IST772 Practice Quiz #4

John Fields

12 Mar 2020

Week 9: Practice Exam #4

Instructions: This is the second of three practice exams leading up to the final exam. Complete the analysis specified below by writing and running the appropriate R-code.

1. Run ?USJudgeRatings and examine the help file for that built-in data frame. Cut and paste the most relevant information from the help screen below.

This dataset contains "Lawyers' Ratings of State Judges in the US Superior Court"

Format A data frame containing 43 observations on 12 numeric variables.

[,1] CONT Number of contacts of lawyer with judge. [,2] INTG Judicial integrity. [,3] DMNR Demeanor. [,4] DILG Diligence. [,5] CFMG Case flow managing. [,6] DECI Prompt decisions. [,7] PREP Preparation for trial. [,8] FAMI Familiarity with law. [,9] ORAL Sound oral rulings. [,10] WRIT Sound written rulings. [,11] PHYS Physical ability. [,12] RTEN Worthy of retention.

Source New Haven Register, 14 January, 1977

```r
?USJudgeRatings
```

2. Copy the built-in data frame USJudgeRatings into a new data frame. Call your new data frame "newJudge".

```r
newJudge <- data.frame(USJudgeRatings)
#View(newJudge)
```

3. Create a binary version of the RTEN variable in newJudge with the following code:

```r
newJudge$retain <- as.integer(newJudge$RTEN > median(newJudge$RTEN))
```

4. Run summary() on retain, CONT, and INTG from your newJudge data frame. Paste in the results below.

```r
summary(newJudge$retain)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4884  1.0000  1.0000
```

```r
summary(newJudge$CONT)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.700   6.850   7.300   7.437   7.900  10.600
```

```r
summary(newJudge$INTG)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.900   7.550   8.100   8.021   8.550   9.200
```

5. Run a logistic regression model using CONT and INTG to predict retain. Report the results below.

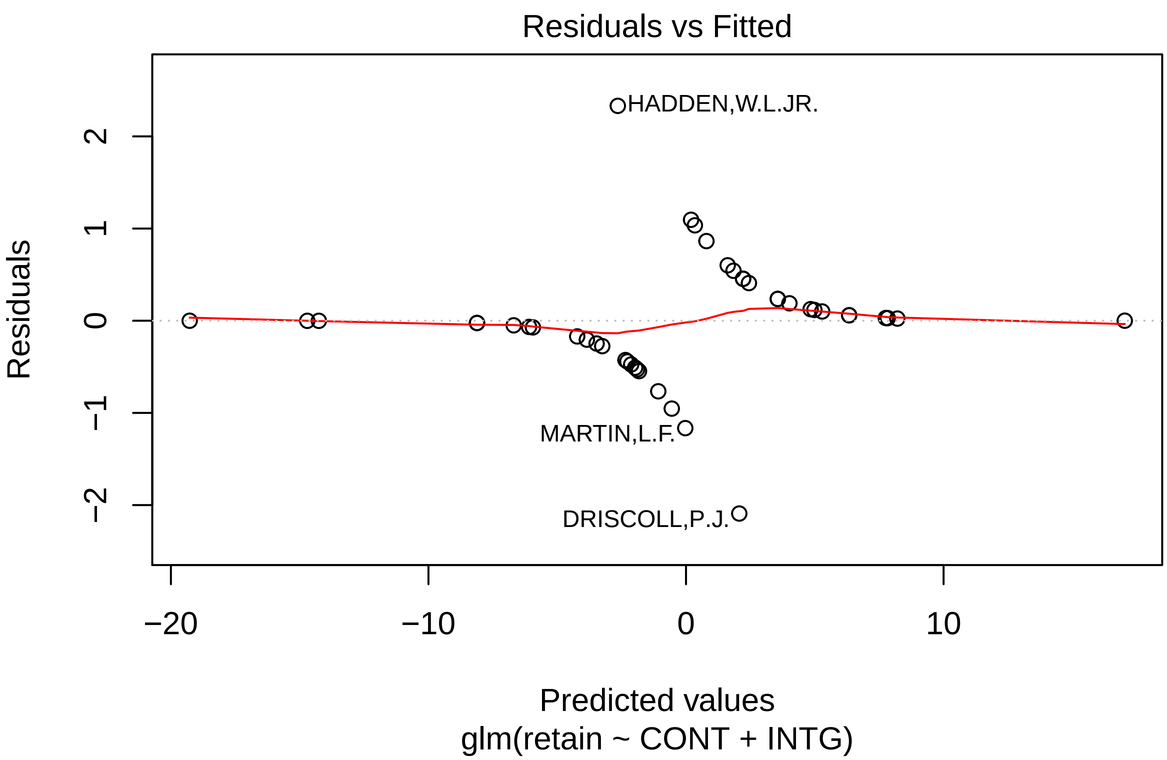
Please see Question 7 for a description of the results.

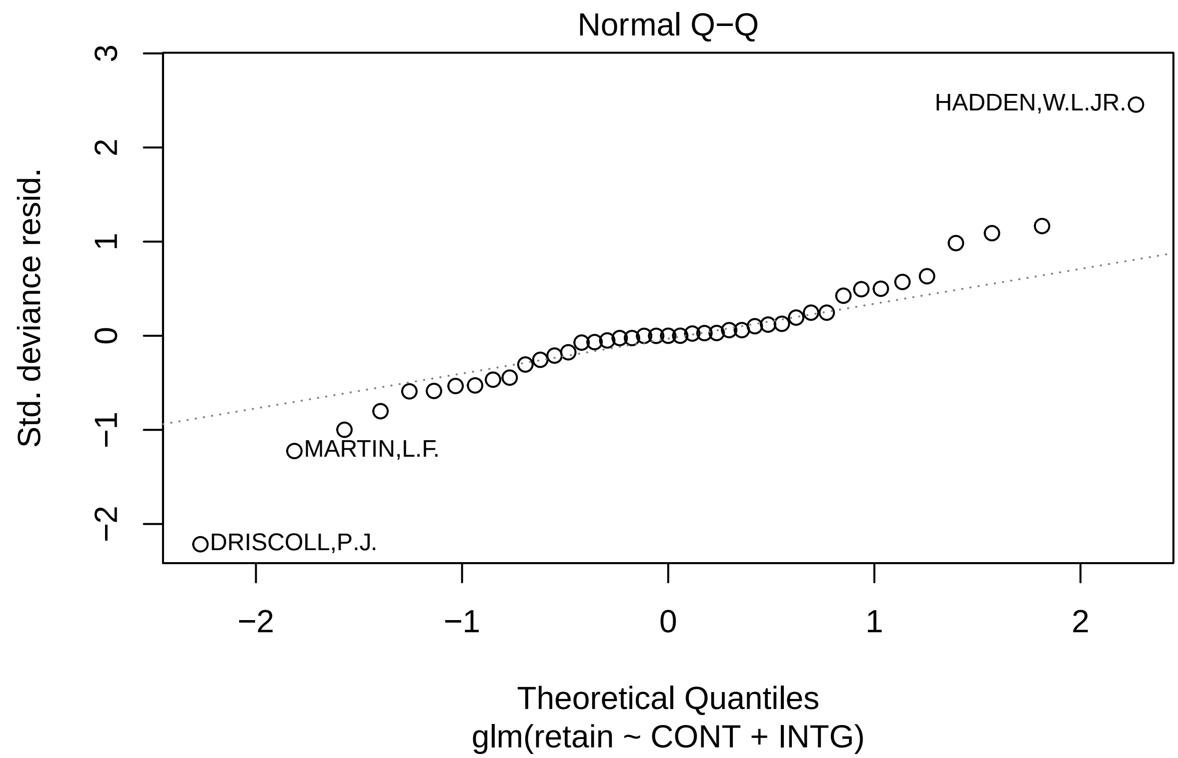
```r
glmOut <- glm(formula = retain ~ CONT + INTG, family = binomial(link="logit"), data = newJudge)
glmOut
```

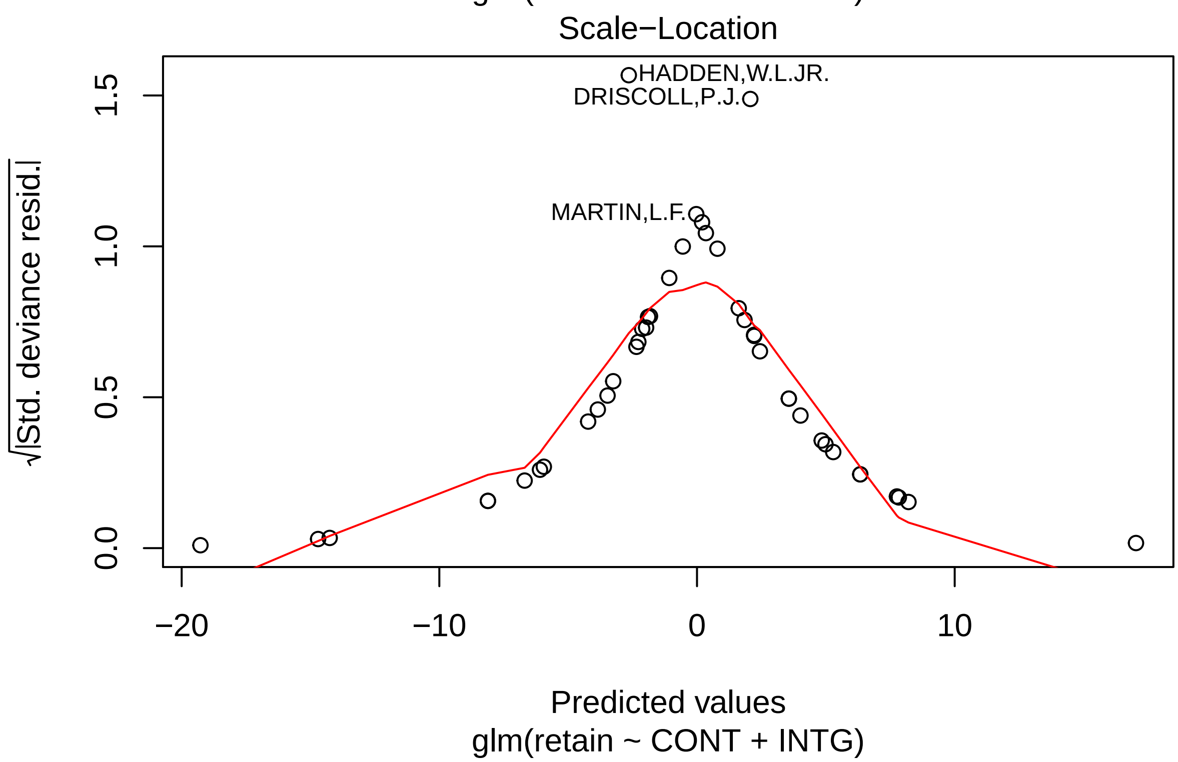
```
##
## Call:  glm(formula = retain ~ CONT + INTG, family = binomial(link = "logit"),
##     data = newJudge)
##
## Coefficients:
## (Intercept)         CONT         INTG
##     -88.816        2.993        8.236
##
## Degrees of Freedom: 42 Total (i.e. Null);  40 Residual
## Null Deviance:       59.59
## Residual Deviance: 18.77     AIC: 24.77
```

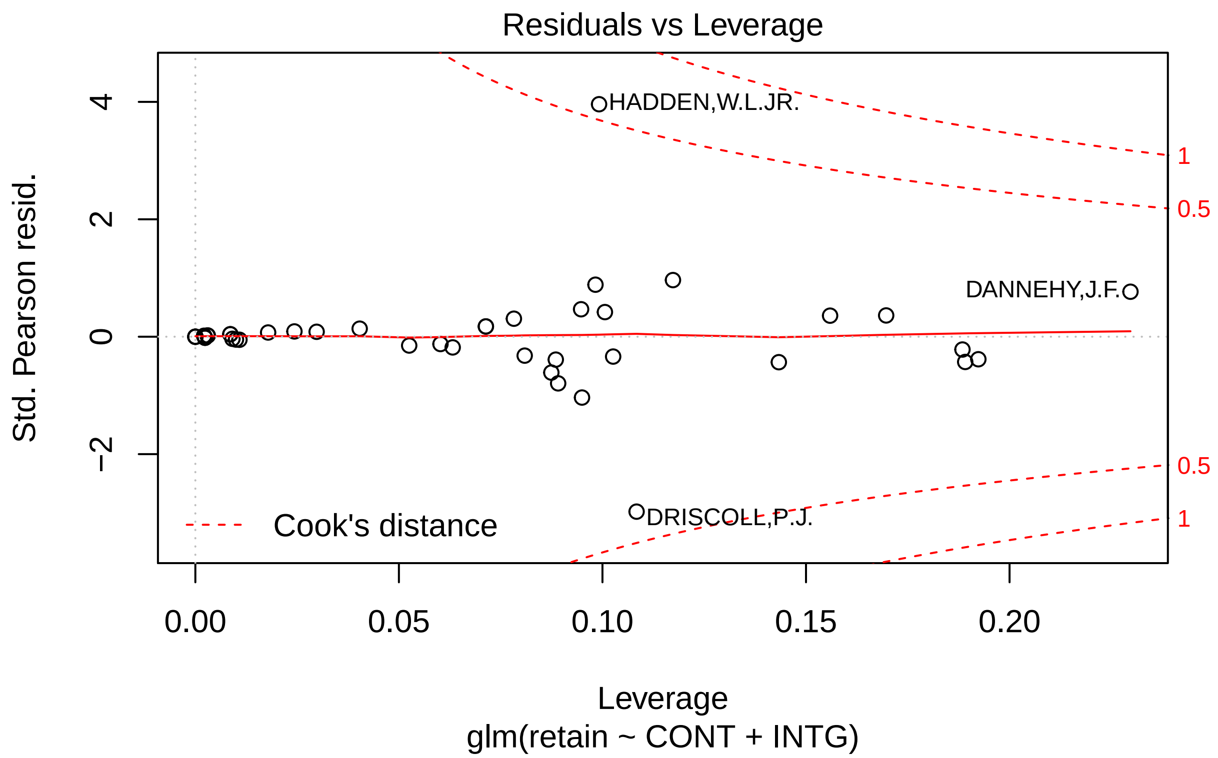
6. Paste in all of your R code below.

```r
plot(glmOut)
```



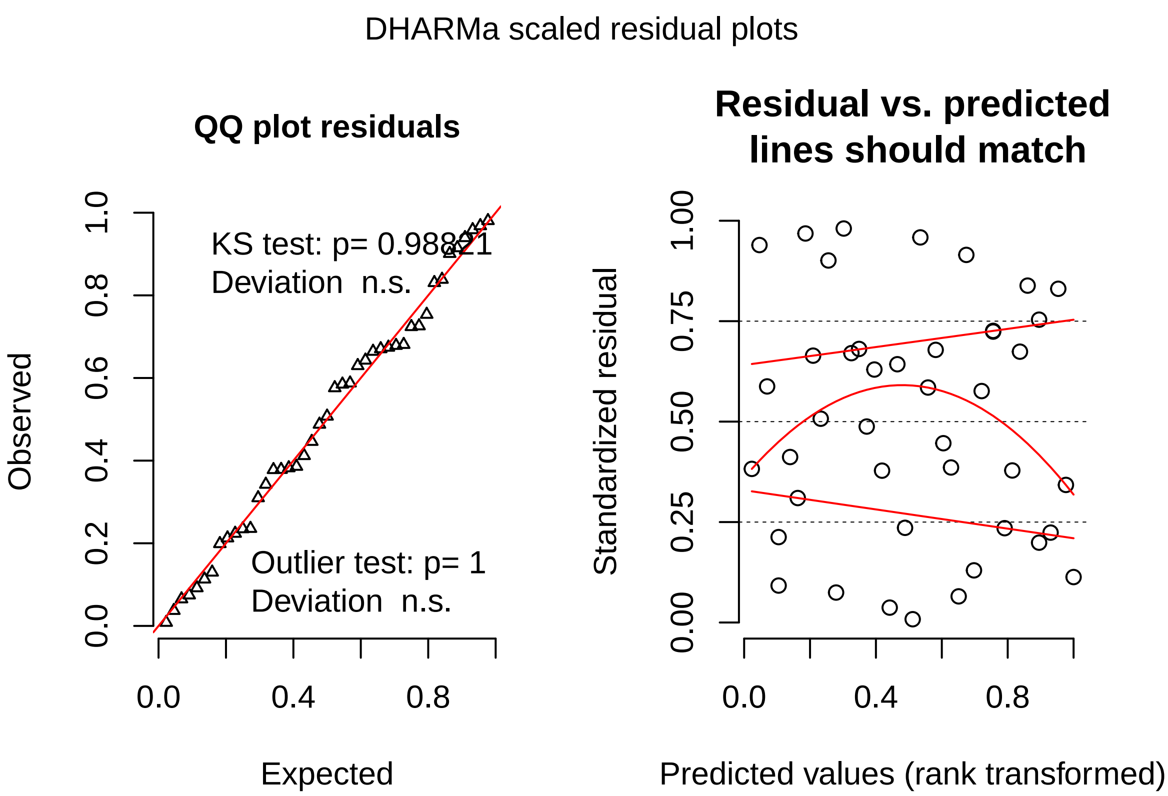

Normal Q−Q
Std. deviance resid.
3
2
1
0
−1
−2
HADDEN,W.L.JR.
MARTIN,L.F.
DRISCOLL,P.J.
−2
−1
0
1
2
Theoretical Quantiles
glm(retain ~ CONT + INTG)


Scale−Location
√|Std. deviance resid.|
1.5
1.0
0.5
0.0
HADDEN,W.L.JR.
DRISCOLL,P.J.
MARTIN,L.F.
−20
−10
0
10
Predicted values
glm(retain ~ CONT + INTG)

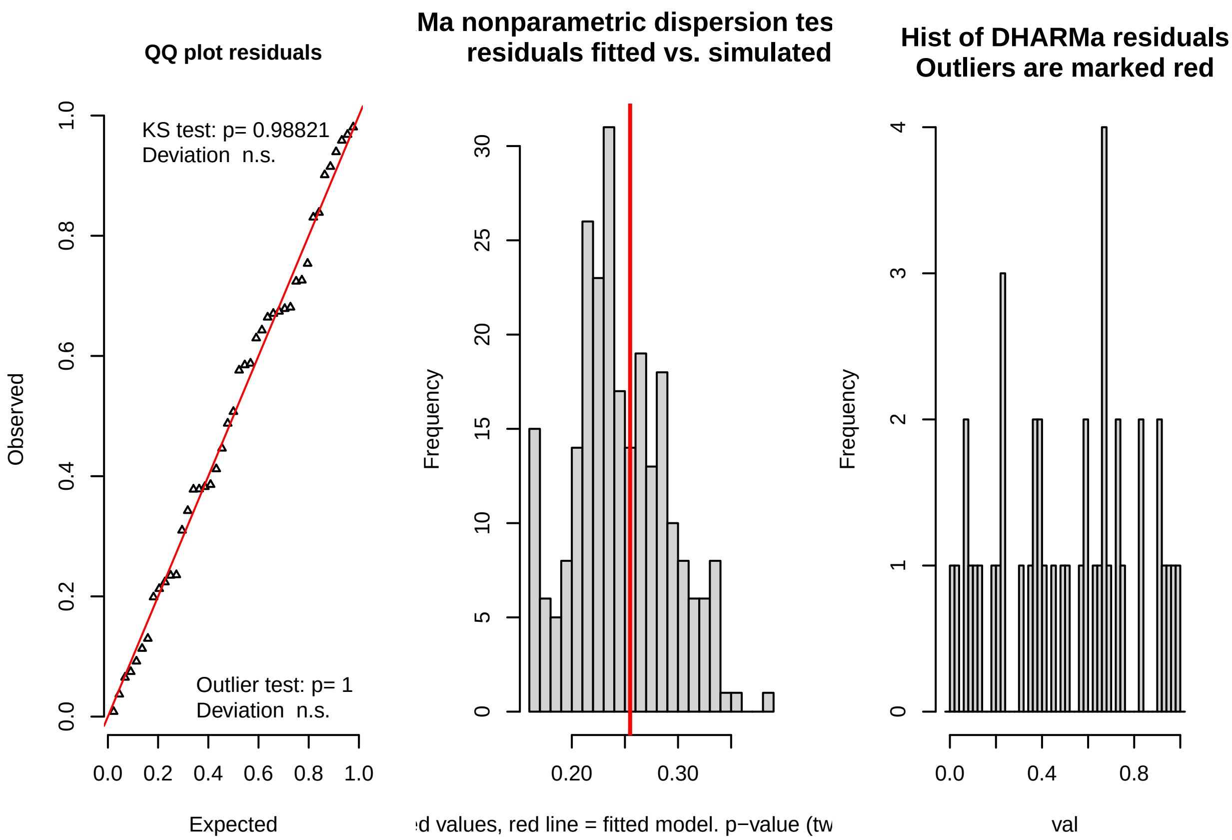


TAKE OUT HADDEEN AND DRISCOLL?

```
library(DHARMa)
simulationOutput <- simulateResiduals(fittedModel = glmOut, n = 250)
plot(simulationOutput)
```



```r
testResiduals(simulationOutput)
```




```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.064853, p-value = 0.9882
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 1.0372, p-value = 0.8
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa outlier test based on exact binomial test
##
## data:  simulationOutput
## outLow = 0.0000000, outHigh = 0.0000000, nobs = 43.0000000, freqH0 =
```

```
## 0.0039841, p-value = 1
## alternative hypothesis: two.sided
```

```
## $uniformity
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  simulationOutput$scaledResiduals
## D = 0.064853, p-value = 0.9882
## alternative hypothesis: two-sided
##
##
## $dispersion
##
##  DHARMa nonparametric dispersion test via sd of residuals fitted vs.
##  simulated
##
## data:  simulationOutput
## ratioObsSim = 1.0372, p-value = 0.8
## alternative hypothesis: two.sided
##
##
## $outliers
##
##  DHARMa outlier test based on exact binomial test
##
## data:  simulationOutput
## outLow = 0.0000000, outHigh = 0.0000000, nobs = 43.0000000, freqH0 =
## 0.0039841, p-value = 1
## alternative hypothesis: two.sided
```

```r
summary(glmOut)
```

```
##
## Call:
## glm(formula = retain ~ CONT + INTG, family = binomial(link = "logit"),
##     data = newJudge)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.09169  -0.26140  -0.00009   0.21278   2.33087
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -88.816     30.359  -2.926  0.00344 **
## CONT           2.993      1.347   2.223  0.02625 *
## INTG           8.236      2.728   3.019  0.00254 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 59.587  on 42  degrees of freedom
## Residual deviance: 18.772  on 40  degrees of freedom
## AIC: 24.772
```

```
##
## Number of Fisher Scoring iterations: 7
```

```r
anova(glmOut, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: retain
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                    42     59.587
## CONT  1    0.636        41     58.952    0.4253
## INTG  1   40.179        40     18.772 2.317e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
exp(coef(glmOut)) #convert log odds to regular oddds
```

```
##  (Intercept)         CONT         INTG
## 2.678132e-39 1.994735e+01 3.774832e+03
```

```r
exp(confint(glmOut)) # Look at confidence intervals
```

```
## Waiting for profiling to be done...

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```
##                      2.5 %       97.5 %
## (Intercept) 1.767315e-73 1.942106e-19
## CONT        2.529851e+00 6.914718e+02
## INTG        6.115934e+01 4.376847e+06
```

```r
library(BaylorEdPsych)
PseudoR2(glmOut)
```

```
##         McFadden     Adj.McFadden        Cox.Snell       Nagelkerke
##        0.6849624        0.5507059        0.6129452        0.8174077
## McKelvey.Zavoina           Effron            Count        Adj.Count
##        0.9279901        0.7460153        0.9534884        0.9047619
##              AIC    Corrected.AIC
##       24.7722685       25.3876531
```

```r
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
predictedRetain<-round(predict(glmOut,type="response"))

sum(predictedRetain) # number we predict to be retained
```

```
## [1] 21
```

```r
sum(newJudge$retain) # number actually retained
```

```
## [1] 21
```

```r
confusion <-table(predictedRetain, newJudge$retain)
addmargins(confusion)
```

```
##
## predictedRetain  0  1 Sum
##             0   21  1  22
##             1    1 20  21
##             Sum 22 21  43
```

```r
confusionMatrix(confusion, positive="1")
```

```
## Confusion Matrix and Statistics
##
##
## predictedRetain  0  1
##               0 21  1
##               1  1 20
##
##                Accuracy : 0.9535
##                  95% CI : (0.8419, 0.9943)
##     No Information Rate : 0.5116
##     P-Value [Acc > NIR] : 2.642e-10
##
##                   Kappa : 0.9069
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9524
```

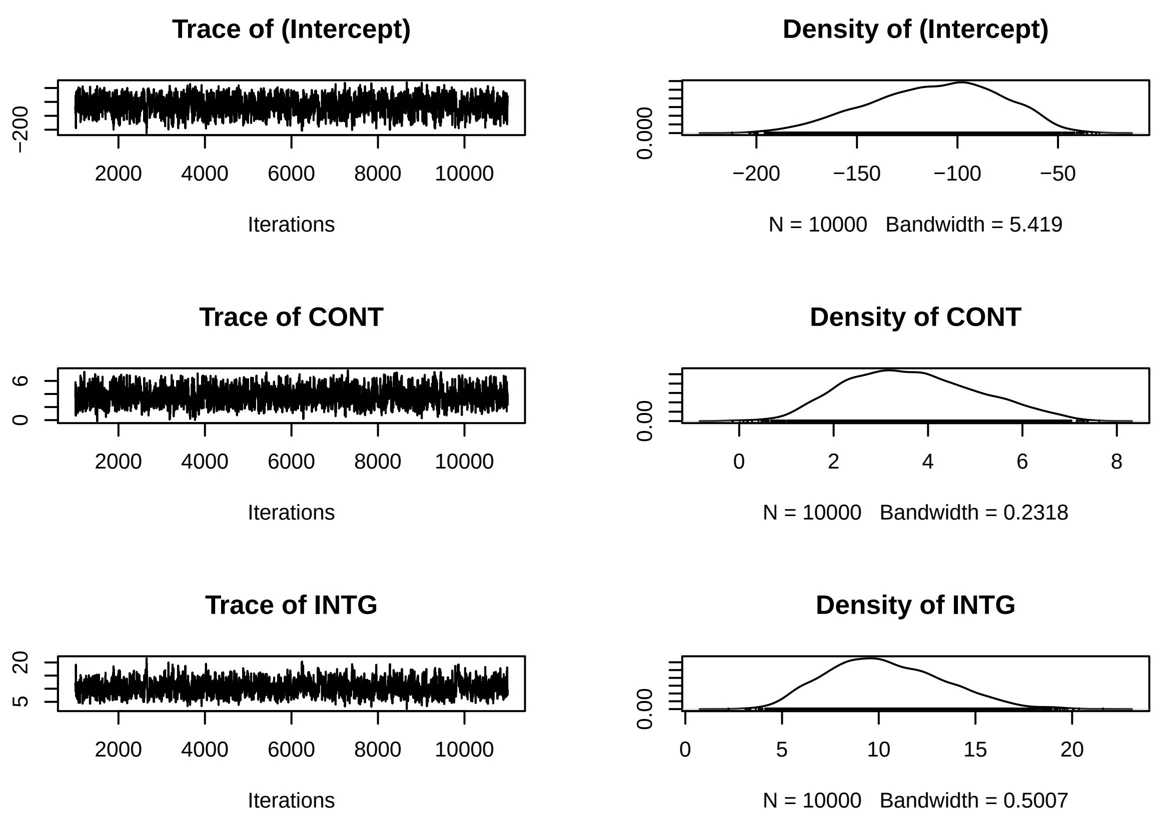
```
##              Specificity : 0.9545
##           Pos Pred Value : 0.9524
##           Neg Pred Value : 0.9545
##               Prevalence : 0.4884
##           Detection Rate : 0.4651
##     Detection Prevalence : 0.4884
##        Balanced Accuracy : 0.9535
##
##         'Positive' Class : 1
##
```

```r
library(MCMCpack)
```

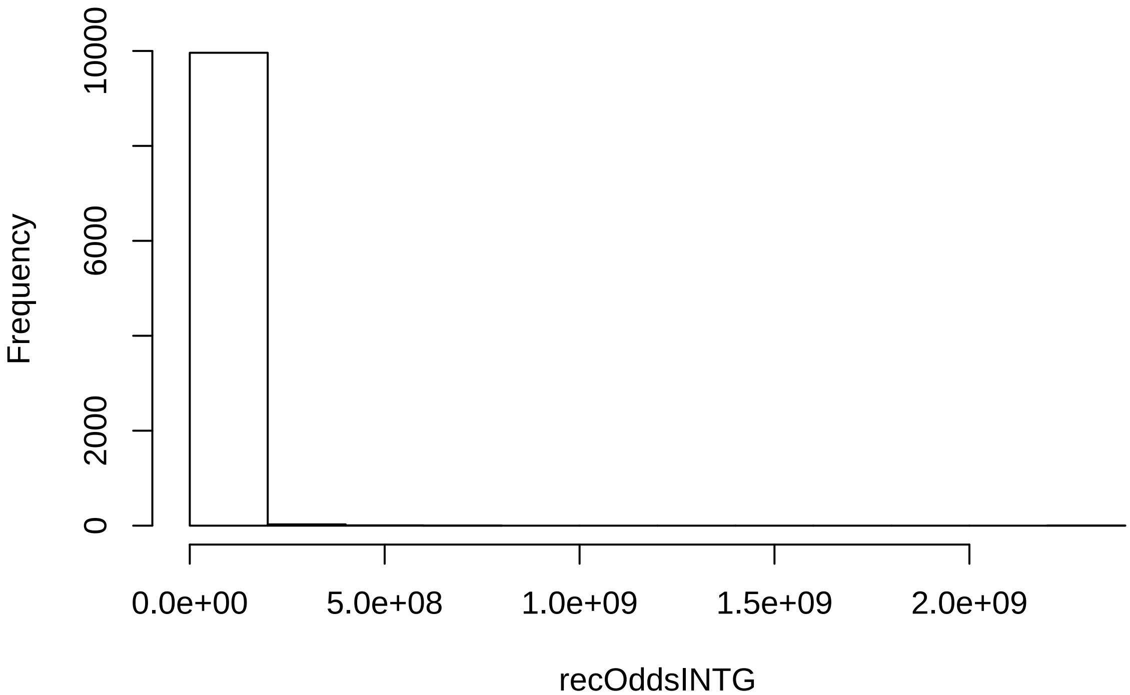
```
## Loading required package: coda

## Loading required package: MASS

## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2020 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park

## ##
## ## Support provided by the U.S. National Science Foundation

## ## (Grants SES-0350646 and SES-0350613)
## ##
```

```r
bayesLogitOut <- MCMClogit(formula = retain ~ CONT + INTG, data = newJudge)
```

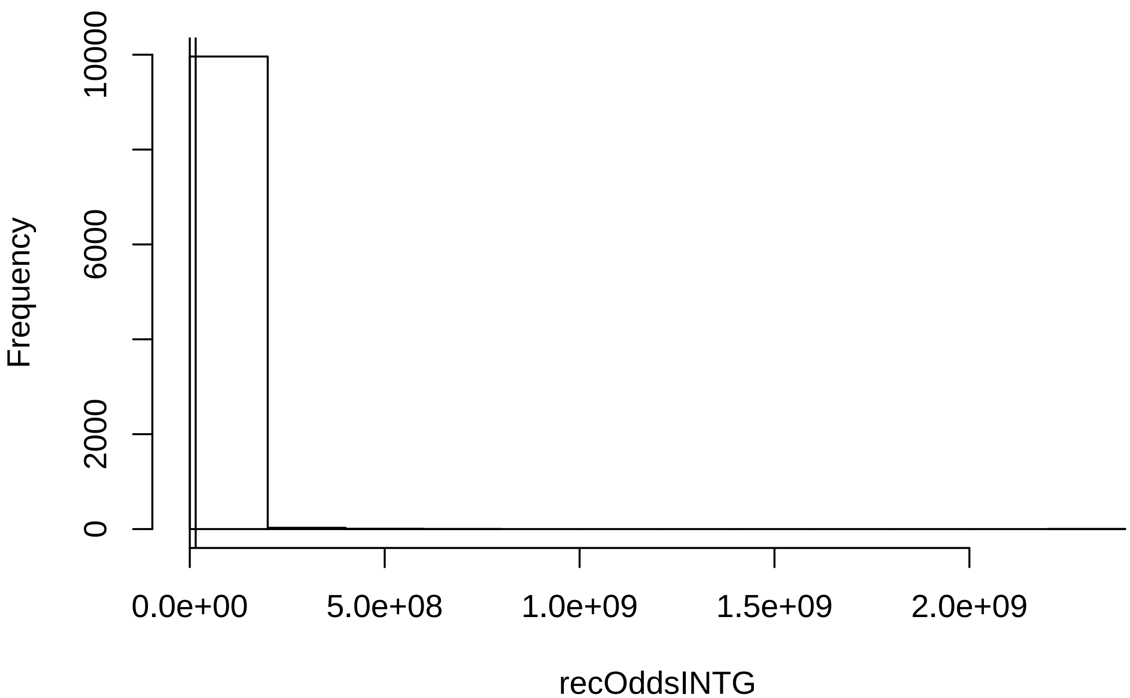
```r
plot(bayesLogitOut)
```



```r
recLogOddsINTG <- as.matrix(bayesLogitOut[,"INTG"])
recOddsINTG <- exp(recLogOddsINTG)
hist(recOddsINTG, main=NULL)
```




```r
hist(recOddsINTG, main=NULL)
abline(v=quantile(recOddsINTG,c(0.025)),col="black")
abline(v=quantile(recOddsINTG,c(0.975)),col="black")
```




```r
quantile(recOddsINTG,c(0.025))
```

```
##     2.5%
## 213.5759
```

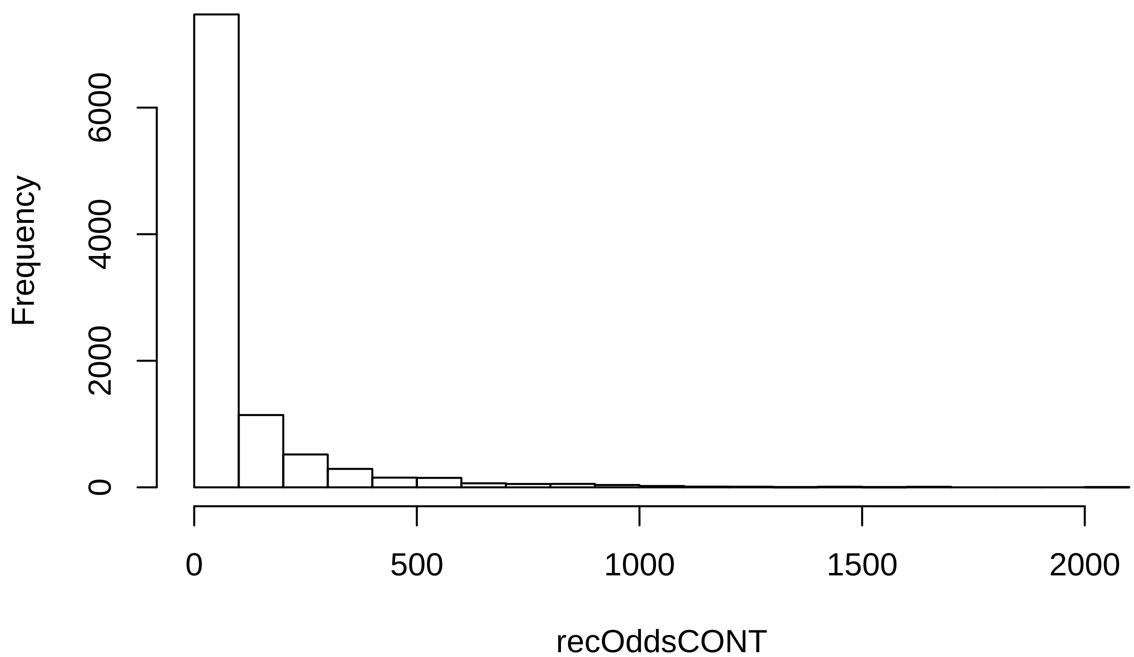
```r
quantile(recOddsINTG,c(0.975))
```

```
##    97.5%
## 15161626
```

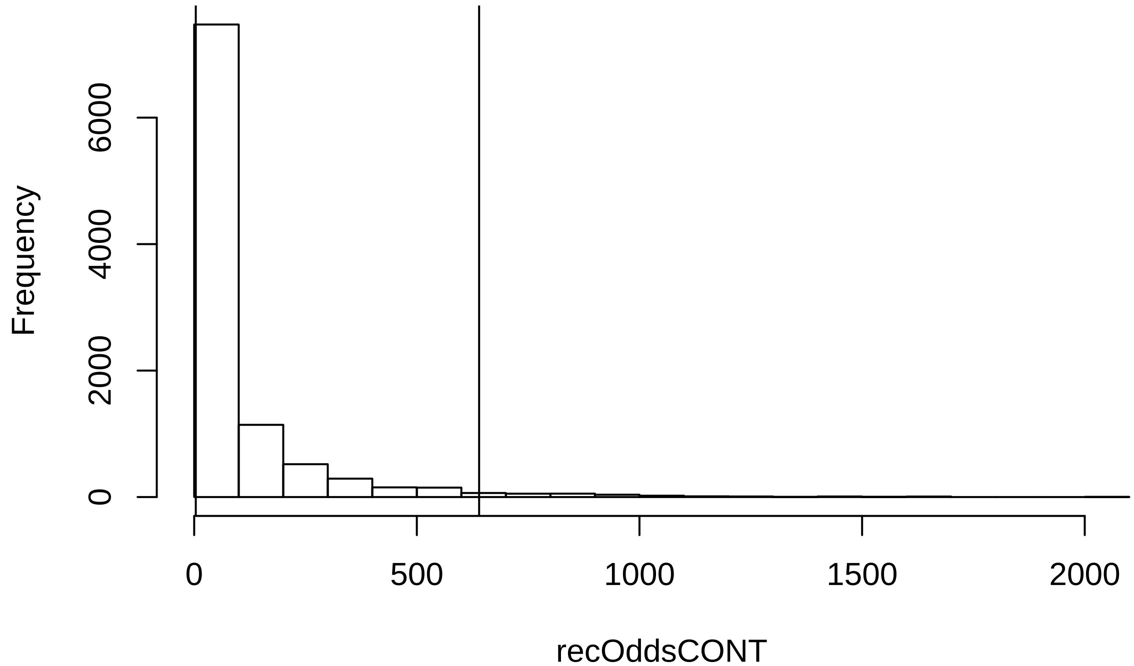
```r
quantile(recOddsINTG,c(0.5))
```

```
##      50%
## 23724.55
```

```r
recLogOddsCONT <- as.matrix(bayesLogitOut[,"CONT"])
recOddsCONT <- exp(recLogOddsCONT)
hist(recOddsCONT, main=NULL)
```




```r
hist(recOddsCONT, main=NULL)
abline(v=quantile(recOddsCONT,c(0.025)),col="black")
abline(v=quantile(recOddsCONT,c(0.975)),col="black")
```




```r
quantile(recOddsCONT,c(0.025))
```

```
##     2.5%
## 3.618981
```

```r
quantile(recOddsCONT,c(0.975))
```

```
##    97.5%
## 639.9281
```

```r
quantile(recOddsCONT,c(0.5))
```

```
##      50%
## 35.28128
```

7. Provide a write up of your results, with an interpretation of the research question, "Can the number of contacts between a lawyer and a judge along with the lawyer's rating of judicial integrity predict retention recommendations about the judge?"

The effects of judicial integrity (INTG) and contacts (CONT) were analyzed to determine if these variable would predict retention recommendations of state judges in the US Superior Court. A chi-square omnibus test on the results of a logistic regression was significant for INTG, chisq(1) = 40.18, $p<0.001$. The chi-square test for CONT was not significant, chisq(1) = 0.64, p=0.43. A Wald's z-test on the INTG coefficient was significant, z = 3.02, $p<0.01$ and the CONT coefficient was also significant, z = 2.22, $p<0.05$. The deviance residuals are near zero at -0.00009 which indicates that there is very little skew in the residuals of the distribution. We can reject the null hypothesis that the log-odds of income and statusquo is equal to 0 in the population.

The regular odds of INTG were 3775:1 which indicates that for every additional unit of INT, judges are 3775 times more likely to be have higher retention ratings. For CONT, the log odds of 20:1 shows that for every additional unit of CONT, judges are 20 times more likely to have more contacts with the lawyers.

A review of the confidence intervals shows that INTG and CONT do not cross zero. This supports the significance of INTG and CONT for these coefficients.

The last step in the traditional analysis was to evaluate the accuracy of a prediction model for predicting retention from INTG and CONT. The confusion matrix shows the model was 95% accurate with 1 false positive and 1 false negative prediction.

A Bayesian logistic regression was also performed and the resulting MCMC plot is shown above. The trace of the log odds for INTG and CONT did not show any anomolies. However, the CONT variable does cross zero with one of the traces. The range of the 95% HDI for INTG converted to plain odds is 213 to 15,161,626. The mean of the INTG HDI is 23,725. The range of the 95% HDI plain odds for CONT is 3.6 to 639.9. The mean of the CONT HDI is 35.3 The plain odds results from the 95% HDI are higher than the traditional analysis but both provide evidence for the strength of this statistical model for predicting retention.

The overall result is that the combination of judicial integrity (INTG) and contacts (CONT) are good predictors of a high retention recommendation by lawyers for US Superior Court Judges in this data.