Question 2

Joseph Froelicher

4/9/2021

```
knitr::opts_chunk$set(echo = TRUE, include = TRUE)
```

Part A

```
skincancer = read.csv("skincancer.csv", header = TRUE)

for (i in 1:dim(skincancer)[1]){
  if (skincancer$city[i] == 'Dallas') {
    skincancer$city[i] = 1
  } else {
    skincancer$city[i] = 0
  }
}

model1 = glm(cases ~ city + offset(log(py1000)) + age.group, family = "poisson", data = skincancer)
table = summary(model1)$coefficients

rownames(table) = c(
  'Minneappolis (B0)', 'Dallas (B1)', '25-34', '35-44', '45-54', '55-64', '65-74', '75-84', '85+'
)

colnames(table) = c('Estimate', 'Std. Error', 'Z-value', 'p-value')
table
```

```
##                     Estimate Std. Error    Z-value      p-value
## Minneappolis (B0) -3.4495873 0.23842183 -14.468421 1.918265e-47
## Dallas (B1)        0.7868014 0.05142561  15.299797 7.669558e-53
## 25-34              1.3045827 0.27346959   4.770486 1.837824e-06
## 35-44              2.5602417 0.24953600  10.260009 1.066966e-24
## 45-54              3.3304993 0.24266935  13.724434 7.249350e-43
## 55-64              3.8266235 0.24134943  15.855117 1.295860e-56
## 65-74              4.3570039 0.24047073  18.118645 2.271189e-73
## 75-84              4.7971321 0.24138453  19.873404 6.915142e-88
## 85+                4.8987949 0.25493186  19.216096 2.714437e-82
```

Part B

- H_0: $e^{\beta_{Dallas}} = 2$ which means $\beta_{Dallas} = \log(2)$
- H_A: $e^{\beta_{Dallas}} \neq 2$
- Test statistic:

The Wald test statistic is given by:

$$Z^2 = \frac{(\hat{\beta}_{Dallas} - \beta_{H_0})^2}{Var(\hat{\beta}_{Dallas})} \sim \chi^2_1$$

$$Z^2 = \frac{(0.787 - \log(2))^2}{0.0514^2} \sim \chi^2_1$$

```
Z = (coef(model1)[2] - log(2))/sqrt(vcov(model1)[2,2])
p = pchisq(Z^2, df = 1, lower.tail = FALSE)
```

We conclude that the there is not evidence to suggest that the age-adjusted non-melanoma skin cancer rates in Dallas for women was *not* double that of women in Minneapolis (p = 0.069).

Part C

- H_0: $\beta_{25to34} = 0$ (15 to 24 is the reference group)
- H_A: $\beta_{25to34} \neq 0$
- Test statistic and Null distribution

The test statistic reported by R is the Wald test statistic, Z, which follows a standard normal distribution under the null hypothesis.

$$Z = 4.77 \sim N(0, 1)$$

We conclude that there is significant evidence to suggest that rates in the 25 to 34 age group are not equal to those in the 15 to 24 age group (p < 0.001).

Part D

```
r = exp(sum(c(1, 0, 0, 0, 1, 0, 0, 0, 0) * coef(model1)))
```

The estimated rate of non-melanoma skin cancer in the 45 to 54 year age group in women is 0.888.