STOR 590: ADVANCED LINEAR MODELS

Instructor: Richard L. Smith

Class Notes:

August 28, 2020

CLASS ANNOUNCEMENTS

- New grading policy — P/F grades allowed
  - Change from last semester — C- or D will translate to "low pass"
  - Must declare by last day of semester
- Option of "CV" grade (for incomplete) still exists. Also "WCV" for a course dropped after 08/31.
- Homeworks: expect about 10 assignments total. **Bottom 2 will be dropped**
- Today's homework deadline extended to 11:55 pm — also a "late" deadline of Wednesday, September 2
- Late homeworks will be deducted 25% (I may not enforce this, but assume that I will)
- HW1 will be graded shortly

Example 2:

Logistic Regression With One Covariate

- $f_i(y_i\ ;\ \theta) = \frac{\exp\{y_i(\beta_0+\beta_1 x_i)\}}{1+\exp(\beta_0+\beta_1 x_i)}$, $\theta = (\beta_0, \beta_1)$
- $\ell = \sum_i \log\{1 + \exp(\beta_0 + \beta_1 x_i)\} - \sum_i y_i(\beta_0 + \beta_1 x_i)$
- $\frac{\partial \ell}{\partial \beta_0} = \sum_i \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} - \sum_i y_i$
- $\frac{\partial \ell}{\partial \beta_1} = \sum_i \frac{x_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} - \sum_i x_i y_i$
- Set $\frac{\partial \ell}{\partial \beta_0} = \frac{\partial \ell}{\partial \beta_1} = 0$, solve for $\beta_0,\ \beta_1$
- Intuition: $\sum_i (y_i - p_i) = 0,\ \sum_i x_i(y_i - p_i) = 0$
- Even so, the equations are nonlinear — solve numerically for $\widehat{\beta}_0,\ \widehat{\beta}_1$

- Rewrite $\frac{\partial \ell}{\partial \beta_0} = \sum_i \left(1 - \frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right) - \sum_i y_i$,
  $\frac{\partial \ell}{\partial \beta_1} = \sum_i x_i \left(1 - \frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right) - \sum_i x_i y_i$

- $\frac{\partial^2 \ell}{\partial \beta_0^2} = \sum_i \{1+\exp(\beta_0+\beta_1 x_i)\}^{-2} \cdot \exp(\beta_0+\beta_1 x_i) > 0$

- Also write as $\sum_i p_i(1-p_i)$

- $\frac{\partial^2 \ell}{\partial \beta_1^2} = \sum_i x_i^2 \{1+\exp(\beta_0+\beta_1 x_i)\}^{-2} \cdot x_i^2 \exp(\beta_0+\beta_1 x_i) > 0$

- $\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = \sum_i x_i \{1+\exp(\beta_0+\beta_1 x_i)\}^{-2} \cdot x_i \exp(\beta_0+\beta_1 x_i)$

- $H = \left[\begin{array}{cc} \sum_i \widehat{p}_i(1-\widehat{p}_i) & \sum_i x_i \widehat{p}_i(1-\widehat{p}_i) \\ \sum_i x_i \widehat{p}_i(1-\widehat{p}_i) & \sum_i x_i^2 \widehat{p}_i(1-\widehat{p}_i) \end{array} \right]$

- The determinant of H is > 0 unless all the x_i are the same — this proves that $(\widehat{\beta}_0, \widehat{\beta}_1)$ is a local minimum of ℓ and H^{-1} is a good approximation to the variance-covariance matrix of $(\widehat{\beta}_0, \widehat{\beta}_1)$

Interpretation as Ratio of Odds

- Example: For the smoking-CHD example in the text, $\widehat{\beta}_2 = 0.02313$. How should this be interpreted?

- One answer: for a person who smokes 20 cigarettes a day, $\log \frac{p}{1-p}$ is $20 \times 0.02313 = 0.4626$ larger than for a person who smokes none (p: probability of CHD)

- Alternatively: for a person who smokes 20 cigarettes a day, $\frac{p}{1-p}$ is multiplied by $e^{0.4626} = 1.59$

- In common probability terminology, $\frac{p}{1-p}$ is the *odds*.
  - Example: One bookmaker gives odds of 37:20 that the Patriots will win the Superbowl.
  - Equivalent to: probability of winning is $\frac{37}{37+20} = 0.65$.

- For a 20-a-day smoker, the risk of CHD is increased by about 54%. Not much different from saying the odds is increased 59%.

Profile Likelihood

- Sometimes we're primarily interested in one parameter — all the rest are "nuisance parameters"
- Say θ_1 is interest parameter, $\theta_2, ..., \theta_p$ are nuisance
- Define

$$\ell_P(\theta_1^*) \;=\; \min\{\ell(\theta_1, ..., \theta_p) : \; \theta_1 = \theta_1^*\}$$

- This is called the *profile (log) likelihood* of θ_1
- Can test a specific value for θ_1^* by using LRT with χ_1^2 distribution

Deviance Residuals

- Define the *deviance* as

$$
\begin{aligned}
D &= 2\ell(\widehat{\theta}) \\
&= 2\sum_i \left[\log\{1+\exp(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\} - y_i(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\right] \\
&= \sum_i r_i^2
\end{aligned}
$$

where

$$
r_i^2 = 2\left[\log\{1+\exp(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\} - y_i(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\right]
$$

Ensure correct sign by defining

$$
r_i = \text{sign}(y_i - \widehat{p}_i)\sqrt{r_i^2}.
$$

We call r_i the i'th *deviance residual* (text, page 36).

In R: `residuals(lmod)`

Side Comment

- We defined

$$r_i^2 \;=\; 2\left[\log\{1+\exp(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\} - y_i(\widehat{\beta}_0+\widehat{\beta}_1 x_i)\right]$$

  Do we know this is > 0?

- Claim: $\log(1+e^z) - yz > 0$ when $-\infty < z < \infty,\ y = 0$ or 1

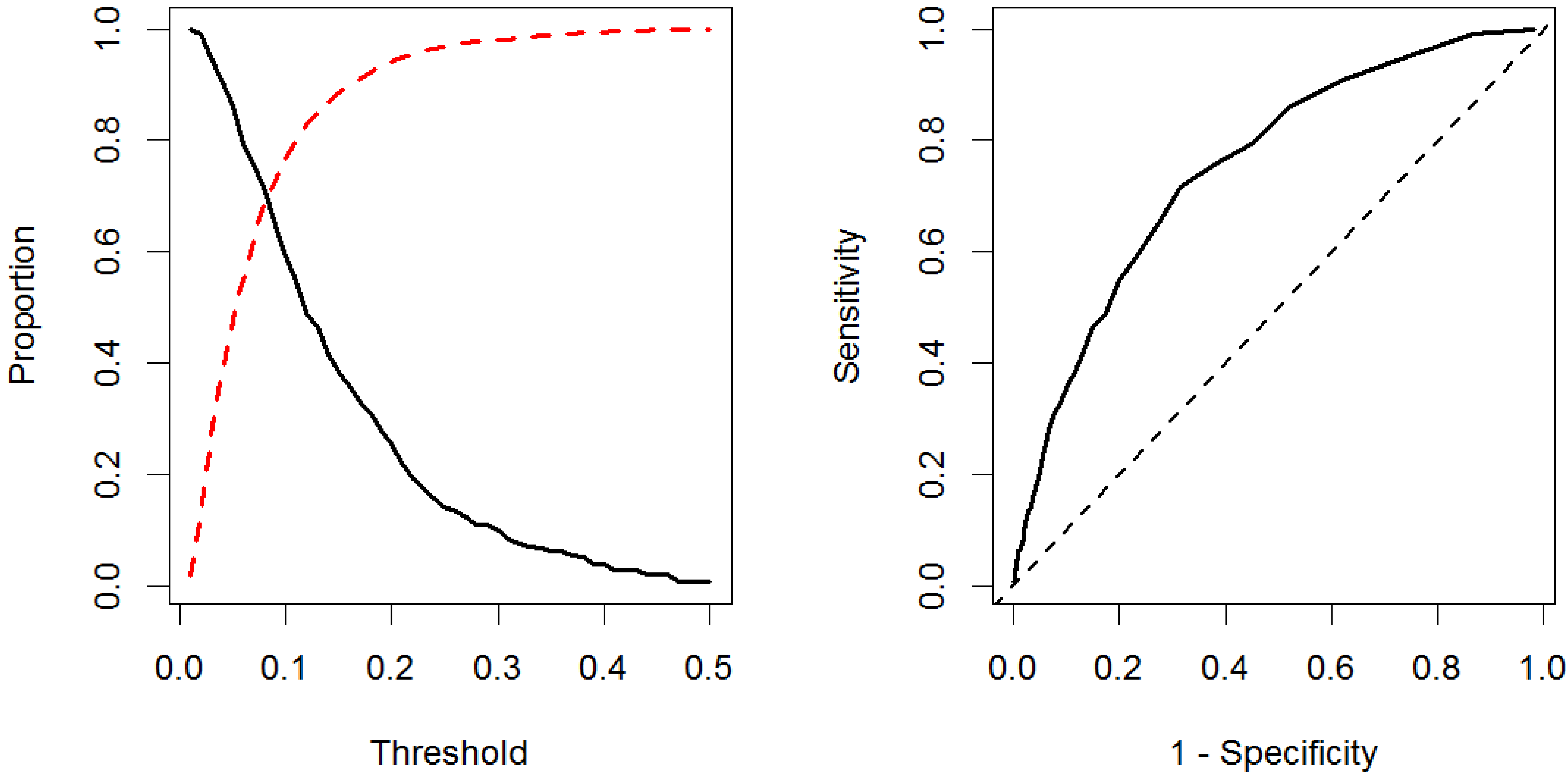
- $y = 0:\ \log(1+e^z) > \log(1) > 0$

- $y = 1:\ \log(1+e^z) - z > \log(e^z) - z = 0$

- So OK either way.

Sensitivity and Specificity

- Assume we are testing for a disease or some specific health outcome, and we use a diagnostic test to predict the outcome
- Specificity: the probability that a person who *does not have* the disease is correctly predicted to not have the disease
- Sensitivity: the probability that a person who *does have* the disease is correctly predicted to have the disease
- After subtracting from 1, these are analogous to type I error and type II error, respectively
- Sensitivity is also the *power of the test*
- As the threshold for detection rises, the specificity increases but the sensitivity decreases
- The plot of Sensitivity against 1-Specificity is called the *Receiver Operating Characteristic* or ROC curve


Proportion
Threshold
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0