R Notebook

Code

Horizontal differentiation notebook

Function definitions and data loading

Hide

```
# change wd for your own computer, this path is mine
setwd('C:/Gergo-mappa/projects/programming/projects/jads/sbm/jads_2021_g9_sbm')
```

```
Warning: The working directory was changed to C:/Gergo-mappa/projects/programming/projects/ja
ds/sbm/jads_2021_g9_sbm inside a notebook chunk. The working directory will be reset when the
chunk is finished running. Use the knitr root.dir option in the setup chunk to change the wor
king directory for notebook chunks.
```

Hide

```
#loading the data in formats we like

games_src <- read.csv('./data/final_games.csv', row.names = 1, header= TRUE)
tech_src <- read.csv('./data/final_tech.csv', row.names = 1, header= TRUE)
design_src <- read.csv('./data/final_design.csv', row.names = 1, header= TRUE)

setup_data <- function(dataframe){

  dataframe$location <- as.factor(dataframe$location)
  dataframe$Category <- as.factor(dataframe$Category)
  dataframe$launch_date <- as.Date(dataframe$launch_date)
  dataframe$degree_of_diff <- as.numeric(gsub("\\[|\\]", "", dataframe$degree_of_diff))
  dataframe$top_country <- as.factor(dataframe$top_country)
  dataframe$Staff_recommended <- as.logical(dataframe$Staff_recommended)
  dataframe$pledged_binary <- as.logical(dataframe$pledged_binary)
  dataframe$pledged_percentage <- NULL

  degree <- dataframe[, c(1:9)]

  return(list('degree'= degree, 'full' = dataframe))
}

binned_plot <- function(model){
  bootcamp2021::binnedplot(fitted(model),
            residuals(model, type = "response"),
            nclass = NULL,
            xlab = "Expected Values",
            ylab = "Average residual",
            main = "Binned residual plot",
            cex.pts = 0.8,
            col.pts = 1,
            col.int = "gray")

}
```

Design dataset analysis

Hide

```
design <- setup_data(design_src)
vec_model <- glm(pledged_binary ~ . - location - top_country, data = design$full, family = bi
nomial(link = 'logit'))
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Hide

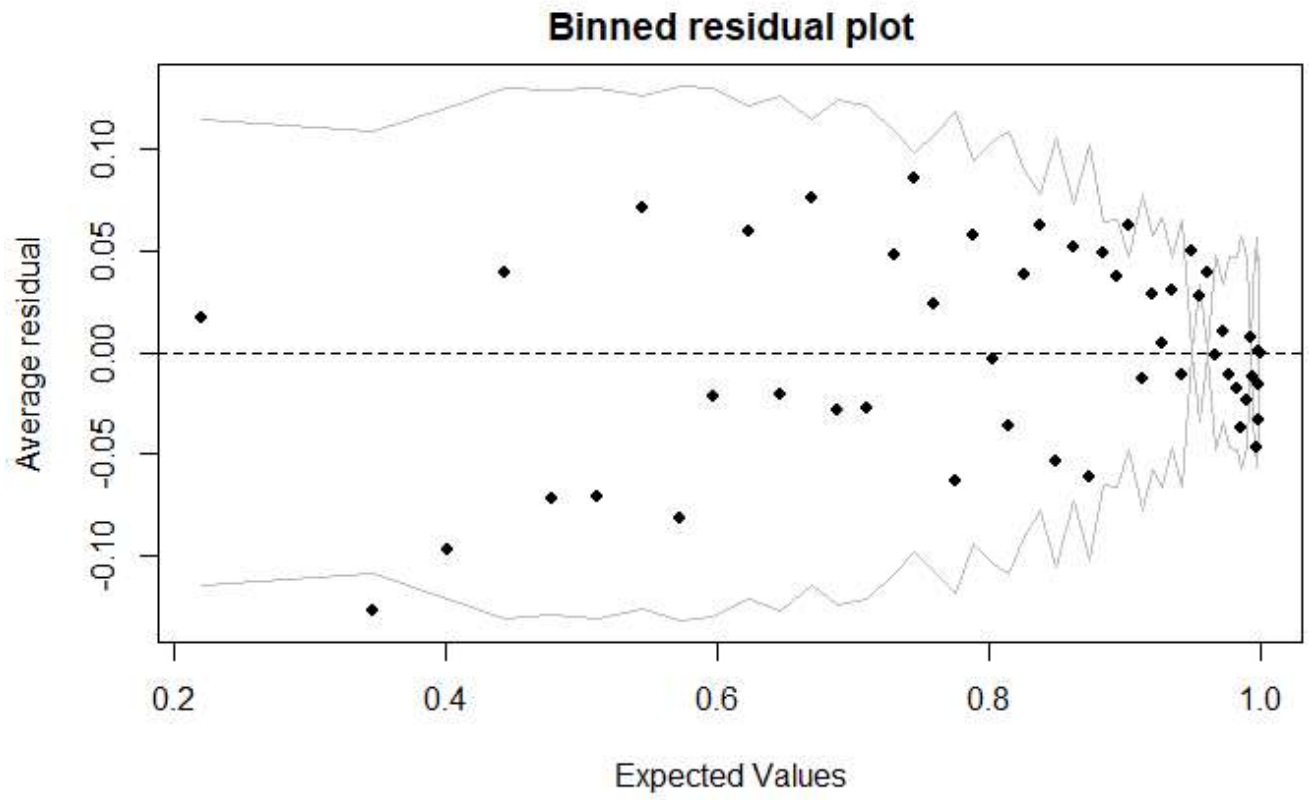
```
summary(vec_model)
```

```
Call:
glm(formula = pledged_binary ~ . - location - top_country, family = binomial(link = "logit"),
    data = design$full)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8354   0.0003   0.2218   0.5739   2.1317

Coefficients: (1 not defined because of singularities)
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                1.258e+00  2.277e+00   0.552  0.58062
CategoryCivic Design      -2.774e+00  1.315e+00  -2.110  0.03489 *
CategoryDesign            -5.952e-01  7.303e-01  -0.815  0.41509
CategoryGraphic Design    -2.101e+00  7.220e-01  -2.911  0.00361 **
CategoryInteractive Design -3.302e-01  1.355e+00  -0.244  0.80749
CategoryProduct Design    -3.891e-01  7.084e-01  -0.549  0.58283
CategoryTypography         1.327e+01  4.473e+02   0.030  0.97633
launch_date               -8.554e-05  9.954e-05  -0.859  0.39014
degree_of_diff             3.007e-01  3.368e-01   0.893  0.37201
Number_Backers             1.429e-02  1.129e-03  12.662  < 2e-16 ***
Creator_nb_projects        1.504e-01  2.407e-02   6.247 4.18e-10 ***
Staff_recommendedTRUE      7.812e+00  3.481e+02   0.022  0.98209
X1                         5.595e+00  3.243e+00   1.725  0.08454 .
X2                        -2.138e+00  1.727e+00  -1.238  0.21575
X3                         1.239e+00  1.572e+00   0.788  0.43069
X4                         1.926e+00  1.645e+00   1.170  0.24181
X5                        -8.953e-01  2.891e+00  -0.310  0.75682
X6                         2.551e+00  1.272e+00   2.006  0.04485 *
X7                         1.755e+00  1.885e+00   0.931  0.35176
X8                        -9.222e-01  1.830e+00  -0.504  0.61441
X9                         3.580e+00  1.349e+00   2.653  0.00797 **
X10                       -4.975e-02  2.152e+00  -0.023  0.98156
X11                        2.906e+00  1.872e+00   1.552  0.12059
X12                        1.661e+00  1.348e+00   1.232  0.21798
X13                        1.167e+00  1.502e+00   0.777  0.43721
X14                       -2.235e+00  1.301e+00  -1.718  0.08576 .
X15                        1.219e+00  1.649e+00   0.739  0.45993
X16                        3.660e+00  1.876e+00   1.951  0.05106 .
X17                        5.840e-01  1.458e+00   0.401  0.68865
X18                        9.393e-01  1.479e+00   0.635  0.52537
X19                        9.104e-01  1.250e+00   0.728  0.46657
X20                        2.309e+00  1.383e+00   1.670  0.09499 .
X21                       -4.949e-01  1.520e+00  -0.326  0.74469
X22                       -6.486e-01  1.412e+00  -0.459  0.64600
X23                        3.551e-01  1.682e+00   0.211  0.83280
X24                       -6.797e-01  1.555e+00  -0.437  0.66200
X25                       -1.322e-01  1.317e+00  -0.100  0.92005
X26                        3.801e+00  2.129e+00   1.785  0.07426 .
X27                       -3.261e-01  1.690e+00  -0.193  0.84699
X28                       -9.598e-02  1.302e+00  -0.074  0.94121
X29                       -3.034e-01  1.406e+00  -0.216  0.82912
X30                        4.862e-02  1.426e+00   0.034  0.97280
X31                        1.918e+00  2.067e+00   0.928  0.35337
X32                        5.996e-01  1.935e+00   0.310  0.75670
X33                        4.741e+00  1.805e+00   2.627  0.00860 **
X34                        1.124e+00  1.211e+00   0.928  0.35325
```

```
X35                      -2.347e+00  1.466e+00  -1.601  0.10940
X36                      -5.501e-01  1.894e+00  -0.290  0.77153
X37                      -5.180e-01  1.331e+00  -0.389  0.69704
X38                       6.502e-01  1.296e+00   0.502  0.61589
X39                       2.808e-01  1.670e+00   0.168  0.86646
X40                       1.523e-01  1.292e+00   0.118  0.90619
X41                       1.155e+00  1.497e+00   0.772  0.44006
X42                       4.883e-01  1.308e+00   0.373  0.70898
X43                       1.699e+00  1.449e+00   1.173  0.24089
X44                       1.281e+00  1.231e+00   1.041  0.29791
X45                       2.559e+00  1.302e+00   1.965  0.04944 *
X46                       6.759e-01  1.662e+00   0.407  0.68431
X47                       9.719e-01  1.295e+00   0.750  0.45313
X48                       1.850e+00  1.694e+00   1.092  0.27468
X49                       8.697e-02  1.289e+00   0.067  0.94619
X50                              NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3149.3  on 3498  degrees of freedom
Residual deviance: 2193.0  on 3438  degrees of freedom
AIC: 2315

Number of Fisher Scoring iterations: 14
```

Hide

```
binned_plot(vec_model)
```



Games dataset analysis

Hide

```
games <- setup_data(games_src)
vec_model <- glm(pledged_binary ~ . - location - top_country, data = games$full, family = bin
omial(link = 'logit'))
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Hide

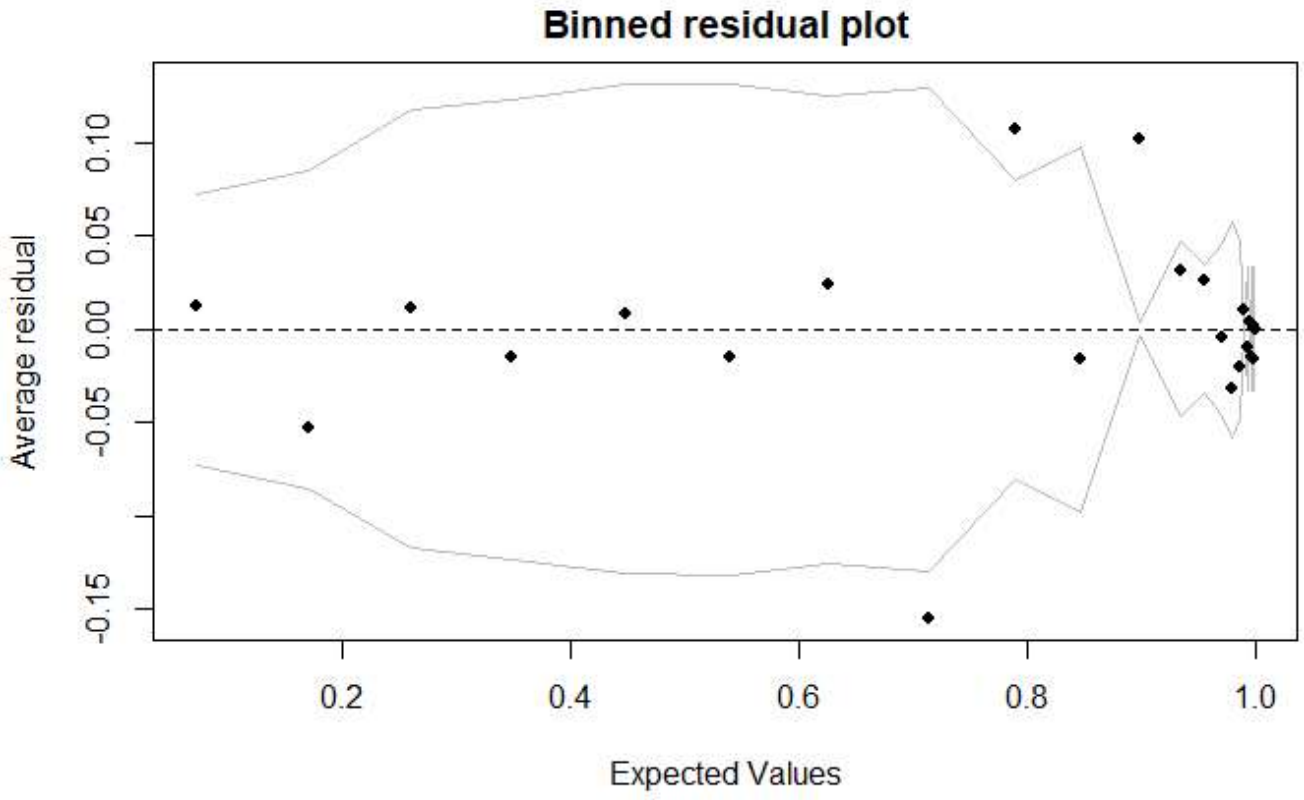
```
summary(vec_model)
```

```
Call:
glm(formula = pledged_binary ~ . - location - top_country, family = binomial(link = "logit"),
    data = games$full)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8188   0.0000   0.0021   0.0707   2.8883

Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             455.723812  37.655128  12.103  < 2e-16 ***
CategoryGaming Hardware   0.230406   1.720451   0.134 0.893465
CategoryLive Games       -1.606189   1.036901  -1.549 0.121375
CategoryMobile Games     -3.255795   1.267487  -2.569 0.010208 *
CategoryPlaying Cards    -1.956334   0.670756  -2.917 0.003539 **
CategoryPuzzles          -0.179621   1.327623  -0.135 0.892378
CategoryTabletop Games   -1.642350   0.542225  -3.029 0.002454 **
CategoryVideo Games      -2.174393   0.672529  -3.233 0.001224 **
launch_date              -0.025496   0.002095 -12.167  < 2e-16 ***
degree_of_diff            1.724105   0.632434   2.726 0.006408 **
Number_Backers            0.009290   0.001023   9.085  < 2e-16 ***
Creator_nb_projects       0.085967   0.023015   3.735 0.000188 ***
Staff_recommendedTRUE     4.453361 638.851358   0.007 0.994438
X1                        2.585717   2.372877   1.090 0.275847
X2                        6.761188   4.639101   1.457 0.144996
X3                        9.550256   8.715138   1.096 0.273156
X4                       -0.277795   3.823473  -0.073 0.942081
X5                        0.458678   2.855202   0.161 0.872372
X6                        7.564295   3.905181   1.937 0.052747 .
X7                        1.547887   2.876046   0.538 0.590439
X8                       12.213634   4.910183   2.487 0.012868 *
X9                        1.611566   2.062349   0.781 0.434554
X10                       5.979906   3.123789   1.914 0.055580 .
X11                       2.105092   1.947881   1.081 0.279827
X12                       4.713629   3.192813   1.476 0.139857
X13                       5.112099   2.080363   2.457 0.013998 *
X14                      13.870013   8.862239   1.565 0.117567
X15                       7.579224   3.969970   1.909 0.056244 .
X16                       0.768358   3.158752   0.243 0.807814
X17                      -1.240985   2.352379  -0.528 0.597816
X18                       1.522220   2.034297   0.748 0.454293
X19                       5.779114   2.230793   2.591 0.009581 **
X20                       1.876797   2.070564   0.906 0.364715
X21                       7.727680   3.341712   2.312 0.020751 *
X22                       6.773453   3.126640   2.166 0.030283 *
X23                      11.465084   6.898094   1.662 0.096500 .
X24                      -0.396139   1.962171  -0.202 0.840004
X25                      -1.019933   3.253530  -0.313 0.753912
X26                       1.204919   2.290796   0.526 0.598900
X27                      10.672888   3.654292   2.921 0.003493 **
X28                       5.699126   4.014324   1.420 0.155696
X29                      -2.006569   3.913792  -0.513 0.608167
X30                       2.210319   2.479811   0.891 0.372754
X31                      -1.015588   2.167479  -0.469 0.639386
X32                       1.217404   1.997725   0.609 0.542263
X33                       5.347793   3.138212   1.704 0.088364 .
```

```
X34                      6.235453   2.625851   2.375 0.017566 *
X35                      1.214860   2.343439   0.518 0.604173
X36                      3.715504   2.593429   1.433 0.151955
X37                      7.979125   2.297252   3.473 0.000514 ***
X38                      3.126595   2.917636   1.072 0.283891
X39                      0.062257   2.450807   0.025 0.979734
X40                      1.468454   1.891674   0.776 0.437588
X41                      5.773990   2.427190   2.379 0.017365 *
X42                      0.119798   2.368518   0.051 0.959661
X43                     -0.793382   2.824526  -0.281 0.778795
X44                      0.857083   2.059886   0.416 0.677349
X45                      9.259344   3.790512   2.443 0.014575 *
X46                      2.488117   2.936520   0.847 0.396827
X47                      1.753236   1.942480   0.903 0.366751
X48                      1.698361   1.887388   0.900 0.368202
X49                     -3.496808   2.446696  -1.429 0.152948
X50                            NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2163.30  on 3497  degrees of freedom
Residual deviance:  799.78  on 3436  degrees of freedom
AIC: 923.78

Number of Fisher Scoring iterations: 16
```

Hide

```
binned_plot(vec_model)
```



Tech dataset analysis

Hide

```
tech <- setup_data(tech_src)
vec_model <- glm(pledged_binary ~ . - location - top_country, data = tech$full, family = bino
mial(link = 'logit'))
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Hide

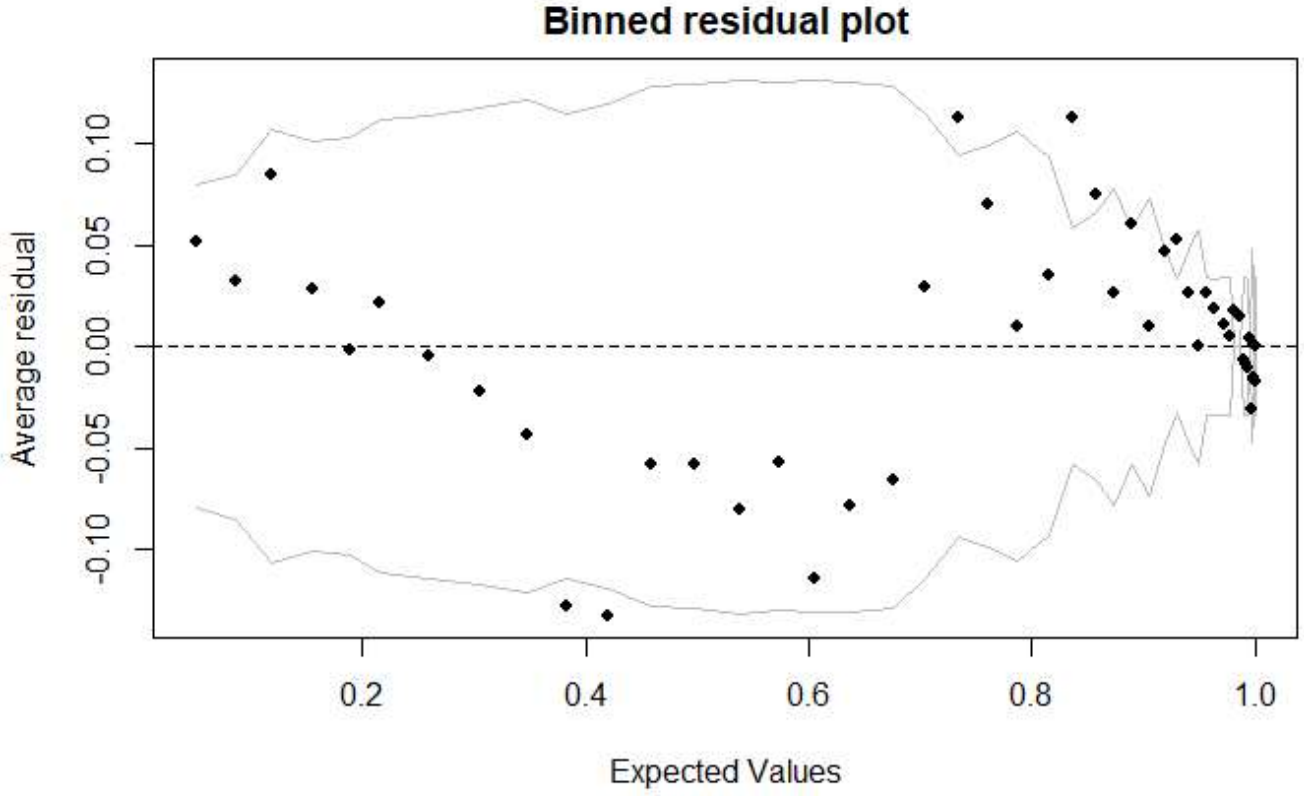
```
summary(vec_model)
```

```
Call:
glm(formula = pledged_binary ~ . - location - top_country, family = binomial(link = "logit"),
    data = tech$full)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5251   0.0000   0.0618   0.4214   2.6892

Coefficients: (2 not defined because of singularities)
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)               27.349032   2.898280   9.436  < 2e-16 ***
CategoryApps              -0.902008   0.913856  -0.987  0.32363
CategoryCamera Equipment  -0.381846   1.075823  -0.355  0.72264
CategoryDIY Electronics    0.447076   1.021980   0.437  0.66178
CategoryFabrication Tools -0.672035   1.259402  -0.534  0.59361
CategoryFlight            -0.504919   1.492382  -0.338  0.73511
CategoryGadgets           -0.434055   0.886978  -0.489  0.62458
CategoryHardware          -3.154897   0.868464  -3.633  0.00028 ***
CategoryMakerspaces        0.016887   1.239669   0.014  0.98913
CategoryRobots             0.463862   1.180358   0.393  0.69433
CategorySoftware          -3.920905   0.901679  -4.348 1.37e-05 ***
CategorySound              0.237971   1.011662   0.235  0.81403
CategorySpace Exploration 13.157617 393.956874   0.033  0.97336
CategoryTechnology        -0.131378   0.886695  -0.148  0.88221
CategoryWearables         -0.201293   0.944240  -0.213  0.83119
CategoryWeb               -1.525024   0.935628  -1.630  0.10311
launch_date               -0.001538   0.000120 -12.819  < 2e-16 ***
degree_of_diff             0.737246   0.377114   1.955  0.05059 .
Number_Backers             0.015298   0.001024  14.946  < 2e-16 ***
Creator_nb_projects        0.185060   0.044780   4.133 3.59e-05 ***
Staff_recommendedTRUE            NA         NA      NA       NA
X1                        -0.311583   2.021344  -0.154  0.87749
X2                         2.363508   4.319218   0.547  0.58424
X3                        -3.980614   2.282905  -1.744  0.08122 .
X4                         1.985853   3.616153   0.549  0.58289
X5                        -1.121259   2.062651  -0.544  0.58672
X6                         2.849890   2.519945   1.131  0.25808
X7                         0.597492   2.760250   0.216  0.82863
X8                         4.520501   3.540042   1.277  0.20162
X9                        -1.812601   2.125948  -0.853  0.39388
X10                        3.842372   3.457229   1.111  0.26640
X11                        0.630535   1.917110   0.329  0.74223
X12                       -0.004283   1.842596  -0.002  0.99815
X13                       -1.867951   2.321412  -0.805  0.42102
X14                       -3.295665   2.188282  -1.506  0.13205
X15                        0.577618   1.969960   0.293  0.76936
X16                        0.394143   1.947324   0.202  0.83960
X17                       -1.828707   3.053435  -0.599  0.54924
X18                        0.140670   1.841665   0.076  0.93912
X19                       -1.099626   2.053745  -0.535  0.59236
X20                       -1.632768   2.018717  -0.809  0.41862
X21                        0.911805   1.933535   0.472  0.63723
X22                        4.023747   2.897442   1.389  0.16492
X23                       -0.597899   1.915918  -0.312  0.75499
X24                       -3.078152   1.902531  -1.618  0.10568
X25                       -0.131122   2.026187  -0.065  0.94840
```

```
X26                         -0.880420   1.952169  -0.451  0.65199
X27                          1.043540   1.839754   0.567  0.57057
X28                         -0.903189   2.412037  -0.374  0.70807
X29                         -0.307961   2.406310  -0.128  0.89816
X30                          0.197072   1.883549   0.105  0.91667
X31                          0.626572   2.095939   0.299  0.76498
X32                         -2.906407   2.629076  -1.105  0.26895
X33                         -0.682269   2.025555  -0.337  0.73624
X34                         -1.759895   1.884705  -0.934  0.35042
X35                         -2.228714   1.968101  -1.132  0.25746
X36                          1.164705   2.354247   0.495  0.62079
X37                         -0.763472   1.845162  -0.414  0.67904
X38                         -1.831544   1.824077  -1.004  0.31533
X39                         -1.263576   1.804271  -0.700  0.48372
X40                         -1.932640   1.929106  -1.002  0.31642
X41                         -0.755646   2.248675  -0.336  0.73684
X42                         -1.971591   1.964055  -1.004  0.31546
X43                         -0.066385   1.840364  -0.036  0.97123
X44                          3.531294   2.998725   1.178  0.23896
X45                          1.959050   1.957661   1.001  0.31697
X46                         -1.017613   1.917765  -0.531  0.59568
X47                          2.923453   2.129165   1.373  0.16974
X48                          0.950167   2.239769   0.424  0.67140
X49                         -0.121028   1.819847  -0.067  0.94698
X50                               NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3799.7  on 3496  degrees of freedom
Residual deviance: 1981.1  on 3428  degrees of freedom
AIC: 2119.1

Number of Fisher Scoring iterations: 15
```

Hide

```
binned_plot(vec_model)
```


Binned residual plot
Average residual
0.10
0.05
0.00
-0.05
-0.10
0.2
0.4
0.6
0.8
1.0
Expected Values