Solution to Exercise 3: Multivariable analysis in R part 1: Logistic regression

Key points:

a. Logistic regression is the standard and most commonly used approach to multivariable analysis if the questions cannot be answered by adjusting in a stratified analysis using Mantel-Haenszel estimation of the odds ratio

b. Use stepwise exclusion of variables to improve the model fit, by observing p values and the AIC summary for quality of model fit

c. Beware of heterogeneity, test formally for heterogeneity using Woolf's test

Task:

- ***Examine whether there are interactions between age and sex and fluoroquinolone resistant and age, so that all possibilities have been checked before we accept the model.***

Solution:

The e_ex03.r script:

```
# Exercise 3: Multivariable analysis in R part 1: Logistic regression

rm(list=ls())

e_ex03 <- read.table("e_ex02_02.dat")
attach(e_ex03)

# Logistic regression with AGE as conintuous variable
mylogit <- glm(formula = outcome02 ~ factor(fq02) + factor(sex) + age, data=e_ex03,
     family=binomial)
summary(mylogit)
lreg.or <-exp(cbind(OR = coef(mylogit), confint(mylogit)))
round(lreg.or, digits=4)

# Drop variables
names(e_ex03)
e_ex03b <- e_ex03[c(1:4, 7:13)]
detach(e_ex03)
attach(e_ex03b)
names(e_ex03b)

# Create new variable for Quartiles of age
# Age as categorical variables in quartiles
quantile(age,  probs = c(25, 50, 75)/100)
## 25% 50% 75%
## 23  31  42

attach(e_ex03b)
e_ex03b$agecat[age >= 00 & age < 23] <- "Q1"
e_ex03b$agecat[age >= 23 & age < 31] <- "Q2"
e_ex03b$agecat[age >= 31 & age < 42] <- "Q3"
e_ex03b$agecat[age >= 42]            <- "Q4"
e_ex03b$out02[outcome02 == "Success"] <- 0
e_ex03b$out02[outcome02 == "Failure"] <- 1
detach(e_ex03b)
```

```
e_ex03c <- data.frame(e_ex03b)
write.table(e_ex03c, file="e_ex03c", row.names=TRUE)
attach(e_ex03c)
names(e_ex03c)
table(agecat, out02)

# Logistic regression with AGE as categorical variable
mylogit2 <- glm(formula = out02 ~ factor(fq02) + factor(sex) + factor(agecat), data=
     e_ex03c, family=binomial)
summary(mylogit2)
lreg.or <-exp(cbind(OR = coef(mylogit2), confint(mylogit2)))
round(lreg.or, digits=4)

# Full model logistic regression
attach(e_ex03c)
e_ex03c$fq03[fq04 == "Susceptible"]           <- "1-Susceptible"
e_ex03c$fq03[fq04 == "Low-level resistance"]  <- "2-Low-level resistance"
e_ex03c$fq03[fq04 == "High-level resistance"] <- "3-High-level resistance"

detach(e_ex03c)
e_ex03d <- data.frame(e_ex03c)
write.table(e_ex03d, file="e_ex03d", row.names=TRUE)
e_ex03e <- read.table("e_ex03d")

# Full model
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age + factor(pza02) +
     factor(kmy02) + factor(pth02) + factor(cxr02), data=e_ex03e, family=binomial)
summary(mylogit3)
# AIC 407.95

# Remove kmy02
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age + factor(pza02) +
     factor(pth02) + factor(cxr02), data=e_ex03e, family=binomial)
summary(mylogit3)
# AIC 405.96

# Remove cxr02
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age + factor(pza02) +
     factor(pth02), data=e_ex03e, family=binomial)
summary(mylogit3)
# AIC 403.99

# Remove pth02
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age + factor(pza02),
     data=e_ex03e, family=binomial)
summary(mylogit3)
# AIC 402.54

# Remove pza02
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age, data=e_ex03e,
     family=binomial)
summary(mylogit3)
# AIC 401.55

# Test for interaction, using the script for
# Woolf's test provided by Mark Myatt
load("c:/epidata_course/woolf.test.r")
attach(e_ex03e)
tabofl <- table(fq03, out02, sex)
woolf.test(tabofl)
# p-value : 0.6336425
tabofl <- table(fq03, out02, agecat)
woolf.test(tabofl)
# p-value : 0.9898
tabofl <- table(sex, out02, agecat)
```

```
woolf.test(tabofl)
# p-value : 0.5075


# Put in the interaction term anyhow
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age +
     factor(sex):factor(fq03), data=e_ex03e, family=binomial)
summary(mylogit3)
# AIC 405.45
# => AIC worsens

# Final model:
mylogit3 <- glm(formula = out02 ~ factor(fq03) + factor(sex) + age, data=e_ex03e,
     family=binomial)
summary(mylogit3)
```

The Woolf's test for heterogeneity:

```
tabofl <- table(fq03, out02, sex)
woolf.test(tabofl)
tabofl <- table(fq03, out02, agecat)
woolf.test(tabofl)
tabofl <- table(sex, out02, agecat)
woolf.test(tabofl)
```

```
tabofl <- table(fq03, out02, sex)
Woolf's X2 : 0.2271536
p-value : 0.6336425

tabofl <- table(fq03, out02, agecat)
Woolf's X2 : 0.115918
p-value : 0.9898611

tabofl <- table(sex, out02, agecat)
Woolf's X2 : 2.326063
p-value : 0.5075462
```

Conclusion:

No interaction between fluoroquinolone resistance and sex, fluoroquinolone resistance and age, nor sex and age. Therfore the final model does not need an interaction term.