HW10_852_EpiStats

2022-12-01

```
skin <- read.csv("/home/elkip/Datasets/skinca.csv")
head(skin)
```

```
##       city sexf rural population cases
## 1   Boston    0     0      284.1   250
## 2   Boston    1     0      333.5   273
## 3 Worchest    0     0       95.9    57
## 4 Worchest    1     0       85.1    48
## 5    Adams    0     1        4.5     5
## 6    Adams    1     1        4.0     1
```

Question 1

Introduction

The investigation is observing the relationship between rural vs urban setting and skin cancer using a Poisson regression. Data was collected from the Massachusetts cancer registry and the Massachusetts Census to get data on the incidence of skin cancer over a 4 year peroid by town and population information.

The variables in this data set are

1. town/city name;
2. sex, coded 1 for females, 0 for males;
3. an indicator coded 1 for rural towns, 0 for urban towns;
4. the sex-specific town population (population of females for lines relating to females, population of males for lines relating to males), in 1,000 of people;
5. the sex-specific number of cancers from the town.

1a. Preliminary analysis

- Find the total number of cancers and the total population size for the rural towns and urban towns, ignoring gender.
- Calculate the 4-year cancer rate per 1000 population for urban towns and for rural towns
- Give a confidence interval for each of these rates
- Calculate the incidence rate ratio comparing the cancer rates for rural vs urban towns (no CI)

```
# Total number of cancers and population
rural = skin[which(skin$rural == 1),]
urban = skin[which(skin$rural == 0),]
tot_cancer.rural <- sum(rural$cases)
tot_pop.rural <- sum(rural$population)*1000
tot_cancer.urban <- sum(urban$cases)
tot_pop.urban <- sum(urban$population)*1000
print(paste("Total cancers rural: ", tot_cancer.rural))
```

```
## [1] "Total cancers rural:  39"
```

```
print(paste("Total population rural: ", tot_pop.rural))
```

```
## [1] "Total population rural:  25100"
```

```
print(paste("Total cancers urban: ", tot_cancer.urban))
```

```
## [1] "Total cancers urban:  628"
```

```
print(paste("Total population urban: ", tot_pop.urban))
```

```
## [1] "Total population urban:  798600"
```

```
# Four Year cancer rate per 1000 population (d)
rate.rural <- (tot_cancer.rural / tot_pop.rural) * 1000
rate.urban <- (tot_cancer.urban / tot_pop.urban) * 1000
print(paste("4 Year Cancer Rate Rural: ", rate.rural, " per 1000 persons"))
```

```
## [1] "4 Year Cancer Rate Rural:  1.55378486055777  per 1000 persons"
```

```
print(paste("4 Year Cancer Rate Urban: ", rate.urban, " per 1000 persons"))
```

```
## [1] "4 Year Cancer Rate Urban:  0.786376158276985  per 1000 persons"
```

```
# CI For Rate
# (1000 / n) (d +/- (1.96 x square root of d)
se.rural <- sqrt(tot_cancer.rural)
se.urban <- sqrt(tot_cancer.urban)
rate_ci_u.rural <- (tot_cancer.rural + 1.96*se.rural)*(1000/tot_pop.rural)
rate_ci_l.rural <- (tot_cancer.rural - 1.96*se.rural)*(1000/tot_pop.rural)
rate_ci_u.urban <- (tot_cancer.urban + 1.96*se.urban)*(1000/tot_pop.urban)
rate_ci_l.urban <- (tot_cancer.urban - 1.96*se.urban)*(1000/tot_pop.urban)
print(paste("4 Year Cancer Rate 95% CI Rural: ", rate_ci_l.rural, ", ", rate_ci_u.rural))
```

```
## [1] "4 Year Cancer Rate 95% CI Rural:  1.0661276463402 ,  2.04144207477533"
```

```
print(paste("4 Year Cancer Rate 95% CI Urban: ", rate_ci_l.urban, ", ", rate_ci_u.urban))
```

```
## [1] "4 Year Cancer Rate 95% CI Urban:  0.724871701455453 ,  0.847880615098517"
```

```
# Incidence rates rural vs urban
rate.ratio <- rate.rural / rate.urban
print(paste("Incidence rate ratio: ", rate.ratio))
```

```
## [1] "Incidence rate ratio:  1.9758799198112"
```

1b. Poisson regression crude

Fit a Poisson regression model predicting the rate of skin cancer cases from the rural / urban variable, using the log population as the 'offset' (so you will be analyzing cases per 1000 population)

- What is the rate ratio for skin cancer for rural vs urban people, and how does the rate ratio from this Poisson model compare to the incidence rate ratio you calculated in 1a?
- From the Poisson model give a 95% CI for this rate ratio.
- Are the rates of skin cancer significantly different between rural towns and urban towns?

```
# Poisson reg
poi.crude <- glm(cases ~ rural, family = poisson(link="log"), offset = log(population), data = skin)
summary(poi.crude)
```

```
##
## Call:
## glm(formula = cases ~ rural, family = poisson(link = "log"),
##     data = skin, offset = log(population))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.60313  -1.14985  -0.04372   0.76502   1.74902
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.2403     0.0399  -6.022 1.72e-09 ***
## rural         0.6810     0.1650   4.127 3.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 40.361  on 11  degrees of freedom
## Residual deviance: 26.333  on 10  degrees of freedom
## AIC: 83.251
##
## Number of Fisher Scoring iterations: 4
```

```
# Rate ratio and 95% CI
exp(cbind(IRR = coef(poi.crude), confint.default(poi.crude)))
```

```
##                   IRR     2.5 %    97.5 %
## (Intercept) 0.7863762 0.7272165 0.8503486
## rural       1.9758799 1.4298626 2.7304032
```

The rate ratio for rural vs urban is 1.98 (95% CI: 1.43, 2.73), this is the same as the IRR I calculated above. It would appear that the rate of cancer is significantly higher in those who live in rural areas than urban areas.

1c. Poisson regression adjusted for sex

Fit a Poisson model predicting the risk of skin cancer for rural vs. urban people, adjusting for sex.

- What is the rate ratio for skin cancer for rural vs. urban people after adjusting for sex?
- From the Poisson model, give a 95% CI for this rate ratio. Is sex a confounder here? Provide evidence to support your conclusion.

```
# Adjusted Poisson model
poi.adj <- glm(cases ~ rural + sexf, family = poisson(link="log"), offset = log(population), data = skin
summary(poi.adj)
```

```
##
## Call:
## glm(formula = cases ~ rural + sexf, family = poisson(link = "log"),
##     data = skin, offset = log(population))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.54926  -1.20595  -0.01998   1.04474   1.81927
##
## Coefficients:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.21216    0.05618  -3.776 0.000159 ***
## rural        0.68024    0.16502   4.122 3.76e-05 ***
## sexf        -0.05443    0.07746  -0.703 0.482223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 40.361  on 11  degrees of freedom
## Residual deviance: 25.839  on  9  degrees of freedom
## AIC: 84.758
##
## Number of Fisher Scoring iterations: 4
```

```
# IRR and 95% CI
exp(cbind(IRR = coef(poi.adj), confint.default(poi.adj)))
```

```
##                   IRR     2.5 %   97.5 %
## (Intercept) 0.8088363 0.7245007 0.902989
## rural       1.9743515 1.4287460 2.728311
## sexf        0.9470237 0.8136350 1.102280
```

```
# Test for confounding in beta (crude - adjusted) / crude
print(paste("Percent diff in rate: ", abs(.6810 - .6802) / .6810))
```

```
## [1] "Percent diff in rate:  0.00117474302496332"
```

The rate ratio for skin cancer in rural vs urban settings adjusted for sex is .6802 (95% CI: 1.43, 2.70). The difference in the crude and adjusted rate is .11%, which is less than 10% so we fail to identify sex as a confounder in the relationship between rural vs urban setting and cancer risk.

Conclusion

Question 2

```
hypo <- read.csv("/home/elkip/Datasets/hypoglycemia.episodes.csv")
head(hypo)
```

```
##   personid hypoepisodes days female age ownmonitor yearssincedx
## 1        1            7    7      0  40          1           34
## 2        2            5    7      0  31          1           25
## 3        3            8    7      1  28          0           20
## 4        4            4    7      0  25          1           14
## 5        5           12    7      1  60          0           45
## 6        6            5    7      0  25          1           12
```

Introduction

127 patients at a clinical practice with type I diabetes were monitored with a continuous glucose monitor (CGM) all day for 7 days. The aim is to study the number of hypoglycemic episodes which occur (blood glucose $<$ 70mg/dL).

Other factors recorded:

- total length of time CGM was function
- gender

- age
- years since diagnosis
- if the patient already owned a CGM

2a. Poisson Regression

```
hypo.poi <- glm(hypoepisodes ~ female + age + yearssincedx + ownmonitor,
                family = poisson(link="log"), offset = log(days), data = hypo)
summary(hypo.poi)
```

```
##
## Call:
## glm(formula = hypoepisodes ~ female + age + yearssincedx + ownmonitor,
##     family = poisson(link = "log"), data = hypo, offset = log(days))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.21317  -0.71304  -0.07357   0.59398   2.65867
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.131901   0.199473  -0.661   0.5085
## female        0.009471   0.084495   0.112   0.9107
## age           0.015628   0.010944   1.428   0.1533
## yearssincedx -0.010213   0.010237  -0.998   0.3185
## ownmonitor   -0.208521   0.084988  -2.454   0.0141 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 129.77  on 126  degrees of freedom
## Residual deviance: 111.94  on 122  degrees of freedom
## AIC: 607.12
##
## Number of Fisher Scoring iterations: 4
```

<insert interpretation here>

2b. Patients with high number of hypoglycemic episodes?

We observe the intercepts with the most negative influence

2c. Which variables are significantly associated with lower numbers of hypoglycemic episodes?

2d. Do gender, age, and years since diagnosis as a group confound the association between hypoglycemic episodes and owning CGM monitors?

Conclusion