# Pete Berryman ----- Team Two-Strike Approach Analysis ----- October 11th, 2018

cat("\014")

install.packages("dplyr")

install.packages("magrittr")

install.packages("readxl")

library(dplyr)

library(magrittr)

library(readxl)

#Find which teams' have lowest K% over last 10 years (best contact team)

Kpct <- read.csv("~/Downloads/Teams K pct over last 10.csv")

KpctSummary <- Kpct %>%

group\_by(Team)%>%

summarise(meanKpct = mean(K.), '25th pctile' = quantile(K., probs = 0.25), '75th pctile' = quantile(K., probs = 0.75), SD = sd(K.))

KpctRank <- Kpct %>%

select(Season, Team, K., wRC.)%>%

group\_by(Season)%>%

arrange(K.)%>%

mutate(rank = rank(K.))

KpctRankSummary <- KpctRank%>%

group\_by(Team)%>%

summarise(meanrank = mean(rank), '25th pctile' = quantile(rank, probs = 0.25), '75th pctile' = quantile(rank, probs = 0.75), SD = sd(rank))%>%

arrange(meanrank)

#Possible expansion on study: Analyze 0/1 strike counts contact percentage for comparison

#---------------------------------------------------------------------------------------------------------------

#Find which teams' quality of contact is least hurt with two strikes (Launch Speed & xwOBA)

NotTwoStrike.xwOBA.Results <- read.csv("~/Downloads/NotTwoStrike xwOBA Results.csv")

TwoStrike.xwOBA.Results <- read.csv("~/Downloads/TwoStrike xwOBA Results.csv")

ContactPct <- read.csv("~/Downloads/ContactPct.csv")

Runs <- read.csv("~/Downloads/Runs.csv")

Team\_xwOBA\_2 <- TwoStrike.xwOBA.Results %>%

mutate(pitch\_percent\_2strike = pitch\_percent, ContactRate\_2 = 1 - whiffs/swings, xwOBA\_2 = xwoba, launch\_speed\_2 = launch\_speed) %>%

select(player\_id, pitch\_percent\_2strike, ContactRate\_2, xwOBA\_2, launch\_speed\_2)

Team\_xwOBA <- NotTwoStrike.xwOBA.Results %>%

mutate(ContactRate = 1 - whiffs/swings) %>%

select(player\_id, pitches, ContactRate, xwoba, launch\_speed)

Comparison <- full\_join(Team\_xwOBA, Team\_xwOBA\_2, by = "player\_id") %>%

mutate(Contact\_Rate\_diff = ContactRate\_2 - ContactRate, xwOBA\_diff = xwOBA\_2 - xwoba, launch\_speed\_diff = launch\_speed\_2 - launch\_speed) %>%

select(player\_id, pitch\_percent\_2strike, Contact\_Rate\_diff, xwOBA\_diff, launch\_speed\_diff)

ContactToRuns <-full\_join(ContactPct, Runs, by= "Team") %>%

select(Team, Contact., O.Swing., Z.Contact., R)%>%

#Export Comparison to csv and then to Excel to make better graph

write.table(Comparison, file = "~/Desktop/MLBTeamTwoStrikeComparison.csv", sep = ",", col.names = TRUE)

write.table(ContactToRuns, file = "~/Desktop/ContactToRuns.csv", sep = ",", col.names = TRUE)

#---------------------------------------------------------------------------------------------------------------

# Which players’ have the largest and greatest drops in production (wOBA) with two strikes?

library(dplyr)

library(magrittr)

notTwoStrikes <- read.csv("~/Downloads/0or1 strike.csv")

TwoStrikes <- read.csv("~/Downloads/2 strike.csv")

notTwoStrikes <- notTwoStrikes %>%

mutate(wOBA = woba, no\_in\_play\_0or1 = pitches) %>%

select(player\_id, player\_name, no\_in\_play\_0or1, wOBA)

TwoStrikes <- TwoStrikes %>%

mutate(wOBA\_2 = woba, no\_in\_play\_2 = pitches) %>%

select(player\_id, player\_name, no\_in\_play\_2, wOBA\_2)

HitterComparison <- inner\_join(notTwoStrikes, TwoStrikes)%>%

mutate(wOBAdiff = wOBA - wOBA\_2)%>%

arrange(wOBAdiff)

#T test to determine whether difference between 0/1 strike wOBA and 2 strike wOBA is statistically significant

t.test(HitterComparison$wOBA, HitterComparison$wOBA\_2, alternative = "two.sided", var.equal = FALSE) #H0: wOBA=wOBA\_2; HA: wOBA≠wOBA\_2

t.test(HitterComparison$wOBA, HitterComparison$wOBA\_2, alternative = "greater", var.equal = FALSE) #H0: wOBA=wOBA\_2; HA: wOBA>wOBA\_2

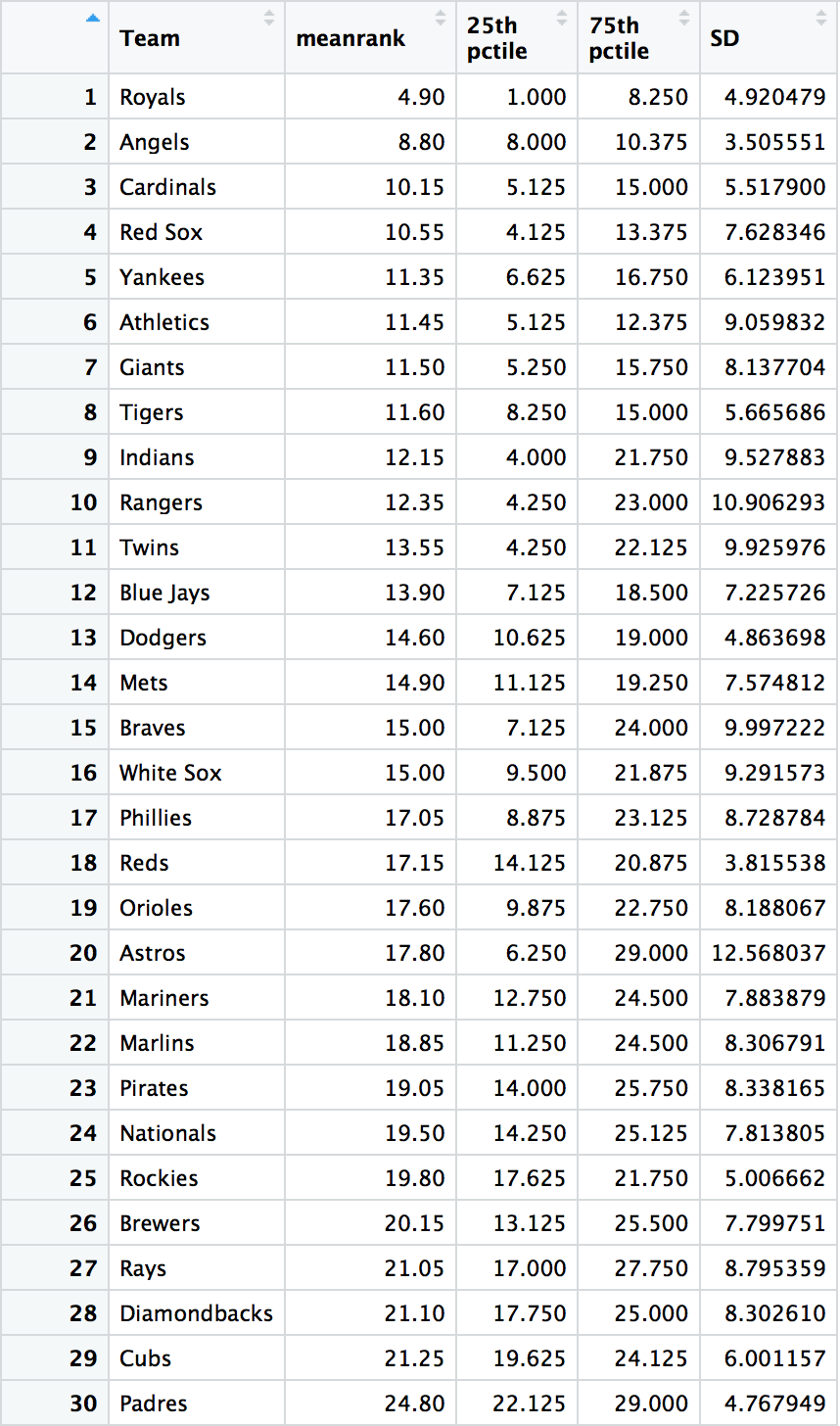
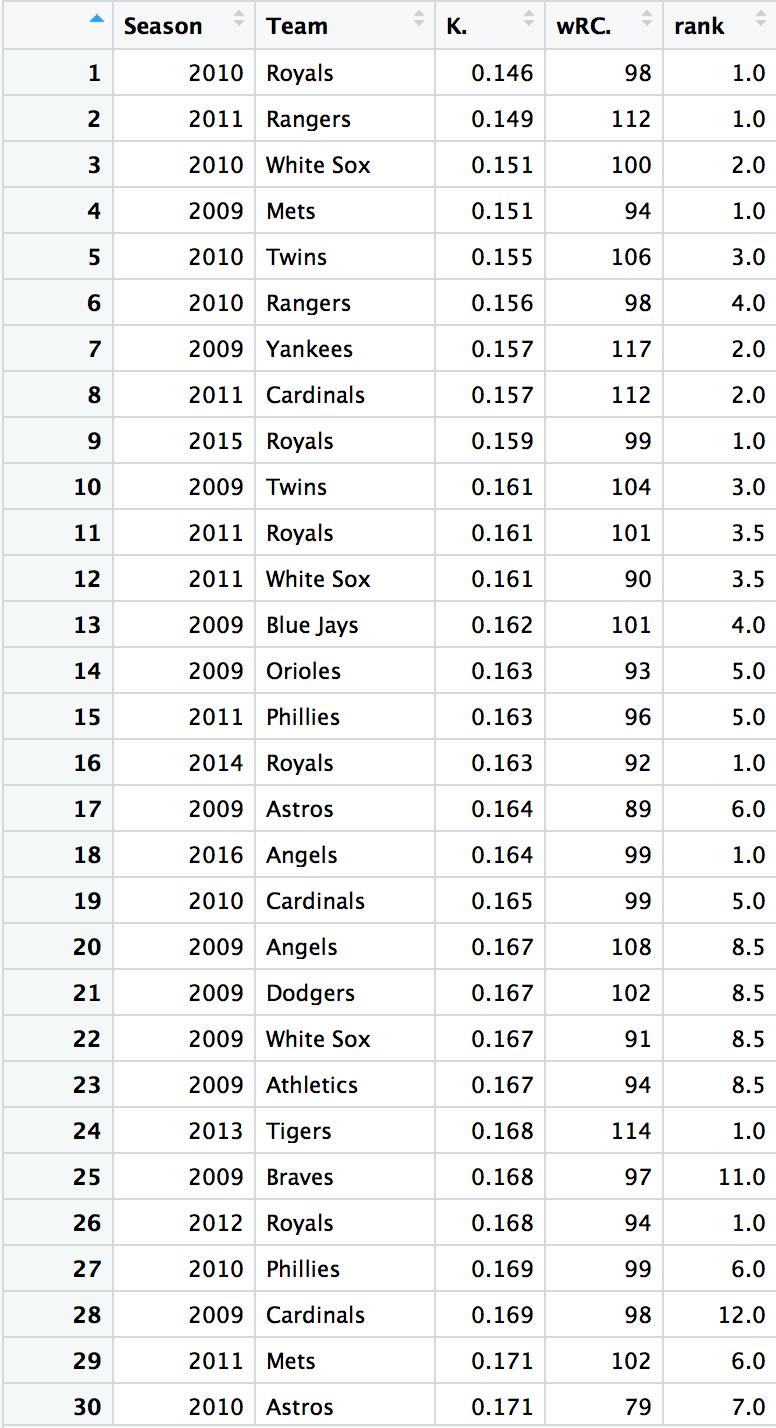
**Pete Berryman | Two Strike Analysis | October 15, 2018**

In today’s MLB, one of the most glaring features of the game is the quantity of strikeouts. This fact is one of the many reasons that experts have cited to explain why the game is decreasing in popularity. It makes sense. The greatest action is when the ball is in play. This is when the fielders showcase their range, slick fielding, and strong arms. This is when base running becomes an art and has a critical effect on the outcome of the game. However, home runs are flying out unlike ever before, which is the ultimate way to wake up a bored crowd. In a nutshell, today’s game is about extremes. My goal with this report was to see if data could help explain the increased strikeouts. An obvious reason for increased strikeouts is that 95 mph fastballs are now the norm. Pitchers now complement their heaters with three to five wicked off-speed pitches. Moreover, pitchers are getting smarter and smarter with their use of data to attack a lineup’s weaknesses. In fact, most pitchers have different variations of their fastball, too, so as the batter loads up, it is harder than ever to know what he is about to see.

With this report, I wanted to know if the hitters’ approaches had something to do with the high strikeouts. Growing up as a baseball player, you often hear phrases such as “choke up with two strikes” or “shorten up with two strikes” or “just put the ball in play”. It is what most of us refer to as a two-strike approach. You forget those dreams of clearing the outfield fence and instead just try to force the defense to make a play. Maybe you will smack a line drive up the middle. As analytics have proven the uselessness of batting average and shown us how much more valuable extra base hits are towards run production, it seems as though the two-strike approach has left us. You see batters swinging out of their shoes and treading back to the dugout shamelessly. Managers do not mind. Heck, the strikeout was no different than a popout, groundout, or flyout. In fact, the batter made the pitcher throw at least three pitches, probably more. MLB defenses are superb (except a few), so errors are not very common in the grand scheme of things. To get to the bottom of what managers, hitting coaches, and batters might be thinking in two-strike situations, I came up with a few questions to guide my research:

1. Question: Are any teams clearly emphasizing a two-strike approach more than others?

Ranking of avg. K% over last 10 years | Best K% seasons of last 10 years

* 1. Royals hitters have been in the top 10 in MLB in K% (lowest %) for 9 of the last 10 years
     1. Angels in last 8 of 10 (also most consistent due to lowest std deviation)
     2. Cubs have been outside the top 10 for each of the last ten years (average rank of 21.25)
        1. Theo Epstein and Joe Maddon have strong history of ingenuity and value finding, so they must be seeing value elsewhere.
  2. Royals have led the MLB in K% in 4 of the last 10 years.
  3. There is a general negative correlation between K% and wRC+, but there are many exceptions, so the correlation is not a strong predictor.

1. Question: Which teams are not holding back their power approaches with two strikes?

Only three teams (CIN, SD, and NYM) managed to have a higher average launch speed with two strikes than with less than two strikes. Interestingly, these teams finished 18th, 28th, and 23rd in the MLB in runs scored this year, respectively. They also finished last, last, and fourth in their respective divisions.

Avg. Contact Rate differential: 0.004927426

Only 17 out of 30 teams had higher contact rates with two strikes than with less than two strikes. It appears that few teams preach a more conservative approach, even if the launch speed drops for 27 out of 30 teams with two strikes. Is it because hitters are tenser? Do pitchers become more confident with two strikes? Is it because of the types of pitches being thrown with two strikes?

At the end of the day, every team’s xwOBA on two-strike contact suffered substantially. This is likely a result of the nearly unanimous decreased launch speed (although launch angle factors in, too).

1. Question: Is the difference in two-strike success statistically significant?
   1. League average for wOBA on balls in play: 0.368
      1. With 0 or 1 strike: 0.377
      2. With 2 strike: 0.353
   2. T-test results:
      1. H0: wOBA on contact with less than 2 strikes = wOBA on contact with 2 strikes

Ha: wOBA on contact with less than 2 strikes ≠ wOBA on contact with 2 strikes

* + 1. Data: all hitters with at least 100 balls in play with less than strikes and 75 balls in play with two strikes (276 hitters; not random sampling but includes entire population of qualified hitters)
    2. with α=0.05, we **reject** the null hypothesis that there **is not** a statistically significant difference between the mean wOBA on contact with less than two strikes and the mean wOBA on contact with two strikes.
    3. While some hitters have more successful contact with two strikes for whatever reason (e.g. Freddie Freeman, Carlos Gonzalez, Mookie Betts, Giancarlo Stanton), there is a statistically significant improvement in hitters’ wOBA values when their contact is with less than two strikes.