

Football Bootstrapping

Dr. Ekstrom

Purpose: In this activity, we explore the probability of a football team getting a first down on a given series. We will use data from the 2018 New England Patriots.

Sports Analytics: Look into a fundamental football question - what is the probability of getting a first down?

Statistics: Bootstrapping simulation.

R: Bootstrapping in *R*

PREPARING THE DATA

We begin by loading the data. The data was downloaded from *Pro Football Reference*. Each week of data contains the offensive plays from the Patriots for the given week. We modified the data to remove turnovers, kneel-downs, penalties, and special team plays.

```
week1=read.csv("W1NEvsTexansW0.csv")
week2=read.csv("W2NEvsJagsW0.csv")
week3=read.csv("W3NEvsLionsW0.csv")
week4=read.csv("W4NEvsDolphinsW0.csv")
week5=read.csv("W5NEvsColtsW0.csv")
week6=read.csv("W6NEvsChiefsW0.csv")
week7=read.csv("W7NEvsBearsW0.csv")
week8=read.csv("W8NEvsBillsW0.csv")
week9=read.csv("W9NEvsPackersW0.csv")
week10=read.csv("W10NEvsTitansW0.csv")
week11=read.csv("W11NEvsJetsW0.csv")
week12=read.csv("W12NEvsVikingsW0.csv")
week13=read.csv("W13NEvsDolphinsW0.csv")
week14=read.csv("W14NEvsSteelersW0.csv")
week15=read.csv("W15NEvsBillsW0.csv")
week16=read.csv("W16NEvsJetsW0.csv")
```

We now combine each weeks data into a dataframe named `season`.

```
season=rbind(week1, week2, week3, week4, week5, week6, week7, week8, week9,
             week10, week11, week12, week13, week14, week15, week16)
summary(season)
```

```
##      Type          Time        Gained        Down
##  Length:1034    Length:1034    Min.   :-11.000  Min.   :1.00
##  Class :character  Class :character  1st Qu.:  0.000  1st Qu.:1.00
##  Mode  :character  Mode  :character  Median :  4.000  Median :2.00
##                                         Mean   :  6.084  Mean   :1.75
##                                         3rd Qu.:  9.000  3rd Qu.:2.00
##                                         Max.   : 63.000  Max.   :4.00
##      ToGo          Location
##  Min.   : 1.000  Length:1034
##  1st Qu.: 6.000  Class  :character
##  Median :10.000 Mode   :character
##  Mean   : 8.132
##  3rd Qu.:10.000
##  Max.   :21.000
```

We divide the plays into first-down plays, second-down plays, third-down plays, and fourth-down plays.

```
FrstDwnPlys=season[which(season$Down==1),]
summary(FrstDwnPlys)
```

```
##      Type          Time        Gained        Down
##  Length:480    Length:480    Min.   :-10.000  Min.   :1
##  Class :character  Class :character  1st Qu.:  1.000  1st Qu.:1
##  Mode  :character  Mode  :character  Median :  4.000  Median :1
##                                         Mean   :  6.519  Mean   :1
##                                         3rd Qu.:  9.000  3rd Qu.:1
##                                         Max.   : 63.000  Max.   :1
##      ToGo          Location
##  Min.   : 1.00  Length:480
##  1st Qu.:10.00 Class  :character
##  Median :10.00 Mode   :character
##  Mean   : 9.96
##  3rd Qu.:10.00
##  Max.   :20.00
```

```
ScdDwnPlys=season[which(season$Down==2),]  
summary(ScdDwnPlys)
```

```
##      Type          Time          Gained          Down  
##  Length:346      Length:346      Min.   :-10.000  Min.   :2  
##  Class :character  Class :character  1st Qu.:  0.250  1st Qu.:2  
##  Mode  :character  Mode  :character  Median  : 4.500  Median :2  
##                                         Mean   : 6.266  Mean   :2  
##                                         3rd Qu.:10.000 3rd Qu.:2  
##                                         Max.   :55.000  Max.   :2  
  
##      ToGo          Location  
##  Min.   : 1.000  Length:346  
##  1st Qu.: 4.000  Class :character  
##  Median : 7.000  Mode  :character  
##  Mean   : 7.104  
##  3rd Qu.:10.000  
##  Max.   :20.000
```

```
ThdDwnPlys=season[which(season$Down==3),]  
summary(ThdDwnPlys)
```

```
##      Type          Time          Gained          Down  
##  Length:194      Length:194      Min.   :-11.000  Min.   :3  
##  Class :character  Class :character  1st Qu.:  0.000  1st Qu.:3  
##  Mode  :character  Mode  :character  Median  : 2.000  Median :3  
##                                         Mean   : 4.763  Mean   :3  
##                                         3rd Qu.: 7.000 3rd Qu.:3  
##                                         Max.   :55.000  Max.   :3  
  
##      ToGo          Location  
##  Min.   : 1.000  Length:194  
##  1st Qu.: 2.000  Class :character  
##  Median : 5.000  Mode  :character  
##  Mean   : 5.747  
##  3rd Qu.: 8.000  
##  Max.   :21.000
```

```
FthDwnPlys=season[which(season$Down==4),]  
summary(FthDwnPlys)  
  
##          Type             Time            Gained            Down  
##  Length:14        Length:14      Min.   : 0    Min.   :4  
##  Class :character  Class :character  1st Qu.: 0    1st Qu.:4  
##  Mode  :character  Mode  :character  Median  : 1    Median :4  
##                                         Mean   : 5    Mean   :4  
##                                         3rd Qu.: 2    3rd Qu.:4  
##                                         Max.  :32    Max.  :4  
  
##          ToGo            Location  
##  Min.   : 1.000  Length:14  
##  1st Qu.: 1.000  Class :character  
##  Median : 1.500  Mode  :character  
##  Mean   : 3.929  
##  3rd Qu.: 5.500  
##  Max.   :15.000
```

CREATING THE SIMULATION

In this simulation, we will sample a play from the first down plays, then a play from the second down plays (if needed), then a play from the third down plays (if needed) to see if we can get a first down. We will not consider penalties, nor we will go for it on fourth down.

```
# keep track of downs - must be reset each time we run a series
down=1
# keep track of yards to go - must be reset each time we run a series
ydstogo=10
# keep track if we get a first down - must be reset each time we run a series
success=0
# Begin the series - sample from a first down play
fd=sample(nrow(FrstDwnPlys), 1)
# determine how many yards were gained on this first down play
gain=FrstDwnPlys[fd, ]$Gained
# print the results
print(paste("Gained", gain, "on 1st down"), quote = FALSE)

## [1] Gained 5 on 1st down

# update how many yards to go
ydstogo=ydstogo-gain
# update the down
down=down+1
# if statement that checks if we made a first down
if(ydstogo<=0){
  # print success statement
  print("Made it!")
  # updates success variable
  success=1
} else {
  # Else - Continue Series - sample from a second down play
  sd=sample(nrow(ScdDwnPlys), 1)
  # determine how many yards were gained on this second down play
  gain=ScdDwnPlys[sd, ]$Gained
  # print the results
  print(paste("Gained", gain, "on 2nd down"), quote = FALSE)
  # update how many yards to go
  ydstogo=ydstogo-gain
  # update the down
  down=down+1
  # if statement that checks if we made a first down
  if(ydstogo<=0){
    # print success statement
    print("Made it!")
    # updates success variable
    success=1
  }
}
```

```

} else {
  # Continue Series - sample from a second down play
  td=sample(nrow(ThdDwnPlys), 1)
  # determine how many yards were gained on this third down play
  gain=ThdDwnPlys[td, ]$Gained
  # print the results
  print(paste("Gained", gain, "on 3rd down"), quote = FALSE)
  # update how many yards to go
  ydstogo=ydstogo-gain
  # update the down
  down=down+1
  # if statement that checks if we made a first down
  if(ydstogo<=0){
    # print success statement
    print("Made it!")
    # updates success variable
    success=1
  } else {
    # print failure statement
    print("Didn't make it...")
  }
}
}

## [1] Gained 4 on 2nd down
## [1] Gained 4 on 3rd down
## [1] "Made it!"

success

## [1] 1

```

Now we wrap this in a `for` loop, so we can run many series. Note that once we have it running - we comment out the print statements.

```
# Series is the number of Series we will run
Series=10000
# vector to keep track of successes
result=rep(-1,Series)
# for loop to simulate many series
for(i in 1:Series){
  # print out which series we are on
  # print(paste("Start of Series", i), quote = FALSE)
  # keep track of downs - must be reset each time we run a series
  down=1
  # keep track of yards to go - must be reset each time we run a series
  ydstogo=10
  # keep track if we get a first down - must be reset each time we run a series
  success=0
  # Begin the series - sample from a first down play
  fd=sample(nrow(FrstDwnPlys), 1)
  # determine how many yards were gained on this first down play
  gain=FrstDwnPlys[fd, ]$Gained
  # print the results
  #print(paste("Gained", gain, "on 1st down"), quote = FALSE)
  # update how many yards to go
  ydstogo=ydstogo-gain
  # update the down
  down=down+1
  # if statement that checks if we made a first down
  if(ydstogo<=0){
    # print success statement
    #print("Made it!")
    # updates success variable
    success=1
  } else {
    # Continue Series - sample from a second down play
    sd=sample(nrow(ScdDwnPlys), 1)
    # determine how many yards were gained on this second down play
    gain=ScdDwnPlys[sd, ]$Gained
    # print the results
    #print(paste("Gained", gain, "on 2nd down"), quote = FALSE)
    # update how many yards to go
    ydstogo=ydstogo-gain
    # update the down
    down=down+1
    # if statement that checks if we made a first down
    if(ydstogo<=0){
      # print success statement
```

```

#print("Made it!")
# updates success variable
success=1
} else {
  # Continue Series - sample from a third down play
  td=sample(nrow(ThdDwnPlys), 1)
  # determine how many yards were gained on this third down play
  gain=ThdDwnPlys[td, ]$Gained
  # print the results
  #print(paste("Gained", gain, "on 3rd down"), quote = FALSE)
  # update how many yards to go
  ydstogo=ydstogo-gain
  # update the down
  down=down+1
  # if statement that checks if we made a first down
  if(ydstogo<=0){
    # print success statement
    #print("Made it!")
    # updates success variable
    success=1
  } else {
    success=0
    #print("Didn't make it...")
  }
}
}
# stores success into the result vector
result[i]=success

}
# computes the mean of the result vector
mean(result)

## [1] 0.691

```

When I simulated 10,000 series, I got a value of 0.6951, which means that the team would get a first down on approximately 69.5% of series.

GROUP ACTIVITES

Passing on First Down: What if we modified our simulation so that we only passed the ball on first down? Adjust the simulation to account for this modification, and run 10000 series. Give the approximate probability of getting a first down in at most three plays when we pass on first down.

This will be the class activity that will be due by the end of class on February 27th. Turn in your notebook with your simulation code, and the result of running 10000 series to the Gradescope assignment *Class Activity 2/27*.

Running on First Down: What if we modified our simulation so that we only ran the ball on first down? Adjust the simulation to account for this modification, and run 10000 series. Give the approximate probability of getting a first down in at most three plays when we run on first down.

Turn in your notebook with your simulation code, and the result of running 10000 series.

Adjusting 2nd and 3rd Down Based on Yards to Go: Adjust the model to select only plays with (about) the same yards to go. For example, if it is 2nd down with 6 yards to go, randomly select a play (either running or passing) from the season that was 2nd down and between 5 and 7 yards to go. Do the same for 3rd down. Adjust the simulation to account for this modification, and run 10000 series. Give the approximate probability of getting a first down in at most three plays when we run on first down.

Turn in your notebook with your simulation code, and the result of running 10000 series.