What to do today (Feb 27)?

4. Analysis with Count Response (Chp 4)

4.1 Poisson Model for Counts (Chp 4.1)

4.1.1 Poisson distribution

4.1.2 Inference with Poisson distribution

4.2 Poisson Regression Analysis (Chp 4.2)

4.2.1 Introduction to Poisson regression models

4.2.2 Poisson regression analysis

4.2.3 Categorical explanatory variables

4.2.4 Poisson regression with contingency tables

4.3 Additional Topics on Count Responses (Chp 4.3-4)

4.1.1 Poisson distribution

What are commonly-used distn models for a r.v. Y, the count of random events?

- *Bionomial Distn.* $Y \sim B(n, \pi)$ with the pmf

$$P(Y = y) = p(y) = \begin{cases} \binom{n}{y} \pi^y (1-\pi)^{n-y}, & y = 0, 1, \ldots, n \\ 0, & otherwise \end{cases}$$

  e.g. the Bernoulli distn: $B(1, \pi)$.

- **Poisson Distn.** Y has a Poisson distribution, denoted by $Y \sim Poisson(\mu)$, if its pmf is

$$P(Y = y) = p(y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \ldots$$

  The distn is named after S.D. Poisson (1781-1840).

4.1.1 Poisson distribution

Remarks:

- The Poisson distn is especially good at modelling rare events.
- $P(X = 0) = e^{-\mu}$.
- R functions with Poisson distn: $dpois(), rpois(), qpois()$

Properties: For r.v. $Y \sim Poisson(\mu)$,

- $E(Y) = \mu$; $Var(X) = \mu$
- If $Y_1 \perp\!\!\!\perp Y_2$ and $Y_1 \sim Poisson(\mu_1)$, $Y_2 \sim Poisson(\mu_2)$,

$$Y_1 + Y_2 \sim Poisson(\mu_1 + \mu_2)$$

4.1.1 Poisson distribution

Recall that the binomial pmf may be difficult to calculate. It turns out the binomial distn can sometimes be approximated by a distn easier to calculate.

Proposition. $Bin(n, \theta) \approx Poisson(n\theta)$ if n is much larger than $n\theta$.
Rationale (without being rigorous): for a fixed y,

$$\left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y} = \frac{n^{(y)}}{y!} \frac{(n\theta)^y}{n^y} \left(1 - \frac{n\theta}{n}\right)^{n-y} \to \frac{\mu^y}{y!} e^{-\mu}$$

as $n \to \infty$ and $n\theta$ *fixed as* μ.

Proposition. $Poisson(\mu) \approx N(\mu, \sigma^2)$ for large μ with $\sigma^2 = \mu$.

4.1.2 Inference with Poisson distribution

Given a random sample $Y_1, \ldots, Y_n$ from $Poisson(\mu)$,

$$L(\mu|data) = \prod_{i=1}^{n} \frac{\mu^{y_i} e^{-\mu}}{y_i!} \propto \mu^{\sum y_i} e^{-n\mu}$$

$\Longrightarrow$ MLE of μ is $\hat{\mu} = \bar{Y}$.

- MLE of μ is the same as the sample mean.
- $Var(\hat{\mu}) = \mu/n \Longrightarrow \hat{Var}(\hat{\mu}) = \hat{\mu}/n$.

4.1.2 Inference with Poisson distribution

Given a random sample $Y_1, \ldots, Y_n$ from $Poisson(\mu)$, $100(1-\alpha)\%$ CIs for μ

- Wald CI:

$$\hat{\mu} \pm Z_{1-\alpha/2}\sqrt{\hat{\mu}/n}$$

- score-type: based on the score statistic $Z = \frac{\hat{\mu}-\mu_0}{\sqrt{\mu_0/n}}$ for $H_0 : \mu = \mu_0 \Longrightarrow$

$$\left(\hat{\mu} + \frac{Z_{1-\alpha/2}^2}{2n}\right) \pm Z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\hat{\mu} + Z_{1-\alpha/2}^2/4n\right)}$$

- Clopper-Pearson CI: (the exact CI)

$$\frac{\chi^2_{2n\hat{\mu},\alpha/2}}{2n} < \mu < \frac{\chi^2_{2(n\hat{\mu}+1),1-\alpha/2}}{2n}$$

4.2.1 Introduction to Poisson regression models

Poisson Regression Model (Loglinear Model): Consider the response $Y|x \sim Poisson(\mu(x))$ and assume

$$\log[\mu(x)] = \alpha + \beta x$$

equivalently to $\mu(x) = \exp(\alpha + \beta x)$.

- $\mu(x) \geq 0$ for all x
- β's interpretation: when $x \uparrow x + \Delta x$,
  - $\beta = 0 \Rightarrow \mu(x) = e^{\alpha}$;
  - $\beta > 0 \Rightarrow \mu(x) \uparrow \mu(x)e^{\beta\Delta x}$;
  - $\beta < 0 \Rightarrow \mu(x) \downarrow \mu(x)e^{\beta\Delta x}$

4.2.1 Introduction to Poisson regression models

Poisson Regression Model (Loglinear Model): Consider the response $Y|x_1, \ldots, x_K \sim Poisson(\mu(x_1, \ldots, x_K))$ and assume

$$\log[\mu(x_1, \ldots, x_K)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K$$

equivalently to $\mu(x_1, \ldots, x_K) = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K)$.

- $\mu(x_1, \ldots, x_K) \geq 0$ for all $x_1, \ldots, x_K$
- β_1's interpretation with fixed $x_2, \ldots, x_K$: when $x_1 \uparrow x_1 + \Delta x_1$,
  - $\beta_1 = 0 \Rightarrow \mu(x_1, \ldots, x_K)$ stays the same $e^{\beta_0 + \beta_2 x_2 + \ldots + \beta_K x_K}$;
  - $\beta_1 > 0 \Rightarrow \mu(x_1, \ldots, x_K) \uparrow \mu(x_1, \ldots, x_K) e^{\beta_1 \Delta x_1}$;
  - $\beta_1 < 0 \Rightarrow \mu(x_1, \ldots, x_K) \downarrow \mu(x_1, \ldots, x_K) e^{\beta_1 \Delta x_1}$

4.2.2 Poisson regression analysis

Statistical Inference with Poisson Regression (Loglinear) Models

For the count response Y and explanatory variables $\mathbf{X}$ with $Y|\mathbf{X} = \mathbf{x} \sim Poisson\big(\mu(\mathbf{x})\big)$, and data $\big\{(y_i, \mathbf{x}_i) : i = 1, \ldots\big\}$ from n indpnt individuals: the likelihood function is

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{\mu(\mathbf{x}_i)^{y_i}}{y_i!} e^{-\mu(\mathbf{x}_i)}.$$

- Estimation of α, β: MLE and CI?
- Estimation of $\mu(\mathbf{x})$: MLE and CI?
- Testing on hypotheses about α, β: e.g. $H_0 : \beta = 0$?
- Model Comparison/Checking? [to be studied]

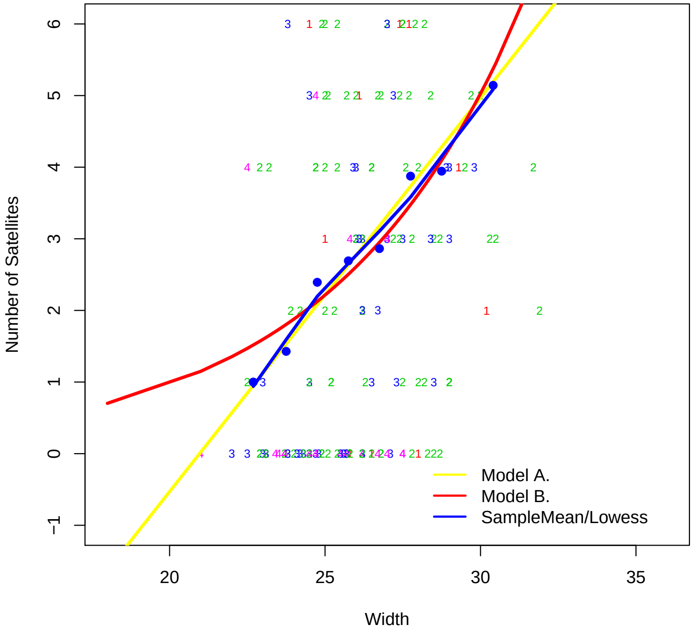
4.2.2 Poisson regression analysis

Example. Revisit 4 to the Horseshoe Crab Study: (Regression with Count Response) Y=number of satellites; X=width

Assume $Y \sim Poisson(\mu(x))$ with

- Model A. Linear mean model: $\mu(x) = \alpha + \beta x$;
  $\hat{\alpha} = -11.53(2.832), \hat{\beta} = .55(.107)$
  with R function $glm(formula, family = gaussian)$

- Model B. Loglinear mean model: $\log\big(\mu(x)\big) = \alpha + \beta x$;
  $\hat{\alpha} = -3.305(.542), \hat{\beta} = .164(.020)$
  with R function $glm(formula, family = poisson)$



What if to consider other explanatory variables?

What will we study next?

1. *Introduction and Preparation*
2. *Analysis with Binary Variables (Chp 1-2)*
3. *Analysis with Multicategory Variables (Chp 3)*
4. **Analsyis with Count Response (Chp 4)**
   - *4.1 Poisson Model for Count Data (Chp 4.1)*
   - **4.2 Poisson Regression Analysis (Chp 4.2)**
   - *4.3 Additional Topics on Count Responses (Chp 4.3-4)*