exrcise 4

2023-04-19

Datastep

```r
library(psych); library(dplyr); library(magrittr);
library(readxl); library(haven)

# Read data

opirx = read_excel("opirx.xls")

# Loading data

one =
  opirx %>%
  mutate(age5 = age / 5,
         opiate = ifelse(opiate == 1, "heroin", "other opioid"),
         newrx = ifelse(newrx == 1, "yes", "no"),
         female = ifelse(female == 1, "yes", "no"),
         homeless = ifelse(homeless == "1_yes", "yes", ifelse(homeless == "0_no", "no", homeless)),
         satreat = as.factor(satreat))

# Applying labels

library(labelled)

var_label(one) =
  list(age = "age at baseline",
       age5 = "age divided by 5",
       opiate = "opiate type",
       newrx = "new medication",
       female = "female",
       homeless = "homeless",
       satreat = "any prior substance abuse treatment",
       racebl = "Black vs. White",
       socsupp = "Social support scale",
       substuse = "use of any substance post detox")

# Viewing contents
str(one)
```

```
## tibble [244 x 13] (S3: tbl_df/tbl/data.frame)
##  $ id        : num [1:244] 1 2 3 4 5 6 8 9 10 12 ...
##  $ age       : num [1:244] 37 37 26 39 32 47 28 50 39 58 ...
```

```
##   ..- attr(*, "label")= chr "age at baseline"
## $ homeless : chr [1:244] "no" "yes" "no" "no" ...
##   ..- attr(*, "label")= chr "homeless"
## $ socsupp  : num [1:244] 0 1 13 11 10 5 4 5 0 13 ...
##   ..- attr(*, "label")= chr "Social support scale"
## $ satreat  : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 2 1 2 2 ...
##   ..- attr(*, "label")= chr "any prior substance abuse treatment"
## $ substuse : num [1:244] 1 1 1 1 1 1 1 1 1 0 ...
##   ..- attr(*, "label")= chr "use of any substance post detox"
## $ daysto   : num [1:244] 177 2 3 180 2 31 47 31 115 180 ...
## $ female   : chr [1:244] "no" "no" "no" "yes" ...
##   ..- attr(*, "label")= chr "female"
## $ newrx    : chr [1:244] "yes" NA "no" "no" ...
##   ..- attr(*, "label")= chr "new medication"
## $ racebl   : chr [1:244] "Black" "White" "Black" "White" ...
##   ..- attr(*, "label")= chr "Black vs. White"
## $ newrxhern: num [1:244] 0 NA 0 0 0 0 0 0 0 0 ...
## $ opiate   : chr [1:244] "other opioid" "other opioid" "heroin" "heroin" ...
##   ..- attr(*, "label")= chr "opiate type"
## $ age5     : num [1:244] 7.4 7.4 5.2 7.8 6.4 9.4 5.6 10 7.8 11.6 ...
##   ..- attr(*, "label")= chr "age divided by 5"
```

```r
# Freq table
  table(one$female)
```

```
##
##  no yes
## 187  57
```

```r
  table(one$homeless)
```

```
##
##  no yes
## 126 118
```

```r
  table(one$racebl)
```

```
##
## Black White
##   127   117
```

```r
  table(one$satreat)
```

```
##
##   0   1
## 176  68
```

```r
  library(tableone)
  CreateTableOne(data = one)
```

2

```
##
##                               Overall
##   n                             244
##   id (mean (SD))         214.77 (143.27)
##   age (mean (SD))           36.26 (7.99)
##   homeless = yes (%)          118 (48.4)
##   socsupp (mean (SD))        6.70 (4.00)
##   satreat = 1 (%)              68 (27.9)
##   substuse (mean (SD))       0.77 (0.42)
##   daysto (mean (SD))       71.60 (72.87)
##   female = yes (%)             57 (23.4)
##   newrx = yes (%)             105 (44.3)
##   racebl = White (%)          117 (48.0)
##   newrxhern (mean (SD))      0.09 (0.28)
##   opiate = other opioid (%)   183 (75.0)
##   age5 (mean (SD))           7.25 (1.60)
```

Q1. logistic models

```r
fit.female = glm(substuse ~ female, data = one, family = binomial(link = "logit"))
summary(fit.female)
```

```
##
## Call:
## glm(formula = substuse ~ female, family = binomial(link = "logit"),
##     data = one)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7706   0.6839   0.6839   0.6839   0.8416
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.3337     0.1800   7.409 1.27e-13 ***
## femaleyes    -0.4780     0.3409  -1.402    0.161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 260.97  on 242  degrees of freedom
## AIC: 264.97
##
## Number of Fisher Scoring iterations: 4
```

```r
exp(coef(fit.female)[2]) #odd ratio
```

```
## femaleyes
## 0.6200318
```

```
exp(confint(fit.female)) #95% CI
```

```
##                 2.5 %    97.5 %
## (Intercept) 2.6965865 5.471694
## femaleyes   0.3207841 1.228767
```

```
fit.homeless = glm(substuse ~ homeless, data = one, family = binomial(link = "logit"))
summary(fit.homeless)
```

```
##
## Call:
## glm(formula = substuse ~ homeless, family = binomial(link = "logit"),
##     data = one)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.734   0.709   0.709   0.736   0.736
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.25276    0.21429   5.846 5.03e-09 ***
## homelessyes -0.08516    0.30454  -0.280     0.78
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 262.80  on 242  degrees of freedom
## AIC: 266.8
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(fit.homeless)[2])
```

```
## homelessyes
##   0.9183673
```

```
exp(confint(fit.homeless))
```

```
##                 2.5 %    97.5 %
## (Intercept) 2.3337556 5.424193
## homelessyes 0.5044573 1.671454
```

```
fit.satreat = glm(substuse ~ satreat, data = one, family = binomial(link = "logit"))
summary(fit.satreat)
```

```
##
## Call:
## glm(formula = substuse ~ satreat, family = binomial(link = "logit"),
```

```
##       data = one)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8626   0.6231   0.7585   0.7585   0.7585
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0986     0.1741   6.311 2.77e-10 ***
## satreat1      0.4418     0.3626   1.218    0.223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 261.32  on 242  degrees of freedom
## AIC: 265.32
##
## Number of Fisher Scoring iterations: 4
```

```
  exp(coef(fit.satreat)[2])
```

```
## satreat1
## 1.555556
```

```
  exp(confint(fit.satreat))
```

```
##                  2.5 %   97.5 %
## (Intercept) 2.1517187 4.264971
## satreat1    0.7837098 3.281859
```

```
# Following function works too
  library(DescTools)
  OddsRatio(fit.female)
```

```
##
## Call:
## glm(formula = substuse ~ female, family = binomial(link = "logit"),
##       data = one)
##
## Odds Ratios:
##                 or or.lci or.uci Pr(>|z|)
## (Intercept) 3.795  2.697  5.472 1.27e-13 ***
## femaleyes   0.620  0.321  1.229   0.1609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Brier Score: 0.175     Nagelkerke R2: 0.012
```

Q2

```
fit.age = glm(substuse ~ age, data = one, family = binomial(link = "logit"))
summary(fit.age)
```

```
##
## Call:
## glm(formula = substuse ~ age, family = binomial(link = "logit"),
##     data = one)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8789   0.6128   0.6857   0.7407   0.9243
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.12632    0.71620   2.969  0.00299 **
## age         -0.02495    0.01886  -1.323  0.18588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 261.13  on 242  degrees of freedom
## AIC: 265.13
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(fit.age)[2]) #odd ratio
```

```
##       age
## 0.9753587
```

```
exp(confint(fit.age)) #95% CI
```

```
##                  2.5 %    97.5 %
## (Intercept) 2.0943832 35.083066
## age         0.9398001  1.012265
```

Q3

```
fit.age5 = glm(substuse ~ age5, data = one, family = binomial(link = "logit"))
summary(fit.age5)
```

```
##
## Call:
```

```
## glm(formula = substuse ~ age5, family = binomial(link = "logit"),
##     data = one)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8789   0.6128   0.6857   0.7407   0.9243
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.1263     0.7162   2.969  0.00299 **
## age5         -0.1247     0.0943  -1.323  0.18588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 261.13  on 242  degrees of freedom
## AIC: 265.13
##
## Number of Fisher Scoring iterations: 4
```

```
  exp(coef(fit.age5)[2]) #odd ratio
```

```
##      age5
## 0.8827176
```

```
  exp(confint(fit.age5)) #95% CI
```

```
##                   2.5 %    97.5 %
## (Intercept) 2.0943832 35.083066
## age5        0.7331241  1.062848
```

Q4

```
# create binary social support var
  one =
  one %>%
  mutate(socsupp.bi = ifelse(socsupp > mean(one$socsupp), 1, 0))
```

```
  fit.socbi = glm(substuse ~ socsupp.bi, data = one, family = binomial(link = "logit"))
  summary(fit.socbi)
```

```
##
## Call:
## glm(formula = substuse ~ socsupp.bi, family = binomial(link = "logit"),
##     data = one)
##
```

```
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.7344   0.7090   0.7090   0.7341   0.7341
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.17351    0.20891   5.617 1.94e-08 ***
## socsupp.bi   0.07925    0.30511   0.260    0.795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 262.81  on 242  degrees of freedom
## AIC: 266.81
##
## Number of Fisher Scoring iterations: 4
```

```
  exp(coef(fit.socbi)[2]) #odd ratio
```

```
## socsupp.bi
##   1.082474
```

```
  exp(confint(fit.socbi)) #95% CI
```

```
##                  2.5 %    97.5 %
## (Intercept) 2.1756079 4.949773
## socsupp.bi  0.5954595 1.977538
```

Q5

```
  fit.all = glm(substuse ~ age5 + female + homeless + satreat + racebl + socsupp.bi,
                data = one, family = binomial(link = "logit"))
  summary(fit.all)
```

```
##
## Call:
## glm(formula = substuse ~ age5 + female + homeless + satreat +
##     racebl + socsupp.bi, family = binomial(link = "logit"), data = one)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.0904   0.4958   0.6522   0.7474   1.1248
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.10856    0.76044   2.773  0.00556 **
## age5        -0.10119    0.09617  -1.052  0.29272
```

8

```
## femaleyes   -0.56694    0.35540  -1.595  0.11067
## homelessyes -0.12049    0.31850  -0.378  0.70521
## satreat1     0.47824    0.37052   1.291  0.19681
## raceblWhite -0.28432    0.31247  -0.910  0.36287
## socsupp.bi   0.14664    0.31741   0.462  0.64408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 262.88  on 243  degrees of freedom
## Residual deviance: 256.45  on 237  degrees of freedom
## AIC: 270.45
##
## Number of Fisher Scoring iterations: 4
```

```
  library(DescTools)
  OddsRatio(fit.all)
```

```
##
## Call:
## glm(formula = substuse ~ age5 + female + homeless + satreat +
##     racebl + socsupp.bi, family = binomial(link = "logit"), data = one)
##
## Odds Ratios:
##                 or or.lci or.uci Pr(>|z|)
## (Intercept) 8.236  1.874 37.366   0.0056 **
## age5        0.904  0.748  1.092   0.2927
## femaleyes   0.567  0.284  1.153   0.1107
## homelessyes 0.886  0.474  1.658   0.7052
## satreat1    1.613  0.800  3.454   0.1968
## raceblWhite 0.753  0.406  1.388   0.3629
## socsupp.bi  1.158  0.622  2.170   0.6441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Brier Score: 0.172     Nagelkerke R2: 0.039
```