HW 5: NFL Analytics

Stats and sports class

Preliminary notes for doing HW

1. All files should be knit and compiled using R Markdown. Knit early and often! I do not recommend waiting until the end of the HW to knit.
2. All questions should be answered completely, and, wherever applicable, code should be included.
3. If you work with a partner or group, please write the names of your teammates.
4. Copying and pasting of code is a violation of the Skidmore honor code

Part I

HW Grade

Return to Homework 4 and assign yourself a grade:

- 1-3 out of 5 points: Most questions attempted, minimal effort
- 4 of 5 points: All questions attempted, complete effort, graded questions incorrect
- 4.5 of 5 points: All questions attempted, complete effort, graded questions partially correct
- 5 of 5 points: All questions attempted, graded questions perfect

Solutions to HW 4 posted on Github

Homework questions

Return to data from our lab that used `nflfastR` data to look at player metrics in football

```
library(tidyverse)
pbp_20 <- readRDS(url("https://raw.githubusercontent.com/guga31bb/nflfastR-data/master/data/play_by_pla
dim(pbp_20)

pbp_scrimmage <- pbp_20 %>%
  select(game_id, posteam, defteam, play_type, play_id, down, ydstogo, yardline_100, home_score, away_s
         passer_player_name, receiver_player_name, air_yards, complete_pass, desc)  %>%
  filter(play_type == "pass"|play_type == "run")
```

Quarterback metrics

Expected points are nice because it allows us to compare how plays helped or hurt an offense put points on the board.

A similar metric can be used with players at each position. Let's look with quarterbacks.

```
passes <- pbp_scrimmage %>%
  filter(play_type == "pass", !is.na(air_yards), air_yards >= -10)
```

```r
passes %>%
  group_by(passer_player_name) %>%
  summarise(completion_pct = mean(complete_pass),
            n_passes = n()) %>%
  filter(n_passes >= 30) %>%
  arrange(-completion_pct)
```

The above code ranks quarterbacks with at least 30 attempts by their completion percentage.

However, not all completions are equal. It's much easier for quarterbacks to complete 5 yard passes than to complete 25 yard passes. Which quarterbacks tend to throw shorter and deeper passes?

We can look at the `air_yards` variable

```r
passes %>%
  group_by(passer_player_name) %>%
  summarise(completion_pct = mean(complete_pass),
            avg_pass_length = mean(air_yards),
            n_passes = n()) %>%
  filter(n_passes >= 30)
```

1. Which quarterbacks have tended to throw the shortest and longest passes?

QB performance versus expected

Not surprisingly, it's more difficult to complete passes that travel further in the air.

```r
fit_1 <- glm(complete_pass ~ air_yards, data = passes, family = "binomial")
library(broom)
tidy(fit_1)
```

We can use the model above to estimate probabilities for each pass being complete.

```r
passes <- passes %>%
  mutate(complete_hat = predict(fit_1, type = "response"))
```

The predict command provides a probability of being complete for each pass in the data set, at least using the model above.

Let's take a look at the first pass in the data set, from Jimmy Garoppolo to George Kittle.

```r
passes %>% head(1) %>% print.data.frame()
```

The pass to Kittle traveled four yards in the air, which comes with an estimated completion probability of about 72 percent.

2. The pass to Kittle was complete. How many completions over expected was this pass worth?
3. Identify the average expected completion percentage for each quarterback – that is the average of their `complete_hat`. This represents how an average quarterback would do if they were given the same pass lengths as was given to each quarterback.
4. Using your average above, as well as the observed completion percentages (See earlier code), which quarterbacks have the highest completion percentage above their expectation? The lowest?

Visualizing QB performance

5. Use the `passes` data set, and make a plot of air-yards (x-axis) versus `complete_hat` on the y-axis. How do you feel about this association?

Here's code to make a comparison of QBs. Use for Questions 6-9

```
passes %>%
  group_by(passer_player_name) %>%
  summarise(complete_rate_obs = mean(complete_pass),
            complete_rate_exp = mean(complete_hat),
            complete_above_exp = complete_rate_obs - complete_rate_exp,
            n_passes = n()) %>%
  filter(n_passes >= 30) %>%
  ggplot(aes(complete_rate_exp, complete_rate_obs)) +
  geom_text(aes(label = passer_player_name)) +
  geom_point()
```

6. Identify the quarterbacks doing the best and worst at the most difficult passes
7. Identify the quarterbacks doing the best and worst at the easiest passes
8. Add the line `+ geom_abline(aes(slope = 1, intercept = 0))` to the code above. What does this line represent?
9. Which quarterback is *most* outperforming expectations? Why?

Open ended

Read the following post (credit: Mike Irene) on the creation and evaluation of a field goal model in the NFL

https://www.opensourcefootball.com/posts/2020-09-28-nflfastr-ep-wp-and-cp-models/

10. Compare the variables in the authors final model to the ones we have looked at in class. What additional variables does the author include?
11. Describe the authors' metric "Field Goals over expected". How are expected field goals' calculated?
12. Justin Tucker is listed as having 3.9 field goals over expected. How many additional points has that been worth?
13. A skeptic notes that evaluating field goal kickers on their field goal kicks alone may miss the value that they provide (even ignoring other plays in the game). What's the additional benefit of having a top-notch field goal kicker that isn't considered when only looking at their observed kicks?