

Each of three different post-hoc tests, Bonferroni, Scheffe, and Tukey's test, reach the same conclusion. The robustness of the conclusions leads us to be confident in the results of these tests. Analysis of the post-hoc tests brings the conclusion that Bonferroni's Test is the most appropriate post-hoc test for our study due to its large critical distance, strong power for large sample sizes, and purpose for testing correcting Type I error[9]. Further, even though we are currently using a balanced data set if we added more variables the data set may become unbalanced and we would still be able to use the Bonferroni post-hoc test.

While ANOVA does not produce a regression model where we can plug independent variables into to obtain our response variable output, there are other ways we can gain useful information from our analysis. Given the calculated treatment means and separation into significantly different groups we are able to estimate the response variable based on the independent variable predictors. We estimate that if a soccer player is Young and in the Low Wage group, they will have an overall rating score of about 64.1. If a soccer playing is Old and in the Low Wage group, we estimate they will have an overall rating score of about 67.8. If a soccer player in Young and in the High Wage group, we estimate that the player will have an overall rating score of about 77.7. We estimate that a soccer player in the Old and High Wage group will have an overall rating score of about 79. Lastly, if a soccer player is Young or Old but in the Medium Wage group, they will have an overall rating score in the range of about 73.6 to 74.

We further tested our conclusions by plugging in real data to the treatment groups. Our goal was to be able to take a random assessment of our conclusions to judge its adequacy with a new technique. We used a random entry that was in the original data set of 19,000 players to be sure that the data we entered was reasonable. This player was in the Young age group and High Wage group. Our ANOVA analysis indicated that the player should have an overall rating of about 79 while the player's actual overall rating was 85. However, the actual overall rating is only 1.04 standard deviations away from the estimated rating showing that while we are not able to estimate the overall rating score perfectly, we are able to obtain a reasonably close value.

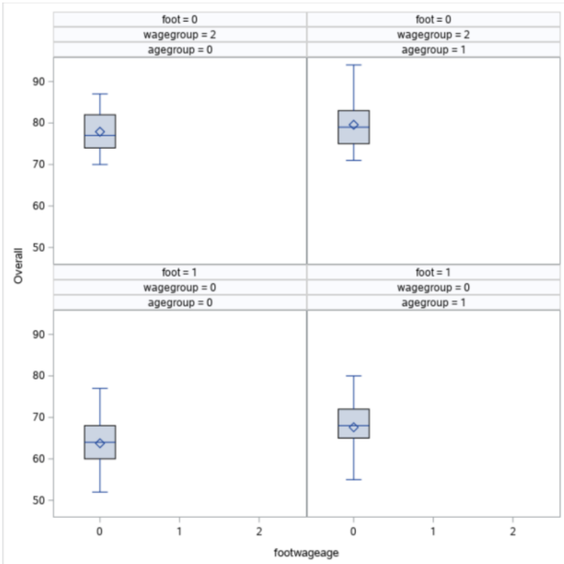
Limitations and Future Directions

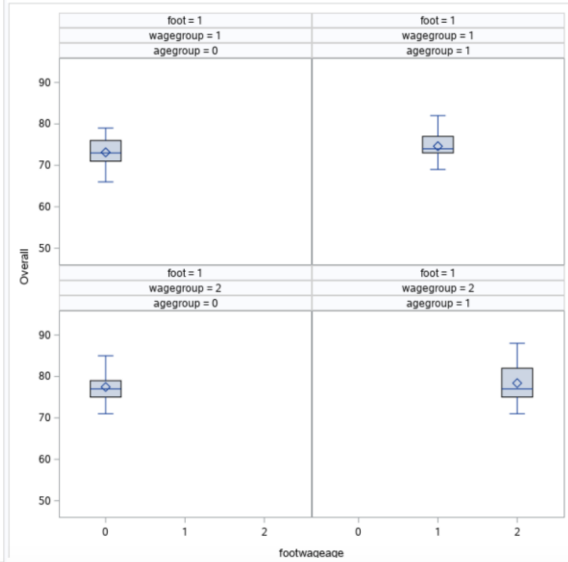
Our analysis has several limitations that must be considered when evaluating our conclusions. First and foremost, this analysis may not by applicable for a continuous time period or may not be fit for inference if the characteristics of soccer players change drastically. Since this data is based on a past version of the FIFA video game, if the game releases a new set of players with different statistics than the variables used for analysis, our conclusions may not be applicable to the new players.

In the future, we could complete further post-hoc testing to verify if the same groups are consistently deemed significantly different across multiple post-hoc tests. We could also complete variable screening to choose variables that may be better predictors of our response variable. Another important avenue we could explore with this data is the potential effect of influential points. In the beginning of our analysis we examined and identified potential influential points, but did not have enough evidence to remove them from the data set. In the future we could remove these points and determine the effect these points have on the final conclusions drawn from our three-way ANOVA.

Appendix

Influential Analysis - Three Way Interaction Treatment Group Box Plots





These box plots show no outlying points that could be potentially classified as influential.

SAS Code

*sample selection;

proc print data=mydata.fifadata;

```
run;

data fifadata1;
	set mydata.fifadata;
	foot=0;
	agegroup=0;
	wagenumber=compress(wage, 'K');
	wagequan=compress(wagenumber, '€');
	wagegroup=0;

	if 15<=wagequan<=30 then
		wagegroup=1;

	if wagequan>30 then
		wagegroup=2;

	if age>=30 then
		agegroup=1;

	if Preferred_Foot='Right' then
		foot=1;
run;

proc sort data=fifadata1;
	by agegroup foot wagegroup;
run;

proc surveyselect data=fifadata1 method=srs n=50 seed=48 out=mydata.samplefifa;
	strata agegroup foot wagegroup;
run;

*influential points;

proc sgplot data=mydata.samplefifa;
	vbox overall/category=foot;
run;

proc sgplot data=mydata.samplefifa;
	vbox overall/category=agegroup;
run;
```

```
proc sgplot data=mydata.samplefifa;
	vbox overall/category=wagegroup;
run;

data fifainfluential;
	set mydata.samplefifa;
	wageage=wagegroup*agegroup;
	wagefoot=wagegroup*foot;
	agefoot=agegroup*foot;
	footwageage=foot*wagegroup*agegroup;
run;

proc sgpanel data=fifainfluential;
	panelby agegroup wagegroup;
	vbox overall/category=wageage;
run;

proc sgpanel data=fifainfluential;
	panelby agegroup foot;
	vbox overall/category=agefoot;
run;

proc sgpanel data=fifainfluential;
	panelby foot wagegroup;
	vbox overall/category=wagefoot;
run;

proc sgpanel data=fifainfluential;
	panelby foot wagegroup agegroup;
	vbox overall/category=footwageage;
run;

*check assumptions;

proc sgpanel data=mydata.samplefifa;
	panelby foot agegroup wagegroup;
	histogram overall;
	density overall /type=normal;
run;
```

```
proc sgpanel data=mydata.samplefifa;
	panelby foot agegroup;
	histogram overall;
	density overall /type=normal;
run;

proc sgpanel data=mydata.samplefifa;
	panelby foot wagegroup;
	histogram overall;
	density overall /type=normal;
run;

proc sgpanel data=mydata.samplefifa;
	panelby agegroup wagegroup;
	histogram overall;
	density overall /type=normal;
run;

proc sgpanel data=mydata.samplefifa;
	panelby foot;
	histogram overall;
	density overall /type=normal;
run;

proc sgpanel data=mydata.samplefifa;
	panelby agegroup;
	histogram overall;
	density overall /type=normal;
run;

proc sgpanel data=mydata.samplefifa;
	panelby wagegroup;
	histogram overall;
	density overall /type=normal;
run;

proc anova data=mydata.samplefifa;
	class foot agegroup wagegroup;
	model overall=foot*agegroup*wagegroup;
```

```
	means foot*agegroup*wagegroup/ hovtest=bartlett;
	run;

proc anova data=mydata.samplefifa;
	class foot agegroup;
	model overall=foot*agegroup;
	means foot*agegroup/ hovtest=bartlett;
	run;

proc anova data=mydata.samplefifa;
	class agegroup wagegroup;
	model overall=agegroup*wagegroup;
	means agegroup*wagegroup/ hovtest=bartlett;
	run;

proc anova data=mydata.samplefifa;
	class foot wagegroup;
	model overall=foot*wagegroup;
	means foot*wagegroup/ hovtest=bartlett;
	run;
	*anova;

proc anova data=mydata.samplefifa;
	class foot agegroup wagegroup;
	model overall=foot agegroup wagegroup foot*agegroup foot*wagegroup
		agegroup*wagegroup foot*agegroup*wagegroup;
	run;

proc anova data=mydata.samplefifa;
	class foot agegroup wagegroup;
	model overall=agegroup wagegroup agegroup*wagegroup;
	run;
	*post hoc;

proc anova data=mydata.samplefifa;
	class agegroup wagegroup;
	model overall=agegroup*wagegroup;
	means ageagegroup*wagegroup/ bon cldiff;
	run;
```

```
proc anova data=mydata.samplefifa;
	class agegroup wagegroup;
	model overall=agegroup*wagegroup;
	means agegroup*wagegroup/scheffe cldiff;
	run;
	*tukey;

proc anova data=mydata.samplefifa;
	class agegroup wagegroup;
	model overall=agegroup*wagegroup;
	means agegroup*wagegroup/ tukey cldiff;
	run;
```

References

1. Kim, Tae Kyun. “Understanding One-Way ANOVA Using Conceptual Figures.” *Korean Journal of Anesthesiology*, The Korean Society of Anesthesiologists, Feb. 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5296382.
2. Lazic, Stanley E. “Why We Should Use Simpler Models If the Data Allow This: Relevance for ANOVA Designs in Experimental Biology.” *BMC Physiology*, BioMed Central, 21 July 2008, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2496911/.
3. Gadiya, Karan. “FIFA 19 Complete Player Dataset.” *Kaggle*, 21 Dec. 2018, https://www.kaggle.com/karangadiya/fifa19.
4. Borjigin, Khachin. “Players FIFA 20 Nov 13, 2019 SoFIFA.” *SoFIFA*, https://sofifa.com/.
5. Smeyers@ncaa.org. “Estimated Probability of Competing in College Athletics.” *NCAA.org - The Official Site of the NCAA*, 9 Apr. 2019, http://www.ncaa.org/about/resources/research/estimated-probability-competing-college-athletics.
6. Mendenall, William. “Chapter 12.8: The Analysis of Variance for Designed Experiments” *A Second Course in Regression Analysis: Seventh Edition*, Pearson Education Inc, 2012
7. Anders, Kallner. “Resolution of Students t-Tests, ANOVA and Analysis of Variance Components from Intermediary Data.” Biochemia Medica, Croatian Society of Medical Biochemistry and Laboratory Medicine, 15 June 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5493167/.
8. Mendenall, William. “Chapter 12: The Analysis of Variance for Designed Experiments” *A Second Course in Regression Analysis: Seventh Edition*, Pearson Education Inc, 2012
9. Lee, Sangseok, and Dong Kyu Lee. “What is the proper way to apply the multiple comparison test?.” *Korean journal of anesthesiology* vol. 71,5 (2018): 353-360. doi:10.4097/kja.d.18.00242