Lesson 23 NFL Field Goals (continued)

LTC J.K. Starling

Last compiled on 28 March, 2022

```r
# Load packages
library(tidyverse)
library(table1)

### Data management
#loading play by play for the 2021 season
# library(nflfastR)
# data <- load_pbp(2021)
# field_goal <- data %>% filter(is.na(field_goal_result)!=T)


# field_goal <- read.csv("https://raw.githubusercontent.com/jkstarling/MA376/main/fg_2021.csv",header=T

# setwd("C:/Users/james.starling/OneDrive - West Point/Teaching/MA376/JimsLessons/Block III/LSN23/")
# field_goal <- read.csv("fg_2021.csv", header = TRUE, stringsAsFactors = FALSE)
```

Background

(0) Last time we looked at the 2021 NFL Season and field goal kicks. In this lesson we will build our previous models.

First we cleaned up the variables `field_goal_result` and `surface`:

```r
# # Selecting only field goals and changing blocked field goals to be missed.
# fg <- field_goal %>%
#   mutate(field_goal_result = case_when(field_goal_result=="blocked"~"missed",
#                                        TRUE~field_goal_result)) %>%
#   mutate(field_goal_result = fct_rev(as.factor(field_goal_result)))
#
# #changing the variations of turf to just turf
# fg <- fg %>%
#   mutate(surface = case_when(surface=="astroturf"~"turf",
#                              surface=="fieldturf"~"turf",
#                              surface=="fieldturf "~"turf",
#                              surface=="matrixturf"~"turf",
#                              surface=="sportturf"~"turf",
#                              TRUE~surface)) %>%
#   mutate(surface = as.factor(surface))
#
# # updated dataframe
```

```
# table(fg$field_goal_result)
# table(fg$surface)
```

Then we created a logistic regression model on the surface and distance.

```
# contrasts(fg$surface)
# surface.glm <- fg %>% glm(field_goal_result~surface, data=., family='binomial')
# distance.glm <- glm(field_goal_result ~ kick_distance, data=fg, family='binomial')
# summary(surface.glm)
# summary(distance.glm)
```

(1) We then took a look at the coefficients for each model and found the confidence intervals on those coefficients. How did we interpret these coefficients?

What conclusions can we draw based on the confidence intervals?

```
# coef(surface.glm)
# confint(surface.glm)
#
# coef(distance.glm)
# confint(distance.glm)
```

(2) We then took the exponential of the coefficients and their CIs.

Why did we do this? What is the interpretation of these results?

```
# exp(coef(surface.glm))
# exp(confint(surface.glm))
#
# exp(coef(distance.glm))
# exp(confint(distance.glm))
```

(3) Create a new logistic regression model that predicts the field goal result, taking into account the surface type and the kick distance. Discuss the results of this model.

```
# surf.dist.glm <- glm(....)
# summary(surf.dist.glm)
# confint(surf.dist.glm)
```

(4) Create a new logistic regression model that predicts the field goal result, taking into account the *interaction* between surface type and kick distance. Discuss the results of this model.

```

```
int.glm <- glm(...)
summary(int.glm)
confint(int.glm)
```

(5) What else could we try? Create a different model with some other variables from the data. These can be found at the site: `https://www.nflfastr.com/reference/fast_scraper.html#value`

What are the results of adding these other variables in the model?

- `kick_distance`
- `surface`
- `play_clock` # time on clock when ball snapped
- `season_type` # post/regular
- `weather` # issues with this one?
- `score_differential` # difference in score between team kicking vs. team on defense
- `qtr` # quarter of the game (5 is OT)

```
# fg.glm <- fg %>% glm(...)
#
# summary(fg.glm)
# confint(fg.glm)
```

(6) Calculate:

- the overall percentage rate for field goals made
- correct classification rate with `surface.lm`
- correct classification rate with `distance.lm`
- correct classification rate with `surf.dist.lm`
- correct classification rate with `int.lm`

(What do we need to specify in order to calculate the percentages/rates? Maybe incorporate 0.5 somehow. . . ?)

```
# using surface model


# using distance model


# using surf + distance model
```

(7) How can we check overall model performance?

References Nathan Tintle et al.(2019). Intermediate Statistical Investigations for U.S. Military Academy at West Point.