

Sports Pay Comparison

Jim Turman

October 27, 2016

Background

This statistical analysis is done to find if there is a difference in pay between the top 15 salaries in the NFL (National Football League), MLB (Major League Baseball) and the NBA (National Basketball Association). I perform below an ANOVA to test my hypothesis that there is a statistically significant difference between the average pay of the top 15 salaries in each sport.

Summary Statistics

Below are the summary statistics for the three major sports in the United States, the NFL, NBA and MLB.

```
## $mlb
##          sum          mean          median
##      "400.6"  "26.7066666666667"  "25"
##          mode          var          sd
##      "numeric"  "13.6678095238095"  "3.69700007084251"
##          n          kurtosis          skew
##      "15"  "-0.378451593384341"  "0.922506092204073"
##          max          min          range
##      "34.5"          "22.7"          "11.8"
##      quartiles.0%      quartiles.25%      quartiles.50%
##      "22.7"          "24.5"          "25"
##      quartiles.75%      quartiles.100%          iqr
##      "28"          "34.5"          "3.5"
##
## $nba
##          sum          mean          median
##      "356.5"  "23.7666666666667"  "23.2"
##          mode          var          sd
##      "numeric"  "6.93095238095238"  "2.6326701998071"
##          n          kurtosis          skew
##      "15"  "0.957128923277273"  "1.25151000102872"
##          max          min          range
##      "30.9"          "21.2"          "9.7"
##      quartiles.0%      quartiles.25%      quartiles.50%
##      "21.2"          "22.1"          "23.2"
##      quartiles.75%      quartiles.100%          iqr
##      "24.8"          "30.9"          "2.7"
##
## $nfl
##          sum          mean          median
##      "390.6"          "26.04"          "23.3"
##          mode          var          sd
##      "numeric"  "65.1182857142857"  "8.06959018254866"
##          n          kurtosis          skew
##      "15"  "1.80289829518881"  "1.61468849924971"
```

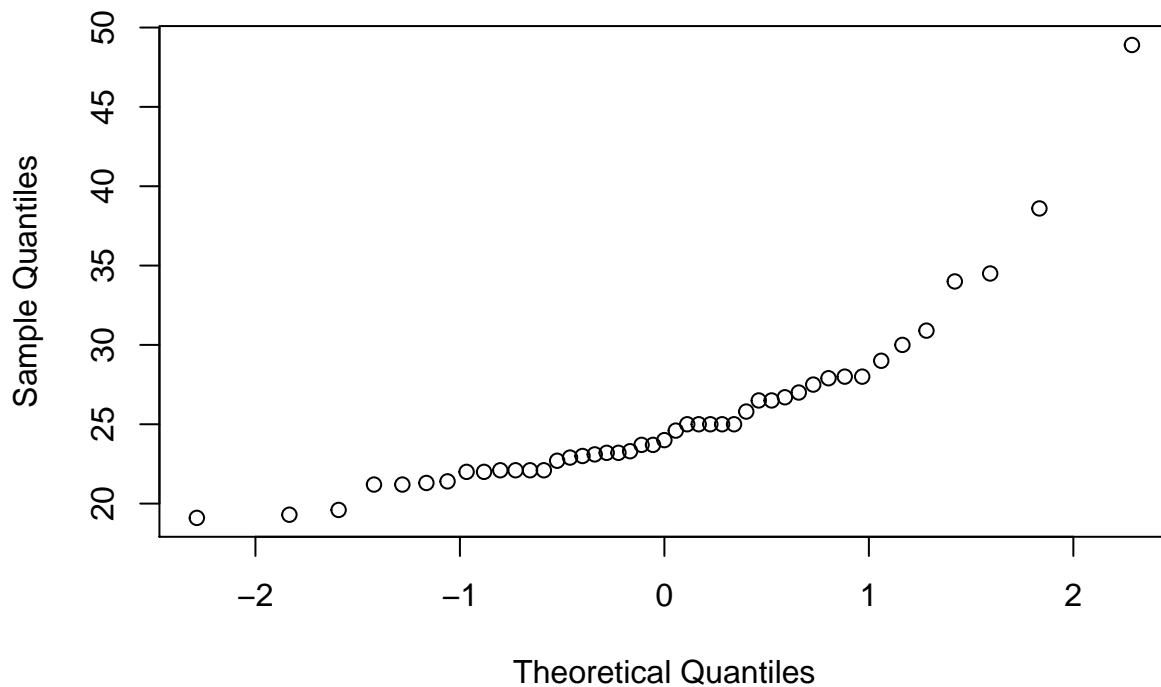
```
##           max           min           range
##      "48.9"      "19.1"      "29.8"
##   quartiles.0%   quartiles.25%   quartiles.50%
##      "19.1"      "21.7"      "23.3"
##   quartiles.75%   quartiles.100%           iqr
##      "27.45"      "48.9"      "5.75"
```

Shapiro-Wilk Test

Below is the Shapiro-Wilk test for normality. To test the hypothesis that the data is normally distributed I am using a .05 level of significance. If the P value is less than .05 then the null hypothesis that the data are normally distributed will be rejected. Because my P value is 1.474e-06 I must reject the null hypothesis and conclude that my data are not normally distributed. To confirm this in a visual below the Shapiro-Wilk results is a plot showing the data points according to a standard normal distribution curve. If they were normally distributed they would look like a bell curve centered on the zero point of the x axis.

```
##
##  Shapiro-Wilk normality test
##
## data:  tot
## W = 0.79041, p-value = 1.474e-06
```

Normal Q-Q Plot



Anova Hypothesis Test

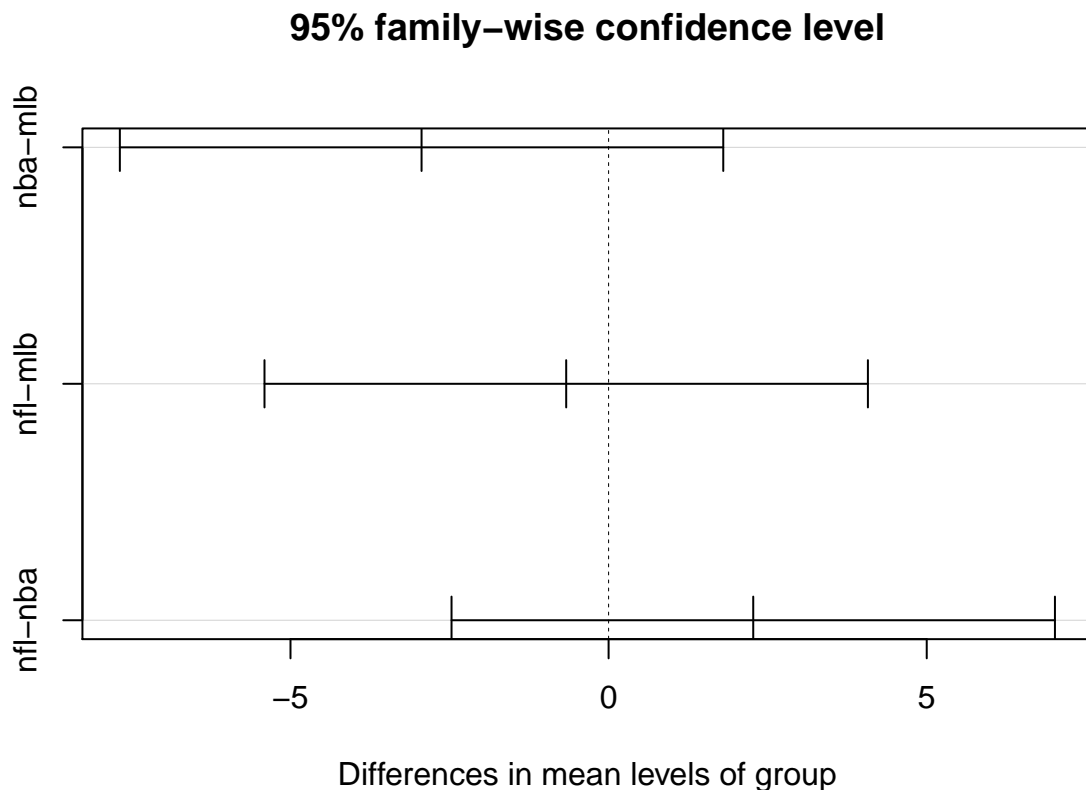
This test is to decide whether or not there is a statistically significant difference between the average of the top 15 salaries of the three sports leagues. For this test I again use the standard .05 level of significance to determine whether or not to reject the null hypothesis that there is no significant difference between the average top 15 salary's of the different leagues. Because my P value is $.298 > .05$ the data does not suggest that there is a statistically significant difference between the average top salary based on league.

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      2    71.3    35.64   1.247  0.298
## Residuals 42 1200.0    28.57
```

Tukey Post Hoc Test

This test is to determine which means may have a difference between each other. The left hand column shows the two leagues that are being compared and the right hand shows the p value. The lower the P value the more significant the difference between the means of the leagues. I use a significance value of .05 to determine if the difference is statistically significant. This test shows that the least difference between averages is NFL to MLB and the largest difference is between NBA and MLB. Though this is interesting the P values are still greater than .05 so I cannot say that there is a statistically significant difference between the average top 15 salaries of these leagues.

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = fit)
##
## $group
##           diff          lwr          upr          p adj
## nba-mlb -2.9400000 -7.681963  1.801963  0.2983102
## nfl-mlb -0.6666667 -5.408629  4.075296  0.9378129
## nfl-nba  2.2733333 -2.468629  7.015296  0.4805387
```



Levene's Test

This test is used to determine the homogeneity of variance for the three leagues. I use an alpha value of .05 to determine the level of significance. If P is less than .05 then I can conclude that I am 95% certain that the variance between the leagues salaries is significantly different. The P value for the group is .1229 > .05 so I cannot say that there is a statistically significant difference in variance between the leagues.

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 2   2.205 0.1229
##      42
```

Conclusion

Through the findings of the tests above I can conclude that there is no statistically significant difference between the average salaries of the top 15 players in each league. What this really means is that the top 15 salaries of the NFL, NBA and MLB are very similar to each other. If people are going to choose between sports to play professionally simply based on pay, there is no real difference for the highest paid athletes. To find the difference for the average pay of athletes in each league the sample sizes could be expanded for each league, because of the difficulty of obtaining that data I stuck with the top 15 salaries.