STAT 380: Classification: Logistic Regression

Subhadip Pal

Outlook of the unit

- **Prediction and Classification Approaches**
  - Classification Techniques
    - Logistic regression
    - Discriminant analysis
  - Evaluating Performance of a Classification Technique
  - Tree-based methods: Decision trees
    - Classification trees
    - Regression trees

Classification Technique: Logistic Regression

A Few Examples of classification

Banking: determining whether a transaction is fraudulent.

E-commerce: forecasting whether a particular order will be paid for or which customer will buy a specific product.

National security: identifying whether a certain behavior indicates a possible thread.

Tourism: determining the rating a hotel should be awarded.

Medicine: diagnosing whether a particular disease is present.

Assumptions

- Extracting knowledge from massive data sets assumes that they contain non-random, useful information.
- Classification algorithms, especially predictions, do not have further assumptions about how the data was generated.
- Most classical statistical methods are not valid if their **assumptions** are not fulfilled.

This does not mean that big data is always good data.

Notations

n: number of observations in a sample

p: number of predictor variables (features)

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$:matrix of predictor variables

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$ vector of response (dependent / target) variable

Data : $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, where x_i is a vector of length p, represents the observed data from which a statistical model is built.

Classification Rule

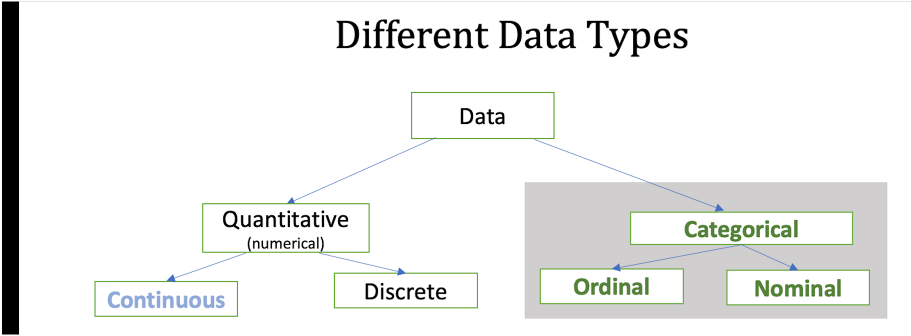
Classification employs a set of p inputs to explain a value of an outcome variable y.

- x is an object in an object space $\mathbb{X}$
- $y = y(x)$ is the response in a response space $\mathbb{Y}$
- Algorithm $a(x)$ is a function from $\mathbb{X}$ to $\mathbb{Y}$

In classification, $\mathbb{Y}$ is a set of discrete categories, $\{C_1, C_2, ..., C_K\}$, exclusively and exhaustively defining possible states of an element.
$\mathbb{Y} = \{C_1, C_2\}$ binary classification

$\mathbb{Y} = \{C_1, C_2, ..., C_K\}$ multiclass classification

Reminder (Types of Data)



Bernouli Distribution: A probability distribution for the Binary random Variables

$Y \sim \text{Bernouli}(\pi)$

Support of the random variable: $Y \in \{0, 1\}$
Probabilities: $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$

$E(Y) = \pi$ and $Var(Y) = \pi(1 - \pi)$.

Terminology: Often, in terms of notation we say:

"$Y = 1$" $\equiv$ "*Success*" / "*Survived*" / "*Accepted*" / "*Win*"

"$Y = 0$" $\equiv$ "*Failure*" / "*NotSurvived*" / "*NotAccepted*" / "*Loss*"

Bernouli Distribution: A probability distribution for the Binary random Variables

$Y \sim \text{Bernouli}(\pi)$

Support of the random variable: $Y \in \{0, 1\}$
Probabilities: $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$

$E(Y) = \pi$ and $Var(Y) = \pi(1 - \pi)$

.

Terminology: Often, in terms of notation we say:

"$Y = 1$" $\equiv$ "*Success*" / "*Survived*" / "*Accepted*" / "*Win*"

"$Y = 0$" $\equiv$ "*Failure*" / "*NotSurvived*" / "*NotAccepted*" / "*Loss*"

.

Definition: Odds of “Success"

Odds: The odds of success are defined as the ratio of the probability of success (say, π) over the probability of failure ($1 - \pi$).

$$\text{Odds(“Success")} = \frac{\pi}{1-\pi}$$

Example: Odds

Example: In a Game of rolling a fair dice twice. A player wins if both throws result in same number.

$Prob(Win) = \pi = \frac{6}{36} = \frac{1}{6}$, $Prob(Loss) = 1 - \pi = 1 - \frac{1}{6}$

$$\text{Odds(“Win")} = \frac{Prob(Win)}{Prob(Loss)} = \frac{\pi}{1-\pi} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}$$

Example: Odds

Example: In a Game of rolling a fair dice twice. A player wins if both throws result in same number.

$Prob(Win) = \pi = \frac{6}{36} = \frac{1}{6}$, $Prob(Loss) = 1 - \pi = 1 - \frac{1}{6}$

$$\text{Odds(“Win")} = \frac{Prob(Win)}{Prob(Loss)} = \frac{\pi}{1-\pi} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}$$

Example: Odds

Example: Probability of 15 years "survival" of a patient after a critical medical procedure is "0.9"

$$Prob(Survival) = \pi = 0.9, \qquad 1 - \pi = 1 - 0.9 = 0.1$$

$$\text{Odds("Survival")} = \frac{Prob(Win)}{Prob(Loss)} = \frac{\pi}{1-\pi} = \frac{0.9}{0.1} = 9$$

Odds of 15 year survival of the patient after the medical procedure is 9 to 1

Example: Odds

Example: Probability of 15 years “survival" of a patient after a critical medical procedure is “0.9”

$$Prob(Survival) = \pi = 0.9, \qquad 1 - \pi = 1 - 0.9 = 0.1$$

$$\text{Odds(“Survival")} = \frac{Prob(Win)}{Prob(Loss)} = \frac{\pi}{1-\pi} = \frac{0.9}{0.1} = 9$$

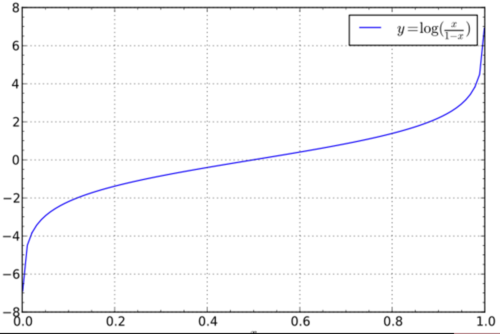
Odds of 15 year survival of the patient after the medical procedure is 9 to 1

Definition: Logit Function

Logit: Logit is the natural log of an odds; That is, the logit of a number π between 0 and 1 is given by the formula:

$$\boxed{\text{Logit}(\pi)} = \log\left(\text{Odds}(\pi)\right) = \boxed{\log\left(\frac{\pi}{1-\pi}\right)}$$

The range of Logit(π) is $-\infty$ to ∞ when the range of π is between 0 and 1.

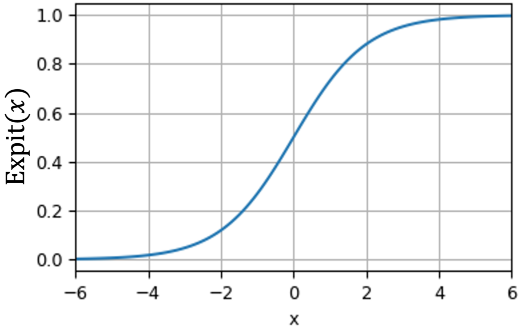


Definition: Expit Function

Expit: Expit of a number x between $-\infty$ and ∞ is given by:

$$\boxed{\text{Expit}(x)} = \boxed{\frac{e^x}{1+e^x}}$$

The range of Expit(x) is 0 to 1 when the range of x is between $-\infty$ and ∞.



‘Expit’ and ’Logit’ are Inverse Function to each other

$$\text{Expit}(x) = \pi \implies x = \text{Logit}(\pi)$$

$$\text{Logit}(\pi) = x \implies \pi = \text{Expit}(x)$$

‘Expit’ and ’Logit’ are Inverse Function to each other

In a logistic regression, Logit(π) is modelled instead of a direct modelling of π.

Logistic Regression

Why There is a Need Logistic Regression ?

We have already discussed the **linear regression model and their estimation.**

Why do we need additional regression methods?

- In standard linear regression we assume that the **Response variable is continuous.**
- What if the Response variable is **categorical?** in nature?

Examples of Binary Variables

Let's imagine independent observations $y_1, \ldots, y_n$ of a variable of interest Y, which can take **only two values** (e.g. 0 and 1), e.g.:

- Workpiece is defective(1)/not defective(0)
- Banking: A customer is considered creditworthy: yes(1)/no(0)
- Health Science: Therapy is successful: yes(1)/no(0).

: Is the standard linear regression **appropriate** for modeling binary response/dependent variables?

Reminder: The SLR and Response Type

- The random sample of size n is denoted as $\{\underset{\sim}{\mathbf{y}}_i, bx_i\}_{i=1}^{n}$ where the respone $\underset{\sim}{\mathbf{y}}_i \in \mathbb{R}$.
- The simple linear regression line is given by

$$Y = \alpha + \beta X + \varepsilon. \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2)$$

- The simple linear regression is not applicable when $\underset{\sim}{\mathbf{y}}_i$ NOT continuous.
- Specifically, we use the logistic regression when the data is of following type: random sample of size n is denoted as $\{\underset{\sim}{\mathbf{y}}_i, \underset{\sim}{\mathbf{x}}_i\}_{i=1}^{n}$ where $\underset{\sim}{\mathbf{y}}_i \in \{\mathbf{0}, \mathbf{1}\}$ are responses and $\underset{\sim}{\mathbf{x}}_i$ pertains to numerical or continuous covariates.

Reminder: The SLR and Response Type

- The random sample of size n is denoted as $\{\underset{\sim}{\mathbf{y}}_i, bx_i\}_{i=1}^{n}$ where the respone $\underset{\sim}{\mathbf{y}}_i \in \mathbb{R}$.

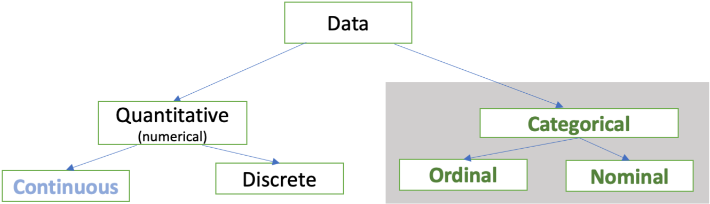
- The simple linear regression line is given by

$$Y = \alpha + \beta X + \varepsilon. \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2)$$

- The simple linear regression is not applicable when $\underset{\sim}{\mathbf{y}}_i$ NOT continuous.

- Specifically, we use the logistic regression when the data is of following type: random sample of size n is denoted as $\{\underset{\sim}{\mathbf{y}}_i, \underset{\sim}{\mathbf{x}}_i\}_{i=1}^{n}$ where $\underset{\sim}{\mathbf{y}}_i \in \{\mathbf{0}, \mathbf{1}\}$ are responses and $\underset{\sim}{\mathbf{x}}_i$ pertains to numerical or continuous covariates.

Different Data Types



- For standard Linear regression: Response variable is Continuous.
- Methods are available for Discrete random variables. Also, the variations of the continuous random variables are also possible.

- The Logistic regression is applicable when: Response variable is Categorical.
- Multicategory logistic regression are available when more than two categories are present in the Ordinal, Nominal Response variable.

- Standard Logistic regression is commonly used to analyze binary data.
- Other Binary regression models: Probit regression model, Complementary log-log model.

Reminder: Bernouli Distribution to model the Response

$Y \sim \text{Bernouli}(\pi)$

Support of the random variable: $Y \in \{0, 1\}$
Probabilities: $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$

In-stead of modeling Y directly we model the probabilities π

.

In particular: In a logistic regression, $\text{Logit}(\pi)$ is assumed to be a function of the covariates.

The Logistic Regression

- The random sample of size n is denoted as $\{\underset{\sim}{\mathbf{y}}_i, bx_i\}_{i=1}^{n}$ where $\underset{\sim}{\mathbf{y}}_i \in \{0, 1\}$ and $\underset{\sim}{\mathbf{x}}_i$ can be of any type numerical or categorical.

$\underset{\sim}{\mathbf{y}}_i \sim \text{Bernouli}(\pi_i), 0 < \pi_i < 1.$

$\pi_i = Prob(Y_i = 1)$

Model: $\boxed{\text{Logit}\,(\pi_i) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}}$

A Few Results & Terminology in Logistic Regression

‘Logit Link’ and Probability of “Success” Logistic Regression

‘Logit’ and ’Expit’

$$\text{Logit}\left(\pi_i\right) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}$$

$$\implies \pi_i = \text{Expit}\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right)$$

‘Logit’ and ’Expit’

$$\text{Logit}\,(\pi_i) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}$$

$$\implies \quad \boxed{\pi_i = \text{Expit}\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right)}$$

Explanation of 'Logit' Link

$$\pi_{\boldsymbol{i}} = \text{Expit}\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right) = \frac{\exp\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right)}{1 + \exp\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right)}$$

‘Logit’ and ’Odds’ in Logistic Regression

Reminder: Odds of “Success"

Odds: The odds of success are defined as the ratio of the probability of success (say, π) over the probability of failure $(1 - \pi)$.

$$\text{Odds(“Success")} = \frac{\pi}{1 - \pi}$$

‘Logit’ and ’Oddds’

$\text{Logit}\,(\pi_i) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta} \implies \log\left(\frac{\pi_i}{1-(\pi_i)}\right) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}$

$\implies \log\left(\text{Odds}(\pi_i)\right) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta} \implies \text{Odds}(\pi_i) = e^{\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}}$

$\text{Logit}\,(\pi_i) = \underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta} \implies \text{Odds}(\pi_i) = \exp\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\beta}\right)$

'Odds' and the 'Logit' Link

$$\boxed{\text{Odds}(\pi_i)} = \exp\left(\underset{\sim}{\mathbf{x}}_i^T \underset{\sim}{\boldsymbol{\beta}}\right)$$

Interpretation of a Regression Coefficient: Intercept

Interpretation of the Intercept

Let the fitted model be

$$\text{Logit}\,(\pi_i) = \boxed{\beta_0} + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\boxed{\beta_0} + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}\right)$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\boxed{\beta_0} + \beta_1 \times 0 + \ldots + \beta_p \times 0\right) = \exp\left(\boxed{\beta_0}\right)$$

when $x_{i,1} = 0, x_{i,2} = 0, \ldots, x_{i,p} = 0$.

e^{β_0} : Odds that the corresponding Response is 1 when all the covariate values are zero.

Interpretation of the Intercept

Let the fitted model be

$$\text{Logit}\left(\pi_i\right) = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}\right)$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\beta_0 + \beta_1 \times 0 + \ldots + \beta_p \times 0\right) = \exp\left(\beta_0\right),$$

when $x_{i,1} = 0, x_{i,2} = 0, \ldots, x_{i,p} = 0$.

e^{β_0} : Odds that the corresponding Response is 1 when all the covariate values are zero.

Interpretation of the Intercept

Let the fitted model be

$$\text{Logit}(\pi_i) = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p}\right)$$

$$\implies \text{Odds}(\pi_i) = \exp\left(\beta_0 + \beta_1 \times 0 + \ldots + \beta_p \times 0\right) = \exp\left(\beta_0\right),$$

when $x_{i,1} = 0, x_{i,2} = 0, \ldots, x_{i,p} = 0$.

e^{β_0} : Odds that the corresponding Response is 1 when all the covariate values are zero.

Interpretation of a Regression Coef.: A Numerical Covariate

Interpretation of a regression coefficient: numerical variable

$$\text{Logit}(\pi_i) = \beta_0 + \boxed{\beta_1}\, x_{i,1} + \ldots + \beta_p x_{i,p}$$

Assuming: $x_{i,1}$ to be a numerical covariate.

$$\text{Odds}(\pi_i)\,\big|_{x_{i,1}=\,x} = \exp\left(\beta_0 + \beta_1\, x + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\text{Odds}(\pi_i)\,\big|_{x_{i,1}=\,x+1} = \exp\left(\beta_0 + \beta_1\,(x+1) + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\frac{\text{Odds}(\pi_i)\,\big|_{x_{i,1}=\,x+1}}{\text{Odds}(\pi_i)\,\big|_{x_{i,1}=\,x}} = \exp(\beta_1).$$

Interpretation of a regression coefficient: numerical variable

$$\text{Logit}\,(\pi_i) = \beta_0 + \beta_1\, x_{i,1} + \ldots + \beta_p x_{i,p}$$

Assuming: $x_{i,1}$ to be a numerical covariate.

$$\text{Odds}(\pi_i)\,\big|_{X_{i,1}=\,x} = \exp\left(\beta_0 + \beta_1\, x + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\text{Odds}(\pi_i)\,\big|_{X_{i,1}=\,x+1} = \exp\left(\beta_0 + \beta_1 (\, x+1\,) + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\frac{\text{Odds}(\pi_i)\,\big|_{X_{i,1}=\,x+1}}{\text{Odds}(\pi_i)\,\big|_{X_{i,1}=\,x}} = \exp(\beta_1).$$

Interpretation of a regression coefficient corresponding to a numerical variable

$$\frac{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x+1}}{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x}} = \exp(\beta_1).$$

$$\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x+1} = \exp(\beta_1)\,\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x}.$$

e^{β_1} : If all other variables remain fixed, the Odds of "Succcess ($Y_i = 1$)" is changed by the factor e^{β_1} when the $x_{1,i}$ is **increased by** 1.

Interpretation of a regression coefficient corresponding to a numerical variable

$$\frac{\text{Odds}(\pi_i)\mid_{x_{i,1}=\,x+1}}{\text{Odds}(\pi_i)\mid_{x_{i,1}=\,x}} = \exp(\beta_1).$$

$$\text{Odds}(\pi_i)\mid_{x_{i,1}=\,x+1} = \exp(\beta_1)\,\text{Odds}(\pi_i)\mid_{x_{i,1}=\,x}.$$

e^{β_1} : If all other variables remain fixed, the Odds of "Succcess ($Y_i = 1$)" is changed by the factor e^{β_1} when the $x_{1,i}$ is **increased by** 1.

Finally: Interpretation of a regression coefficient corresponding to a numerical variable

$$\frac{\text{Odds}(\pi_i)\mid_{x_{i,1}=\ x+1}}{\text{Odds}(\pi_i)\mid_{x_{i,1}=\ x}} = \exp(\boxed{\beta_1}).$$

$$\frac{\text{Odds}(\pi_i)\mid_{x_{i,1}=\ x+1} - \text{Odds}(\pi_i)\mid_{x_{i,1}=\ x}}{\text{Odds}(\pi_i)\mid_{x_{i,1}=\ x}} \times 100\% = \left(\boxed{e^{\beta_1}} - 1\right) \times 100\%.$$

$(e^{\beta_1} - 1) \times 100\%$: If all other variables remain fixed, the Odds of "Success ($Y_i = 1$)" is increased (decreased if it is a negative number) by $(e^{\beta_1} - 1) \times 100\%$ when the variable $x_{1,i}$ is **increased by** 1.

Finally: Interpretation of a regression coefficient corresponding to a numerical variable

$$\frac{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x+1}}{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x}} = \exp(\beta_1).$$

$$\frac{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x+1} - \text{Odds}(\pi_i)\,|_{x_{i,1}=\,x}}{\text{Odds}(\pi_i)\,|_{x_{i,1}=\,x}} \times 100\% = \left(e^{\beta_1} - 1\right) \times 100\%.$$

$(e^{\beta_1} - 1) \times 100\%$: If all other variables remain fixed, the Odds of "Success ($Y_i = 1$)" is increased (decreased if it is a negative number) by $(e^{\beta_1} - 1) \times 100\%$ when the variable $x_{1,i}$ is **increased by** 1.

Interpretation of Regression Coefficient: A Categorical Covariate

Interpretation of a reg. coeff. : Categorical Covariate

$$\text{Logit}(\pi_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2\, x_{i,2} + \ldots + \beta_p x_{i,p}$$

Assume: $x_{i,2}$ to be a Categorical Gender variable. $x_{i,2} = 1$ for Female and $x_{i,2} = 0$ for Male.

$$\text{Odds}(\pi_i)\ |_{'Male'} = \exp\left(\beta_0 + \beta_1 x_{i,1} + \beta_2\, 0 + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\text{Odds}(\pi_i)\ |_{'Female'} = \exp\left(\beta_0 + \beta_1 x_{i,1} + \beta_2\,(1) + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\frac{\text{Odds}(\pi_i)\ |_{'Female'}}{\text{Odds}(\pi_i)\ |_{'Male'}} = \exp(\beta_2).$$

$$\text{Logit}\,(\pi_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2\, x_{i,2} + \ldots + \beta_p x_{i,p}$$

Assume: $x_{i,2}$ to be a Categorical Gender variable. $x_{i,2} = 1$ for Female and $x_{i,2} = 0$ for Male.

$$\text{Odds}(\pi_i)\mid_{'Male'} = \exp\left(\beta_0 + \beta_1 x_{i,1} + \beta_2\, 0 + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\text{Odds}(\pi_i)\mid_{'Female'} = \exp\left(\beta_0 + \beta_1 x_{i,1} + \beta_2\,(1) + \beta_2 x_{i,2} \ldots + \beta_p x_{i,p}\right)$$

$$\frac{\text{Odds}(\pi_i)\mid_{'Female'}}{\text{Odds}(\pi_i)\mid_{'Male'}} = \exp(\beta_2).$$

Interpretation: Regression Coefficient for a Categorical Covariate

$$\frac{\text{Odds}(\pi_i)\ |_{\text{'}Female\text{'}}}{\text{Odds}(\pi_i)\ |_{\text{'}Male\text{'}}} = \exp(\beta_2).$$

$$\text{Odds}(\pi_i)\ |_{\text{'}Female\text{'}} = \exp(\beta_2)\ \text{Odds}(\pi_i)\ |_{\text{'}Male\text{'}}.$$

e^{β_2} : Assuming all other variables remain fixed, the Odds of "Success ($Y_i = 1$)" for a datapoint corresponding to the category '*Female*' is e^{β_2} times the Odds of a data point with the category '*Male*'

Interpretation: Regression Coefficient for a Categorical Covariate

$$\frac{\text{Odds}(\pi_i)\,\big|_{\text{'}Female\text{'}}}{\text{Odds}(\pi_i)\,\big|_{\text{'}Male\text{'}}} = \exp(\beta_2).$$

$$\text{Odds}(\pi_i)\,\big|_{\text{'}Female\text{'}} = \exp(\beta_2)\,\text{Odds}(\pi_i)\,\big|_{\text{'}Male\text{'}}.$$

e^{β_2} : Assuming all other variables remain fixed, the Odds of "Success ($Y_i = 1$)" for a datapoint corresponding to the category '*Female*' is e^{β_2} times the Odds of a data point with the category '*Male*'

Finally: Interpretation: Regression Coefficient for a Categorical Covariate

$$\frac{\text{Odds}(\pi_i)\mid_{\text{'Female'}}}{\text{Odds}(\pi_i)\mid_{\text{'Male'}}} = \exp(\beta_2).$$

$$\frac{\text{Odds}(\pi_i)\mid_{\text{'Female'}} - \text{Odds}(\pi_i)\mid_{\text{'Male'}}}{\text{Odds}(\pi_i)\mid_{\text{'Male'}}} \times 100\% = \left(e^{\beta_1} - 1\right) \times 100\%.$$

$(e^{\beta_1} - 1) \times 100\%$: If all other variables remain fixed, the Odds/Chance of "Success ($Y_i = 1$)" for a *'Female'* is $(e^{\beta_1} - 1) \times 100\%$ **more** (**less** if it is a negative number) compared to the corresponding Odds/Chance for a *'Male'*.

Finally: Interpretation: Regression Coefficient for a Categorical Covariate

$$\frac{\text{Odds}(\pi_i)\,|_{\text{'}Female\text{'}}}{\text{Odds}(\pi_i)\,|_{\text{'}Male\text{'}}} = \exp(\beta_2).$$

$$\frac{\text{Odds}(\pi_i)\,|_{\text{'}Female\text{'}} - \text{Odds}(\pi_i)\,|_{\text{'}Male\text{'}}}{\text{Odds}(\pi_i)\,|_{\text{'}Male\text{'}}} \times 100\% = \left(e^{\beta_1} - 1\right) \times 100\%.$$

$(e^{\beta_1} - 1) \times 100\%$: If all other variables remain fixed, the Odds/Chance of "Success ($Y_i = 1$)" for a '*Female*' is $(e^{\beta_1} - 1) \times 100\%$ **more** (**less** if it is a negative number) compared to the corresponding Odds/Chance for a '*Male*'.

Interpretation of coefficients: Summary

- The direct interpretation of the $\hat{\beta}$ is highly nontrivial and often non-intuitive.
- Therefore, in Logistic regression, it is common practice to interpret the $\exp(\hat{\beta})$ which has a straightforward interpretation with Odds.

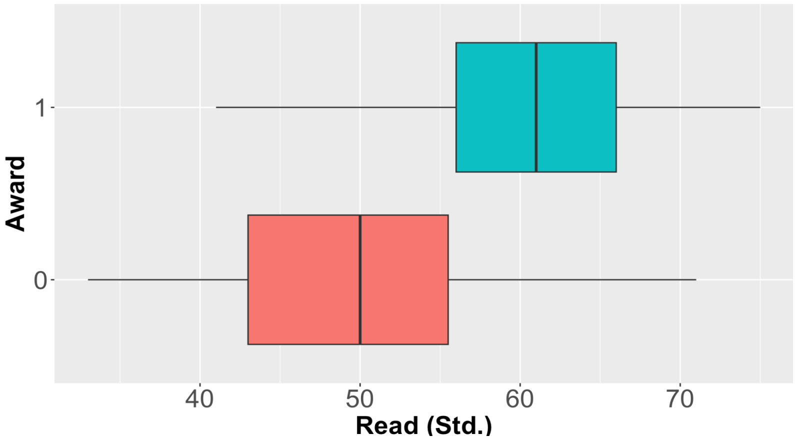
$\hat{\beta}$	$\text{Odds}(\pi_i) = \frac{P(Y_i=1)}{P(Y_i=0)}$	%change in Odds	$P(Y_i = 1)$
$\hat{\beta} > 0$	increases by $\exp(\hat{\beta})$	$(\exp(\hat{\beta}) - 1) \times 100\%$	increases
$\hat{\beta} < 0$	decreases by $\exp(\hat{\beta})$	$(\exp(\hat{\beta}) - 1) \times 100\%$	decreases
$\hat{\beta} = 0$	remains the same	0%	same

A Data Example

Awards Data

The Dataset have two columns. 'Award' and 'Read'

- $'Award'$: Indicates whether a student obtained the Award or not.
- $'Read'$: Hours the student have spend in additional literature study.
- We want to model the response 'Award' based on the numerical covariate 'Read'.


Award
1
0
40
50
60
70
Read (Std.)

$model.lr$<-$glm(Award \sim Read, data = performance, family = ''binomial'')$

$summary(model.lr)$

```
Call:
glm(formula = Award ~ Read, family = "binomial", data = performance)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0332  -0.6785  -0.3506  -0.1565   2.6143

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.79394    1.48174  -6.610 3.85e-11 ***
Read         0.15634    0.02561   6.105 1.03e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 222.71  on 199  degrees of freedom
Residual deviance: 167.07  on 198  degrees of freedom
AIC: 171.07

Number of Fisher Scoring iterations: 5
```

Reminder: Interpretation of coefficients

- The direct interpretation of the $\hat{\beta}$ is highly nontrivial and often non-intuitive.
- Therefore, in Logistic regression, it is common practice to interpret the $\exp(\hat{\beta})$ which has a straightforward interpretation with Odds.

$\hat{\beta}$	$\text{Odds}(\pi_i) = \frac{P(Y_i=1)}{P(Y_i=0)}$	%change in Odds	$P(Y_i = 1)$
$\hat{\beta} > 0$	increases by $\exp(\hat{\beta})$	$(\exp(\hat{\beta}) - 1) \times 100\%$	increases
$\hat{\beta} < 0$	decreases by $\exp(\hat{\beta})$	$(\exp(\hat{\beta}) - 1) \times 100\%$	decreases
$\hat{\beta} = 0$	remains the same	0%	same

Illustration: interpretation of results

- Model to be estimated:
  $P(AWARD = 1 \mid READ) = \frac{\exp(\beta_0 + \beta_1 * Read)}{1 + \exp(\beta_0 + \beta_1 * Read)}$

- Estimated parameter: $\hat{\beta}_1 = 0.15634$ (Regression coefficient corresponding to the variable 'Read')

$e^{\hat{\beta}_1} = \exp(0.15634) = 1.1692$: For each hour increment in the variable 'Read' (Study time aiming the award), on average, the Odds/chance for a student to win the prize in increased by a factor of 1.1692 times.

$(e^{\hat{\beta}_1 - 1}) \times 100\% = 16.92\%$: For each hour increment in the variable 'Read' (Study time aiming the award), on average, a student have 16.92 % more Odds/chance to win the prize.

Illustration: interpretation of results

- Model to be estimated:
  $P(AWARD = 1 \mid READ) = \frac{\exp(\beta_0 + \beta_1 * Read)}{1 + \exp(\beta_0 + \beta_1 * Read)}$

- Estimated parameter: $\hat{\beta}_1 = 0.15634$ (Regression coefficient corresponding to the variable 'Read')

$e^{\hat{\beta}_1} = \exp(0.15634) = 1.1692$: For each hour increment in the variable 'Read' (Study time aiming the award), on average, the Odds/chance for a student to win the prize in increased by a factor of 1.1692 times.

$(e^{\hat{\beta}_1 - 1}) \times 100\% = 16.92\%$: For each hour increment in the variable 'Read' (Study time aiming the award), on average, a student have 16.92 % more Odds/chance to win the prize.

Illustration: interpretation of results

- Model to be estimated:
  $P(AWARD = 1 \mid READ) = \frac{\exp(\beta_0 + \beta_1 * Read)}{1 + \exp(\beta_0 + \beta_1 * Read)}$
- Estimated parameter: $\hat{\beta}_0 = -9.79394$ (Regression coefficient corresponding to the variable 'Read')

$e^{\hat{\beta}_0} = \exp(-9.79394) < 0.00006$: On an average, the Odds/Chance for a student to win the Award is less than 0.00006 if he/she spends 0 hours of Reading.

Illustration: interpretation of results

What is the estimated probability of wining the prize for a student, if she/she spends 70 hours of Read?

$\frac{\exp(-9.8+0.16*70)}{1+\exp(-9.8+0.16*70)} = 0.8$: The estimated probability of wining the prize for a student, if she/she spends 70 hours of Read is 0.80.

THANK YOU