Exam II Review

Dr. Tackett

12.04.2018
Announcements

- Exam II on Thursday

- Poster Session on December 15, 2p - 5p
Exam II Outline

- Short answer questions

- Permitted to bring one sheet of handwritten notes (front and back)
  - Must turn in notes with exam

- Calculator **not** permitted on exam

- Please use pencil!
Topics

- Review: Multiple Linear Regression

- Model Selection

- Logistic Regression

- Multinomial Logistic Regression

- Ordinal & Poisson Regression
  - Main ideas - no calculations
Not on Exam

- ANOVA
- Missing Data
- Time Series
- R Code
Variable Selection

- Consider the main objective:
  - Prediction
  - Adjusting for many variables
  - Explanation

- Forward, backward, stepwise selection
  - Optimize some criteria at each step

- *Ex:* Minimize \( \text{AIC} = n \log(SSE) - n \log(n) + 2(p + 1) \)
Proportional Odds Model

- Use for response variable that is categorical and
  - Has more than 2 levels
  - Levels have natural ordering

- **Model:**

\[
\log \left( \frac{P(Y \leq 1)}{P(Y > 1)} \right) = \beta_{01} - [\beta_1 X_1 + \cdots + \beta_p X_p]
\]

\[
\log \left( \frac{P(Y \leq 2)}{P(Y > 2)} \right) = \beta_{02} - [\beta_1 X_1 + \cdots + \beta_p X_p]
\]

\[\vdots\]

\[
\log \left( \frac{P(Y \leq (J - 1))}{P(Y > (J - 1))} \right) = \beta_{0,J-1} - [\beta_1 X_1 + \cdots + \beta_p X_p]
\]
Proportional Odds Model

- **Slope**: When $X_1$ increases by one unit, the odds of falling at or below category $j$ multiply by a factor of $\exp\{-\beta_1\}$

- **Intercept**: When all explanatory variables equal 0, the odds of falling at or below category $j$ are $\exp\{\beta_{0j}\}$
Poisson Regression Model

- Use for response variable that is a count

- **Model:**

  \[ \log(\mu \{Y|X\}) = \beta_0 + \beta_1 X \]

- **Slope:** When \( X \) increases by one unit, the expected count of \( Y \) changes by a multiplicative factor of \( \exp\{\beta_1\} \)

- **Intercept:** When \( X \) is 0, the expected count of \( Y \) is \( \exp\{\beta_0\} \)
Multinomial Logistic Regression

- Use for response variable that is categorical and
  - Has more than 2 levels

- **Model:**
  - Choose a baseline category. Let's choose $Y = 1$. Then,

\[
\log \left( \frac{\pi_2}{\pi_1} \right) = \beta_{02} + \beta_{12}X \\
\vdots \\
\log \left( \frac{\pi_J}{\pi_1} \right) = \beta_{0J} + \beta_{1J}X
\]
Multinomial Logistic Regression

- **Slope:** When $X$ increases by one unit, the odds of $Y = j$ versus $Y = 1$ are expected to multiply by a factor of $\exp(\beta_{1j})$.

- **Intercept:** When $X = 0$, the odds of $Y = j$ versus $Y = 1$ are expected to be $\exp(\beta_{0j})$. 
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

- Explanatory:
  - **Force:** Closing force of claw (newtons)
  - **Height:** Propodus height (mm)

- Response:
  - **Species:** Hemigrapsus nudus (Hn), Lophopanopeus bellus (Lb), Cancer productus (Cp)
Data Description

Exploratory Data Analysis

For each category of the response variable, examine plots of the explanatory variables vs. the proportion in that category.
Exploratory Data Analysis

- For each category of the response variable, examine plots of the explanatory variables vs. the proportion in that category.

![Proportion Hemigrapsus vs. Force](image1)

![Proportion Hemigrapsus vs. Height](image2)
Exploratory Data Analysis

- For each category of the response variable, examine plots of the explanatory variables vs. the proportion in that category.
## Model

<table>
<thead>
<tr>
<th>y.level</th>
<th>term</th>
<th>estimate</th>
<th>std.error</th>
<th>statistic</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hn</td>
<td>(Intercept)</td>
<td>3.220</td>
<td>3.251</td>
<td>0.990</td>
<td>0.322</td>
</tr>
<tr>
<td>Hn</td>
<td>Force</td>
<td>-0.494</td>
<td>0.196</td>
<td>-2.514</td>
<td>0.012</td>
</tr>
<tr>
<td>Hn</td>
<td>Height</td>
<td>0.179</td>
<td>0.474</td>
<td>0.378</td>
<td>0.706</td>
</tr>
<tr>
<td>Lb</td>
<td>(Intercept)</td>
<td>6.822</td>
<td>2.900</td>
<td>2.352</td>
<td>0.019</td>
</tr>
<tr>
<td>Lb</td>
<td>Force</td>
<td>0.095</td>
<td>0.101</td>
<td>0.941</td>
<td>0.346</td>
</tr>
<tr>
<td>Lb</td>
<td>Height</td>
<td>-0.902</td>
<td>0.429</td>
<td>-2.103</td>
<td>0.035</td>
</tr>
</tbody>
</table>

1. Write the model for the odds of Hn vs. Cp species.

2. Interpret the intercept for this model.

3. Interpret the slope of Force for this model.
Residuals

- For each category of the response, examine the binned residual plots vs the explanatory variables.
Residuals

- For each category of the response, examine the binned residual plots vs the explanatory variables.
Residuals

- For each category of the response, examine the binned residual plots vs the explanatory variables.

- What do we learn from the residuals?
Force, Height, and Species

Height vs. Force by Species

- Species legend:
  - Cp
  - Hn
  - Lb
Lb species?

- Suppose we only wish to determine whether or not a crab is from the Lophopanopeus bellus (Lb) species.

- What type of model should we use?
Logistic Regression Model

<table>
<thead>
<tr>
<th>term</th>
<th>estimate</th>
<th>std.error</th>
<th>statistic</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.944</td>
<td>0.316</td>
<td>2.984</td>
<td>0.005</td>
</tr>
<tr>
<td>Force</td>
<td>0.027</td>
<td>0.011</td>
<td>2.543</td>
<td>0.016</td>
</tr>
<tr>
<td>Height</td>
<td>-0.108</td>
<td>0.043</td>
<td>-2.537</td>
<td>0.016</td>
</tr>
</tbody>
</table>

- Write the equation for the odds of crab being from the Lb species.

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

- What does **specificity** mean in the context of this model?
# Area under the curve: 0.7724
Questions?