250C HW 1

Esther Chung

February 1, 2018

```
setwd("/Users/estherchung/Documents/Graduate/Spring 2018/250C Advanced Epi Methods/Homework/Homework 1")
load("frmgham_recoded.wide.Rdata")

# Load required packages:
library(geepack)
library(doBy)
library(stats4) # needed for MLE function library(blm)
library(xtable)
```

Calculate the proportion of the sample who experienced the event:

```
prop <- table(frmgham_recoded.wide$hyperten) / sum(table(frmgham_recoded.wide$hyperten))
prop
```

```
##
##         0         1
## 0.3424598 0.6575402
```

The proportion of the sample who experienced hypertension was 65.8%.

Estimate a logistic regression for the association of BMI on incident hypertension, adjusted for current smoking (binary), age (continuous), sex (binary), education (4-level):

```
# Make 2nd category (BMI 18.5-24.9, ideal weight) the referent group:
frmgham_recoded.wide$bmi_cat <- relevel(as.factor(frmgham_recoded.wide$bmi_cat),"2")

# Estimate logistic regression:
logistic.frmgham <- glm(hyperten ~ factor(bmi_cat) + cursmoke + age + factor(sex) +
                        factor(educ), data=frmgham_recoded.wide, family=binomial)

summary(logistic.frmgham)
```

```
##
## Call:
## glm(formula = hyperten ~ factor(bmi_cat) + cursmoke + age + factor(sex) +
##     factor(educ), family = binomial, data = frmgham_recoded.wide)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1608  -1.2652   0.7276   0.9496   1.4494
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.17500    0.33312  -3.527 0.000420 ***
## factor(bmi_cat)1 -0.33155    0.33201  -0.999 0.317976
## factor(bmi_cat)3  0.45022    0.10001   4.502 6.74e-06 ***
## factor(bmi_cat)4  1.26440    0.22206   5.694 1.24e-08 ***
## cursmoke         -0.18592    0.09451  -1.967 0.049156 *
## age               0.03639    0.00614   5.927 3.09e-09 ***
## factor(sex)2      0.19609    0.09689   2.024 0.042987 *
```

```
## factor(educ)2    -0.09546    0.11323  -0.843 0.399170
## factor(educ)3    -0.16995    0.13273  -1.280 0.200396
## factor(educ)4    -0.53273    0.14286  -3.729 0.000192 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2957.5  on 2300  degrees of freedom
## Residual deviance: 2821.1  on 2291  degrees of freedom
## AIC: 2841.1
##
## Number of Fisher Scoring iterations: 4
```

```
coef.logistic <- coef(logistic.frmgham)

ci.logistic <- confint(logistic.frmgham)
```

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

```
logistic <- round(exp(cbind(coef.logistic, ci.logistic))[2:4,],2)
```

Estimate a log-binomial model for the BMI-hypertension association.

Estimate a modified Poisson model for the BMI-hypertension association.

```
# Estimate modified Poisson regression:
poiss.frmgham <- geeglm(formula=hyperten ~ factor(bmi_cat) + cursmoke + age +
                        factor(sex) + factor(educ), data=frmgham_recoded.wide,
                        id=randid,
                        family=poisson(link="log"),
                        corstr = "exchangeable")
summary(poiss.frmgham)
```

```
##
## Call:
## geeglm(formula = hyperten ~ factor(bmi_cat) + cursmoke + age +
##     factor(sex) + factor(educ), family = poisson(link = "log"),
##     data = frmgham_recoded.wide, id = randid, corstr = "exchangeable")
##
##  Coefficients:
##                   Estimate   Std.err   Wald Pr(>|W|)
## (Intercept)      -1.001931  0.104090 92.653  < 2e-16 ***
## factor(bmi_cat)1 -0.145207  0.157683  0.848  0.35712
## factor(bmi_cat)3  0.149684  0.033001 20.572 5.74e-06 ***
## factor(bmi_cat)4  0.318544  0.040572 61.643 4.11e-15 ***
## cursmoke         -0.060456  0.031086  3.782  0.05180 .
## age               0.011153  0.001828 37.222 1.05e-09 ***
## factor(sex)2      0.056085  0.030884  3.298  0.06937 .
## factor(educ)2    -0.024293  0.035783  0.461  0.49721
## factor(educ)3    -0.047818  0.042671  1.256  0.26245
## factor(educ)4    -0.191890  0.056477 11.544  0.00068 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Estimated Scale Parameters:
##             Estimate Std.err
## (Intercept)   0.3423 0.01165
##
## Correlation: Structure = exchangeable  Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha        0       0
## Number of clusters:   2301   Maximum cluster size: 1
```

```
coef.poiss <- coef(poiss.frmgham)

# RR CIs with robust SEs:
require(doBy)
fram.RR.coefci.poiss <- esticon(poiss.frmgham, diag(length(coef(poiss.frmgham))))

fram.RR.expci.poiss <- exp(cbind(fram.RR.coefci.poiss$Estimate, fram.RR.coefci.poiss$Lower,
                                 fram.RR.coefci.poiss$Upper))

rownames(fram.RR.expci.poiss) <- names(coef(poiss.frmgham))
colnames(fram.RR.expci.poiss) <- c("RR", "95% LL", "95% UL")

# RR from Poisson model
fram.RR.expci.poiss
```

```
##                      RR 95% LL 95% UL
## (Intercept)      0.3672 0.2994 0.4503
## factor(bmi_cat)1 0.8648 0.6349 1.1780
## factor(bmi_cat)3 1.1615 1.0887 1.2391
## factor(bmi_cat)4 1.3751 1.2700 1.4889
## cursmoke         0.9413 0.8857 1.0005
## age              1.0112 1.0076 1.0148
## factor(sex)2     1.0577 0.9956 1.1237
## factor(educ)2    0.9760 0.9099 1.0469
## factor(educ)3    0.9533 0.8768 1.0365
## factor(educ)4    0.8254 0.7389 0.9220
```

```
poisson <- round(fram.RR.expci.poiss[2:4,],2) # only interested in BMI categories rows
```

```
row1 <- c("0.72 (0.37, 1.38)", "--", "0.86 (0.63, 1.18)")
row2 <- c("ref", "--", "ref")
row3 <- c("1.57 (1.29, 1.91)", "--", "1.16 (1.09, 1.24)")
row4 <- c("3.54 (2.33, 5.58)", "--", "1.38 (1.27, 1.49)")

table1 <- as.data.frame(rbind(row1, row2, row3, row4))
row.names(table1) <- c("<18.5", "18.5-24.9", "25.0-29.9", ">30.0")
names(table1) <- c("Logistic OR (95% CI)", "Log-binomial RR (95% CI)", "Poisson RR (95% CI)")
```

```
print(xtable(table1, caption="Table 2", align = "l|ccc"))
```

```
## % latex table generated in R 3.4.3 by xtable 1.8-2 package
## % Sun Feb  4 15:07:23 2018
## \begin{table}[ht]
## \centering
```

```
## \begin{tabular}{l|ccc}
##   \hline
##  & Logistic OR (95\% CI) & Log-binomial RR (95\% CI) & Poisson RR (95\% CI) \\
##   \hline
## $<$18.5 & 0.72 (0.37, 1.38) & -- & 0.86 (0.63, 1.18) \\
##   18.5-24.9 & ref & -- & ref \\
##   25.0-29.9 & 1.57 (1.29, 1.91) & -- & 1.16 (1.09, 1.24) \\
##   $>$30.0 & 3.54 (2.33, 5.58) & -- & 1.38 (1.27, 1.49) \\
##    \hline
## \end{tabular}
## \caption{Table 2}
## \end{table}
```

Model-Based Standardization with Logistic Model

```r
# Create copies of the original dataset:
frmgham_recoded.wide.obese <- frmgham_recoded.wide.ideal <- frmgham_recoded.wide

# Set BMI to obese (in p1) and ideal weight (in p0):
frmgham_recoded.wide.obese$bmi_cat <- 4 # Framingham population w/ all obese
frmgham_recoded.wide.ideal$bmi_cat <- 2 # Framingham population w/ all ideal weight

# Obtain predicted individual risk of hypertension under each new dataset:
rhat.obese <- predict(logistic.frmgham, type="response", newdata=frmgham_recoded.wide.obese)
rhat.ideal <- predict(logistic.frmgham, type="response", newdata=frmgham_recoded.wide.ideal)

# Calculate the average risk of hypertension in each hypothetical population:
mu.rhat.obese <- mean(rhat.obese)
mu.rhat.ideal <- mean(rhat.ideal)

# Estimate the risk ratio
RR <- mu.rhat.obese/mu.rhat.ideal
RR
```

```
## [1] 1.39
```

```r
# Estiamte the risk difference
RD <- mu.rhat.obese - mu.rhat.ideal
RD
```

```
## [1] 0.2355
```

Model-Based Standardization with Poisson Model

```r
# Obtain predicted individual risk of hypertension under each new dataset:
rhat.obese.poiss <- predict(poiss.frmgham, type="response", newdata=frmgham_recoded.wide.obese)
rhat.ideal.poiss <- predict(poiss.frmgham, type="response", newdata=frmgham_recoded.wide.ideal)

# Calculate the average risk of hypertension in each hypothetical population:
mu.rhat.obese.poiss <- mean(rhat.obese.poiss)
mu.rhat.ideal.poiss <- mean(rhat.ideal.poiss)
```

```r
# Estimate the risk ratio
RR.poiss <- mu.rhat.obese.poiss/mu.rhat.ideal.poiss
RR.poiss
```

```
## [1] 1.375
```

```r
# Estimate the risk difference
RD.poiss <- mu.rhat.obese.poiss - mu.rhat.ideal.poiss
RD.poiss
```

```
## [1] 0.2261
```

Questions:

1. Using the notation for generalized linear models presented in class, write out the equations for each of the 3 models, in terms of the variables in the dataset. Clearly define all parameters in each of the models.