

Topic 6: Intro to GLM

NOV11 MON

Note: Some of the Latex formatted math functions may not be displayed well on Github. Try use Typora on Mac instead.

Questions to be solved

1. pp 57 structural defect of linear probability model
2. pp60 same relative spacing

Review:

1. Coding in SAS

- Effect Coding 1 -1, Reference Coding Ref = first or Ref = Last

2. Model Overfitting

- Too many variables relative to the sample size or number of events results in numerically unstable estimates.

3. Multicollinearity

- Strong Linear dependencies among the explanatory vars
- vars highly correlated to each other -> hard to estimate vars' distinct effects on some dependent vars
- make coefficients unstable

4. EPV less than 10

- Penalized regression(Ridge Regression,LASSO)

5. Model Selection

- Forward, Backward
- p-value, AIC($-2 \times (\text{maxloglike} - \text{number of parameters})$)
- simple model can be preferred b/c provide better estimates of certain characteristics
- predictive power: c-index, ROC

6. Model Validation

1. Calibration: agreement of observed outcomes and predictions
2. Discrimination: ability to discriminate b/t low/high risk patients

Intro to GLM

1. Restrictions of linear models

1. Y is restricted(eg. binary or counts)
 2. when var of Y depends on mean
2. All Models listed have the distn in the Exponential Dispersion Family

3. limitations

1. Linear function
2. Responses must be independent

4. Components

1. random component
 2. Systematic component
 3. Link function
 1. $g(\mu) = \mu$ called identity link
 2. Canonical Link: transform the mean to the natural parameter
5. GLM is a linear model for a transformed mean of a response variable that has a distribution in the natural exponential family.
6. $f(y_i; \theta_i) = a(\theta_i)b(y_i)\exp(y_i Q(\theta_i))$ Q_i : natural parameter

7. Bernoulli

1. $f = \pi^y (1 - \pi)^{1-y} = (1 - \pi) \exp * y \log\left(\frac{\pi}{1-\pi}\right)$

8. Poisson

1. $f = \frac{e^{-\mu} \mu^y}{y!} = \exp(-\mu) \frac{1}{y!} \exp(y \log \mu)$

9. Transformation VS GLM

Transform:

1. Response var has changed
2. transformation must simultaneously improve linearity and homogeneity of variance
3. transformation may not be defined on the boundaries of the sample space

GLM:

1. unnecessary to transform data because GLM fitting process uses ML methods for random component, we are not restricted to normality of that choice
 2. link function is separate from the choice of random component, is not chosen to produce normality or stabilize the variance.
10. **Exponential Family Restriction:** A weighted version of least squares for ML parameter estimates applies to this entire family for any choice of link.
11. **Binary Case:** $\pi(x) = \alpha + \beta x$'s structural defect: only deal with response 0 and 1
12. **Why not 0 1 2 3 for those levels:** distance between different levels(x variables) may not be the same.
1. Different schemes for x say 0 1 2 3 instead of 0 2 4 5 will give different fitted values
 2. Same relative spacing 0 4 8 10 will give different fitted values

13. **Probit Model:** Use normal as the link function, logistic: logit function

1. Mean $\mu = -\frac{\alpha}{\beta}$
2. standard deviation $\sigma = \frac{1}{|\beta|}$

14. **Effect of X:** $\beta = \text{link}[\pi(1)] - \text{link}[\pi(0)]$ ie. measures of association for 2x2 tables are effect parameters in GLMs for binary data

1. identity link: $\pi(1) - \pi(0)$
2. Log link: $\log[\pi(1)] - \log[\pi(0)] = \log[\pi(1)/\pi(0)]$
3. Logit link = $\text{logit}[\pi(1)] - \text{logit}[\pi(0)] = \log\left(\frac{\pi(1)}{1-\pi(1)}\right) - \log\left(\frac{\pi(0)}{1-\pi(0)}\right) = \log\left[\frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))}\right]$

4. **Count Case :**

1. $\log\mu(x) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$
2. $\mu(x) = \exp(\alpha + \beta_1 x_1 + \dots + \beta_p x_p)$: A 1-unit increase in x_j has a multiplicative impact of e^{β_j}
3. $\exp(\hat{\beta}) = \exp(0.164) = 1.18$ positive multiplicative effect on $\hat{\mu}$.

5. **Poisson GLM over-dispersion :**

1. defn: in Poisson case, assumption: variance equals mean But sometimes variance >> mean: over dispersion.
2. Caused by Heterogeneity among subjects. if variance equals the mean when all relevant variables are controlled, it exceeds the mean when only a subset of those variables is controlled.

6. **Negative Binomial GLMs:**

1. $f = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1 - \frac{k}{\mu+k}\right)^y, y = 0, 1, 2, \dots$
2. $E(Y) = \mu, \text{Var}(Y) = \mu + \mu^2/k$
3. NBGLM has a larger SE than Poisson

7. **Poisson process:**

Assumptions

1. The prob of at least one occurrence of the event in a given time interval is proportional to the length of the interval
2. the prob of two or more occurrences of the event in a very small time interval is negligible
3. the # of occurrences of the event in disjoint time intervals are mutually independent

Mean: $\mu = \lambda t$

8. TYPE OF GLM:

Random Component	Link	Systematic Component	Model
normal	identity	continuous	regression
normal	identity	categorical	analysis of variance
normal	identity	mixed	analysis of covariance
binomial	logit	mixed	Logistic regression
poisson	log	mixed	loglinear
multinomial	generalized logit	mixed	multinomial response

9. Events over time, or other index of size ----> we are interested in the **rate**

$$\log(\mu_i/t_i) = \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \log(\mu_i) - \log(t_i),$$

$-\log(t_i)$ is called **offset**

10. **LR is better than Wald**

1. Wald: when $\beta = 0$, $z = \frac{\hat{\beta}}{SE}$

2. LR: l_0 under $H_0 : \beta = 0$, l_1 : when β not equal to 0

$$-2\log(l_0/l_1) = -2[\log(l_0) - \log(l_1)] = -2(L_0 - L_1)$$

Under H_0 , test stat has a large-sample chi-squared distribution with $df = 1$

3. $\hat{\pi} = 0.0172 + 0.0198x$: p value < 0.0001 : extremely strong evidence of a positive snoring effect on the probability of heart disease

11. **Deviance of GLM**

1. Definition: $-2 [LM - LS]$

2. Deviance $\sim \chi^2_{N-p}$ N, p are the number of model parameters. Saturated - Unsaturated

3. provides a test of model fit (intuition: test if the the extra parameters are significant)

4. Difference of Two models M_0 and M_1 use $D_0 - D_1 = -2 [L_0 - L_1] \sim \chi^2_{df}$, df = diff in parameters

12. **Fitting using numerical methods:**

1. Iteration continue until there is convergence

1. when gradient of loglikelihood is close to 0

2. estimates do not change after one step

2. $\xi_n = D_n \Upsilon_n$

1. $\Upsilon_n = \frac{\partial \ln L}{\partial \theta_n}$

2. D_n is direction matrix ie how rapidly the gradient is changing

13. **Newton-Raphson** NOTE: technical part will not be tested on final exam

1. Hessian Matrix H as bellow

2.

$\frac{\partial^2 \ln L}{\partial \alpha \partial \alpha}$	$\frac{\partial^2 \ln L}{\partial \alpha \partial \beta}$
$\frac{\partial^2 \ln L}{\partial \beta \partial \alpha}$	$\frac{\partial^2 \ln L}{\partial \beta \partial \beta}$

3. $H_{11} > H_{22}$, the gradient is changing more rapidly as alpha changes as beta changes

4. $\theta_{n+1} = \theta_n - H^{-1} \frac{\partial \ln L}{\partial \theta_n}$

1. Note Inverse 2x2 matrix = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

5. Variance = $(-E(H))^{-1}$ = information matrix

one estimator: $Var(\hat{\theta}) = -(sum H_i)^{-1}$ When second derivative is smaller, the likelihood function is flatter, harder to find the maximum, less confidence. Very flat -> var will be large.

14. ML estimation for GLM: **iteratively reweighted least squares**

Applied Corner to be reviewed.
