STATISTICS 538, Lecture #2

Logistic Regression

October 27, 2010

Space shuttle data: on i-th launch, y_i out of $n_i = 6$ ‘O-Rings’ sustained some damage, $i = 1, \ldots, n = 23$

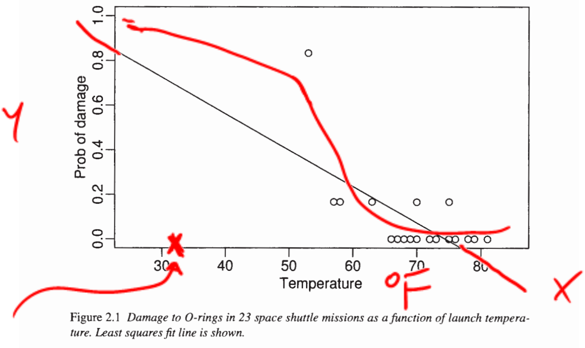

Figure 2.1 *Damage to O-rings in 23 space shuttle missions as a function of launch temperature. Least squares fit line is shown.*

Two equivalent data representations: R happy with either!

‘Bernoulli’

OUTCOME	TEMP
1	53
1	53
1	53
1	53
1	53
0	53
1	57
0	57
0	57
0	57
0	57
0	57
⋮	⋮

‘Binomial’

‘S’	‘F’	TEMP
5	1	53
1	5	57
	⋮	

Logistic regression based on **logit** (actually log-odds!) function: $\text{logit}(p) = \log\{p/(1-p)\} = \log p - \log(1-p)$

p maps $(0,1) \to (-\infty, \infty)$

$(Y_i|X^{(i)}) \sim \text{Binomial}(n_i, p_i)$,

$p_i = \text{logit}^{-1}(\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_{q-1} x_{q-1}^{(i)})$
$\equiv \text{expit}(\ \text{''}\ )$

where

$$\text{logit} p_i \quad = \quad \beta_0 + \beta_1 X_1^{(i)} + \ldots + \beta_{q-1} X_{q-1}^{(i)}$$

Note - completely specified - no variance to worry about.

Interpretation?

$(X_1, X_2, \ldots X_{j-1}, X_{j+1}, \ldots X_{q-1})$

LM: β_j describes change in $E(Y|X)$ w.r.t. X_j, with X_{-j} held fixed.

Logistic regression?

where $\text{expit}(z) = \frac{1}{1+e^{-z}}$

Very natural if willing to think in terms of **odds**

odds = 3 → prob = .75
prob = .6 → odds = 1.5

e.g., Bernoulli data, X_j also binary:

odds-ratio

$$\frac{\overbrace{\text{Odds}(Y=1|X_j=1,X_{-j})}^{e^{\beta_j \times 1 + \text{stuff}}}}{\underbrace{\text{Odds}(Y=1|X_j=0,X_{-j})}_{e^{\beta_j \times 0 + \text{stuff}}}} = \exp(\beta_j),$$

provided X_j not involved in interactions...

Continuous X_j?

$$\frac{\text{odds}(Y=1|X_j=a+\Delta, X_{-j})}{\text{odds}(Y=1|X_j=a, X_{-j})} = \exp(\Delta\beta_j)$$

This is the basis for interpreting logistic regression models, and explains the popularity of the logit function, rather than some other 'appropriate' function, for setting up a binary regression model.

invertible, map $(0,1) \to (-\infty, \infty)$

Interpretation gets pushed a bit further, particularly in health sciences...

Mathematically, if p_a, p_b close to zero,

$$\frac{p_a/(1-p_a)}{p_b/(1-p_b)} \approx p_a/p_b$$

So if $Y=1$ is <u>rare</u> (i.e., "rare disease" assumption)

$$\text{Relative risk} \equiv \frac{\Pr(Y=1 \mid X_j=1, X_{-j})}{\Pr(Y=1 \mid X_j=0, X_{-j})} \approx \frac{\text{Odds}(Y=1 \mid X_j=1, X_{-j})}{\text{Odds}(Y=0 \mid X_j=0, X_{-j})} = e^{\beta_j}$$

easier to think about

Mechanics of fitting: simple as $\hat{\beta} = (X^T X)^{-1} X^T Y$ for LM?

iteratively reweighted least squares

Not quite ...

Newton-Raphson to get MLE $\hat{\beta}$ (more later...)

Usual ML asymptotics to get $\hat{\text{Var}}\left(\hat{\beta}\right)$, $\leftarrow \left[-\ell''(\hat{\underset{\sim}{\beta}})\right]^{-1}$

hence standard errors and confidence intervals.

log-likelihood

$$\ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \log\left\{\binom{n_i}{y_i}\left[\text{expit}(x^{(i)}\underset{\sim}{\beta})\right]^{y_i}\left[1-\text{expit}(x^{(i)}\beta)\right]^{n_i-y_i}\right\}$$

$= \dots$

then $\hat{\underset{\sim}{\beta}}$ solves $\ell'(\underset{\sim}{\beta}) = \underset{\sim}{0}$

R syntax: lm(y~x, ...) $\rightarrow$ glm(y~x, family=binomial, ...)

Shuttle data again

data frame

y

```
> logitmod <- glm(cbind(damage,6-damage) ~ temp, family=binomial, orings)
> summary(logitmod)
Deviance Residuals:
   Min       1Q   Median       3Q      Max
-0.953   -0.735   -0.439   -0.208    1.957

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  11.6630     3.2963    3.54    4e-04
temp         -0.2162     0.0532   -4.07   4.8e-05

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 38.898  on 22  degrees of freedom
Residual deviance: 16.912  on 21  degrees of freedom
AIC: 33.67
```

← later

$\hat{\beta}$

odds of damage to an O-ring goes down by a factor of $e^{-.2162} \approx .81$ when temp goes up by $1°F$

odds of damage $e^{2.162} \approx 8.7$ times higher when $10°F$ colder

Intercept only model

Fig 2.2 ← steep

Goodness-of-fit (GOF) - easier than for LM!

Recall data structure $Y_i \sim \text{Binomial}(n_i, p_i)$, $i = 1, \ldots, n$.

Have (somehow) chosen regressors $(1, X_1, \ldots, X_{q-1})$.

Do likelihood ratio test:

H_0: $\text{logit}\, p_i = \beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_{q-1} x_{q-1}^{(i)}$

H_1: $p_1, \ldots, p_n$ unknown parameters

saturated model

Recall in general for a pair of **nested** models with likelihood functions $L_S(\theta_S)$ and $L_L(\theta_L)$:

Look at $D = 2\left\{\log L_L(\hat{\theta}_L) - \log L_S(\hat{\theta}_S)\right\}$.

If the smaller model is correct then

$D \overset{\text{approx}}{\sim} ?$ $\chi^2_{q_L - q_S}$ resp. # params

special case - larger model is the saturated model

- call D the <u>deviance</u> of the smaller model
- the smaller model fails GOF if D is too large

More on GOF

($D = 16.9$ compared to χ^2_{23-2}

Shuttle data:

- no evidence of lack of fit for $\overbrace{\text{logit}\, p = \beta_0 + \beta_1 \text{TEMP}}$
- evidence of lack of fit for $\underbrace{\text{logit}\, p = \beta_0}$ $D = 38.9$ compared to χ^2_{23-1}

Very important caveat: conditions for the χ^2 approximation to the sampling distribution of the the deviance to be good?

n_i's not too small ($n_i \geq 5$ rule-of-thumb?))

$n_i = 1$ problematic (though procedure still gets used))

Conceptual question: why can't the same approach to GOF work in the LM setting?