STAT2014/6014 Lecture Week 7 Binary regression with logistic link

Lucy Hu
Email: yunxi.hu@anu.edu.au [1]

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

Australian National University

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 β_j
  - MLE (Maximum likelihood estimator)
  - Prediction (Example 1)
  - Deviance
  - Hypothesis testing (Example 4)

Today

Next week

$$\begin{cases} W1 - W6 : LR \underset{\boxed{TD}}{:} Y_i \mid X_i = x \sim ND \\ W7 - W12 : GLM : Y_i \mid X_i = x \sim \text{expo. family dsb (EFD)} \end{cases}$$

$$\Downarrow$$

$$\boxed{TI} : Y_i \mid X_i = x \sim \text{Binary} = \text{Bernoulli}$$

$$(\underset{s}{\underset{\parallel}{1}}, \underset{f}{\underset{\parallel}{0}})$$

Today:

$\boxed{T1}$ $Y_i \mid X_i = x$ ~ Binary dsb (0/1)

the relation b/w X_i & Y_i ~ Binary regression

Examples of binary regression

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:
  1. 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**.

(skip)

Example of $Y_i \sim$ Binary $\boxed{?}$

① $\begin{cases} Y_i = \text{whether to buy a new product} \begin{cases} 0 \text{ (no)} \sim \text{failure} \\ 1 \text{ (yes)} \sim \text{success} \end{cases} \\ x_i = \text{age} \end{cases}$

② Calculus example below

$\begin{cases} Y_i = \text{whether is able to answer the quiz correctly} \begin{cases} 0 \text{ (no)} \sim \text{failure} \\ 1 \text{ (yes)} \sim \text{success} \end{cases} \\ x_i = \text{time to study Calculus} \end{cases}$

- $Y_i \sim Bin(M_i = 1, \pi_i) = Bern(\pi_i)$ (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).

(skip)

Binary regression features (Calculus example)

- $Y_i \mid X_i = x \rightarrow Y_i \sim \text{Bin}(M_i = 1, \pi_i) = \text{Bern}(\pi_i)$

  where $\pi_i = P_i = E(Y_i \mid X_i = x) = P(Y_i = 1 \mid X_i = x)$

- $Y_i =$
  - 0 (failure)
  - 1 (success)

  ← coded as 0/1

- non-aggregate data:

  (x_i, Y_i)

X_i	Y_i
⋮	0
⋮	1
⋮	0
⋮	1
⋮	⋮

1 row = 1 trail = 1 obs

rows = # obs.

Y_i = response (non-aggregate)

/

Example 1 - Calculus Data

(Wattle)

- 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.

(skip)

X_i Y_i

Non-Aggregate data

"Time" Time to study (Xi)	"Response" Answer correctly (Yi)
1.25	1
1.25	1
1.25	1
1.25	1
1.25	1
1.25	1
1.25	0
1	1
1	1
1	1
1	1
1	0
0.833333333	1
0.833333333	1
0.833333333	0
0.833333333	0
0.833333333	0
0.833333333	0
0.714285714	1
0.714285714	1
0.714285714	0
0.714285714	0
0.714285714	0
0.714285714	0
0.625	0
0.625	0
0.625	0
0.625	0
0.4	0
0.4	0

7# 6# have '1'

Aggregate data

Time to study (Xi)	# of trials that answer correctly (Zi)	# of trials (#Mi)
1.25	6	7
1	4	5
0.833333333	2	6
0.714285714	2	6
0.625	0	4
0.4	0	2

- $n=30$ # students (1 student = 1 trial)
- X = Time to study (hrs)
- $Y = \begin{cases} 1 \sim \text{Quiz ANS } \checkmark = \text{a success} \\ 0 \sim \cdots\cdots\cdots \times = \text{a failure} \end{cases}$

2 levels under 1 trial (1 student)

```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
```

n=30

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

← 1st 14 # students

X Y

```
##          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
```

$$\frac{1\times 6+0\times 1}{7} = \frac{6}{7}$$

(7# students, $X_i = 1.25$ hours
Prob of $Y_i=1$ is $\frac{6}{7} = 0.857$)

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: $(X_i = \mathit{Time}, Y_i = 0/1)$.
- $Y_i = 0/1$ 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 $X_i = x$)** $\sim \mathit{Bin}(1, \pi_i) = \mathit{Bern}(\pi_i)$.

(skip)

non-agg data (each student)

$n = 30 \rightarrow 1$ student $= 1$ trail $\rightarrow Y_i = \begin{cases} 0 \\ 1 \end{cases}$ (2 outcomes)

Y_i (under 1 trail, given $X_i = x$) $\sim Bin(\underbrace{1}_{\text{under 1 trail (1 student)}}, \pi_i) = Bern(\pi_i)$

$\pi_i = P(Y_i = 1 \mid X_i = x)$

$= p_i$

(some textbook)

Issues

X_i	Y_i (0,1)
⋮	⋮

Why can't we use linear regression?

$\boxed{1}$
- Issue 1: $Y_i = 0/1$ only, which is not normally distributed.

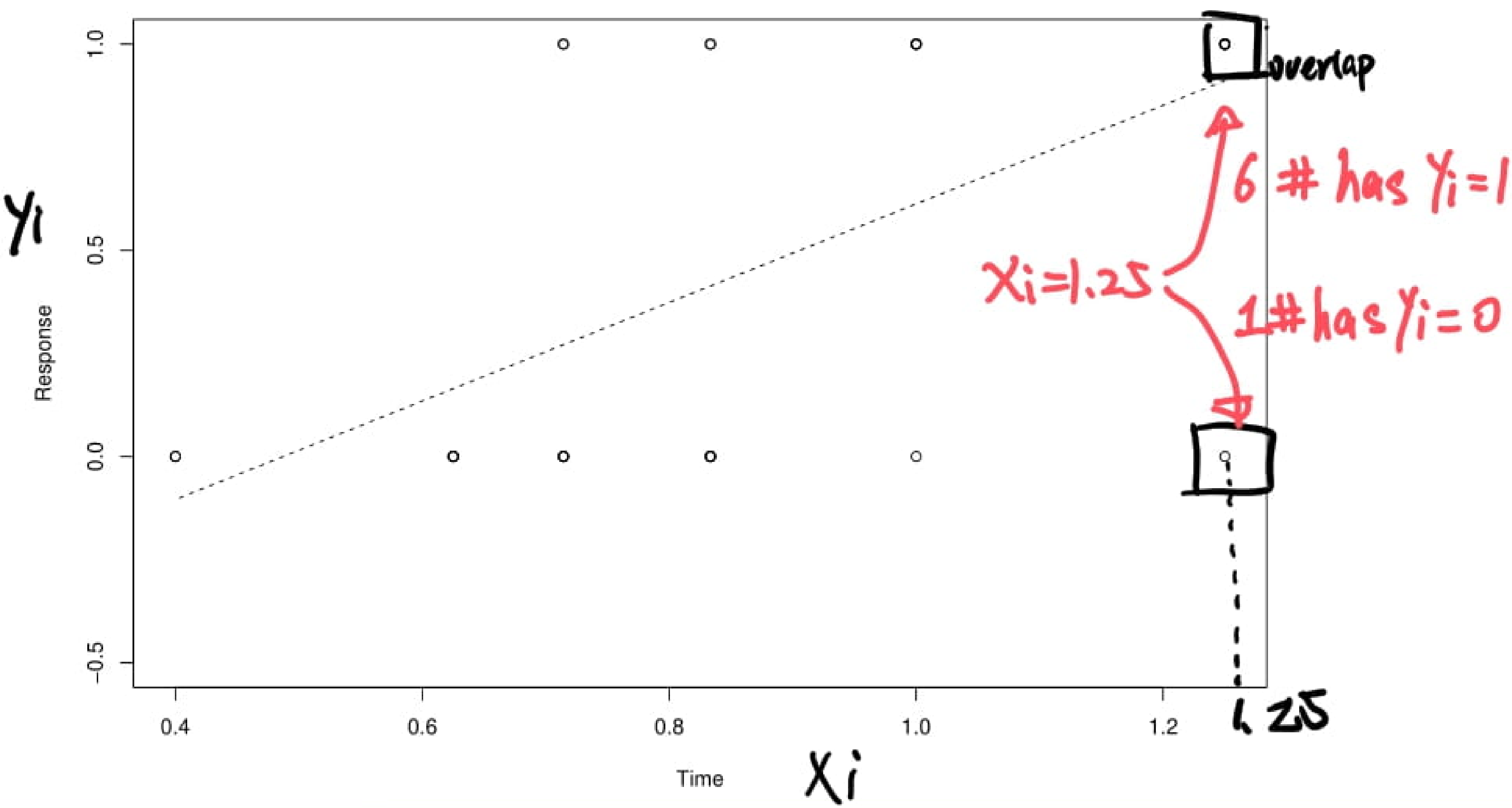
$\boxed{2}$
- Issue 2: For LR, $\mu_{Y_i} = \beta_0 + \beta_1 X_{1_i} + ... + \beta_k X_{k_i} = x_i^T \beta$.
  - Range for LHS: [0,1].
  - Range for RHS: Real number.
  - Not match!

```
# 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)
```

← Plot (X_i, Y_i); $Y_i = 0/1$ (not ND)

[1]

e.g.: $Y_i \mid X_i = 1.25$



$\boxed{2}$

```
# 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)
```

LHS

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

$X_i = 1.25 \begin{cases} Y_i = 1 \to 6\# \\ Y_i = 0 \to 1\# \end{cases}$

$\therefore \frac{6\times 1 + 1\times 0}{7} = \hat{\mu}_{Y|X=1.25}$

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

RHS

```
##               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
```

$b_0 + b_1 \cdot X = -0.58 + 1.195 \cdot X$

```
## [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)
```

- "⋰" : LR: lm(Y~X)

LR: $\mu_{Y|X=x} = \Sigma \beta X$

proxy: $\hat{\mu}_{Y|X=x}$ (sample mean of Y @ $X=x$)

e.g.: $\hat{\mu}_{Y|X=1.25} = 0.85$ (7#)

· proxy: $\Sigma \hat{\beta} X = b_0 + b_1 X$
$= -0.58 + 1.195 \cdot x$

· e.g $= \begin{cases} -0.1044, & x=0.4 \\ 0.911, & x=1.25 \end{cases}$

- $Y=0/1 \rightarrow \hat{\mu}_{Y|X=x} \in [0,1] \overset{\otimes}{\longleftrightarrow}$ · $b_0<0$, $\notin [0,1]$, $\in \mathbb{R}$

- Solution? $\begin{cases} LR: & \mu_Y = \Sigma \beta X \\ GLM: & g(\mu_Y) = \Sigma \beta X \quad \text{(link fun)} \end{cases}$

If $Y_i \mid X_i \sim \text{Bern}(\pi_i) = \text{Binary}(\pi_i)$

why $E(Y_i \mid X_i = x) = P(Y_i = 1 \mid X_i = x)$ $\boxed{?}$

$$\mu_{Y_i \mid X_i} = \mu_{Y_i}$$

π_i

(or p_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 π_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 $P(Y_i = y_i) = f(y_i) = \pi_i^{y_i} \times (1 - \pi_i)^{1-y_i}$.
  - $E(Y_i) = 1 \times \pi_i + 0 \times (1 - \pi_i) = \pi_i = \pi_i \in [0, 1]$.
  - $V(Y_i) = (1 - \pi_i)^2 \times \pi_i + (0 - \pi_i)^2 \times (1 - \pi_i) = \pi_i \times (1 - \pi_i)$.

(skip)

- 1 obs = 1 trial = 1 student $\sim 1\ y_i$

- This 1 y_i = (2 outcomes)

$$y_i = \begin{cases} \boxed{1} \longrightarrow s \longrightarrow p(s) = \pi_i = P(Y_i = 1 \mid X_i = x) \\ \boxed{0} \longrightarrow f \longrightarrow p(f) = 1 - \pi_i = P(Y_i = 0 \mid X_i = x) \end{cases}$$

- $E(Y_i \mid X_i = x) = \sum^{\text{2 outcomes}} pr. \cdot y_i = 1 \cdot \pi_i + 0 \cdot (1 - \pi_i) = \pi_i$

Therefore: $\underbrace{E(Y_i \mid X_i = x)}_{= \mu_{Y_i \mid X_i = x}} = \underbrace{P(Y_i = 1 \mid X_i = x)}_{= \pi_i}$

$\mu_{Y_i} = \mu_{Y_i \mid X_i = x}$

$\pi_i = p_i$

Link & inverse link fun for

Binary regression $= \boxed{?}$

Australian National University

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.

$= \pi_i = P(Y_i = 1 \mid X_i = x)$

- [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!

(skip)

(w6)

- [1]: Link : $g(\mu_y) = \underbrace{\Sigma\beta x = \theta}_{\text{canonical link}}$ $= \ln\left(\frac{\mu_y}{1-\mu_y}\right) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln\left(\text{odds ratio}\right)$

  [= [1] logistic link

  $\mu_y = \pi_i$

  [2]: Inverse link: solve for μ_y

$\downarrow$

[1]: canonical link : $\underbrace{\Sigma\beta x}_{\in \mathbb{R}} = \underbrace{\ln\left(\frac{\mu_y}{1-\mu_y}\right)}_{\in \mathbb{R}}$

[2]: Inverse canonical link : $\underset{\substack{\in[0,1] \\ y=0,1}}{\mu_y} = \underset{\in[0,1]}{\frac{e^{\Sigma\beta x}}{1+e^{\Sigma\beta x}}} = \pi_i = P(Y_i = 1 \mid X_i = x)$ ☆

- Model fitting logic:

  $n\#\ (x_i, y_i)$

  $\downarrow$

  GLM: obtain b_is

  $\downarrow$

  sub X_{new} to the equations (prediction)

[1] canonical link fun: $\sum b_i \cdot X_{new} = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln\left(\frac{\mu_Y}{1-\mu_Y}\right) = \ln\left(\text{odds ratio}\right)$

$\rightarrow \text{odds ratio} = \frac{\pi_i}{1-\pi_i} = \exp([1])$

[2] inverse canonical link fun: $\frac{\exp(\sum b_i \cdot X_{new})}{1+\exp(\sum b_i \cdot X_{new})} = \mu_Y = \pi_i = P(Y_i = 1 \mid X_i)$

$\rightarrow \hat{Y}_i = 0 \text{ or } 1$ (target) odds ratio $= \boxed{?}$

Australian National University

- $\pi_i = \pi_i|X = P(Y_i = 1|X)$.
- [1] The canonical link function: $\theta_i = \ln(\boxed{\frac{\pi_i}{1-\pi_i}}) \overset{=}{=} \ln(\boxed{Odds}) = x_i^T\beta$.

exp

- Odds ratio of $(Y_i = 1|X) = \exp([1]) = \exp(x_i^T\beta)$
  - $= \frac{\pi_i}{1-\pi_i}$
  - $= \frac{P(Y_i=1|X)}{1-P(Y_i=1|X)} = \frac{P(Y_i=1|X)}{P(Y_i=0|X)}$ } = relative pr of $Y_i=1$ to $Y_i=0$

    $\pi_i = P(Y_i = 1 | X_i = x)$
  - Odds that $Y_i = 1$ given X
- Odds ratio of $(Y_i = 1|X) = \frac{P(Y_i=1|X)}{P(Y_i=0|X)} = \frac{\pi_i}{1-\pi_i}$
  - $= 1 \rightarrow P(Y_i = 1|X) = P(Y_i = 0|X) = 0.5 \rightarrow$ a 50% chance that $Y_i = 1$ will occur.
  - $> 1 \rightarrow P(\overset{\pi_i}{Y_i = 1}|X) > P(\overset{1-\pi_i}{Y_i = 0}|X) \rightarrow$ 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.

(skip)

- $\pi_i = p_i = P(Y_i=1 \mid X=x) = \text{Prob}(s)$

  (some textbook)

  $\rightarrow 1-\pi_i = P(Y_i=0 \mid X=x) = \text{Prob}(f)$

- odds ratio of $(Y_i=1 \mid X=x)$ (= odds)

$$\frac{P(Y_i=1 \mid X)}{P(Y_i=0 \mid X)} \begin{matrix} s \\ f \end{matrix} = \frac{\pi_i}{1-\pi_i} = \text{relative prob of } Y_i=1 \text{ over } Y_i=0 = \exp([1])$$

- [1]: canonical link:

$$\Sigma\beta X = \ln\left(\frac{\mu_Y}{1-\mu_Y}\right) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln(\text{odds})$$

- odds ratio of $(Y_i=1 \mid X=x) = \frac{P(Y_i=1 \mid X)}{P(Y_i=0 \mid X)} = \frac{\pi_i}{1-\pi_i}$

$$\begin{cases} =1 \rightarrow \ldots \\ >1 \rightarrow P(Y_i=1 \mid X) > P(Y_i=0 \mid X) \rightarrow \pi_i > 1-\pi_i \rightarrow \pi_i > 0.5 \rightarrow \text{a } >50\% \text{ chance that } Y_i=1 \text{ will occur.} \\ <1 \rightarrow \ldots \end{cases}$$

eg: sub X_{new} to model (bi given) $\rightarrow \widehat{\text{odds}} = 1.5 > 1 \rightarrow$ Prob of $Y_i=1$ $> 50\% \rightarrow \hat{Y}_i = 1$

- What if:

given: odds of $(Y_i=1 \mid X_i=x) = \dfrac{\pi_i}{1-\pi_i} = \dfrac{P(Y_i=1 \mid X_i=x)^{\,s}}{P(Y_i=0 \mid X_i=x)_{\,f}}$

$$\Downarrow$$

$\boxed{Q}$ odds of $(Y_i=0 \mid X_i=x) = ?$

$$= \frac{P(Y_i=0 \mid X_i=x)^{\,f}}{P(Y_i=1 \mid X_i=x)_{\,s}} = \frac{1-\pi_i}{\pi_i} = \frac{1}{\text{odds}\,(Y_i=1 \mid X_i=x)}$$

Australian National University

- 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 β_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 $\exp(\beta_j)$.

(skip)

- β_1 Interpretation:

  - $\text{odds}(Y_i=1 \mid X_i=x) = \dfrac{P(Y_i=1 \mid X_i=x)^{\,s}}{P(Y_i=0 \mid X_i=x)_{\,f}} = \dfrac{\pi_i}{1-\pi_i}$
  - Canonical link: $\ln(\text{odds}_1) = \beta_0 + \beta_1 \cdot x$

  $\hookrightarrow \uparrow x$ by 1: $\ln(\text{odds}_2) = \beta_0 + \beta_1 (x+1) = \beta_0 + \beta_1 \cdot x + \beta_1$

  $= \ln(\text{odds}_1) + \beta_1$

  $\therefore \ln(\text{odds}_2) - \ln(\text{odds}_1) = \beta_1$

  $\therefore \ln\left(\dfrac{\text{odds}_2}{\text{odds}_1}\right) = \beta_1 \longrightarrow \dfrac{\text{odds}_2}{\text{odds}_1} = e^{\beta_1} \rightarrow \boxed{\underbrace{\text{odds}_2}_{X=x+1} = \underbrace{\text{odds}_1}_{X=x} \cdot e^{\beta_1}}$

  - $\boxed{\beta_1:}$ $\uparrow x$ by 1, odds of $(Y_i=1)$ will change by a multiplicative factor of e^{β_1}

/

M LE

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.

⇒bis

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 π_i, the inverse link function [2]. Hence $f(y_i)$ is a function of β_js).
- 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.

Model Prediction

- prediction (LR):
  - $\boxed{S1}$ Data $(x_i, y_i) \leftarrow n\#$

    $\downarrow$
  - $\boxed{S2}$ build a LR via OLSE $(n\#)$

    $\downarrow$
  - $\boxed{S3}$ $\hat{y} = b_0 + b_1 \cdot x$ = fitted model

    sab $\boxed{X_{new}}$ to $\hat{y} = b_0 + b_1 \cdot x_{new} \Rightarrow \hat{y} = \ldots$

- prediction: (GLM):

  - $\boxed{S1}$ Given n # (X_i, Y_i) $Y_i = 0/1$

    ↓

  - $\boxed{S2}$ Build the model

    canonical link: $\sum \hat{\beta} x = \ln\left(\frac{\hat{\mu}_Y}{1-\hat{\mu}_Y}\right)$

    Inverse: $\hat{\mu}_Y = \frac{e^{\sum \hat{\beta} x}}{1+e^{\sum \hat{\beta} x}} = \hat{\pi}_i$

    ↓

  - $\boxed{S3}$ Xnew ~ $\hat{Y} = ?$

Logic (GLM)

S1 · Data = non-agg (x_i, y_i) = $\boxed{n\#}$ (training)   (y_i: 0, 1)

↓

S2 · MLE : build model (GLM)

↓

bis & $\hat{[1]}$ & $\hat{[2]}$ model ← $b_i = \hat{\beta} \xrightarrow{\text{proxy}} \beta$

} build a GLM model via MLE (n# obs)

S3 · sub $\boxed{X_{new}}$ (testing) to model (bis) ⇒ $\hat{y}_i = 1 \text{ or } 0$

} fcst: sub $\boxed{X_{new}}$ to model to get $\hat{y}_i = 1/0$

W1

fitted canonical link fun =

$\hat{[1]} = \hat{\theta}_i = \Sigma \hat{\beta} X_{new} = \ln\left(\frac{\hat{\mu}_{y|x_{new}}}{1-\hat{\mu}_{y|x_{new}}}\right)$

$\hat{[2]} = \hat{\mu}_{y_i|X=X_{new}} = \hat{\pi}_i = \frac{e^{\Sigma \hat{\beta} x_{new}}}{1+e^{\Sigma \hat{\beta} x_{new}}}$

fitted inverse canonical link "

$= P(\hat{y}_i = 1 \mid X = x_{new})$

→ $> 0.5 \rightarrow \hat{y}_i = 1$

$< 0.5 \rightarrow \hat{y}_i = 0$

$= 0.5 \rightarrow \hat{y}_i = 1/0$

→ interpret $\hat{y}_i$

W2

$\widehat{odds}(y_i=1) = \frac{P(\hat{y}_i = 1 \mid X_{new})}{\hat{P}(y_i = 0 \mid X_{new})}$ (s / f)

$= \exp(\hat{[1]})$ ← fitted link fun.

→ $> 1 \rightarrow \hat{y}_i = 1$

$< 1 \rightarrow \hat{y}_i = 0$

$= 1 \rightarrow \hat{y}_i = 0/1$

→ interpret $\hat{y}_i$

S2

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: $\widehat{[2]} = \widehat{\mu_{Y_i}} = \frac{e^{\widehat{\theta_i}}}{1+e^{\widehat{\theta_i}}} = \frac{e^{\widehat{\beta_0}+\widehat{\beta_1}X_{1_i}+...+\widehat{\beta_k}X_{k_i}}}{1+e^{\widehat{\beta_0}+\widehat{\beta_1}X_{1_i}+...+\widehat{\beta_k}X_{k_i}}}$.

(skip)

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

S1

n=30

non-agg (x_i, y_i)

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

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

x_i y_i

```
##    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
```

S2 MLE → bis { by default: link = logit; $b_0, b_1 = \cdots$; $\hat{[1]} = \cdots, \hat{[2]} = \cdots$

b_0 b_1

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

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

b_0 b_1

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

$\hat{[1]}\ \hat{\theta}_i = -6.12 + 6.8X = \ln\left(\frac{\hat{\mu}_{Y|x}}{1-\hat{\mu}_{Y|x}}\right)$

$\hat{[2]}\ \hat{\mu}_{Y|x} = \frac{e^{-6.12+6.8X}}{1+e^{-6.12+6.8X}} = P(\hat{Y}_i = 1 | X_i = x) = \hat{\pi}_i$

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

S3 W1 $\boxed{X_{new} = 1.25}$ (1st obs)

$\rightarrow \hat{[2]} = P(\hat{Y}_i = 1 | X_i = 1.25) = \frac{e^{-6.12+6.8\cdot 1.25}}{1+e^{-6.12+6.8\cdot 1.25}}$

$= 0.92 > 0.5$

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

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

$\therefore \hat{Y}_i = 1$ @ $X = 1.25$

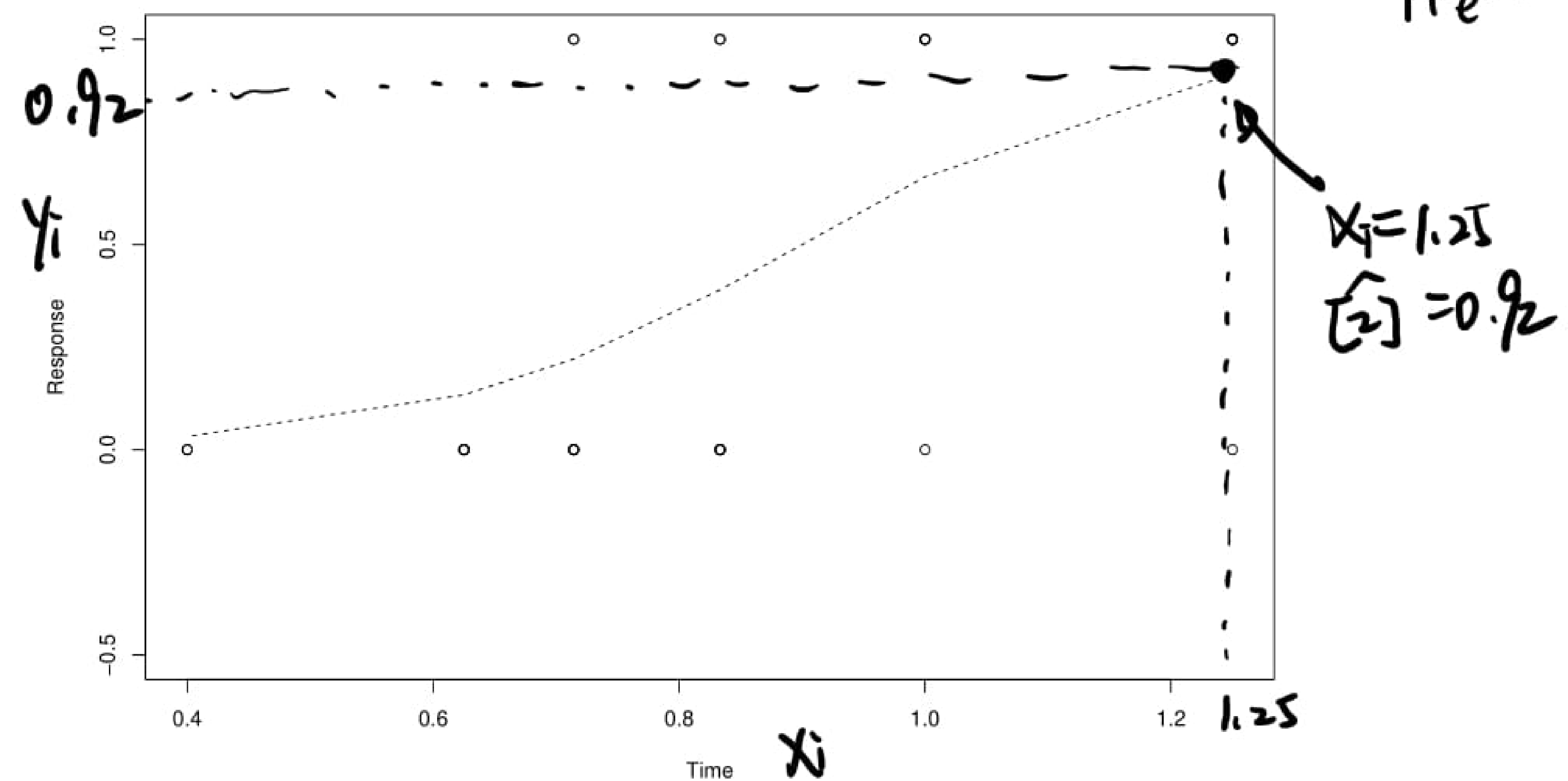
$\therefore$ Quiz ANSW

```
##         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)
```

$\leftarrow (X_i, Y_i)$ 0/1

vertical $= \hat{[z]} = \text{inv link} = P(\hat{Y}_i = 1 \mid X_i = x) = \hat{\mu}_{Y_i \mid X_i = x} = \dfrac{e^{\Sigma bx}}{1 + e^{\Sigma bx}}$



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

— S3 W2 $\boxed{X_{new}=1.25}$ → $\hat{[1]} = b_0 + b_1 \cdot X = 2.389$

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

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

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

→ odds $(Y_i=1 \mid X=1.25) = \exp(\hat{[1]}) = 10.9 > 1$

→ $P(\hat{Y_i}=1 \mid X=1.25) > P(\hat{Y_i}=0 \mid X=1.25)$

→ $\hat{Y_i}=1$ (ANSW)

→ odds: another way of expressing probability

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

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

$\ln(odds_1) = b_0 + b_1 x$

$\ln(odds_2) = b_0 + b_1(x+1) = \ln(odds_1) + b_1$

← previous $\therefore \frac{odds_2}{odds_1} = e^{b_1}$

```
# 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
```

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

another eg

S3 | W1 | Xnew=1.5hr

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

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

[2] = inv link = $P(\hat{Y}_i=1|X=1.5) = 0.98 > 0.5$

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

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

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

$\hat{Y}_i = 1$

```
# 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)
```

Recap:

- $Y_i \sim Bin(1, \pi_i) = Bern(\pi_i) = 0/1$
- examples: $glm(Y_{0/1} \sim X) \rightarrow$

X_i	Y_i
⋮	0 ⋮

non-aggregate data
binary regression

- $\pi_i = p_i = P(Y_i = 1 | X_i = x)$
  $\mu_{Y_i} = \mu_{Y_i | x_i = x} = E(Y_i | X_i = x)$) θ

canonical: logit = logistic link

- Binary regression {
  (canonical) link fun [1]: $g(\mu_Y) = \sum \beta X = \theta = \ln\left(\frac{\mu_{Y_i}}{1-\mu_{Y_i}}\right) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln(\text{odds})$
  inverse (canonical) link fun: [2] $\mu_Y = \frac{\exp(\sum \beta X)}{1+\exp(\sum \beta X)} = \pi_i = P(Y_i = 1 | X_i = x)$

- odds ratio: $\frac{\pi_i}{1-\pi_i} = \frac{P(s)}{P(f)} = \exp([1])$

- MLE → coefficients

Target ☆

- Prediction {
  $\boxed{w1}$ sub X_{new} to [2] $= \hat{P}(Y_i = 1 | X_i = X_{new}) \rightarrow \boxed{\hat{Y}_i = 0/1}$
  $\boxed{w2}$ sub X_{new} to $\hat{\text{odds}} = \hat{\exp}([1])$