

Behave_Project_Modeling

Yu Zhu

12/1/2020

```
library(readxl)
Behave <- read_excel("~/Desktop/ucsc/courses/stat 204/project/ProjectDataset_behave.xlsx")

Behave$ParticipantNum = as.factor(Behave$ParticipantNum)
Behave$Confidence = as.factor(Behave$Confidence)
Behave$StrengthLevel = as.factor(Behave$StrengthLevel)
```

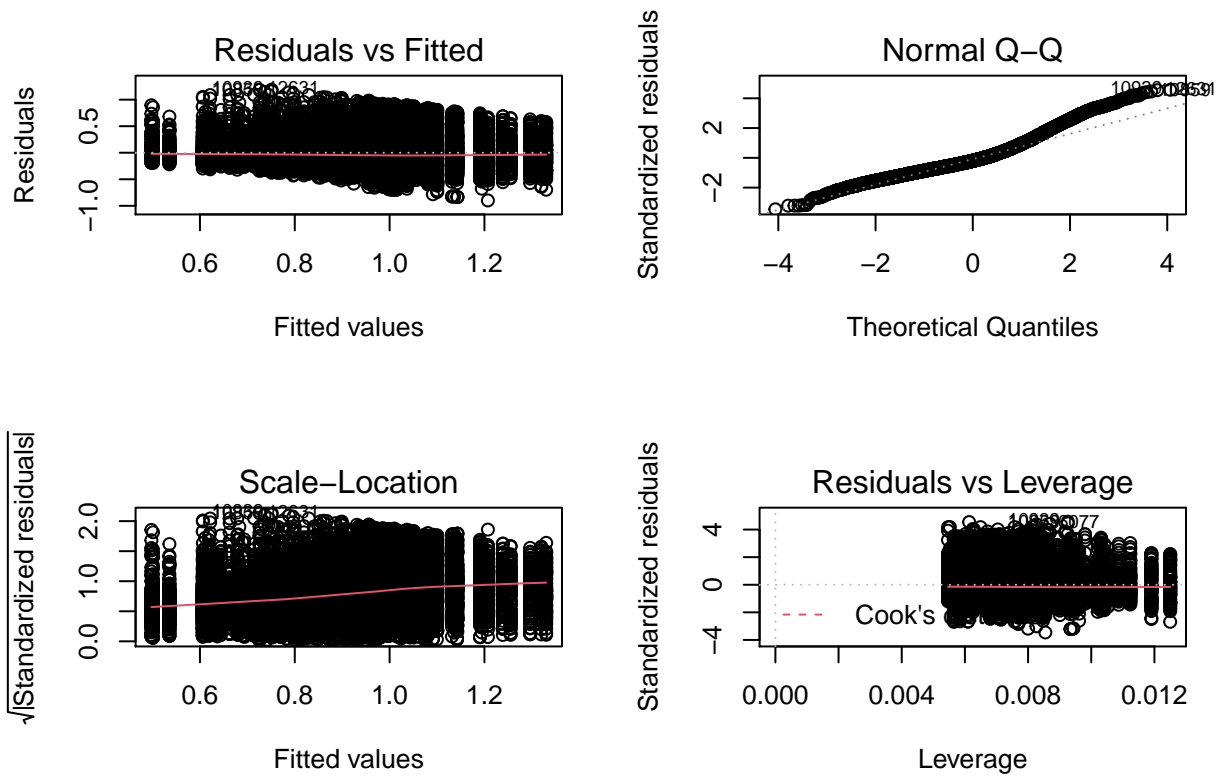
Models

1. Randomized Block Model

```
L_aov = aov(ResponseTime~StrengthLevel*ParticipantNum, data = Behave)
summary(L_aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## StrengthLevel    5  106.2   21.246   311.63 <2e-16 ***
## ParticipantNum   24  396.5   16.520   242.31 <2e-16 ***
## StrengthLevel:ParticipantNum 120   34.4    0.287    4.21 <2e-16 ***
## Residuals       20297 1383.8    0.068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(L_aov)
```



Diagnostics

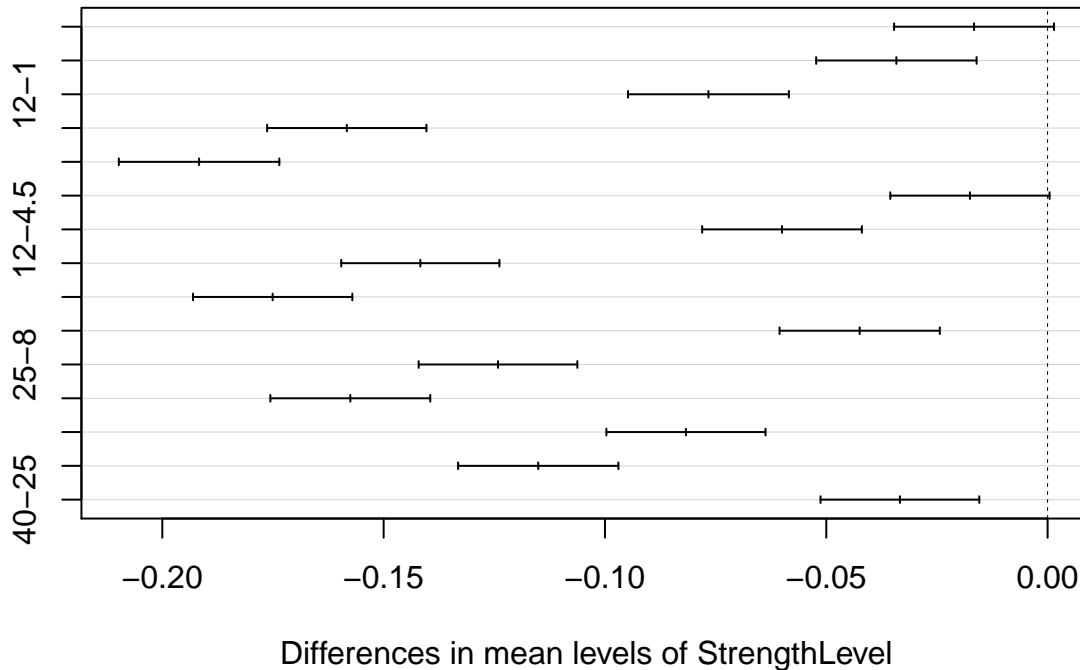
```
CIs_strLevel = TukeyHSD(L_aov, which = 1)
CIs_strLevel
```

Tukey's HSD

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = ResponseTime ~ StrengthLevel * ParticipantNum, data = Behave)
##
## $StrengthLevel
##          diff          lwr          upr      p adj
## 4.5-1 -0.01661997 -0.03467940  0.0014394563 0.0918041
## 8-1   -0.03416539 -0.05228362 -0.0160471528 0.0000011
## 12-1  -0.07662106 -0.09478733 -0.0584547915 0.0000000
## 25-1  -0.15832734 -0.17632047 -0.1403342062 0.0000000
## 40-1  -0.19169213 -0.20981964 -0.1735646275 0.0000000
## 8-4.5 -0.01754542 -0.03553467  0.0004438399 0.0607700
## 12-4.5 -0.06000109 -0.07803873 -0.0419634553 0.0000000
## 25-4.5 -0.14170736 -0.15957061 -0.1238441168 0.0000000
## 40-4.5 -0.17507216 -0.19307075 -0.1570735684 0.0000000
## 12-8   -0.04245567 -0.06055219 -0.0243591624 0.0000000
## 25-8   -0.12416195 -0.14208465 -0.1062392510 0.0000000
## 40-8   -0.15752675 -0.17558434 -0.1394691482 0.0000000
## 25-12  -0.08170627 -0.09967753 -0.0637350176 0.0000000
## 40-12  -0.11507107 -0.13317686 -0.0969652765 0.0000000
## 40-25  -0.03336480 -0.05129687 -0.0154327277 0.0000017
```

```
plot(CIs_strLevel)
```

95% family-wise confidence level



```
library(DescTools)
ScheffeTest(x=L_aov, which="StrengthLevel")
```

Scheffe's method

```
##
## Posthoc multiple comparisons of means: Scheffe Test
## 95% family-wise confidence level
##
## $StrengthLevel
##      diff      lwr.ci      upr.ci    pval
## 4.5-1 -0.01661997 -0.1013099  0.068069988 1.0000
## 8-1    -0.03416539 -0.1191311  0.050800346 1.0000
## 12-1   -0.07662106 -0.1618121  0.008569938 0.5888
## 25-1   -0.15832734 -0.2427064 -0.073948283 < 2e-16 ***
## 40-1   -0.19169213 -0.2767013 -0.106682925 < 2e-16 ***
## 8-4.5  -0.01754542 -0.1019063  0.066815467 1.0000
## 12-4.5 -0.06000109 -0.1445888  0.024586669 1.0000
## 25-4.5 -0.14170736 -0.2254773 -0.057937399 < 2e-16 ***
## 40-4.5 -0.17507216 -0.2594768 -0.090667493 < 2e-16 ***
## 12-8    -0.04245567 -0.1273195  0.042408191 1.0000
## 25-8    -0.12416195 -0.2082107 -0.040113191 < 2e-16 ***
## 40-8    -0.15752675 -0.2422081 -0.072845374 < 2e-16 ***
## 25-12   -0.08170627 -0.1659827  0.002570201 0.1389
## 40-12   -0.11507107 -0.1999785 -0.030163679 2.5e-15 ***
## 40-25   -0.03336480 -0.1174575  0.050727909 1.0000
```

```
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2. Logistic Regression Model

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
logistic_fit1 = glmer(Accuracy~StrengthLevel+ResponseTime+(1|ParticipantNum), data = Behave, family = binomial)
summary(logistic_fit1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial   ( logit )
## Formula: Accuracy ~ StrengthLevel + ResponseTime + (1 | ParticipantNum)
##   Data: Behave
##
##           AIC          BIC    logLik deviance df.resid
## 18768.7 18832.1 -9376.3 18752.7    20439
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -8.2560  0.1420  0.2696  0.6002  2.2668
##
## Random effects:
##   Groups                Name                Variance Std.Dev.
## ParticipantNum (Intercept) 0.1201    0.3466
## Number of obs: 20447, groups: ParticipantNum, 25
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.09428    0.09994  10.950 < 2e-16 ***
## StrengthLevel4.5  0.41038    0.05054   8.119 4.69e-16 ***
## StrengthLevel8    0.78687    0.05254  14.975 < 2e-16 ***
## StrengthLevel12   1.21858    0.05649  21.572 < 2e-16 ***
## StrengthLevel25   2.41070    0.07746  31.122 < 2e-16 ***
## StrengthLevel40   3.15879    0.10408  30.348 < 2e-16 ***
## ResponseTime     -0.97004    0.06346 -15.287 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) StL4.5 StrnL8 StrL12 StrL25 StrL40
## StrngthL4.5 -0.249
## StrngthLv18 -0.242  0.471
## StrngthLv12 -0.247  0.439  0.424
## StrngthLv25 -0.211  0.321  0.311  0.293
## StrngthLv40 -0.166  0.239  0.232  0.219  0.165
## ResponseTim -0.627  0.003  0.007  0.041  0.078  0.073
```

```
logistic_fit2 = glmer(Accuracy~StrengthLevel+(1|ParticipantNum), data = Behave, family = binomial)
summary(logistic_fit2)
```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Accuracy ~ StrengthLevel + (1 | ParticipantNum)
## Data: Behave
##
##      AIC      BIC    logLik deviance df.resid
## 19000.8 19056.3 -9493.4 18986.8    20440
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.4729  0.1535  0.2689  0.6013  1.5089
##
## Random effects:
## Groups          Name          Variance Std.Dev.
## ParticipantNum (Intercept) 0.1026  0.3203
## Number of obs: 20447, groups: ParticipantNum, 25
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.13669    0.07306   1.871  0.0613 .
## StrengthLevel4.5 0.42021    0.05014   8.381 <2e-16 ***
## StrengthLevel8   0.80604    0.05210  15.470 <2e-16 ***
## StrengthLevel12  1.27367    0.05600  22.744 <2e-16 ***
## StrengthLevel25  2.53187    0.07681  32.964 <2e-16 ***
## StrengthLevel40  3.30620    0.10344  31.964 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) StL4.5 StrnL8 StrL12 StrL25
## StrngthL4.5 -0.336
## StrngthLv18 -0.323  0.472
## StrngthLv12 -0.301  0.439  0.424
## StrngthLv25 -0.220  0.320  0.310  0.289
## StrngthLv40 -0.163  0.238  0.231  0.216  0.159
logistic_fit3 = glmer(Accuracy~ResponseTime+(1|ParticipantNum), data = Behave, family = binomial)
summary(logistic_fit3)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Accuracy ~ ResponseTime + (1 | ParticipantNum)
## Data: Behave
##
##      AIC      BIC    logLik deviance df.resid
## 21340.5 21364.3 -10667.2 21334.5    20444
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.6528  0.3554  0.4632  0.5551  1.6600
##
## Random effects:
## Groups          Name          Variance Std.Dev.

```

```
## ParticipantNum (Intercept) 0.1192 0.3453
## Number of obs: 20447, groups: ParticipantNum, 25
##
## Fixed effects:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.63998 0.09080 29.07 <2e-16 ***
## ResponseTime -1.54439 0.05876 -26.28 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## ResponseTim -0.621
```

```
library(stargazer)
```

output the table

```
##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
stargazer(logistic_fit1, logistic_fit2, logistic_fit3, title='Logistic Regression Model', header = FALSE)

##
## \begin{table}[!htbp] \centering
##   \caption{Logistic Regression Model}
##   \label{tab:02005}
## \begin{tabular}{@{\extracolsep{5pt}}lccc}
## \hline
## \hline \hline
## & \multicolumn{3}{c}{\textit{Dependent variable:}} & \\
## \cline{2-4}
## \hline & \multicolumn{3}{c}{Accuracy} & \\
## \hline & (1) & (2) & (3) & \\
## \hline
## StrengthLevel4.5 & 0.410$^{***}$ & 0.420$^{***}$ & & \\
## & (0.311, 0.509) & (0.322, 0.518) & & \\
## & & & & \\
## StrengthLevel8 & 0.787$^{***}$ & 0.806$^{***}$ & & \\
## & (0.684, 0.890) & (0.704, 0.908) & & \\
## & & & & \\
## StrengthLevel12 & 1.219$^{***}$ & 1.274$^{***}$ & & \\
## & (1.108, 1.329) & (1.164, 1.383) & & \\
## & & & & \\
## StrengthLevel25 & 2.411$^{***}$ & 2.532$^{***}$ & & \\
## & (2.259, 2.563) & (2.381, 2.682) & & \\
## & & & & \\
## StrengthLevel40 & 3.159$^{***}$ & 3.306$^{***}$ & & \\
## & (2.955, 3.363) & (3.103, 3.509) & & \\
## & & & & \\
## ResponseTime & $-0.970$^{***}$ & & $-1.544$^{***}$ & \end{table}
```

```
## & ($-1.094, $-$0.846) & & ($-1.660, $-$1.429) \\
## & & & \\
## Constant & 1.094$^{***}$ & 0.137$^{*}$ & 2.640$^{***}$ \\
## & (0.898, 1.290) & ($-$0.006, 0.280) & (2.462, 2.818) \\
## & & & \\
## \hline \\[-1.8ex]
## Observations & 20,447 & 20,447 & 20,447 \\
## Log Likelihood & $-$9,376.330 & $-$9,493.403 & $-$10,667.240 \\
## Akaike Inf. Crit. & 18,768.660 & 19,000.810 & 21,340.480 \\
## Bayesian Inf. Crit. & 18,832.070 & 19,056.280 & 21,364.260 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:} & \multicolumn{3}{r}{${}^{*}$p$<$0.1; ${}^{**}$p$<$0.05; ${}^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
library(lmtest)
```

Likelihood Ratio Tests (Goodness-of-fit)

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
lrtest(logistic_fit2, logistic_fit1)
```

```
## Likelihood ratio test
##
## Model 1: Accuracy ~ StrengthLevel + (1 | ParticipantNum)
## Model 2: Accuracy ~ StrengthLevel + ResponseTime + (1 | ParticipantNum)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    7 -9493.4
## 2    8 -9376.3  1 234.15 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(logistic_fit3, logistic_fit1)
```

```
## Likelihood ratio test
##
## Model 1: Accuracy ~ ResponseTime + (1 | ParticipantNum)
## Model 2: Accuracy ~ StrengthLevel + ResponseTime + (1 | ParticipantNum)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -10667.2
## 2    8 -9376.3  5 2581.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(dplyr)
```

train test split

```
##
## 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

set.seed(1234)
train = Behave %>% group_by(ParticipantNum, StrengthLevel) %>% sample_n(80)
test = dplyr::anti_join(Behave, train)

## Joining, by = c("ParticipantNum", "StrengthLevel", "ResponseTime", "Accuracy", "Confidence")
```

```
library(ROCR)
logistic_fit4 = glmer(Accuracy~StrengthLevel+ResponseTime+(1|ParticipantNum), data = train, family = binomial)
# Compute AUC for predicting Class with the model
prob <- predict(logistic_fit4, newdata=test, type="response")
pred <- prediction(prob, test$Accuracy)
Behave_fit_pred = rep(0, dim(test)[1])
Behave_fit_pred[prob > 0.5] = 1

test_error = mean(Behave_fit_pred != test$Accuracy)
test_error
```

test error

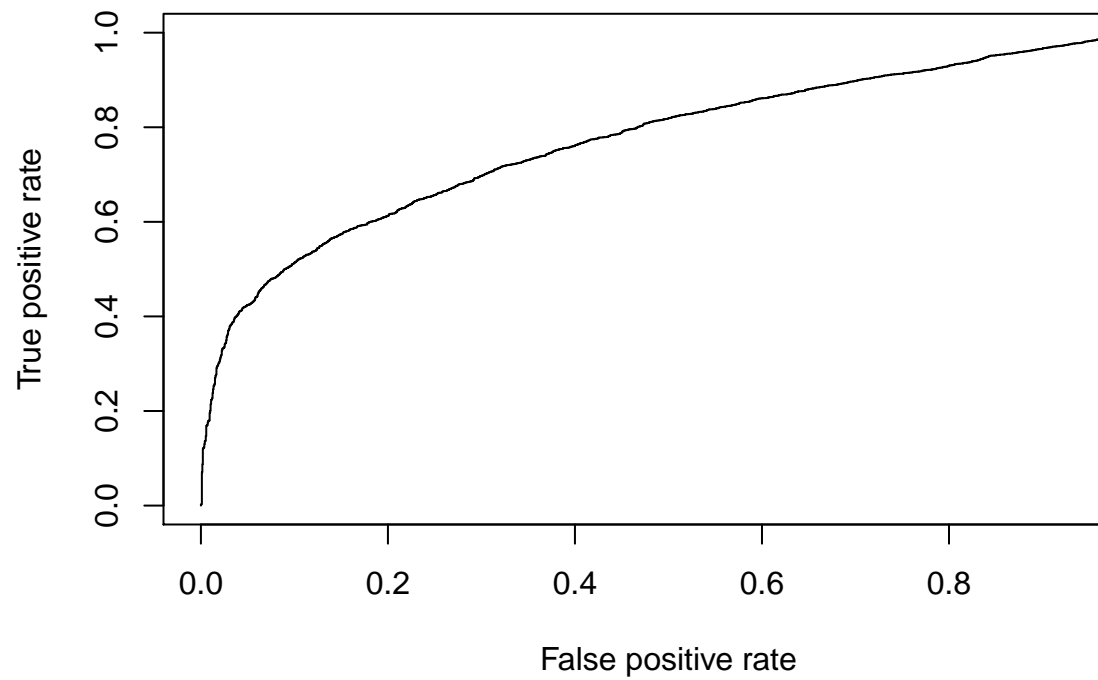
```
## [1] 0.2310243
```

```
table(Behave_fit_pred, test$Accuracy)
```

Confusion Matrix

```
##
## Behave_fit_pred    0    1
##                   0 281 304
##                   1 1647 6213
```

```
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf)
```

ROC curve and AUROC

```
auc <- performance(pred, measure = "auc")
auc <- auc@y.values[[1]]
auc
```

```
## [1] 0.7703337
```