

# Linear Regression Models

## P8111

Lecture 25

Jeff Goldsmith  
April 26, 2016



THE DEPARTMENT OF  
**BIostatISTICS**



Columbia University  
MAILMAN SCHOOL  
OF PUBLIC HEALTH

# Today's Lecture

- Logistic regression / GLMs
  - ▶ Model framework ✓
  - ▶ Interpretation ✓
  - ▶ Estimation ✓

# Linear regression

Course started with the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where

$$\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

In particular,  $y_i$  has been continuous throughout the course

$$y_i | x_i \sim \mathcal{N}(x_i \beta, \sigma_\epsilon^2)$$

# Binary responses

Binary outcomes are common in practice; usually indicate some event

- Yes vs no
- Transplant vs no transplant
- Death vs no death

*Count*

*PROP [0, 1]*

# Binary responses

$$y|x \sim \mathcal{N}(x\beta, \sigma^2)$$

$$E(y|x)$$

How should we deal with binary (0/1) y's?

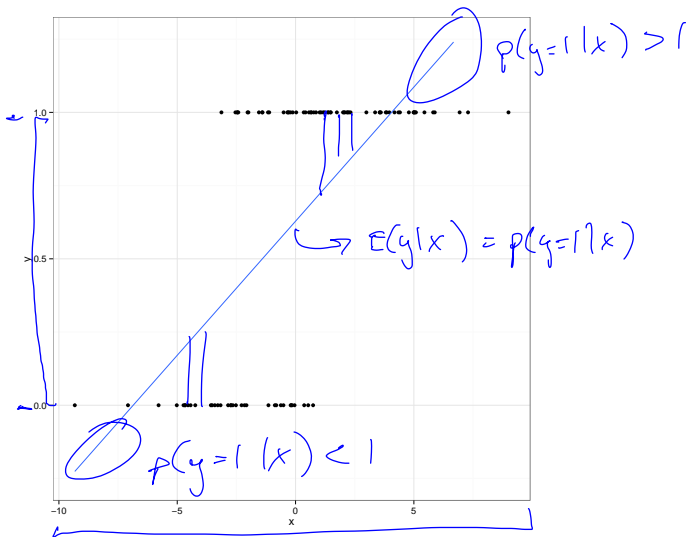
- Regression focuses on  $E(y|x) = x\beta + \epsilon$
- For binary outcomes, we want  $E(y|x) = p(y = 1|x)$
- Does  $p_i = p(y = 1|x) = \beta_0 + \beta_1 x_i$  work?

$$p_i = x\beta \quad X$$


$$p_i = x\beta \quad \text{? ?}$$

$$y_i = x\beta + \epsilon$$

# Linear regression for binary outcome



# What we need for binary outcomes

- Fitted probabilities should be between 0 and 1
- Use an invertible function  $g : (0, 1) \rightarrow (-\infty, \infty)$  to link probabilities to the real line
- Build a model for  $\underline{g(p_i)} = \underline{\beta_0 + \beta_1 x_i}$   


$$g^{-1}(x\beta) = p_i$$

# Link functions

- Lots of possible link functions: logit, probit, complimentary log-log
- By far, most common is the logit link:

$$g(p_i) = \text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = z_i$$

- The inverse link function is also useful:

$$g^{-1}(z_i) = \frac{\exp(z_i)}{1 + \exp(z_i)} = p_i$$

$\Phi^{-1}(p_i)$





# Logistic regression

Model is now

$$\begin{aligned}E(y_i|x_i) &= p_i \\g(p_i) &= \log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_i\end{aligned}$$

Using the logit link, we have

$$p_i = g^{-1}(\beta_0 + \beta_1 x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

# Parameter interpretation

Suppose we can estimate  $\beta_0, \beta_1$ ; what do they mean?

For a binary predictor ...

# Parameter interpretation

For a continuous predictor ...

# Parameter estimation

- For linear regression, we used least squares and found that this corresponded to ML
- Try using maximum likelihood for logistic regression; need a likelihood ...

# ML for logistic regression

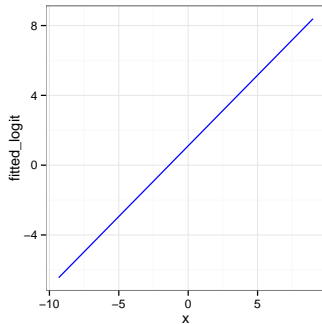
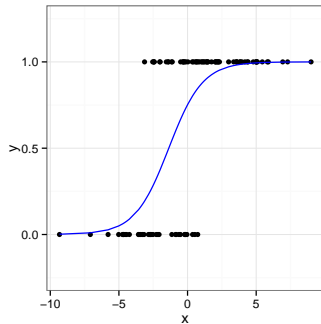
- Assume that  $[y_i|x_i] \sim \text{Bern}(p_i)$
- Density function is  $p(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$
- As before, use that  $\text{logit}(p_i) = \beta_0 + \beta_1 x_i$
- Likelihood is

$$L(\beta_0, \beta_1; \mathbf{y}) = \prod_{i=1}^n p_i^{y_i}(1 - p_i)^{1-y_i}$$

# ML for logistic regression

- Log likelihood is easier to work with, but it is typically not possible to find a closed-form solution
- Iterative algorithms are used instead (Newton-Raphson, Iteratively Reweighted Least Squares)
- These are implemented for a variety of link functions in  $R$

# Example



# Code

```
> model = glm(y~x, family = binomial(link = "logit"), data = data)
> summary(model)

Call:
glm(formula = y ~ x, family = binomial(link = "logit"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9360  -0.4631   0.1561   0.5564   1.8131

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.1072     0.3357   3.298 0.000974 ***
x             0.8097     0.1664   4.865 1.15e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 129.49  on 99  degrees of freedom
Residual deviance:  73.24  on 98  degrees of freedom
AIC: 77.24

Number of Fisher Scoring iterations: 6
```



# Multiple predictors

- Essentially everything that worked for linear models works for logistic models:
  - ▶ Multiple predictors of various types
  - ▶ Interactions
  - ▶ Polynomials
  - ▶ Piecewise, splines
  - ▶ (Penalization, random effects, Bayesian models)

# Testing in Logistic

- In linear models, many of our inferential procedures (ANOVA, F tests, ...) were based on RSS
- For logistic regression (and GLMs), we'll use the asymptotic Normality of MLEs:

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N[0, V]$$

with  $V = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$  and weight matrix  $W$  to construct Wald tests

- Likelihood ratio tests can be used to compare nested models

# Wald tests

For individual coefficients

- We can use the test statistic

$$T = \frac{\hat{\beta}_j - \beta_j}{\widehat{se}(\hat{\beta}_j)}$$

- This is compared to a Normal distribution, trusting that the asymptotics have kicked in
- Recall that coefficients are on the logit scale ...

# Confidence intervals

- A confidence interval with coverage  $(1 - \alpha)$  is given by

$$\beta_j \pm t_{1-\alpha/2, n-p-1} \widehat{se}(\hat{\beta}_j)$$

- To create a confidence interval for the  $\exp(\hat{\beta}_j)$ , the estimated odds ratio, exponentiate:

$$(\exp(\hat{\beta}_j - 2\widehat{se}(\hat{\beta}_j)), \exp(\hat{\beta}_j + 2\widehat{se}(\hat{\beta}_j)))$$

# Wald tests for multiple coefficients

- Define  $H_0 : c^T \beta = c^T \beta_0$  or  $H_0 : c^T \beta = 0$
- We can use the test statistic

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\widehat{se}(c^T \hat{\beta})} = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{c^T \text{Var}(\hat{\beta}) c}}$$

- Useful for some tests, looking at fitted values

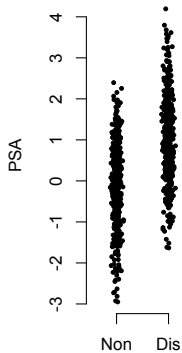
# Model building

- Can define a model building strategy (at least for nested models) using these
- Other tools, like AIC and BIC, can compare non-nested models

# ROC curves

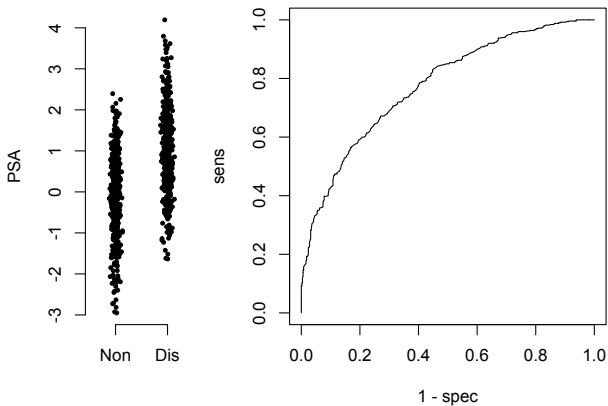
- Forget logistic for a minute
- Suppose you have some test to classifying subjects as diseased or non-diseased
- You can describe that test using sensitivity  $P(+|D)$  and specificity  $P(-|D')$
- These values depend on what threshold you use for your test

# Threshold effect on sens, spec

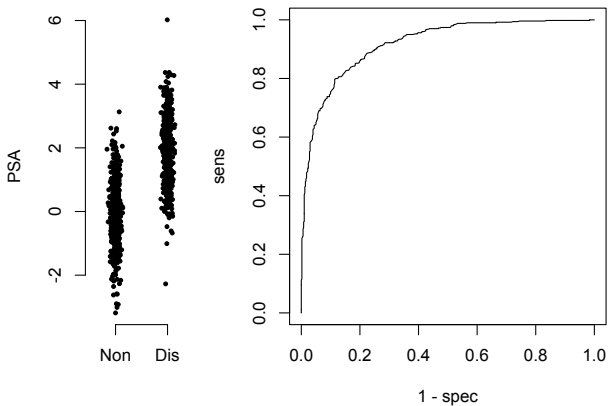




# Threshold effect on sens, spec



# Better tests give better ROCs



# Summarizing ROCs

- Area under the curve is a useful summary of an ROC
- AUC shouldn't be less than .5; can't be more than 1
- Bigger AUC indicates better classification
- Useful alternative to AIC, BIC, etc

# Connection with logistic regression

- Your “test” might be  $\hat{\mu}_i = \hat{p}(y_i = 1|x_i)$
- You can model this probability using logistic regression
- Cross-validated ROCs are a way to compare the predictive performance of different models:
  - ▶ Based on fitted model (from training set) you construct fitted probabilities  $\hat{\mu}_i = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$  for subjects in the validation set
  - ▶ Validation subjects test “positive” or “negative” based on their fitted value; compare to the observed value

# Generalizing this approach

Suppose instead of binary data, we have

$$y_i \sim EF(\mu_i, \theta)$$

where

$$E(y_i|x_i) = \mu_i$$

and

$$\text{Var}(y_i|x_i) = a(\phi)V(x_i)$$

with known variance function  $V(\cdot)$  and dispersion parameter  $\phi$

# Generalized Linear Model

Model components are the

- Probability distribution
- Link function
- Linear predictor

# Linear regression as a GLM

# Comparing linear and logistic

- ▶ Comparing linear, logistic, and Poisson regression models:

	Linear	Logistic	Poisson
Outcome	Continuous	Binary	Count
Distribution	Normal	Binomial	Poisson
Parameter	$E(Y) = \mu$	$E(Y) = p$	$E(Y) = \lambda$
Range of mean	$-\infty < \mu < \infty$	$0 < p < 1$	$0 < \lambda < \infty$
Variance	$\sigma^2$	$p(1 - p)$	$\lambda$
“Natural” Link	identity	logit	log



Other link functions?

# Other GLMs

Framework holds for any member of the exponential family

- Probability distribution
- Link function
- Linear predictor

# Exponential family distribution

Any distribution whose density can be expressed as

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta + b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

where  $b'(\theta) = \mu$  and  $b''(\theta) = V$

- Can take some effort to convert usual density to this form
- Includes Normal, Bern, Poisson, Gamma, Multinomial, ...

# Exponential family examples

Normal:

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right) \\ &= \exp\left((y\mu - \mu^2/2)/\sigma^2 - \frac{1}{2}(y^2/\sigma^2 + \log(2\pi\sigma^2))\right) \end{aligned}$$

# Exponential family examples

Bernoulli:

$$\begin{aligned} f(y; p) &= \exp(y \log(p) + (1 - y) \log(1 - p)) \\ &= \exp\left(y \log \frac{p}{1 - p} + (-\log(1 - p))\right) \end{aligned}$$

# Today's big ideas

- Logistic regression and GLMs

- 
- Suggested reading: ISLR Ch 4.2 and 4.3