stat431 a3 q1

Yiming Shen 20891774

16/11/2023

```
library(GLMsData) # Load the library (if unavailable use install.packages())

data(danishlc) # Danish lung cancer incidence data

# By default ">74" is the comparison group for Age. Here we redefine the levels of Age so they are orde
danishlc$Age = C(factor(danishlc$Age, levels=c("40-54", "55-59", "60-64", "65-69", "70-74", ">74"), ord

# Fit the main effects log linear model
model1 = glm(Cases ~ Age + City + offset(log(Pop)), family=poisson, data=danishlc)


summary(model1)
```

```
##
## Call:
## glm(formula = Cases ~ Age + City + offset(log(Pop)), family = poisson,
##     data = danishlc)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.63573  -0.67296  -0.03436   0.37258   1.85267
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.6321     0.2003 -28.125  < 2e-16 ***
## Age55-59      1.1010     0.2483   4.434 9.23e-06 ***
## Age60-64      1.5186     0.2316   6.556 5.53e-11 ***
## Age65-69      1.7677     0.2294   7.704 1.31e-14 ***
## Age70-74      1.8569     0.2353   7.891 3.00e-15 ***
## Age>74        1.4197     0.2503   5.672 1.41e-08 ***
## CityHorsens  -0.3301     0.1815  -1.818   0.0690 .
## CityKolding  -0.3715     0.1878  -1.978   0.0479 *
## CityVejle    -0.2723     0.1879  -1.450   0.1472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 129.908  on 23  degrees of freedom
## Residual deviance:  23.447  on 15  degrees of freedom
## AIC: 137.84
##
```

```
## Number of Fisher Scoring iterations: 5
```

(a)

```
estimate_beta0 <- -5.6321
estimate_beta1 <- 1.1010
estimate_beta7 <- -0.3715
m <- 1050
estimate_miu <- exp(estimate_beta0+estimate_beta1+estimate_beta7)*m
estimate_miu
```

```
## [1] 7.798609
```

(b)

```
estimate_beta4 <- 1.8569
relative_increase <- exp(estimate_beta4)
relative_increase
```

```
## [1] 6.403854
```

```
# 95% CI for beta4
c <- 1.96
se <- 0.2353
L <- estimate_beta4 - c*se
U <- estimate_beta4 + c*se

# 95% CI for Relative Increase
exp(L)
```

```
## [1] 4.037848
```

```
exp(U)
```

```
## [1] 10.15624
```

(c)

```
D <- 23.447
df <- 15
p_value <- 1-pchisq(D,df)
p_value
```

```
## [1] 0.07509937
```

(d)

```
estimate_miu
```

```
## [1] 7.798609
```

```
m
```

```
## [1] 1050
```

```
# rate per person-year in 1970
estimate_miu/ (3*m)
```

```
## [1] 0.002475749
```

```
# rate per 100000 person-years in 1970
p <- 100000
estimate_miu/ (3*m) * p
```

```
## [1] 247.5749
```

(e)

```
# Fit the main effects logistic regression model based on Cases/Pop
model2 = glm(Cases/Pop ~ Age + City, weights = Pop,
             family=binomial(link = logit), data=danishlc)


summary(model2)
```

```
##
## Call:
## glm(formula = Cases/Pop ~ Age + City, family = binomial(link = logit),
##     data = danishlc, weights = Pop)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.64532  -0.67472  -0.03449   0.37480   1.85912
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.6262     0.2008 -28.021  < 2e-16 ***
## Age55-59      1.1070     0.2490   4.445 8.77e-06 ***
## Age60-64      1.5291     0.2325   6.577 4.81e-11 ***
## Age65-69      1.7819     0.2305   7.732 1.06e-14 ***
## Age70-74      1.8727     0.2365   7.918 2.42e-15 ***
## Age>74        1.4289     0.2512   5.688 1.29e-08 ***
## CityHorsens  -0.3345     0.1827  -1.830   0.0672 .
## CityKolding  -0.3764     0.1890  -1.991   0.0465 *
## CityVejle    -0.2760     0.1891  -1.459   0.1444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 130.999  on 23  degrees of freedom
## Residual deviance:  23.638  on 15  degrees of freedom
## AIC: 137.74
##
## Number of Fisher Scoring iterations: 5
```

Comments: Comparing to the Poisson GLM fit above (model1), we found that all the estimates of coefficient, standard errors are very similar in both model1 and model2, and the significance levels for each factors in both models are the same.