Math 651 Final Project

Mary Peng, Hillary Dunn, and Max Kearns

December 17, 2018

```
base = read.csv("data/base_data.csv")
head(base)
```

```
  X        country count year          gdp        pop host comm_soviet
1 1  United States   110 2008 1.471858e+13  304093966    0           0
2 2          China   100 2008 4.598206e+12 1324655000    1           1
3 3         Russia    72 2008 1.660844e+12  142742350    0           1
4 4 United Kingdom    47 2008 2.890564e+12   61806995    1           0
5 5      Australia    46 2008 1.052585e+12   21249200    1           0
6 6        Germany    41 2008 3.752366e+12   82110097    0           0
```

```
base.total = base[which(base$year!=2016),]
Olympic = base.total[,c(2,3,4,5,6,7,8)]

#Clean the data (Mary to add)
head(Olympic)
```

```
         country count year          gdp        pop host comm_soviet
1  United States   110 2008 1.471858e+13  304093966    0           0
2          China   100 2008 4.598206e+12 1324655000    1           1
3         Russia    72 2008 1.660844e+12  142742350    0           1
4 United Kingdom    47 2008 2.890564e+12   61806995    1           0
5      Australia    46 2008 1.052585e+12   21249200    1           0
6        Germany    41 2008 3.752366e+12   82110097    0           0
```

```
attach(Olympic)

#Make new dataframe with GDP / capita
Olympic_v2 <- data.frame(year, country, count, log_pop = log(pop), log_gdp_pcap = log(gdp/pop), host, c
head(Olympic_v2)
```

```
  year        country count  log_pop log_gdp_pcap host comm_soviet
1 2008  United States   110 19.53285    10.787285    0           0
2 2008          China   100 21.00442     8.152269    1           1
3 2008         Russia    72 18.77655     9.361795    0           1
4 2008 United Kingdom    47 17.93953    10.752946    1           0
5 2008      Australia    46 16.87183    10.810440    1           0
6 2008        Germany    41 18.22357    10.729836    0           0
```

```
attach(Olympic_v2)
```

```
The following objects are masked from Olympic:

    comm_soviet, count, country, host, year
```

```
Olympic.pois<-glm(count~log_pop + log_gdp_pcap + host + comm_soviet, family = poisson)
summary(Olympic.pois)
```

```

Call:
```

```
glm(formula = count ~ log_pop + log_gdp_pcap + host + comm_soviet,
    family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.9512  -1.8745  -0.6271   0.8053  10.5042

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.28882    0.25392 -44.458  < 2e-16 ***
log_pop       0.50479    0.01024  49.277  < 2e-16 ***
log_gdp_pcap  0.52117    0.01392  37.450  < 2e-16 ***
host          0.31070    0.04157   7.474 7.81e-14 ***
comm_soviet   1.02332    0.03693  27.713  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 7063.4  on 386  degrees of freedom
Residual deviance: 2211.0  on 382  degrees of freedom
AIC: 3588.8

Number of Fisher Scoring iterations: 5
```

```
P__disp(Olympic.pois)
```

```
[1] 2529.182464    6.620897
```

```
olympic.nb <- glm.nb(count~log_pop + log_gdp_pcap + host + comm_soviet)
summary(olympic.nb)
```

```
Call:
glm.nb(formula = count ~ log_pop + log_gdp_pcap + host + comm_soviet,
    init.theta = 2.104346269, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3544  -0.9235  -0.3356   0.3913   2.8745

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.16577    0.64686 -15.716  < 2e-16 ***
log_pop       0.49848    0.02881  17.300  < 2e-16 ***
log_gdp_pcap  0.40239    0.03174  12.677  < 2e-16 ***
host          0.69267    0.15932   4.348 1.38e-05 ***
comm_soviet   1.03376    0.09849  10.496  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.1043) family taken to be 1)

    Null deviance: 1039.7  on 386  degrees of freedom
Residual deviance:  386.8  on 382  degrees of freedom
```

```
AIC: 2321.2

Number of Fisher Scoring iterations: 1

              Theta:  2.104
          Std. Err.:  0.193

 2 x log-likelihood:  -2309.233
```

```
olympic.nb_leap <- leaps(y=count, x=Olympic_v2[,4:7])
Cp.nb<-round(olympic.nb_leap$Cp, 2)

cbind(olympic.nb_leap$which, olympic.nb_leap$size, Cp.nb)
```

```
  1 2 3 4    Cp.nb
1 1 0 0 0 2 198.46
1 0 0 1 0 2 215.98
1 0 1 0 0 2 336.41
1 0 0 0 1 2 370.52
2 1 0 1 0 3 108.70
2 1 1 0 0 3 119.18
2 1 0 0 1 3 180.88
2 0 1 1 0 3 199.48
2 0 0 1 1 3 206.80
2 0 1 0 1 3 318.21
3 1 1 0 1 4  58.39
3 1 1 1 0 4  60.95
3 1 0 1 1 4  87.77
3 0 1 1 1 4 178.38
4 1 1 1 1 5   5.00
```

```
xList <- names(Olympic_v2)[4:7]
vec <- olympic.nb_leap$which

#Name the columns in the grid
names(vec) <- paste("X", 1:4, sep="")

#Build matrix of formula for every row
allModelsList <- apply(vec, 1, function(x) as.formula(
  paste(c("count ~ 1", xList[x]), collapse = "+")))

#Calculate the coefficients for all 16 models
allModelsResults <- lapply(allModelsList,
                           function(x) glm.nb(x, data=Olympic_v2))


AIC.nb<-matrix(unlist(lapply(allModelsResults, function(x) round(extractAIC(x),2))), ncol = 2, byrow = T

cbind(olympic.nb_leap$which, olympic.nb_leap$size, Cp.nb, AIC.nb)
```

```
  1 2 3 4    Cp.nb  AIC.nb
1 1 0 0 0 2 198.46 2497.82
1 0 0 1 0 2 215.98 2633.68
1 0 1 0 0 2 336.41 2658.69
1 0 0 0 1 2 370.52 2695.53
```

```
2 1 0 1 0 3 108.70 2474.29
2 1 1 0 0 3 119.18 2428.33
2 1 0 0 1 3 180.88 2489.15
2 0 1 1 0 3 199.48 2601.80
2 0 0 1 1 3 206.80 2626.84
2 0 1 0 1 3 318.21 2614.08
3 1 1 0 1 4  58.39 2338.79
3 1 1 1 0 4  60.95 2417.52
3 1 0 1 1 4  87.77 2457.19
3 0 1 1 1 4 178.38 2556.83
4 1 1 1 1 5   5.00 2319.23
```

```
#PRESS
olympic.nb = PRESS(olympic.nb)
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1 = PRESS(glm.nb(count~log_pop))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX2 = PRESS(glm.nb(count~log_gdp_pcap))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX3 = PRESS(glm.nb(count~host))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX4 = PRESS(glm.nb(count~comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X2 = PRESS(glm.nb(count~log_pop+log_gdp_pcap))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X3 = PRESS(glm.nb(count~log_pop+host))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X4 = PRESS(glm.nb(count~log_pop+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX2X3 = PRESS(glm.nb(count~log_gdp_pcap+host))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX2X4 = PRESS(glm.nb(count~log_gdp_pcap+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX3X4 = PRESS(glm.nb(count~host+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X2X3 = PRESS(glm.nb(count~log_pop+log_gdp_pcap+host))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X2X4 = PRESS(glm.nb(count~log_pop+log_gdp_pcap+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX2X3X4 = PRESS(glm.nb(count~log_gdp_pcap+host+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
olympic.nbX1X3X4 = PRESS(glm.nb(count~log_pop+host+comm_soviet))
```

```
.........10.........20.........30.........40.........50
.........60.........70.........80.........90.........100
.........110.........120.........130.........140.........150
```

```
.........160.........170.........180.........190.........200
.........210.........220.........230.........240.........250
.........260.........270.........280.........290.........300
.........310.........320.........330.........340.........350
.........360.........370.........380.......
```

```
PRESS.nb <- rbind(olympic.nbX1$stat,
                          olympic.nbX3$stat,
                          olympic.nbX2$stat,
                          olympic.nbX4$stat,
                          olympic.nbX1X3$stat,
                          olympic.nbX1X2$stat,
                          olympic.nbX1X4$stat,
                          olympic.nbX2X3$stat,
                          olympic.nbX3X4$stat,
                          olympic.nbX2X4$stat,
                          olympic.nbX1X2X3$stat,
                          olympic.nbX1X2X4$stat,
                          olympic.nbX1X3X4$stat,
                          olympic.nbX2X3X4$stat,
                          olympic.nb$stat)

cbind(olympic.nb_leap$which, Size = olympic.nb_leap$size, Cp.nb, AIC.nb, 'PRESS.nb'=PRESS.nb)
```

```
  1 2 3 4 Size  Cp.nb  AIC.nb
1 1 0 0 0    2 198.46 2497.82 165158.9
1 0 0 1 0    2 215.98 2633.68 166715.3
1 0 1 0 0    2 336.41 2658.69 167645.3
1 0 0 0 1    2 370.52 2695.53 168281.8
2 1 0 1 0    3 108.70 2474.29 164471.6
2 1 1 0 0    3 119.18 2428.33 164383.2
2 1 0 0 1    3 180.88 2489.15 164950.4
2 0 1 1 0    3 199.48 2601.80 166271.2
2 0 0 1 1    3 206.80 2626.84 166615.0
2 0 1 0 1    3 318.21 2614.08 167153.7
3 1 1 0 1    4  58.39 2338.79 163997.8
3 1 1 1 0    4  60.95 2417.52 163451.0
3 1 0 1 1    4  87.77 2457.19 164085.9
3 0 1 1 1    4 178.38 2556.83 165880.7
4 1 1 1 1    5   5.00 2319.23 162957.3
```