Question 3

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)
model2 = glm(cases ~ city + offset(log(py1000)) + age.group + city * age.group, family = "poisson", data

a = anova(model1, model2, test = 'LRT')
b = a$`Pr(>Chi)`[2]
```

- H_0: $\beta_{agegroupxcity} = 0$
- H_A: $\beta_{agegroupxcity} \neq 0$

There is not evidence to suggest that the rate ratios for non-melanoma skin cancer does not depend on on age group (p = 0.080717).

Part B

```
model1$deviance # test statistic
```

```
## [1] 12.6645
```

```
model1$df.residual # degrees of freedom
```

```
## [1] 7
```

In this case, since there are only two covariates, the interaction model is the saturated model (the model containing all possible covariates). The residual deviance of the model with both main effect terms gives the difference in deviance between the saturated model and the model without the interaction (which is the test statistic for the LRT test we just performed).