

Exercise day 4 – Poisson regression

In this exercise, we examine whether mosquito breeding inside and outside houses is associated with the weekly number of dengue fever cases in the district of Manjung in the state of Perak, Malaysia, over the years 2003 to 2006. Manjung is one of the 9 districts of Perak and is highly endemic for dengue. The outcome is the number of dengue fever cases reported each week; the exposures are the numbers of Aedes breeding sites found inside and outside houses each week; rainfall, humidity and mean temperature are treated as confounders. The data used in this exercise are based on real surveillance data but have been modified for teaching purposes.

Variables in the data file Manjung.csv :

year: 2003, 2004, 2005, 2006

epidtime: Week 1, week 2, ..., week 52, week 53, ..., week 208

cases: Number of cases of dengue observed weekly

inside: Number of breeding mosquitoes Aedes found weekly inside the house

outside: Number of breeding mosquitoes Aedes found weekly in the environment outside the house

meantemp: The mean temperature for the week

humidity: The average weekly humidity (in %)

rainfall: The average weekly rainfall (in mm)

Load packages and import the data:

```
library(tidyverse)
library(MASS)

setwd("___directory path___")
manjung <- read_csv("Manjung.csv")
glimpse(manjung)
summary(manjung)
```

Part A: Exploratory data analysis

Before fitting any model, describe the outcome (weekly dengue cases), the two exposures (breeding sites inside and outside houses), and the three confounders (rainfall, humidity, and mean temperature) identified by DAG.

1) Outcome distribution:

- a) What is the distribution of the number of dengue cases?

```
ggplot(manjung, aes(x = cases)) +
  geom_histogram(color = "black") +
  labs(
    x = "Dengue cases",
    y = "Frequency"
  ) +
  theme_minimal()
```

- b) How are the numbers of dengue cases distributed weekly?

```
manjung |>
  summarise(
    sum_cases = sum(cases, na.rm = TRUE),
```

```

    .by = epidweek
  ) |>
  ggplot(aes(x = epidweek, y = sum_cases)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_x_continuous(breaks = seq(1, 52, by = 4)) +
  labs(
    x = "Week",
    y = "Count of dengue cases",
  ) +
  theme_minimal()

```

- 2) Describe the exposures. What are the average weekly numbers of breeding mosquitoes found inside and outside houses?

```

manjung |>
  dplyr::select(inside, outside) |>
  tidyr::pivot_longer(
    cols = everything(),
    names_to = "Variable",
    values_to = "value"
  ) |>
  dplyr::summarise(
    Mean = mean(value, na.rm = TRUE),
    SD = sd(value, na.rm = TRUE),
    Median = median(value, na.rm = TRUE),
    IQR = IQR(value, na.rm = TRUE),
    Min = min(value, na.rm = TRUE),
    Max = max(value, na.rm = TRUE),
    .by = Variable
  ) |>
  dplyr::mutate(
    dplyr::across(where(is.numeric), ~ round(.x, 2))
  )

```

- 3) Describe the confounders (rainfall, humidity and mean temperature).

```

manjung |>
  dplyr::select(meantemp, humidity, rainfall) |>
  tidyr::pivot_longer(
    cols = everything(),
    names_to = "Variable",
    values_to = "value"
  ) |>
  dplyr::summarise(
    Mean = mean(value, na.rm = TRUE),
    SD = sd(value, na.rm = TRUE),
    Median = median(value, na.rm = TRUE),
    IQR = IQR(value, na.rm = TRUE),
    Min = min(value, na.rm = TRUE),
    Max = max(value, na.rm = TRUE),
    .by = Variable
  ) |>
  dplyr::mutate(
    dplyr::across(where(is.numeric), ~ round(.x, 2))
  )

```

Part B: Poisson regression

Research question: Are mosquitoes breeding inside and outside houses associated with the weekly number of dengue cases?

- 5) Since the cases of dengue fever are count data, it is reasonable to fit the number of cases by a generalized linear model (R's function glm). Perform univariate Poisson regressions for inside, outside, meantemp, humidity and rainfall and report incidence rate ratios.

```
irr <- function(model) {  
  out <- cbind(IRR = exp(coef(model)), exp(confint.default(model)),  
              p = summary(model)$coefficients[, 4])  
  round(out, 4)  
}
```

```
m_inside <- glm(cases ~ inside, family = poisson(link = "log"), data = manjung)  
m_outside <- glm(cases ~ outside, family = poisson(link = "log"), data = manjung)
```

```
irr(m_inside)  
irr(m_outside)
```

- a. Explain the outputs. What does the IRR mean?

- 6) Fit the multiple Poisson regression for the two exposures, inside and outside breeding sites, adjusting for the confounders identified from the DAG: rainfall, humidity and mean temperature. Interpret the results.

```
m_multi1 <- glm(cases ~ inside + meantemp + humidity + rainfall, family = poisson(link  
= "log"), data = manjung)  
irr(m_multi1)
```

```
m_multi2 <- glm(cases ~ outside + meantemp + humidity + rainfall, family = poisson(link  
= "log"), data = manjung)  
irr(m_multi2)
```