4939_project

Zijiang Yan

4/9/2020

R Markdown

To install packages

```
## Warning: package 'devtools' was built under R version 3.6.3
```

```
## Loading required package: usethis
```

```
## Warning: package 'usethis' was built under R version 3.6.3
```

the 8th column V8 is convert logit value success to probability value. because in logit regression, the response variable must be positive.

```
## Warning: package 'car' was built under R version 3.6.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 3.6.1
```

```
##         judge            nation   rater     decision     language
##  MacGuigan :70   Lebanon    :71   no :254   no :270   English:253
##  Hugessen  :62   China      :68   yes:130   yes:114   French :131
##  Desjardins:46   Sri.Lanka  :63
##  Pratte    :42   Bulgaria   :36
##  Heald     :36   Somalia    :29
##  Stone     :33   El.Salvador:26
##  (Other)   :95   (Other)    :91
##      location      success
##  Montreal:138   Min.   :-2.0907
##  other   : 55   1st Qu.:-1.0986
##  Toronto :191   Median :-0.9946
##                 Mean   :-1.0204
##                 3rd Qu.:-0.7538
##                 Max.   : 0.4055
##
```

```
## 'data.frame':    384 obs. of  7 variables:
##  $ judge   : Factor w/ 10 levels "Desjardins","Heald",..: 2 2 2 5 1 9 8 5 5 8 ...
##  $ nation  : Factor w/ 17 levels "Argentina","Bulgaria",..: 11 17 5 4 11 11 7 16 16
3 ...
##  $ rater   : Factor w/ 2 levels "no","yes": 1 1 1 1 2 2 1 1 2 1 ...
##  $ decision: Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 1 1 ...
##  $ language: Factor w/ 2 levels "English","French": 1 1 1 2 2 1 1 1 2 1 ...
##  $ location: Factor w/ 3 levels "Montreal","other",..: 3 3 3 1 1 3 3 3 1 2 ...
##  $ success : num  -1.099 -0.754 -1.046 0.405 -1.099 ...
```

```
##     judge    nation     rater  decision  language  location   success
##  "factor"  "factor"  "factor"  "factor"  "factor"  "factor" "numeric"
```

```
##         judge         nation rater decision language location  success
## 13      Heald        Lebanon    no       no  English  Toronto -1.09861
## 15      Heald      Sri.Lanka    no       no  English  Toronto -0.75377
## 19      Heald    El.Salvador    no      yes  English  Toronto -1.04597
## 30  MacGuigan Czechoslovakia    no      yes   French Montreal  0.40547
## 36 Desjardins        Lebanon   yes      yes   French Montreal -1.09861
## 42      Stone        Lebanon   yes      yes  English  Toronto -1.09861
```

```
## [1] 384   7
```

We introduced the dummy variable in rater and decision for the further approach.

the binomial GLM regression in this dataset

```
##         judge         nation rater decision language location  success
## 13      Heald        Lebanon    no       no  English  Toronto -1.09861
## 15      Heald      Sri.Lanka    no       no  English  Toronto -0.75377
## 19      Heald    El.Salvador    no      yes  English  Toronto -1.04597
## 30  MacGuigan Czechoslovakia    no      yes   French Montreal  0.40547
## 36 Desjardins        Lebanon   yes      yes   French Montreal -1.09861
## 42      Stone        Lebanon   yes      yes  English  Toronto -1.09861
```

```
##
## Call:
## glm(formula = decision ~ judge + nation + rater + language +
##     location + success2, family = binomial, data = Greene)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9715  -0.6739  -0.3489   0.6588   2.5846
##
## Coefficients: (1 not defined because of singularities)
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.15828    1.33790  -0.866 0.386632
## judgeHeald             -1.40620    0.55184  -2.548 0.010829 *
## judgeHugessen          -1.41868    0.55890  -2.538 0.011137 *
## judgeIacobucci         -2.69319    0.74246  -3.627 0.000286 ***
## judgeMacGuigan         -1.36944    0.48117  -2.846 0.004426 **
## judgeMahoney           -0.81159    0.55213  -1.470 0.141579
## judgeMarceau            1.50382    0.65534   2.295 0.021750 *
## judgePratte            -2.10748    0.62924  -3.349 0.000810 ***
## judgeStone             -1.83682    0.59001  -3.113 0.001851 **
## judgeUrie              -0.08193    0.77469  -0.106 0.915779
## nationBulgaria         -0.73411    1.30616  -0.562 0.574088
## nationChina            -0.15081    1.20044  -0.126 0.900024
## nationCzechoslovakia    2.77967    1.27626   2.178 0.029407 *
## nationEl.Salvador      -0.25468    1.24694  -0.204 0.838165
## nationFiji            -16.34334 3956.18056  -0.004 0.996704
## nationGhana            -1.83422    1.64276  -1.117 0.264190
## nationGuatemala        -0.04252    1.65654  -0.026 0.979522
## nationIndia            18.43742 2085.16307   0.009 0.992945
## nationIran             -0.56407    1.35362  -0.417 0.676890
## nationLebanon          -0.01163    1.19484  -0.010 0.992235
## nationNicaragua        -0.85140    1.49871  -0.568 0.569977
## nationNigeria          -1.70344    1.66652  -1.022 0.306706
## nationPakistan         -0.21638    1.61055  -0.134 0.893125
## nationPoland          -16.65578 1080.81933  -0.015 0.987705
## nationSomalia          -0.04774    1.25904  -0.038 0.969753
## nationSri.Lanka         0.35742    1.19019   0.300 0.763941
## rateryes                1.49508    0.29526   5.064 4.11e-07 ***
## languageFrench         -0.28821    0.65762  -0.438 0.661202
## locationother           1.35424    0.73336   1.847 0.064803 .
## locationToronto         1.09464    0.66303   1.651 0.098744 .
## success2                     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: 467.09  on 383  degrees of freedom
```

```
## Residual deviance: 336.49  on 354  degrees of freedom
## AIC: 396.49
##
## Number of Fisher Scoring iterations: 16
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: decision
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                       383     467.09
## judge     9   40.403       374     426.69
## nation   16   52.848       358     373.84
## rater     1   24.734       357     349.11
## language  1    9.067       356     340.04
## location  2    3.551       354     336.49
## success2  0    0.000       354     336.49
```

```
## Gineralized Linear Model : calculate the p-value for the deviance goodness of fit
```

```
## Warning in pchisq(model1$deviance, df = model1$residuals, lower.tail = FALSE):
## NaNs produced
```

```
##           13           15           19           30           36           42
##          NaN          NaN 4.341948e-69 9.500276e-74 7.708645e-74 3.565552e-73
```

In this model ,similar to the multivariate modeling selection, the judge and judge contributes more in this model.

Due to the all pvalue is small in the pearson chi square test table, we assume these variables are independent except location and language. Location and language are highly correlation.

For example, we ususally use English in Toronto and use French in Montreal.

We will continue to discuss correlation matrix in the facorial design parts.

```
##
## Call:
## glm(formula = decision ~ judge + nation + rater + location, family = binomial,
##     data = Greene)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.971  -0.673  -0.347   0.658   2.581
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.412e+00  1.209e+00  -1.167 0.243056
## judgeHeald           -1.401e+00  5.513e-01  -2.540 0.011071 *
## judgeHugessen        -1.454e+00  5.511e-01  -2.638 0.008348 **
## judgeIacobucci       -2.691e+00  7.415e-01  -3.629 0.000285 ***
## judgeMacGuigan       -1.380e+00  4.796e-01  -2.877 0.004014 **
## judgeMahoney         -8.055e-01  5.513e-01  -1.461 0.143959
## judgeMarceau          1.489e+00  6.551e-01   2.273 0.023054 *
## judgePratte          -2.124e+00  6.291e-01  -3.377 0.000734 ***
## judgeStone           -1.833e+00  5.895e-01  -3.109 0.001878 **
## judgeUrie            -6.119e-02  7.758e-01  -0.079 0.937130
## nationBulgaria       -7.221e-01  1.309e+00  -0.552 0.581190
## nationChina          -1.250e-01  1.203e+00  -0.104 0.917239
## nationCzechoslovakia  2.793e+00  1.278e+00   2.185 0.028870 *
## nationEl.Salvador    -2.154e-01  1.247e+00  -0.173 0.862844
## nationFiji           -1.628e+01  3.956e+03  -0.004 0.996717
## nationGhana          -1.802e+00  1.643e+00  -1.097 0.272818
## nationGuatemala      -1.155e-02  1.658e+00  -0.007 0.994441
## nationIndia           1.846e+01  2.086e+03   0.009 0.992938
## nationIran           -5.656e-01  1.358e+00  -0.416 0.677104
## nationLebanon         3.263e-03  1.198e+00   0.003 0.997827
## nationNicaragua      -8.162e-01  1.500e+00  -0.544 0.586372
## nationNigeria        -1.626e+00  1.662e+00  -0.978 0.327891
## nationPakistan       -2.004e-01  1.615e+00  -0.124 0.901238
## nationPoland         -1.665e+01  1.080e+03  -0.015 0.987698
## nationSomalia        -4.332e-02  1.262e+00  -0.034 0.972610
## nationSri.Lanka       3.864e-01  1.192e+00   0.324 0.745793
## rateryes              1.493e+00  2.949e-01   5.060 4.19e-07 ***
## locationother         1.579e+00  5.288e-01   2.987 0.002819 **
## locationToronto       1.320e+00  4.234e-01   3.117 0.001828 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 467.09  on 383  degrees of freedom
## Residual deviance: 336.68  on 355  degrees of freedom
## AIC: 394.68
```

```
##
## Number of Fisher Scoring iterations: 16
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: decision
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                       383     467.09
## judge     9   40.403       374     426.69
## nation   16   52.848       358     373.84
## rater     1   24.734       357     349.11
## location  2   12.425       355     336.68
```

```
## Gineralized Linear Model : calculate the p-value for the deviance goodness of fit
```

```
## Warning in pchisq(model1$deviance, df = model1$residuals, lower.tail = FALSE):
## NaNs produced
```

```
##           13           15           19           30           36           42
##          NaN          NaN 4.341948e-69 9.500276e-74 7.708645e-74 3.565552e-73
```

chisquare test demonstrate interaction is small. In order to simplify this GLM model, we introduced a model selection here.

```
## Start:  AIC=396.49
## decision ~ judge + nation + rater + location + language + success
##
##
## Step:  AIC=396.49
## decision ~ judge + nation + rater + location + language
##
##            Df Deviance    AIC
## - language  1   336.68 394.68
## - location  2   340.04 396.04
## <none>          336.49 396.49
## - nation   16   390.20 418.20
## - rater     1   363.83 421.83
## - judge     9   391.17 433.17
##
## Step:  AIC=394.68
## decision ~ judge + nation + rater + location
##
##            Df Deviance    AIC
## <none>          336.68 394.68
## + language  1   336.49 396.49
## - location  2   349.11 403.11
## - nation   16   390.54 416.54
## - rater     1   363.98 419.98
## - judge     9   391.54 431.54
```

```
##
## Call:  glm(formula = decision ~ judge + nation + rater + location, family = binomia
l,
##     data = Greene)
##
## Coefficients:
##          (Intercept)            judgeHeald         judgeHugessen
##            -1.411836             -1.400514             -1.453742
##       judgeIacobucci        judgeMacGuigan          judgeMahoney
##            -2.690672             -1.379900             -0.805550
##         judgeMarceau           judgePratte            judgeStone
##             1.488844             -2.124165             -1.832683
##            judgeUrie        nationBulgaria           nationChina
##            -0.061193             -0.722086             -0.125001
## nationCzechoslovakia     nationEl.Salvador            nationFiji
##             2.792988             -0.215372            -16.279901
##          nationGhana       nationGuatemala           nationIndia
##            -1.801683             -0.011552             18.464771
##           nationIran         nationLebanon       nationNicaragua
##            -0.565565              0.003263             -0.816159
##        nationNigeria        nationPakistan          nationPoland
##            -1.625930             -0.200360            -16.653716
##        nationSomalia       nationSri.Lanka              rateryes
##            -0.043322              0.386417              1.492501
##        locationother       locationToronto
##             1.579410              1.319671
##
## Degrees of Freedom: 383 Total (i.e. Null);  355 Residual
## Null Deviance:       467.1
## Residual Deviance: 336.7     AIC: 394.7
```

Based on this model selection. the model is selected by deviance test and AIC, that is we would like to find the highest deviance difference and smallest AIC.

The original glm(formula = decision ~ judge + nation + rater + language + location + success2, family = binomial, data = Greene)

The reduced model we have

glm(formula = decision ~ judge + nation + rater + location, family = binomial, data = Greene)

in this reduced model, we find decision only concerned about judge nation rater and location. we remove the other variables.

```
##
## Call:
## glm(formula = decision ~ judge + nation + location, family = binomial,
##     data = Greene)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7518  -0.7669  -0.4598   0.8212   2.4795
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.19147    1.05125  -0.182 0.855474
## judgeHeald             -1.36576    0.52520  -2.600 0.009310 **
## judgeHugessen          -1.34747    0.52438  -2.570 0.010181 *
## judgeIacobucci         -2.60665    0.71289  -3.656 0.000256 ***
## judgeMacGuigan         -1.33559    0.45737  -2.920 0.003499 **
## judgeMahoney           -0.64986    0.52379  -1.241 0.214720
## judgeMarceau            1.14586    0.62567   1.831 0.067040 .
## judgePratte            -2.03398    0.61134  -3.327 0.000878 ***
## judgeStone             -1.76671    0.56019  -3.154 0.001612 **
## judgeUrie              -0.27016    0.73985  -0.365 0.714992
## nationBulgaria         -1.48590    1.15919  -1.282 0.199896
## nationChina            -0.80124    1.05244  -0.761 0.446465
## nationCzechoslovakia    2.08799    1.11601   1.871 0.061353 .
## nationEl.Salvador      -0.44489    1.11013  -0.401 0.688599
## nationFiji            -17.42160 3956.18050  -0.004 0.996486
## nationGhana            -2.34631    1.55282  -1.511 0.130789
## nationGuatemala        -0.72025    1.53258  -0.470 0.638383
## nationIndia            18.11702 2271.12069   0.008 0.993635
## nationIran             -1.21643    1.23399  -0.986 0.324249
## nationLebanon          -0.65461    1.04764  -0.625 0.532073
## nationNicaragua        -0.99825    1.39812  -0.714 0.475230
## nationNigeria          -1.77604    1.52489  -1.165 0.244140
## nationPakistan         -0.09671    1.45262  -0.067 0.946920
## nationPoland          -17.30549 1068.45533  -0.016 0.987077
## nationSomalia          -0.45199    1.11367  -0.406 0.684849
## nationSri.Lanka        -0.03326    1.04860  -0.032 0.974698
## locationother           1.39447    0.51482   2.709 0.006755 **
## locationToronto         1.13856    0.41077   2.772 0.005576 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 467.09  on 383  degrees of freedom
## Residual deviance: 363.98  on 356  degrees of freedom
## AIC: 419.98
```

```
##
## Number of Fisher Scoring iterations: 16
```

In this model, we find judge and location contributes more. So the appeal decision depends on judge and location.

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: decision
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                       383     467.09
## judge     9   40.403       374     426.69
## nation   16   52.848       358     373.84
## location  2    9.866       356     363.98
```

In this analysis table

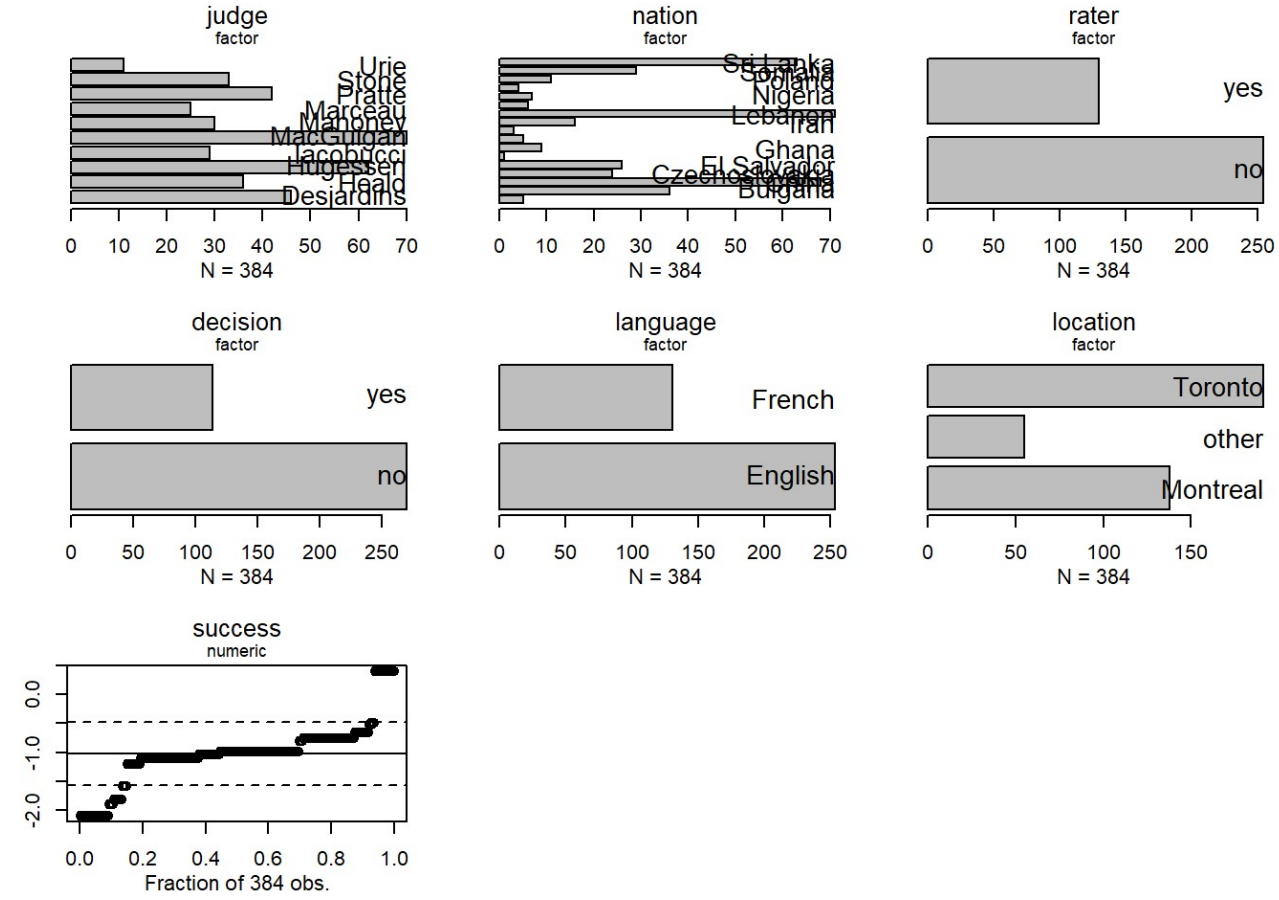
Type I SS: fits the SS sequentially. order of the variables matters. Type II SS: hierarchical, or partially sequential. order still matters a bit. Type III SS: Marginal or orthogonal. Order does not matter at all.

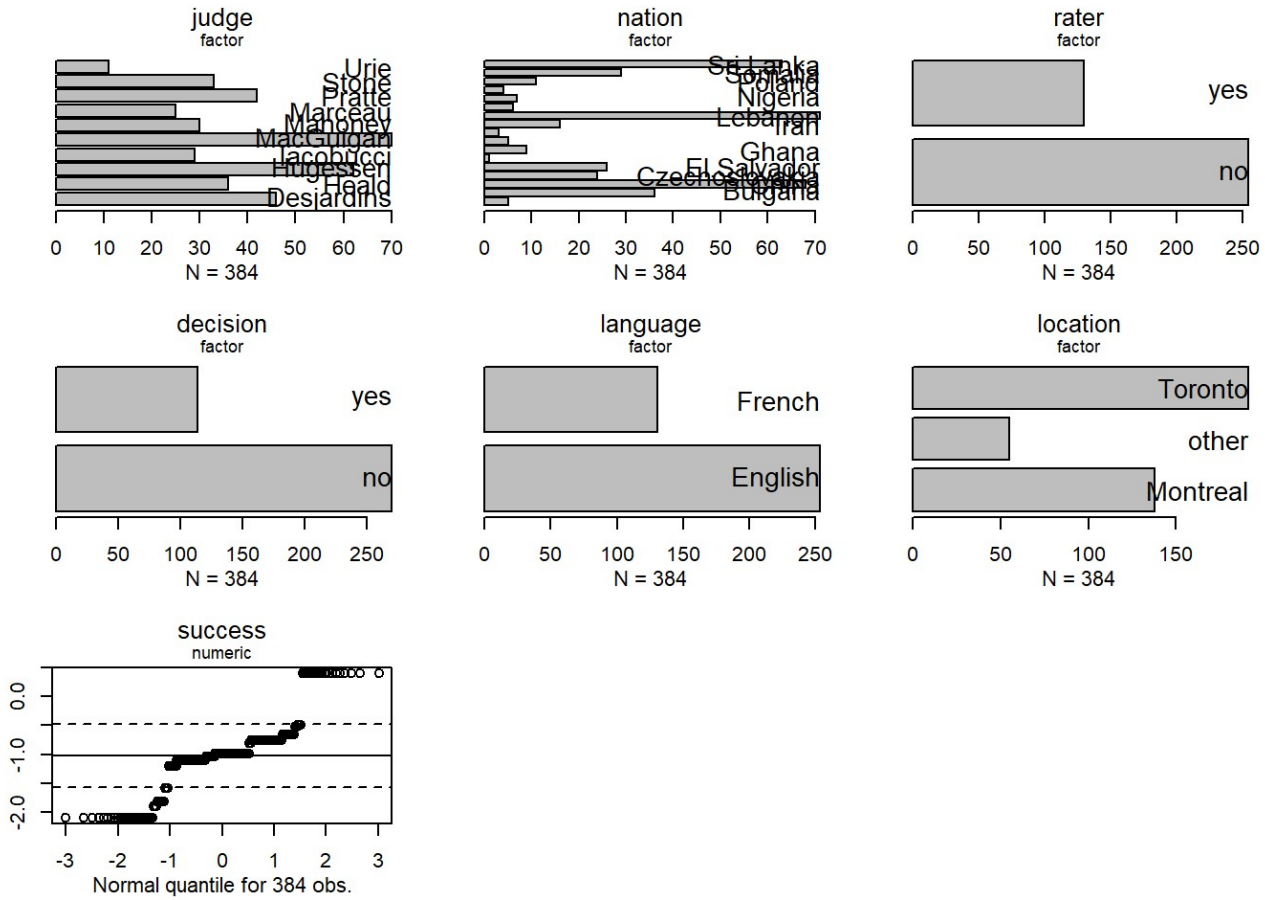
For likelihood Ratio Test, judge and nation matters a lot.

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: decision
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                       383     467.09
## judge     9   40.403       374     426.69 6.423e-06 ***
## nation   16   52.848       358     373.84 7.986e-06 ***
## location  2    9.866       356     363.98  0.007203 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From chisquare test and C statistics draws a similar conclusion, judge,nation and location matters a lot.

Extended Quantile Plots





In the Extended Quantile Plots, this is the plot for overall dataset. we can simply conclude

the rater is more optimistic than the judge, because the number of rator said yes is greater than thwe final appeal passed.

the number of cases occured in Toronto and Montreal is higher than in the other place.

Some judge Judges hear many cases. The appeals is not equalized distribute to the judges.

Some nations are eager to appeal.

```
## character(0)
```

```
## [1] "Argentina"
```

```
##
##        no       yes
## 0.6614583 0.3385417
```

```
## the proportion
```

the plot shows the the probability to pass the appeal vs the country

We assume there are two models 1. rater decision vs the country 2. judge decision vs the country

The judge is more strict than rator. The judge has preference or have some prejudice to pass in some specific country,like Czechoslovakia.

#Factorial design we defined countries as different numbers Argentina 1 Ghana 2 Nigeria 3 Bulgaria 4 Guatemala 5 Pakistan 6 China 7 India 8 Poland 9 Czechoslovakia 10 Iran 11 Somalia 12 El.Salvador 13 Lebanon 14 Sri.Lanka 15 Fiji 16 Nicaragua 17

we defined judge as different index Desjardins 1 Heald 2 Hugessen 3 Iacobucci 4 MacGuigan 5 Mahoney 6 Marceau 7 Pratte 8 Stone 9 Urie 10

defined the dummy variable as rater and decision Yes 1 No 0

defined the dummy variable as locations Toronto 1 Montreal 2 Other 0

defined the dummy variable as language English 0 French 1

```
## # A tibble: 384 x 7
##    judge nation rater decision language location success
##    <dbl>  <dbl> <dbl>    <dbl>    <dbl>    <dbl>   <dbl>
##  1     2      5     0        0        1        1  -1.10
##  2     2     15     0        0        1        1  -0.754
##  3     2      4     0        1        1        1  -1.05
##  4     5      4     0        1        0        2   0.405
##  5     1      5     1        1        0        2  -1.10
##  6     9      5     1        1        1        1  -1.10
##  7     8      4     0        0        1        1  -1.21
##  8     5     14     0        0        1        1  -0.995
##  9     5     14     1        0        0        2  -0.995
## 10     8      4     0        0        1        0  -0.995
## # ... with 374 more rows
```

```
head(greene_factorial) #List heading 6 rows.
```

```
## The following objects are masked from 'package:spida2':
##
##     fillin, Lag, na.include
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
##          judge nation rater decision language location success
## judge     1.00   0.00 -0.06    -0.05     0.10    -0.08   -0.04
## nation    0.00   1.00  0.09     0.07     0.13    -0.05    0.15
## rater    -0.06   0.09  1.00     0.27     0.00     0.03    0.13
## decision -0.05   0.07  0.27     1.00     0.11    -0.07    0.28
## language  0.10   0.13  0.00     0.11     1.00    -0.77   -0.01
## location -0.08  -0.05  0.03    -0.07    -0.77     1.00    0.09
## success  -0.04   0.15  0.13     0.28    -0.01     0.09    1.00
##
## n= 384
##
##
## P
##          judge  nation rater  decision language location success
## judge           0.9556 0.2441 0.3379   0.0478   0.1151   0.4911
## nation   0.9556        0.0753 0.1769   0.0098   0.3155   0.0025
## rater    0.2441 0.0753        0.0000   0.9369   0.5345   0.0133
## decision 0.3379 0.1769 0.0000          0.0363   0.1537   0.0000
## language 0.0478 0.0098 0.9369 0.0363            0.0000   0.8848
## location 0.1151 0.3155 0.5345 0.1537   0.0000           0.0689
## success  0.4911 0.0025 0.0133 0.0000   0.8848   0.0689
```

From the correlation table, we find language and location are highly correlated.we need to drop one of these variables. Doing a regression analysis in this factorial design.

```
##
## Call:
## glm(formula = decision ~ judge + nation + rater + location +
##     language + success, family = binomial, data = greene_factorial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6656  -0.8110  -0.5486   1.0089   2.5818
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.1141013  0.6992446   0.163    0.870
## judge       -0.0350472  0.0460443  -0.761    0.447
## nation       0.0002251  0.0264892   0.008    0.993
## rater        1.1771754  0.2480543   4.746 2.08e-06 ***
## location    -0.2814176  0.2831967  -0.994    0.320
## language     0.4813728  0.4134257   1.164    0.244
## success      1.3457803  0.2702803   4.979 6.38e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 467.09  on 383  degrees of freedom
## Residual deviance: 404.99  on 377  degrees of freedom
## AIC: 418.99
##
## Number of Fisher Scoring iterations: 4
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: decision
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                       383     467.09
## judge     1   0.9272       382     466.17
## nation    1   1.7923       381     464.37
## rater     1  25.7350       380     438.64
## location  1   2.7969       379     435.84
## language  1   1.8151       378     434.03
## success   1  29.0315       377     404.99
```