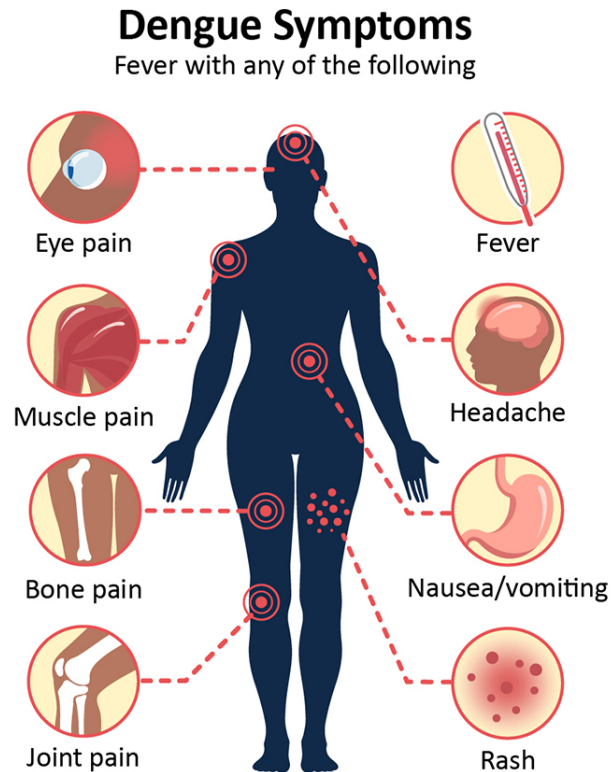


Introduction to Logistic Regression

Motivating example: Dengue fever

Dengue fever: a mosquito-borne viral disease affecting 400 million people a year



Motivating example: Dengue data

Data: Data on 5720 Vietnamese children, admitted to the hospital with possible dengue fever. Variables include:

- + *Sex*: patient's sex (female or male)
- + *Age*: patient's age (in years)
- + *WBC*: white blood cell count
- + *PLT*: platelet count
- + other diagnostic variables...
- + *Dengue*: whether the patient has dengue (0 = no, 1 = yes)

Research questions:

- + How well can we predict whether a patient has dengue?
- + Which diagnostic measurements are most useful?
- + Is there a significant relationship between WBC and dengue?

Research questions

- + How well can we predict whether a patient has dengue?
- + Which diagnostic measurements are most useful?
- + Is there a significant relationship between WBC and dengue?

How can I answer each of these questions? Discuss with a neighbor for 2 minutes, then we will discuss as a class.

Fitting a model: initial attempt

What if we try a linear regression model?

Y_i = dengue status of i th patient

$$Y_i = \beta_0 + \beta_1 WBC_i + \varepsilon_i \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

What are some potential issues with this linear regression model? Go to <https://pollev.com/ciaranevans637> to respond.

Second attempt

Let's rewrite the linear regression model:

Second attempt

$$Y_i \sim \text{Bernoulli}(p_i) \quad p_i = \mathbb{P}(Y_i = 1 | WBC_i)$$

$$p_i = \beta_0 + \beta_1 WBC_i$$

Are there still any potential issues with this approach?

Don't fit linear regression with a binary response