HW 4 Solutions

Your name here

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

For this homework, we will be using data provided in the `nfl.kick` data set, as was done during class. Our goals will be to confirm our knowledge of logistic regression, our interpretations of slopes, as well as kicker-specific analysis.

Part I: Exploratory data analysis

1. Use R to find the **kicker** with the best percentage of successful field goals. Why might one argue that this specific kicker may not be the most accurate, even though he has the highest percentage?

```r
library(RCurl)
library(mosaic)
url <- getURL("https://raw.githubusercontent.com/statsbylopez/StatsSports/master/Kickers.csv")
nfl.kick <- read.csv(text = url)
tally(Success ~ Kicker, data = nfl.kick, format = "proportion")
```

```
##        Kicker
## Success      Akers   Andersen     Andrus     Bailey      Barth    Bironas
##       0 0.19940476 0.11764706 0.60000000 0.10493827 0.15060241 0.14487633
##       1 0.80059524 0.88235294 0.40000000 0.89506173 0.84939759 0.85512367
##        Kicker
## Success    Boswell      Brien    Brindza      Brown     Bryant    Buehler
##       0 0.07692308 0.75000000 0.50000000 0.17827869 0.13961039 0.25000000
##       1 0.92307692 0.25000000 0.50000000 0.82172131 0.86038961 0.75000000
##        Kicker
## Success    Bullock     Carney  Carpenter  Catanzaro      Coons     Cortez
##       0 0.19354839 0.15107914 0.15175097 0.10606061 0.12500000 0.29411765
##       1 0.80645161 0.84892086 0.84824903 0.89393939 0.87500000 0.70588235
##        Kicker
## Success      Coutu     Crosby    Cundiff     Dawson   Edinger       Elam
##       0 1.00000000 0.19314642 0.22471910 0.14156627 0.26470588 0.15753425
##       1 0.00000000 0.80685358 0.77528090 0.85843373 0.73529412 0.84246575
##        Kicker
## Success     Elling      Feely       Folk    Forbath     France     Franks
##       0 1.00000000 0.15248227 0.19475655 0.15853659 0.22222222 0.18750000
##       1 0.00000000 0.84751773 0.80524345 0.84146341 0.77777778 0.81250000
##        Kicker
## Success     Freese       Gano Gostkowski      Gould     Graham  Gramatica
##       0 0.57142857 0.20100503 0.12280702 0.14285714 0.14285714 0.23076923
##       1 0.42857143 0.79899497 0.87719298 0.85714286 0.85714286 0.76923077
##        Kicker
## Success       Hall     Hanson   Hartley  Hauschka     Henery     Hocker
##       0 0.17857143 0.15384615 0.16964286 0.12626263 0.18279570 0.28571429
```

```
##       1 0.82142857 0.84615385 0.83035714 0.87373737 0.81720430 0.71428571
##        Kicker
## Success    Hopkins Janikowski    Kaeding      Kasay     Koenen      Lambo
##       0 0.10344828 0.20059880 0.15151515 0.15137615 0.69230769 0.18750000
##       1 0.89655172 0.79940120 0.84848485 0.84862385 0.30769231 0.81250000
##        Kicker
## Success    Lindell   Longwell       Mare    McManus    Medlock   Mehlhaff
##       0 0.16194332 0.15591398 0.19512195 0.15517241 0.33333333 0.25000000
##       1 0.83805668 0.84408602 0.80487805 0.84482759 0.66666667 0.75000000
##        Kicker
## Success     Murray      Myers     Nedney      Novak     Nugent     Parkey
##       0 0.16666667 0.13333333 0.13422819 0.18023256 0.18819188 0.12500000
##       1 0.83333333 0.86666667 0.86577181 0.81976744 0.81180812 0.87500000
##        Kicker
## Success   Peterson    Pettrey     Potter     Prater    Rackers     Rayner
##       0 0.08000000 0.50000000 0.25000000 0.17226891 0.16101695 0.27777778
##       1 0.92000000 0.50000000 0.75000000 0.82773109 0.83898305 0.72222222
##        Kicker
## Success       Reed     Santos    Schmitt    Scifres     Scobee    Stitser
##       0 0.16304348 0.16666667 0.33333333 0.00000000 0.19927536 0.12500000
##       1 0.83695652 0.83333333 0.66666667 1.00000000 0.80072464 0.87500000
##        Kicker
## Success     Stover    Sturgis     Succop    Suisham     Tucker      Tynes
##       0 0.13071895 0.21505376 0.17553191 0.16791045 0.11538462 0.18777293
##       1 0.86928105 0.78494624 0.82446809 0.83208955 0.88461538 0.81222707
##        Kicker
## Success Vanderjagt  Vinatieri      Walsh    Wilkins   Zuerlein
##       0 0.17777778 0.13864307 0.14965986 0.17000000 0.21848739
##       1 0.82222222 0.86135693 0.85034014 0.83000000 0.78151261
```

```
tally(Success ~ Kicker, data = nfl.kick)
```

```
##        Kicker
## Success Akers Andersen Andrus Bailey Barth Bironas Boswell Brien Brindza
##       0    67        6      3     17    25      41       3     3       6
##       1   269       45      2    145   141     242      36     1       6
##        Kicker
## Success Brown Bryant Buehler Bullock Carney Carpenter Catanzaro Coons
##       0    87     43       8      18     21        39         7     4
##       1   401    265      24      75    118       218        59    28
##        Kicker
## Success Cortez Coutu Crosby Cundiff Dawson Edinger Elam Elling Feely Folk
##       0      5     1     62      40     47       9   23      1    43   52
##       1     12     0    259     138    285      25  123      0   239  215
##        Kicker
## Success Forbath France Franks Freese Gano Gostkowski Gould Graham
##       0      13      2      3      4   40         42    47     37
##       1      69      7     13      3  159        300   282    222
##        Kicker
## Success Gramatica Hall Hanson Hartley Hauschka Henery Hocker Hopkins
##       0         6    5     34      19       25     17      4       3
##       1        20   23    187      93      173     76     10      26
##        Kicker
## Success Janikowski Kaeding Kasay Koenen Lambo Lindell Longwell Mare
```

```
##       0         67       30     33       9      6       40        29    40
##       1        267      168    185       4     26      207       157   165
##        Kicker
## Success McManus Medlock Mehlhaff Murray Myers Nedney Novak Nugent Parkey
##       0       9       4        1      4     4     20    31     51      5
##       1      49       8        3     20    26    129   141    220     35
##        Kicker
## Success Peterson Pettrey Potter Prater Rackers Rayner Reed Santos Schmitt
##       0        2       2      1     41      38     25   30     12       1
##       1       23       2      3    197     198     65  154     60       2
##        Kicker
## Success Scifres Scobee Stitser Stover Sturgis Succop Suisham Tucker Tynes
##       0       0     55       1     20      20     33      45     18    43
##       1       1    221       7    133      73    155     223    138   186
##        Kicker
## Success Vanderjagt Vinatieri Walsh Wilkins Zuerlein
##       0          8        47    22      17       26
##       1         37       292   125      83       93
```

```
tally(Success ~ Team, data = nfl.kick, format = "proportion")
```

```
##        Team
## Success       ARI       ATL       BAL       BUF       CAR       CHI
##       0 0.1485714 0.1560694 0.1460396 0.1522388 0.1622419 0.1536232
##       1 0.8514286 0.8439306 0.8539604 0.8477612 0.8377581 0.8463768
##        Team
## Success       CIN       CLE       DAL       DEN       DET        GB
##       0 0.1618497 0.1529052 0.1783626 0.1555556 0.1753846 0.2036554
##       1 0.8381503 0.8470948 0.8216374 0.8444444 0.8246154 0.7963446
##        Team
## Success       HOU       IND       JAC        KC       MIA       MIN
##       0 0.1857923 0.1340782 0.1856678 0.2023810 0.1970588 0.1558824
##       1 0.8142077 0.8659218 0.8143322 0.7976190 0.8029412 0.8441176
##        Team
## Success        NE        NO       NYG       NYJ       OAK       PHI
##       0 0.1250000 0.2056075 0.1412742 0.1830601 0.2005988 0.1721311
##       1 0.8750000 0.7943925 0.8587258 0.8169399 0.7994012 0.8278689
##        Team
## Success       PIT        SD       SEA        SF       STL        TB
##       0 0.1466667 0.1468927 0.1389646 0.1639785 0.1902017 0.1945289
##       1 0.8533333 0.8531073 0.8610354 0.8360215 0.8097983 0.8054711
##        Team
## Success       TEN       WAS
##       0 0.1433022 0.2111437
##       1 0.8566978 0.7888563
```

```
tally(Grass ~ Team, data = nfl.kick, format = "proportion")
```

```
##        Team
## Grass         ARI       ATL       BAL       BUF       CAR       CHI
##   TRUE  0.7428571 0.3005780 0.3836634 0.2805970 0.7433628 0.7130435
##   FALSE 0.2571429 0.6994220 0.6163366 0.7194030 0.2566372 0.2869565
##        Team
```

```
## Grass         CIN       CLE       DAL       DEN       DET        GB
##   TRUE  0.3034682 0.7553517 0.2777778 0.8750000 0.2584615 0.2845953
##   FALSE 0.6965318 0.2446483 0.7222222 0.1250000 0.7415385 0.7154047
##       Team
## Grass         HOU       IND       JAC        KC       MIA       MIN
##   TRUE  0.7841530 0.3631285 0.7719870 0.8422619 0.7588235 0.2617647
##   FALSE 0.2158470 0.6368715 0.2280130 0.1577381 0.2411765 0.7382353
##       Team
## Grass          NE        NO       NYG       NYJ       OAK       PHI
##   TRUE  0.2994792 0.3489097 0.2853186 0.2377049 0.8712575 0.7240437
##   FALSE 0.7005208 0.6510903 0.7146814 0.7622951 0.1287425 0.2759563
##       Team
## Grass         PIT        SD       SEA        SF       STL        TB
##   TRUE  0.7546667 0.8615819 0.2915531 0.7231183 0.2968300 0.7325228
##   FALSE 0.2453333 0.1384181 0.7084469 0.2768817 0.7031700 0.2674772
##       Team
## Grass         TEN       WAS
##   TRUE  0.8286604 0.6950147
##   FALSE 0.1713396 0.3049853
```

Scifres hit 100% of his field goals, but this is a silly measurement as he only attempted one field goal! Also, some kickers may take FGs from different distances.

2. Use R to find the **team** with the best percentage of successful field goals. Why might one argue that this team may not have had the best kickers even though they've posted the highest overall percentage?

New England hit the most FG's: because we don't know about distances and surfaces, they may not have the best kickers (for example, the Patriots could attempt shorter field goals)

3. Identify the teams that have kicked the highest percentage of their field goals on grass (recall: the `Grass` variable is a TRUE/FALSE indicator for whether or not each kick was kicked on a grass surface.).

Denver's attempted 87.5% of FGs on grass.

Part II: Logistic regression

4. There are several variables in this data set. Using the AIC criteria, identify the logistic regression model that is the best fit for our `Success` outcome.

```r
fit1 <- glm(Success ~ Grass + ScoreDiff + Distance +
              GameMinute+Year, family = binomial(), data= nfl.kick)

fit2 <- glm(Success ~ Grass  + Distance +
              Year, family = binomial(), data= nfl.kick)
AIC(fit1)
```

```
## [1] 8710.012
```

```r
AIC(fit2)
```

```
## [1] 8706.263
```

I get the lowest AIC when using `Success`, `Distance` and `Year`, but answers can vary.

5. Using your model from (4), interpret the coefficient for `Distance` as an odds ratio.

```
exp(-0.104)
```

```
## [1] 0.9012253
```

Field goals that are a yard longer have about 10% lower odds of success, given our logistic regression model that adjusts for field type and year.

6. Odds ratios are multiplicative. That is, if the odds of a successful outcome are e^{β_1} given a one-unit increase in x_1, the odds of a successful outcome are $e^{c*\beta_1}$ given a c-unit increase in x_1. Given your model in (4) what are the odds of making a field goal that is 10 yards longer?

```
exp(10*-0.1047)
```

```
## [1] 0.3509891
```

Odds drop by about 65%.

7. Using the model below, estimate the probability of a successful 40-yard field goal, kicked on a non-grass surface in 2015.

```
fit.3 <- glm(Success ~ Distance + Grass + Year, data = nfl.kick, family ="binomial")
summary(fit.3)
```

Answer:

```
df <- data.frame(Distance = 40, Grass = FALSE, Year = 2015)
predict(fit.3, df, type = "response")
```

About 88%

Part III: Expected points

8. Use your answer to Question (7) to estimate the expected points of a 40-yard field goal, kicked on a non-grass surface in 2015.

0.88 * 3 = 2.64 expected points

9. Kicker A hits the field goal in Question (8) while Kicker B misses it. How many expected points has Kicker A added to his team given this single kick? How about Kicker B?

Kicker A was worth 0.36 over expectation, while Kicker B was worth -2.64 points.

10. It is straightforward to estimate the value of kickers using expected points.

First, we generate predicted probabilities for each field goal using `fit.3`. Next, we use that to estimate the expected points for each field goal (`predict.points`). Finally, we use the result of the field goal (`Success` = 0 or a 1) and the value of the kick (3 points) to get an expected points added (`EPA`) for each kicker on each kick.

```
nfl.kick <- mutate(nfl.kick, predict.Success = predict(fit.3, nfl.kick, type = "response"),
                   predict.points = 3*predict.Success,
                   EPA = Success*3 - predict.points)
```

The first row corresponds to a David Akers kick in 2005. What was the predicted success rate for Akers on this kick? What relative worth (in terms of `EPA`) did Akers provide on this kick?

```
nfl.kick[1,]
```

```
##   Team Year GameMinute Kicker Distance ScoreDiff Grass Success
## 1  PHI 2005          3  Akers       49         0 FALSE       0
```

Akers missed the field goal, costing his team 1.86 points.

11. One metric we may be interested in is the relative worth, in terms of total `EPA`, among all kickers in our data set. The `dplyr` function makes it simple.

```
options(dplyr.print_max = 1e9)
kick.summary <- nfl.kick %>%
  group_by(Kicker) %>%
  summarize(percent.success = mean(Success), total.kicks = length(EPA), total.EPA = sum(EPA)) %>%
  arrange(total.EPA)
kick.summary
```

The above function calculates kicker-specific percentages, each kicker's total number of kicks, and each kickers total EPA.

Since 2005, who has been worth the most (and least) total `EPA` to their teams?

Rob Bironas (37.2 points) and Billy Cundiff (-41.8 points) have been worth the most and the least.

12. Interpret the R-squared calculated below. What does it suggest about the fraction of unexplained variability when it comes to kicker EPA?

```
xyplot(total.EPA ~ percent.success, data = filter(kick.summary, total.kicks >=50 ))
cor(total.EPA ~ percent.success, data = filter(kick.summary, total.kicks >=50 ))^2
```

About half of the variability in `total.EPA` can be explained by a linear association with success rate. In other words, about half of how we judge kickers comes down to factors besides success rate.

13. Given your readings, are there any other variables that you would want to account for when measuring field goal success that aren't in the current data set? How may it effect the ranking of kickers in Question Question (12)?

Yes! Temperature: it's more difficult to kick in the cold/wind. This would increase rankings of kickers who tend to kick in the cold.