

A Recreation of Paul Pepodesta's Linear Regression Analysis

Chelsea Lancaster

6/29/2021



Introduction

All sports require numbers to be recorded as statistics but there is something about baseball that sets its statistics apart from other sports: sports statisticians can and have quantified every aspect of the game of baseball. We used to just look at batting average or earned run average. Now-a-days, we are recording launch angle and spray pattern from the batters and spin rate and launch speed for pitchers. We have given value for different scenarios, so, for example bunting with no one on base is valued differently than bunting with a person on first. This means we are able to quantify every single scenario in the game of baseball. There is even a special name for baseball statistics: sabermetrics. Sabermetrics is the search for objective knowledge about baseball.¹ In other words, sabermetrics is the study that breaks down baseball to quite literally, a science.

Sabermetrics is a relatively new concept to the sport of baseball but statistical analysis has been deployed as long as the game has been played. Everyone seeks to gain an edge in any game, which can arguably be attributed to our quirky human nature. Though, in the 1980's a sport statistician by the name of Bill James popularized sabermetrics by writing several books on baseball history and statistics.² He is most famous

¹Birnbaum, P. (n.d.). A Guide to Sabermetric Research. Society for American Baseball Research. <https://sabr.org/sabermetrics>.

²Beth, A. (2016, March 31). How Esquire Discovered Bill James. Esquire Classic. <https://web.archive.org/web/20161230182832/http://classic.esquire.com/editors-notes/how-esquire-discovered-bill-james/>.

for the Pythagorean Winning Percentage.³ Sabermetrics was then used to revolutionize the entirety of the sports world in the early 2000's by Paul DePodesta and Billy Beane.

Now today, outfielders can be seen taking pieces of paper out of their back pockets to look at the spray patterns of the current batter, batters themselves will immediately go into the dugout and check their launch angle and distance of the ball they just hit, and pitchers are concerned with their launch speed and spins rates. You know what they say, knowledge is power, and more power to them.

Data

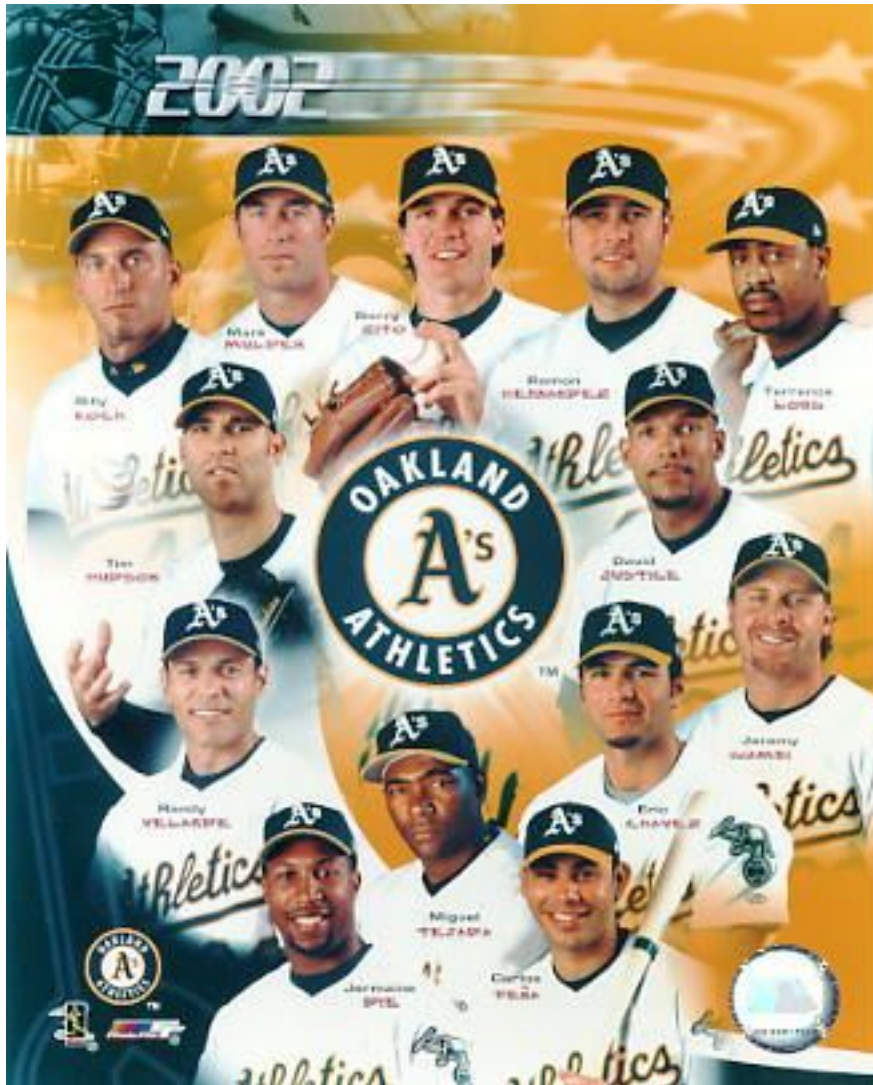
In 2002, the A's lost all their star players to other teams and did not have the budget to buy more. The A's manager, Billy Beane, gets in touch with Paul DePodesta, a graduate in Economics from Harvard University, who joined the A's franchise in 1999.⁴

³Unknown. (2021, June 4). Pythagorean Theorem of Baseball. Pythagorean Theorem of Baseball - BR Bullpen. https://www.baseball-reference.com/bullpen/Pythagorean_Theorem_of_Baseball.

⁴Banerjee, S. (2018, April 16). Linear Regression: Moneyball - Part 1. Medium. <https://towardsdatascience.com/linear-regression-moneyball-part-1-b93b3b9f5b53>.



DePodesta, with the help of linear regression, convinces Beane to build a team based off of on base percentage (OBP) and slugging percentage (SLG) instead of batting average (BA), like all the other teams were basing their scouting off of at the time. Since DePodesta determined that statistics like OBP and SLG were highly correlated with scoring runs (RS) while BA has far less of a correlation, Beane was able to scout for players and get their contracts for a fraction of the price of what it cost to get players with high BA's. It makes sense as I was always told that in order to score runs, you need to get on base. Needless to say, it gave the Oakland A's the edge that season.



The data set is from Kaggle.com.⁵ It contains a set of variables from years 1962-2012 that Beane and DePodesta focused heavily on. For the analysis, I used only the years 1962-2001 to recreate DePodesta's analysis for the Oakland A's. The variables are as follows:

- Team
- League
- Year
- Runs Scored (RS)
- Runs Allowed (RA)
- Wins (W)
- On-Base Percentage (OBP)
- Slugging Percentage (SLG)
- Batting Average (BA)
- Playoffs (binary)
- RankSeason
- Games Played (G)
- Opponent On-Base Percentage (OOBP)
- Opponent Slugging Percentage (OSLG)

⁵Duckett, W. (2017). Moneyball. <https://www.kaggle.com/wduckett/moneyball-mlb-stats-19622012>.

Motivation

The Prime Minister of Great Britain, Benjamin Disraeli, is famous for the saying, “There are three types of lies: lies, damned lies, and statistics.” I noticed early on in my life people often times manipulate data analysis to fit their own narrative and their own agenda rather than doing what is right for the greater good. Furthermore, this turned out to be a nice little exercise with ample opportunity to practice creating detail-intensive data visualizations.

Lastly, I admire DePodesta because he was able to seek and find truth in his data set and answer some specific questions:

- What’s the goal of the team? How does a team make it to playoffs?
- How does a team win more games? How many wins does it take to make it to playoffs?
- How does a team score more runs? How many more runs do we need to score than runs we allow to win 95 games in the regular season.
- How do we score 133 more runs than we allow? Which statistics are significant when trying to score runs?
- What OBP and SLG do we need to achieve a run differential of +133?

Load Libraries

```
# load libraries
library(readr)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Load Data

```
# Load data
moneyball <- read_csv("moneyball.csv")

##
## -- Column specification -----
## cols(
##   Team = col_character(),
##   League = col_character(),
##   Year = col_double(),
##   RS = col_double(),
##   RA = col_double(),
##   W = col_double(),
##   OBP = col_double(),
```

```
## SLG = col_double(),
## BA = col_double(),
## Playoffs = col_double(),
## RankSeason = col_double(),
## RankPlayoffs = col_double(),
## G = col_double(),
## OOBP = col_double(),
## OSLG = col_double()
## )
```

```
head(moneyball)
```

```
## # A tibble: 6 x 15
##   Team League Year   RS   RA   W   OBP   SLG   BA Playoffs RankSeason
##   <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 ARI   NL    2012  734  688  81 0.328 0.418 0.259      0      NA
## 2 ATL   NL    2012  700  600  94 0.32  0.389 0.247      1       4
## 3 BAL   AL    2012  712  705  93 0.311 0.417 0.247      1       5
## 4 BOS   AL    2012  734  806  69 0.315 0.415 0.26       0      NA
## 5 CHC   NL    2012  613  759  61 0.302 0.378 0.24       0      NA
## 6 CHW   AL    2012  748  676  85 0.318 0.422 0.255      0      NA
## # ... with 4 more variables: RankPlayoffs <dbl>, G <dbl>, OOBP <dbl>,
## #   OSLG <dbl>
```

Create Run Differential Column

```
moneyball <- moneyball %>%
  mutate(RD = moneyball$RS - moneyball$RA)
```

Best Practice: Know what your problem is and ask the right questions

I was taught that **the most important notion of any data-driven project is the question being asked**. In 2002, Billy Beane was facing financial troubles and all his best players were taken away. He knew what his problem was: “I don’t have a team that can make it to play offs.” Paul DePodesta loved baseball and therefore knew what questions to ask but he also knew how to extract the answers from the data. Thus, they asked themselves: What is the goal of the team?

Question 1:

What’s the goal of the team?

They decided that getting to the playoffs was good enough because, according to Beane, the post season is based solely on luck.⁶ If they could build a skilled-team throughout the season that could get them to playoffs, then perhaps the scouted-skills of the players would shine through and perhaps win the 2002 World Series. The reasoning was that a regular season is 162 games while the World Series is only 7 games. With decreased games comes an increased chance for players to get lucky – it becomes less about skill and more about luck when the number of games, (sample size), are reduced.

⁶Kirby, N. (2019, May 28). Billy Beane: The most overrated exec in sports history. Bronx Pinstripes | BronxPinstripes.com. <http://bronxpinstripes.com/opinion/billy-beane-the-most-overrated-exec-in-sports-history/>.

Question 2:

How does a team make it to playoffs?

Now that Beane and DePodesta narrowed down their goal to making it to playoffs, it laid the foundation for their next question. The very obvious answer is to win more games, but exactly how many games do they need to win? Thus, finalizing our jumping off point: We will focus on wins (W).

Question 3:

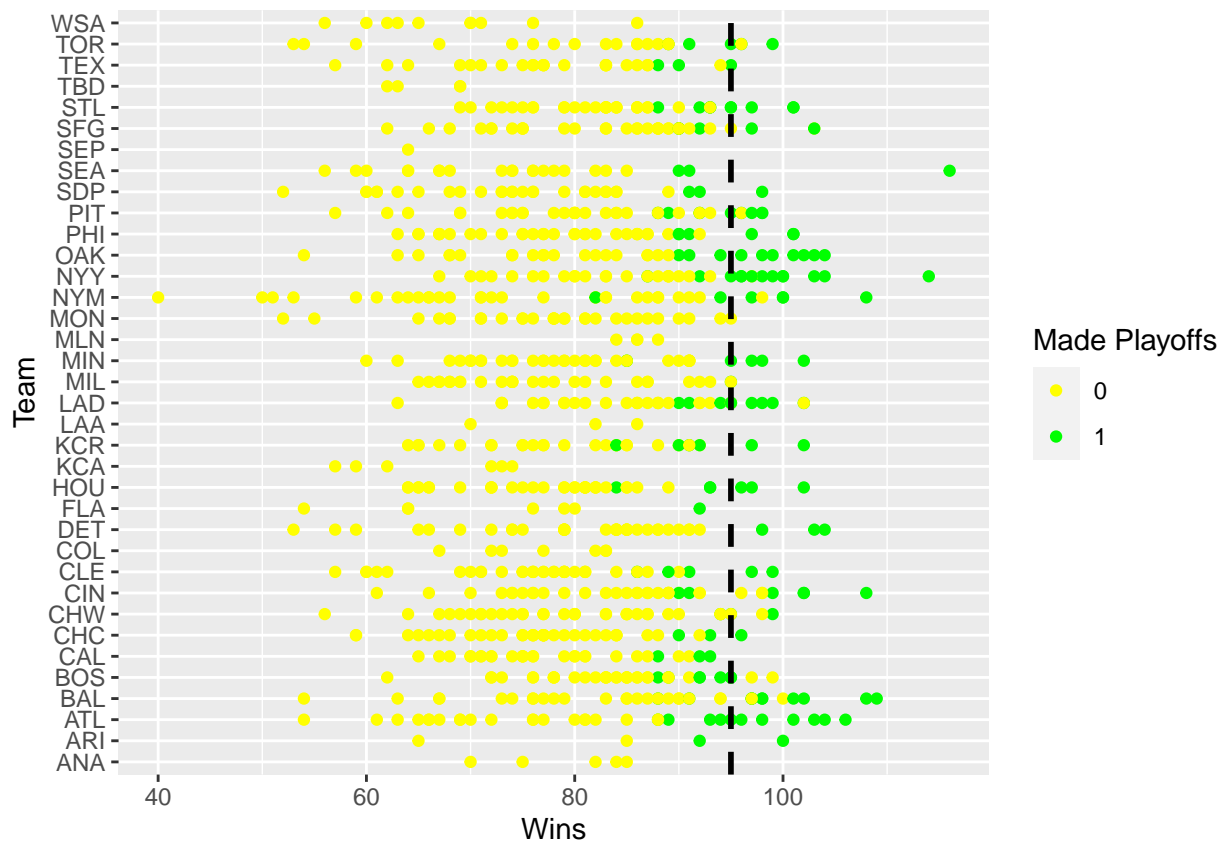
How many wins does it take to make it to playoff?

DePodesta used data from 1962-2001 on to determine the average number of wins (W). First we will need to subset the data so we run only the years 1962-2001 to ensure we get the same results.

```
# subset to use data prior to 2002
moneyball_subset <- moneyball %>%
  subset(moneyball$Year < 2002)
```

Now, the graph of the number of teams who made it to playoffs, (and those teams who did not make it to playoffs), for all MLB teams would look something like this:

```
# create a scatterplot of x = wins, y = teams to see who made playoffs in
library(ggplot2)
playoff_plot <- ggplot(data = moneyball_subset,
  aes(x=W,
      y = Team,
      color = factor(Playoffs))) +
  geom_point() + scale_color_manual(values = c("yellow", "green"),
    name = "Made Playoffs")
playoff_plot + geom_vline(aes(xintercept = 95),
  color = "black",
  linetype = "dashed",
  size=1) + xlab("Wins")
```



In the above graph, ‘1’ in green denotes the teams that made playoffs while ‘0’ in yellow denotes the teams that did not make it to playoffs. The dotted line illustrates the wins (W) threshold at which a team will almost certainly make the playoffs. The dotted line is located at 95 wins on the x-axis.

So how did DePodesta determine the win threshold?

Simply by eyeing the graph. The minimum threshold of teams who make it to playoffs can be visually seen as the graph physically changes from yellow to green. We would be able to see the threshold even if I did not include the dashed line. If the dashed line were placed at, let’s say, 90 games, there would be too many yellow dots to the right of the dashed line. Similarly, if the dashed line were placed at 100 games, there would be too many green dots to the left of the dashed line.

Now we know the answer to question 3: the A’s had to win at least 95 games or more throughout the regular season.

Question 4:

Okay, so the Oakland A’s needed 95 wins (W) to make the playoffs. This leads us to the next question:

How many more runs do we need to score than we allow in order to win 95 games in a regular season?

If runs scored are not the cause of wins (W), in the very least, runs are the currency in which wins are purchased.⁷ A team must score runs to trade in for wins.

Fortunately for the A’s, DePodesta knew how to answer that question with the notion of run differential (RD). According to the MLB glossary, run differential (RD) is determined by subtracting the total number of runs (both earned and unearned) it has allowed (RA) from the number of runs scored (RS). Examining a team’s

⁷Gleeman, A., & Bonnes, J. (2020, May 24). Episode 477: Baseball Stats 101. Baseball Stat 101. broadcast, Minneapolis, Minnesota; Spotify.

run differential (RD) can help to identify teams that are overachieving and teams that are underachieving.⁸ Knowing that, it makes sense DePodesta would use it as a metric in which to measure the A's success.

The following is the run differential equation:

$$RD = RS - RA$$

In layman's terms, the run differential is positive if a team scores more runs than it allows and is negative when a team allows more runs than it scores.

We already created a run differential (RD) column with the entire data set so it is ready to use.

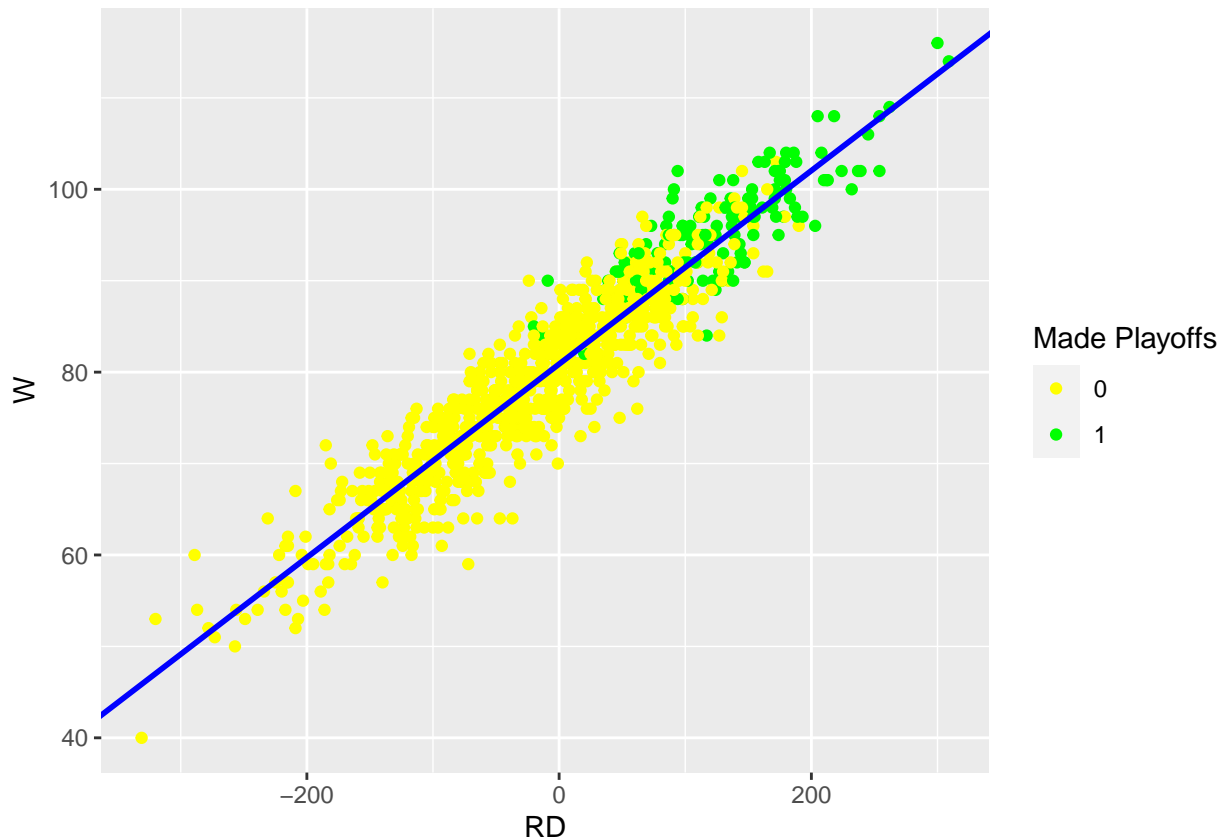
```
#Linear regression model to predict wins needed to go to playoffs
W_reg <- lm(W ~ RD, data = moneyball_subset)
W_reg

##
## Call:
## lm(formula = W ~ RD, data = moneyball_subset)
##
## Coefficients:
## (Intercept)          RD
##    80.8814         0.1058

# create scatterplot to visualize the relationship between x=RD, y = W
rundiff_plot <- ggplot(data = moneyball_subset,
                      aes(x=RD,
                          y = W,
                          color = factor(Playoffs))) +
  geom_point() + scale_color_manual(values= c("yellow", "green"),
                                   name = "Made Playoffs")

rundiff_plot + geom_abline(intercept = 80.8814,
                          slope = 0.1058,
                          color = "blue",
                          linetype = "solid",
                          size = 1)
```

⁸MLB. (n.d.). Run Differential. <https://www.mlb.com/glossary/standard-stats/run-differential>.



The linear regression model above analyzes run differential (RD) and the number of wins (W) the A's have to determine if there is a relationship. The scatter plot visualizes the relationship. The relationship is linear indicating a high linear correlation that is positive and increasing.⁹

In the graph, I included the color codes for playoffs to reinforce our idea that a team with 95 or more wins (W) will almost certainly make it to the playoffs. The residuals in the graph appear to be pretty close to the linear regression line I plotted meaning it appears we have an unbiased estimate and can proceed. Now let us look at the summary of our linear regression.

```
# summary of predicted wins
summary(W_reg)
```

```
##
## Call:
## lm(formula = W ~ RD, data = moneyball_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2662  -2.6509   0.1234   2.9364  11.6570
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  80.881375   0.131157   616.67  <2e-16 ***
## RD           0.105766   0.001297    81.55  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

⁹Unknown. (n.d.). Scatter Diagrams and Regression Lines. Stastics Regression. <http://www.ltcconline.net/greenl/courses/201/regression/scatter.html>

```
##
## Residual standard error: 3.939 on 900 degrees of freedom
## Multiple R-squared:  0.8808, Adjusted R-squared:  0.8807
## F-statistic: 6651 on 1 and 900 DF,  p-value: < 2.2e-16
```

The linear regression equation is:

$$W = 80.881375 + 0.1058(RD)$$

Where 80.881375 is the y-intercept and 0.1058 is the slope. Run differential (RD) is the independent or predictor variable and wins (W) is the dependent or response variable. We interpret the response variable as the value of the y-intercept when the predictor variable is zero and we interpret the slope as the amount that the response variable increases when the predictor variable increases by one. In other words, when run differential (RD) is set to zero, wins (W) equal 80.881375. When run differential (RD) increases by one, wins will equal 80.881375 plus 0.1058(1), just to give you a mental picture.

The p-value for the run differential (RD) is incredibly small and much less than a significance level of 0.05. This indicates a relationship is present between the number of wins (W) and run differential (RD).

The adjusted R-sq is 0.8807, which is relatively high and indicates our model fits relatively well with our data.

In order to find the run differential (RD) necessary to win 95 games in one regular season, we let $W = 95$:

$$95 = 80.881375 + 0.1058(RD)$$

$$95 - 80.881375 = 0.1058(RD)$$

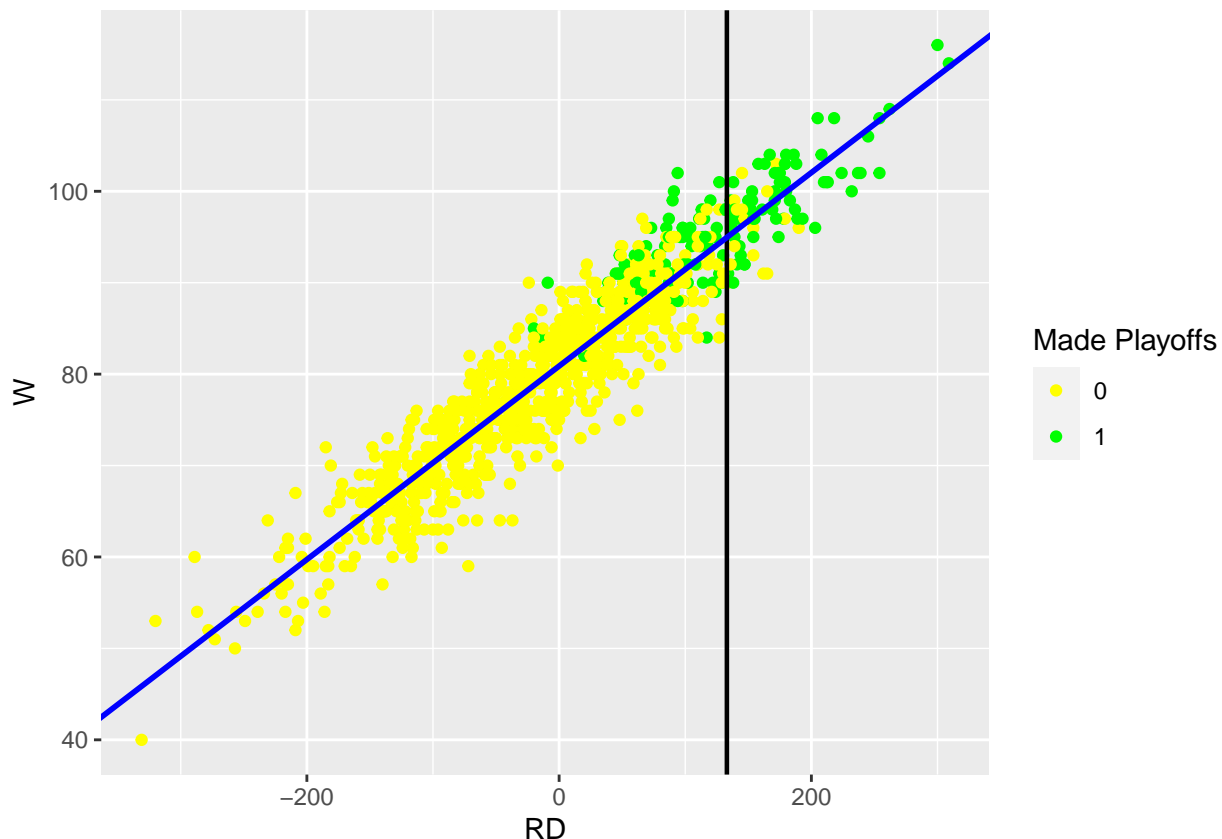
$$14.12 = 0.1058(RD)$$

$$RD = 133.45$$

$$RD \approx +133$$

Thus, the A's needed a run differential (RD) of +133 in order to win 95 games in one season. We will recreate the previous residuals plot and add a reference line at 133 on the x-axis.

```
# plot the x=RD to y=W with a reference line at 133 wins
RDref_plot <- ggplot(data = moneyball_subset,
                     aes(x=RD,
                         y = W,
                         color = factor(Playoffs))) +
  geom_point() + scale_color_manual(values= c("yellow", "green"),
                                    name = "Made Playoffs") +
  geom_vline(aes(xintercept = 133),
             color = "black",
             linetype = "solid",
             size = 0.8)
RDref_plot + geom_abline(intercept = 80.8814,
                        slope = 0.1058,
                        color = "blue",
                        linetype = "solid",
                        size = 1)
```



The threshold looks reasonable. It was not an exact science, but again, the whole point of the calculations was simply to give the Oakland A's the edge during that 2002 season. At the time, it was probably the most sure-fire, shot-in-the-dark anyone has ever taken.

Question 5:

Now Beane and DePodesta once again had another question to answer:

How do we score 133 more runs than we allow?

To answer this, DePodesta needed to find the predicted run differential (RD). So he ran a linear regression model on runs scored (RS) and also on runs allowed (RA). He also conducted this analysis to determine which statistics were the most statistically significant when attempting to predict the number of runs scored in a season.

In short, DePodesta found that on-base percentage (OBP) and slugging percentage (SLG) were much better at predicting runs scored (RS). Even further, years later, and we now know On-Base Percentage + Slugging Percentage (OPS) is an even better at predicting runs scored (RS). DePodesta performed this analysis with many variables but for the sake of simplicity, I will start with a regression model predicting the number of runs scored by using on-base percentage (OBP), slugging percentage (SLG), and batting average (BA).

```
# Linear regression model for RS using OBP, SLG, & BA
RS_reg <- lm(RS ~ OBP + SLG + BA, data = moneyball_subset)
summary(RS_reg)
```

```
##
## Call:
## lm(formula = RS ~ OBP + SLG + BA, data = moneyball_subset)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -70.941 -17.247  -0.621  16.754  90.998
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -788.46      19.70  -40.029 < 2e-16 ***
## OBP           2917.42     110.47   26.410 < 2e-16 ***
## SLG           1637.93      45.99   35.612 < 2e-16 ***
## BA            -368.97     130.58   -2.826  0.00482 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.69 on 898 degrees of freedom
## Multiple R-squared:  0.9302, Adjusted R-squared:  0.93
## F-statistic: 3989 on 3 and 898 DF,  p-value: < 2.2e-16
```

The linear regression equation is:

$$RS = -788.46 + 2917.42(OBP) + 1637.93(SLG) - 368.97(BA)$$

From the linear regression model summary, the p-values for on-base percentage (OBP) and slugging percentage (SLG) are incredibly small, $2e-16$, and much less than a significance level of 0.05. This indicates on-base percentage (OBP) and slugging percentage (SLG) have a strong relationship with runs scored (RS). On the other hand, the p-value for batting average (BA) is 0.00482. While that is less than a significance level of 0.05, it indicates a moderately strong relationship.

The adjusted R-sq is 0.93, which is great and indicates our model fits the data incredibly well. Although, let us see what removing batting average (BA) does to our model.

```
# regression analysis for RS using OBP & SLG
RS_regnoBA <- lm(RS ~ OBP + SLG, data = moneyball_subset)
summary(RS_regnoBA)
```

```
##
## Call:
## lm(formula = RS ~ OBP + SLG, data = moneyball_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -70.838 -17.174  -1.108  16.770  90.036
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -804.63      18.92  -42.53 <2e-16 ***
## OBP           2737.77      90.68   30.19 <2e-16 ***
## SLG           1584.91      42.16   37.60 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.79 on 899 degrees of freedom
## Multiple R-squared:  0.9296, Adjusted R-squared:  0.9294
## F-statistic: 5934 on 2 and 899 DF,  p-value: < 2.2e-16
```

The linear regression equation is:

$$RS = -804.63 + 2737.77(OBP) + 1584.91(SLG)$$

The p-values for on-base percentage (OBP) and slugging percentage (SLG) are incredibly small, $2e-16$, and much less than a significance level of 0.05. The result is the same before, on-base percentage and slugging percentage (SLG) are highly correlated with runs scored (RS).

The adjusted R-sq is 0.9294. Remember, our goal is to have the highest adjusted R-sq with the least amount of variables. In this case, the adjusted R-sq is only slightly better by a minuscule amount with batting average (BA) included. Therefore, to follow with best practice and DePodesta's original hypothesis, I'll stick with the second model.

Now we need to create a regression model for the number of runs allowed (RA) in a season using the opponent's on-base percentage (OOBP) and the opponent's slugging percentage (OSLG).

```
# Linear regression model for RA
RA_reg <- lm(RA ~ OOBP + OSLG, data = moneyball_subset)
summary(RA_reg)

##
## Call:
## lm(formula = RA ~ OOBP + OSLG, data = moneyball_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -82.397 -15.178  -0.129   17.679   60.955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -837.38      60.26  -13.897  < 2e-16 ***
## OOBP          2913.60     291.97    9.979 4.46e-16 ***
## OSLG          1514.29     175.43    8.632 2.55e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.67 on 87 degrees of freedom
## (812 observations deleted due to missingness)
## Multiple R-squared:  0.9073, Adjusted R-squared:  0.9052
## F-statistic: 425.8 on 2 and 87 DF,  p-value: < 2.2e-16
```

The linear regression equation is:

$$RA = -837.38 + 2913.60(OOBP) + 1514.29(OSLG)$$

The adjusted R-sq is 0.9052, which is high and exactly what we want: our model fits the data well. Now we can proceed:

Question 6:

What OBP and SLG do we need to achieve a run differential of +133?

Using the opponent's on-base percentage (OOBP) and the opponent's slugging percentage (OSLG) for the A's from the 2001 season and by plugging it into the linear regression equation, DePodesta assumed that the numbers would be roughly the same for the 2002 season. He used the runs allowed (RA) regression model to predict how many runs the A's would allow. First we need to extract that information from the data set:

```
OAK_2001stats <- moneyball_subset %>%
  filter(Team == "OAK", Year == 2001) %>%
  select(Team, Year, OBP, SLG, OOBP, OSLG)
OAK_2001stats
```

```
## # A tibble: 1 x 6
##   Team   Year   OBP   SLG  OOBP  OSLG
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 OAK    2001 0.345 0.439 0.308 0.38
```

Plugging the 2001 values in, we get:

$$RA = -837.38 + 2913.60(OOBP) + 1514.29(OSLG)$$

$$RA = -837.38 + 2913.60(0.308) + 1514.29(0.38)$$

$$RA = -837.38 + 897.39 + 575.43$$

$$RA = 635.42$$

$$RA \approx 600$$

So now if the A's are predicted to allow 600 runs in 2002, and they need a run differential of +133, then the math is as follows:

$$133 = RS - 600$$

$$RS = 733$$

The A's will need to score 733 runs. As far as the story goes, this is where the true recruiting process began. Beane and DePodesta needed to recruit players that would raise the team's on-base percentage (OBP) and slugging percentage (SLG) to the point where the regression equation resulted in a value greater than 733 runs. It helped out Billy Beane because he and Paul DePodesta were able to build a playoff worthy team on a tight budget.

The Payoff

Now we just need to look at the 2002 season and see if all DePodesta's hard work paid off. We will use the regression equations created above and plug in statistics from the 2002 season and see how accurately they match up to what their real stats ended up being.

```
OAK_2002stats <- moneyball %>%
  filter(Team == "OAK", Year == 2002) %>%
  select(Team, Year, OBP, SLG, OOBP, OSLG)
OAK_2002stats
```

```
## # A tibble: 1 x 6
##   Team   Year   OBP   SLG  OOBP  OSLG
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 OAK    2002 0.339 0.432 0.315 0.384
```

Plugging the 2002 values in to the runs allowed regression equation, we get:

$$RA = -837.38 + 2913.60(0.315) + 1514.29(0.384)$$

$$RA = -837.38 + 917.784 + 581.49$$

$$RA = 661.9$$

$$RA \approx 662$$

The actual runs allowed (RA) in 2002 was 654.¹⁰ By DePodesta's calculations, the Oakland A's did better than expected.

Plugging the 2002 values into the runs scored (RS) regression equation, we get:

$$RS = -804.63 + 2737.77(OBP) + 1584.91(SLG)$$

$$RS = -804.63 + 2737.77(0.339) + 1584.91(0.432)$$

$$RS = -804.63 + 928.1 + 684.7$$

$$RS = 808.2$$

$$RS \approx 808$$

The actual runs scored (RS) in 2002 was 800.¹¹ Again, the A's did better than what was expected.

Finally, this leads us to the actual run differential (RD) in 2002:

$$RD = 800 - 654 = 146$$

As we can see, our equations turned out to be pretty accurate, and the A's ended up with greater than a +133 run differential.

Using our regression equation for the number of games won (W), we get:

$$W = 80.881375 + 0.1058(RD)$$

$$W = 80.881375 + 0.1058(146)$$

$$W = 80.881375 + 15.45$$

$$W = 96.33$$

$$W \approx 96$$

The actual games won in the 2002 season was 103 games won.¹² All of the math adds up and exceeds the thresholds we set earlier. The Oakland A's did indeed make the playoffs in 2002, in fact, they finished first in the American League Western Division. Unfortunately, the A's did not win the world series that season, but they made it to playoffs!

¹⁰Sports Reference Inc. (2007, July 4). 2002 Oakland Athletics Statistics and Roster - Baseball-Reference.com. http://spiff.rit.edu/richmond/baseball/order/team_stats/OAK_2002.html.

¹¹Sports Reference Inc. (2007, July 4). 2002 Oakland Athletics Statistics and Roster - Baseball-Reference.com. http://spiff.rit.edu/richmond/baseball/order/team_stats/OAK_2002.html.

¹²Sports Reference Inc. (2007, July 4). 2002 Oakland Athletics Statistics and Roster - Baseball-Reference.com. http://spiff.rit.edu/richmond/baseball/order/team_stats/OAK_2002.html.

