Non-Linear Regression

Dr. Supaporn Erjongmanee

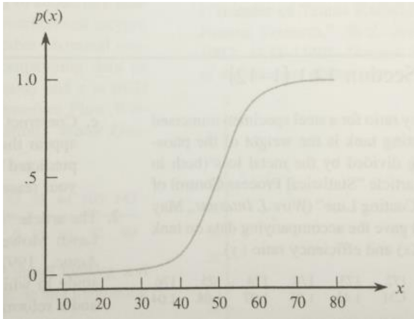
Department of Computer Engineering
Kasetsart University
fengspe@ku.ac.th

Outline

- Logistic regression
- Non-linear regression

Logistic Probabilistic Model

- When y is either 0 or 1, linear model is not appropriate since value is between 0 and 1
  - We will apply another model called “logistic” probabilistic model
- For example:
  - Probability that a car needs maintenance service depends on the car’s mileage



- P(y = 1) or P(y=0) depends on value of x

Image source: Figure 12.7 [1]

3

Logistic Probabilistic Model (cont.)

- P(y = 1) or P(y=0) depends on value of x
- $\mu_{Y\,x}$ = E(Y|x) = 1*p(x) + 0*[1-p(x)] = p(x)

Mean of probability = p(x) = [0, 1]

- Here, we apply *logit function*

$$\boxed{p(x) = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}}$$

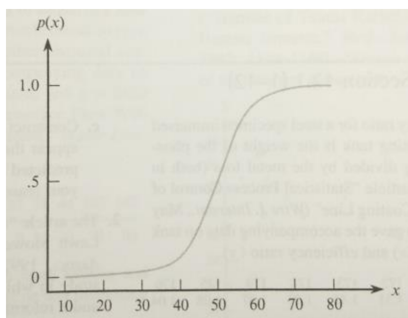


Image source: Figure 12.7 [1]

Unit of x is thousand miles

4

Logistic Probabilistic Model (cont.)

- From the logit function,

$$p(x) = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

$$p(x) = \frac{1}{e^{-(b_0 + b_1 x)} + 1}$$

$$\frac{1}{p(x)} = e^{-(b_0 + b_1 x)} + 1$$

$$\frac{1}{p(x)} - 1 = e^{-(b_0 + b_1 x)}$$

$$\frac{1 - p(x)}{p(x)} = e^{-(b_0 + b_1 x)}$$

$$\boxed{\frac{p(x)}{1 - p(x)}} = e^{(b_0 + b_1 x)}$$

Odds ratio

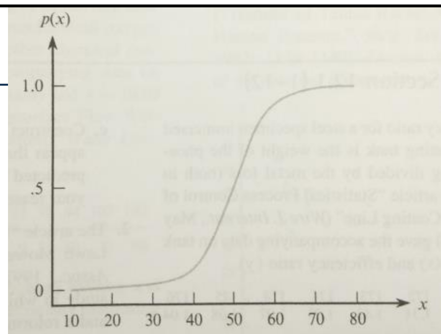
$$\ln\left(\boxed{\frac{p(x)}{1 - p(x)}}\right) = b_0 + b_1 x$$

5

Odds Ratio

Odds ratio

$$\boxed{\frac{p(x)}{1 - p(x)}} = e^{(b_0 + b_1 x)}$$



Unit of x is thousand miles

- Odds ratio = ratio between prob(success) and prob(failure)

If $p(60) = 3/4$, what is odds ratio?

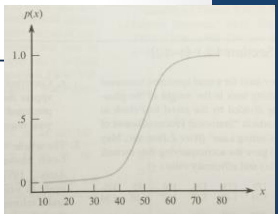
$$\frac{p(x)}{1 - p(x)} = \frac{3/4}{1/4} = 3$$

At 60 thousands miles,
prob(a car needs service) is 3 times more than prob(a car does not need service)

Image source: Figure 12.7 [1]

6

Logistic Regression Example



- $p(x) = \frac{e^{b_0 + b_1 x}}{1+e^{b_0 + b_1 x}}$ P(x) = probability that an event occurs
- <u>Example</u>:
- There are 5 sample results $\{x_1, x_2, x_3, x_4, x_5\}$ where x_2, x_4, x_5 are successes.

Likelihood function = $(1-p(x_1))\ (p(x_2))\ (1-p(x_3))\ (p(x_4))\ (p(x_5))$

- It is very complicated to find b_0 and b_1 that maximize the likelihood function
- Thus, it is recommended to use function available in software tool

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

7

Logistic Regression Example (cont.)

- On 28 January 1986, Challenger shuttle broke apart 73 seconds after its launch

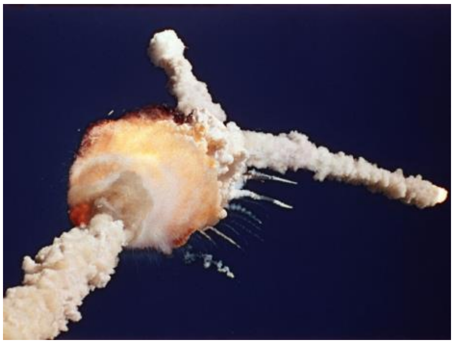


Image sources:
http://www.telegraph.co.uk/news/science/space/12124466/30th-Anniversary-of-the-Space-Shuttle-Challenger-disaster-in-pictures.html
http://www.cbsnews.com/pictures/challenger-shuttle-disaster/

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

8

Logistic Regression Example (cont.)

- Cape Canaveral, Florida
- Average temperature in January = 49 °F (9 °C)
- The temperature on the that day = 31 °F (-0.56 °C)
  - "Unusually cold"
- Part of shuttle is called O-Ring
- O-Ring is used to seal and protect high-pressured gas.

Location in the United States

Source: https://en.wikipedia.org/wiki/Cape_Canaveral,_Florida

Logistic Regression Example (cont.)

- Data on prior launch temperature (°F) and O-ring failures

Temperature	Failure	Temperature	Failure	Temperature	Failure
53	Y	68	N	75	N
57	Y	69	N	75	Y
58	Y	70	N	76	N
63	Y	70	N	76	N
66	N	70	Y	78	N
67	N	70	Y	79	N
67	N	72	N	81	N
67	N	73	N		

Lower bound for outlier: Q1 – 1.5IQR = 55

From temperature on the Challenger's launch (31 °F) , would O-ring fail or not?

31 = Q1 – 4.5IQR

Answer: Since 31 °F is far outside the above data set, we should not predict the outcome from the above data set.

Logistic Regression Example (cont.)

Let's find out whether temperature affects on O-Ring

- Data on prior launch temperature (°F) and O-ring failures

Temperature	Failure	Temperature	Failure	Temperature	Failure
53	Y	68	N	75	N
57	Y	69	N	75	Y
58	Y	70	N	76	N
63	Y	70	N	76	N
66	N	70	Y	78	N
67	N	70	Y	79	N
67	N	72	N	81	N
67	N	73	N		

Apply tool and derive b0 = 15.0429 , b1 = 0.2322
R^2 = 0.5465

$$p(x) = \frac{e^{15.0429 - 0.2322\,x}}{1 + e^{15.0429 - 0.2322\,x}}$$

Logistic Regression Example (cont.)

Let's find out whether temperature affects on O-Ring

- Data on prior launch temperature (°F) and O-ring failures

Temperature	Failure	Temperature	Failure	Temperature	Failure
53	Y	68	N	75	N
57	Y	69	N	75	Y
58	Y	70	N	76	N
63	Y	70	N	76	N
66	N	70	Y	78	N
67	N	70	Y	79	N
67	N	72	N	81	N
67	N	73	N		

```
model <- glm(Failure1 ~ Temperature, data = df, family = "binomial")
summary(model)
```

Logistic Regression Example (cont.)

- Data on prior launch temperature (°F) and O-ring failures

```
> summary(model)

Call:
glm(formula = Failure1 ~ Temp, family = "binomial", data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0611  -0.7613  -0.3783   0.4524   2.2175

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  15.0429     7.3786   2.039   0.0415 *
Temp         -0.2322     0.1082  -2.145   0.0320 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28.267  on 22  degrees of freedom
Residual deviance: 20.315  on 21  degrees of freedom
AIC: 24.315

Number of Fisher Scoring iterations: 5

>
```

$$\boxed{p(x) = \frac{e^{15.0429 - 0.2322\,x}}{1 + e^{15.0429 - 0.2322\,x}}}$$

Outline

- Logistic regression
- Non-linear regression

General Concept

- Let Y represent by nonlinear function of x
  - $Y = f(X, \mathbf{b}) + \varepsilon$

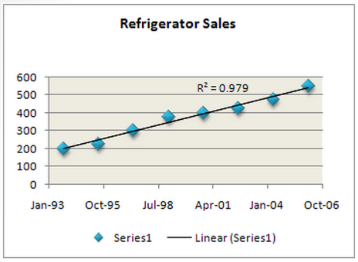
  where
    - $\mathbf{b}$ = set of parameters = $\{b_0, b_1, .., b_p\}$
    - ε is normally distributed with $\mu = 0$, var = σ^2
  - Sample collection: $y_1, y_2, \ldots, y_n$
  - Error = $Q = \sum_{i=1}^{n}(y_i - f(x, \mathbf{b}))^2$
  - Objective: to find $\hat{b}_j = \arg\min_b Q$

General Concept (cont.)

- Objective: to find $\hat{b}_i = \arg\min_b Q$
  - $\frac{\partial Q}{\partial b_j} = -2\sum_{i=1}^{n}(y_i - f(x_i, \mathbf{b}))\left[\frac{\partial f(x_i, \mathbf{b})}{\partial b_j}\right] = 0$

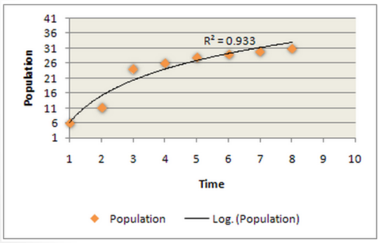
  Solve for each b_j
  - Example of nonlinear f function
    - Exponential, Logarithm
    - Power
    - Polynomials
    - Others

Data Trend



Linear trend $f(x) = b\,x$

- Data looks like a straight line

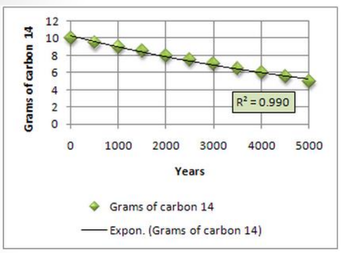


Logarithm trend $f(x) = \log(x)$

- Data changes quickly and reaches stability

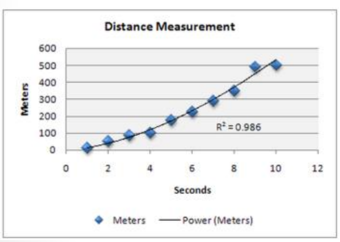
Source: https://support.office.com/en-sg/article/Add-change-or-remove-a-trendline-in-a-chart-072d130b-c60c-4458-9391-3c6e4b5c5812

Data Trend (cont.)



Exponential trend $f(x) = c^{x}$

- Data increases/decreases at <u>constantly increasing rate</u>

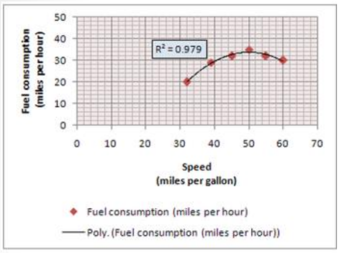


Power trend $f(x) = x^{c}$

- Data increases at specific rate
- Do not work with negative or zero data

Source: https://support.office.com/en-sg/article/Add-change-or-remove-a-trendline-in-a-chart-072d130b-c60c-4458-9391-3c6e4b5c5812

Data Trend (cont.)



Polynomial trend

$$f(x) = x^n + x^{n-1} + \cdots + x + 1$$

- Data fluctuates
- Degree (n) of polynomial
  - Depends on #bends + 1
    - Bend : peak or valley
  - As example, 1 bend, choose polynomial degree = 2

Moving average trend

May be difficult to fit with regression

- Select when the goal is to smooth out data that fluctuates a lot
- Use average computed from subset (period) of data
  - Allow to choose subset size

Source: https://support.office.com/en-sg/article/Add-change-or-remove-a-trendline-in-a-chart-072d130b-c60c-4458-9391-3c6e4b5c5812

19

Nonlinear as Linear Regression

- Some of nonlinear regression models can be trained as linear regression models
- Example
  - Exponential
  - Logarithm
  - Power

20

Exponential Function

- Let *f(X, **b**)* = $e^{b_0+b_1x} = e^{b_0+b_1x}$
- Thus, $\hat{y} = e^{b_0+b_1x}$

  $\ln(\hat{y}) = \boxed{b_0 + b_1x}$

  Linear

% of change in Y when X changes by 1 unit

- To solve for b_0 and b_1 in "linear regression", use
  - log values of $y_1, y_2, \ldots, y_n$
  - $x_1, x_2, \ldots, x_n$
- Example: growth or decay equation

21

Logarithm Function

- Let *f(X, **b**)* = $b_0 + b_1\ln(x)$
- Thus, $\hat{y} = \boxed{b_0 + b_1\ln(x)}$

  Linear

Amount of change in Y when X changes by 1%

- To solve for b_0 and b_1 in "linear regression", use
  - $y_1, y_2, \ldots, y_n$
  - log values of $x_1, x_2, \ldots, x_n$

22

Power Function (Elasticity)

- Let $f(x, b) = b_0 x^{b_1}$
- Thus, $\hat{y} = b_0 x^{b_1}$

% of change in Y when X changes by 1%

X and Y must be positive

$$\ln(\hat{y}) = \boxed{\ln(b_0) + b_1 \ln(x)}$$

Constant

Linear

- To solve for $\ln(b_0)$ and b_1 in "linear regression", use
  - log values of $y_1, y_2, \ldots, y_n$
  - log values of $x_1, x_2, \ldots, x_n$

$b_0 = e^{\text{Constant}}$

- Example: supply, demand, cost, production functions

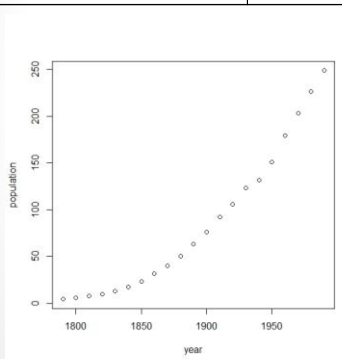
Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Example

- U.S. Population

Year	1790	1800	1810	…	1990
U.S. Population (Y)	3.929	5.308	7.240	…	248.710



Population increases at constantly increasing rate

Select exponential function to fit this data

- Let $f(x, b) = e^{b_0 + b_1 x}$

Source: https://onlinecourses.science.psu.edu/stat501/node/399

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

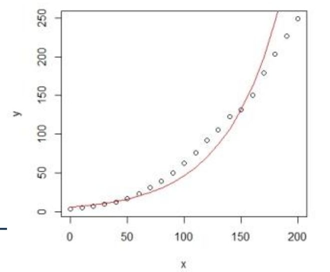
Example (cont.)

- Let $f(x, b) = e^{b_0 + b_1 x}$

Year	1790	1800	1810	...	1990
U.S. Population (Y)	3.929	5.308	7.240	...	248.710

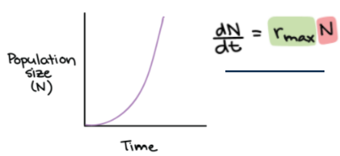
X (Year – 1790)	0	10	20	...	200
ln (Y)	1.368	1.669	1.980	...	5.516

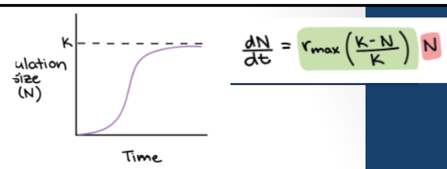
- Solved in linear regression, obtain
  - $b_0 = 1.766$
  - $b_1 = 0.021$
  - $y = e^{1.766 + 0.021x}$



Source: https://onlinecourses.science.psu.edu/stat501/node/399

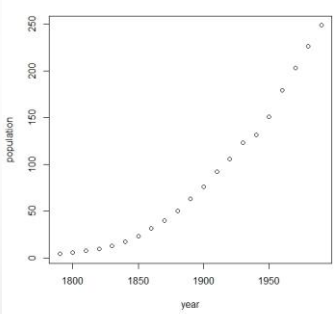
Example (cont.)





- U.S. Population

Year	1790	1800	1810	...	1990
U.S. Population (Y)	3.929	5.308	7.240	...	248.710



To find a model with decreasing rate and finally population size reaches an asymptote

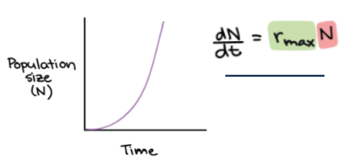
Select logistic model

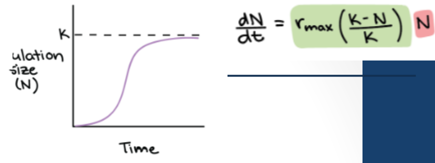
- Let $f(x, b) = \dfrac{b_0}{1 + e^{b_1 + b_2 x}}$

b_0 = asymptote value
b_1 = Population at x = 0
b_2 = growth rate

Source: https://onlinecourses.science.psu.edu/stat501/node/399

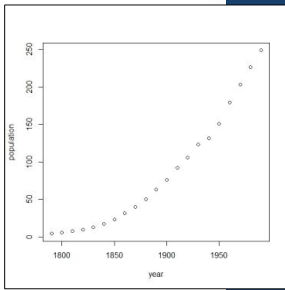
Example (cont.)





- U.S. Population

X (Year – 1790)	0	10	20	...	200
U.S. Population (Y)	3.929	5.308	7.240	...	248.710



Select logistic model

- Let $f(x,b) = \frac{b_0}{1+e^{b_1+b_2 x}}$

b_0 = asymptote value → $b_0 = 350$

b_1 = Population at x = 0 → $3.929 = \frac{350}{1+e^{b_1+b_2(0)}}$ → $b_1 = \log\left(\frac{350}{3.929} - 1\right)$ = 4.478

b_2 = growth rate → $5.308 = \frac{350}{1+e^{4.478+b_2(1)}}$ = -0.30

Source: https://onlinecourses.science.psu.edu/stat501/node/399

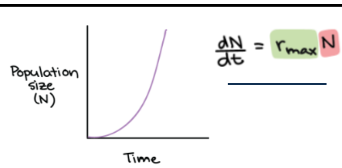
Supaporn Erjongmanee
fengspe@ku.ac.th

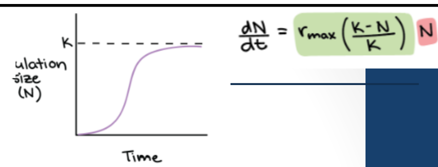
Statistics in Computer Engineering
Slide 27

Department of Computer Engineering
Kasetsart University

27

Example (cont.)





- U.S. Population

X (Year – 1790)	0	10	20	...	200
U.S. Population (Y)	3.929	5.308	7.240	...	248.710

Select logistic model

- Let $f(x,b) = \frac{b_0}{1+e^{b_1+b_2 x}}$

$b_0 = 350$

$b_1 = 4.478$

$b_2 = -0.30$

```
> model <- nls(y~b0/(1+exp(b1+b2*x)),
+              start=list(b0=b0_start,
+                         b1=b1_start,
+                         b2=b2_start))
> summary(model)

Formula: y ~ b0/(1 + exp(b1 + b2 * x))

Parameters:
    Estimate Std. Error t value Pr(>|t|)
b0 389.165811  30.812041   12.63  2.2e-10 ***
b1   3.990345   0.070321   56.74  < 2e-16 ***
b2  -0.022662   0.001086  -20.87  4.6e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.45 on 18 degrees of freedom

Number of iterations to convergence: 13
Achieved convergence tolerance: 2.656e-06
```

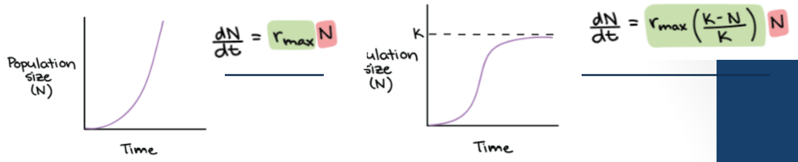
Source: https://onlinecourses.science.psu.edu/stat5

Supaporn Erjongmanee
fengspe@ku.ac.th

Statistics in Co

28

Example (cont.)



- U.S. Population

X (Year – 1790)	0	10	20
U.S. Population (Y)	3.929	5.308	7.240

```
> model <- nls(y~b0/(1+exp(b1+b2*x)),
+              start=list(b0=b0_start,
+                         b1=b1_start,
+                         b2=b2_start))
> summary(model)

Formula: y ~ b0/(1 + exp(b1 + b2 * x))

Parameters:
     Estimate Std. Error t value Pr(>|t|)
b0 389.165811  30.812041   12.63  2.2e-10 ***
b1   3.990345   0.070321   56.74  < 2e-16 ***
b2  -0.022662   0.001086  -20.87  4.6e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.45 on 18 degrees of freedom

Number of iterations to convergence: 13
Achieved convergence tolerance: 2.656e-06
```

- Let $f(x,b) = \frac{b_0}{1+e^{b_1+b_2x}}$

  $b_0 = 350$

  $b_1 = 4.478$

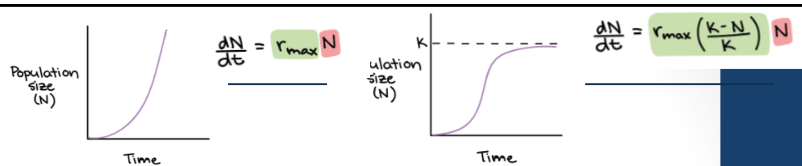
  $b_2 = -0.30$

$$y = \frac{389.166}{1+e^{3.99-0.02x}}$$

Source: https://onlinecourses.science.psu.edu/stat501/node/399

29

Example (cont.)



- U.S. Population

X (Year – 1790)	0	10	20	…	200
U.S. Population (Y)	3.929	5.308	7.240	…	248.710

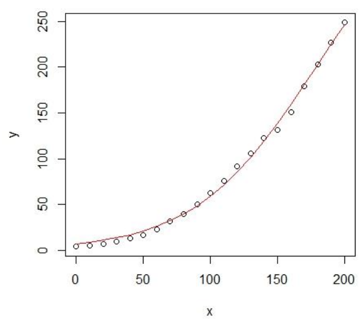
- Let $f(x,b) = \frac{b_0}{1+e^{b_1+b_2x}}$

  $b_0 = 350$

  $b_1 = 4.478$

  $b_2 = -0.30$

$$y = \frac{389.166}{1+e^{3.99-0.02x}}$$



Source: https://onlinecourses.science.psu.edu/stat501/node/399

30

References

1. J.L. Devore and K.N.Berk, Modern Mathematical Statistics with Applications, Springer, 2012.
2. S. Few, Now You See It: Simple Visualization Techniques for Quantitative Analysis, Analytics Press, 2009