MV Regression: Indicators and Interaction Terms

Consider a variable that takes value 0 or 1 that want to add to a regression model

Ex: Sex of penguins → 1.7% of world pop^n estimated to be intersex (people, not penguins)

Example: Male 0 or 1

$$y_i = b_0 + b_1 x + b_2 I_{Male_i} + \varepsilon_i$$

I_{Male_i}: value 1 for males, 0 otherwise

y: weight; x: bill length; ε_i: error $N(0,\sigma)$

This is an indicator variable

For a male penguin the equation becomes

$$y_i = b_0 + b_1 X_i + b_2 \times 1 + \varepsilon_i \Rightarrow y_i = (b_0 + b_2) + b_1 X_i + \varepsilon_i$$

For a female (non-male) the equation becomes

$$y_i = b_0 + b_1 X_i + b_2 \times 0 + \varepsilon_i \Rightarrow y_i = b_0 + b_1 X_i + \varepsilon_i$$

Essentially, male penguins have a different intercept in the equation.

What if we want to use more than 2 categories?

Species: Adelie, Gentoo, Chinstrap

Make 3 indicators $\quad y_i = b_0 + b_1 X_i + b_2 I_{A_i} + b_3 I_{G_i} + b_4 I_{C_i} + \varepsilon_i$

Adelie | Gentoo | Chinstrap

$y_i = (b_0 + b_3) + b_1 X_i + \varepsilon_i$ | $y_i = (b_0 + b_4) + b_1 X_i + \varepsilon_i$

Adelie: $Y_i = (b_0 + b_2) + b_1 X_i + \varepsilon_i$

Gentoo: $Y_i = (b_0 + b_3) + b_1 X_i + \varepsilon_i$

Chinstrap: $Y_i = (b_0 + b_4) + b_1 X_i + \varepsilon_i$

But.... we dont need 4 different parameters to describe 3 different intercepts.

We choose a baseline (Adelie) and remove the indicator for the baseline. [We roll the old b_2 into the intercept term)

$$Y_i = b_0 + b_1 X_i + b_2 I_{G_i} + b_3 I_{C_i} + \varepsilon_i$$

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

What if we want different slopes for each species?

We can use interaction terms. (value of 0 or 1)

Equation becomes $Y_i = b_0 + b_1 I_{G_i} + b_2 I_{C_i} + b_3 X_i + b_4 X_i \cdot I_G + b_5 X_i \cdot I_C + \varepsilon_i$

3 equations

A: $Y_i = b_0 + b_3 X_i + \varepsilon_i$

G: $Y_i = (b_0 + b_1) + (b_3 + b_4) \cdot X_i + \varepsilon_i$ (b_4: represents diff in slope for Gentoo penguins compared to Adelie.)

C: $Y_i = (b_0 + b_2) + (b_3 + b_5) \cdot X_i + \varepsilon_i$ (b_5: diff for chinstrap compared to Adelie.)

We can also have interactions between continuous variables

Ex bike rentals we might wonder how temp and humidity affect rentals on their own, [illegible] they interact together.

and humidity affect rentals [illegible]

but also how they interact together.

$$Y_i = b_0 + b_1 \text{Temp} + b_2 \text{Humidity} + b_3 \text{Temp} * \text{Humidity} + \varepsilon_i$$