# STAT2014/6014 Lecture Week 7 Binary regression with logistic link

Lucy Hu
Email: yunxi.hu@anu.edu.au

[1]Research School of Finance, Actuarial Studies and Statistics
Australian National University

# Overview

**Week 7: Binary regression with logistic link**

- Motivation
- Binary regression with logistic link
  - Mean and variance
  - Canonical link function and inverse canonical link function
  - Odds ratio and interpretation of $\beta_j$
  - MLE (Maximum likelihood estimator)
  - Prediction (Example 1)
  - Deviance
  - Hypothesis testing (Example 4)

What if. . .

- The response variable $Y_i$ is a categorical variable which only takes **two possible values 0 and 1 (i.e. a binary variable)**. For example:
  - In a study on the effectiveness of advertising, the response might be whether a given customer is willing to buy the new product.
  - In a study of home ownership, the response variable is whether a given individual owns a home.
- The response variable $Y_i$ is **bernoulli (binary) distributed**.

# T1 - Binary regression with logistic link

- $Y_i \sim Bin(M_i = 1, \pi_i) = Bern(\pi_i)$ (Bernoulli/Binary distribution).
- $Y_i$ is discrete (categorical), which has two levels.
- $Y_i = 0$ (a failure under one trial) or 1 (a success under one trial).
- Non-aggregate data: $(X_i, Y_i)$.

# Example 1: Calculus Data (Non-aggregate data)

- Suppose there are **30** students attempting a Calculus quiz question.
- We are interested in studying the relationship between the study time that each student spends in Calculus (in hours) and whether one student can answer the Calculus quiz question correctly.
- Assume there is no partial correct answer for this quiz question. i.e., one student's quiz question answer is either right or wrong.
- The independent variable ($X_i$) is **Time** (i.e. the study time that each student spends (in hours) in Calculus).
- The **response variable ($Y_i$)** is an indication of whether one student's quiz question answer is right or wrong.
  - $Y_i = 1$ if one student's quiz question answer is right.
  - $Y_i = 0$ if one student's quiz question answer is wrong.

```r
calculus <-read.table("calculus.csv", header = TRUE, sep = ",")
attach(calculus)
# n=30, X=Time, Y = 1/0 = Answer the question correctly/not
length(calculus$Time)
```

```
## [1] 30
```

```r
head(calculus,n=14)
```

```
##         Time Response
## 1  1.2500000        1
## 2  1.2500000        1
## 3  1.2500000        1
## 4  1.2500000        1
## 5  1.2500000        1
## 6  1.2500000        1
## 7  1.2500000        0
## 8  1.0000000        1
## 9  1.0000000        1
## 10 1.0000000        1
## 11 1.0000000        1
## 12 1.0000000        0
## 13 0.8333333        1
## 14 0.8333333        1
```

# Example 1: Calculus Data (Non-aggregate data)

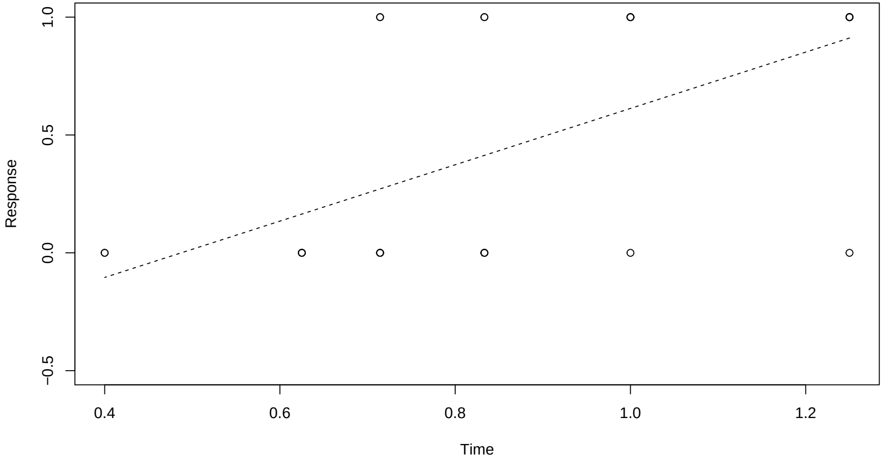
- This is an example of **non-aggregate** data.
- There are 30 students in the sample, each student = one observation, $n = 30$.
- Data: $(X_i = Time, Y_i = 0/1)$.
- $Y_i = 0/1$ indicator under one trial.
    - indication of whether one student's quiz question answer is right.
    - $Y_i = 1$ if answer is right or 0 if wrong.
    - Two levels under 1 trial (whether one student's answer is right/wrong).
- $Y_i$ **(under 1 trial, given** $X_i = x$**)** $\sim Bin(1, \pi_i) = Bern(\pi_i)$.

# Issues

Why can't we use linear regression?

- Issue 1: $Y_i = 0/1$ only, which is not normally distributed.
- Issue 2: For LR, $\mu_{Y_i} = \beta_0 + \beta_1 X_{1_i} + ... + \beta_k X_{k_i} = x_i^T \beta$. + Range for LHS: [0,1]. + Range for RHS: Real number. + Not match!

```r
# For Issue 1 => Create plot between Xi and Yi - points
plot(Time,Response,ylim=c(-0.5,1))
# If use LR: mu_Y = B0+B1*Xi - dashed line
fit<-lm(Response~Time); lines(Time,fit$fitted,lty=2)
```



```
# Issue 2: if use LR: mu_Y = B0+B1*Xi to fit the data:
# LHS proxy: sample mean of Y under each Xi level
# Since all Y=0/1 -> its sample mean belongs to [0,1]
tapply(Response,Time,mean)
```

```
##         0.4       0.625 0.714285714 0.833333333           1        1.25
##   0.0000000   0.0000000   0.3333333   0.3333333   0.8000000   0.8571429
```

```
# RHS proxy: yi hat = bo+b1*Xi => can be >0 or <0
summary(fit)$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -0.5824669  0.2698848 -2.158205 0.0396358463
## Time         1.1951073  0.2953612  4.046257 0.0003711471
```

```
Time1<-c(0.4,0.625,0.714285714,0.833333333,1,1.25)
summary(fit)$coefficients[1,1]+summary(fit)$coefficients[2,1]*Time1
```

```
## [1] -0.1044239  0.1644752  0.2711813  0.4134559  0.6126405  0.9114173
```

```
# Not fit!May need LHS transformation: from mu_Y to g(mu_Y)
```

## $E(Y_i), V(Y_i)$ for Binary variable $Y_i \sim Bin(1, \pi_i) = Bern(\pi_i)$

- Assume there are $n$ number of observations in the sample, each observation $(X_i, Y_i)$ is independent.
- There are $n$ number of $Y_i$, where $Y_i \sim Bin(1, \pi_i) = Bern(\pi_i)$ (i.e. by changing rows, the $\pi_i$ value would change.).
- $\pi_i = \pi_i | X = P(Y_i = 1|X)$, where $X = (X_{1_i}, ...X_{k_i})$ (k different types of X for $i^{th}$ observation), and $i = 1, 2, ..., n$.
- For each $Y_i$ (there are $n$ number of $Y_i$ in total), it can be either 0 or 1
    - Scenario 1: $P(Y_i = 1|X) = \pi_i = \pi_i^1 \times (1 - \pi_i)^{1-1}$, where $Y_i = 1$.
    - Scenario 2: $P(Y_i = 0|X) = 1 - \pi_i = \pi_i^0 \times (1 - \pi_i)^{1-0}$, where $Y_i = 0$.
    - Therefore, the PMF (probability mass function) is $P(Y_i = y_i) = f(y_i) = \pi_i^{y_i} \times (1 - \pi_i)^{1-y_i}$.
    - $E(Y_i) = 1 \times \pi_i + 0 \times (1 - \pi_i) = \pi_i = \pi_i \in [0, 1]$.
    - $V(Y_i) = (1 - \pi_i)^2 \times \pi_i + (0 - \pi_i)^2 \times (1 - \pi_i) = \pi_i \times (1 - \pi_i)$.

# The canonical link function and the inverse link function

If $Y_i \sim Bin(1, \pi_i) = Bern(\pi_i)$

- [1] The canonical link function:
  $\theta_i = g(\mu_{Y_i}) = \ln(\frac{\mu_{Y_i}}{1-\mu_{Y_i}}) = \ln(\frac{\pi_i}{1-\pi_i}) = \ln(Odds) = logit(\pi_i) = x_i^T \beta$.
  - Previously we know that $E(Y_i) = \mu_{Y_i} = \pi_i$.
  - Range: Real number.
- [2] The inverse canonical link function: $\mu_{Y_i} = \frac{e^{\theta_i}}{1+e^{\theta_i}} = \frac{e^{x_i^T \beta}}{1+e^{x_i^T \beta}}$.
  - Range: [0,1].
- The previous issues in LR (T0) have been solved!

# Odds ratio of $(Y_i = 1|X)$

- $\pi_i = \pi_i|X = P(Y_i = 1|X)$.
- [1] The canonical link function: $\theta_i = \ln(\frac{\pi_i}{1-\pi_i}) = \ln(Odds) = x_i^T\beta$.
- Odds ratio of $(Y_i = 1|X) = \exp([1]) = \exp(x_i^T\beta)$
  - $= \frac{\pi_i}{1-\pi_i}$
  - $= \frac{P(Y_i=1|X)}{1-P(Y_i=1|X)} = \frac{P(Y_i=1|X)}{P(Y_i=0|X)}$
  - Odds that $Y_i = 1$ given $X$
- Odds ratio of $(Y_i = 1|X) = \frac{P(Y_i=1|X)}{P(Y_i=0|X)}$
  - $= 1 \rightarrow P(Y_i = 1|X) = P(Y_i = 0|X) = 0.5 \rightarrow$ a 50% chance that $Y_i = 1$ will occur.
  - $> 1 \rightarrow P(Y_i = 1|X) > P(Y_i = 0|X) \rightarrow$ a >50% chance that $Y_i = 1$ will occur.
  - $< 1 \rightarrow P(Y_i = 1|X) < P(Y_i = 0|X) = 0.5 \rightarrow$ a <50% chance that $Y_i = 1$ will occur.
- Hence, odds is another way to describe probability.

## Interpretation of $\beta_j$

- Odds ratio of $(Y_i = 1|X) = \exp(\theta_i) = \exp(x_i^T \beta) = \exp(\beta_0 + \beta_1 X_{1_i} + ... + \beta_k X_{k_i})$.
- Odds ratio of $(Y_i = 1|X)$ for $(X_{1_i} = x_{1_i}, ..., X_{j_i} = x_{j_i}, ..., X_{k_i} = x_{k_i}) = \exp(\beta_0 + \beta_1 x_{1_i} + ... + \beta_j x_{j_i} + ... + \beta_k x_{k_i}) = A$.
- Odds ratio of $(Y_i = 1|X)$ for $(X_{1_i} = x_{1_i}, ..., X_{j_i} = x_{j_i} + 1, ..., X_{k_i} = x_{k_i}) = \exp(\beta_0 + \beta_1 x_{1_i} + ... + \beta_j (x_{j_i} + 1) + ... + \beta_k x_{k_i}) = B = A \times \exp(\beta_j)$.
- Interpretation of $\beta_j$: With the other variables held constant, if $X_{j_i}$ increases by 1 unit, the odds that $Y_i = 1$ will change by a multiplicative factor of $\exp(\beta_j)$.

# MLE: Obtaining estimated coefficients $\widehat{\beta}_i$s

Assume there are $n$ number of observations in the sample (non-aggregate data), each observation is independent.

In **GLM**, we use **MLE (Maximum Likelihood Estimator)** method to obtain estimated coefficients $\widehat{\beta}_i$s $(\widehat{\beta}_0, ..., \widehat{\beta}_k)$.

- $Y_i \sim Bern(\pi_i)$ with the PMF $P(Y_i = y_i) = f(y_i) = \pi_i^{y_i} \times (1-\pi_i)^{1-y_i}$.
- The likelihood function: $L = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n (\pi_i^{y_i} \times (1-\pi_i)^{1-y_i})$ (each observation is independent, and there are $n$ observations in total).(Note that each $f(y_i)$ is a function of $\pi_i$, the inverse link function [2]. Hence $f(y_i)$ is a function of $\beta_j$s).
- Log-likelihood:
  $l = \ln(L) = \ln(\prod_{i=1}^n f(y_i)) = \sum_{i=1}^n \ln(f(y_i)) = \sum_{i=1}^n \ln(\pi_i^{y_i} \times (1-\pi_i)^{1-y_i})$.
- To maximize $L$ is equivalent to maximize $l = \ln(L)$. Therefore, $\frac{\partial l}{\partial \widehat{\beta}_0} = 0, ..., \frac{\partial l}{\partial \widehat{\beta}_k} = 0$.
- Solve for $\widehat{\beta}_0, ..., \widehat{\beta}_k$.

In R's MLE calculation, it uses IRLS (Iterative Re-Weighted Least Squares) method instead, which is a method that will obtain equivalent estimated coefficients as MLE does. The detail for IRLS is not required in this course.

Substitute $\beta_0, ..., \beta_k$ by $\widehat{\beta_0}, ..., \widehat{\beta_k}$ we have

- [1] The fitted link function: $\widehat{[1]} = \widehat{\theta_i} = g(\widehat{\mu_{Y_i}}) = \ln(\frac{\widehat{\mu_{Y_i}}}{1-\widehat{\mu_{Y_i}}}) = \ln(\frac{\widehat{\pi_i}}{1-\widehat{\pi_i}}) = logit(\widehat{\pi_i}) = \ln(\widehat{Odds}) = x_i{}^T\widehat{\beta} = \widehat{\beta_0} + \widehat{\beta_1}X_{1_i} + ... + \widehat{\beta_k}X_{k_i}$.
- [2] The fitted inverse link function: $\widehat{[2]} = \widehat{\mu_{Y_i}} = \frac{e^{\widehat{\theta_i}}}{1+e^{\widehat{\theta_i}}} = \frac{e^{\widehat{\beta_0}+\widehat{\beta_1}X_{1_i}+\cdots+\widehat{\beta_k}X_{k_i}}}{1+e^{\widehat{\beta_0}+\widehat{\beta_1}X_{1_i}+\cdots+\widehat{\beta_k}X_{k_i}}}$.

# Example 1 - Calculus (Non-aggregate data) - Prediction:

```r
# n=30, X=Time, Y = 1/0 = Answer the question correctly/not
length(calculus$Time)
```

```
## [1] 30
```

```r
head(calculus,n=10)
```

```
##    Time Response
## 1  1.25        1
## 2  1.25        1
## 3  1.25        1
## 4  1.25        1
## 5  1.25        1
## 6  1.25        1
## 7  1.25        0
## 8  1.00        1
## 9  1.00        1
## 10 1.00        1
```

```r
# [1] fitted: logit(pi_hat) =ln(pi_hat/(1-pi_hat))= b0+b1*X1
# [2] fitted: pi_hat = exp([1] fitted)/(1+exp([1] fitted))
calculus.glmt1<-glm(Response ~ Time, family = binomial)
calculus.glmt1$coefficients
```

```
## (Intercept)        Time
##   -6.125247    6.811302
```

```r
calculus.glmt1<-glm(Response ~ Time, family = binomial(link='logit'))
calculus.glmt1$coefficients
```

```
## (Intercept)        Time
##   -6.125247    6.811302
```

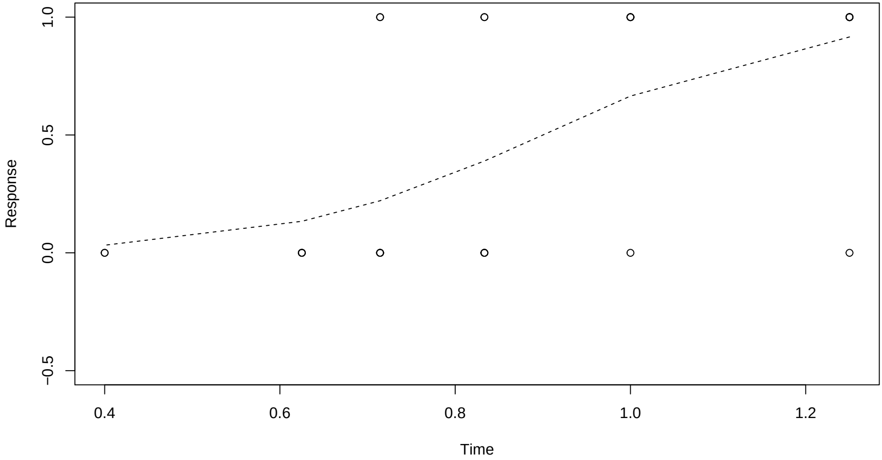
```r
# [2] fitted = exp([1] fitted)/(1+exp([1] fitted))
fitted(calculus.glmt1)[1]
```

```
##         1
## 0.9159755
```

```r
calculus.glmt1$fitted.values[1]
```

```
##         1
## 0.9159755
```

```r
# Create plot between Xi and Yi
plot(Time,Response,ylim=c(-0.5,1))
# Plot T1 (Binary model) => much better than T0 (LR model)
lines(Time,calculus.glmt1$fitted.values,lty=2)
```



```r
# Odds
# [1] logit(pi_hat) =ln(pi_hat/(1-pi_hat))=ln(odds_hat)= b0+b1*X1
predict(calculus.glmt1)[1]
```

```
##        1
## 2.388881
```

```r
-6.125247+6.811302*Time[1]
```

```
## [1] 2.388881
```

```r
# odds = exp([1])
exp(predict(calculus.glmt1))[1]
```

```
##        1
## 10.90129
```

```r
# B1 interpretation:
# With the other variables constant, increase X1 by 1 (from x to x+1)
# => odds of Y=1 will change by a multiplicative factor of exp(B1)
# exp(b1)
exp(summary(calculus.glmt1)$coefficients[2,1])
```

```
## [1] 908.0528
```

```r
# Prediction
# Let Xi=1.5 hours
# [2] fitted = exp([1])/(1+exp([1]))
Xnew<-data.frame(Time=1.5)
Xnew
```

```
##   Time
## 1  1.5
```

```r
predict(calculus.glmt1,Xnew,type='response')
```

```
##        1
## 0.983564
```

```r
exp(-6.125247+6.811302*Xnew)/(1+exp(-6.125247+6.811302*Xnew))
```

```
##       Time
## 1 0.983564
```

```r
# mu_Y hat = pi hat = 0.98 >1 => Yi hat = 1 =>
# for Xi =1.5 (student who studies 1.5 hours' calculus)
# => will have Yi fitted = 1 (answer the quiz question correctly)
```

# Example 4: Calculus Data with Gender (Non-aggregate data)

- Suppose the context for Example 4 is exactly the same as Example 1.
- However, instead of **30** students we now have **another sample**, which contains **100** students attempting the Calculus quiz question.
- We also record the gender of each student.
  - Gender is a dummy variable which has 2 levels $\rightarrow$ use $(0/1)$ to fit
  - Let $X_2$ = Gender.
    - $X_2 = 1$ if this student is a male.
    - $X_2 = 0$ if this student is a female.
  - Rule: For the $X$ variable which has $q$ levels requires $q - 1$ number of dummy variables $(0/1)$ to parameterize this $X$ variable (assume includes the intercept).
- Moreover, we also want to investigate on the interaction effect between the study time and the gender for each student.
  - $X_3 = X_1 \times X_2$, the interaction effect of the gender and the study time

## The Question

- We are interested in studying the relationship between **three factors** and whether one student can answer the Calculus quiz question correctly. Those three factors are:
    - $X_1$ = Time = the study time that each student spends in Calculus (in hours).
    - $X_2$ = the gender.
        - $X_2 = 1$ if this student is a male, and $X_2 = 0$ if this student is a female.
    - $X_3 = X_1 \times X_2$, the interaction effect of the gender and the study time.
        - $X_3 = X_1 \times X_2 = X_1 \times 1 = X_1$ if this student is a male.
        - $X_3 = X_1 \times X_2 = X_1 \times 0 = 0$ if this student is a female.
- The **response variable ($Y_i$)** is still an indication of whether one student's quiz question answer is right or wrong.
    - $Y_i = 1$ if one student's quiz question answer is right.
    - $Y_i = 0$ if one student's quiz question answer is wrong.
- $n$=100.
- Non-aggregate data: $(X_1, X_2, X_3, Y)$

```r
# n=100,X1=Time,X2=gender(0~F,1~M),X3=X1*X2,Y=0/1(answer correctly)
# non-aggregate data
calculus <-read.table("calculus,gender.csv", header = TRUE, sep = ",")
length(calculus$Time)
```

```
## [1] 100
```

```r
# X2 = Gender (2 levels - M/F => M=1, F=0 (baseline level))
contrasts(factor(calculus$Gender))
```

```
##   1
## 0 0
## 1 1
```

```r
# Convert notations to nonaggregate data form
x1_nonagg<-calculus$Time
x2_nonagg<-calculus$Gender
x3_nonagg<-x1_nonagg * x2_nonagg
y_nonagg<-calculus$Response
```

```r
# Non-aggregate data (X1,X2,X3,Y)
# n=100,X1=Time,X2=gender(0~F,1~M),X3=X1*X2,Y=0/1(answer correctly)
nonagg<-data.frame(x1_nonagg,x2_nonagg,x3_nonagg,y_nonagg)
head(nonagg,n=17)
```

```
##    x1_nonagg x2_nonagg x3_nonagg y_nonagg
## 1       1.25         1      1.25        1
## 2       1.25         1      1.25        1
## 3       1.25         1      1.25        1
## 4       1.25         1      1.25        1
## 5       1.25         0      0.00        0
## 6       1.25         1      1.25        0
## 7       1.25         1      1.25        1
## 8       1.25         0      0.00        1
## 9       1.25         1      1.25        1
## 10      1.25         1      1.25        0
## 11      1.25         1      1.25        1
## 12      1.25         0      0.00        0
## 13      1.14         1      1.14        0
## 14      1.14         0      0.00        1
## 15      1.14         1      1.14        1
## 16      1.14         0      0.00        1
## 17      1.14         0      0.00        0
```

As the response variable $Y$ is a binary variable, we fit the Binary regression with logistic link (T1).

- $X_1$ = Time, $X_2$ = the gender (=1 if male), $X_3 = X_1 \times X_2$, $Y$ = whether the quiz question answer is right or wrong.
- $k = 3$ (slope number),$p = k + 1 = 4$ (parameter number), $n$=100.
- We use $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3} \rightarrow \beta_0, \beta_1, \beta_2, \beta_3$.
- [1] The canonical link function: $\theta_i = g(\mu_{Y_i}) = \ln(\frac{\mu_{Y_i}}{1-\mu_{Y_i}}) = \ln(\frac{\pi_i}{1-\pi_i}) = logit(\pi_i) = \ln(Odds) = {x_i}^T \beta = \beta_0 + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i}$.
- [1] The fitted canonical link function: $\widehat{\theta_i} = g(\widehat{\mu_{Y_i}}) = \ln(\frac{\widehat{\mu_{Y_i}}}{1-\widehat{\mu_{Y_i}}}) = \ln(\frac{\widehat{\pi_i}}{1-\widehat{\pi_i}}) = logit(\widehat{\pi_i}) = \ln(\widehat{Odds}) = {x_i}^T \widehat{\beta} = \widehat{\beta_0} + \widehat{\beta_1} X_{1_i} + \widehat{\beta_2} X_{2_i} + \widehat{\beta_3} X_{3_i}$.

```r
# [1] logit(pi) =ln(pi/(1-pi))= B0+B1*X1+B2*X2+B3*X3
# [2] pi = exp([1])/(1+exp([1]))
calculus.glmt1<-glm(y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg,
                    family = binomial)
calculus.glmt1$coefficients
```

```
## (Intercept)   x1_nonagg   x2_nonagg   x3_nonagg
##  -1.0756181   1.7650778  -0.8174754   0.7326969
```

```r
# Dispersion (assumed) = 1
summary(calculus.glmt1)$dispersion
```

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

```r
# (Unscaled) Deviance = D
summary(calculus.glmt1)$deviance
```

```
## [1] 126.6875
```

# Deviance

- (Unscaled) Deviance measures the goodness of fit for GLM.
  - How far the model deviates from the data(observation).
  - (Unscaled) Deviance $= D = D(y, \widehat{y}) = constant - 2 \times l$
  - log-likelihood $= l = \ln(L) = \ln(\prod_{i=1}^{n} f(y_i)) = \sum_{i=1}^{n} \ln(f(y_i)) =$ a function of ($\beta_i$s).
  - a lower deviance: model fits the data better.
  - In R: summary(model)$deviance
- Scaled deviance $= D^* = D(y, \widehat{y})^* = \frac{D(y, \widehat{y})}{\phi_{assumed}}$
- $\phi_{assumed}$
  - $= 1$ for T1 (Binary)/2(Binomial)/3(Poisson)
  - $= MSE = \sigma^2$ for T0 (LR)
  - $= cv = 1/$shape parameter$= \frac{1}{\alpha}$ for T5 (Gamma).
  - In R: summary(model)$dispersion

# Hypothesis testing

(Please refer to the extra hand-written paper)

4 Types of hypothesis testing

- Hypothesis 1 - Individual hypothesis - Wald test (1 single beta) $\beta_j$ $(j = 0, 1, 2, ..., k)$
- Hypothesis 2 - Drop in deviance test - 1 single beta $\beta_j$ $(j = 1, 2, ..., k)$
- Hypothesis 3 - Drop in deviance test - All the $\beta_j$s $(j = 1, 2, ..., k)$
- Hypothesis 4 - Drop in deviance test - any combination of $\beta_j$s $(j = 1, 2, ..., k)$

# Hypothesis 1 - Individual hypothesis - Wald test (1 single beta) $\beta_j$ ($j = 0, 1, 2, ..., k$)

```r
# Ho: Bi=0; H1:Bi != 0
summary(calculus.glmt1)$coefficients
```

```
##                Estimate Std. Error    z value   Pr(>|z|)
## (Intercept) -1.0756181  0.6644921 -1.6187071 0.10551030
## x1_nonagg    1.7650778  0.9142947  1.9305348 0.05354061
## x2_nonagg   -0.8174754  1.0117823 -0.8079557 0.41911606
## x3_nonagg    0.7326969  1.3311888  0.5504080 0.58203959
```

```r
# vs N(0,1)'s critical value
c(qnorm(0.025), qnorm(0.975))
```

```
## [1] -1.959964  1.959964
```

```r
# If p-value < 5% => Reject Ho
# Assume the sample size is relatively large.
```

# Hypothesis 2 - Drop in deviance test - 1 single beta $\beta_j$ ($j = 1, 2, ..., k$)

```r
# F:[1] logit(pi) =ln(pi/(1-pi))= B0+B1*X1+B2*X2+B3*X3
calculus.glmt1<-glm(y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg,
                    family = binomial)
# R:[1] logit(pi) =ln(pi/(1-pi))= B0+B1*X1+B2*X2
calculus.glmt12<-glm(y_nonagg ~ x1_nonagg + x2_nonagg,
                     family = binomial)
# Ho: B3=0 ; H1: >=1 B3 !=0
# Shows unscaled change in deviance and scaled change
# in deviance (by dispersion assumed) p-value from Chisquare
anova(calculus.glmt12,calculus.glmt1,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: y_nonagg ~ x1_nonagg + x2_nonagg
## Model 2: y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        97     126.99
## 2        96     126.69  1   0.3036   0.5816
```

```r
# TS = Change in deviance/dispersion(assumed) of full
anova(calculus.glmt12,calculus.glmt1,test="Chisq")$Deviance[2]/
      summary(calculus.glmt1)$dispersion
```

```
## [1] 0.3036019
```

```r
ts2<-(summary(calculus.glmt12)$deviance-summary(calculus.glmt1)$deviance)/
      summary(calculus.glmt1)$dispersion
ts2
```

```
## [1] 0.3036019
```

```r
# Vs Chisq,one side,df=# in Ho = 1
# cv (RHS RR=5%)
qchisq(0.95, 1)
```

```
## [1] 3.841459
```

```r
# p-values = values from ANOVA table (scaled by dispersion assumed)
1 - pchisq(ts2, 1)
```

```
## [1] 0.5816331
```

# Hypothesis 3 - Drop in deviance test - All the $\beta_j$s ($j = 1, 2, ..., k$)

```r
# F:[1] logit(pi) =ln(pi/(1-pi))= B0+B1*X1+B2*X2+B3*X3
calculus.glmt1<-glm(y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg,
                    family = binomial)
# R:[1] logit(pi) =ln(pi/(1-pi))= B0
calculus.glmt13<-glm(y_nonagg ~ 1, family = binomial)
# Ho: B1=B2=B3=0; H1: >=1 Bi !=0
# Shows unscaled change in deviance and scaled change
# in deviance (by dispersion assumed) p-value from Chisquare
anova(calculus.glmt13,calculus.glmt1,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: y_nonagg ~ 1
## Model 2: y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        99     138.63
## 2        96     126.69  3   11.942 0.007585 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# TS = Change in deviance/dispersion(assumed) of full
anova(calculus.glmt13,calculus.glmt1,test="Chisq")$Deviance[2]/
      summary(calculus.glmt1)$dispersion
```

```
## [1] 11.94198
```

```r
ts3<-(summary(calculus.glmt13)$deviance-summary(calculus.glmt1)$deviance)/
      summary(calculus.glmt1)$dispersion
ts3
```

```
## [1] 11.94198
```

```r
# Vs Chisq,one side,df=# in Ho = 3
# cv (RHS RR=5%)
qchisq(0.95, 3)
```

```
## [1] 7.814728
```

```r
# p-values = values from ANOVA table (scaled by dispersion assumed)
1 - pchisq(ts3, 3)
```

```
## [1] 0.00758457
```

# Hypothesis 4 - Drop in deviance test - any combination of $\beta_j$s ($j = 1, 2, ..., k$)

```
# F:[1] logit(pi) =ln(pi/(1-pi))= B0+B1*X1+B2*X2+B3*X3
calculus.glmt1<-glm(y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg,
                    family = binomial)
# R:[1] logit(pi) =ln(pi/(1-pi))= B0+B2*X2
calculus.glmt14<-glm(y_nonagg ~ x2_nonagg, family = binomial)
# Ho: B1=B3=0 ; H1: >=1 Bi !=0
# Shows unscaled change in deviance and scaled change
# in deviance (by dispersion assumed) p-value from Chisquare
anova(calculus.glmt14,calculus.glmt1,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: y_nonagg ~ x2_nonagg
## Model 2: y_nonagg ~ x1_nonagg + x2_nonagg + x3_nonagg
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        98     138.47
## 2        96     126.69  2   11.782 0.002764 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# TS = Change in deviance/dispersion(assumed) of full
anova(calculus.glmt14,calculus.glmt1,test="Chisq")$Deviance[2]/
      summary(calculus.glmt1)$dispersion
```

```
## [1] 11.78194
```

```r
ts4<-(summary(calculus.glmt14)$deviance-summary(calculus.glmt1)$deviance)/
      summary(calculus.glmt1)$dispersion
ts4
```

```
## [1] 11.78194
```

```r
# Vs Chisq,one side,df=# in Ho = 2
# cv (RHS RR=5%)
qchisq(0.95, 2)
```

```
## [1] 5.991465
```

```r
# p-values = values from ANOVA table (scaled by dispersion assumed)
1 - pchisq(ts4, 2)
```

```
## [1] 0.002764294
```