Exam 2, Sports Analytics take home

Name: _____

November, 2018

NHL applications

```
library(RCurl)
library(tidyverse)
url <- getURL("https://raw.githubusercontent.com/statsbylopez/StatsSports/master/NHL.csv")
nhl.data <- read_csv(url)
nhl.data <- nhl.data %>% filter(TOI > 500)
nhl.data <- na.omit(nhl.data)
nhl.data$ShP <- nhl.data$Goals/nhl.data$Shots
```

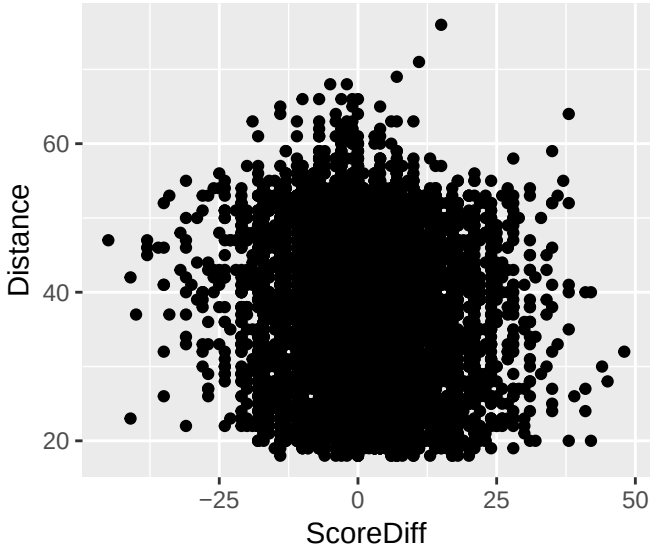
1. Which three players recorded seasons with the highest PDO?
2. Average each players PDO by season. Which three players have the highest average PDO over their career?
3. Average each players PDO by season. Among players with at least 5 seasons, which three players have the highest average PDO over their career?
4. In two sentences, explain to a general manager what Derek Stepan's high average PDO (103) means about how you should value his past performances.

NFL applications

```
url <- getURL("https://raw.githubusercontent.com/statsbylopez/StatsSports/master/Kickers.csv")
nfl.kick <- read.csv(text = url)
head(nfl.kick)
```

```
##   Team Year GameMinute Kicker Distance ScoreDiff Grass Success
## 1  PHI 2005          3  Akers       49         0 FALSE       0
## 2  PHI 2005         29  Akers       49        -7 FALSE       0
## 3  PHI 2005         51  Akers       44        -7 FALSE       1
## 4  PHI 2005         14  Akers       43        14  TRUE       0
## 5  PHI 2005         60  Akers       23         0  TRUE       1
## 6  PHI 2005         39  Akers       34        -3  TRUE       1
```

5. Identify the team whose kickers averaged the longest kicks during the 2015 season
6. Identify the team whose kickers kicked the greatest percentage of kicks on grass during the 2015 season
7. Make the following plot and summarize the small but noticeable association between score difference (positive when leading) and field goal distance



8. In exactly one sentence, interpret the results that follow from the calculations below.

```
nfl.kick %>%
  filter(Distance >= 59) %>%
  group_by(Grass) %>%
  summarise(success.rate = mean(Success))
```

9. A coach identifies that your results above don't exactly make sense. Give an alternative explanation that the coach may be thinking of that explains your findings above.

James-Stein estimate

Use the following code to get started. Complete the questions using the `first.defenders` data set.

```
url <- getURL("https://raw.githubusercontent.com/statsbylopez/StatsSports/master/NHL.csv")
nhl.data <- read_csv(url)
nhl.data <- nhl.data %>% filter(TOI > 500)
nhl.data <- na.omit(nhl.data)
nhl.data$ShP <- nhl.data$Goals/nhl.data$Shots
first.season <- nhl.data %>% filter(Season> 20132014)
first.defenders <- first.season %>%
  group_by(Name) %>%
  filter(Shots > 20, Position == "D") %>%
  select(Name,  Position, Goals, Shots, ShP)
dim(first.defenders)
head(first.defenders)
```

10. Estimate the MLE and the James-Stein estimator for all of the players in the `first.defenders` data set. This uses data from the 13-14 NHL season
11. Interpret the shrinkage constant (e.g., the `c`) for Karl Alzner. Note that Karl is the second row in the data
12. What is the MLE for Karl Alzner? What is his JS estimate?
13. Karl's shooting percentage over the next two years is 7 percent. Identify the absolute error for Karl using his MLE and his JS estimate. Which yielded a lower error?

NFL expected points

Use the following data set: all rushes in the 2015 season. **Note**: please put the url on one line.

```
url.15 <- getURL("https://raw.githubusercontent.com/ryurko/
                  nflscrapR-data/master/play_by_play_data/regular_season/reg_pbp_2015.csv")
pbp <- read.csv(text = url.15)
rushes <- pbp %>% filter(play_type == "run")
rushes <- rushes %>%
  select(game_id, home_team, away_team, posteam, yardline_100, game_seconds_remaining,
         qtr, down, ydstogo, yards_gained, shotgun, no_huddle, run_gap,
         run_location, epa, wpa, rusher_player_name) %>% na.omit()
```

14. Make side by side boxplots of EPA for each rush team (`posteam`). Note: use the code `+ ylim(c(-2.5, 2.5))` to zoom in on the y-axis and `+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5))` to improve your x-axis. Which team has the highest median EPA?
15. Use the summarize command to double check your answer in No. 14.
16. Facet your boxplots by the shotgun variable (`shotgun`). Which team stands out as having a very high 75th percentile when running from shotgun?

Honor code

Write the following on the top of your exam, and sign

While taking this examinations, I have not witnessed any wrongdoing, nor have I personally violated any conditions of the Skidmore College honor code.