Association between Dallas Cowboys Offensive and Defensive Season Rankings and Margin of Victory

Rithvik Saravanan

October 16, 2020

Data

```
# load data
data <- read.csv('./dallas_cowboys_season_data.csv')
mydata <- data[c("MoV", "PtsScoredRank", "PtsAllowedRank", "YdsGainedRank", "YdsAllowedRank")]
```

This dataset includes various season statistics for the Dallas Cowboys football team of the NFL. These statistics are calculated and measured out of the full 16 games that the team plays each season. In this analysis, we are specifically interested in the Dallas Cowboys season rank for points scored, points allowed, yards gained, and yards allowed measured on a scale of 1 to 32 (representing the 32 teams in the NFL).

```
kable(head(mydata))
```

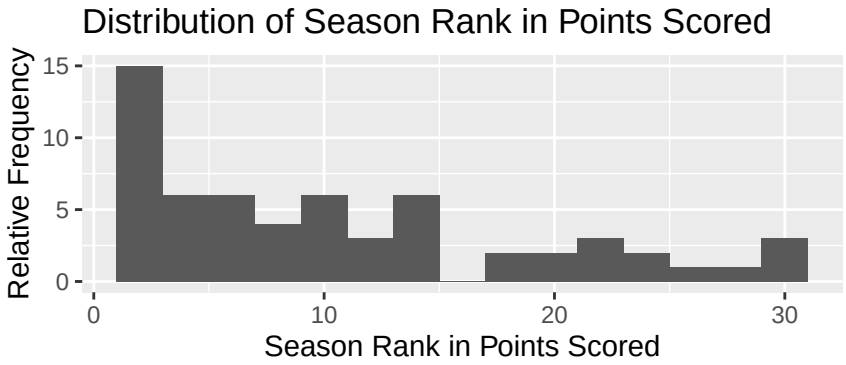
MoV	PtsScoredRank	PtsAllowedRank	YdsGainedRank	YdsAllowedRank
7.1	6	11	1	9
0.9	22	6	22	7
1.4	14	13	14	8
7.2	5	5	5	14
-6.2	31	16	22	17
7.2	5	15	7	19

`PtsScoredRank` indicates the season rank in points scored and `YdsGainedRank` indicates the season rank in yards gained. `PtsAllowedRank` indicates the season rank in points allowed and `YdsAllowedRank` indicates the season rank in yards allowed. `PtsScoredRank` and `YdsGainedRank` are both primarily offensive statistics and `PtsAllowedRank` and `YdsAllowedRank` are both primarily defensive statistics. These rankings are accurate measures of how the team performed over a typical season relative to the other 31 NFL teams.

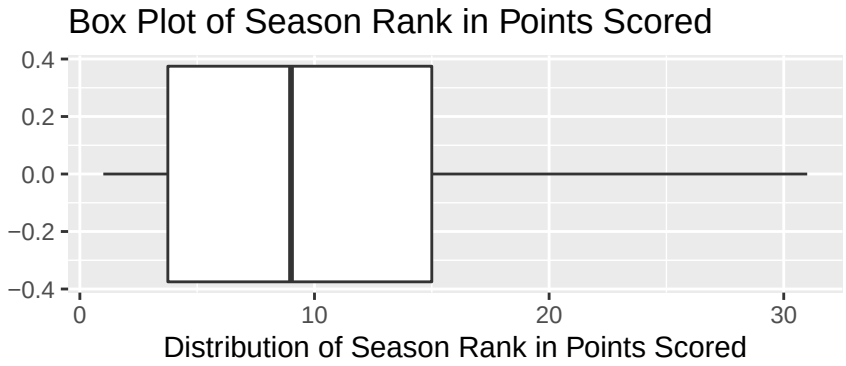
In this analysis, using these predictor variables, we analyze the relationship between offensive and defensive rankings and `MoV`, the Dallas Cowboys average margin of victory over a season. `MoV` can be a positive or negative value and is calculated by summing the score margins of a team's games and dividing by the total number of games played.

Exploratory Data Analysis

```
# visualize predictor variable #1: PtsScoredRank
ggplot(data, aes(x = PtsScoredRank)) +
  geom_histogram(binwidth = 2) +
  labs(title = "Distribution of Season Rank in Points Scored",
       x = "Season Rank in Points Scored",
       y = "Relative Frequency")
```

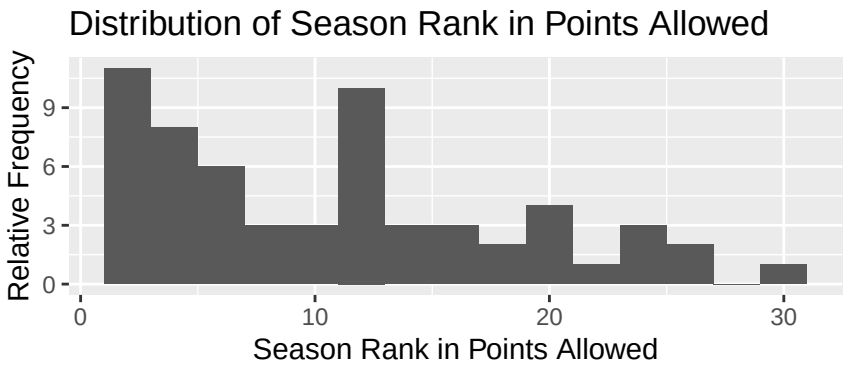



```
ggplot(data, aes(y = PtsScoredRank)) +
  geom_boxplot() +
  coord_flip() +
  labs(y = "Distribution of Season Rank in Points Scored") +
  ggtitle("Box Plot of Season Rank in Points Scored")
```

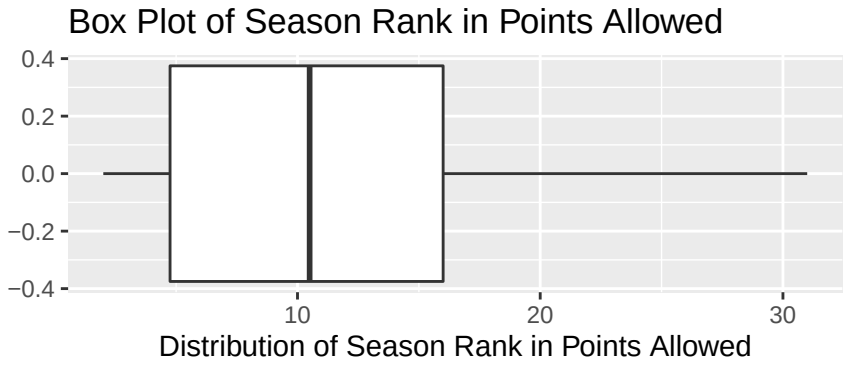


From the histogram, we observe that the frequency distribution of the Dallas Cowboys season rank in points scored (offense) is unimodal and skewed right with most values falling between ranks of 1 and 7. From the box plot, we also note that the distribution is skewed right with an average season rank in points scored of approximately 9 and 50% of all season ranks falling between 4 and 15.

```r
# visualize predictor variable #2: PtsAllowedRank
ggplot(data, aes(x = PtsAllowedRank)) +
  geom_histogram(binwidth = 2) +
  labs(title = "Distribution of Season Rank in Points Allowed",
       x = "Season Rank in Points Allowed",
       y = "Relative Frequency")
```

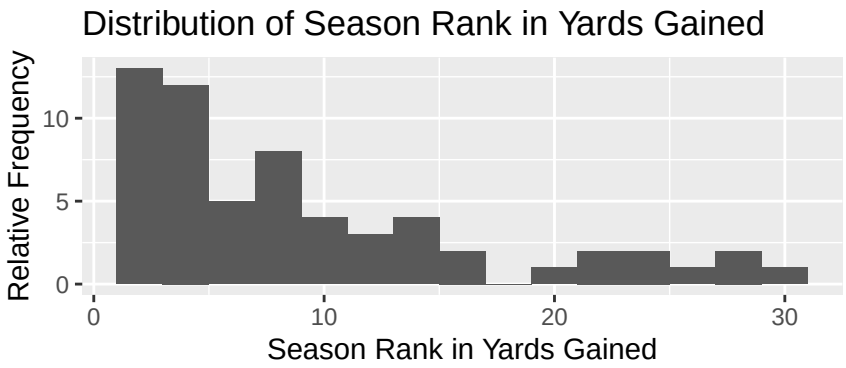



```r
ggplot(data, aes(y = PtsAllowedRank)) +
  geom_boxplot() +
  coord_flip() +
  labs(y = "Distribution of Season Rank in Points Allowed") +
  ggtitle("Box Plot of Season Rank in Points Allowed")
```

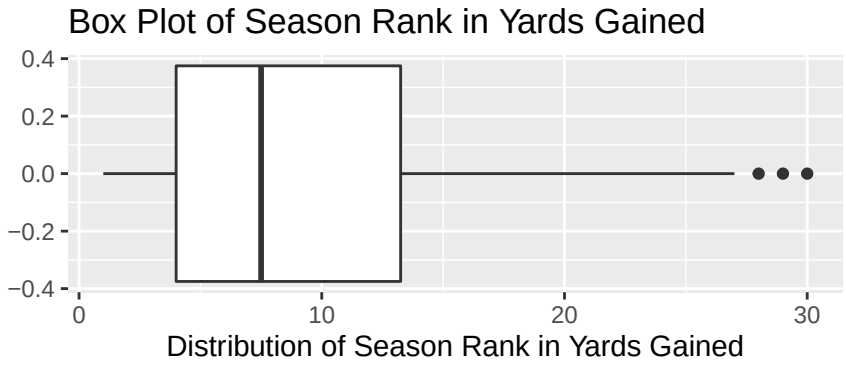


From the histogram, we observe that the frequency distribution of the Dallas Cowboys season rank in points allowed (defense) is bimodal with two separate peaks around ranks of 2 and 12 and skewed right with most values falling between ranks of 1 and 6. From the box plot, we also note that the distribution is skewed right with an average season rank in points allowed of approximately 11 and 50% of all season ranks falling between 4 and 16.

```r
# visualize predictor variable #3: YdsGainedRank
ggplot(data, aes(x = YdsGainedRank)) +
  geom_histogram(binwidth = 2) +
  labs(title = "Distribution of Season Rank in Yards Gained",
       x = "Season Rank in Yards Gained",
       y = "Relative Frequency")
```

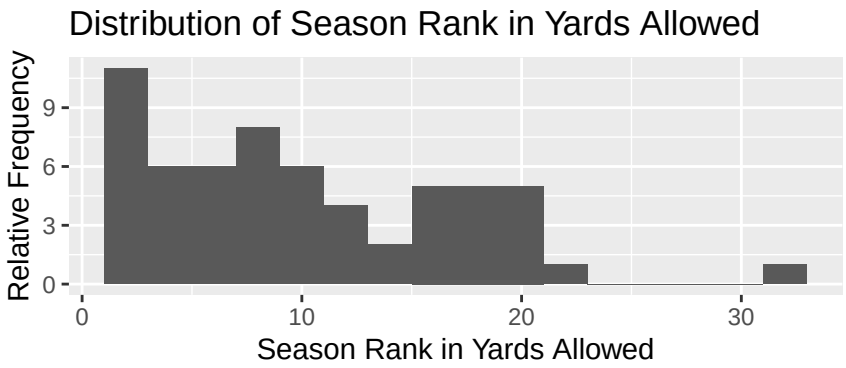



```r
ggplot(data, aes(y = YdsGainedRank)) +
  geom_boxplot() +
  coord_flip() +
  labs(y = "Distribution of Season Rank in Yards Gained") +
  ggtitle("Box Plot of Season Rank in Yards Gained")
```

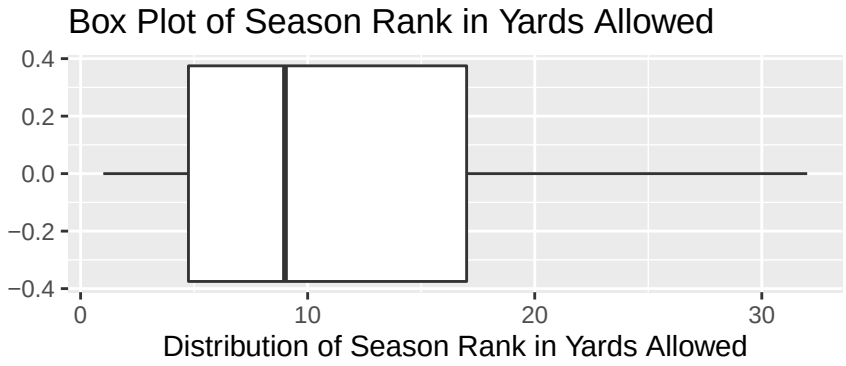


From the histogram, we observe that the frequency distribution of the Dallas Cowboys season rank in yards gained (offense) is unimodal and skewed right with most values falling between ranks of 1 and 4. From the box plot, we also note that the distribution is skewed right with an average season rank in points allowed of approximately 7 and 50% of all season ranks falling between 4 and 13. The box plot also depicts that this distribution has 3 outliers indicating three different seasons where the Cowboys season rank in yards gained was among the worst teams in the NFL.

```r
# visualize predictor variable #4: YdsAllowedRank
ggplot(data, aes(x = YdsAllowedRank)) +
  geom_histogram(binwidth = 2) +
  labs(title = "Distribution of Season Rank in Yards Allowed",
       x = "Season Rank in Yards Allowed",
       y = "Relative Frequency")
```



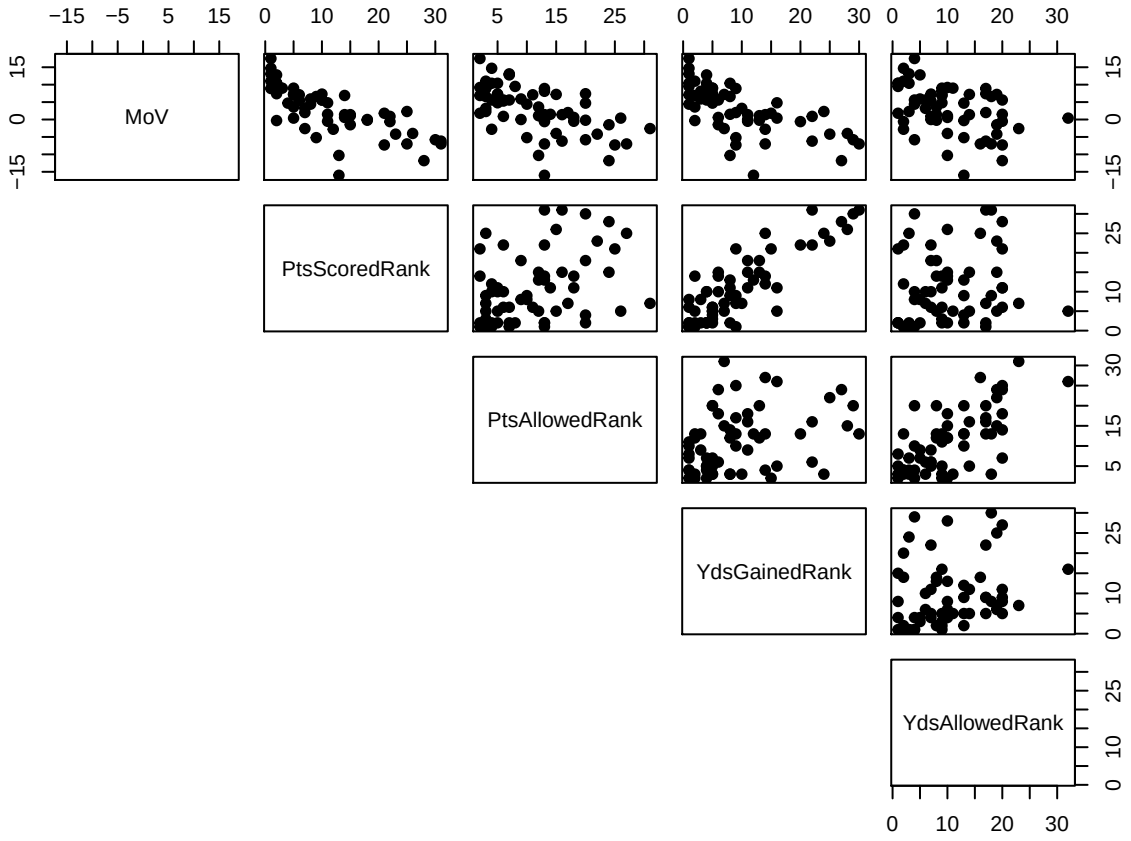

```r
ggplot(data, aes(y = YdsAllowedRank)) +
  geom_boxplot() +
  coord_flip() +
  labs(y = "Distribution of Season Rank in Yards Allowed") +
  ggtitle("Box Plot of Season Rank in Yards Allowed")
```



From the histogram, we observe that the frequency distribution of the Dallas Cowboys season rank in yards allowed (defense) is unimodal and skewed right with most values falling between ranks of 1 and 11. We also note a single data point in the far right of the histogram (more than 10 greater than the closest data point), indicating that the Cowboys allowed close to a league-worst number of yards allowed. From the box plot, we also note that the distribution is skewed right with an average season rank in points allowed of approximately 9 and 50% of all season ranks falling between 5 and 17.

Correlation

```
# relationships between the response and each predictor variable and between predictors
pairs(mydata, pch = 19, lower.panel = NULL)
```




```
# calculate the correlation coefficients for each relationship
cor(mydata)
```

```
##                      MoV PtsScoredRank PtsAllowedRank YdsGainedRank
## MoV            1.0000000    -0.7422503     -0.5990449    -0.6581934
## PtsScoredRank -0.7422503     1.0000000      0.3688277     0.8586953
## PtsAllowedRank -0.5990449     0.3688277      1.0000000     0.3399740
## YdsGainedRank -0.6581934     0.8586953      0.3399740     1.0000000
## YdsAllowedRank -0.4270423     0.1541712      0.6814417     0.2210897
##                YdsAllowedRank
## MoV                -0.4270423
## PtsScoredRank       0.1541712
## PtsAllowedRank      0.6814417
## YdsGainedRank       0.2210897
## YdsAllowedRank      1.0000000
```

From the scatterplot of `PtsScoredRank` and `MoV`, we notice that the form of the association between these two variables is linear, negative, and strong with a correlation of -0.74.

From the scatterplot of `PtsAllowedRank` and `MoV`, we notice that the form of the association between these two variables is linear, negative, and moderate with a correlation of -0.60.

From the scatterplot of `YdsGainedRank` and `MoV`, we notice that the form of the association between these two variables is linear, negative, and moderate with a correlation of -0.66.

From the scatterplot of `YdsAllowedRank` and `MoV`, we notice that the form of the association between these two variables is linear, negative, and moderate with a correlation of -0.43.