

# APPLIED STATISTICS

## Inferential Tools and Variable Selection for Generalised Linear Model

Dr Tao Zou

Research School of Finance, Actuarial Studies & Statistics  
The Australian National University

Last Updated: Fri Sep 29 11:02:41 2017

# Overview

- Maximum Likelihood Estimation (MLE) for Generalised Linear Models
- Hypothesis Testing
  1. z-Test
  2.  $\chi^2$ -Test
- Confidence Intervals
- Variable Selection

# References

1. **F.L. Ramsey and D.W. Schafer** (2012)  
Chapter 20 of *The Statistical Sleuth*
2. The slides are made by **R Markdown**.  
<http://rmarkdown.rstudio.com>

## Review: Generalised Linear Model (GLM)

A generalised linear model (GLM) is a model where the mean of the response is related to the explanatory variables via the following relationship:

$$\overline{g}(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ for } X = (X_1, \cdots, X_k).$$

The function  $g(\cdot)$  is called the link function.

The choice of link function  $g(\cdot)$  depends on the type of the response variable.

We have introduced the link function for two-category/binary response  $Y$  (in such situations the response can be represented by a binary indicator variable taking on values 0 and 1). The link function is the logit link

$$g(u) = \log \left\{ \frac{u}{1-u} \right\}.$$

# Review: Estimation of Binary Logistic Regression Parameters

Given the independent observations for binary response  $Y \in \{0, 1\}$ ,

$$(X_{1,1}, \dots, X_{k,1}, Y_1 = y_1),$$

$\vdots$

$$(X_{1,n}, \dots, X_{k,n}, Y_n = y_n),$$

we can derive the likelihood

$$\mathcal{L}(\beta_0, \dots, \beta_k) = \text{P}(Y_1 = y_1, \dots, Y_n = y_n \mid \text{given all } X\text{s}),$$

which is the probability that we observe  $Y_1 = y_1, \dots, Y_n = y_n$  given all  $X$ s.

For all the other generalised linear models introduced later in this course, we can also obtain the likelihood  $\mathcal{L}(\beta_0, \dots, \beta_k)$  given all  $X$ s.

We only show the derivation of the likelihood for the binary response, since it is complicated for the other generalised linear models. But the idea is very similar.

# Maximum Likelihood Estimation (MLE) for Generalised Linear Models and Its Standard Error

The MLE  $\hat{\beta}_0, \dots, \hat{\beta}_k$  for generalised linear models numerically maximize the likelihood  $\mathcal{L}(\beta_0, \dots, \beta_k)$ .

The standard errors of MLE  $\hat{\beta}_j, j = 0, \dots, k$  can be given if the corresponding generalised linear model is correctly specified.

The formula is complicated and will not be introduced in this course. If you are interested, please enroll STAT3013 Statistical Inference course.

Moreover, the standard error formula can only be given **when the sample size  $n$  is large enough**. Sometimes we call it approximate standard error  $SE_a(\hat{\beta}_j)$ .  
 $\downarrow n > 30, \text{ so } \dots$

The summary() output of R shows the standard errors.

# Practical Sampling Distributions of $\hat{\beta}_j$ , $j = 0, \dots, k$

If the generalised linear model is correctly specified, it can be shown mathematically that

$$\frac{\hat{\beta}_j - \beta_j}{\text{SE}_a(\hat{\beta}_j)} \overset{a}{\sim} N(0, 1) \text{ for } j = 0, \dots, k.$$

*Handwritten notes:*  $\hat{\beta}_j$  is the MLE;  $\text{SE}_a(\hat{\beta}_j)$  is the output of R.

Here "a" means that it is an approximation when the sample size  $n$  is large enough.

It is worth noting that we no longer have the  $t$  distribution, but the approximate standard normal distribution  $N(0, 1)$  for the generalised linear model.

This practical sampling distribution indicates that the MLE estimators are roughly unbiased. *roughly normal distributed.*

This theory leads directly to tests and confidence intervals for individual regression coefficients, in the familiar way.

When the sample size  $n$  is small, we will introduce bootstrap method later to obtain the standard error and the practical sampling distribution.

## Hypothesis Testing for $\beta_j$ : z-Test

$\searrow$  t-Test

$$H_0 : \beta_j = 0 \leftrightarrow H_a : \beta_j \neq 0.$$

$$\text{Test Statistic} = TS = \frac{\hat{\beta}_j - 0}{SE_a(\hat{\beta}_j)} \rightarrow t \text{ statistic}$$

which should be compared to the  $N(0, 1)$  distribution.  $\rightarrow$  CLM

The  $p$ -value is

$$p\text{-value} = 2 \times P(\underline{Z} > |TS|), \text{ where } Z \sim N(0, 1).$$

If  $p\text{-value} < \alpha \Rightarrow$  reject  $H_0$ ;  $p\text{-value} \geq \alpha \Rightarrow$  not reject  $H_0$ .



## Drop-in-Deviance $\chi^2$ -Test

Similar to the  $F$ -test in MLR,  $\chi^2$ -Test is used to test whether or not a **subgroup** of  $\beta_j$ ,  $j = 1, \dots, k$  (**several** regression coefficients) in GLM are all zeros, e.g., for the model

$$g(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

we consider the test *example of  $\chi^2$ -Test :*

$$H_0 : \beta_1 = \beta_3 = 0 \leftrightarrow H_a : \text{at least one of } \beta_1 \text{ and } \beta_3 \text{ is not } 0.$$

**Remark:** compare to the  $z$ -test  $H_0 : \beta_j = 0$  (whether or not a **single**  $\beta_j$  in GLM is zero).

$z$ -tests cannot be used to test a hypothesis involving more than one parameter.

In order to propose the test statistic for the  $\chi^2$ -test, we introduce the following terminology.

# Deviance

The deviance for a GLM

$$g(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ where } X = (X_1, \cdots, X_k),$$

is defined by

$$\text{deviance} = \underline{-2 \log} \left[ \mathcal{L}(\hat{\beta}_0, \cdots, \hat{\beta}_k) \right]$$

where  $\hat{\beta}_0, \cdots, \hat{\beta}_k$  are MLE of the above GLM.

Deviance measures the goodness of fit for GLM. *numerically max  $\mathcal{L}(\beta_0 \cdots \beta_k)$*

Similar to SSE in MLR, the smaller the deviance is, the better fitting of a model.

## Full Model and Reduced Model

For ~~MLR~~ GLM

$$g(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \text{ where } X = (X_1, X_2, X_3),$$

the following  $\chi^2$ -test is considered

$$H_0 : \beta_1 = \beta_3 = 0 \leftrightarrow H_a : \text{at least one of } \beta_1 \text{ and } \beta_3 \text{ is not 0;}$$

$\Leftrightarrow$

$$H_0 : \text{reduced model } g(\mu\{Y|X_2\}) = \beta_0 + \beta_2 X_2 \text{ is appropriate} \leftrightarrow$$

$$H_a : \text{full model } g(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \text{ is appropriate.}$$

This is in contrast to the  $F$ -test adopted for MLR.

We use  $d$  to denote the number of  $\beta$ s being tested. Here  $d = 2$ . Note that

$d$  = number of regression coefficients in the full model

– number of regression coefficients in the reduced model.

# Drop in Deviance

We can compute deviance<sub>full</sub> for the full model and deviance<sub>reduced</sub> for the reduced model.

One can verify that

$$\text{deviance}_{\text{full}} \leq \text{deviance}_{\text{reduced}}.$$

That means the full model has a better fitting compared to the reduced model.

Drop in deviance is defined by

*Large*  $\nearrow$

$\text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}}$

$\nwarrow$  *Small*

The drop in deviance provides an indication of the importance of the variables that have been excluded from the full model.)

## Drop-in-Deviance $\chi^2$ -Test

The  $\chi^2$ -test statistic is

$$TS = \text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}}.$$

page 7

This test statistic should be compared to approximately  $\chi_d^2$  distribution, where the number of degrees of freedom  $d$  is the number of  $\beta$ s being tested. Note that

$d$  = number of regression coefficients in the full model

– number of regression coefficients in the reduced model.

Here “approximately” means that it is an approximation **when the sample size  $n$  is large enough.**

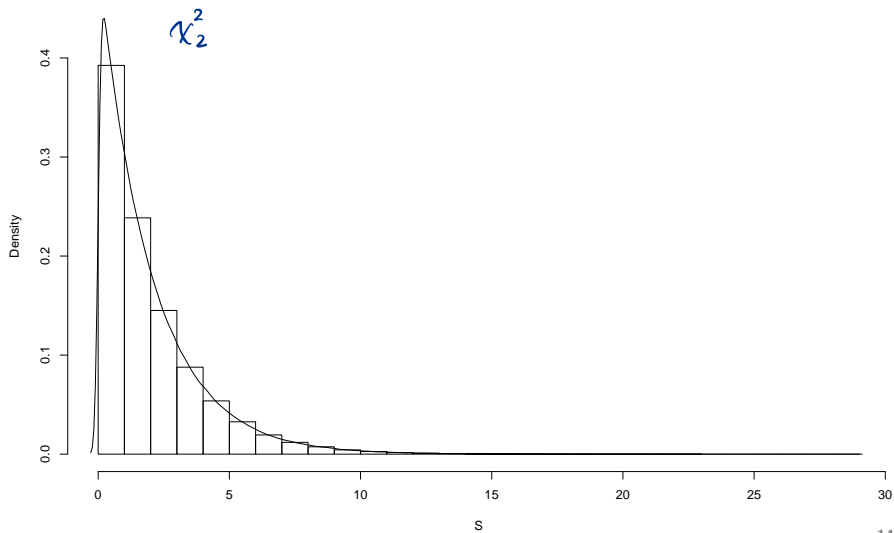
The  $p$ -value is

$$p\text{-value} = P(\underline{S} > TS), \text{ where } S \sim \chi_d^2.$$

If  $p\text{-value} < \alpha \Rightarrow$  reject  $H_0$ ;  $p\text{-value} \geq \alpha \Rightarrow$  not reject  $H_0$ .

# $\chi^2$ Distribution

Chi-Squared Distribution



## Special Cases of $\chi^2$ -Tests

Consider GLM

$$g(\mu\{Y|X\}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \text{ where } X = (X_1, X_2, X_3).$$

Testing the “overall significance”:

The following  $\chi^2$ -test

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \leftrightarrow H_a : \text{otherwise}$$

involves testing whether all the coefficients except  $\beta_0$ , equal to 0 or not.

This **null** hypothesis proposes that none of the considered explanatory variables is useful in explaining the mean of response.

The deviances for this test are given in the R summary(glm()) output.

# Special Cases of $\chi^2$ -Tests (Con'd)

## Testing a single coefficient:

$\chi^2$ -test can also be used to test

$$H_0 : \beta_j = 0 \leftrightarrow H_a : \beta_j \neq 0.$$

Test Statistic =  $TS = \text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}}$

which should be compared to the  $\chi_d^2$  distribution, where  $d = 1$ .

For testing that a single coefficient is zero, the  $\chi^2$ -test and z-test are **similar**, i.e., they return similar p-values. Usually the significance/insignificance conclusions by using the two tests are the same.

*approximation*



## Confidence Intervals (CI) for $\beta_j$

Recall the practical sampling distributions of  $\hat{\beta}_j$ :

$$\frac{\hat{\beta}_j - \beta_j}{\text{SE}_a(\hat{\beta}_j)} \overset{a}{\sim} N(0, 1), \text{ for } j = 0, \dots, k.$$

Using this information, a  $(1 - \alpha)$  CI for  $\beta_j$  is  $\alpha = 0.05$   
 $z_{0.025} = 1.96$

$$\hat{\beta}_j \mp z_{\alpha/2} \times \text{SE}_a(\hat{\beta}_j)$$

where  $z_{\alpha/2}$  is the  $1 - \alpha/2$  quantile of  $N(0, 1)$ , namely

$$P(Z \leq z_{\alpha/2}) = 1 - \alpha/2 \text{ or } P(Z > z_{\alpha/2}) = \alpha/2$$

for  $Z \sim N(0, 1)$ .

## Idea of Variable Selection in GLM

Similar to variable selection in MLR, recall that we pursue good fitting for a GLM, and deviance measures the goodness of fit for GLM.

Based on the definition of deviance, the smaller the deviance is, the better fitting of a model.

Hence, one goal for GLM is to find an appropriate model with smaller deviance.

However, the goal for variable selection is to find a small number of explanatory variables if possible to construct GLM.

The above two goals are contradicted since more explanatory variables results in the decrease in deviance.

Hence, an appropriate variable selection criterion should provide a compromise between how well the model fits the data and the number of explanatory variables  $\Rightarrow$  a standard to determine statistical measures.

# Idea of Variable Selection by Using AIC and BIC

AIC (Akaike Information Criterion) and BIC (Bayesian) can be considered.  
For the following GLM

$g(\mu\{Y|X_1, \dots, X_j\}) = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j$ , we define

$$\text{AIC} = \text{deviance} + 2 \times (j + 1) \text{ and}$$

$$\text{BIC} = \text{deviance} + \log(n) \times (j + 1).$$

For two candidate models, if their number of the explanatory variables  $j$  is the same, then the model with smaller deviance, or equivalently **smaller AIC (or BIC)**, is preferred.

For two candidate models, if their deviance is the same, then the model with smaller number of explanatory variables  $j$ , or equivalently **smaller AIC (or BIC)**, is preferred.

Hence, the variable selection criterion based on AIC (or BIC) is: the model with **smaller AIC (or BIC) is preferred**.

## Idea of Variable Selection by Using AIC and BIC (Con'd)

AIC and BIC both compromise how well the model fits the data (deviance) and the number of explanatory variables ( $j$ ).

Compared to AIC, BIC assigns a larger weight to the number of explanatory variables  $j$  in its expression (usually the sample size  $n$  is large such that  $\log(n) > 2$ ).

Hence BIC usually prefers the model with less explanatory variables compared to AIC.

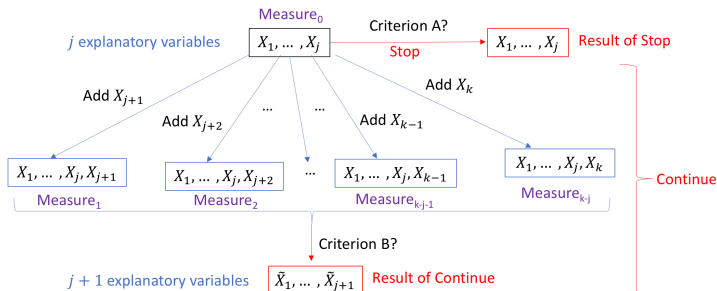
In the following, we will introduce the sequential variable selection procedure based on either of the above two measures.

Let Measure = AIC, or Measure = BIC in the following.

Then, the variable selection criterion based on "Measure" is: the model with **smaller "Measure" is preferred**.

# Forward Selection Step by Using AIC and BIC

## Forward Selection Step for $j$ Explanatory Variables



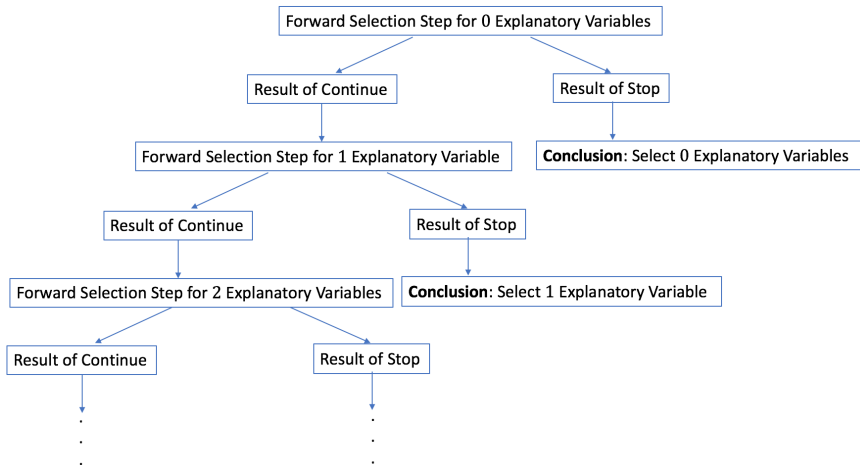
Suppose we have  $k$  explanatory variables  $X_1, \dots, X_k$  in total.

### Criterion A: if

$\min\{\text{Measure}_1, \text{Measure}_2, \dots, \text{Measure}_{k-j-1}, \text{Measure}_{k-j}\} > \text{Measure}_0$  or  $j = k$ , then Stop; otherwise Continue.

**Criterion B:**  $\tilde{X}_1, \dots, \tilde{X}_{j+1}$  are those variables such that the corresponding "Measure" is the smallest.

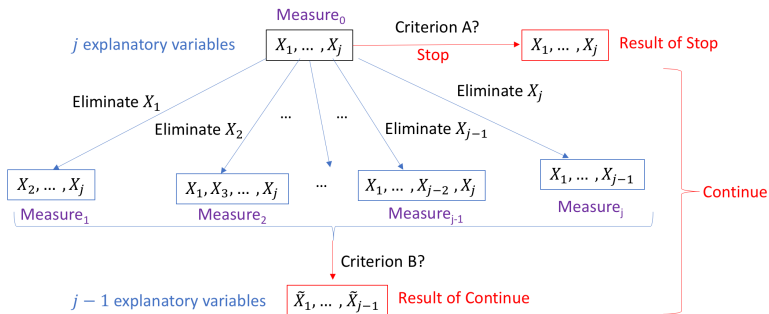
# Complete Forward Selection Procedure by Using Other Statistical Measures



Keep doing on the above procedures, until the first time we obtain the **Result of Stop**.

# Backward Elimination Step by AIC and BIC

## Backward Elimination Step for $j$ Explanatory Variables

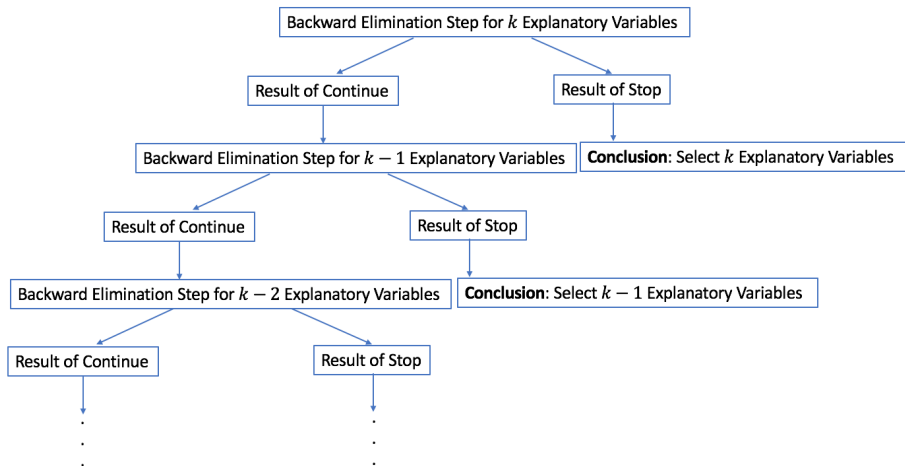


### Criterion A: if

$\min\{\text{Measure}_1, \text{Measure}_2, \dots, \text{Measure}_{j-1}, \text{Measure}_j\} > \text{Measure}_0$  **or**  $j = 0$ ,  
then Stop; otherwise Continue.

**Criterion B:**  $\tilde{X}_1, \dots, \tilde{X}_{j+1}$  are those variables such that the corresponding  
“Measure” is the smallest.

# Complete Backward Elimination Procedure by AIC and BIC



Keep doing on the above procedures, until the first time we obtain the **Result of Stop**.



# Stepwise Selection Procedure by AIC and BIC

This constitutes a combination of backward elimination and forward selection. Each step consists of the following steps. (Any starting model can be chosen, e.g., the model with no explanatory variables, or the model with all the explanatory variables.)

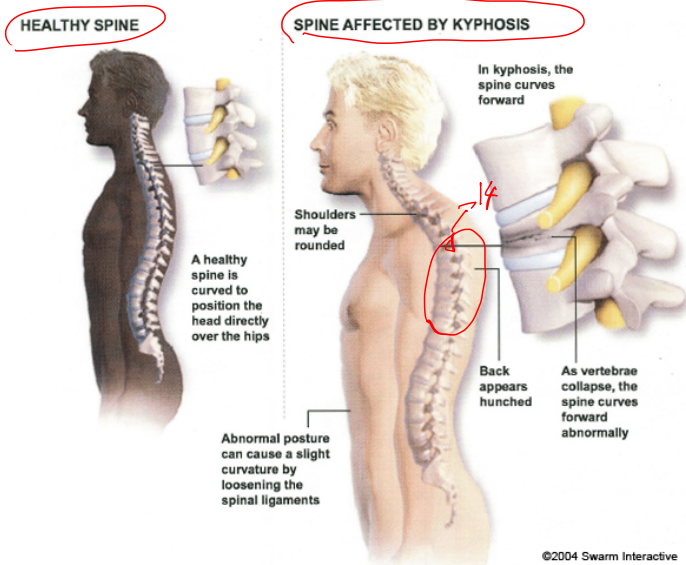
1. Do one Forward Selection Step.
2. Do one Backward Elimination Step.

Repeat Steps 1 and 2 until no explanatory variables can be added or removed.

**Remark:** The above three methods will sometimes lead to different variable selection results.

## Example: Kyphosis

The R library “rpart” contains a built in dataset called “kyphosis”. This dataset contains data on 81 children who have had corrective spinal surgery.



## Example: Kyphosis (Con'd)

For each individual there are four variables:

- Kyphosis: a binary response indicating whether kyphosis (a spinal condition) was present after the operation.
- Age: the age of the child in months.
- Number: the number of vertebrae involved in the spinal operation.
- Start: the beginning of the range of the vertebra involved in the operation.

In this example we will see whether there is a relationship between whether a child has kyphosis after the operation and the covariates Age, Number, and Start.

We use this binary response example to show the inferences of GLM. The inferences for other GLMs introduced later in this course are similar and we will not talk about the above theory for inferences again.

# R Code

```
library(rpart) #install.packages("rpart")
```

```
## Warning: package 'rpart' was built under R version 3.3.2
```

```
names(kyphosis)
```

```
## [1] "Kyphosis" "Age"      "Number"    "Start"
```

```
pairs(kyphosis[,-1]) #pairwise scatterplots for the covariates
```



## R Code (Con'd)

```
attach(kyphosis)
indKyp=ifelse(Kyphosis=="present",1,0)
kyph.logit=glm(indKyp~Age+Number+Start,family=binomial(link=logit))
summary(kyph.logit)
```

$X_1$   $X_2$   $X_3$

```
##
## Call:
## glm(formula = indKyp ~ Age + Number + Start, family = binomial(link = logit))
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.3124  -0.5484  -0.3632  -0.1659   2.1613
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.036934   1.449575  -1.405  0.15996
## Age          0.010930   0.006446   1.696  0.08996 .
## Number       0.410601   0.224861   1.826  0.06785 .
## Start       -0.206510   0.067699  -3.050  0.00229 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
## Null deviance: 83.234 on 80 degrees of freedom
```

```
## Residual deviance: 61.380 on 77 degrees of freedom
```

```
## AIC: 69.38
```

```
##
```

```
## Number of Fisher Scoring iterations: 5
```

MLE  $SE_a(\beta_j)$

TS in z-test

## R Code (Con'd)

$$H_0 : \beta_1 = \beta_3 = 0 \leftrightarrow H_a : \text{otherwise.}$$

```
#drop in deviance test
#reduced model
kyph.logitr=glm(indKyp~Number,family=binomial(link=logit))
anova(kyph.logitr,kyph.logit,test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: indKyp ~ Number
```

```
## Model 2: indKyp ~ Age + Number + Start
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1          79      73.357      0      73.357 0.000000e+000
## 2          77      61.380      2      11.977 0.002507e+000 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

125

## R Code (Con'd)

$$H_0 : \underline{\underline{\beta_1}} = \underline{\underline{\beta_2}} = \underline{\underline{\beta_3}} = 0 \leftrightarrow H_a : \text{otherwise.}$$

*#drop in deviance test*

*#reduced model*

```
kyph.logitr=glm(indKyp~1,family=binomial(link=logit))  
anova(kyph.logitr,kyph.logit,test='Chisq')
```

## Analysis of Deviance Table

##

## Model 1: indKyp ~ 1

## Model 2: indKyp ~ Age + Number + Start

##    Resid. Df Resid. Dev Df Deviance Pr(>Chi)

## 1	80	83.234			
## 2	77	61.380	3	21.855	6.994e-05 ***

## ---

## Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*Compare to  
Page 29*

*deviance reduced*

*deviance full*

## R Code (Con'd)

$$H_0 : \beta_3 = 0 \leftrightarrow H_a : \text{otherwise.}$$

```
#drop in deviance test  
#reduced model  
kyph.logitr=glm(indKyp~Age+Number,family=binomial(link=logit))  
anova(kyph.logitr,kyph.logit,test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: indKyp ~ Age + Number
```

```
## Model 2: indKyp ~ Age + Number + Start
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1          78      71.627
```

```
## 2          77      61.380  1    10.247 0.001369 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The data suggests that the coefficient of Start is significantly different from 0. Increases in Start are associated with a decrease in the probability that kyphosis is present after the operation.



## R Code (Con'd)

The estimated coefficient of Start is

```
summaryresult=summary(kyph.logit)  
summaryresult$coefficients[4,1] → MLE
```

```
## [1] -0.20651
```

page 29

The corresponding approximate standard error is

```
summaryresult$coefficients[4,2] → SE
```

```
## [1] 0.06769886
```

page 29

A 95% CI for the coefficient of Start ( $\beta_3$ ) is 1.96

```
summaryresult$coefficients[4,1] - 1.96 * summaryresult$coefficients[4,2]  
summaryresult$coefficients[4,1] + 1.96 * summaryresult$coefficients[4,2]
```

## [1] -0.3391974

$\beta_3$        $z_{0.025}$        $SE(\hat{\beta}_3)$

```
summaryresult$coefficients[4,1] - 1.96 * summaryresult$coefficients[4,2]  
summaryresult$coefficients[4,1] + 1.96 * summaryresult$coefficients[4,2]
```

## [1] -0.07382272

## R Code (Con'd)

Recall the interpretation of  $\beta_3$  is: with the other variables held constant, if Start is increased by 1 unit, ~~the odds~~ that  $Y = 1$  will change by a multiplicative factor of  $e^{\beta_3}$ .

It might also be of interest to calculate a 95% CI for  $e^{\beta_3}$ , which gives the multiplicative amount that the odds of  $Y = 1$  if Start is increased by 1 unit. A 95% CI for  $e^{\beta_3}$  is obtained by exponentiating the 95% CI for  $\beta_3$ , which is

```
exp(summaryresult$coefficients[4,1]-qnorm(0.975)*summaryresult$coefficients[4,2])
```

⏏

```
## [1] 0.7123418
```

```
exp(summaryresult$coefficients[4,1]+qnorm(0.975)*summaryresult$coefficients[4,2])
```

⏏

```
## [1] 0.9288363
```

## R Code (Con'd)

Note that Age and Number are not significant. Are these variables required?

$$H_0 : \beta_1 = \beta_2 = 0 \leftrightarrow H_a : \text{otherwise.}$$

```
kyph.logitr=glm(indKyp~Start,family=binomial(link=logit))  
anova(kyph.logitr,kyph.logit,test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: indKyp ~ Start
```

```
## Model 2: indKyp ~ Age + Number + Start
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1          79      68.072
```

```
## 2          77      61.380  2    6.6923 0.03522 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The output from the  $\chi^2$ -test shows that at least one of these two explanatory variables is needed. We then try the variable selection method.

## Example: R Code (Con'd)

```
#install.packages('MASS')  
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.3.2
```

```
#Backward AIC  
a=stepAIC(kyph.logit, direction="backward")
```

```
summary(a)
```

```
##  
## Call:  
## glm(formula = indKyp ~ Age + Number + Start, family = binomial(link = logit))  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.3124  -0.5484  -0.3632  -0.1659   2.1613   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)      
## (Intercept) -2.036934   1.449575  -1.405  0.15996      
## Age          0.010930   0.006446   1.696  0.08996 .      
## Number       0.410601   0.224861   1.826  0.06785 .      
## Start       -0.206510   0.067699  -3.050  0.00229 **     
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##      Null deviance: 83.234  on 80  degrees of freedom  
## Residual deviance: 61.380  on 77  degrees of freedom  
## AIC: 69.38  
##  
## Number of Fisher Scoring iterations: 5
```

## Example: R Code (Con'd)

```
#Backward BIC
n=length(indKyp)
a=stepAIC(kyph.logit,direction="backward", k=log(n))
```

```
summary(a)
```

```
##
## Call:
## glm(formula = indKyp ~ Start, family = binomial(link = logit))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4729  -0.5176  -0.4211  -0.3413   2.1305
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.89007     0.62996   1.413 0.157686
## Start       -0.21789     0.06044  -3.605 0.000312 ***
## --
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 83.234  on 80  degrees of freedom
## Residual deviance: 68.072  on 79  degrees of freedom
## AIC: 72.072
##
## Number of Fisher Scoring iterations: 5
```

## Example: R Code (Con'd)

```
#Stepwise AIC  
a=stepAIC(kyph.logit,direction="both")
```

```
summary(a)
```

```
##  
## Call:  
## glm(formula = indKyp ~ Age + Number + Start, family = binomial(link = logit))  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.3124  -0.5484  -0.3632  -0.1659   2.1613   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)      
## (Intercept) -2.036934   1.449575  -1.405  0.15996      
## Age          0.010930   0.006446   1.696  0.08996 .      
## Number      0.410601   0.224861   1.826  0.06785 .      
## Start       -0.206510   0.067699  -3.050  0.00229 **     
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 83.234  on 80  degrees of freedom  
## Residual deviance: 61.380  on 77  degrees of freedom  
## AIC: 69.38  
##  
## Number of Fisher Scoring iterations: 5
```

```
detach(kyphosis)
```