EE4211: Data Science for the Internet of Things

Logistic Regression

Biplab Sikdar

Agenda

- Examples
- Logit function
- Fitting a logistic regression
- Understanding the parameters

Example

- Regression analysis: describes the relationship between two (or more) variables
- Examples:
  - Income and educational level
  - Demand for electricity and the weather
  - Tire tread depth and mileage
- Linear regression:
  - The data is represented as ($\epsilon_i \sim N(0, \sigma^2)$):

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

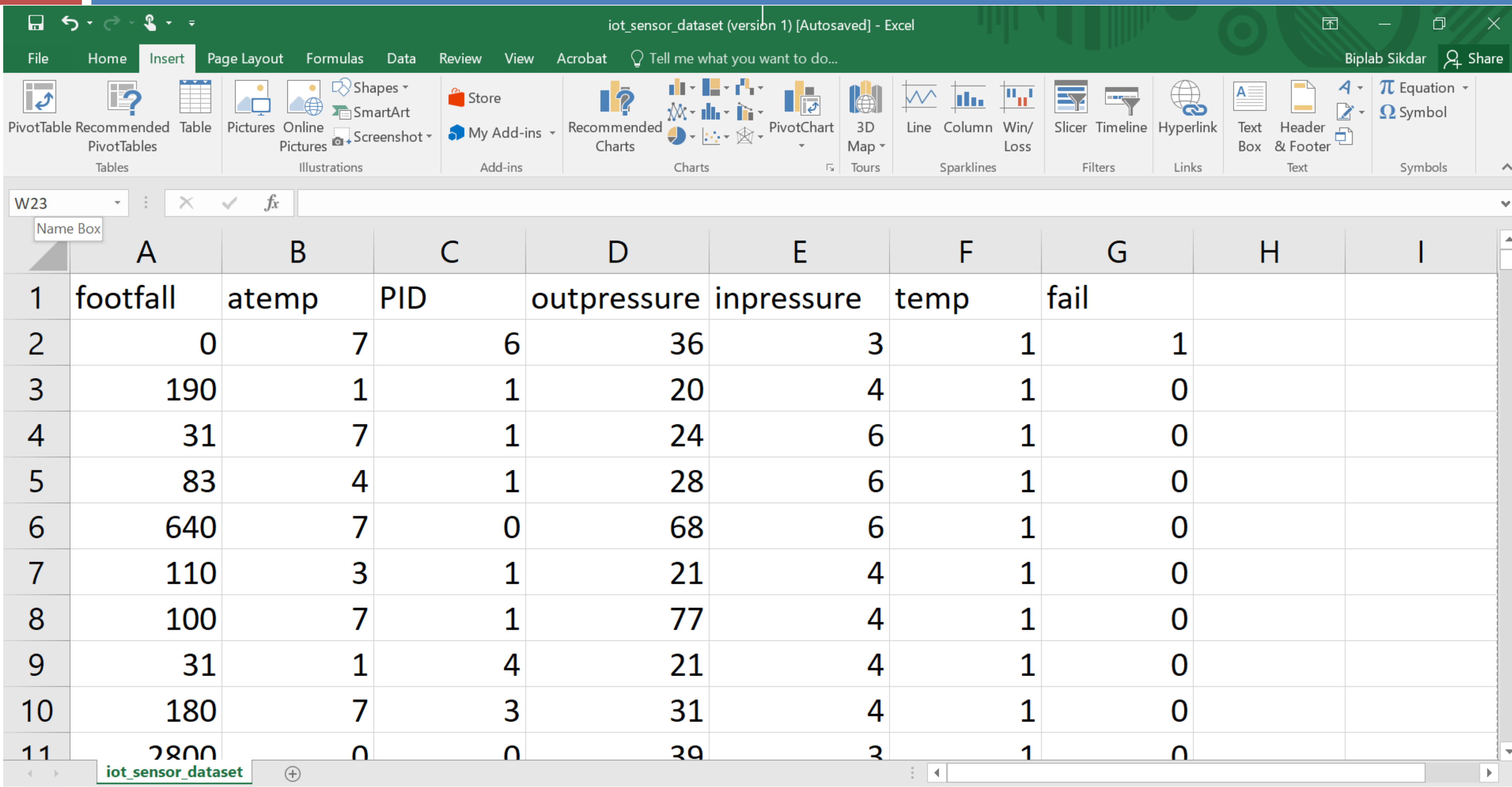
  - The fitted model:

$$\hat{y} = \beta_0 + \beta_1 x$$

Example

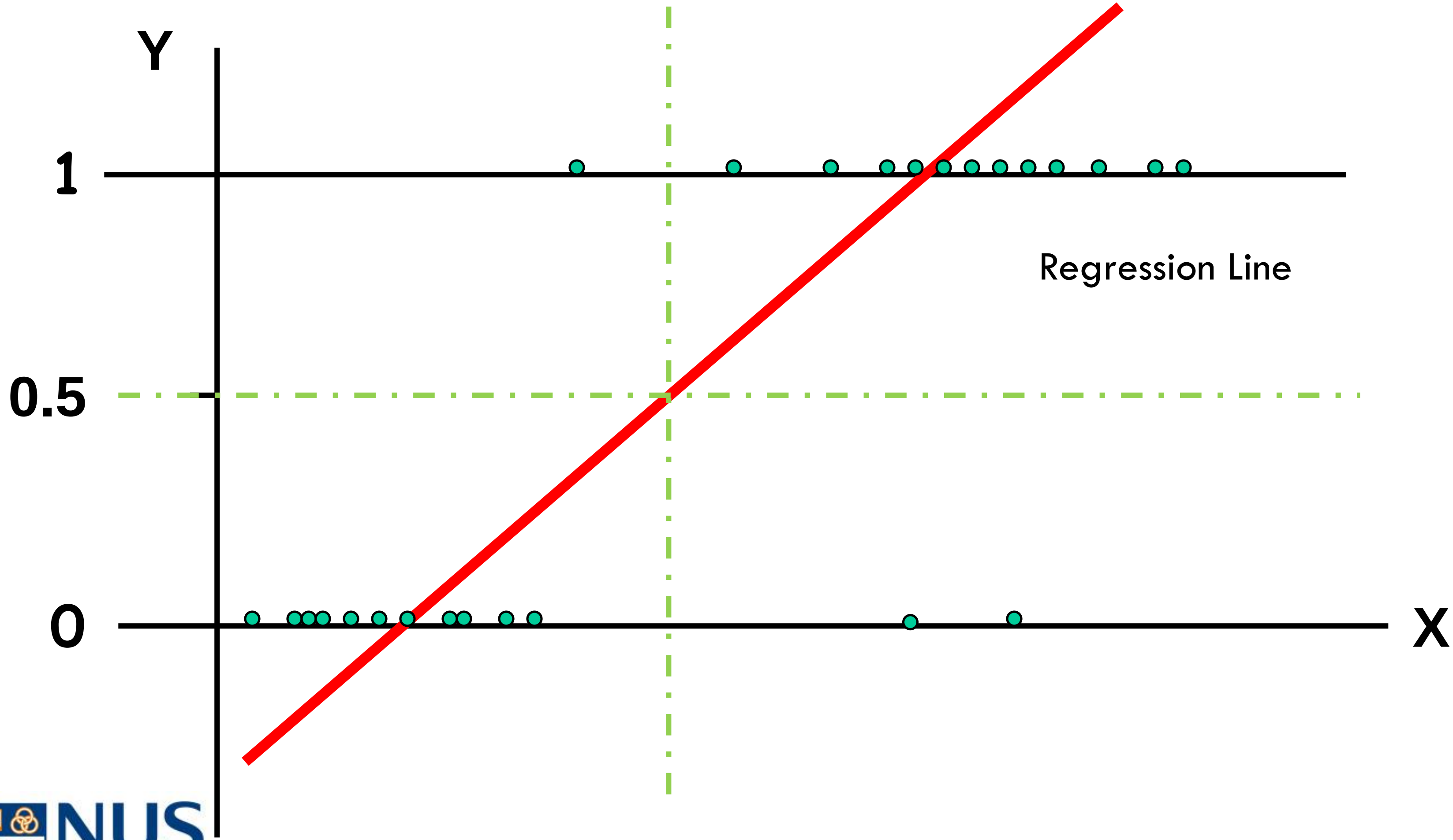
- Typical retail/restaurant chain: hundreds of pieces of equipment in each store, spread across thousands of store locations across geographic locations.
- Preventive maintenance in such scenarios: difficult problem.
- Sensors mounted on IoT devices, equipment, and process monitoring controllers continuously collect and transmit data.
- Rule based systems raise false alarms that turn out to be less serious alerts upon actual inspection, such as an open door on a refrigeration unit.
- Use analytics to learn, enhance, and implement a new method for predicting equipment failure using IoT sensor data.

Example

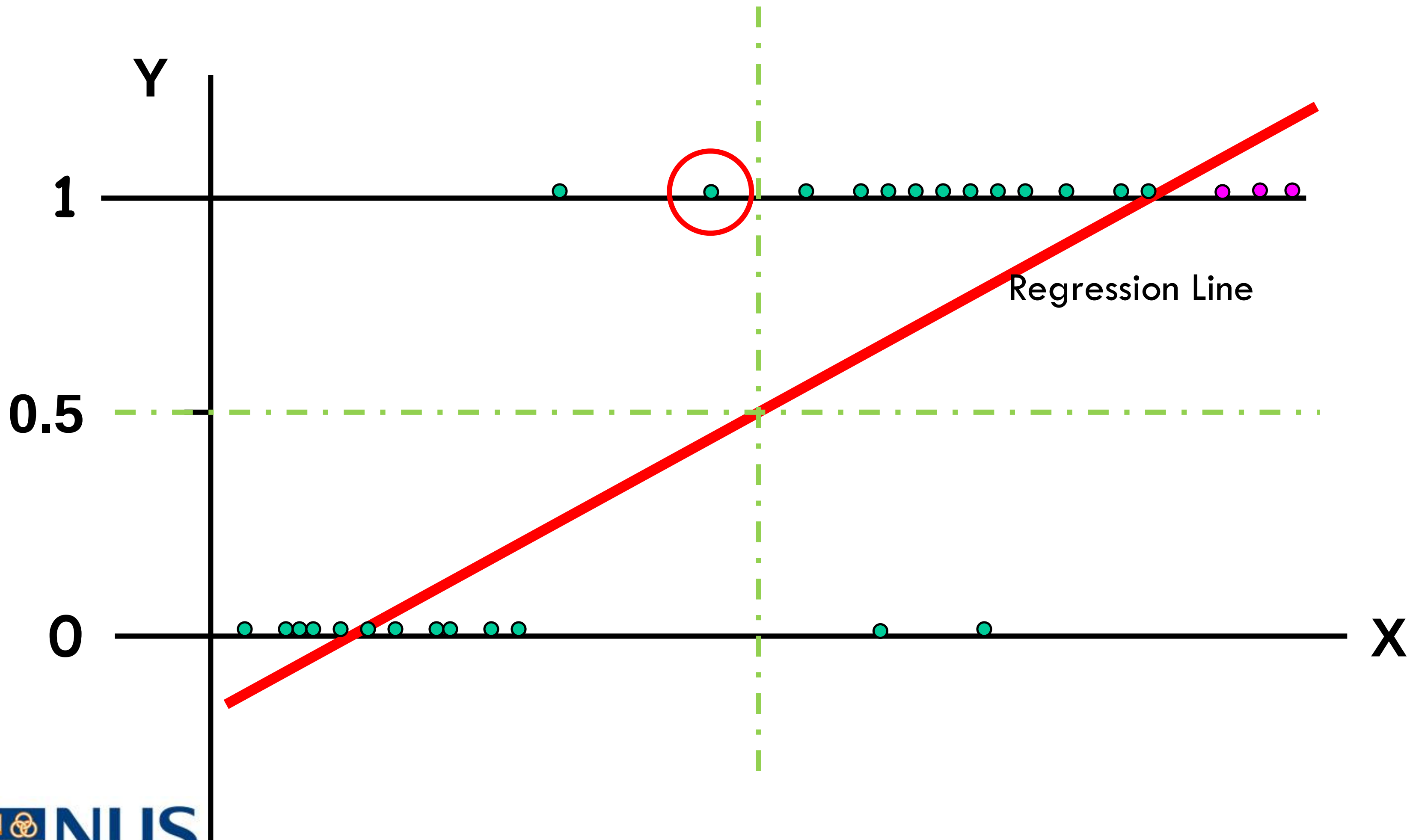


	A	B	C	D	E	F	G
1	footfall	atemp	PID	outpressure	inpressure	temp	fail
2	0	7	6	36	3	1	1
3	190	1	1	20	4	1	0
4	31	7	1	24	6	1	0
5	83	4	1	28	6	1	0
6	640	7	0	68	6	1	0
7	110	3	1	21	4	1	0
8	100	7	1	77	4	1	0
9	31	1	4	21	4	1	0
10	180	7	3	31	4	1	0
11	2800	0	0	39	3	1	0

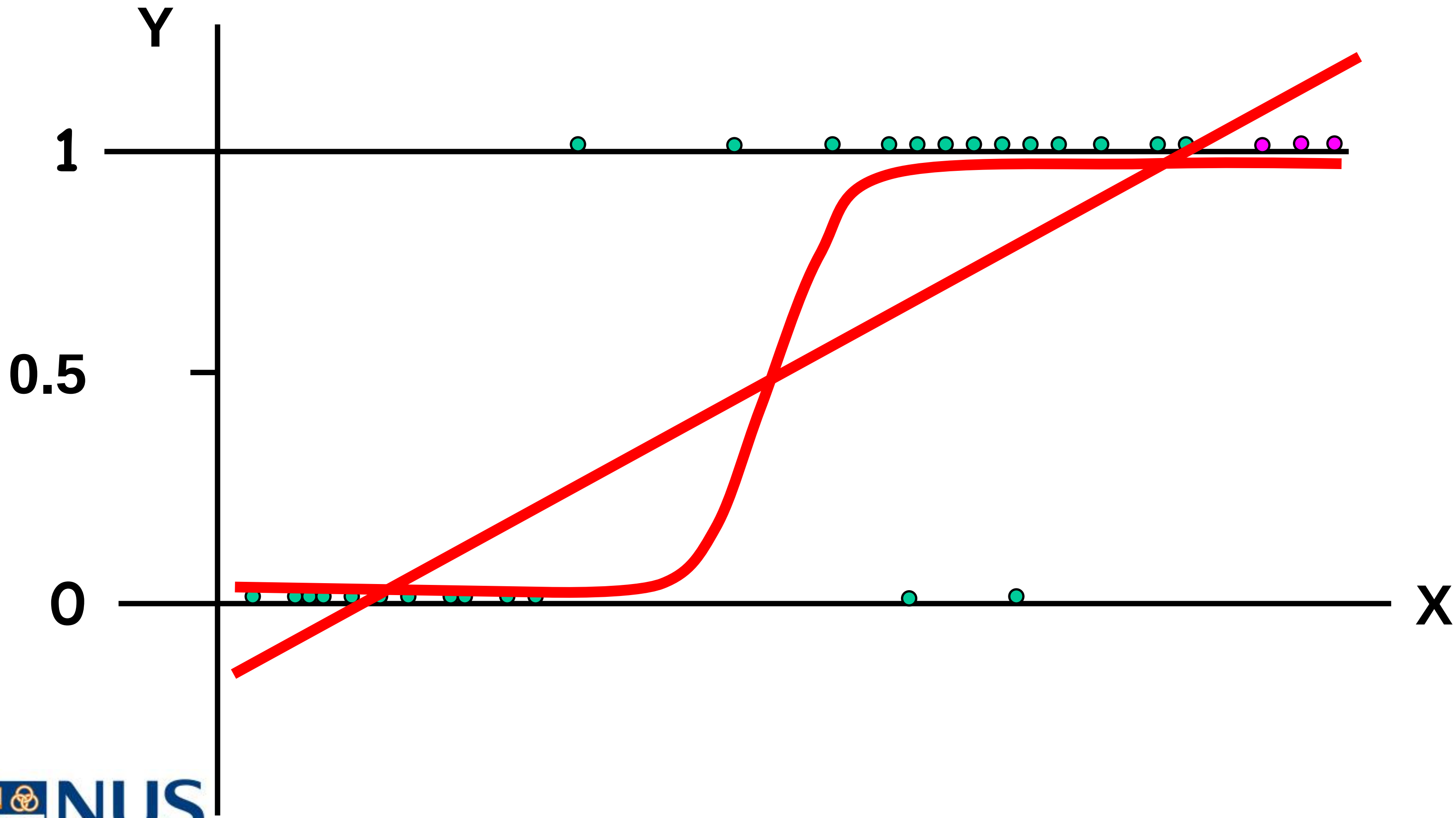
Example



Example

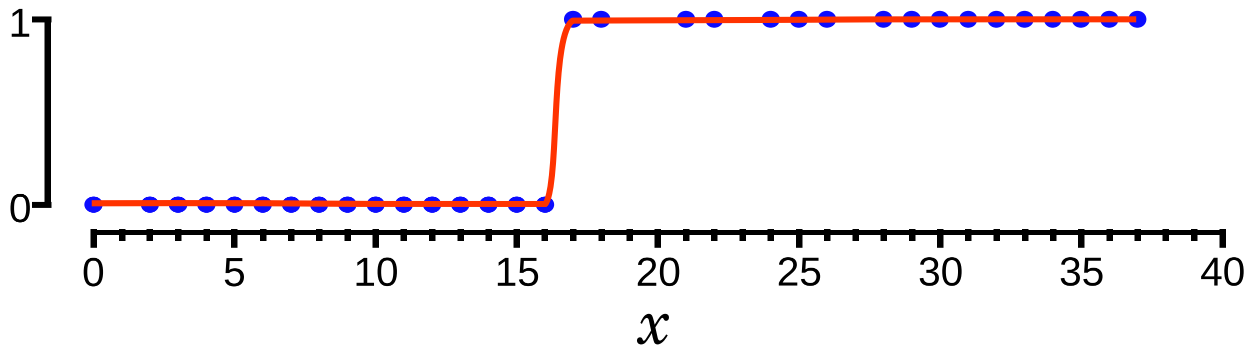


Example

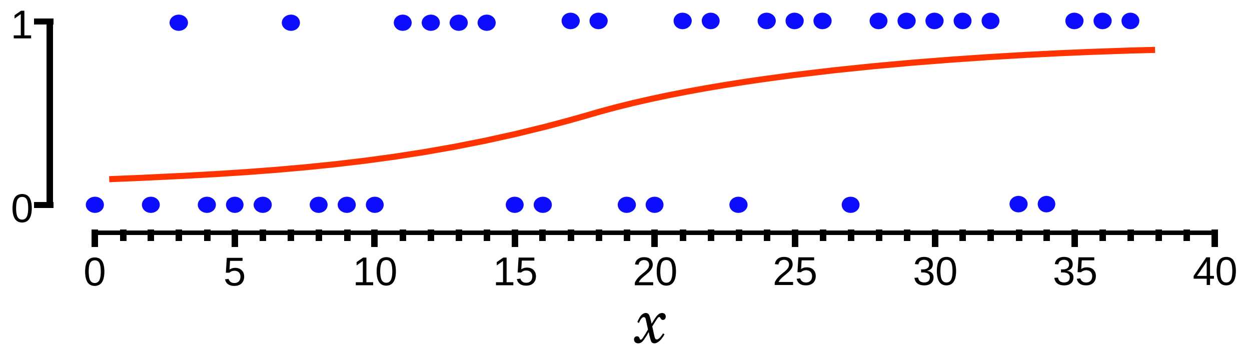


Best Fit Curve

- Data that has a sharp cut off point between categories:



- Data with a lengthy transition between categories:



Logistic Regression

- Response/outcome/dependent variable: y
- Predictor/explanatory/independent variable: x
- Logistic regression extends the ideas of linear regression to the situation where the dependent variable, y, is categorical.
- Categorical variable: divides the observations into classes.
  - y: state of a machine (normal, alarm, failure)
  - Each of the measurements in the dataset (the observations) may be classified as belonging to one of three classes.

Logistic Regression

- Uses a logarithmic transformation to allow a non-linear relationship to be expressed in a linear way.
- Logistic regression can be used for classifying a new observation into one of the classes, based on the values of its predictor variables (called "classification").
- It can also be used in data (where the class is known) to find similarities between observations within each class in terms of the predictor variables (called "profiling").

Logistic Regression Model

- Example:
  - We are interested evaluating the success (y) of a new fire-retardant spray is at preventing "fire" and how it depends on the amount (x) that is applied.
  - The values of y are 1 (Success) or 0 (Failure).
  - The values of x range over a continuum

Logistic Regression Model

Logistic Regression Model

- $p \triangleq P[y = 1] = P[success]$
- p in general will increase with x
- Odds ratio: $\frac{p}{1-p}$
- The odds ration also increases with the value of x, ranging from zero to infinity
- Log odds ratio (logit): $\ln\left(\frac{p}{1-p}\right)$

Example: Odds Ratio

- We roll a regular, six-faced die
- Success: "roll a six"
- $P[success] = p = \frac{1}{6}$
- Odds ratio: $\frac{p}{1-p} = \frac{\frac{1}{6}}{1-\frac{1}{6}} = \frac{1}{5}$
- Log odds ratio (logit): $\ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{1}{5}\right) = -1.69$

NUS National University of Singapore

Better Example: Odds Ratio

MANCHESTER CITY	4/6
LIVERPOOL	11/4
CHELSEA	10/1
TOTTENHAM	16/1
MANCHESTER UTD	25/1

Logistic Regression Model

- The logistic regression model assumes that the log odds ratio is linearly related to x:

$$z \longrightarrow \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

- This can be solved using the technique used solving an ordinary (linear) regression!
- Fit observed data to obtain β_0 and β_1

Logistic Regression Model

- In terms of the odd ratio:

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

- Solving for p in terms of x:

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

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

$$\Rightarrow p + pe^{\beta_0+\beta_1 x} = e^{\beta_0+\beta_1 x}$$

$$\Rightarrow p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

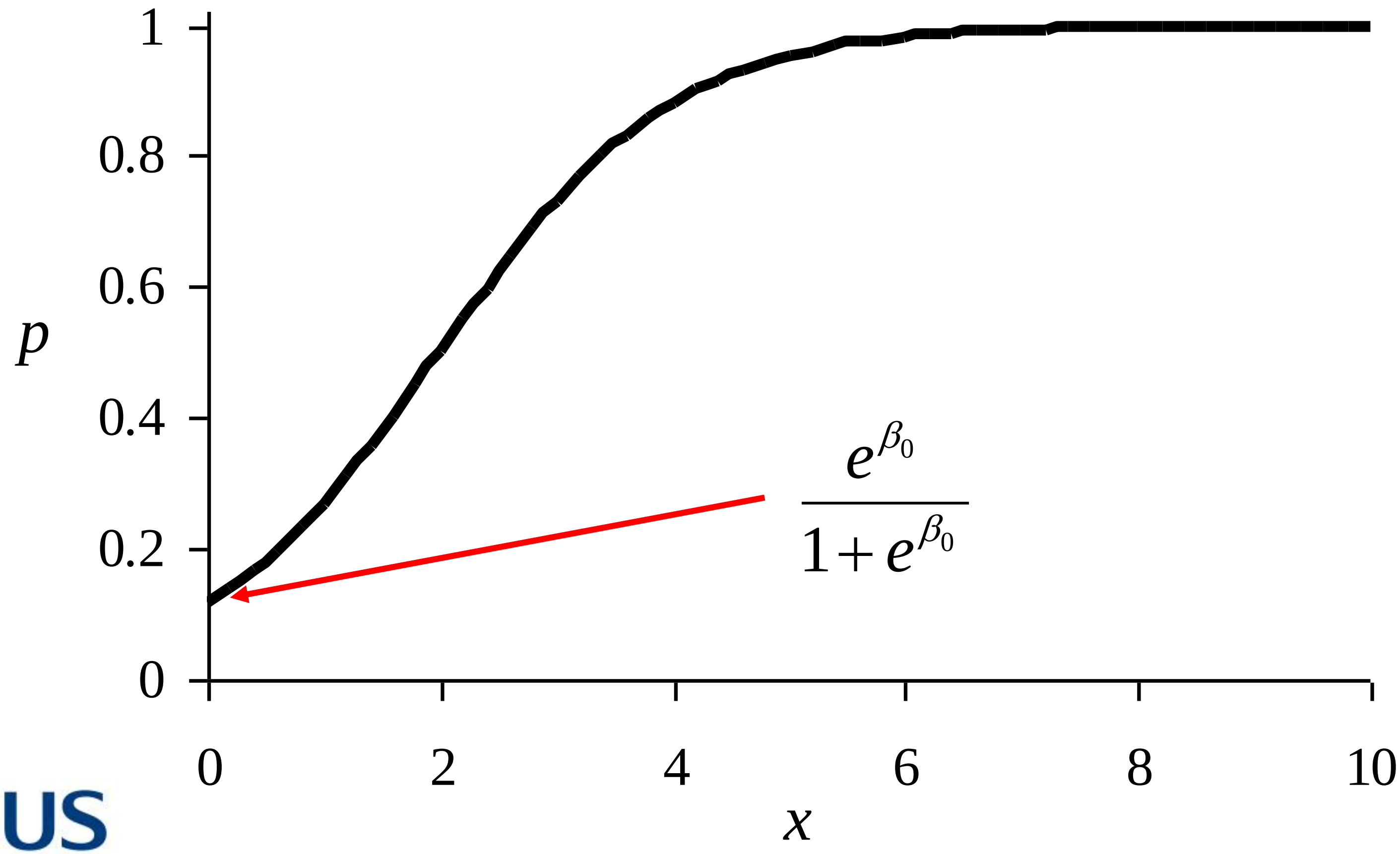
$$\boxed{\Rightarrow p = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}$$

Sigmoid function

NUS National University of Singapore

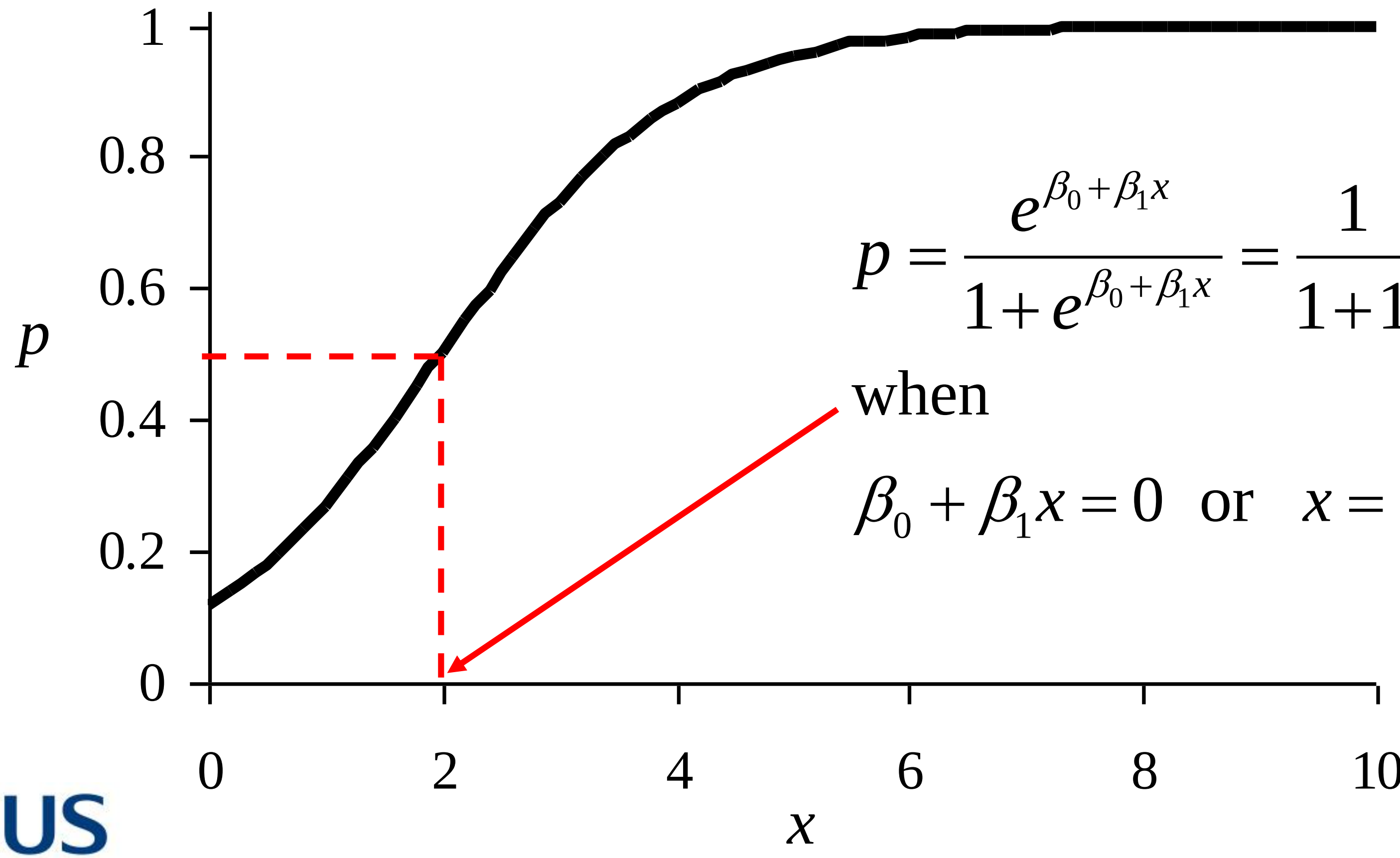
Interpretation: β_0

- β_0: determines the intercept



Interpretation: β_1

- β_1: determines when $p = 0.5$ (along with β_0)



$$p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+1} = \frac{1}{2}$$

when

$$\beta_0 + \beta_1 x = 0 \quad \text{or} \quad x = -\frac{\beta_0}{\beta_1}$$

Interpretation: β_1

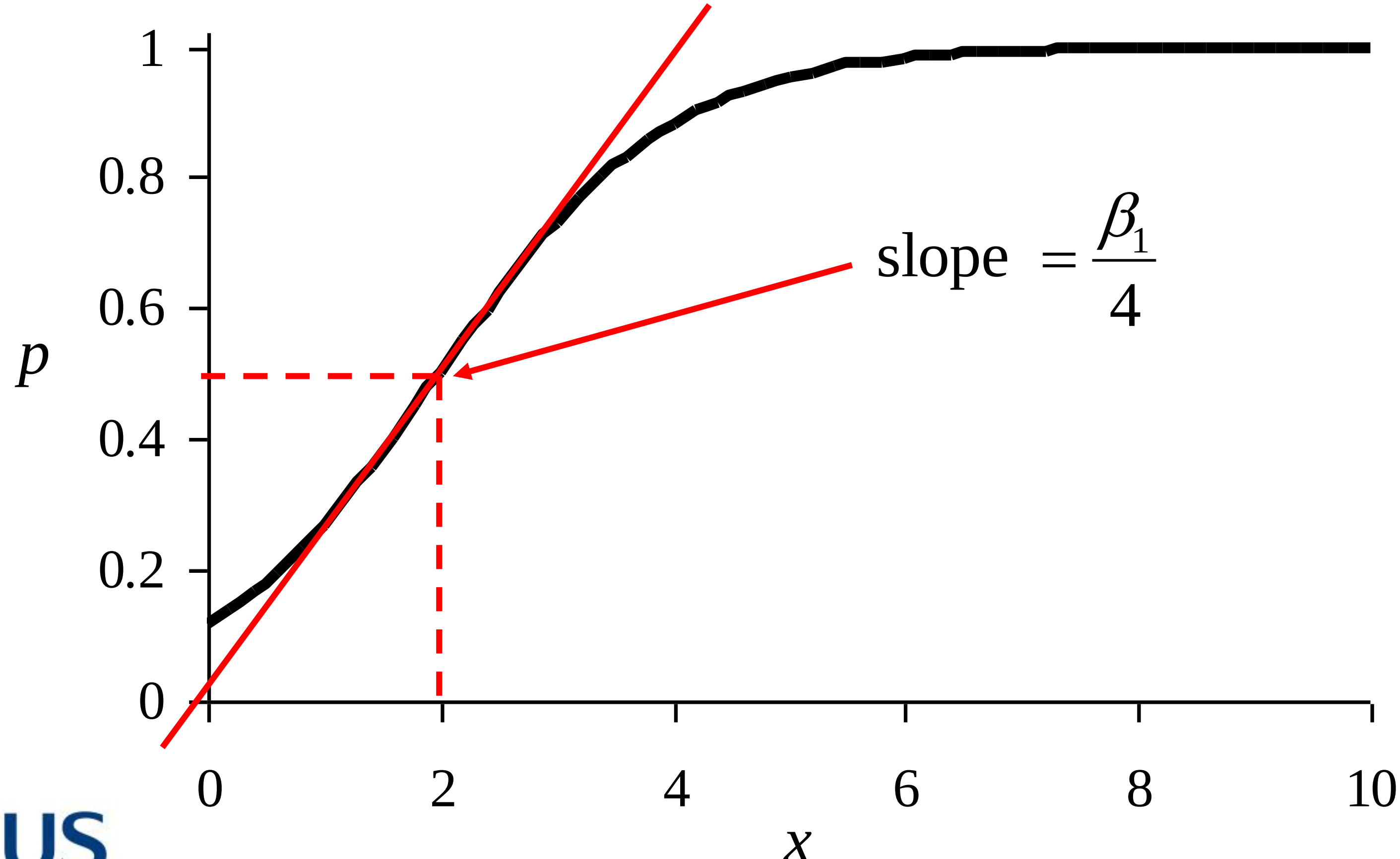
- Also:

$$\frac{dp}{dx} = \frac{d}{dx}\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$
$$= \frac{e^{\beta_0+\beta_1 x}\beta_1\left(1+e^{\beta_0+\beta_1 x}\right) - e^{\beta_0+\beta_1 x}\beta_1 e^{\beta_0+\beta_1 x}}{(1+e^{\beta_0+\beta_1 x})^2}$$
$$= \frac{e^{\beta_0+\beta_1 x}\beta_1}{(1+e^{\beta_0+\beta_1 x})^2}$$
$$= \frac{\beta_1}{4} \qquad \text{when } x = -\frac{\beta_0}{\beta_1}$$

- $\frac{\beta_1}{4}$: rate of increase in p with respect to x when $p = 0.5$

Interpretation: β_1

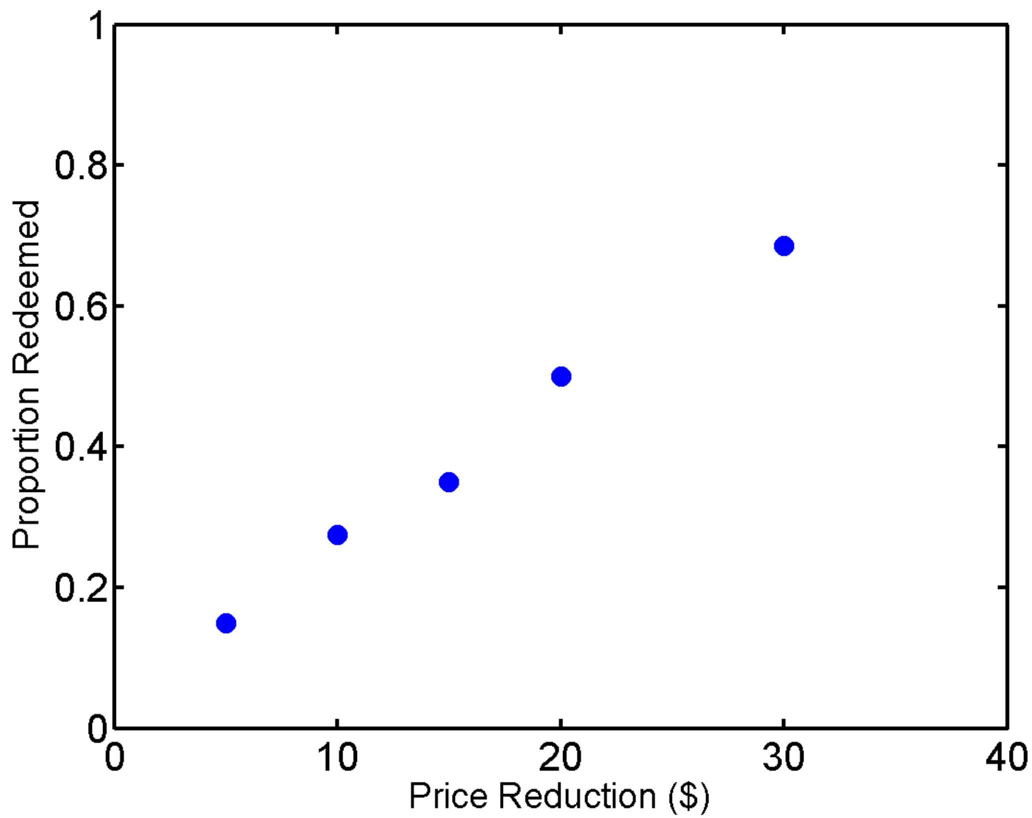
- β_1: determines slope when $p = 0.5$



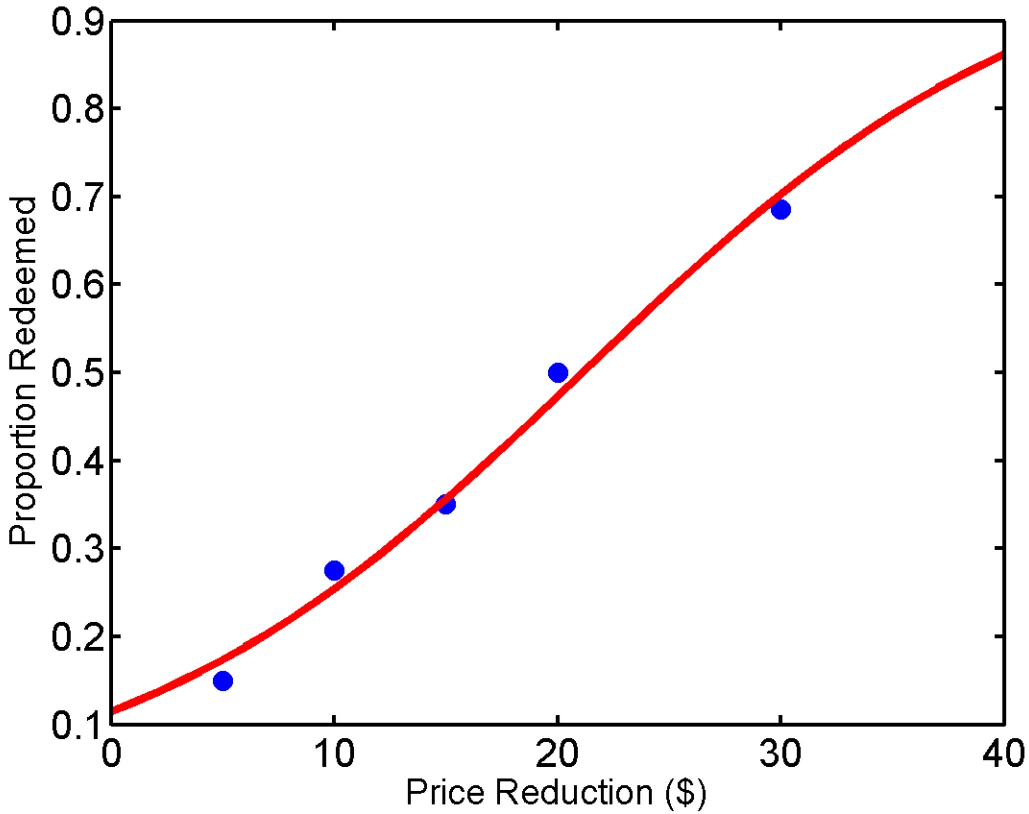
Example 1

- In a study of the effectiveness of coupons offering a price reduction, 1,000 homes were selected and coupons mailed
- Coupon price reductions: 5, 10, 15, 20, and 30 dollars
- 200 homes assigned at random to each coupon value
- x: amount of price reduction
- y: binary variable indicating whether or not coupon was redeemed

Example 1


Proportion Redeemed
Price Reduction ($)
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40

Example 1



- Fitted response function: $\beta_0 = -2.04$ and $\beta_1 = 0.097$

Example 2

- For each value of x, we may not have probability but rather a number of $\{x, y\}$ pairs
- Step 1: Extract frequencies from the pairwise data and hence probabilities
  - Raw data: $\{12, 0\}$, $\{12, 1\}$, $\{14, 0\}$, $\{12, 1\}$, $\{14, 1\}$, $\{14, 1\}$, $\{12,\ 0\}$, $\{12, 0\}$
  - Probability data: ($p = 1$ for each x, third entry is number of occurrences in raw data): $\{12, 0.4, 5\}$, $\{14, 0.67, 2\}$
- Step 2: Obtain the odds ratio data
- Step 3: Obtain β_0 and β_1 using linear regression model

Example 2

Number of Trips	Require Maintenance?		Total
	No	Yes	
1	9	1	10
2	13	2	15
3	9	3	12
4	10	5	15
5	7	6	13
6	3	5	8
7	4	13	17
8	2	8	10
Total	57	43	100

Example 2

Number of Trips	P(maintenance)	odds	Log odds	Number of occurrences
1	0.1000	0.1111	-2.1972	10
2	0.1333	0.1538	-1.8718	15
3	0.2500	0.3333	-1.0986	12
4	0.3333	0.5000	-0.6931	15
5	0.4615	0.8571	-0.1542	13
6	0.6250	1.6667	0.5108	8
7	0.7647	3.2500	1.1787	17
8	0.8000	4.000	1.3863	10

Example 2

x (trips)	y (log odds)	x^2	xy	Number of occurrences
1	-2.1972	1	-2.1972	10
2	-1.8718	4	-3.7436	15
3	-1.0986	9	-3.2958	12
4	-0.6931	16	-2.7726	15
5	-0.1542	25	-0.7708	13
6	0.5108	36	3.0650	8
7	1.1787	49	8.2506	17
8	1.3863	64	11.0904	10
448	**-37.6471**	**2504**	**106.3981**	**100**

Bottom row: weighted sum $(\sum_{i=1}^{8} value_i \times occurances_i)$

Example 2

- Results from regression
  - $\beta_0 = -2.856$ and $\beta_1 = 0.5535$

Number of Trips	P(maintenance)	$\hat{p}$
1	0.1000	0.0909
2	0.1333	0.1482
3	0.2500	0.2323
4	0.3333	0.3448
5	0.4615	0.4788
6	0.6250	0.6142
7	0.7647	0.7346
8	0.8000	0.8280

SSE 0.0028

SST 0.5265

R^2 0.9946

Logistic vs Linear

$$y = a + bx$$

‘y’ represents the dependent variable (what we are trying to predict)

‘a’ represents the intercept (where the regression line crosses the vertical ‘y’ axis) aka the constant

‘b’ represents the slope of the line (the association between ‘y’ & ‘x’)

‘x’ represents the independent variable (what we are using to predict ‘y’)

Probability

$$P(y) = 1/(1 + e^{-(a + bx)})$$

Logarithmic Transformation

National University of Singapore

Poisson Regression

- Y is a non-negative integer: $0, 1, 2, \cdots$ (usually small values)
- Model:

$$\ln(E[Y]) = \beta_0 + \beta_1 x$$

- This implies:

$$E[Y] = e^{\beta_0 + \beta_1 x}$$

Acknowledgements

- A number of the slides in this lecture are based on material from various sources:
  - Luke Sloan
  - Luiz Pessoa