Second assignment

Ahmad Saquib Sina

February 25, 2019

Question one

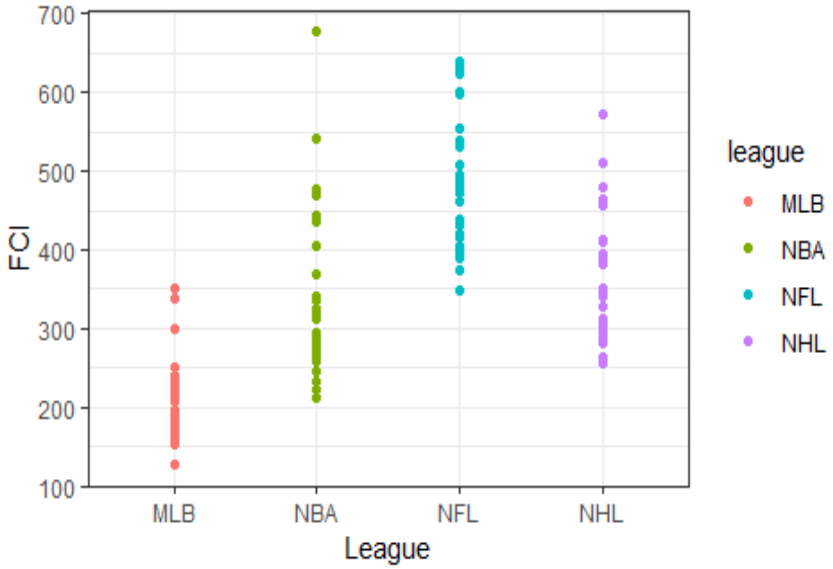

Figure 1:The scatterplot of the relationship between FCI and sporitng league

From the scatterplot, it is visible that the homogeneity of variance is likely violated. There is more varation in NBA leagues' FCI compared to MLB, NFL, and NHL leagues' FCI. Variation in NHL leagues' FCI seems slightly higher than variation in NFL leagues' FCI. MLB leagues' FCI has less variation compared to NFL and NHL leagues' FCI. Therefore, the scatterplot likely suggests the violation of the homogeneity of variance.

Question two

```
##
## Call:
## lm(formula = lfci ~ 1 + MLB + NFL + NHL, data = FCI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48823 -0.16561 -0.00802  0.11489  0.74784
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.76898    0.04058 142.179  < 2e-16 ***
## MLB         -0.43743    0.05738  -7.623 6.89e-12 ***
## NFL          0.39239    0.05648   6.948 2.18e-10 ***
```

```
## NHL         0.10122    0.05738   1.764   0.0803 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2222 on 118 degrees of freedom
## Multiple R-squared:  0.6521, Adjusted R-squared:  0.6433
## F-statistic: 73.73 on 3 and 118 DF,  p-value: < 2.2e-16
```

The NHL coefficient from the fitted model is 0.10. The interpretation of the NHL model is that the slope associated with NHL, $\hat{\beta}\ (NHL) = 0.10$, indicates that NHL league has a log-FCI that is 0.10 higher than NBA leagues' FCI, on average.

Question three

```
## (Intercept)        MLB        NFL        NHL
## 320.2097643   0.6456951   1.4805106   1.1065153
```

The back-transformed NHL coefficient is 1.11. The back-transformed interpretation is that NHL league has a FCI that is 1.11 TIMES the estimated FCI for NBA league, on average.

Question four

The trend that is observed from the loess smoother indicates that there is non-linear realtionship between salary and log-fci. From the scatter plot, it is apparent that the average log-FCI both increases and decreases with respect to player salary. Therefore, it indicates that there is a non-linear relationship between log-fci and player salary. There are two ways to model this nonlinearity. They are polynomial effect (power transformation such as quadratic, cubic) and another method of modeling nonlinearity is to transform the predictor (or outcome) using a nonlinear transformation, logarithm. Using a power tranformation of player salary (quadratic,cubic) is more appropriate than using a log transformation because quadratic or cubic function changes direcion. On the other hand, logarithmic function does not change direction. The quadratic function shows continuous and diminishing growth followed by continuous and increasing loss, while the logarithmic function models continuous, albeit diminishing, growth.Therefore,as the non-linear relationshship between salary and log-fci changes direction, power transformation is more appropriate than log tranformation.

Question five

```
## Analysis of Variance Table
##
## Model 1: lfci ~ 1 + salary
## Model 2: lfci ~ 1 + salary + Quadraticsalary
## Model 3: lfci ~ 1 + salary + Quadraticsalary + Cubicsalary
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    120 14.871
## 2    119 13.507  1    1.3637 13.789 0.0003137 ***
## 3    118 11.670  1    1.8375 18.580 3.395e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The nested F test suggests that there is a statistically signficant cubic effect of player salary on log-FCI, $F(1,118) = 11.67$, $p < 0.001$. Therefore, the results suggest the higher order terms that are tested are warranted.

Question six

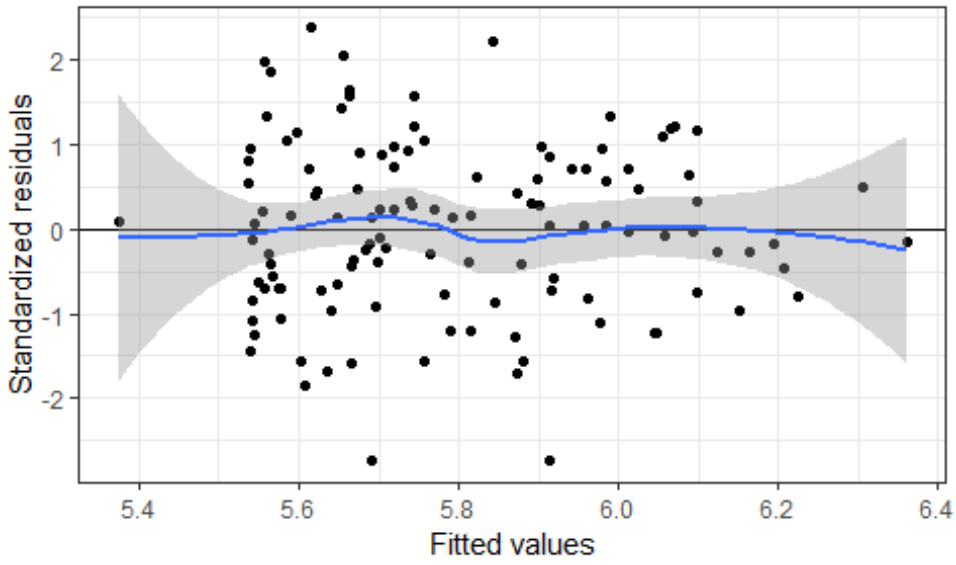

Figure 2:The scatterplot of the standardized residuals versus the fitted values using the cubic effect of player salary to predict log-fci

A horizontal line at Y=0 shows the expected mean residual under the linearity assumption. The loess line (blue) and uncertainty bands (grey shaded area) indicate that the average conditional residuals for each of the fitted values are not statistically different from 0. Therefore, the linearity assumption seems tenable. Therefore, the trasnformation fixes the non-linearity.

Question seven

```
##     (Intercept)          salary Quadraticsalary     Cubicsalary
##     338.0763621       0.9999996       1.0000000       1.0000000
##             MLB             NFL             NHL
##       0.6952038       1.9455869       1.4166726
```

The back-transformed NHL coefficient is 1.42. The back-transformed interpretation is that NHL league has a FCI that is 1.42 TIMES the estimated FCI for NBA leagues, on average.

Question eight

```
## Analysis of Variance Table
##
## Model 1: lfci ~ 1 + salary + Quadraticsalary + Cubicsalary + MLB + NFL +
##     NHL
## Model 2: lfci ~ 1 + salary + Quadraticsalary + Cubicsalary + MLB + NFL +
##     NHL + MLB_salary + NFL_salary + NHL_salary
## Model 3: lfci ~ 1 + salary + Quadraticsalary + Cubicsalary + MLB + NFL +
##     NHL + MLB_salary + NFL_salary + NHL_salary + Quadsalary_Mlb +
```

```
##     Quadsalary_Nfl + Quadsalary_Nhl
## Model 4: lfci ~ 1 + salary + Quadraticsalary + Cubicsalary + MLB + NFL +
##     NHL + MLB_salary + NFL_salary + NHL_salary + Quadsalary_Mlb +
##     Quadsalary_Nfl + Quadsalary_Nhl + Cubicsalary_Mlb + Cubicsalary_Nfl +
##     Cubicssalary_Nhl
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    115 4.8104
## 2    112 4.3867  3   0.42374 3.6144 0.01566 *
## 3    109 4.3130  3   0.07369 0.6285 0.59817
## 4    106 4.1424  3   0.17060 1.4552 0.23103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the results from the Nested F test, it is apparent that model Model 5 (the interaction model with linear and quadratic interaction effects) and Model 6 (the interaction model with linear, quadratic, and cubicinteraction effects) are not statistically significant. However, Model 4 is statistically significant, $F(3,112)=3.61, p<0.05$. Therefore, the interaction model with linear interaction effects that is Model 4 should be adopted based on the results of this test.

Question nine

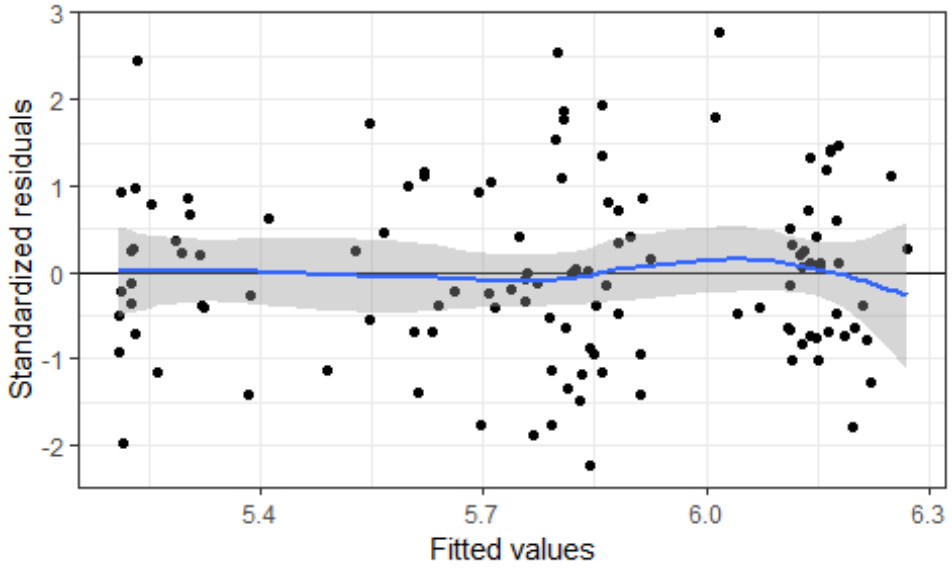

Fig 3: The scatterplot of the standardized residuals versus the fitted values using the interaction model with linear interaction effect

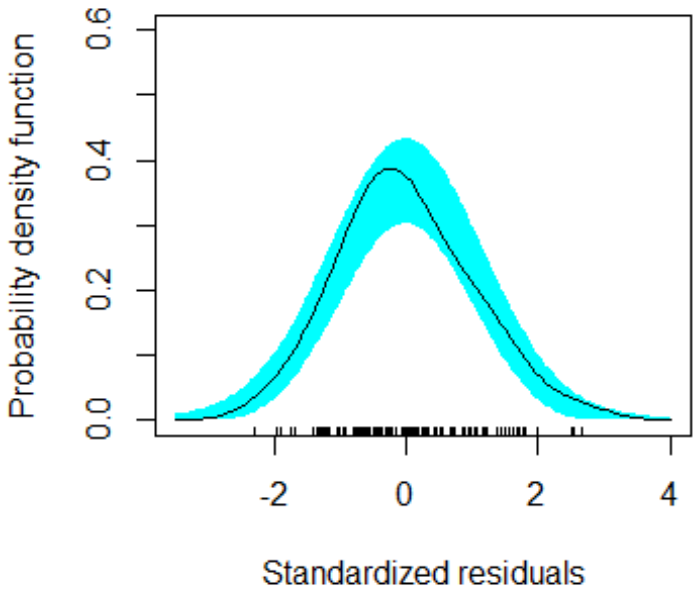

Fig 4: A density plot of the residuauls form the fitted regression model

A horizontal lines at Y=0 shows the expected mean residual under the linearity assumption. The loess line (blue) and uncertainty bands (grey shaded area) indicate that the average coditional residuals for each of the fitted values are not statistically different from 0. Therefore, the linearity assumption seems tenable. Since the actual density curve of the residuals lies within the confidence envelope, we conclude that the normality assumption seems reasonably met. The homogeneity assumption seems likely OK. However, there is a little fanning of residuals in the midle of this residual plot.

Question 12

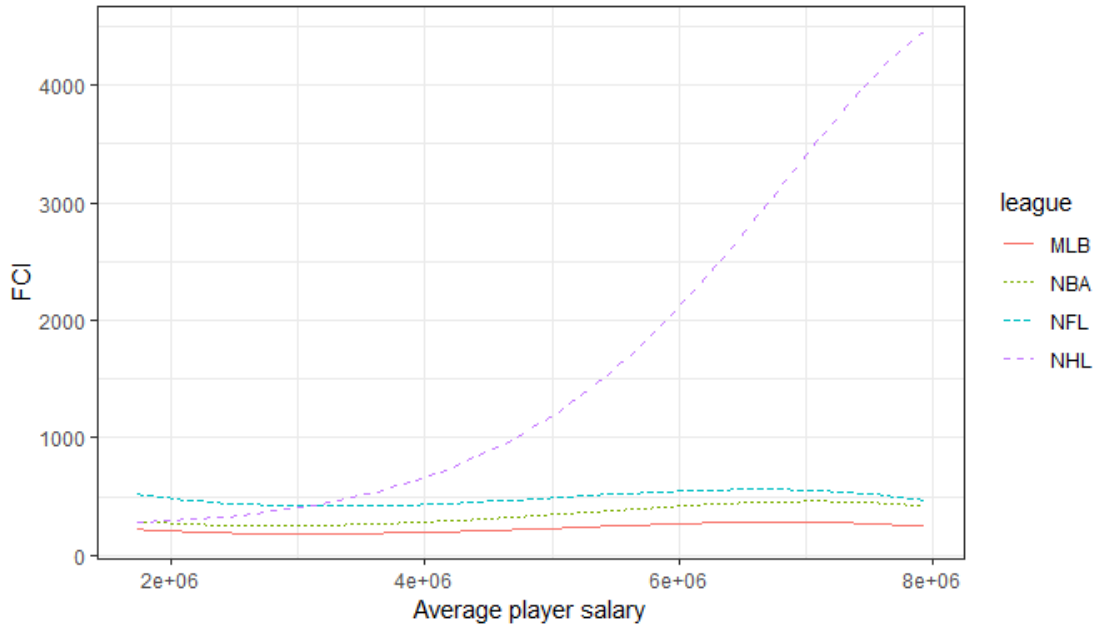

Fig 5: The scatterplot of the final adoptted model to interpret the effects of sporing league and average player salary on FCI

Question 13

Based on the figure, it is appareent the MLB leagues have the lowest level FCI than other leagues. NFL leagues' predicted FCI is higher than NBA and MLB leagues' FCI. NHL leagues' FCI is pretty interesting. In the case of lower average player salaries, NHL leagues FCI is lower than NFL leagues' FCI. However, the estimated FCI for NHL league' increases exponentially after its intersection with NFL league' and at higher average salaries. This indicates the interaction effct. It means that the effect of NHL league on FCI depends on the level of average player salaries. NHL leagues' FCI has the highest predicted value compared to other leagues.

Question 14

Baed on the figures, it is visible that the there is a non-linear effect of average player salary on FCI for different leagues.In case of NHL league, the avearge FCI increases with the average increase in player salary. It indicates the interaction effect that is the effect of avearge player salary on FCI depends on league. FOr, NHL league, the highest average salary has the highest average FCI. For the NFL, MLB, and NBA leagues, there is a directional change in average FCI with respect to the average player salary. For these three leagues, the average FCI decreases with lower player salary. After that the average FCI increases and at higher average salaries, the average FCI decreases again for these three leagues.