Exam 02 Review

Dr. Maria Tackett

11.20.19

Click for PDF of slides

Announcements

- HW 06 due **today at 11:59p**
- Project regression analysis due **today at 11:59p**
- Thursday's Lab: Exam 02 office hours

Exam 02 Outline

- Mostly short answer questions
- Permitted to bring one sheet of **handwritten** notes (front and back)
  - I will check your notes when you turn in your exam
- Calculator **not** permitted on exam
- Please use black or blue pen (and write neatly!)

How to prepare

- Review lecture notes and readings
- Review HW and lab assignments
- Practice problem sets on Sakai
- Utilize office hours and help hours during lab
- Study while making your page of notes

Topics

- **Review:** Multiple Linear Regression
- **Models**:
  - Logistic Regression
  - Multinomial Logistic Regression
  - Poisson Regression (main ideas, no zero-inflated poisson)
- Model Selection
- Modeling in practice
  - Model validation (main ideas, why it's important)
  - Dealing with missing data (main ideas, why it's important)

Logistic Regression

- Use for response variable y that is categorical with 2 levels

$$\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_i + \cdots + \hat{\beta}_p x_p$$

- **Slope:** As x_j increases by 1 unit, the odds of y are expected to multiply by a factor of $\exp\{\beta_j\}$, holding all else constant
- **Intercept:** When $x_1 = \ldots x_p = 0$, odds of y are expected to be $\exp\{\beta_0\}$

Multinomial Logistic Regression

- Use for response variable that is categorical with more than 2 levels
- Suppose we have a categorical variable with $k > 2$ levels. Let $y = 1$ be the baseline category

$$\log\left(\frac{\hat{p}_2}{\hat{p}_1}\right) = \hat{\beta}_{02} + \hat{\beta}_{12}x_1 + \cdots + \hat{\beta}_{p2}x_p$$

$$\vdots$$

$$\log\left(\frac{\hat{p}_k}{\hat{p}_1}\right) = \hat{\beta}_{0k} + \hat{\beta}_{1k}x_1 + \cdots + \hat{\beta}_{pk}x_p$$

- **Slope**: When x increases by one unit, the odds of $y = k$ versus $y = 1$ are expected to multiply by a factor of $\exp\{\hat{\beta}_{1k}\}$, holding all else constant.
- **Intercept**: When $x_1 = \ldots x_p = 0$, the odds of $y = k$ versus $y = 1$ are expected to be $\exp\{\hat{\beta}_{0k}\}$

Model Selection

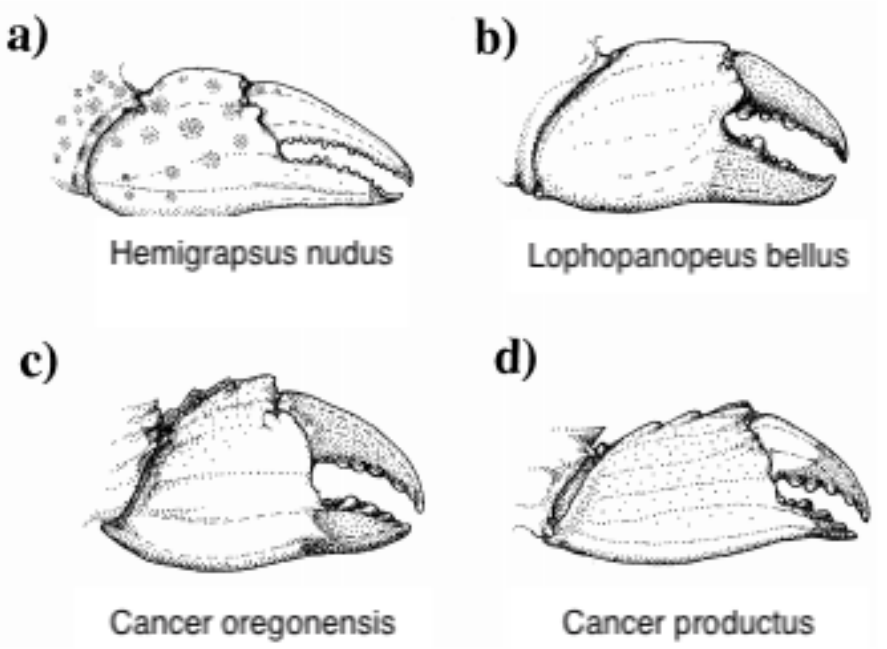
- Consider the main objective:
  - Prediction
  - Adjusting for many variables
  - Explanation
- Forward, backward, stepwise selection
  - Optimize some criteria at each step
- *Example*: Minimize **AIC** = $n\log(SSE) - n\log(n) + 2(p+1)$

Questions?

Data Description

- We would like to identify crab species based on the closing force and propodus height of claws
  - `ex0722` data set in the `Sleuth3` R package
- **Predictors:**
  - **Force:** Closing force of claw (newtons)
  - **Height:** Propodus height (mm)
- **Response:**
  - **Species:** Hemigrapsus nudus (Hn), Lophopanopeus bellus (Lb), Cancer productus (Cp)

Data Description



Source: Yamada, S. and Boulding E., 1998, Claw morphology, prey size selection and foraging efficiency in generalist and specialist shell-breaking crabs, *Journal of Experimental Marine Biology and Ecology*, 220: 191-211.

Exam 02 review questions