# Predictive Analytics

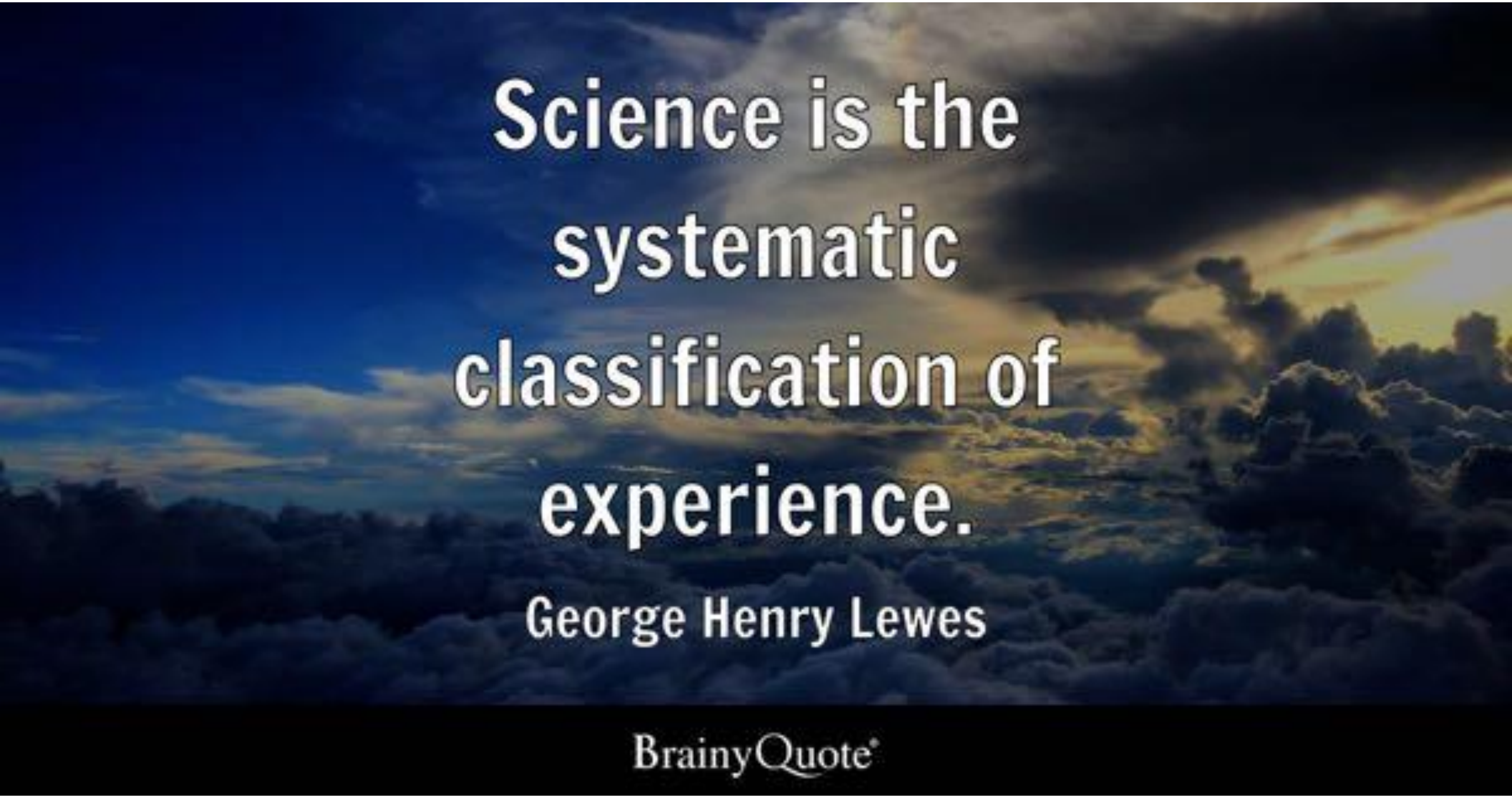
## Classification Modeling

**Dr. Tanujit Chakraborty**

@ Sorbonne

tanujitisi@gmail.com

# Quote of the day..



# This presentation includes…

- Logistic Regression
  - Binomial logistic regression
  - Multinomial logistic regression

# Logistic Regression

# Regression Analysis and Logistic Regression

- **Regression analysis**
  - Based on the principle of **Least Square Estimation (LSE)**
    - The parameters are chosen to minimize the sum of squared errors (SSE)
  - Minimizes error in prediction
    - If the error distribution is normal with constant variance, the LSE estimates the parameters accurately; that is, model is the best possible and with the smallest standard errors
  - Applicable when dependent variable follows normal distribution
- **Logistic regression analysis**
  - When a dependent variable does not follow normal distribution
  - Value of a dependent variable  may be with 2, 3 or a fewer more outcomes

# Regression Analysis and Logistic Regression

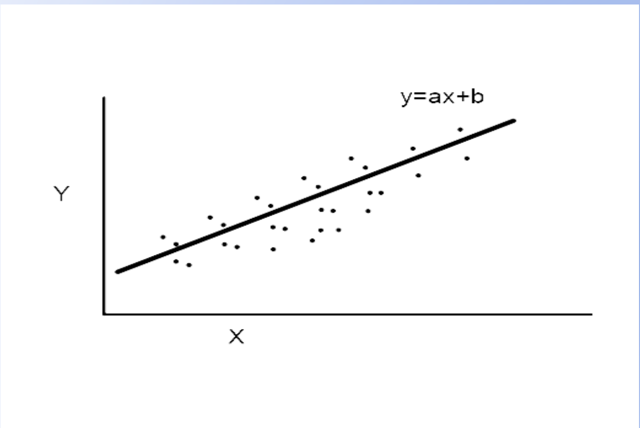
- **Logistic regression**
  - Considers maximum likelihood estimation (MLE) to give better result.
  - In MLE, the likelihood is the probability of the observed data set given a set of proposed values for the parameters.
  - The principle of MLE is to estimate parameters by choosing parameter values that give the largest possible likelihood.

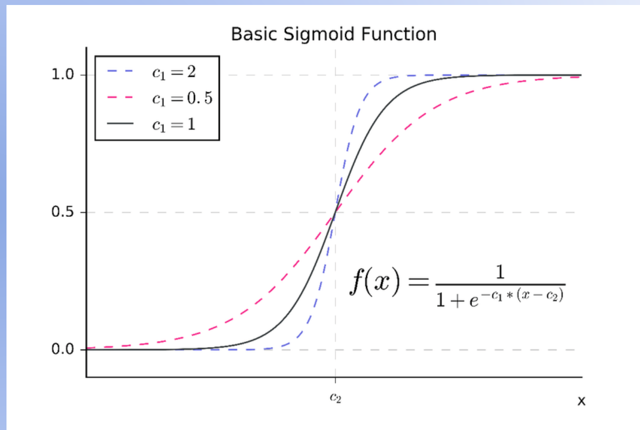
**Note**

- Regression analysis  predicts a value of a dependent variable
- Logistic regression predicts the probability of a given value of a dependent variable
- Both estimates their respective model parameters

# Regression Analysis and Logistic Regression

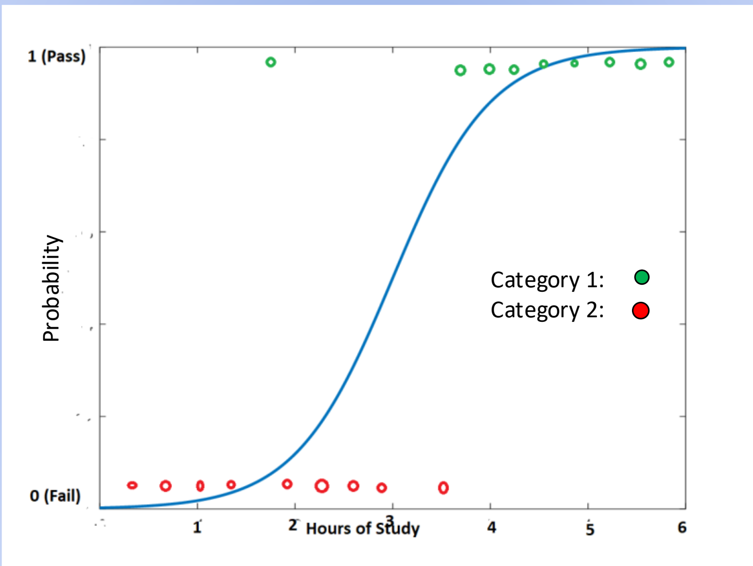
## A Regression model and Logistic Regression model



# An Example

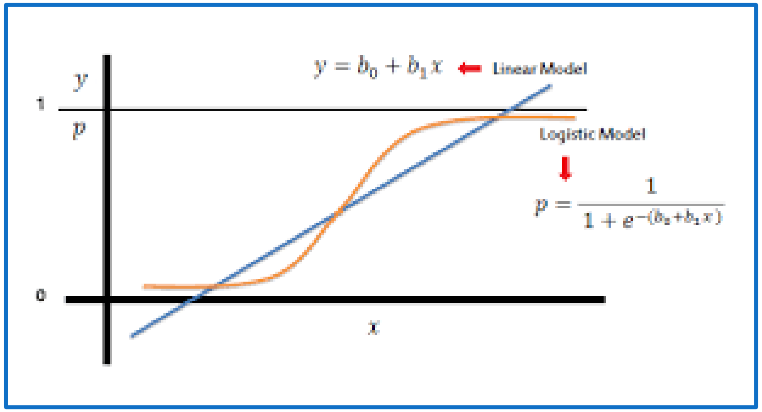
| Hours ($x_i$) | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pass ($y_i$) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |



# Concept of Logistic Regression

## Introduction

- Developed and popularized primarily by *Joseph Berkson* in *(1944)*, where he coined the term **logit.**
- Logistic regression is a statistical method.
  - uses a logistic function to model a binary dependent variable.
  - although many more complex extensions exist.
- Logistic regression (or logit regression) estimates the parameters of a logistic model.



# Concept of Logistic Regression

## Introduction

- A binary logistic model has a dependent variable with two possible values, such as *Pass* ***or*** *Fail*, *Happy* ***or*** *Sad* etc.
- It is represented by an indicator variable, where the two values are labeled '1' and '0'


1 0

# Concept of Logistic Regression

## What is logistic regression?

- The logistic function, takes the following form

$$p(x_1,\ldots,x_m) = p_x = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_m x_m}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_m x_m}}$$

- Binary regression: A case when $p$ has two outcomes: success and failure
- It defines **odds**, which is the ratio of the probability of success and failure

$$odds = \frac{p_x}{1-p_x} = e^{\beta_0+\beta_a x_1+\cdots+\beta_m x_m}$$

- The logarithm of the odds (called logit) is

$$t = ln(odds) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

$$\boldsymbol{t = ln\left(\frac{p_x}{1-p_x}\right)}$$

$$\boldsymbol{p_x = \frac{e^t}{1+e^t}}$$

# An Illustration

A sample is collected to examine the effect of toxic substance on tumor. A subject is examined for the toxic content in the body and then the presence (1) or absence (0) of tumors. The independent variable is the concentration of the toxic substance "Conc" . The number of subjects at each concentration (N) and the number having tumors "Tumor" is shown in the table.

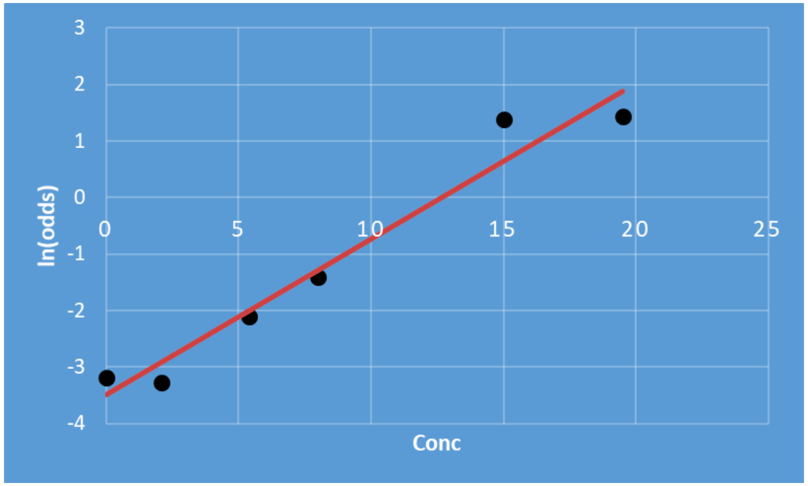
Odds and ln(odds) are also included in the table.

| Conc | N | Tumor | Odds | ln(odds) |
|---|---|---|---|---|
| 0.0 | 50 | 2 | 0.0417 | -3.18 |
| 2.1 | 54 | 5 | 0.1020 | -3.28 |
| 5.4 | 46 | 5 | 0.1220 | -2.10 |
| 8.0 | 51 | 10 | 0.2439 | -1.41 |
| 15.0 | 50 | 40 | 4.0000 | +1.39 |
| 19.5 | 52 | 42 | 4.2000 | +1.44 |

$$t = ln\frac{p}{1-p}$$

# An Illustration

| Conc | N | Tumor | Odds | ln(odds) |
|---|---|---|---|---|
| 0.0 | 50 | 2 | 0.0417 | -3.18 |
| 2.1 | 54 | 5 | 0.1020 | -3.28 |
| 5.4 | 46 | 5 | 0.1220 | -2.10 |
| 8.0 | 51 | 10 | 0.2439 | -1.41 |
| 15.0 | 50 | 40 | 4.0000 | +1.39 |
| 19.5 | 52 | 42 | 4.2000 | +1.44 |




Here, we find a relation between t (ln(odds) and x (Conc).

$$t = \beta_0 + \beta_1 x$$

Thus, logit is a linear function

For this, it can be calculated as

$\beta_0$= -3.204 and $\beta_1$= 0.2628

# An Illustration

Here, we find a relation between t (ln(odds) and x (Conc).

$$t = \beta_0 + \beta_1 \mathrm{x}$$

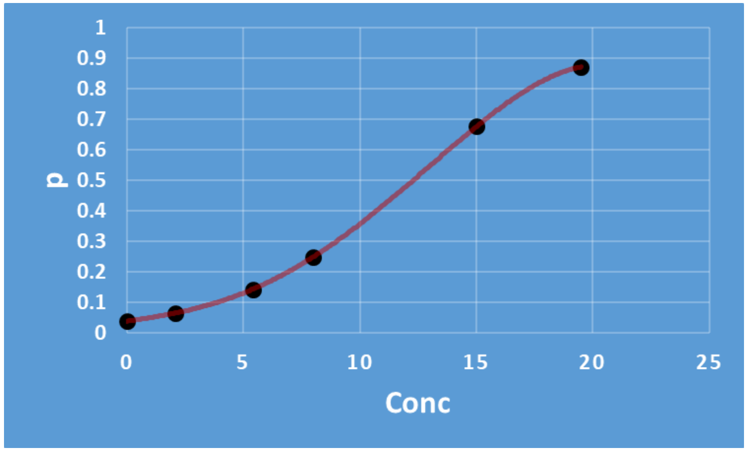
Thus, logit is a linear function

For this, it can be calculated as

$\beta_0$= -3.204 and $\beta_1$= 0.2628

For an example, a subject exposed to a concentration of 10, has an estimated probability of tumor is

$$p_{10} = \frac{e^t}{1+e^t} = 0.36$$



# Concept of Logistic Regression

### Logit in Logistic Regression

- The log-odds (the logarithm of the odds) is a linear combination of
  - one or more independent variables ("predictors").
  - the independent variables can each be a continuous variable (any real value).
- The probability of the value can vary between 0 (such as certainly false) and 1 (such as certainly true).

# Concept of Logistic Regression

## Output of logistic function

- Increasing one of the independent variables, multiplicatively scales the odds of the given outcome at a constant rate, with each independent variable having its own parameter;
  - for a binary dependent variable this generalizes the odds ratio.
- Binary logistic regression: The dependent variable has two levels (categorical).
- Multinomial logistic regression: Outputs with more than two levels.
- Ordinal logistic regression: if the multiple categories are ordered, then it is called ordinal logistic regression.

# Logistic Regression as Classifier

Logistic regression models the probabilities for classification problems with possible outcomes.

- It's an extension of the linear regression model for classification problems.

- The logistic regression simply models probability of output in terms of input and does not perform statistical classification (it is not a classifier).
- Although it can be used to make a classifier.
  - By choosing a cut-off value and classifying inputs with probability greater than the cutoff as one class, below the cut-off as the other

# Logistic Regression Techniques

# Types of Logistic Regression

$$p(x_1, ..., x_m) = p = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m}}$$

- Logistic regression
  - Binary logistic regression
    - One explanatory variable, two categories
    - Many explanatory variable, two categories
  - Multinomial logistic regression
    - Many explanatory variables, many categories

# Types of Logistic Regression

**Case 1: One explanatory variable, two categories**

- Explanatory variable:
  X: Hours Study
- Outcome: Pass or Fail

**Case 2: Many explanatory variable, two categories**

- Explanatory variable
  X1: Hours Study     X2: 12th % Marks
- Outcome: Pass or Fail

**Case 3: Many explanatory variable, many categories**

- Explanatory variable:
  X1: Hours Study     X2: 12th % Marks     X3: Age
- Outcome: Bad, Good, Excellent

# Binary Logistic Regression

- Logistic Regression Analysis
  - Binary Logistic Regression
    - One explanatory variable, two categories

# One explanatory variable, two categories

# One Explanatory Variable, Two Categories

A group of 20 students spends between 0 and 6 hours studying for an exam.

The table shows the number of hours each student spent studying, and whether they passed (1) or failed (0).

| Hours ($x_i$) | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Pass ($y_i$) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

| Hours ($x_i$) | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Pass ($y_i$) | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

**How does the number of hours spent studying affect the probability of the student passing the exam?**

# One Explanatory Variable, Two Categories

**How does the number of hours spent studying affect the probability of the student passing the exam?**

| Hours ($x_i$) | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pass ($y_i$) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

## Note

- The reason for using logistic regression for this problem is that the values of the dependent variable, pass and fail, while represented by "1" and "0", are not cardinal numbers.
- If the problem was changed so that pass/fail was replaced with the grade 0 to 100 (cardinal numbers), then simple regression analysis could be used.

# One Explanatory Variable, Two Categories

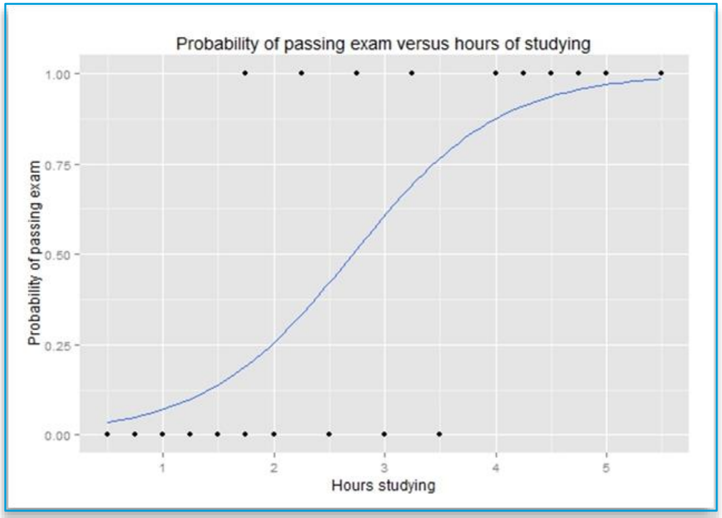
| Hours ($x_i$) | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Pass ($y_i$)** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

- We wish to fit a logistic function to the data consisting of the hours studied ($x_i$) and the outcome of the test ($y_i = 1$ for pass, 0 for fail).
- The data points are indexed by the subscript i which runs from $i = 1$ to $20$ $(= n)$. The $x$ variable is called the "explanatory variable", and the $y$ variable is called the "**categorical variable**" consisting of two categories: "**pass**" or "**fail**" corresponding to the categorical values 1 and 0, respectively.

# One Explanatory Variable, Two Categories

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

$$t = log_b \frac{p}{1 - p}$$




- Graph of a logistic regression curve fitted to the $(x_i, y_i)$ data. The curve shows the probability of passing an exam versus hours studying.

# One Explanatory Variable, Two Categories

- The logistic function is of the form:

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

  where $\mu$ is a location parameter (the midpoint of the curve, where $p(\mu) = {}^1/_2$ ) and $s$ is a scale parameter.

- This expression may be rewritten as:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

  where $\beta_0 = -{}^{\mu}/_{s}$ and is known as the intercept, and $\beta_1 = {}^1/_s$ known as slope.

- We may define the “fit” to $y_i$ at a given $x_i$ as:

$$p_i = p(x_i)$$

# One Explanatory Variable, Two Categories

- The $p_i$ are the probabilities that the corresponding $y_i$ will be unity and $1 - pi$ are the probabilities that they will be zero.

- We wish to find the values of $\beta_0$ and $\beta_1$ which give the *"best fit"* to the data.
  - In the case of linear regression, the sum of the squared deviations of the fit from the data points ($y_i$) is taken as a measure of the goodness of fit, and the best fit is obtained when that function is minimized.

# One Explanatory Variable, Two Categories

- In the case of logistic regression, the measure of goodness of fit is given by the likelihood function, which is the probability that the given data set is produced by a particular logistic function:

$$L = \prod_{i:y_i=1} p_i \prod_{i:y_i=0} (1 - p_i)$$

and the best fit is obtained for those choices of $\beta_0$ and $\beta_1$ where $L$ is maximized.

# One Explanatory Variable, Two Categories

- The maximum of $L$ will also be the maximum of the log-likelihood $\ell$, defined as the logarithm of $L$ :

$$\ell = \sum_{i:y_i=1} \ln(p_i) + \sum_{i:y_i=0} \ln(1-p_i)$$

$$= \sum_{i=1}^{n} \left(y_i \ln(p_i) + (1-y_i)\ln(1-p_i)\right)$$

$$p_i = p(xi) = \frac{1}{1+e^{-}(\beta_0 + \beta_1 x_i)}$$

- Here, $\ell$ is nonlinear in $\beta_0$ and $\beta_1$. Determining their optimum values will require numerical methods.

# One Explanatory Variable, Two Categories

- Note that one method of maximizing $\ell$ is to require the derivatives of $\ell$ with respect to $\beta_0$ and $\beta_1$ to be zero:

$$0 = \frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^{n} (y_i - p_i)$$

$$0 = \frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} (y_i - p_i) x_i$$

  and the maximization procedure can be accomplished by solving the above two equations for $\beta_0$ and $\beta_1$, which again, will generally require the use of numerical methods.

$$p_i = p(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

# One Explanatory Variable, Two Categories

| Hours ($x_i$) | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 | 3.00 | 3.25 | 3.50 | 4.00 | 4.25 | 4.50 | 4.75 | 5.00 | 5.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Pass ($y_i$)** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

The values of $\beta_0$ and $\beta_1$ which maximize $\ell$ using the above data are found to be:

$$\beta_0 = -4.07771 \text{ and } \beta_1 = 1.50465$$

which yields a value for $\mu$ of:

$$\mu = -\beta_0/\beta_1 = 2.71008$$

The logistic regression analysis gives the following output.

| | Coefficient | Std. Error | z-value | p-value (Wald) |
|---|---|---|---|---|
| **Intercept ($\beta_0$)** | −4.0777 | 1.7610 | −2.316 | 0.0206 |
| **Slope ($\beta_1$)** | 1.5046 | 0.6287 | 2.393 | 0.0167 |

- By the Wald test, the output indicates that hours studying is significantly associated with the probability of passing the exam ($p = 0.0167$)

# One Explanatory Variable, Two Categories

| | Coefficient | Std. Error | z-value | p-value (Wald) |
|---|---|---|---|---|
| Intercept ($\beta_0$) | −4.0777 | 1.7610 | −2.316 | 0.0206 |
| Slope ($\beta_1$) | 1.5046 | 0.6287 | 2.393 | 0.0167 |

- The $\beta_0$ and $\beta_1$ coefficients may be entered into the logistic regression equation to estimate the probability of passing the exam.
- For example, for a student who studies 2 hours, entering the value $x = 2$ into the equation gives the estimated probability of passing the exam of 0.26 :

$$t = \beta_0 + 2\beta_1 = -4.0777 + 2 \times 1.5046 = -1.0685$$

$$p = \frac{1}{1 + e^{-t}} = 0.26 = \text{ Probability of passing exam}$$

# One Explanatory Variable, Two Categories

- Similarly, for a student who studies 4 hours, the estimated probability of passing the exam is 0.87:

$$t = \beta_0 + 4\beta_1 = -4.0777 + 4 \times 1.5046 = 1.9407$$

$$p = \frac{1}{1+e^{-t}} = 0.87 = \text{ Probability of passing exam}$$

- This table shows the probability of passing the exam for several values of hours studying.

| Hours of study (x) | Passing exam | | |
|---|---|---|---|
| | Log-odds (t) | Odds ($e^t$) | Probability (p) |
| **1** | −2.57 | 0.076 ≈ 1:13.1 | 0.07 |
| **2** | −1.07 | 0.34 ≈ 1:2.91 | 0.26 |
| **μ=2.71…** | 0 | 1 | 0.5 |
| **3** | 0.44 | 1.55 | 0.61 |
| **4** | 1.94 | 6.96 | 0.87 |
| **5** | 3.45 | 31.4 | 0.97 |

# One Explanatory Variable, Two Categories

- This simple model is known as *"binary logistic regression"*, and has one explanatory variable and a categorical variable which can assume one of two categorical values.

- The above example of binary logistic regression on one explanatory variable can be generalized to binary logistic regression on any number of explanatory variables $\boldsymbol{x_1, x_2, \ldots}$ and any number of categorical values $\boldsymbol{y = 0, 1, 2, \ldots}$.

# Many explanatory variable, two categories

- Logistic Regression Analysis
  - Binary Logistic Regression
    - Many explanatory variable, two categories

# Many Explanatory Variable, Two Categories

- To begin with, we may consider a logistic model with $M$ explanatory variables, $x_1, x_2$ ... $x_M$ and two categorical values ($y = 0$ and $1$).
- For the simple binary logistic regression model, we assumed a linear relationship between the predictor variable and the log-odds (also called logit) of the event that $y = 1$.
- This linear relationship may be extended to the case of $M$ explanatory variables:

$$t = \log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_M x_M$$

where $t$ is the log-odds and $\beta_i$ are parameters of the model.

- An additional generalization has been introduced in which the base of the model ***b*** is not restricted to the **Euler number *e***.
  - In most applications, the base ***b*** of the logarithm is usually taken to be ***e***.
  - However, in some cases it can be easier to communicate results by working in **base 2** or **base 10**.

# Many Explanatory Variable, Two Categories

- For a more compact notation, we will specify the explanatory variables and the $\boldsymbol{\beta}$ coefficients as $(\boldsymbol{M}+\mathbf{1})$-dimensional vectors:

$$\boldsymbol{x} = \{x_0, x_1, x_2, \dots, x_M\}$$

$$\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2, \dots, \beta_M\}$$

with an added explanatory variable $\boldsymbol{x_0} = \mathbf{1}$. The logit may now be written as:

$$t = \sum_{m=0}^{M} \beta_m x_m = \beta \cdot x$$

# Many Explanatory Variable, Two Categories

- Solving for the probability $\boldsymbol{p}$ that $\boldsymbol{y} = \mathbf{1}$ yields:

$$p(\boldsymbol{x}) = \frac{b^{\beta \cdot \boldsymbol{x}}}{1 + b^{\beta \cdot x}} = \frac{1}{1 + b^{-\beta \cdot x}} = S_b(t)$$

  where $\boldsymbol{S_b}$ is the sigmoid function with base $\boldsymbol{b}$.

- The above formula shows that once the $\beta$ is fixed, we can easily compute the log-odds that $y = 1$ for a given observation.

- The main use-case of a logistic model is given an observation $\boldsymbol{x}$, estimate the probability $p(\boldsymbol{x})$ that $y = 1$.

# Many Explanatory Variable, Two Categories

- The optimum beta coefficients may again be found by maximizing the log-likelihood.

- For $K$ measurements, defining $x_k$ as the explanatory vector of the $k-th$ measurement, and $y_k$ as the categorical outcome of that measurement, the log likelihood may be written in a form very similar to the simple case of $M = 1$:

$$\ell = \sum_{k=1}^{K} y_k \log_b\left(p(\boldsymbol{x}_k)\right) + \sum_{k=1}^{K} (1 - y_k)\log_b\left(1 - p(\boldsymbol{x}_k)\right)$$

- Finding the optimum $\beta$ parameters will require numerical methods.

# Many Explanatory Variable, Two Categories

- One useful technique is to equate the derivatives of the log likelihood with respect to each of the $\beta$ parameters to zero yielding a set of equations which will hold at the maximum of the log likelihood:

$$\frac{\partial \ell}{\partial \beta_m} = 0 = \sum_{k=1}^{K} y_k x_{mk} - \sum_{k=1}^{K} p(\boldsymbol{x}_k) x_{mk}$$

where $x_{mk}$ is the value of the $x_m$ explanatory variable from the *k-th* measurement.

# Many Explanatory Variable, Two Categories

- Consider an example with $M = 2$ explanatory variables, $b = 10$, and coefficients $\beta_0 = -3$, $\beta_1 = 1$, and $\beta_2 = 2$ which have been determined.

- To be concrete, the model is:

$$t = \log_{10} \frac{p}{1-p} = -3 + x_1 + 2x_2$$

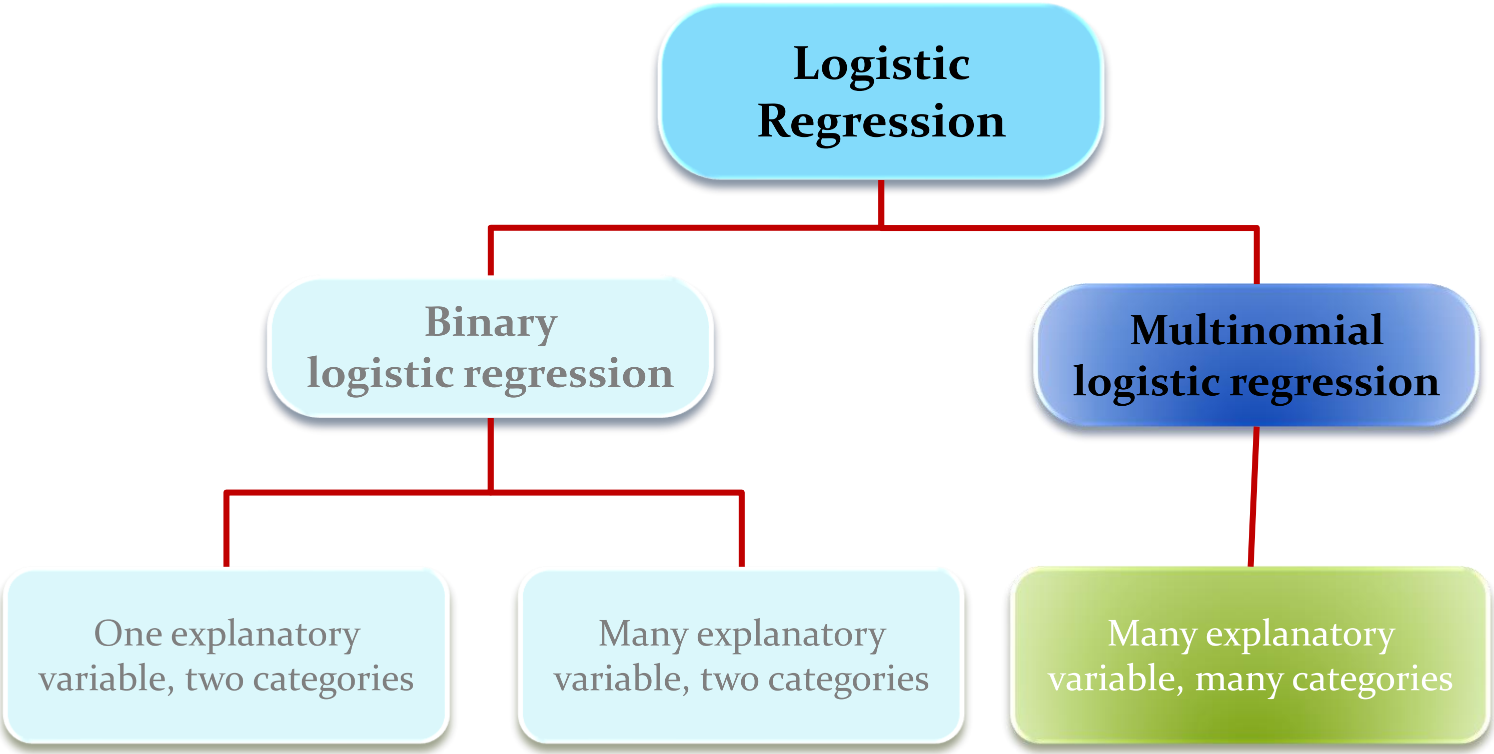
$$p = \frac{b^{\beta \cdot x}}{1 + b^{\beta \cdot x}} = \frac{b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}} = \frac{1}{1 + b^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

  where $p$ is the probability of the event that $y = 1$.

- This can be interpreted as follows:
  - $\beta_0 = -3$ is the *y*-intercept.
  - $\beta_1 = 1$ means that increasing $x_1$ by 1 increases the log-odds by 1.
  - $\beta_2 = 2$ means that increasing $x_2$ by 1 increases the log-odds by 2.

# Multinomial Logistic Regression

# Multinomial Logistic Regression



# Many Explanatory Variable, Many Categories

- In the cases of two categories *(binomial logistic regression)*, the categories were indexed by "0" and "1", and we had two probability distributions:
  - The probability that the outcome was in *category 1* was given by $p(x)$ and
  - The probability that the outcome was in *category 0* was given by $1 - p(x)$.
  - The sum of both probabilities is equal to unity, as they must be.

- In general, if we have $M + 1$ explanatory variables (including $x_0$) and $N + 1$ categories, we will need $N + 1$ separate probability distributions, one for each category, indexed by $n$, which describe the probability that the categorical outcome $y$ for explanatory vector $x$ will be in category $y_n$.

# Many Explanatory Variable, Many Categories

- It will also be required that the sum of these probabilities over all categories be equal to unity.
- Using the mathematically convenient base $\boldsymbol{e}$, these probabilities are:

$$p_n(\boldsymbol{x}) = \frac{e^{\beta_n \cdot \boldsymbol{x}}}{1 + \sum_{u=1}^{N} e^{\beta_u \cdot x}} \text{ for } n = 1,2, \ldots, \mathrm{N}$$

$$p_0(\boldsymbol{x}) = 1 - \sum_{n=1}^{N} p_n(\boldsymbol{x}) = \frac{1}{1 + \sum_{u=1}^{N} e^{\beta_u \cdot x}}$$

- It can be seen that, as required, the sum of the $p_n(\boldsymbol{x})$ over all categories is unity.
- Note that the selection of $p_0(\boldsymbol{x})$ to be defined in terms of the other probabilities is artificial.
- Any of the probabilities could have been selected to be so defined.

# Many Explanatory Variable, Many Categories

- This special value of $n$ is termed the "pivot index," and the log-odds $(t_n)$ are expressed in terms of the pivot probability and are again expressed as a linear combination of the explanatory variables:

$$t_n = \ln\left(\frac{p_n(\boldsymbol{x})}{p_0(\boldsymbol{x})}\right) = \boldsymbol{\beta}_n \cdot \boldsymbol{x}$$

**Note :**

For the simple case of $N = 1$, the two-category case is recovered, with

$$p(x) = p_1(x) \text{ and}$$

$$p_0(x) = 1 - p_1(x)$$

# Many Explanatory Variable, Many Categories

- The log-likelihood that a particular set of $K$ measurements or data points will be generated by the above probabilities can now be calculated.

- Indexing each measurement by $k$, let the $k-th$ set of measured explanatory variables be denoted by $x_k$ and their categorical outcomes be denoted by $y_k$ which can be equal to any integer in $[0, N]$.

# Many Explanatory Variable, Many Categories

- The log-likelihood is then:

$$\ell = \sum_{k=1}^{K} \sum_{n=0}^{N} \Delta(n, y_k) ln\big(p_n(x_k)\big)$$

where $\Delta(n, y_k)$ is an indicator function which is equal to unity if $y_k = n$ and zero otherwise.

- In the case of two explanatory variables, this indicator function was defined as $y_k$ when $n = 1$ and $1 - yk$ when $n = 0$.

# Many Explanatory Variable, Many Categories

- Again, the optimum beta coefficients may be found by maximizing the log-likelihood function generally using numerical methods.

- A possible method of solution is to set the derivatives of the log-likelihood with respect to each beta coefficient equal to zero and solve for the beta coefficients:

$$\frac{\partial \ell}{\partial \beta_{nm}} = 0 = \sum_{k=1}^{K} \Delta(n, y_k) x_{mk} - \sum_{k=1}^{K} p_n(\boldsymbol{x}_k) x_{mk}$$

where $\beta_{nm}$ is the $m$-th coefficient of the $\beta_n$ vector and $x_{mk}$ is the $m-th$ explanatory variable of the $k-th$ measurement.

# Applications of Logistic Regression

# Applications of Logistic Regression

- **In medical domains:**
  - The Trauma and Injury Severity Score (TRISS), which is widely used to predict mortality in injured patients, was originally developed by **Boyd *et al*.** using logistic regression.
  - Many other medical scales used to assess severity of a patient have been developed using logistic regression.
  - Logistic regression may be used to predict the risk of developing a given disease (e.g. diabetes; coronary heart disease), based on observed characteristics of the patient (age, gender, body mass index, results of various blood tests, etc.)

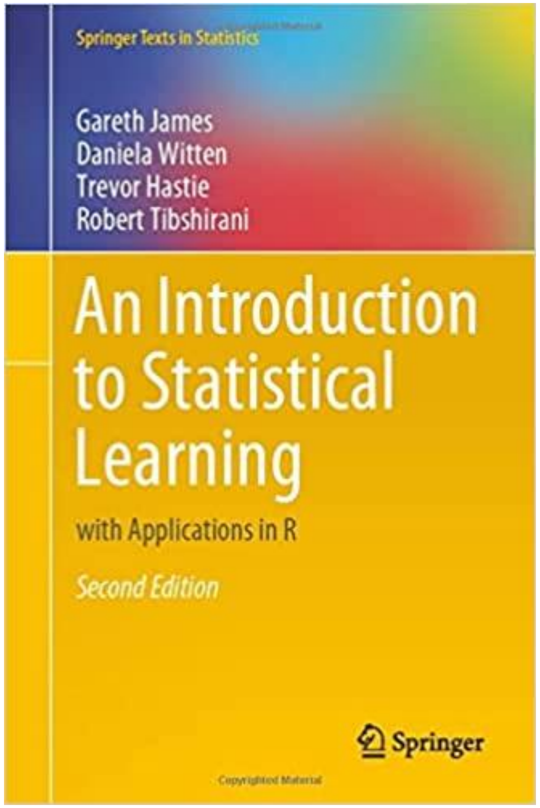
# Applications of Logistic Regression

- **In social sciences:**
  - Another example might be to predict whether an Indian voter will vote National Congress or Communist Party or BJP or Any other party, based on age, income, gender, race, state of residence, votes in previous elections, etc.

- **In engineering:**
  - The technique can also be used in engineering, especially for predicting the probability of failure of a given process, system or product.

- **In marketing:**
  - It is also used in marketing applications, such as prediction of a customer's propensity to purchase a product or halt a subscription, etc.

# References


Springer Texts in Statistics
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani
An Introduction to Statistical Learning
with Applications in R
Second Edition
Springer

# End of Course