Class Activity#15

Sunny Lee

2021-04-09

Poisson Log-Linear Regression: Detecting Cancer Clusters. Calculating the Cancer Rate per Person-Year

1) If the 325 homes in Randolph have an average of 3.5 people in each home (1138 people), then, based on the 67 reported cases, what proportion of people reported a cancer diagnosis?

```
prop <- 67/1138
prop
```

```
## [1] 0.05887522
```

Since there are 67 cases out of 1138 people, the proportion of people who reported a cancer diagnosis is $\frac{67}{1138}$ or 5.89%.

2) If the 325 homes in Randolph have an average of 3.5 people in each home (1138 people) and if we assume people have lived in the community for an average of 25 years each, then what is the total number of person-years at risk in the community?

```
person_year <- 25 * 1138
```

To calculate the person-years, we multiply the average time people lived in the community by the amount of people and get get 28450 person years.

3) Given your answer to Question 2 and the fact that 67 cancer cases were reported, what is the observed cancer rate (number of cases per person-year) in Randolph? What is the cancer incidence rate?

```
rate <- 67/person_year
randolph_incident_rate <- 235.5/100000
rate
```

```
## [1] 0.002355009
```

```
randolph_incident_rate
```

```
## [1] 0.002355
```

The observed cancer rate is $\frac{67}{28450}$ or .2355%, thus our incident rate is 235.5 cancer incidents per 100,000 people.

4) Using the national cancer incidence rate of 326 cases per 100,000 people in a year and the perosn-years of exposure for Randolph in Question 2, calculate the expected count of cancer cases. Does the observed value of 67 cases out of 325 homes seem unusually high to you?

```
univ_incidence_rate <- 326/100000
expected_count <- univ_incidence_rate * person_year
expected_count
```

```
## [1] 92.747
```

Since Randolph observed cancer is much lower than what we expect, the observed value of 67 is not unusually high.

5) Day and colleagues actually made the assumption that residents lived in Randolph for an average of 12.5 years each. Using this assumption, does the observed value of 67 cases out of 325 homes seem unusually high to you?

```
person_year <- 12.5*1138
rate <- 67/person_year
expected_count <- univ_incidence_rate * person_year
expected_count
```

```
## [1] 46.3735
```

Changing our average years lived in Randolph, we now observe an expected count of 46.3735, which is still lower than 92.747. Thus, our cases does not seem unusually high.