PH241 HW12

Kunal Mishra

4/14/2018

Basic Setup and Minimal EDA

```
library(dplyr)
library(dummies)

rawdata = read.csv(file="HW12.csv", header=TRUE) %>% select(-X)
colnames(rawdata) = c("Activity_Level", "Low_Fat_Diet", "All_Cause_Mortality")

# Dropping one of the columns to use as a reference variable
data = dummy.data.frame(data=rawdata, names=c("Activity_Level")) %>% select(-Activity_Level0)

data %>% nrow
```

```
## [1] 500
```

```
data %>% head
```

```
##   Activity_Level1 Activity_Level2 Activity_Level3 Low_Fat_Diet
## 1               0               0               0            0
## 2               0               0               0            0
## 3               0               0               0            0
## 4               0               0               0            0
## 5               0               0               0            0
## 6               0               0               0            0
##   All_Cause_Mortality
## 1                   1
## 2                   0
## 3                   1
## 4                   1
## 5                   0
## 6                   0
```

Fitting 3 Logistic Regression Models

- Model 1: B0 + B1*Diet
- Model 2: B0 + B1*Diet + B2*act1 + B3*act2 + B4*act3
- Model 3: B0 + B1*Diet + B2*act1 + B3*act2 + B4*act3 + B5*Diet*act1 + B6*Diet*act2 + B7*Diet*act3

```
fit1 = glm(formula=All_Cause_Mortality~Low_Fat_Diet,
           family="binomial",
           data=data)

fit2 = glm(formula=All_Cause_Mortality~Low_Fat_Diet + Activity_Level1 + Activity_Level2 + Activity_Level
           family="binomial",
           data=data)
```

```r
fit3 = glm(formula=All_Cause_Mortality~Low_Fat_Diet + Activity_Level1 + Activity_Level2 + Activity_Level
           family="binomial",
           data=data)

fit.report = function(fit) {
    # Reporting Log Odds Ratios (Model fit)
    fit %>% summary %>% print
    writeLines("\n\nORs")
    # Reporting Odds Ratios (Coefficients)
    fit %>% coef %>% exp %>% print
    writeLines("\n\nConfidence Intervals of ORs")
    #Reporting Confidence Intervals of Odds Ratios
    fit %>% confint %>% exp %>% print
    writeLines("\n\n")
}

fit.report(fit1)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet, family = "binomial",
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0477  -1.0477  -0.8974   1.3129   1.4861
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.3131     0.1128  -2.775  0.00552 **
## Low_Fat_Diet  -0.3885     0.1951  -1.991  0.04650 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 668.75  on 499  degrees of freedom
## Residual deviance: 664.73  on 498  degrees of freedom
## AIC: 668.73
##
## Number of Fisher Scoring iterations: 4
##
##
##
## ORs
##  (Intercept) Low_Fat_Diet
##    0.7311828    0.6780771
##
##
## Confidence Intervals of ORs
## Waiting for profiling to be done...
##                   2.5 %    97.5 %
## (Intercept)  0.5852476 0.9111922
```

```
## Low_Fat_Diet 0.4608566 0.9912398
```

```
fit.report(fit2)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet + Activity_Level1 +
##     Activity_Level2 + Activity_Level3, family = "binomial", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1760  -1.0752  -0.6258   1.1788   1.9646
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.003233   0.151715  -0.021 0.982999
## Low_Fat_Diet    -0.241964   0.207388  -1.167 0.243324
## Activity_Level1 -0.068444   0.224628  -0.305 0.760595
## Activity_Level2 -0.965307   0.282951  -3.412 0.000646 ***
## Activity_Level3 -1.527868   0.362650  -4.213 2.52e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 668.75  on 499  degrees of freedom
## Residual deviance: 633.16  on 495  degrees of freedom
## AIC: 643.16
##
## Number of Fisher Scoring iterations: 4
##
##
##
## ORs
##     (Intercept)    Low_Fat_Diet Activity_Level1 Activity_Level2
##       0.9967723       0.7850843       0.9338458       0.3808663
## Activity_Level3
##       0.2169979
##
##
## Confidence Intervals of ORs
```

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

```
##                     2.5 %    97.5 %
## (Intercept)     0.7399446 1.3424847
## Low_Fat_Diet    0.5215322 1.1771964
## Activity_Level1 0.6008294 1.4507387
## Activity_Level2 0.2157219 0.6562617
## Activity_Level3 0.1019135 0.4272937
```

```
fit.report(fit3)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet + Activity_Level1 +
```

```
##     Activity_Level2 + Activity_Level3 + Low_Fat_Diet * Activity_Level1 +
##     Low_Fat_Diet * Activity_Level2 + Low_Fat_Diet * Activity_Level3,
##     family = "binomial", data = data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.177  -1.073  -0.629   1.177   1.973
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -4.953e-16  1.633e-01   0.000 1.000000
## Low_Fat_Diet                 -2.578e-01  3.619e-01  -0.713 0.476152
## Activity_Level1              -1.178e-01  2.718e-01  -0.433 0.664794
## Activity_Level2              -8.873e-01  3.571e-01  -2.485 0.012960 *
## Activity_Level3              -1.520e+00  4.481e-01  -3.392 0.000694 ***
## Low_Fat_Diet:Activity_Level1  1.243e-01  4.916e-01   0.253 0.800385
## Low_Fat_Diet:Activity_Level2 -1.632e-01  5.990e-01  -0.272 0.785275
## Low_Fat_Diet:Activity_Level3 -1.410e-02  7.725e-01  -0.018 0.985432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 668.75  on 499  degrees of freedom
## Residual deviance: 632.91  on 492  degrees of freedom
## AIC: 648.91
##
## Number of Fisher Scoring iterations: 4
##
##
##
## ORs
##                   (Intercept)                  Low_Fat_Diet
##                     1.0000000                     0.7727273
##               Activity_Level1               Activity_Level2
##                     0.8888889                     0.4117647
##               Activity_Level3 Low_Fat_Diet:Activity_Level1
##                     0.2187500                     1.1323529
## Low_Fat_Diet:Activity_Level2 Low_Fat_Diet:Activity_Level3
##                     0.8494208                     0.9859944
##
##
## Confidence Intervals of ORs
## Waiting for profiling to be done...
##                                    2.5 %    97.5 %
## (Intercept)                   0.72560642 1.3781576
## Low_Fat_Diet                  0.37609412 1.5657927
## Activity_Level1               0.52064997 1.5140975
## Activity_Level2               0.19929770 0.8147570
## Activity_Level3               0.08429673 0.5000593
## Low_Fat_Diet:Activity_Level1 0.43234484 2.9821729
## Low_Fat_Diet:Activity_Level2 0.25950837 2.7431680
## Low_Fat_Diet:Activity_Level3 0.20222254 4.3847318
```

Question 1C

```
activityLevel0 = rawdata %>% filter(Activity_Level==0)
activityLevel1 = rawdata %>% filter(Activity_Level==1)
activityLevel2 = rawdata %>% filter(Activity_Level==2)
activityLevel3 = rawdata %>% filter(Activity_Level==3)

fit.activityLevel0 = glm(formula=All_Cause_Mortality~Low_Fat_Diet, family="binomial", data=activityLevel
fit.activityLevel1 = glm(formula=All_Cause_Mortality~Low_Fat_Diet, family="binomial", data=activityLevel
fit.activityLevel2 = glm(formula=All_Cause_Mortality~Low_Fat_Diet, family="binomial", data=activityLevel
fit.activityLevel3 = glm(formula=All_Cause_Mortality~Low_Fat_Diet, family="binomial", data=activityLevel

fit.report(fit.activityLevel0)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet, family = "binomial",
##     data = activityLevel0)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.177  -1.177  -1.070   1.177   1.289
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.618e-17  1.633e-01   0.000    1.000
## Low_Fat_Diet -2.578e-01  3.619e-01  -0.713    0.476
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 261.88  on 188  degrees of freedom
## Residual deviance: 261.37  on 187  degrees of freedom
## AIC: 265.37
##
## Number of Fisher Scoring iterations: 3
##
##
##
## ORs
##  (Intercept) Low_Fat_Diet
##    1.0000000    0.7727273
##
##
## Confidence Intervals of ORs
## Waiting for profiling to be done...
##                  2.5 %   97.5 %
## (Intercept)  0.7256064 1.378158
## Low_Fat_Diet 0.3760941 1.565793
```

```
fit.report(fit.activityLevel1)
```

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

```
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet, family = "binomial",
##     data = activityLevel1)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.128  -1.128  -1.073   1.228   1.286
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.1178     0.2173  -0.542    0.588
## Low_Fat_Diet  -0.1335     0.3327  -0.401    0.688
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 205.42  on 148  degrees of freedom
## Residual deviance: 205.26  on 147  degrees of freedom
## AIC: 209.26
##
## Number of Fisher Scoring iterations: 3
##
##
##
## ORs
##  (Intercept) Low_Fat_Diet
##    0.8888889    0.8750000
##
##
## Confidence Intervals of ORs

## Waiting for profiling to be done...

##                  2.5 %   97.5 %
## (Intercept)  0.5786216 1.360649
## Low_Fat_Diet 0.4540917 1.678720
```

```
fit.report(fit.activityLevel2)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet, family = "binomial",
##     data = activityLevel2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8305  -0.8305  -0.6917   0.4390   1.7593
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.8873     0.3176  -2.794   0.0052 **
## Low_Fat_Diet  -0.4210     0.4774  -0.882   0.3778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.39  on 94  degrees of freedom
```

```
## Residual deviance: 106.60  on 93  degrees of freedom
## AIC: 110.6
##
## Number of Fisher Scoring iterations: 4
##
##
##
## ORs
##  (Intercept) Low_Fat_Diet
##    0.4117647    0.6563707
##
##
## Confidence Intervals of ORs
## Waiting for profiling to be done...
##                   2.5 %    97.5 %
## (Intercept)  0.2139123 0.7507005
## Low_Fat_Diet 0.2518408 1.6617848
```

```
fit.report(fit.activityLevel3)
```

```
##
## Call:
## glm(formula = All_Cause_Mortality ~ Low_Fat_Diet, family = "binomial",
##     data = activityLevel3)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6290  -0.6290  -0.5553  -0.5553   1.9728
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.5198     0.4173  -3.642  0.00027 ***
## Low_Fat_Diet  -0.2719     0.6825  -0.398  0.69030
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 59.836  on 66  degrees of freedom
## Residual deviance: 59.674  on 65  degrees of freedom
## AIC: 63.674
##
## Number of Fisher Scoring iterations: 4
##
##
##
## ORs
##  (Intercept) Low_Fat_Diet
##    0.2187500    0.7619048
##
##
## Confidence Intervals of ORs
## Waiting for profiling to be done...
```

```
##                   2.5 %   97.5 %
## (Intercept)  0.08862441 0.466893
## Low_Fat_Diet 0.18222471 2.824566
```

Question 1F

```
stratifiedData = array(
    c(c(17, 75, 22, 75),
      c(28, 40, 36, 45),
      c(10, 14, 37, 34),
      c(4, 7, 24, 32)),
    dim=c(2,2,4),
    dimnames=list( c("Low-Fat Diet", "Other Diet"),
                   c("Dead", "Alive"),
                   c("AL0", "AL1", "AL2", "Al3")
    )
)
mantelhaen.test(stratifiedData)
```

```
##
##  Mantel-Haenszel chi-squared test with continuity correction
##
## data:  stratifiedData
## Mantel-Haenszel X-squared = 1.1249, df = 1, p-value = 0.2889
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.5228808 1.1789039
## sample estimates:
## common odds ratio
##         0.7851281
```