IDS 702: Module 2.3

Logistic regression with one predictor (illustration)

Dr. Olanrewaju Michael Akande

Predicting NBA wins

- Let's fit a logistic regression with one predictor to NBA data for four seasons from the 2014/2015 season to the 2017/2018 season.
- Suppose we want to see how the amount of points a team let's the opponents score, affects their odds of winning.
- For this simple example, we will focus on data from one team: SAS (San Antonio Spurs).
- The data is in the file `nba_games_stats_reduced.csv` on Sakai.
- Ideally, we should use more information (and that data is actually available) to predict wins but let's continue for illustrative purposes.
- You will get to practice with the full data soon.

IDS 702

Predicting nba wins

```
nba <- read.csv("data/nba_games_stats_reduced.csv",header=T)
nba <- nba[nba$Team=="SAS",]
colnames(nba)[3] <- "Opp"
nba$win <- rep(0,nrow(nba)); nba$win[nba$WINorLOSS=="W"] <- 1
nba$win <- as.factor(nba$win)
head(nba); dim(nba)
```

```
##     Team WINorLOSS Opp win
## 165  SAS         W 100   1
## 166  SAS         L  94   0
## 167  SAS         W  92   1
## 168  SAS         L  98   0
## 169  SAS         L 100   0
## 170  SAS         W  85   1
```

```
## [1] 328   4
```

```
summary(nba)
```

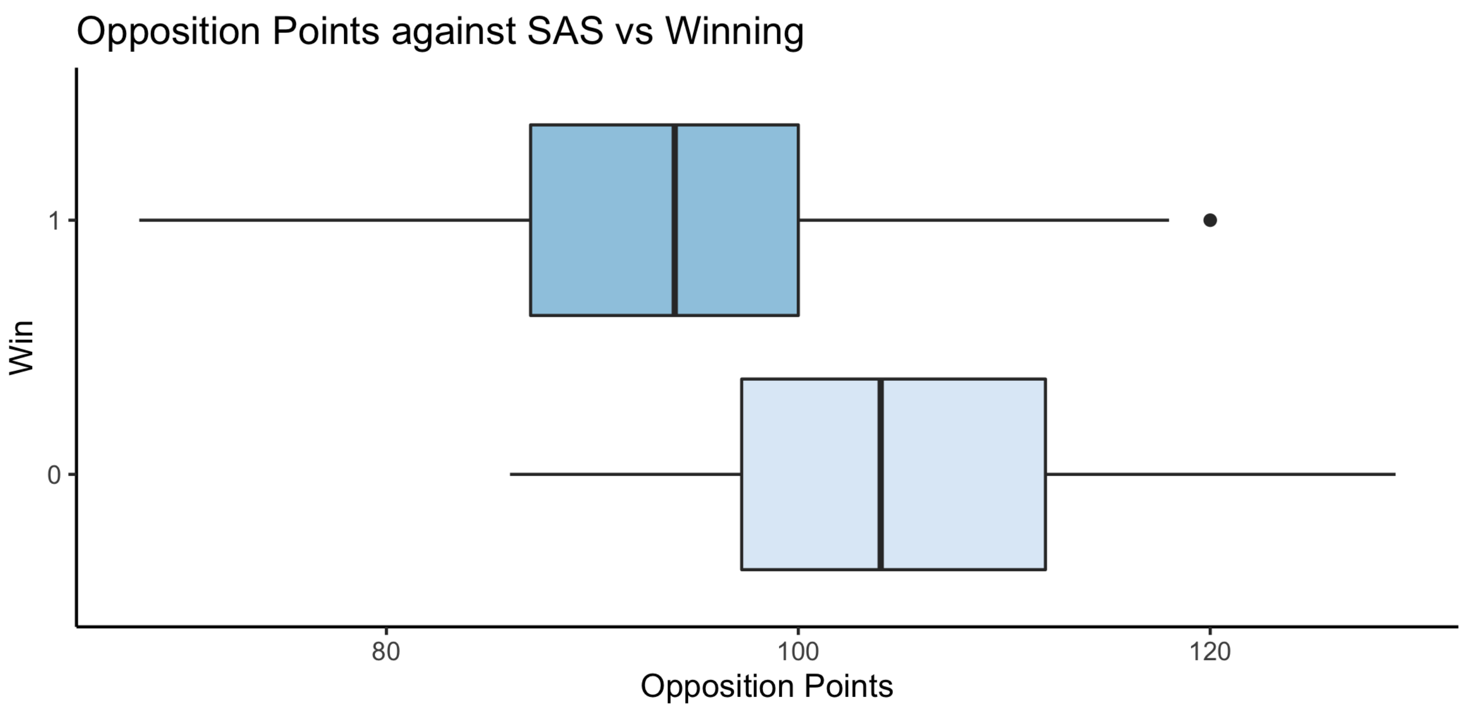
```
##   Team     WINorLOSS      Opp           win
##  CLE:  0   L: 98     Min.   : 68.00   0: 98
##  GSW:  0   W:230     1st Qu.: 90.00   1:230
##  SAS:328             Median : 97.00
##  TOR:  0             Mean   : 96.97
##                      3rd Qu.:104.00
##                      Max.   :129.00
```

IDS 702

Predicting NBA wins

Only one predictor so not much to do in terms of EDA. We can look at

```
ggplot(nba,aes(x=win, y=Opp, fill=win)) +
  geom_boxplot() + coord_flip() +
  scale_fill_brewer(palette="Blues") +
  labs(title="Opposition Points against SAS vs Winning",y="Opposition Points",x="Win") +
   theme_classic() + theme(legend.position="none")
```



IDS 702

Predicting NBA wins

$$\text{win}_i | \text{Opp}_i \sim \text{Bernoulli}(\pi_i); \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \text{Opp}_i$$

```r
nbareg <- glm(win~Opp,family=binomial(link=logit),data=nba); summary(nbareg)
```

```
##
## Call:
## glm(formula = win ~ Opp, family = binomial(link = logit), data = nba)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2760  -0.7073   0.4454   0.7902   1.9593
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.31989    1.66935   7.979 1.47e-15
## Opp         -0.12567    0.01655  -7.594 3.11e-14
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 400.05  on 327  degrees of freedom
## Residual deviance: 313.42  on 326  degrees of freedom
## AIC: 317.42
##
## Number of Fisher Scoring iterations: 5
```

Predicting NBA wins

Same output re-presented:

```
stargazer(nbareg,type = "html", header = FALSE,single.row = TRUE)
```

	Dependent variable:
	win
Opp	-0.126*** (0.017)
Constant	13.320*** (1.669)
Observations	328
Log Likelihood	-156.709
Akaike Inf. Crit.	317.417
Note:	*p<0.1; **p<0.05; ***p<0.01

For every additional point an opponent scores against SAS in a game, the odds of winning decreases by approximately 12%, since exp(-0.126) = 0.88.

IDS 702

PREDICTING NBA WINS

```r
#Let's mean-center the temperature for interpretation.
nba$Opp_cent <- nba$Opp - mean(nba$Opp)
nbareg <- glm(win~Opp_cent,family=binomial(link=logit),data=nba)
stargazer(nbareg,type = "html", header = FALSE,single.row = TRUE)
```

	Dependent variable:
	win
Opp_cent	-0.126*** (0.017)
Constant	1.134*** (0.151)
Observations	328
Log Likelihood	-156.709
Akaike Inf. Crit.	317.417
Note:	*p<0.1; **p<0.05; ***p<0.01

IDS 702

The odds of SAS winning an nba game during this period, when the opposition scores approximately 97 points, is approximately 3.11, that is, exp(1.134).

The Challenger analysis

Confidence intervals for the coefficients. Remember that this is on the log-odds scale.

```
confint.default(nbareg) #Asymptotic
```

```
##                  2.5 %      97.5 %
## (Intercept)  0.8370288  1.43070311
## Opp_cent    -0.1581094 -0.09323567
```

```
confint(nbareg) #Based on the profile-likelihood
```

```
## Waiting for profiling to be done...
```

```
##                  2.5 %      97.5 %
## (Intercept)  0.8462671  1.44156134
## Opp_cent    -0.1599671 -0.09488784
```

Can you interpret the intervals?

IDS 702

The Challenger analysis

Let's transform to the odds scale.

```
exp(confint.default(nbareg)) #Asymptotic
```

```
##                 2.5 %    97.5 %
## (Intercept) 2.3094947 4.1816383
## Opp_cent    0.8537564 0.9109788
```

```
exp(confint(nbareg)) #Based on the profile-likelihood
```

```
## Waiting for profiling to be done...
```

```
##                 2.5 %    97.5 %
## (Intercept) 2.3309296 4.2272909
## Opp_cent    0.8521718 0.9094749
```

Can you interpret the intervals?

The Challenger analysis

We can get the predicted probabilities for the observed cases.

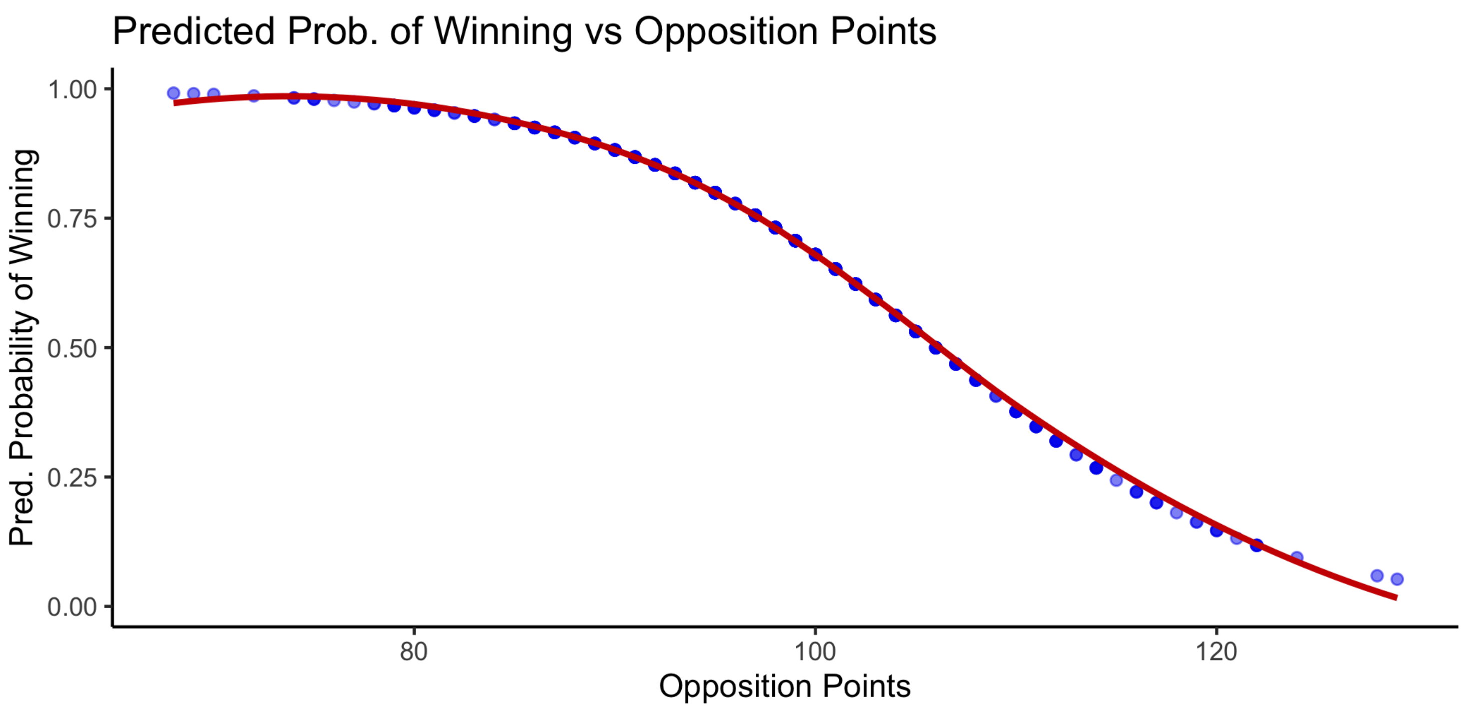
```
nba$predprobs <- predict(nbareg,type="response")
#use predict(logreg, type="link") for the logit scale
nba[1:20,]
```

```
##     Team WINorLOSS Opp win   Opp_cent predprobs
## 165  SAS         W 100   1   3.033537 0.6797523
## 166  SAS         L  94   0  -2.966463 0.8185670
## 167  SAS         W  92   1  -4.966463 0.8529607
## 168  SAS         L  98   0   1.033537 0.7318401
## 169  SAS         L 100   0   3.033537 0.6797523
## 170  SAS         W  85   1 -11.966463 0.9332502
## 171  SAS         W 100   1   3.033537 0.6797523
## 172  SAS         W  80   1 -16.966463 0.9632468
## 173  SAS         L  94   0  -2.966463 0.8185670
## 174  SAS         W  75   1 -21.966463 0.9800514
## 175  SAS         W  90   1  -6.966463 0.8817762
## 176  SAS         W  92   1  -4.966463 0.8529607
## 177  SAS         W  87   1  -9.966463 0.9157825
## 178  SAS         W 100   1   3.033537 0.6797523
## 179  SAS         W 104   1   7.033537 0.5621626
## 180  SAS         W  89   1  -7.966463 0.8942617
## 181  SAS         W 103   1   6.033537 0.5928153
## 182  SAS         L  95   0  -1.966463 0.7991510
## 183  SAS         W 101   1   4.033537 0.6518001
## 184  SAS         W 101   1   4.033537 0.6518001
```

The Challenger analysis

Useful to examine a plot of predicted probabilities by x, that is, opposition points.

```
ggplot(nba,aes(x=Opp, y=predprobs)) +
  geom_point(alpha = .5,colour="blue2") +
  geom_smooth(col="red3") + theme_classic() +
  labs(title="Predicted Prob. of Winning vs Opposition Points",x="Opposition Points",y="Pr
```



IDS 702

What's next?

Move on to the readings for the next module!

IDS 702