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

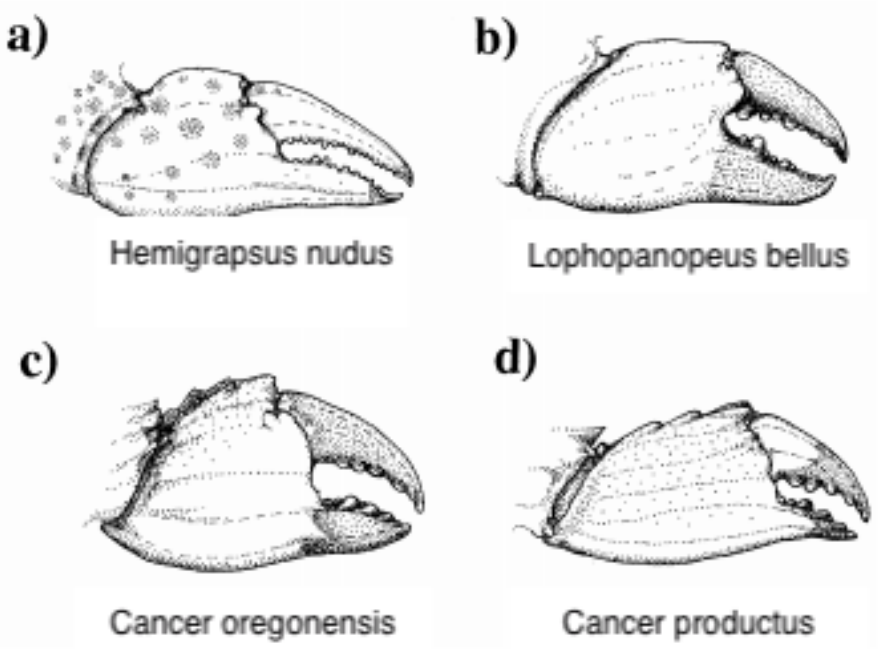
- Consider the main objective:
  - Prediction
  - Adjusting for many variables
  - Explanation
- Forward, backward, stepwise selection
  - Optimize some criteria at each step
- *Example*: Minimize **AIC** = $n\log(SSE) - n\log(n) + 2(p+1)$

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.

Lb species?

- Suppose we want to use `Force` and `Height` to determine whether or not a crab is from the Lophopanopeus bellus (Lb) species.

  - What type of model should we use?
  - What should we do for exploratory data analysis?

Lb species?

- We will use the mean-centered variables for force and height.

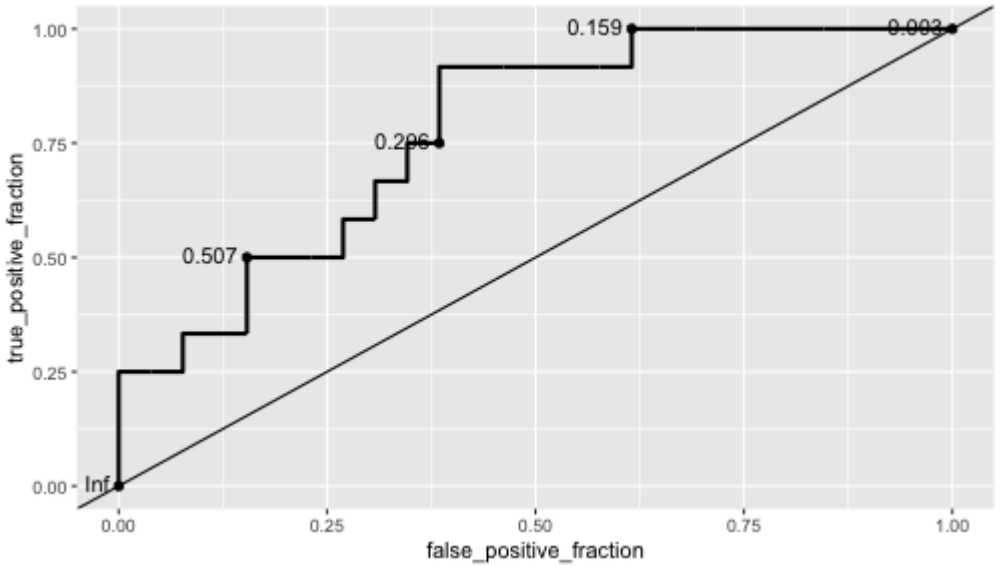
term	estimate	std.error	statistic	p.value
(Intercept)	-1.130	0.463	-2.443	0.015
forceCent	0.211	0.092	2.279	0.023
heightCent	-0.895	0.398	-2.249	0.025

- Write the equation for the odds of a crab being from the Lb species.
- Interpret the intercept in the context of the problem.
- Interpret `forceCent` in the context of the problem.

Lb species?

- What does **sensitivity** mean in the context of this data?
- What does **specificity** mean in the context of this data?

Lb species?




```
## [1] 0.775641
```

Which species?

- Suppose we want to use force and height to determine a crab's species.

y.level	term	estimate	std.error	statistic	p.value
Hn	(Intercept)	-1.193	1.106	-1.079	0.281
Hn	forceCent	-0.494	0.196	-2.514	0.012
Hn	heightCent	0.179	0.474	0.378	0.705
Lb	(Intercept)	0.021	0.602	0.034	0.973
Lb	forceCent	0.095	0.101	0.941	0.347
Lb	heightCent	-0.902	0.429	-2.103	0.035

1. Write the equation for part of the model describing the odds of Hn vs. Cp species.
2. Interpret the intercept for this part of the model.
3. Interpret the coefficient of `forceCent` fort this part of the model.