Final_Project.R

jiayuan

Mon Dec 14 05:18:09 2015

```
### Appendix
## R Code and Output for MA684 Final Project 2015
# Jiayuan Shi

library(psych)
library(GPArotation)
library(ltm)
```

```
## Loading required package: MASS
## Loading required package: msm
## Loading required package: polycor
## Loading required package: mvtnorm
## Loading required package: sfsmisc
##
## Attaching package: 'polycor'
##
## The following object is masked from 'package:psych':
##
##     polyserial
##
##
## Attaching package: 'ltm'
##
## The following object is masked from 'package:psych':
##
##     factor.scores
```

```
#1A
survey <- read.csv("VoterValues2015.csv",header=T)
survey <- survey[-c(231:234),]
attach(survey)
```

```
## The following object is masked from package:ltm:
##
##     Abortion
```

```
princomp(~PrivOwn+GayMarriage+Abortion+GovResp+Compete+AssitSuicide,
         cor=TRUE)
```

```
## Call:
## princomp(formula = ~PrivOwn + GayMarriage + Abortion + GovResp +
##     Compete + AssitSuicide, cor = TRUE)
##
## Standard deviations:
##    Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
```

```
## 1.4193163 1.2266881 0.9855628 0.7868822 0.7151894 0.6154380
##
##  6  variables and  230 observations.
```

```
responses <- data.frame(PrivOwn,GayMarriage,Abortion,GovResp,Compete,AssitSuicide)
principal(responses,nfactors=2,rotate="varimax")
```

```
## Principal Components Analysis
## Call: principal(r = responses, nfactors = 2, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                PC1   PC2   h2   u2 com
## PrivOwn      -0.06  0.82 0.67 0.33   1
## GayMarriage   0.80 -0.02 0.64 0.36   1
## Abortion      0.87 -0.05 0.76 0.24   1
## GovResp       0.05  0.35 0.12 0.88   1
## Compete       0.02  0.84 0.70 0.30   1
## AssitSuicide  0.78  0.12 0.61 0.39   1
##
##                        PC1  PC2
## SS loadings           2.01 1.51
## Proportion Var        0.34 0.25
## Cumulative Var        0.34 0.59
## Proportion Explained  0.57 0.43
## Cumulative Proportion 0.57 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.12
##  with the empirical chi square  100.19  with prob <  9e-21
##
## Fit based upon off diagonal values = 0.79
```

```
#1B
moral <- data.frame(GayMarriage,Abortion,AssitSuicide)
political <- data.frame(PrivOwn,Compete)
alpha(moral)$total[[1]]
```

```
## [1] 0.7523868
```

```
alpha(political)$total[[1]]
```

```
## [1] 0.6132668
```

```
#1C
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:sfsmisc':
```

```
##
##     last
##
## The following object is masked from 'package:MASS':
##
##     select
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
region1 <- subset(survey, Region==1)
region2 <- subset(survey, Region==2)

r1.political <- with(region1, PrivOwn + Compete)
r2.political <- with(region2, PrivOwn + Compete)
t.test(r1.political, r2.political, var.equal = TRUE) # test Political Values
```

```
##
##  Two Sample t-test
##
## data:  r1.political and r2.political
## t = -0.16842, df = 228, p-value = 0.8664
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.0775048  0.9078079
## sample estimates:
## mean of x mean of y
##  10.48182  10.56667
```

```r
#2A
LeadStudy <- read.csv("LeadStudy2015(1).csv",header=T)
attach(LeadStudy)
table(lead)
```

```
## lead
##   0   1
## 373 127
```

```r
mean(age[lead==0])
```

```
## [1] 8.546917
```

```r
sd(age[lead==0])
```

```
## [1] 1.418417
```

```r
mean(age[lead==1])
```

```
## [1] 8.551181
```

```r
sd(age[lead==1])
```

```
## [1] 1.587194
```

```r
table(sexF,lead)
```

```
##      lead
## sexF   0   1
##    0 176  58
##    1 197  69
```

```r
table(race,lead)
```

```
##      lead
## race   0   1
##    1 198  54
##    2  65  26
##    3  50  26
##    4  60  21
```

```r
lowlead <- subset(LeadStudy, lead=="0")
highlead <- subset(LeadStudy, lead=="1")
# calculate p-value for age
t.test(lowlead$age, highlead$age, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  lowlead$age and highlead$age
## t = -0.028371, df = 498, p-value = 0.9774
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2995661  0.2910377
## sample estimates:
## mean of x mean of y
##  8.546917  8.551181
```

```r
# calculate p-value for sex
chisq.test(table(sexF, lead), correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(sexF, lead)
## X-squared = 0.08742, df = 1, p-value = 0.7675
```

```r
# calculate p-value for race
chisq.test(table(race, lead), correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(race, lead)
## X-squared = 5.7059, df = 3, p-value = 0.1268
```

```r
#2B
table(ADHD)
```

```
## ADHD
##   0   1
## 398 102
```

```r
mean(iq)
```

```
## [1] 99.172
```

```r
sd(iq)
```

```
## [1] 14.70316
```

```r
#2C
length(iq[lead==1])
```

```
## [1] 127
```

```r
length(iq[lead==0])
```

```
## [1] 373
```

```r
mean(iq[lead==1])
```

```
## [1] 97.1811
```

```r
sd(iq[lead==1])
```

```
## [1] 13.97698
```

```r
mean(iq[lead==0])
```

```
## [1] 99.84987
```

```r
sd(iq[lead==0])
```

```
## [1] 14.89981
```

```r
t.test(iq ~ lead, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  iq by lead
## t = 1.7705, df = 498, p-value = 0.07725
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2927735  5.6303007
## sample estimates:
## mean in group 0 mean in group 1
##        99.84987        97.18110
```

```r
#2D
reg4 <- lm(iq ~ age + sexF + relevel(factor(race), ref = "1") + lead)
summary(reg4)
```

```
##
## Call:
## lm(formula = iq ~ age + sexF + relevel(factor(race), ref = "1") +
##     lead)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.732 -10.075   0.219  10.280  41.228
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        96.1188     3.9506  24.330   <2e-16 ***
## age                                 0.3791     0.4498   0.843   0.3998
## sexF                               -0.1942     1.3091  -0.148   0.8821
## relevel(factor(race), ref = "1")2   2.0169     1.7909   1.126   0.2606
## relevel(factor(race), ref = "1")3  -3.2688     1.9169  -1.705   0.0888 .
## relevel(factor(race), ref = "1")4   4.2298     1.8655   2.267   0.0238 *
## lead                               -2.5169     1.5039  -1.674   0.0949 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.55 on 493 degrees of freedom
## Multiple R-squared:  0.03197,    Adjusted R-squared:  0.02019
## F-statistic: 2.713 on 6 and 493 DF,  p-value: 0.01332
```

```r
#2F
table(lead, ADHD)
```

```
##     ADHD
```

```
## lead   0   1
##    0 311  62
##    1  87  40
```

```r
chisq.test(table(lead, ADHD), correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(lead, ADHD)
## X-squared = 12.908, df = 1, p-value = 0.0003272
```

```r
#2G
log.out <- glm(ADHD ~ age + sexF + relevel(factor(race), ref = "1") + lead,
               family=binomial(link=logit))
summary(log.out)
```

```
##
## Call:
## glm(formula = ADHD ~ age + sexF + relevel(factor(race), ref = "1") +
##     lead, family = binomial(link = logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1747  -0.7238  -0.5835  -0.3133   2.4940
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        -0.80657    0.67798  -1.190  0.23418
## age                                -0.02484    0.07778  -0.319  0.74940
## sexF                               -0.60745    0.23305  -2.606  0.00915 **
## relevel(factor(race), ref = "1")2 -0.22540    0.29835  -0.756  0.44995
## relevel(factor(race), ref = "1")3 -0.96839    0.38156  -2.538  0.01115 *
## relevel(factor(race), ref = "1")4 -1.40195    0.42997  -3.261  0.00111 **
## lead                                0.97409    0.24776   3.932 8.44e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 505.90  on 499  degrees of freedom
## Residual deviance: 469.02  on 493  degrees of freedom
## AIC: 483.02
##
## Number of Fisher Scoring iterations: 5
```

```r
exp(coef(log.out))
```

```
##                       (Intercept)                               age
##                         0.4463862                         0.9754625
##                              sexF relevel(factor(race), ref = "1")2
##                         0.5447391                         0.7981953
```

```
## relevel(factor(race), ref = "1")3 relevel(factor(race), ref = "1")4
##                         0.3796946                         0.2461167
##                              lead
##                         2.6487438
```

```
exp(confint(log.out))
```

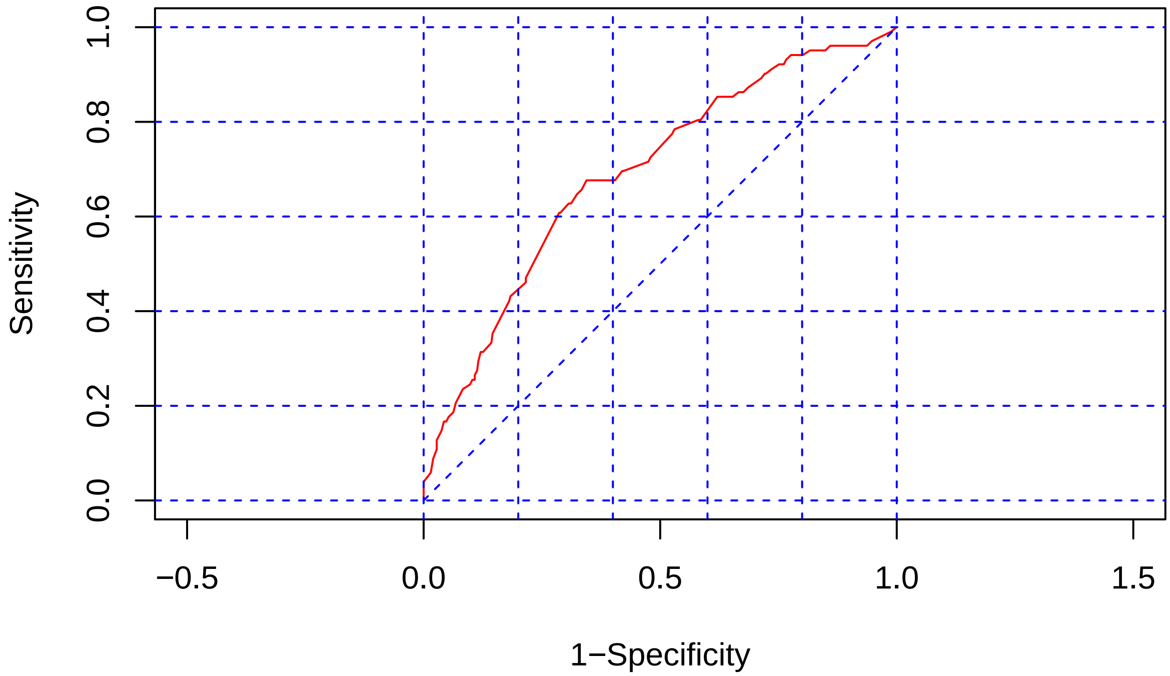
```
## Waiting for profiling to be done...
```

```
##                                          2.5 %    97.5 %
## (Intercept)                         0.11677259 1.6745860
## age                                 0.83717220 1.1362824
## sexF                                0.34330678 0.8575962
## relevel(factor(race), ref = "1")2 0.43728091 1.4145327
## relevel(factor(race), ref = "1")3 0.17074183 0.7719204
## relevel(factor(race), ref = "1")4 0.09747614 0.5380991
## lead                                1.62709265 4.3072478
```

```
library(epicalc)
```

```
## Loading required package: foreign
## Loading required package: survival
## Loading required package: nnet
##
## Attaching package: 'epicalc'
##
## The following object is masked from 'package:dplyr':
##
##     rename
##
## The following objects are masked from 'package:psych':
##
##     alpha, lookup
```

```
lroc(log.out)$auc
```




```
## [1] 0.6914228
```