#6- F2017

Inference for GLMs

Summary

- Estimation for GLM
- IWLS

Reading

- DB Chapter 5
- MN Chapter 2

Hypothesis Testing in GLMs

- Summarize relationships among elements of the data as parsimoniously as possible.
- Determine the fewer number of, and simplest interpretable form of the predictor variables required to adequately explain the outcome of interest.
- General setting:
  - Full model

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \quad + \beta_q X_{iq} + \beta_{q+1} X_{iq+1} + \quad + \beta_p X_{ip}.$$

  - Reduced model: (The model of interest)

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \quad + \beta_q X_{iq}.$$

Two extrems:

- Maximal or saturated model: Each y_i has a different parameter μ_i or η_i, i.e., there are n parameters with n observations.
- Minimal model: there is only one parameter: a common mean $\mu = \mu_i$, i.e., $\eta_i \equiv \beta_0 \quad i = 1, \dots, n$

#6.

intercept
↓

- Testing H_0: $\underbrace{\beta_{q+1} = \quad = \beta_p}_{p-q \text{ parameters}} = 0$ VS. H_a. not H_0.
- Vector form:

$$\beta = \underbrace{\begin{pmatrix} \beta_1 \rightarrow (q+1)\times 1 \\ \beta_2 \rightarrow (p-q)\times 1 \end{pmatrix}}_{(p+1)\times 1}$$

$\beta_1=(\beta_0, \beta_1, \cdots, \beta_q)^T$

$\beta_2=(\beta_{q+1}, \cdots, \beta_{p+q})^T$

The corresponding H_0,

$$H_0 \quad \beta_2 = \mathbf{c},$$

β_1 contains intercept β_0,
(β_2 does not contain intercept.

where $\mathbf{c}$ is any constant vector (often $\mathbf{c} = \mathbf{0}$).

Sampling Distribution

- MLE $\hat{\beta}$.

$$U(\beta) = \frac{\partial \log L}{\partial \beta}, \quad U(\hat{\beta}) = 0.$$

- Asymptotic distribution of MLE $\hat{\beta}$:

$$\hat{\beta} \sim MVN(\beta, I^{-1}(\beta)). \text{ (approximately)},$$

where $I(\beta)$ is a Fisher information,

$$I(\beta) = Cov(U(\beta)) = -E\left[\frac{\partial^2 \log L}{\partial \beta \partial \beta^T}\right]$$

#6 - F2017

Testing Procedure

- Null hypothesis H_0: $\beta_2 = \mathbf{c}$.
- **Wald Test**
  - Asymptotic distribution of β.

$$\hat{\beta} - \beta \xrightarrow{D} MVN(0, I^{-1}(\beta)),$$

i.e.

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} - \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \xrightarrow{D} MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} I_{11}(\beta) & I_{12}(\beta) \\ I_{21}(\beta) & I_{22}(\beta) \end{bmatrix}^{-1} \right)$$

$$= \begin{bmatrix} * & * \\ * & I_{22.1}^{-1}(\beta) \end{bmatrix}$$

Note: $I_{22.1}^{-1}(\beta) \neq I_{22}^{-1}(\beta)$

Thus,

$$\hat{\beta}_2 - \beta_2 \xrightarrow{D} MVN_{p-q}(0, I_{22.1}^{-1}(\beta)),$$

$= (p+1) - (q+1)$, including intercept

and under H_0,

$$(\hat{\beta}_2 - \mathbf{c})^T I_{22.1}(\hat{\beta}^0)(\hat{\beta}_2 - \mathbf{c}) \xrightarrow{D} \chi^2_{p-q},$$

where $\hat{\beta}^0 = (\hat{\beta}_1^0, \mathbf{c})^T$ with $\hat{\beta}_1^0$=MLE of β_1 under H_0.

β_2 is fixed at $\beta_2 = \mathbf{c}$.

Testing Procedure (cont'd)

- **Score Test**
  - Score function: it is a statistic (function of r.v.)
  - So, a sampling distribution of $U(\beta)$.
    - $E[U(\beta)] = 0$
    - Var-Cov matrix

$$Cov[U(\beta)] = E[U(\beta)U(\beta)^T] = -E\left[\frac{\partial \log L(\beta)}{\partial\beta\partial\beta^T}\right] = I(\beta).$$

Thus, by CLT

$$U(\beta) \xrightarrow{D} MVN(0, \underset{(p+1)\times(p+1)}{I(\beta)}).$$

- Under H_0, Denote $U(\beta) = (U_1^T(\beta), U_2^T(\beta))^T$,

$$U_2(\hat{\beta}^0)^T I_{22.1}^{-1}(\hat{\beta}^0) U_2(\hat{\beta}^0) \xrightarrow{D} \chi^2_{p-q},$$

where $\hat{\beta}^0 = (\hat{\beta}_1^0, \mathbf{c})^T$ with $\hat{\beta}_1^0$ = MLE of β_1 under H_0.

If $p = q = 0$ only one parameter β_0 exists, then $U(\beta_0)^T I^{-1}(\beta_0) U(\beta_0) = U^2(\beta_0)/I(\beta_0) \sim \chi^2_{(1)}$

If $q = 0$, test all parameters, $H_0: \beta = \beta^0$, including intercept β_0, then $U(\beta^0)^T I^{-1}(\beta^0) U(\beta^0) \longrightarrow \chi^2_{(p+1)}$

Example: $Y \sim Bin(n, \pi)$, $f(y;\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$,

$\ell(\pi) = y\log\pi + (n-y)\log(1-\pi) + \log\binom{n}{y}$,

$S(\pi) = \frac{\partial \ell(\pi)}{\partial\pi} = \frac{y - n\pi}{\pi(1-\pi)}$, $\frac{\partial^2\ell(\pi)}{\partial\pi^2} = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2}$

Fisher information is $F(\pi) = E\left(-\frac{\partial^2\ell(\pi)}{\partial\pi^2}\right) = \frac{n}{\pi(1-\pi)} = I(\pi)$.

To test $H_0: \pi = \pi_0$,

1. Wald test: $\chi^2_w(1) = (\hat{\pi} - \pi_0)^T F(\pi_0)(\hat{\pi} - \pi_0) = \frac{n(\hat{\pi} - \pi_0)^2}{\pi_0(1-\pi_0)}$
2. The score test: $\chi^2_S(1) = S^T(\pi_0) F^{-1}(\pi_0) S(\pi_0) = \frac{S^2(\pi_0)}{F(\pi_0)} = \frac{(y - n\pi_0)^2}{n\pi_0(1-\pi_0)} = \frac{n(\hat{\pi} - \pi_0)^2}{\pi_0(1-\pi_0)}$

The two test statistics are the same

Try another parameterization: $\eta = \log\frac{\pi}{1-\pi}$ (log-odds) $\Rightarrow \pi = \frac{e^\eta}{1+e^\eta}$,

$\ell(\eta) = \eta y - n\log(1+e^\eta) + \log\binom{n}{y}$ By the invariance property of MLE,

$\hat{\eta} = \log\frac{\hat{\pi}}{1-\hat{\pi}}$. $S(\eta) = \frac{\partial\ell(\eta)}{\partial\eta} = y - n\frac{e^\eta}{1+e^\eta}$, $\frac{\partial^2\ell(\eta)}{\partial\eta^2} = -n\frac{e^\eta}{(1+e^\eta)^2}$,

information $F(\eta) = E\left(-\frac{\partial^2\ell(\eta)}{\partial\eta}\right) = n\frac{e^\eta}{(1+e^\eta)^2} = n\pi(1-\pi)$. We have

1. The Wald test statistic for $H_0: \eta = \eta_0$ is $\chi^2_\eta = \frac{(\hat{\eta} - \eta_0)^2}{F^{-1}(\eta_0)} = \frac{(\hat{\eta} - \eta_0)^2 n e^{\eta_0}}{(1+e^{\eta_0})^2} = (\hat{\eta} - \eta_0)^2 n\pi_0(1-\pi_0) \neq \frac{(\hat{\pi} - \pi_0)^2}{F^{-1}(\pi_0)} = \chi^2_W(1) \Rightarrow$ Wald test **is not** invariant
2. Score test: $\frac{S^2(\eta_0)}{F(\eta_0)} = \frac{(y - n\pi_0)^2}{n\pi_0(1-\pi_0)} = \frac{n(\hat{\pi} - \pi_0)^2}{\pi_0(1-\pi_0)} = \frac{S^2(\pi_0)}{F(\pi_0)} = \chi^2_S(1) \Rightarrow$ Score test is invariant

Testing Procedure (cont'd)

- **Log-likelihood Ratio (LR) Test**
  - Compare a reduced model (or restricted model under H0) with a full model (or saturated, maximal model)
  - Likelihood Ratio

$$\lambda = \frac{L(\hat\beta_1, \hat\beta_2)}{L(\hat\beta_1^0, \mathbf{c})},$$

where $\hat\beta_1$ and $\hat\beta_2$ = MLEs of β_1, β_2 of full model. The Log-likelihood ratio is defined as

$$\log\lambda = \log L(\hat\beta_1, \hat\beta_2) - \log L(\hat\beta_1^0, \mathbf{c}).$$

* Large value of $\log\lambda$ suggests that the model of interest (the reduced model) is a poor description of the data relative to the full model

To make inference, we need its ($\log\lambda$) sampling distribution.

Testing Procedure (cont'd)

- Sampling distribution of $\log\lambda$.
  - Taylor series expansion of the log-likelihood around MLE $\hat\beta$.

$$l(\beta) \approx l(\hat\beta) + U(\hat\beta)(\beta - \hat\beta) - \frac{1}{2}(\beta - \hat\beta)^T \hat I(\hat\beta)(\beta - \hat\beta).$$

i.e.

$$\begin{aligned} 2[l(\hat\beta) - l(\beta)] &\approx (\beta - \hat\beta)^T \hat I(\hat\beta)(\beta - \hat\beta) \\ &\approx (\beta - \hat\beta)^T I(\hat\beta)(\beta - \hat\beta). \end{aligned}$$

Thus, $2[l(\hat\beta) - l(\beta)] \xrightarrow{D} \chi^2_{p+1}$, (including intercept).

  - Under H_0,

$$2[\log L(\hat\beta) - \log L(\hat\beta_1^0, \mathbf{c})] \xrightarrow{D} \chi^2_{p-q}.$$

where $\hat\beta_1^0$ is MLE of β_1 of under H_0. $(p+1)-(q+1) = p-q$

Calculation of Information

- Need to calculate information matrix for test statistic:
  - From IWLS,

$$U_j(\beta) = \sum_{i=1}^{n} X_{ij} W_i (Z_i - \eta_i).$$

Thus,

$$E\left[\frac{\partial U_j(\beta)}{\partial \beta_k}\right] = -\sum_{i=1}^{n} X_{ij} W_i X_{ik}.$$

i.e.

$$\begin{aligned} I(\beta) &= \mathbf{X}^T \mathbf{W} \mathbf{X} \\ Cov(\hat{\beta}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \end{aligned}$$

$*$ In practice, $Cov(\hat{\beta})$ is estimated by $(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1}$ where $\hat{\mathbf{W}}$ is $\mathbf{W}$ evaluated at $\mu = g^{-1}(\mathbf{X}\hat{\beta})$.

$\mathbb{W} = (W_i)$, $W_i = \frac{1}{Var(Y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} = \frac{1}{Var(\eta_i)[g'(\mu_i)]^2} = \frac{1}{V(\mu_i)a(\phi)[g'(\mu_i)]^2}$, $V(\mu_i) = b''(\theta_i)$

$\hat{W}_i = W_i \big|_{\mu_i = \hat{\mu}_i}$, $\hat{\mathbb{W}} = (\hat{W}_i)$

Goodness of Fit

- Fitting a model: replacing a set of data $\mathbf{Y}$ by a set of fitted values $\hat{\mu}$.
  - How discrepant $\mathbf{Y}$ and μ are? $\longrightarrow$ Goodness of Fit (GOF) test.
  - Full model (or saturated model): n parameters (i.e. $\hat{\mu} = \mathbf{Y}$).
  - Current model: model with $q < n$ parameter is the model of interest. (model of interest)
- Two primary measures for GOF are:
  - Deviance
  - Pearson's χ^2 Test.

Deviance

- **Deviance**: log of a ratio of likelihoods; log-likelihood ratio statistic
- Assume that in exponential family,

$$a_i(\phi) = \left\{ \begin{array}{cc} \phi, & \text{Normal} \\ 1/m_i & \text{Binomial} \\ 1 & \text{Poisson} \end{array} \right\} = \phi/m_i \quad \text{(in general)}.$$

For Binomial and Poisson, $\phi = 1$,

For normal, $\phi = \sigma^2$

Thus,

$$\begin{aligned} \log L(\hat{\beta}) &= \sum_{i=1}^{n} \log f_i(Y_i|\hat{\theta}_i, \phi) \\ &= \sum_{i=1}^{n} \{m_i(Y_i\hat{\theta}_i - b(\hat{\theta}_i))/\phi + c_i(Y_i, \phi)\}. \end{aligned}$$

STAT 635-GLM-Lecture Notes 6, Inference for Generalized Linear Models, Fall 2017

Consider $\log L = l$ as a function of $\hat{\mu}$.

$$l(\hat{\mu}, \phi|\mathbf{Y}) = \sum_{i=1}^{n} \{m_i(Y_i\hat{\theta}_i - b(\hat{\theta}_i))/\phi + c_i(Y_i, \phi)\},$$

it is still a log-likelihood maximized over β for a fixed ϕ.

Deviance (cont'd)

- The **deviance** is defined as: This definition is different from that in the textbook where ϕ is not included in D.

$$D(\mathbf{Y}, \hat{\mu}) = 2\phi\{l(\mathbf{Y}, \phi|\mathbf{Y}) - l(\hat{\mu}, \phi|\mathbf{Y})\}$$
$$= 2\sum_{i=1}^{n} m_i\{Y_i(\tilde{\theta}_i - \hat{\theta}_i) - [b(\tilde{\theta}_i) - b(\hat{\theta}_i)]\},$$

where This does not depend on ϕ, $a(\phi) = \phi / m_i$

  - $l(\mathbf{Y}, \phi|\mathbf{Y})$ = log likelihood for the full model, i.e. $\mu = \mathbf{Y}$ and
  - $\tilde{\theta}_i$ = natural parameter for the full model (i.e., $b'(\tilde{\theta}_i) = Y_i$).
- If model fits well (H_0: No LOF), $\Rightarrow H_0: \beta_2 = (\beta_{q+1}, \dots, \beta_p)^T = 0$ (LOF = Lack of Fit)

$$D \ \phi \sim \chi^2_{n-(q+1)}$$

i.e., if fail to reject Ho, we can use the reduced model $\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq}$.

- Note:
  - The Deviance $D(\mathbf{Y}, \mu)$ does not depend on ϕ.

  - Scaled deviance

$$D^*(\mathbf{Y}, \hat{\mu}) = D(\mathbf{Y}, \hat{\mu})/\phi. = 2\{l(Y, \phi|Y) - l(\hat{\mu}, \phi|Y)\}$$

This is the log-likelihood ratio of the reduced model compared to the full model

Note: This definition is different from that in the D and B's textbook, see p. 80. They call D^* Deviance. By this definition, the LR test and Deviance test based on D^* will be equivalent.

Deviance (Example)

For $N(\mu_i, \sigma^2)$,
$\ell(\mu, \phi | y) = -\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu_i)^2 - \frac{1}{2} n \log(2\pi\sigma^2)$

Let $\hat{\mu}_i = x_i^T\hat{\beta}$, the fitted value of $\mu_i = x_i^T\beta$

- Example 1. $\mathbf{Y} = (Y_1, \quad, Y_n)$ independent from Normal(μ_i, σ^2).
  - $\theta_i = \mu_i$, $b(\theta_i) = \theta_i^2/2$, and $m_i = 1$. $= 2\phi\{\ell(y, \phi|y) - \ell(\hat{\mu}, \phi | y)\}$

$$D(\mathbf{Y}, \hat{\mu}) = 2\sigma^2\left[-\frac{1}{2} n \log(2\pi\sigma^2) - \left\{-\frac{1}{2\sigma^2}\sum_i (y_i - \hat{\mu}_i)^2 - \frac{1}{2} n \log(2\pi\sigma^2)\right\}\right]$$
$$= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$$

- Example 2. $\mathbf{Y} = (Y_1, \quad, Y_n)$ independent from Poisson(λ_i).
  - $\theta_i = \lambda_i$, $b(\theta_i) = \exp(\theta_i)$, and $m_i = 1$. $\phi = 1$ $\quad \lambda = \mu$, $\ell(\mu, \phi | y) = \sum_i y_i \log \lambda_i - \sum_i \lambda_i - \sum_i \log y_i!$

then

$$D(\mathbf{Y}, \hat{\mu}) \quad \ell(y, \phi | y) = \sum_i y_i \log y_i - \sum_i y_i - \sum_i \log y_i!$$
$$\ell(\hat{y}, \phi | y) = \sum_i y_i \log \hat{y}_i - \sum_i \hat{y}_i - \sum_i \log y_i!$$

$$D(y, \hat{\mu}) = 2\phi\{\ell(y, \phi|y) - \log(\hat{y}, \phi|y)\}$$
$$= 2\left\{\sum_{i=1}^n y_i \log(y_i/\hat{y}_i) \quad \sum_{i=1}^n (y_i - \hat{y}_i)\right\}.$$

Example 3

Deviance: Model Comparison

For $y \sim Bin(m, p)$
consider new response $y^* = y/m$, $\phi = 1$, $\omega = m$
$$f(y^*; \theta, \phi) = \exp\left\{\frac{\frac{y}{m}\log\frac{p}{1-p} - \log\frac{1}{1-p}}{1/m} + \log\binom{m}{y}\right\}$$

- Deviance generalized SSE for linear models:
  - Could use for hypothesis testing ($H_0 \quad \beta_2 = \mathbf{c}$).
  - Full model

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}, \quad \hat{\mu} = \hat{\mu}_F.$$

$\ell(\mu, \phi | y^*)$
$$= \sum_i\left\{y_i \log\frac{\mu_i}{1-\mu_i} - m_i \log\frac{1}{1-\mu_i} + \log\binom{m_i}{y_i}\right\}$$
$$= \sum_i\left\{y_i \log \mu_i + (m_i - y_i)\log(1-\mu_i) + \log\binom{m_i}{y_i}\right\}$$

  - Reduced model (under H_0)

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1^0 \\ \mathbf{c} \end{pmatrix}, \quad \hat{\mu} = \hat{\mu}_R.$$

For a saturated model, $\mu_i = y_i/m_i$, then
$$\ell(y, \phi | y) = \sum_i\left\{y_i \log\left(\frac{y_i}{m_i}\right) + (m_i - y_i)\log\frac{m_i - y_i}{m_i} + \log\binom{m_i}{y_i}\right\}$$

For the current model, $\hat{y}_i = m_i\hat{\mu}_i$ is the fitted value
$$\ell(\hat{\mu}, \phi | y) = \sum_i\left\{y_i \log\frac{\hat{y}_i}{m_i} + (m_i - y_i)\log\frac{m_i - \hat{y}_i}{m_i} + \log\binom{m_i}{y_i}\right\},$$
so $D(y, \hat{\mu}) = 2\phi\{\ell(y, \phi|y) - \ell(\hat{\mu}, \phi|y)\}$

$$= 2\sum_{i=1}^n\left\{y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - \hat{y}_i}\right)\right\}.$$

- Linear model with normality and constant variance

$$\frac{SSE(R) - SSE(F)}{\sigma^2} \sim \chi^2_{p-q}.$$

- GLM

$$\frac{D(\mathbf{Y}, \hat{\mu}_R) - D(\mathbf{Y}, \hat{\mu}_F)}{\phi} \sim \chi^2_{p-q}.$$

For Normal distribution $N(X\beta, \sigma^2)$, it is $= \dfrac{\sum_i (y_i - \hat{\mu}_{R_i})^2 - \sum_i (y_i - \hat{\mu}_{F_i})^2}{\sigma^2}$

Pearson's χ^2

- (Generalized) Pearson's χ^2

[Textbook: $$X^2 = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{\widehat{\text{Var}}(Y_i)} = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{a(\phi)V(\hat{\mu}_i)} = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{\phi V(\hat{\mu}_i)/m_i}.$$]

we use $\chi^2 = \sum_{i=1}^n \dfrac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)/m_i} = \sum_{i=1}^n \dfrac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)/m_i}$, to be consistent with definition in Notes #7.

- Example (cont'd)
  - Normal $(m_i = 1,\ V(\mu_i) = 1)$

    [Textbook: $X^2 = \sum_{i=1}^n \dfrac{(y_i - \hat{\mu}_i)^2}{\sigma^2}$] we use $X^2 = \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2$
  - Poisson $(m_i = 1,\ V(\mu_i) = \mu_i)$ $\quad X^2 = \sum_{i=1}^n \dfrac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$
  - Binomial $(\underbrace{y/m_i}_{=y^*},\ m_i,\ V(\mu_i) = \mu_i(1-\mu_i))$ $X^2 = \sum_{i=1}^n \dfrac{(y_i^* - \hat{\mu}_i)^2}{\hat{\mu}_i(1-\hat{\mu}_i)/m_i} = \sum_i \dfrac{(m_i y_i^* - m_i \hat{\mu}_i)^2}{m_i \hat{\mu}_i(1-\hat{\mu}_i)} = \sum_{i=1}^n \dfrac{(y_i - m_i\hat{\mu}_i)^2}{m_i(1-\hat{\mu}_i)\hat{\mu}_i}$

See p.135 ← in D&B 3rd Edt.

- For model based on the Normal distribution, if the nuisance parameter ϕ (σ^2) is not estimated, then the deviance may not be fully determined from the data, F-test can be used to overcome the problem. See p.86 in the textbook.

Interval Estimation

- We obtain point estimates for β using IWLS.
- How to get an interval estimator for β?
  - Interval estimation and hypothesis testing are connected.
  - Inverting test gets you interval estimator
- e.g., 95% CI for β_j using a Wald statistic
  - Testing: H_0: $\beta_j = \beta_j^0$ vs. H_a. $\beta_j \neq \beta_j^0$

$$\text{Test Statistic} \quad W = \frac{(\hat{\beta}_j - \beta_j^0)^2}{\widehat{\text{Var}}(\hat{\beta}_j} \sim \chi_1^2 \qquad \widehat{Var}(\hat{\beta}_j) = \widehat{Var}(\beta_j)\Big|_{\beta_j=\hat{\beta}_j}$$

$1.96^2 = 3.84$

$\|$

Reject H_0, if $|W| > \chi^2_{0.05,1}$ In some textbooks, under H_0, $\widehat{\text{Var}}(\beta_j^0)$ is used instead. In some textbooks, if β is a scalar (only one parameter), under Ho, for the test, use $\widehat{Var}(\beta_j^0)$ given β_j^0 ~~in~~ Ho, instead $\widehat{Var}(\hat{\beta}_j)$. For the CI estimation, use $\widehat{Var}(\hat{\beta}_j)$, β_j^0 is an estimate.

- Inverting this test $\longrightarrow$ 95% CI for β_j

$$\frac{(\hat{\beta}_j - \beta_j)^2}{\widehat{\text{Var}}(\hat{\beta}_j} \leq \chi^2_{0.05,1}$$

i.e.

$$\hat{\beta}_j - \{\chi^2_{0.05,1}\widehat{\text{Var}}(\hat{\beta}_j\ \}^{1/2} \leq \beta_j \leq \hat{\beta}_j + \{\chi^2_{0.05,1}\widehat{\text{Var}}(\hat{\beta}_j)\}^{1/2}$$

- A similar approach will work with the score statistic. See the textbook, P. 75

If there is only one parameter β,

95% CI for β: $\{\beta \quad U^T(\beta)\, I^{-1}(\hat{\beta})\, U(\beta) \leq \chi^2_{0.05}(1)\} = \{\beta.\ U^2(\beta)/I(\hat{\beta}) \leq 3.84\}$

If β is a vector ($p = \dim(\beta) \geq 2$),

95% Confidence region (not CI) for β.

$$\{\beta: \quad U^T(\beta)\, I^{-1}(\hat{\beta})\, U(\beta) \leq \chi^2_{0.05}(p)\}$$

A complicated situation in computing