Tutorial Assignment 4 - SOLUTIONS

Instructor: Kevin McGregor

MATH 4330

Question 1:

Fitting the Poisson regression model using `glm()`. First, need to calculate the person time, which is the number of individuals in each group multiplied by 6 (since there are 6 years of follow-up).

```
data <- read.csv("lung.csv")

# Calculating person-time
pt <- 6*data$pop

# Running Poisson regression model with offset
fit <- glm(dead~age+offset(log(pt)),
              data=data, family = poisson)
summary(fit)
```

```
##
## Call:
## glm(formula = dead ~ age + offset(log(pt)), family = poisson,
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.1822  -1.7290  -0.7631   0.8167   4.9551
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.18748    0.05842 -88.795  < 2e-16 ***
## age45-59     0.55603    0.07999   6.951 3.62e-12 ***
## age50-54     0.98815    0.07681  12.864  < 2e-16 ***
## age55-59     1.37145    0.06526  21.017  < 2e-16 ***
## age60-64     1.62900    0.06254  26.049  < 2e-16 ***
## age65-69     1.95715    0.06269  31.218  < 2e-16 ***
## age70-74     2.20577    0.06410  34.409  < 2e-16 ***
## age75-79     2.45779    0.06713  36.610  < 2e-16 ***
## age80+       2.68749    0.07080  37.958  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4055.98  on 35  degrees of freedom
## Residual deviance:  191.72  on 27  degrees of freedom
## AIC: 449.75
##
## Number of Fisher Scoring iterations: 4
```

Question 2:

The reference category is the 40-44 year-old age group. Therefore, to estimate the rate ratio, we extract the beta corresponding to 65-69 year-olds. Confidence interval included as well.

```
# Estimated rate ratio between 65-69 year-olds and 40-44 year-olds.
exp(fit$coefficients[6])
```

```
## age65-69
## 7.079088
```

```
# CI
exp(confint(fit)[6,])
```

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

```
##    2.5 %   97.5 %
## 6.271920 8.019872
```

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

The rate ratio is 7.0791. This means that 65-69 year-olds have a lung cancer rate that's 7.0791 times higher than 40-44 year-olds. The confidence interval is given by $(6.2719, 8.0199)$. Since it does not contain 1, we have that 65-69 year olds have a significanly higher lung cancer rate compared to 40-44 year-olds.

Question 3:

Below is the predicted rate:

```
# Rate per 1000 person-years
ndat <- data.frame(age="65-69", pt=1000)
# Predict with type=response will give the estimated rate.
pred.rate <- predict(fit, newdata = ndat, type="response")
pred.rate
```

```
##        1
## 39.54421
```