NS125 PCW Session 12

2022-10-14

Installing packages and inspecting the data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readxl)
library(ellipse)
```

```
##
## Attaching package: 'ellipse'
```

```
## The following object is masked from 'package:graphics':
##
##     pairs
```

```
df <- read_excel("./data.xls", col_names = TRUE)
```

```
## Warning: Expecting numeric in AJ1369 / R1369C36: got '1 3'
```

```
head(df)
```

```
## # A tibble: 6 x 66
##    caseid start~1 start~2 start~3 endmo~4 endday endyear ongoi~5 ongoi~6 ongoi~7
##     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1  1.95e9      12      16    1945       5     27    1947      NA      NA      NA
## 2  1.95e9       2       8    1946       6     19    1949      NA      NA      NA
## 3  1.95e9       3      11    1946      10      9    1993      NA      NA      NA
## 4  1.95e9       4       9    1946       8      3    1960      NA      NA      NA
## 5  1.95e9       5      10    1946       8      9    1946      NA      NA      NA
## 6  1.95e9       7      21    1946      NA     NA      NA      10      22    1946
## # ... with 56 more variables: sender1 <dbl>, sender2 <dbl>, sender3 <dbl>,
## #   sender4 <dbl>, sender5 <dbl>, primarysender <dbl>, targetstate <dbl>,
## #   institution <dbl>, institutionid <chr>, targetinstitution <dbl>,
## #   otherinstitution <chr>, issue1 <dbl>, issue2 <dbl>, issue3 <dbl>,
## #   otherissue <chr>, threat <dbl>, threatid1 <dbl>, threatid2 <dbl>,
## #   threatid3 <dbl>, sanctiontypethreat <chr>,
## #   othersanctiontypethreatened <chr>, bspecif <dbl>, scommit <dbl>, ...
```

```
dim(df)
```

```
## [1] 1412   66
```

```
colnames(df)
```

```
##  [1] "caseid"                                "startmonth"
##  [3] "startday"                              "startyear"
##  [5] "endmonth"                              "endday"
##  [7] "endyear"                               "ongoingasofmonth"
##  [9] "ongoingasofday"                        "ongoingasofyear"
## [11] "sender1"                               "sender2"
## [13] "sender3"                               "sender4"
## [15] "sender5"                               "primarysender"
## [17] "targetstate"                           "institution"
## [19] "institutionid"                         "targetinstitution"
## [21] "otherinstitution"                      "issue1"
## [23] "issue2"                                "issue3"
## [25] "otherissue"                            "threat"
## [27] "threatid1"                             "threatid2"
## [29] "threatid3"                             "sanctiontypethreat"
## [31] "othersanctiontypethreatened"           "bspecif"
## [33] "scommit"                               "threatenedtargetinterest"
## [35] "dsanctions"                            "carrots"
## [37] "anticipatedtargetcosts"                "anticipatedtargetcostsfigureifav"
## [39] "tcurrency"                             "tyearofestimate"
## [41] "anticipatedsendercosts"                "scurrency"
## [43] "syearofestimate"                       "imposition"
## [45] "sancimpositionstartmonth"              "sancimpositionstartday"
## [47] "sancimpositionstartyear"               "sanctionidentity"
## [49] "sanctiontype"                          "othersanctiontype"
## [51] "implementationofdiplomaticsancti"      "carrotsduringsanction"
## [53] "carrotvalue"                           "carrotcurrency"
## [55] "carrotyear"                            "targetcosts"
## [57] "targeteconomiccostsfigure"             "targeteconomiccostscurrency"
## [59] "targeteconomiccostsyear"               "sendercosts"
## [61] "sendereconomiccostsfigure"             "sendereconomiccostscurrency"
## [63] "sendereconomiccostsyear"               "finaloutcome"
## [65] "settlementnaturesender"                "settlementnaturetarget"
```

Filtering only state 7, 8 and 9 for finaloutcome, and removing descriptive variables

```
chosen_cols <- c("startyear", "endyear", "primarysender", "targetstate", "institution", "targetinstituti
new_df <- df[names(df) %in% chosen_cols]
new_df <- new_df[new_df$finaloutcome %in% c(7, 8, 9), ]
head(new_df)
```

```
## # A tibble: 6 x 17
##   starty~1 endyear prima~2 targe~3 insti~4 targe~5 issue1 threat threa~6 sanct~7
##      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>  <dbl>  <dbl>   <dbl> <chr>
## 1     1945    1947       2     710       1       1      2      1       5 8
## 2     1946    1993     750     560       1       0      8      1       6 2
## 3     1946    1946       2     290       0      NA     13      0      NA <NA>
## 4     1947    1949       2     670       1       1      9      1       8 3
## 5     1947    1949       2     666       1       1      5      1       8 3
## 6     1947    1949       2     645       1       1      2      1       8 3
```

```
## # ... with 7 more variables: bspecif <dbl>, scommit <dbl>,
## #   anticipatedsendercosts <dbl>, sanctiontype <chr>, targetcosts <dbl>,
## #   sendercosts <dbl>, finaloutcome <dbl>, and abbreviated variable names
## #   1: startyear, 2: primarysender, 3: targetstate, 4: institution,
## #   5: targetinstitution, 6: threatid1, 7: sanctiontypethreat
```

```
dim(new_df)
```

```
## [1] 358  17
```

Creating a binary outcome:

```
new_df$outcome <- ifelse(new_df$finaloutcome == 7, 1, 0)
```

```
library(caTools)
library(ROCR)

lm <- glm (outcome ~ targetcosts + sendercosts, data=new_df, family="binomial")
lm
```

```
##
## Call:  glm(formula = outcome ~ targetcosts + sendercosts, family = "binomial",
##     data = new_df)
##
## Coefficients:
## (Intercept)  targetcosts  sendercosts
##     -1.4866       1.0306      -0.6242
##
## Degrees of Freedom: 318 Total (i.e. Null);  316 Residual
##   (39 observations deleted due to missingness)
## Null Deviance:       395.2
## Residual Deviance: 373.8     AIC: 379.8
```

```
summary(lm)
```

```
##
## Call:
## glm(formula = outcome ~ targetcosts + sendercosts, family = "binomial",
##     data = new_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6121  -0.7646  -0.7646   1.1986   1.9347
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4866     0.6301  -2.359   0.0183 *
## targetcosts   1.0306     0.2316   4.449 8.62e-06 ***
## sendercosts  -0.6242     0.5838  -1.069   0.2850
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 395.16  on 318  degrees of freedom
## Residual deviance: 373.80  on 316  degrees of freedom
##   (39 observations deleted due to missingness)
```

```
## AIC: 379.8
##
## Number of Fisher Scoring iterations: 4
```

Looking at a binary outcome for target costs vs sender costs, we can see there is a statistical significance only for target costs, where the higher the costs, the more likely the outcome is to succeed, however no such relationship exists for sender cost. Based on this analysis, it only partially support the hypothesis that cost of target > cost of sender.

```
# Predict test data based on model
predict_reg <- predict(lm, new_df, type = "response")
predict_reg <- ifelse(predict_reg >0.5, 1, 0)

# Evaluating model accuracy
# using confusion matrix
table(new_df$outcome, predict_reg)
```

```
##    predict_reg
##       0   1
##   0 215   5
##   1  90   9
```