Logistic Regression

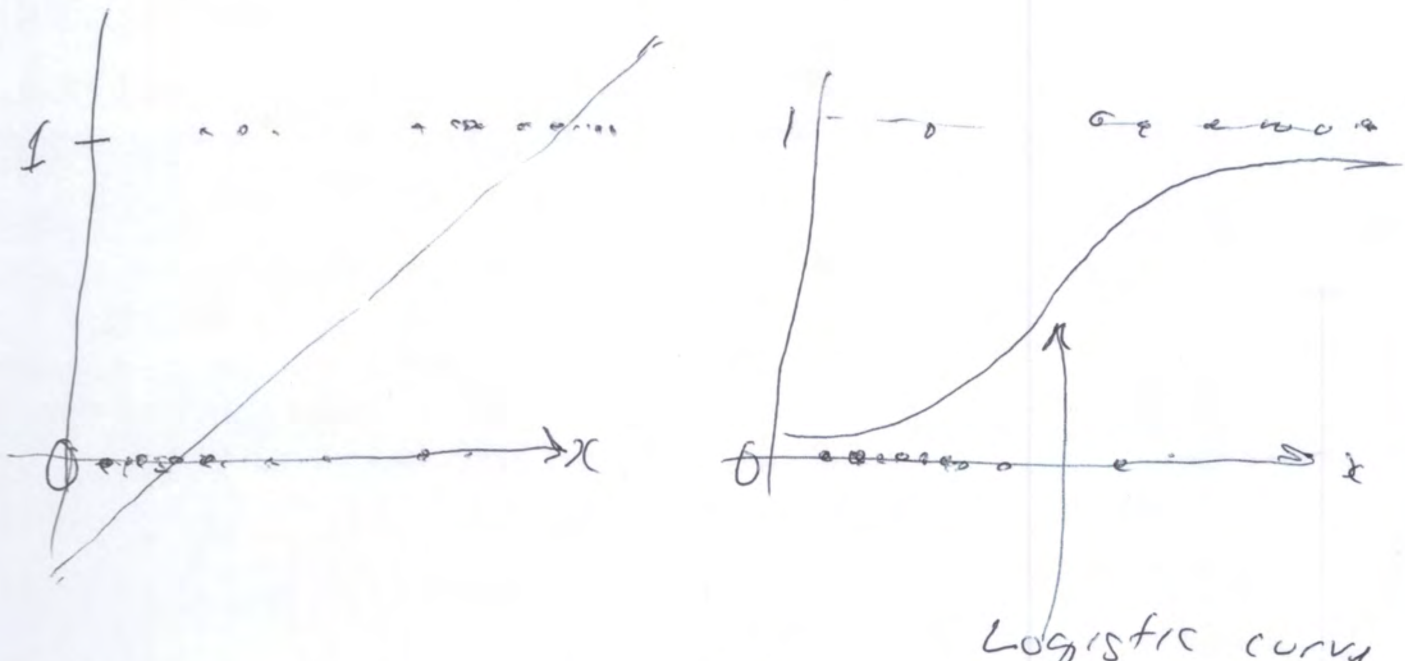
Oct 25 ①

Reading In Ch 3, Skim pp. 65-68, read 68-73
Read Ch 4

Binary Dependent (Response) Variable

$Y=1$ means yes, 0 means no

$$P(Y=1 \mid x_1, \ldots, x_k) = \pi$$



Based on a linear model for LOG ODDS.

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Note π is a conditional probability

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \underset{\sim}{x}'\underset{\sim}{\beta}$$

$$\Leftrightarrow \frac{\pi}{1-\pi} = e^{x'\beta} \Leftrightarrow \pi = e^{x'\beta} - \pi e^{x'\beta}$$

$$\Leftrightarrow \pi + \pi e^{x'\beta} = e^{x'\beta} \Leftrightarrow \pi(1 + e^{x'\beta}) = e^{x'\beta}$$

$$\Leftrightarrow \pi = \frac{e^{x'\beta}}{1+e^{x'\beta}}$$

Increasing function of $x'\beta = \beta_0 + \beta_1 x_1 + \beta_k x_k$

- NON-LINEAR FUNCTION OF THE BETAS
- NON-LINEAR REGRESSION.

$F(x) = \dfrac{e^x}{1+e^x}$ is cumulative distribution function of LOGISTIC DISTRIBUTION

For regression with binary resp var, could use <u>any</u> CDF, $\pi = F(x'\beta)$

For example normal $\longrightarrow$ probit analysis

In terms of log odds, logistic regression is like regular regression. Like hold all other vars constant, increase x_2 by one unit, log odds increases by β_2.

On the odds scale

③

$$\frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}$$
$$= e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}$$

Increase x_2 by one unit, holding x_1 & x_3 constant. Odds are.

$$e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 (x_2+1)} e^{\beta_3 x_3}$$
$$= e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_2} e^{\beta_3 x_3} \leftarrow$$

<u>NOW</u> Look at the ratio of odds when x_2 is increased by one to odds when x_2 is left alone.

Odds Ratio

$$\frac{\cancel{e^{\beta_0}} \cancel{e^{\beta_1 x_1}} \cancel{e^{\beta_2 x_2}} e^{\beta_2} \cancel{e^{\beta_3 x_3}}}{\cancel{e^{\beta_0}} \cancel{e^{\beta_1 x_1}} \cancel{e^{\beta_2 x_2}} \cancel{e^{\beta_3 x_3}}} = e^{\beta_2}$$

Exponential function of regression coeff. is odds ratio: odds of $y=1$ when that x is increased by one unit, compared to leaving it alone.

So for example if you increase smoking by one cigarette a day, holding all other risk factors constant, the odds of death are multiplied by e^{β_4}.

Indicator dummy variables

$X = 1$ means smoker,

$X = 0$ means non-smoker

Log Odds of death $= \beta_0 + \beta_1 x$

Group	x	Odds $= \cancel{\beta_0 + \beta_1 x}$ $e^{\beta_0 + \beta_1 x}$
Smoker	1	$\cancel{\beta_0 + \beta_1}$ $e^{\beta_0} e^{\beta_1}$
Non-smoker	0	$\cancel{\beta_0}$ e^{β_0}

$$\frac{\text{Odds of death given smoker}}{\text{Odds of death given non-smoker}} = \frac{e^{\beta_0} e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Cancer therapy

$$\text{Log Survival odds} = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 x$$

odds of 5-year survival

disease severity

Treatment	d_1	d_2	
Chemotherapy	1	0	$e^{\beta_0} e^{\beta_1} \; e^{\beta_3 x}$
Radiation	0	1	$e^{\beta_0} \; e^{\beta_2} e^{\beta_3 x}$
Both	0	0	$e^{\beta_0} \quad e^{\beta_3 x}$

$$e^{\beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 x}$$

$$= e^{\beta_0} e^{\beta_1 d_1} e^{\beta_2 d_2} e^{\beta_3 x}$$

For any given disease severity x

$$\frac{\text{Survival odds with Chemo}}{\text{Survival odds with both}} = \frac{\cancel{e^{\beta_0}} e^{\beta_1} \, \cancel{e^{\beta_3 x}}}{\cancel{e^{\beta_0}} \quad \cancel{e^{\beta_3 x}}}$$

$$= e^{\beta_1}$$

When x_j is increased by one unit and all other variables remain constant, odds of $Y=1$ are multiplied by e^{β_j} ✓

Maximum Likelihood for logistic Regression

Have, independently for $i = 1, \dots, n$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

$$= x_i'\beta$$

$$\Longleftrightarrow \quad \pi_i = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$$

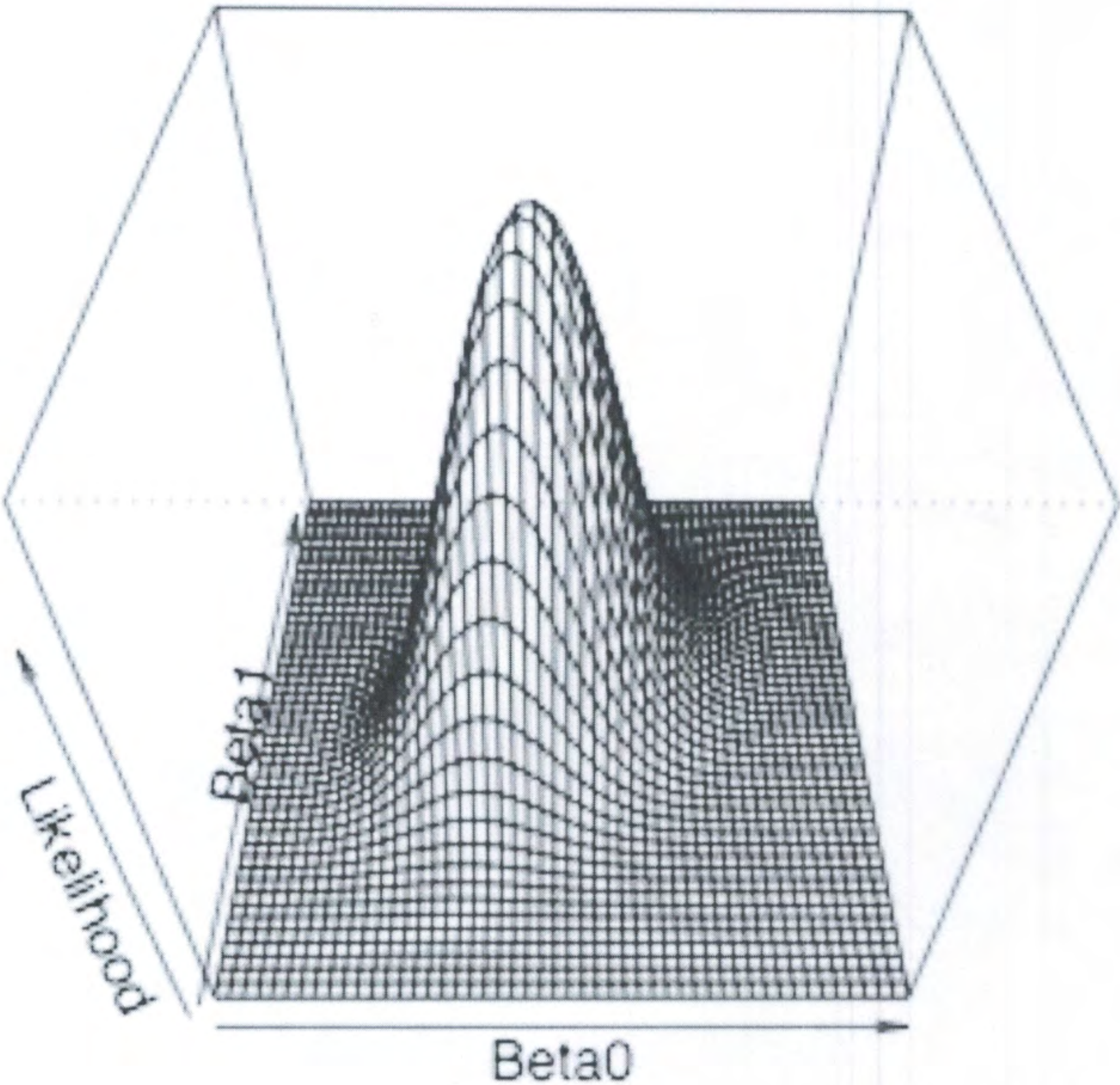
$$y_i \sim \text{Bernoulli}(\pi_i)$$

$$\ell(\underset{\sim}{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right)^{y_i} \left(1 - \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right)^{1-y_i}$$

$$\prod_{i=1}^{n} \left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right)^{y_i} \left(\frac{1 + \cancel{e^{x_i'\beta} - e^{x_i'\beta}}}{1+e^{x_i'\beta}}\right)^{1-y_i}$$

$$= \prod_{i=1}^{n} \frac{e^{x_i'\beta y_i}}{1 + e^{x_i'\beta}}$$


Likelihood
Beta1
Beta0

$$= \frac{e^{\sum_{i=1}^{n} x_i'\beta y_i}}{\prod_{i=1}^{n}(1+e^{x_i'\beta})}$$

Take log, differentiate, no closed-form solution to likelihood equations, Get MLE numerically ("iteratively re-weighted least squares")

For "large" samples MLEs have an approximate multivariate normal distribution. "Asymptotically normal"

Recall $y \sim F_\beta(y)$, $\beta \in \mathcal{B}$, MLE $\hat\beta$ or $\hat\beta_n$

$$\hat\beta_n \mathrel{\dot\sim} N_{k+1}(\beta, V_n)$$

Facts about Multivariate normal. $x \sim N_p(\mu, \Sigma)$

- If $\underset{r\times p}{A}$ is matrix of constants $Ax \sim N_r(A\mu, A\Sigma A')$
- $(x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2(p)$

For logistic regression,

$$Z = \frac{\hat{\beta}_j - \beta_j}{se_{\hat{\beta}_j}} \stackrel{\cdot}{\sim} N(0,1)$$

$se_{\hat{\beta}_j}$ is square root of jth diagonal element of $\hat{V}_n$

Wald Tests for general MLE

Want to test $H_0: \underset{r \times p}{L}\,\underset{p \times 1}{\beta} = \underset{r \times 1}{h}$

For example, in logistic regression

x_1 = HS English mark

x_2 = HS GPA

y = Passed 1st year calculus

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad H_0: \beta_1 = \beta_2 = 0$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{h}$$

$y_i \sim M(1, (\pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5\ \pi_6))$

$H_0: \pi_1 = \cdots = \pi_5 = \frac{1}{6}$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix} = \begin{pmatrix} \frac{1}{6} \\ \vdots \\ \frac{1}{6} \end{pmatrix}$$

Maximum Likelihood for Logistic Reg

Have $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = x_i'\beta$

want to estimate β_j

$$\log \frac{\pi_i}{1-\pi_i} = x_i'\beta \iff \pi_i = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$$

works

$$\ell(\underset{\sim}{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \frac{e^{\sum_{i=1}^{n} x_i'\beta y_i}}{\prod_{i=1}^{n}(1+e^{x_i'\beta})}$$

Said it was hopeless

$$\log \ell(\beta) = \sum_{i=1}^{n} x_i'\beta y_i - \sum_{i=1}^{n} \log\left(1+e^{x_i'\beta}\right)$$

For example, take $x_i'\beta = \beta_0 + \beta_1 x_i$

$$\log \ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n}(\beta_0+\beta_1 x_i)y_i - \sum_{i=1}^{n}\log\left(1+e^{\beta_0+\beta_1 x_i}\right)$$

$$= \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n}\log\left(1+e^{\beta_0+\beta_1 x_i}\right)$$

$$\log l(\beta) = \beta_0 \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \log\left(1 + e^{\beta_0 + \beta_1 x_i}\right)$$ (11)

$$\frac{\partial}{\partial \beta_0} \log l = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \overset{\text{set}}{=} 0$$

$$\frac{\partial}{\partial \beta_1} \log l = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \overset{\text{set}}{=} 0$$

So we have

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

Good luck.

Do it numerically. Minimize Loss Function

$$-\log l(\beta) = \sum_{i=1}^{n} \log\left(1 + e^{x_i' \beta}\right) - \sum_{i=1}^{n} x_i' \beta y_i$$

Get MLE $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_k$

Random variables with joint probability dist.
For large samples, they have an approximate
Multivariate normal distribution.
Say $\hat{\underset{\sim}{\beta}}_n$ is "asymptotically" MV normal

FACTS ABOUT MVN

If $\underset{\sim}{z} \sim N_p(\mu, \Sigma)$

- If $\underset{r \times p}{A}$ is matrix of constants

$$A z \sim N_r(A\mu, A \Sigma A')$$

- $$(z-\mu)' \Sigma^{-1} (z-\mu) \sim \chi^2(p)$$

Compare scalar $\frac{z-\mu}{\sigma}$ standard normal

$$\left(\frac{z-\mu}{\sigma}\right)^2 = \frac{(z-\mu)^2}{\sigma^2} \sim \chi^2(1)$$

$$\hat{\underset{\sim}{\beta}}_n = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \dot{\sim} N_{k+1}(\beta, V_n)$$

Asymptotic covariance matrix

Where, for the record V_n is the inverse of the Fisher information matrix.

$\hat{V}_n$ is easy.

Diagonal elements of $\hat{V}_n$ are estimated variances. Their square roots are standard errors

Using the approximate normality of $\hat{\beta}_n$, get
- z - tests
- Confidence intervals
- Likelihood ratio tests
- Wald tests

Z - tests

Square root of jth diagonal element of $\hat{V}_n$ is $se_{\hat{\beta}_j}$

$$z = \frac{\hat{\beta}_j - \beta_j}{se_{\hat{\beta}_j}} \dot{\sim} N(0,1)$$

Tests of $H_0: \beta_j = 0$ are automatic

Confidence Intervals

$(1-\alpha)100\%$ CI for β_j is $\hat{\beta}_j \pm z_{\alpha/2}\, se_{\hat{\beta}_j}$

As usual

But recall e^{β_j} is odds ratio.

Letting $a = \hat{\beta}_j - z_{\alpha/2}\, se_{\hat{\beta}_j}$

$b = \hat{\beta}_j + z_{\alpha/2}\, se_{\hat{\beta}_j}$

Have $1-\alpha \approx P(a < \beta_j < b)$

$= P(e^a < e^{\beta_j} < e^b)$

odds ratio

Confidence Intervals for conditional Probabilities

$$\hat{\pi}_j = \frac{e^{x_i'\hat{\beta}}}{1+e^{x_i'\hat{\beta}}}$$

Two approaches

1) Because $\hat{\beta}_n \mathrel{\dot\sim} MVN$, and the function is smooth, $\hat{\pi}_i$ is approximately normal and we can figure out the approximate variance

$$\hat{\pi}_i \mathrel{\dot\sim} N(\pi_i, v_i), \text{ \& CI is}$$

$\hat{\pi}_i \pm z_{\alpha/2}\sqrt{\hat{v}_i}$, where $\hat{v}_i$ is a function of V_n

Good, valid as $n \to \infty$, but does not always stay in $(0,1)$

(2) An asymmetric interval that stays in $(0,1)$. (15)

Recall $Az \sim N(A\mu, A\Sigma A')$. In $x_i'\hat{\beta}$, x_i is like A, so

$$x_i'\beta \mathrel{\dot\sim} N(x_i'\beta, \underbrace{x_i' V_n x_i}_{1\times 1}), \text{ and CI}$$

for log odds $x_i'\beta$ is $x_i'\hat{\beta} \pm z_{\alpha/2}\sqrt{x_i' \hat{V}_n x_i}$

Have

$$1-\alpha = P_r\left(x_i'\hat{\beta} - z_{\alpha/2}\sqrt{x_i'\hat{V}_n x_i} < x_i'\beta < x_i'\hat{\beta} + \cdots\right)$$

$$= P_r(a < x_i'\beta < b)$$

$$= P_r\left(\frac{e^a}{1+e^a} < \underbrace{\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}}_{\pi_i} < \frac{e^b}{1+e^b}\right)$$

will definitely stay in $(0,1)$

Likelihood Ratio Tests

Formula sheet: $G^2 = -2 \log \dfrac{l(\hat{\vec{\beta}}_0)}{l(\hat{\vec{\beta}})} \overset{H_0}{\sim} \chi^2(r)$

where $\vec{\beta}_0$ is the estimate of β restricted by some null hypothesis that imposes r restrictions on β.

Usually H_0 is that r of βs $= 0$

R's anova(full, restricted)

Wald tests of $H_0: L\beta = h$

For example if $\frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$H_0: \beta_1 = \beta_2 = 0$

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{h}$$

Derivation of Wald Statistic

$$\hat{\beta}_n \stackrel{.}{\sim} N_{k+1}(\beta, V_n), \text{ so}$$

$$\underset{\substack{\uparrow \\ r \times (k+1)}}{L}\hat{\beta}_n \stackrel{.}{\sim} N_r(L\beta, L V_n L')$$

using $(z-\mu)'\Sigma^{-1}(z-\mu) \sim \chi^2(p)$

$$(L\hat{\beta}_n - L\beta)'(L\hat{V}_n L')^{-1}(L\hat{\beta} - L\beta)$$

If $H_0 : L\beta = h$ is true

$$W_n = (L\hat{\beta}_n - h)'(L\hat{V}_n L')^{-1}(L\hat{\beta} - h)$$

$$\stackrel{.}{\sim} \chi^2(r)$$