Logistic regression interpretation

Recap: logistic regression

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)

Goal: Build a model to predict dengue status

Recap: logistic regression

dengue status (0 = no dengue, 1 = dengue)

(random component) $\quad Y_i \sim Bernoulli(\pi_i)$

π_i: $P(Y_i = 1)$

(systematic component) $$\underbrace{\log\left(\overbrace{\frac{\pi_i}{1-\pi_i}}^{\text{odds}}\right)}_{\text{log odds aka logit}} = \beta_0 + \beta_1\, Age_i$$

natural log

odds $= \dfrac{\text{prob.}}{1 - \text{prob.}}$

prob. $= \dfrac{\text{odds}}{1 + \text{odds}}$

$$\frac{\pi_i}{1-\pi_i} = e^{\beta_0 + \beta_1 Age_i}$$

$$\pi_i = \frac{e^{\beta_0 + \beta_1 Age_i}}{1 + e^{\beta_0 + \beta_1 Age_i}}$$

Class activity

https://sta214-s23.github.io/class_activities/ca_lecture_3.html

- Spend 5 minutes working in pairs on the questions
- We will then discuss as a class

Class activity

$$\log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = -2.45 + 0.22\ Age_i$$

What is the predicted probability of dengue for a 10 year old patient?

$$\frac{\hat{\pi}_i}{1-\hat{\pi}_i} = e^{-2.45+0.22(10)} = e^{-0.25} = 0.78$$

$$\hat{\pi}_i = \frac{e^{-0.25}}{1+e^{-0.25}} = 0.44$$

Similar question on Hw 1!

Class activity

$$\log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = -2.45 + 0.22\ Age_i$$

> Suppose we want to identify patients for whom the predicted probability of dengue is at least 0.5. What age range should we focus on?

want: $\hat{\pi}_i > 0.5 \quad \Rightarrow \quad \dfrac{\hat{\pi}_i}{1-\hat{\pi}_i} > 1 \quad \Rightarrow \quad \log\left(\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) > 0$

$\Rightarrow \quad -2.45 + 0.22 Age_i > 0$

$\Rightarrow \quad Age > 11.14$

Since Age is a whole # in the data, look at patients aged at least 12

Class activity

Increasing age by one year multiplies odds by $e^{0.22} = 1.25$

$$\log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = -2.45 + 0.22\ Age_i$$

> Compare the odds of dengue for a 12 year old patient to the odds of dengue for an 11 year old patient. What do you notice?

$$\frac{\text{odds when age}=12}{\text{odds when age}=11} = 1.25 = \frac{\text{odds when age}=11}{\text{odds when age}=10}$$

In general:

$$\frac{\text{odds when age}=x+1}{\text{odds when age}=x} = \frac{e^{-2.45+0.22(x+1)}}{e^{-2.45+0.22x}} = e^{0.22}$$

$$= 1.25$$

Interpretation

$$\log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = -2.45 + 0.22\ Age_i$$

slope:

odds: A one-year increase in Age is associated with an increase in odds of dengue by a factor of $e^{0.22} = 1.25$

log odds: A one-year increase in age is associated with an increase in log odds of 0.22

Intercept:

log odds: when age = 0, the estimate log odds of dengue is −2.45

odds: when age = 0, predicted odds of dengue are $e^{-2.45} = 0.086$

$0.34 + 0.15 Age_i - 0.31 \underbrace{AbdoPain_i}_{\substack{=1 \text{ if pain} \\ =0 \text{ if no pain}}}$

Adding more variables

The presence of abdominal pain is associated w/ a decrease in odds by a factor of $e^{-0.31}$, holding Age fixed

Now let's add WBC as a variable to the model:

```
m2 <- glm(Dengue ~ Age + WBC, data = dengue,
          family = binomial)
summary(m2)
```

$$\log\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = 0.34 + 0.15\ Age_i - 0.31 WBC_i$$

> How should I interpret each coefficient in the fitted model?

0.15 : A one-year increase in age is associated with an increase in odds of dengue by a factor of $e^{0.15}$, holding WBC fixed

-0.31 : A one-unit increase in WBC is associated with a decrease in odds of dengue by a factor of $e^{-0.31}$, holding Age fixed