What is the Relationship between Offensive Rebounding and Wins?

Introduction:

Teams have a choice between aggressively chasing offensive rebounds, or conservatively dropping back to prevent giving up easy points in transition defense. This topic has been analyzed by those close to the NBA.(1) There is no definitive consensus yet on which approach is more effective, and several teams have already responded by sending more or fewer players towards offensive rebounds than in previous seasons.(2) In the NBA, what is the relationship between a team's league-adjusted offensive rebound rate and a team's record? Is it more common in strong teams than would be expected due to chance? Knowing this relationship would be useful to help construct and evaluate rosters and coaching strategies. Teams that emphasize or de-emphasize offensive rebounding could use this info to gain a competitive advantage and to raise the value of their franchise.

Data:

Basketball-reference collects box-score statistics for all NBA games and categorizes them by regular seasons when the data is available. Using the above link, I copied and pasted all cases since the start of the 1973-1974 season onto a .csv and .xlsx file.(3) Each case is an NBA team's stats during one season. The potential explanatory variable Team OREB% (the formula is ORB / (ORB + Opp DRB) is the percentage of the team's missed shots that they rebounded successfully. It is a continuous numerical variable. However since we are comparing teams from different eras, it would be more appropriate to adjust this variable by a team's average from the year. The reason League Adjusted Team OREB% is used instead of total rebounds, or rebound percent is to control for team differences in pace or league differences due to changes in rules and officiating over the years. I downloaded the average team's season statistics from basketball-reference and saved it both as a new tab on our .xlsx file, and as a new .csv file.(4) I used the below VLookup formula to copy the league average OREB% onto our list of cases as a new variable. Then with the league average from that year listed for every team, I divided that team's OREB% by the league average OREB%. League Adjusted OREB% is the percentage of the team's missed shots that they rebounded successfully relative to the league average, and it is a continuous numerical variable.

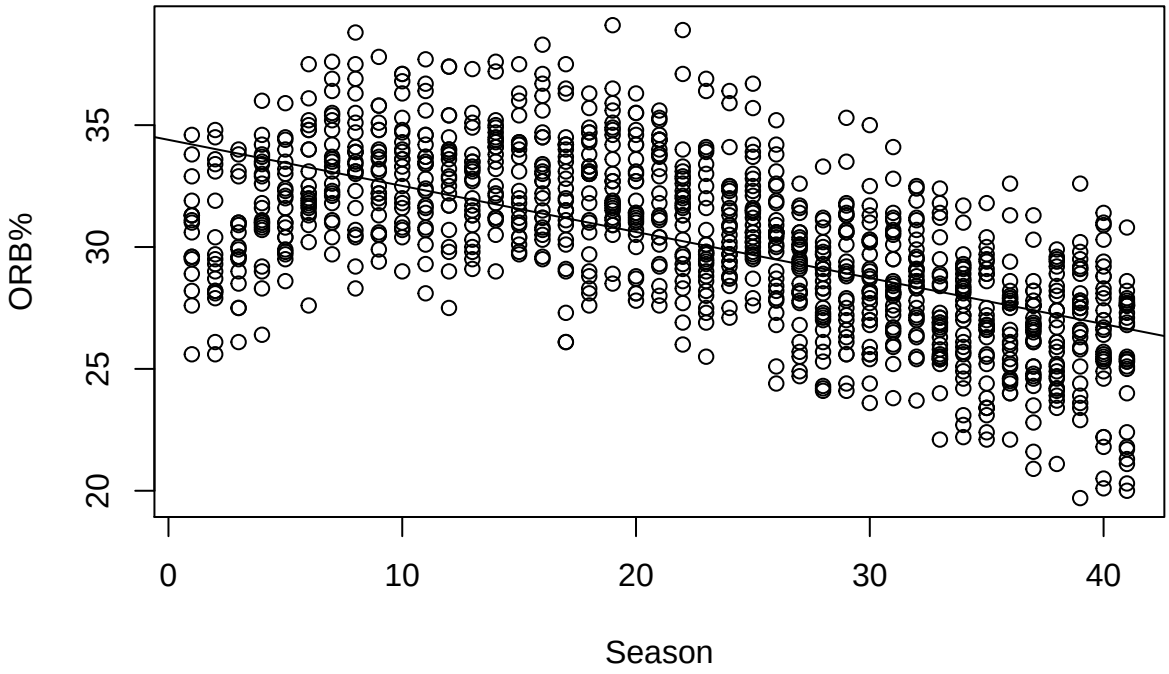
The response variable of interest is Team Strength, which is an ordinal categorical variable containing four levels of Won-Lost Percentage. These four levels broadly separate teams by their chances at a championship, and their performance on the season. The majority of teams with the below range of winning percentage tend to fall into these categories. The Won-Lost Percentage (The formula is W / (W + L) is separated into the below four levels: Elite: won-lost percentage $>$ 66% (rank 3) Playoff: won-lost percentage 50$<=$ 66% (rank 2) Mediocre: won-lost percentage 34.2$<=$ 50% (rank 1) Lottery: won-lost percentage $<$ 34.2% (rank 0) After reading our table into R, the teams were categorized into each of these four levels via subset. This data is collected from recorded events. No variables were controlled or assigned randomly, thus it is an observational study and it cannot be used to infer causality. The random sampling method was used. All NBA team seasons make up the population of interest.

Exploratory data analysis:

I originally intended to find the effect of OREB% on win percentage. However I noticed a consistent downward trend in OREB% over the past 20 years. Some of the reasons for this include the upward trend in 3 point attempts, which are difficult for teams on offense to rebound.(2)

```
#plot oreb, season + linear model
nba$Season <- as.numeric(nba$Season)
```

```r
plot(nba$ORB ~ nba$Season, xlab = "Season", ylab = "ORB%")
reb.win.model <- lm(nba$ORB ~ nba$Season)
abline(reb.win.model)
```



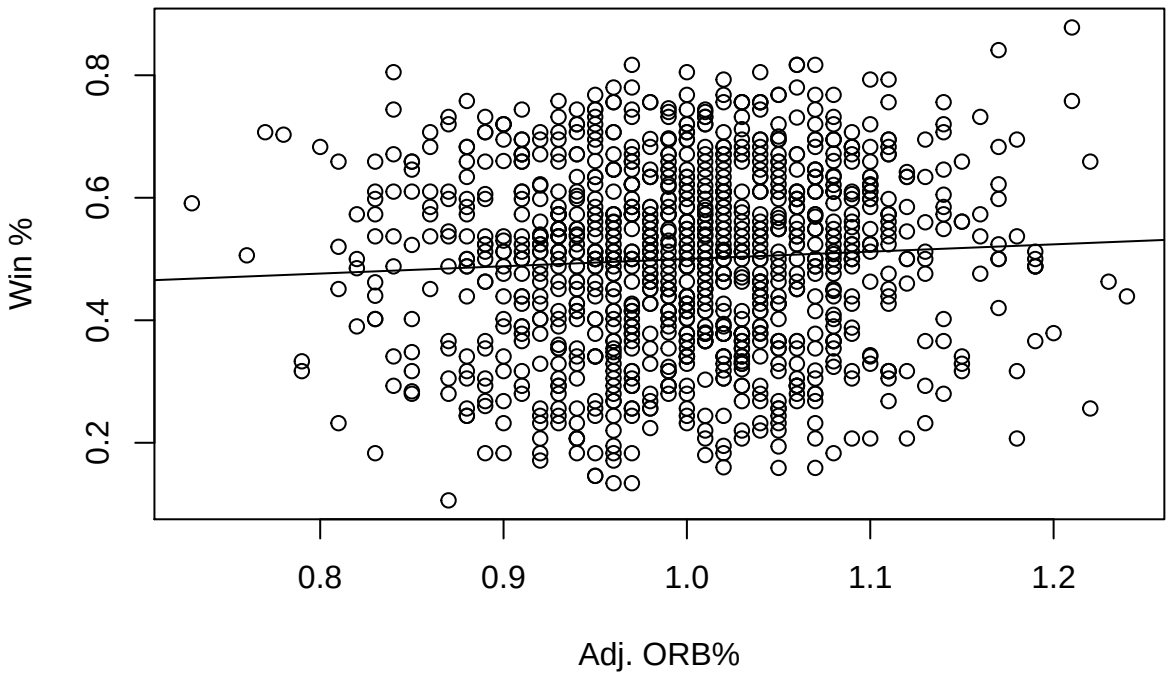

```r
summary(reb.win.model)
```

```
##
## Call:
## lm(formula = nba$ORB ~ nba$Season)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.608 -1.807  0.105  1.805  8.660
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.39652    0.18204   188.9   <2e-16 ***
## nba$Season  -0.18893    0.00717   -26.3   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.71 on 1071 degrees of freedom
## Multiple R-squared:  0.393,  Adjusted R-squared:  0.393
## F-statistic:  694 on 1 and 1071 DF,  p-value: <2e-16
```

Because this looks like a confounding variable, I created a new variable by dividing each team/case's OREB% by the season average OREB% (acquired via Basketball-Reference).4 After this adjustment, OREB% and season are independent. This variable will be referred to as adjusted offensive rebound or ADJORB%.

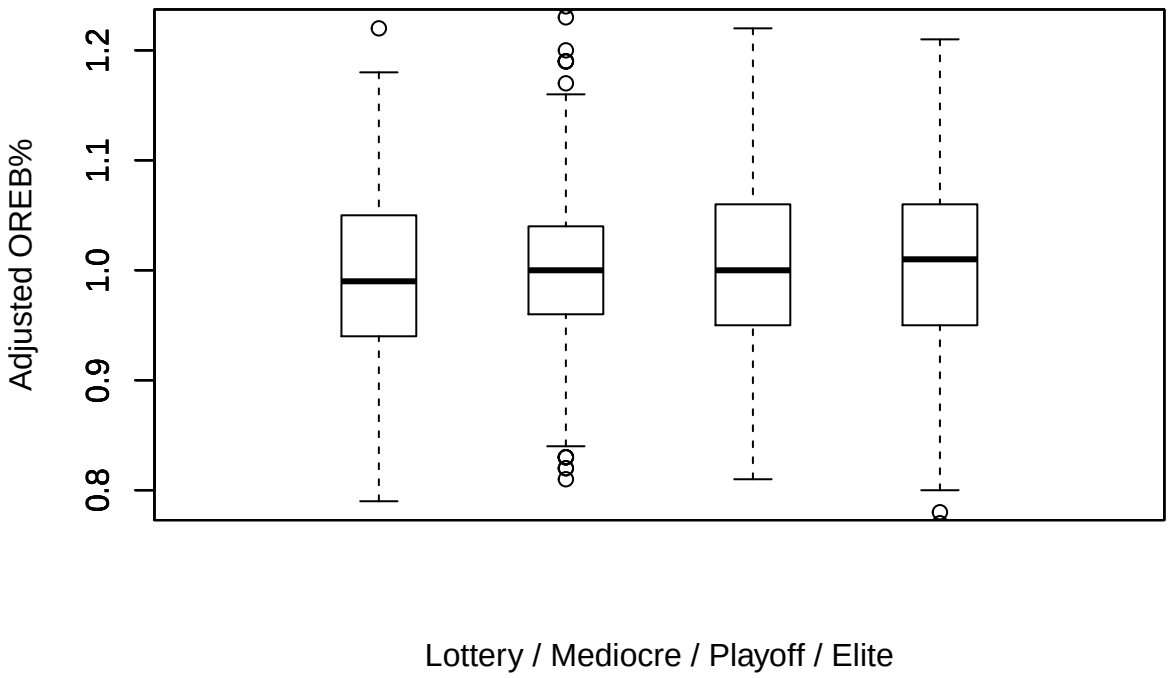
However, plotting this variable against winning percentage shows that the relationship between is certainely not linear.

```
plot(nba$ADJ.ORB, nba$W.L., xlab = "Adj. ORB%", ylab = "Win %")
adj.win.model <- lm(nba$W.L. ~ nba$ADJ.ORB)
abline(adj.win.model)
```



Due to this, I decided to categorize winning percentage into 4 levels which correspond to a team's general strength or expectations. The answer this should bring - do better teams tend to gather more offensive rebounding?

```
# seperate y = win% into 4 levels
elite <- subset(nba, nba$W.L. > .66)
playoff <- subset(nba, nba$W.L. >= .5 & nba$W.L. <= .66)
mediocre <- subset(nba, nba$W.L. >= .342 & nba$W.L. < .5)
lottery <- subset(nba, nba$W.L. < .342)
# horizontal plot oreb by team quality
boxplot(lottery$ADJ.ORB.,  xlab = "Lottery / Mediocre / Playoff / Elite",
        at = 1, xlim = c(0,5), ylab = "Adjusted OREB%")
boxplot(mediocre$ADJ.ORB.,  at = 2, add=TRUE)
boxplot(playoff$ADJ.ORB.,  at = 3,  add=TRUE)
boxplot(elite$ADJ.ORB.,  at = 4, add=TRUE)
```



Adjusted OREB% appears to have a very weak relationship to each level of this team strength variable. However since the variance in Adjusted OREB% is very small, it could be a larger relationship than it appears. More inference methods will need to be employed to tell whether there is a relationship and if so, how strong this effect is.

Inference:

There are two competing claims: 1. "There is no relationship between Adj. OREB rate and team strength" The observed difference in proportions is simply due to chance. H0: Null hypothesis.

2. "There is a relationship between Adj. OREB rate and team strength" The observed difference in proportions is not due to chance. HA: Alternative hypothesis.

To find out, we want to use Analysis of Variance in order to test if there is a statistically significant difference between the mean Adjusted OREB rate for any of the four levels of team strength. If the Offensive Rebound rate varies sufficiently more between teams in different win brackets than it varies with teams in the same win bracket, we can assert a relationship with some confidence. In order to use the One-Way ANOVA method, we must first check these observations for i) independence, ii) approximate normality, and iii) equal variance.

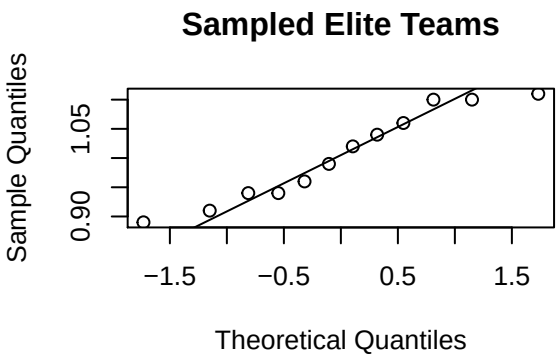
Conditions: Using the entire population of teams dating back to 1973-1974 would not satisfy independence. This is because any sample greater than 10% of the population will have observations dependent on each other. In this case, take for example a successful team that has mostly kept the same players for 5 years in a row. These 5 cases would be dependent on each other, and their win percent and level of offensive rebounding would over-represent the trend and act as leverage points skewing the rest of the data. Taking a random sample of 90 teams would help satisfy the requirement that all observations be independent (it is important to use "set.seed(1)" before taking the sample so the data is reproducible). We would expect independence

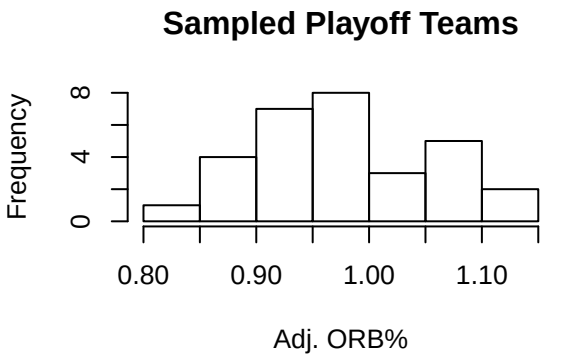
between the team strength groups as well because no single team can appear in multiple team strength levels, and because an NBA schedule is reasonably balanced so the same teams do not always play each other.

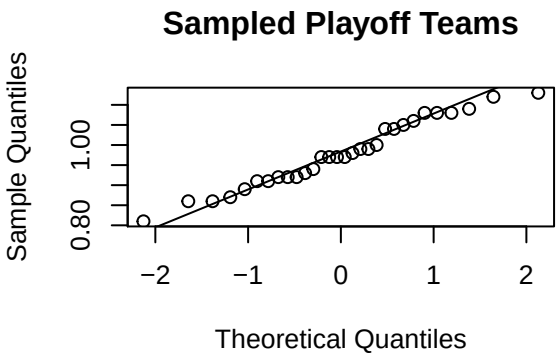
```
set.seed(1)
samp <- nba[sample(nrow(nba), 80), ]
#3a. create 4 levels from samp

s.elite <- subset(samp, samp$W.L. > .66)
s.playoff <- subset(samp, samp$W.L. >= .5 & samp$W.L. <= .66)
s.mediocre <- subset(samp, samp$W.L. >= .342 & samp$W.L. < .5)
s.lottery <- subset(samp, samp$W.L. < .342)

samp$team.quality <- 0
samp$team.quality[samp$W.L. > .66] <- 3
samp$team.quality[samp$W.L. >= .5 & samp$W.L. <= .66] <- 2
samp$team.quality[samp$W.L. >= .34 & samp$W.L. < .5] <- 1
samp$team.quality[samp$W.L. < .34] <- 0
```

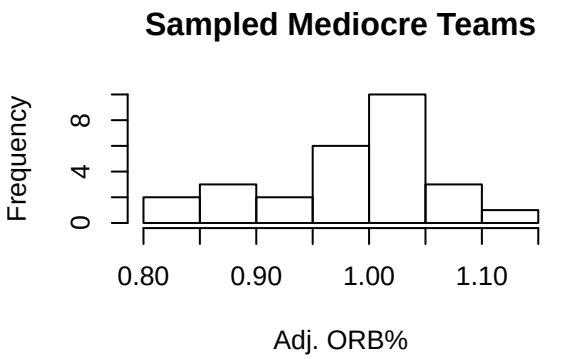
The samples' histograms appear approximately normally distributed. Checking by the normality probability plot, the ends of the data appear fatter, or a bit more extreme than would be predicted, but overall it is approximately normal. However it is very difficult to distinguish whether a dataset is normal or not with only roughly 20 cases per level.

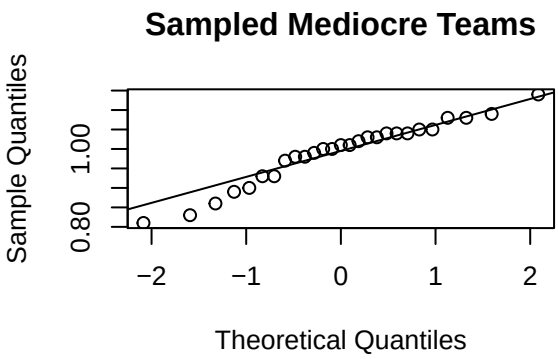

Sampled Elite Teams
Frequency
0.0
1.5
3.0
0.85
0.95
1.05
1.15
Adj. ORB%

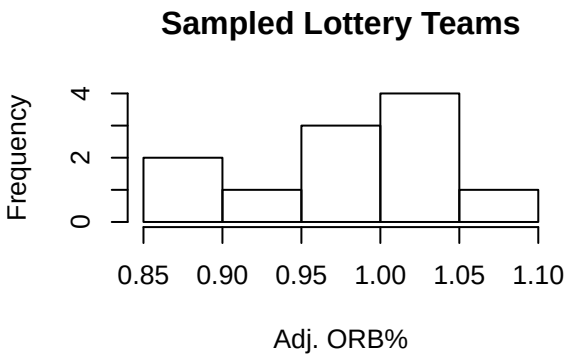

Sampled Elite Teams
Sample Quantiles
0.90
1.05
−1.5
−0.5
0.5
1.5
Theoretical Quantiles

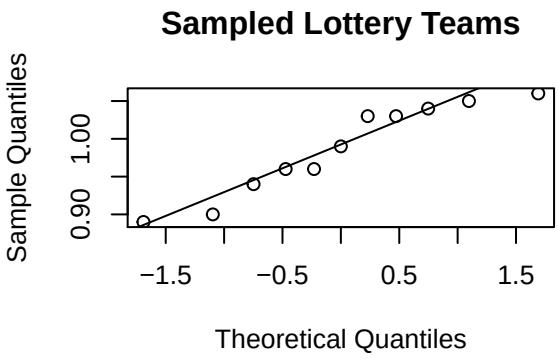

Sampled Playoff Teams
Frequency
0
4
8
0.80
0.90
1.00
1.10
Adj. ORB%


Sampled Playoff Teams
Sample Quantiles
0.80
1.00
−2
−1
0
1
2
Theoretical Quantiles


Sampled Mediocre Teams
Frequency
0
4
8
0.80
0.90
1.00
1.10
Adj. ORB%

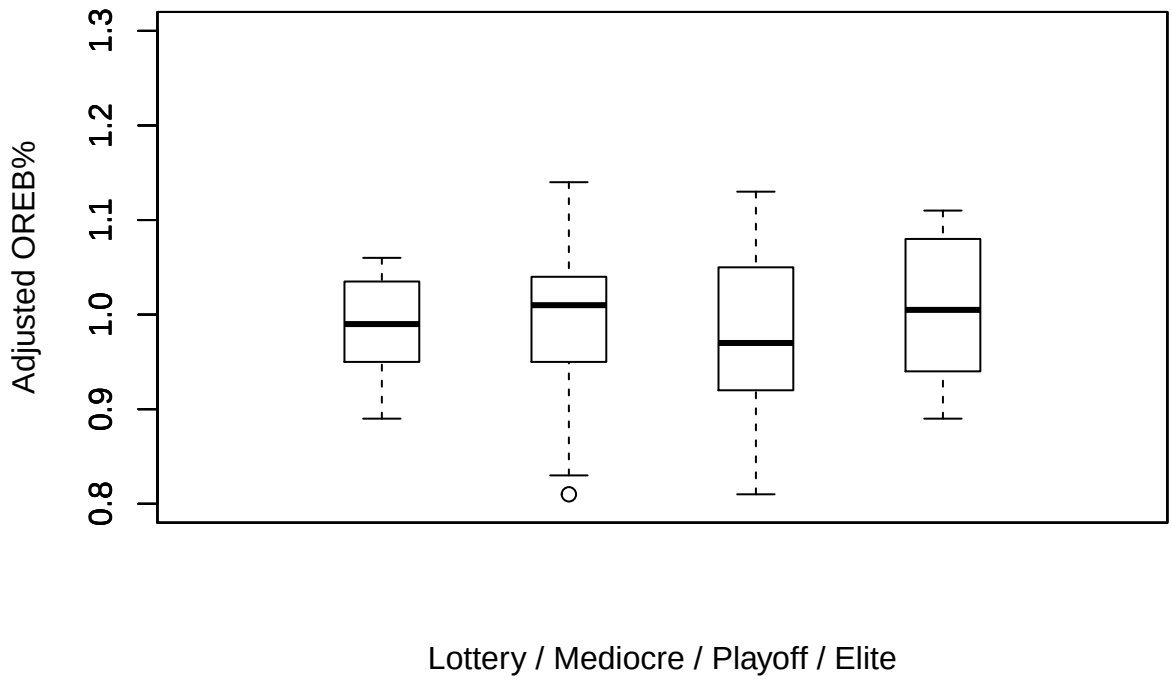

Sampled Mediocre Teams
Sample Quantiles
0.80
1.00
−2
−1
0
1
2
Theoretical Quantiles


Sampled Lottery Teams
Frequency
0
2
4
0.85
0.90
0.95
1.00
1.05
1.10
Adj. ORB%


Sampled Lottery Teams
Sample Quantiles
0.90
1.00
−1.5
−0.5
0.5
1.5
Theoretical Quantiles

The last condition needed to ensure a proper use of ANOVA is homoscedasticity, a constant variance in rebounding across the four levels of team strength. This was the case with the entire population, but at first glance it does not appear to be the case for this sample.

```
#3. variance is constant across levels

boxplot(s.lottery$ADJ.ORB.,  xlab = "Lottery / Mediocre / Playoff / Elite",
        at = 1, xlim = c(0,5), ylab = "Adjusted OREB%", ylim = c(.8, 1.3))
boxplot(s.mediocre$ADJ.ORB.,  at = 2, add=TRUE)
boxplot(s.playoff$ADJ.ORB.,  at = 3,  add=TRUE)
boxplot(s.elite$ADJ.ORB.,  at = 4, add=TRUE)
```



The Levene Test is one way of asserting whether the variances are similar. If we assume that the variances are equal in each group, then the probability of seeing data this or more extreme is 57% per the test below. So it would be safe to assume that there is constant variance across these groups.

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value Pr(>F)
## group  3    0.42   0.74
##       76
```

ANOVA: Now that we've checked our conditions, we can use the ANOVA method to compare these four levels, executed below.

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## samp$team.quality  3  0.007 0.00246    0.39   0.76
## Residuals         76  0.478 0.00629
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = samp$ADJ.ORB. ~ samp$team.quality)
##
## $`samp$team.quality`
##          diff      lwr     upr  p adj
## 1-0  0.011214 -0.06550 0.08793 0.9806
## 2-0 -0.003667 -0.07971 0.07238 0.9993
## 3-0  0.023000 -0.06617 0.11217 0.9053
## 2-1 -0.014881 -0.06960 0.03984 0.8911
## 3-1  0.011786 -0.06007 0.08364 0.9730
## 3-2  0.026667 -0.04447 0.09780 0.7585
```

This tells us that there is less offensive rebounding variance across team strength conditions than there is between teams of the same team strength. Given that the null hypothesis that there is no relationship between these variables, the probability of this F-value occurring is 58%. Since we set a significance level of 5%, we fail to reject the null hypothesis that there is a statistically significant difference between team strength levels given our offensive rebounding data. We can go further and check the differences between each pair of team strength levels. As you can see below, given the assumption of there being no difference in offensive rebounding rates, there is a very high probability of the data or more spread-out data occurring. In other words, given this data, there is a 75% chance that there is no difference between the offensive rebounding rates of elite teams and lottery teams. We have failed to reject the null hypothesis and must conclude that there is no evidence for the offensive rebounding rate affecting the winning percentage of teams.

Conclusion:

This analysis does not suggest any relationship between collecting offensive rebounds and team strength. However, more research into the effects of rebounding on winning would be needed to make any conclusive correlational statements. One restraint of this type of analysis is that the population of team seasons is limited to roughly 1050 teams dating back to 1974. Several of the four groups have small populations, which limits the sample size we can take to maintain independent samples. Some group samples are less than 20, which limits the power to make strong inferences and it limits the certainty that the assumption that the samples are normal and non-skewed. If this analysis could be done on a game-to-game basis instead of by season, that would increase the sample and could help assert how often better offensive rebounding teams won, all else equal.

References:

1. Wiens, Jenna, et al. "To Crash or Not To Crash: A quantitative look at the relationship between offensive rebounding and transition defense in the NBA". MIT Sloan Sports Analytics Conference 2013. March 2013.
2. Lowe, Zach. "Party Crashers: Debunking the Myths of Offensive Rebounding and Transition Defense." Grantland. N.p., 12 Sept. 2013. Web. 22 Apr. 2014. http://grantland.com/the-triangle/party-crashers-debunking-the-myths-of-offensive-rebounding-and-transition-defense/.
3. Kubatko, Justin. Team Season Finder. Basketball-Reference.com - Basketball Statistics and History. http://www.basketball-reference.com/. 18 Mar. 2014. http://bkref.com/tiny/dUJN5
4. Kubatko, Justin. NBA League Averages. Basketball-Reference.com - Basketball Statistics and History. http://www.basketball-reference.com/. 18 Mar. 2014. http://www.basketball-reference.com/leagues/NBA_stats.html#stats::none