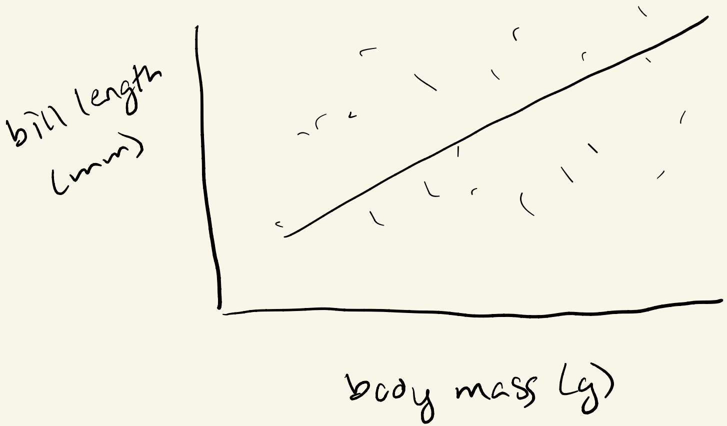
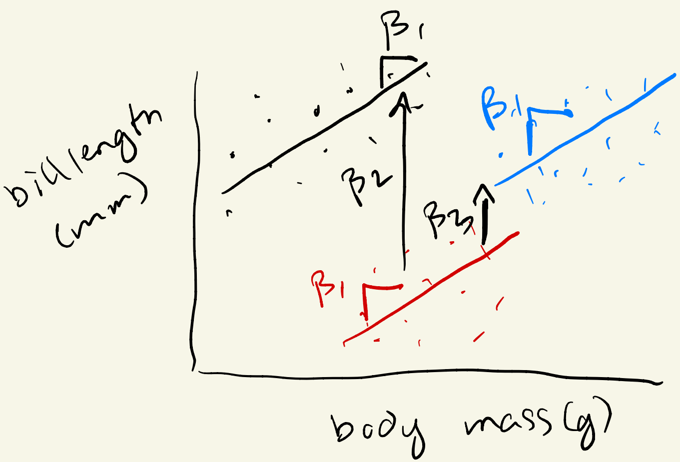
ZIP models

- Challenge 8 posted on course website
- Project 2 released later today
- HW 6 due Monday
- Exam 2 released next Friday
  - Poisson, EDMs, quasi-Poisson, negative binomial, ZIP models



$Y_i =$ bill length

$$Y_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$



● = Adelie
● = Gentoo
● = Chinstrap

$$Y_i = \beta_0 + \beta_1 \text{Mass}_i + \beta_2 \text{Chinstrap}_i + \beta_3 \text{Gentoo}_i + \varepsilon_i$$

For Adelie:

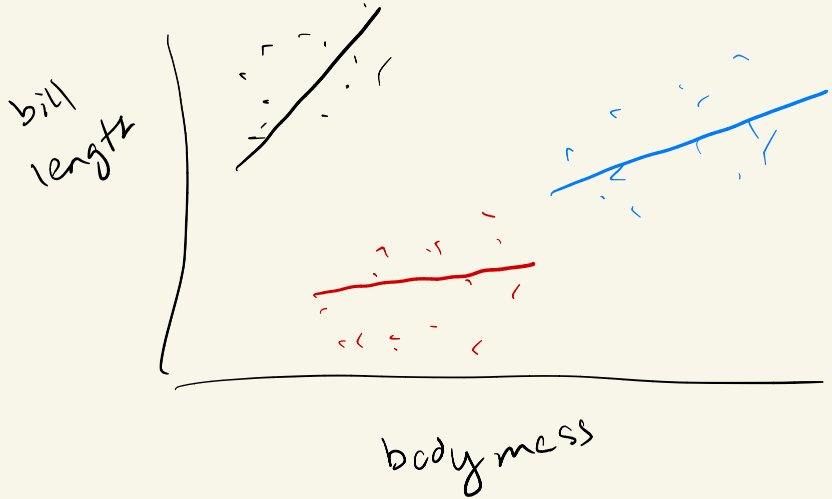
$$Y_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$

Chinstrap:

$$Y_i = (\beta_0 + \beta_2) + \beta_1 \text{Mass}_i + \varepsilon_i$$

Gentoo:

$$Y_i = (\beta_0 + \beta_3) + \beta_1 \text{Mass} + \varepsilon_i$$



= Adelie
= Gentoo
= Chinstrap

$$Y_i = \beta_0 + \beta_1 Mass_i + \beta_2 Chinstrap_i + \beta_3 Gentoo_i + \beta_4 \cdot Mass_i \cdot Chinstrap_i + \beta_5 \cdot Mass_i \cdot Gentoo_i + \varepsilon_i$$

β_1 = slope for Adelies

β_4 = change in slope for Chinstrap vs. Adelies

β_5 = change in slope for Gentoo vs. Adelies

H_0: all species have same slope

$\beta_4 = \beta_5 = 0$

For Adelie:

$$Y_i = \beta_0 + \beta_1 Mass_i + \varepsilon_i$$

Chinstrap:

$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_4) Mass_i + \varepsilon_i$$

Gentoo:

$$Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) Mass_i + \varepsilon_i$$

Recap: Zero-inflated Poisson (ZIP) model

$$P(Y_i = y) = \begin{cases} e^{-\lambda_i}(1-\alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1-\alpha_i) & y > 0 \end{cases}$$

mixture of a point mass at 0 and a Poisson

where

α_i = mixing proportion

logistic component

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma_0 + \gamma_1 FirstYear_i + \gamma_2 OffCampus_i + \gamma_3 Male_i$$

$$\log(\lambda_i) = \beta_0 + \beta_1 FirstYear_i + \beta_2 OffCampus_i + \beta_3 Male_i$$

Poisson component

$\alpha_i = P(Z_i = 1)$

$Y_i \mid (Z_i = 1) \equiv 0$ (point mass)

$Y_i \mid (Z_i = 0) \sim \text{Poisson}(\lambda_i)$

Fitted model

$$P(Y_i = y) = \begin{cases} e^{-\lambda_i}(1-\alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1-\alpha_i) & y > 0 \end{cases}$$

$$\log\left(\frac{\widehat{\alpha}_i}{1-\widehat{\alpha}_i}\right) = -0.40 + 0.89 FirstYear_i - 1.69 OffCampus_i - 0.07 Male_i$$

$$\log(\widehat{\lambda}_i) = 0.80 - 0.16 FirstYear_i + 0.37 OffCampus_i + 0.98 Male_i$$

Warm up: Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_29.html

Class activity

$Y_i = $ # drinks consumed

$Z_i = \mathbb{1}\{$ student i never drinks $\}$

$$P(Y_i = y) = \begin{cases} e^{-\lambda_i}(1 - \alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1 - \alpha_i) & y > 0 \end{cases}$$

$$\log\left(\frac{\widehat{\alpha}_i}{1 - \widehat{\alpha}_i}\right) = \underline{-0.40 + 0.89 FirstYear_i - 1.69 OffCampus_i - 0.07 Male_i}$$

$$\log(\widehat{\lambda}_i) = 0.80 - 0.16 FirstYear_i + 0.37 OffCampus_i + 0.98 Male_i$$

> What is the estimated probability that a male first year student who lives on campus *never* drinks?

$P(Z_i = 1) = \alpha_i$

$$\hat{\alpha}_i = \frac{\exp\{-0.4 + 0.89 - 0.07\}}{1 + \exp\{-0.4 + 0.89 - 0.07\}} = 0.60$$

Class activity

$$P(Y_i = y) = \begin{cases} e^{-\lambda_i}(1-\alpha_i) + \alpha_i & y = 0 \\ \dfrac{e^{-\lambda_i}\lambda_i^y}{y!}(1-\alpha_i) & y > 0 \end{cases}$$

$$\log\left(\frac{\widehat{\alpha}_i}{1-\widehat{\alpha}_i}\right) = -0.40 + 0.89FirstYear_i - 1.69OffCampus_i - 0.07Male_i$$

$$\log(\widehat{\lambda}_i) = 0.80 - 0.16FirstYear_i + 0.37OffCampus_i + 0.98Male_i$$

> What is the estimated probability that a male first year student who lives on campus consumed 3 drinks last weekend?

$$\hat{P}(Y_i = 3) = \frac{e^{-\hat{\lambda}_i}\hat{\lambda}_i^3}{3!}(1-\hat{\alpha}_i)$$

$$\hat{\alpha}_i = 0.60$$

$$\hat{\lambda}_i = \exp\{0.8 - 0.16 + 0.98\} = 5.05$$

$$\hat{P}(Y_i = 3) = 0.055$$

$$P(Y_i, Z_i \mid \gamma, \beta) = \underbrace{P(Z_i \mid \gamma, \beta)}_{P(Z_i \mid \gamma)} \underbrace{P(Y_i \mid Z_i, \gamma, \beta)}_{P(Y_i \mid Z_i, \beta)}$$

$$P(A \cap B) = P(A)P(B \mid A)$$

Fitting ZIP models

Suppose we can actually observe latent variable Z_i

$Z_i \sim \text{Bernoulli}(\alpha_i)$ $\quad$ $Y_i \mid (Z_i = 0) \sim \text{Poisson}(\lambda_i)$

$$\log\left(\frac{\alpha_i}{1-\alpha_i}\right) = \gamma^T X_i \qquad \log(\lambda_i) = \beta^T X_i$$

$$L(\gamma, \beta) = \prod_{i=1}^n P(Y_i, Z_i \mid \gamma, \beta) = \prod_{i=1}^n P(Z_i \mid \gamma) P(Y_i \mid Z_i, \beta)$$

$$= \prod_{i=1}^n \alpha_i^{Z_i} (1-\alpha_i)^{1-Z_i} \left(\frac{e^{-\lambda_i} \lambda_i^{Y_i}}{Y_i!}\right)^{1-Z_i}$$

$$\Rightarrow \ell(\gamma, \beta) = \sum_{i=1}^n \left(Z_i \log(\alpha_i) + (1-Z_i)\log(1-\alpha_i)\right) + \sum_{i=1}^n (1-Z_i)(-\lambda_i + Y_i \log \lambda_i) - \sum_{i=1}^n (1-Z_i) \log(Y_i!)$$

$$\ell(\gamma) = \sum_{i=1}^n \left(Z_i \log(\alpha_i) + (1-Z_i)\log(1-\alpha_i)\right)$$

$$\ell(\beta) = \sum_{i=1}^n (1-Z_i)(-\lambda_i + Y_i \log \lambda_i)$$

<u>if</u> we knew Z_i, I can separately maximize $\ell(\gamma)$ and $\ell(\beta)$