

Quiz2

key

1/28/2022

1. Write out the mathematical notation for standard linear regression model and note the three components of GLMs.

$$y \sim N(\mu, \sigma^2) \quad (1)$$

$$\mu = 1(X\beta) \quad (2)$$

1. the probability distribution is Normal
2. the link function is the identify 1()
3. the linear combination is $X\beta$

2. Consider a linear regression model with one continuous variable and one categorical variable with three levels (such as high, medium, low) with an interaction term.

a. Write out the mathematical notation for this model and clearly articulate the interpretation for all of the parameters. You can assume the continuous variable has been centered.

$$y = \beta_0 + \beta_1 x_{medium} + \beta_2 x_{low} + \beta_3 x_2 + \beta_4 x_2 x_{medium} + \beta_5 x_2 x_{low} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

- β_0 : (intecept for **high** group), expected response for the **high** group when $x_2 = 0$
- β_1 : expected difference from **high** group for a unit with the **medium** level and $x_2 = 0$
- β_2 : expected difference from **high** group for a unit with the **low** level and $x_2 = 0$
- β_3 : (slope for the **high** group), predicted increase in response between units in the high group where x_2 differs by 1
- β_4 : expected difference in slope from **high** group reference case (β_3) to **medium** slope
- β_5 : expected difference in slope from **high** group reference case (β_3) to **low** slope

b. Create a figure of the fitted model and use annotation (`annotate`) to show how the model parameters create the slope and intercepts. Code and figure provided below

```
library(tidyverse)
beta0 <- 0
beta1 <- 3
beta2 <- -1
beta3 <- 1
beta4 <- -2
beta5 <- 1

x_seq <- 0:2

line_data <- tibble(x = rep(x_seq, 3),
                    y = c(beta0 + beta3 * x_seq,
                          beta0 + beta1 + (beta3 + beta4) * x_seq,
                          beta0 + beta2 + (beta3 + beta5) * x_seq),
```

```

group = rep(c('high', 'medium', 'low'), each = length(x_seq))

line_data %>% ggplot(aes(x = x, y = y, color = group)) +
  geom_line() + theme_bw() +
  annotate('text', x=.2, y = .1 ,
          label = expression(beta[0]), color = 'red') +
  annotate('text', x=.2, y = 2.8 ,
          label = expression(beta[0] + beta[1]), color = 'blue') +
  annotate('text', x=.25, y = -0.8 ,
          label = expression(beta[0] + beta[2]), color = 'forestgreen') +
  annotate('text', x=0.5, y = .8 ,
          label = expression(beta[3]), angle = 27, color = 'red') +
  annotate('text', x=1, y = 2.1 ,
          label = expression(beta[3] + beta[4]), angle = 340, color = 'blue') +
  annotate('text', x=.75, y = .3 ,
          label = expression(slope = beta[3] + beta[4]), angle = 45, color = 'forestgreen')

```

