STAT UN2104 – Take Home Quiz

Steven Francis

May 2, 2018

```
job_sat_data <- read.table("jobsat-data.txt", header = T)
```

```
#Changing numerical variables region and age(groups) into factor variables
job_sat_data$region <- factor(job_sat_data$region)
job_sat_data$age <- factor(job_sat_data$age)
```

```
job_sat_data.logit <- glm(cbind(sat,unsat) ~ region + race + age + gender +
                            race*gender,
                          data = job_sat_data, family = binomial)

summary(job_sat_data.logit)
```

```
##
## Call:
## glm(formula = cbind(sat, unsat) ~ region + race + age + gender +
##     race * gender, family = binomial, data = job_sat_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4630  -0.5681   0.1558   0.7102   2.4220
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.005388   0.091138  -0.059  0.95286
## region2        0.436366   0.103868   4.201 2.66e-05 ***
## region3        0.175139   0.073084   2.396  0.01656 *
## region4        0.086916   0.066481   1.307  0.19108
## region5        0.123197   0.067601   1.822  0.06839 .
## region6        0.288263   0.073567   3.918 8.91e-05 ***
## region7        0.411842   0.096729   4.258 2.07e-05 ***
## racew          0.213454   0.090890   2.348  0.01885 *
## age2           0.127823   0.051895   2.463  0.01377 *
## age3           0.363392   0.051602   7.042 1.89e-12 ***
## genderm        0.490254   0.111201   4.409 1.04e-05 ***
## racew:genderm -0.380160   0.123149  -3.087  0.00202 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 204.141  on 83  degrees of freedom
## Residual deviance:  87.941  on 72  degrees of freedom
## AIC: 476.95
##
## Number of Fisher Scoring iterations: 4
```

```
anova(job_sat_data.logit)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(sat, unsat)
##
## Terms added sequentially (first to last)
##
##
##             Df Deviance Resid. Df Resid. Dev
## NULL                         83    204.141
## region       6   34.558        77    169.583
## race         1    3.547        76    166.036
## age          2   54.202        74    111.834
## gender       1   14.317        73     97.516
## race:gender  1    9.575        72     87.941
```

To test the model's goodness of fit

```
pchisq(q=87.941, df=72, lower.tail = F)
```

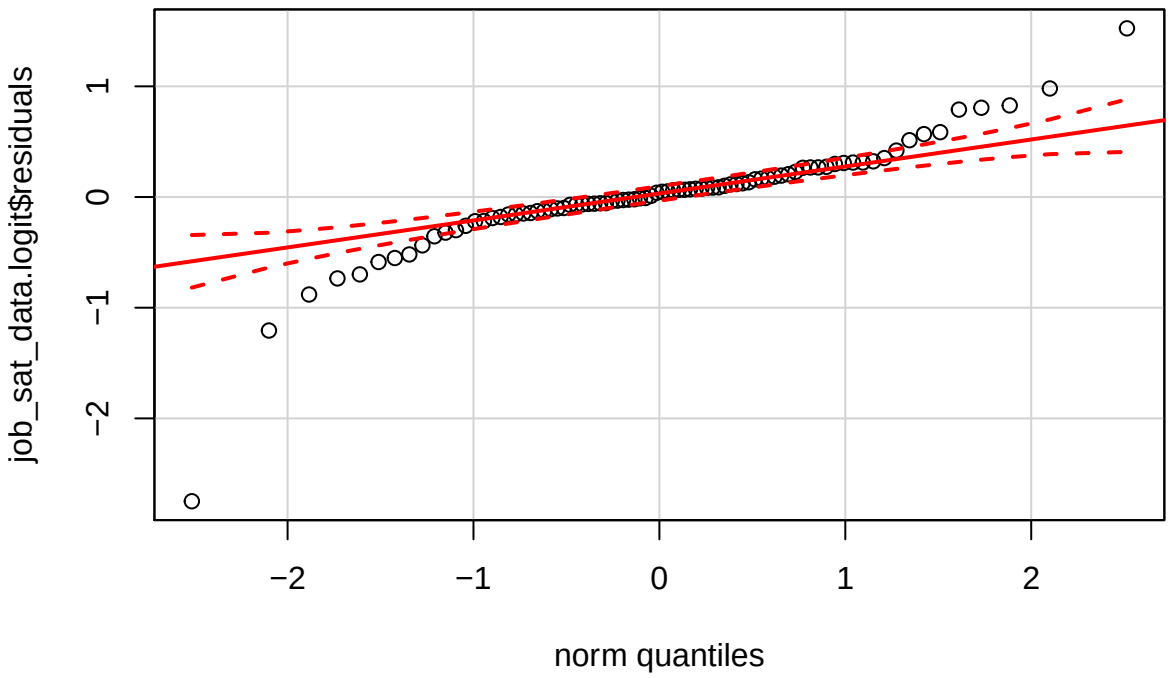
```
## [1] 0.09748707
```

```
#Proving that the residuals follow a relatively follow a standard normal
#distribution

library('car')
```

```
## Warning: package 'car' was built under R version 3.4.1
```

```
qqPlot(job_sat_data.logit$residuals)
```



Adding observed satisfaction probability percentage, fitted satisfaction probability percentage and fitted satisfaction odds percentage

```
job_sat_data.obs.prop <- job_sat_data$sat/(job_sat_data$sat + job_sat_data$unsat)
cbind(job_sat_data, job_sat_data.obs.prop, fitted(job_sat_data.logit))
```

```
##    region race age gender sat unsat job_sat_data.obs.prop
## 1       1    w   1      m 288   177             0.6193548
## 2       1    w   1      f  60    57             0.5128205
## 3       1    w   2      m 224   166             0.5743590
## 4       1    w   2      f  35    19             0.6481481
## 5       1    w   3      m 337   172             0.6620825
## 6       1    w   3      f  70    30             0.7000000
## 7       1    o   1      m  38    33             0.5352113
## 8       1    o   1      f  19    35             0.3518519
## 9       1    o   2      m  32    11             0.7441860
## 10      1    o   2      f  22    20             0.5238095
## 11      1    o   3      m  21     8             0.7241379
## 12      1    o   3      f  15    10             0.6000000
## 13      2    w   1      m  90    45             0.6666667
## 14      2    w   1      f  19    12             0.6129032
## 15      2    w   2      m  96    42             0.6956522
## 16      2    w   2      f  12     5             0.7058824
```

```
## 17      2    w   3      m 124    39           0.7607362
## 18      2    w   3      f  17     2           0.8947368
## 19      2    o   1      m  18     6           0.7500000
## 20      2    o   1      f  13     7           0.6500000
## 21      2    o   2      m   7     2           0.7777778
## 22      2    o   2      f   0     3           0.0000000
## 23      2    o   3      m   9     2           0.8181818
## 24      2    o   3      f   1     1           0.5000000
## 25      3    w   1      m 226   128           0.6384181
## 26      3    w   1      f  88    57           0.6068966
## 27      3    w   2      m 189   117           0.6176471
## 28      3    w   2      f  44    34           0.5641026
## 29      3    w   3      m 156    73           0.6812227
## 30      3    w   3      f  70    25           0.7368421
## 31      3    o   1      m  45    31           0.5921053
## 32      3    o   1      f  47    35           0.5731707
## 33      3    o   2      m  18     3           0.8571429
## 34      3    o   2      f  13     7           0.6500000
## 35      3    o   3      m  11     2           0.8461538
## 36      3    o   3      f   9     2           0.8181818
## 37      4    w   1      m 285   179           0.6142241
## 38      4    w   1      f 110    93           0.5418719
## 39      4    w   2      m 225   141           0.6147541
## 40      4    w   2      f  53    24           0.6883117
## 41      4    w   3      m 324   140           0.6982759
## 42      4    w   3      f  60    47           0.5607477
## 43      4    o   1      m  40    25           0.6153846
## 44      4    o   1      f  66    56           0.5409836
## 45      4    o   2      m  19    11           0.6333333
## 46      4    o   2      f  25    19           0.5681818
## 47      4    o   3      m  22     2           0.9166667
## 48      4    o   3      f  11    12           0.4782609
## 49      5    w   1      m 270   180           0.6000000
## 50      5    w   1      f 176   151           0.5382263
## 51      5    w   2      m 215   108           0.6656347
## 52      5    w   2      f  80    40           0.6666667
## 53      5    w   3      m 269   136           0.6641975
## 54      5    w   3      f 110    40           0.7333333
## 55      5    o   1      m  36    20           0.6428571
## 56      5    o   1      f  25    16           0.6097561
## 57      5    o   2      m   9     7           0.5625000
## 58      5    o   2      f  11     5           0.6875000
## 59      5    o   3      m  16     3           0.8421053
## 60      5    o   3      f   4     5           0.4444444
## 61      6    w   1      m 252   126           0.6666667
## 62      6    w   1      f  97    61           0.6139241
## 63      6    w   2      m 162    72           0.6923077
## 64      6    w   2      f  47    27           0.6351351
## 65      6    w   3      m 199    93           0.6815068
## 66      6    w   3      f  62    24           0.7209302
## 67      6    o   1      m  69    27           0.7187500
## 68      6    o   1      f  45    36           0.5555556
## 69      6    o   2      m  14     7           0.6666667
## 70      6    o   2      f   8     4           0.6666667
```

```
## 71      6    o   3      m  14     5             0.7368421
## 72      6    o   3      f   2     0             1.0000000
## 73      7    w   1      m 119    58             0.6723164
## 74      7    w   1      f  62    33             0.6526316
## 75      7    w   2      m  66    20             0.7674419
## 76      7    w   2      f  20    10             0.6666667
## 77      7    w   3      m  67    21             0.7613636
## 78      7    w   3      f  25    10             0.7142857
## 79      7    o   1      m  45    16             0.7377049
## 80      7    o   1      f  22    15             0.5945946
## 81      7    o   2      m  15    10             0.6000000
## 82      7    o   2      f  10     8             0.5555556
## 83      7    o   3      m   8     6             0.5714286
## 84      7    o   3      f   6     2             0.7500000
##    fitted(job_sat_data.logit)
## 1                   0.5788758
## 2                   0.5518297
## 3                   0.6096837
## 4                   0.5831915
## 5                   0.6640850
## 6                   0.6390995
## 7                   0.6188964
## 8                   0.4986531
## 9                   0.6485540
## 10                  0.5305705
## 11                  0.7002016
## 12                  0.5885571
## 13                  0.6801643
## 14                  0.6557548
## 15                  0.7073088
## 16                  0.6840085
## 17                  0.7536024
## 18                  0.7325941
## 19                  0.7152933
## 20                  0.6061073
## 21                  0.7405935
## 22                  0.6361751
## 23                  0.7832360
## 24                  0.6887689
## 25                  0.6208833
## 26                  0.5946459
## 27                  0.6504736
## 28                  0.6250473
## 29                  0.7019688
## 30                  0.6784366
## 31                  0.6592616
## 32                  0.5423361
## 33                  0.6873648
## 34                  0.5738493
## 35                  0.7356338
## 36                  0.6302158
## 37                  0.5999066
## 38                  0.5732154
## 39                  0.6301590
```

```
## 40                  0.6041542
## 41                  0.6831894
## 42                  0.6588949
## 43                  0.6391744
## 44                  0.5203707
## 45                  0.6681002
## 46                  0.5521474
## 47                  0.7181238
## 48                  0.6094307
## 49                  0.6085823
## 50                  0.5820666
## 51                  0.6385738
## 52                  0.6127972
## 53                  0.6909893
## 54                  0.6670015
## 55                  0.6474989
## 56                  0.5294182
## 57                  0.6760956
## 58                  0.5611010
## 59                  0.7254094
## 60                  0.6180313
## 61                  0.6471245
## 62                  0.6215963
## 63                  0.6757363
## 64                  0.6511622
## 65                  0.7250826
## 66                  0.7026023
## 67                  0.6841975
## 68                  0.5702510
## 69                  0.7111452
## 70                  0.6012552
## 71                  0.7570404
## 72                  0.6561687
## 73                  0.6748057
## 74                  0.6501976
## 75                  0.7022060
## 76                  0.6786840
## 77                  0.7490202
## 78                  0.7277624
## 79                  0.7102726
## 80                  0.6002373
## 81                  0.7358543
## 82                  0.6304800
## 83                  0.7790433
## 84                  0.6834875
```

```
logit.fit.prob <- fitted(job_sat_data.logit)
logit.fit.odds <- logit.fit.prob/(1-logit.fit.prob)

#Creation of a new data frame with added columns
sat_data <- cbind(job_sat_data, logit.fit.prob, logit.fit.odds)

sat_data_region1 <- subset(sat_data, region=="1")
sat_data_region2 <- subset(sat_data, region=="2")
```

```
sat_data_region3 <- subset(sat_data, region=="3")
sat_data_region4 <- subset(sat_data, region=="4")
sat_data_region5 <- subset(sat_data, region=="5")
sat_data_region6 <- subset(sat_data, region=="6")
sat_data_region7 <- subset(sat_data, region=="7")
```

Contingency Tables by Race and Region

```
sat_data_race_w <- subset(sat_data, race=="w")
sat_data_race_o <- subset(sat_data, race=="o")
```

Contingency Tables by Age Group, and Region

```
sat_data_age1 <- subset(sat_data, age=="1")
sat_data_age2 <- subset(sat_data, age=="2")
sat_data_age3 <- subset(sat_data, age=="3")
```

Contingency Table by Gender and Region

```
sat_data_f <- subset(sat_data, gender=="f")
sat_data_m <- subset(sat_data, gender=="m")
```

```
sat_data_w_m <- subset(sat_data, race=="w" & gender =="m")
sat_data_w_f <- subset(sat_data, race=="w" & gender =="f")
sat_data_o_m <- subset(sat_data, race=="o" & gender =="m")
sat_data_o_f <- subset(sat_data, race=="o" & gender =="f")
```