Introduction to Regression Modeling – Part I

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 1 – Lecture 1B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

What is Regression Modeling ?

- Regression modeling is a statistical technique for exploring and investigating the relationship between variables.

An Example of Regression Modeling

An industrial engineer working for a soft drink beverage bottler suspected:

- the time (y) required by a route deliveryman to load and service a machine

is related to:

- the number (x) of cases of product delivered

Reference: page 1 of Textbook

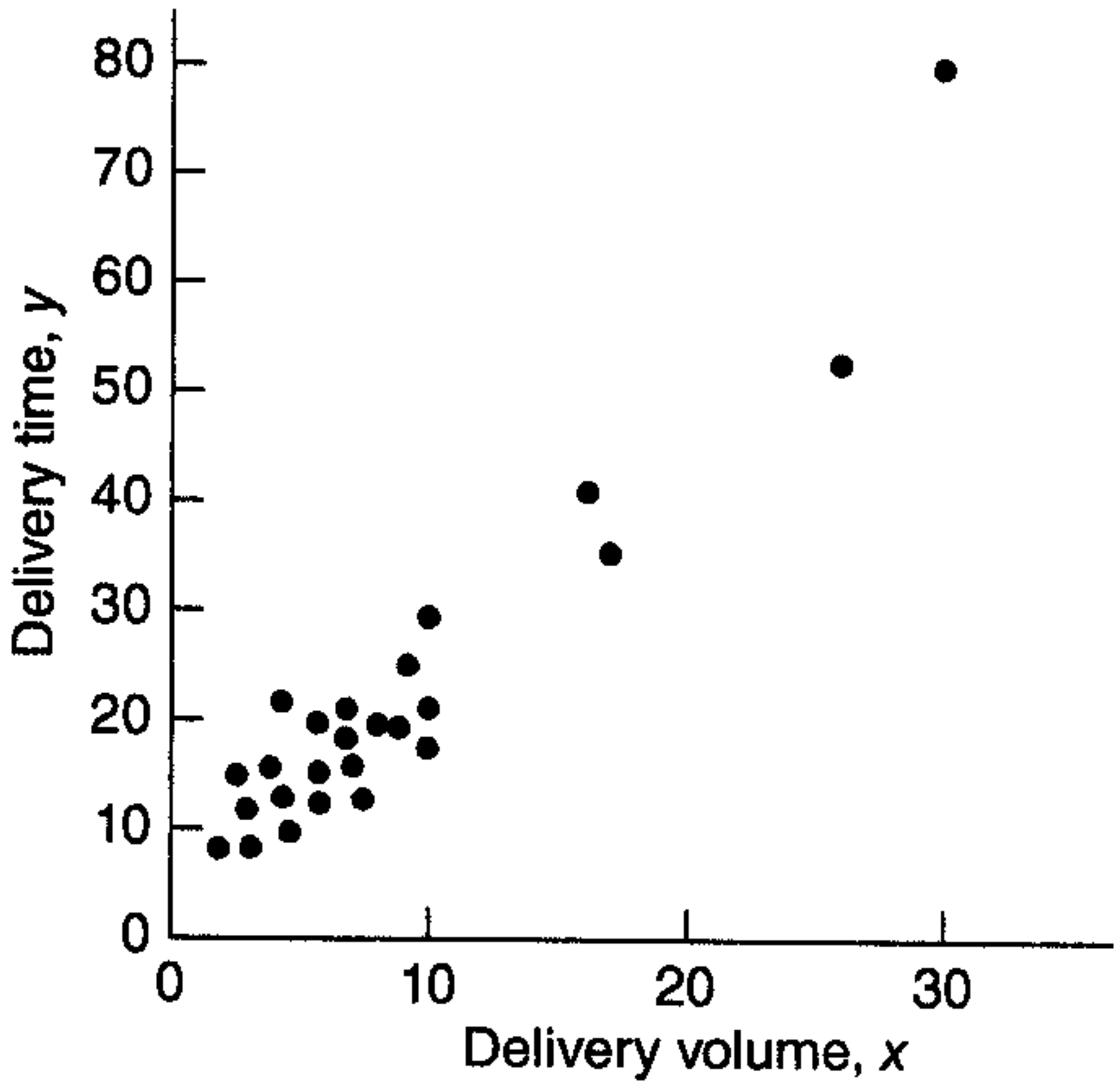
An Example of Regression Modeling

The engineer visited 25 *randomly selected* retail stores having vending machines, and collected data on the delivery time (in minutes) and the volume of product delivered (in cases) .

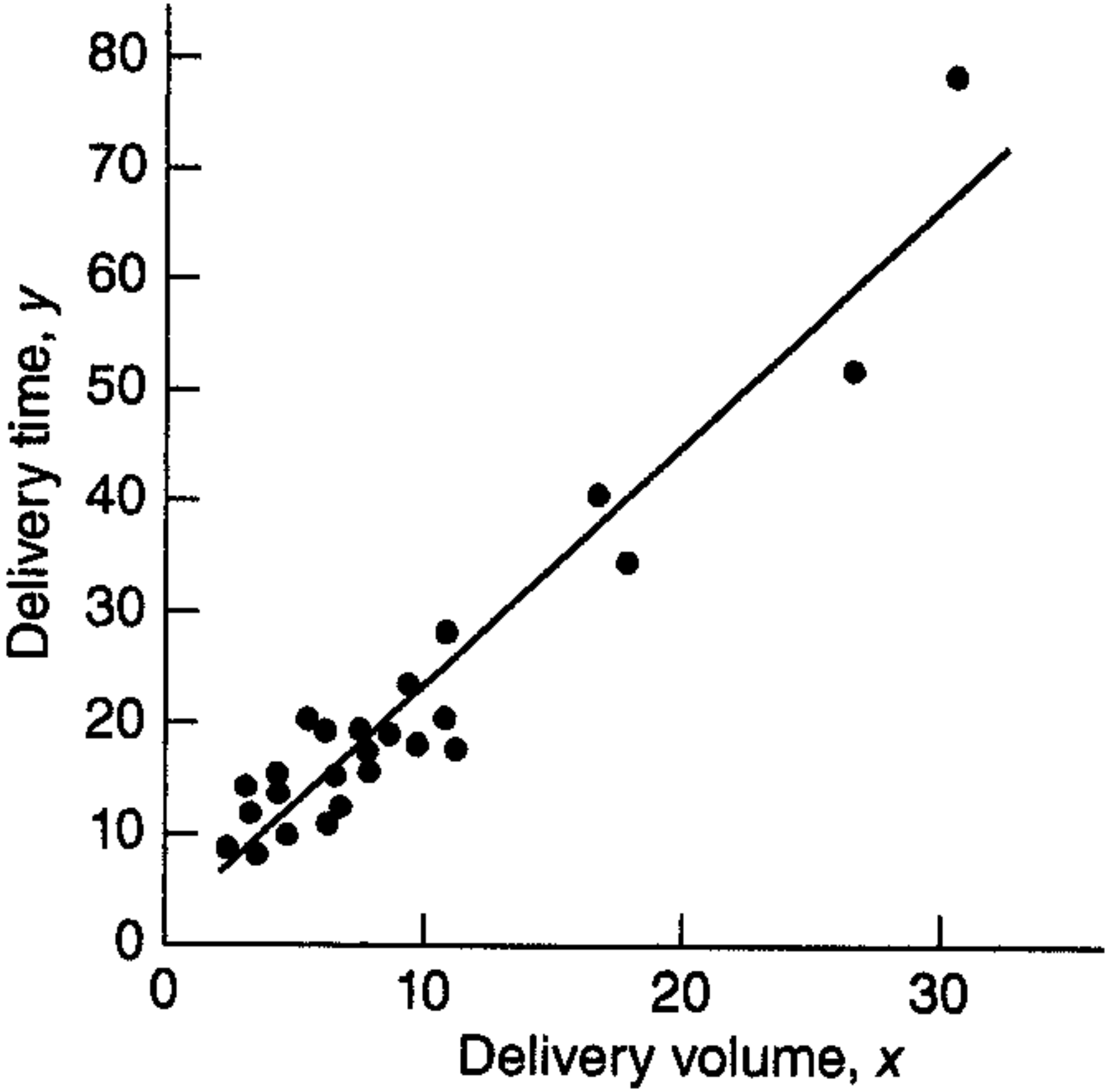
Plotted the 25 paired observations
- scatter diagram (slide 5, Fig 1.1a of Textbook)

$\Rightarrow$ Suggests a straight-line relationship
(slide 6, Fig 1.1b of Textbook)

The Example - Scatter Diagram



The Example - Scatter Diagram with Straight-Line Fit



Mathematical Model for Straight Line Relationship

y : deliver time
x : deliver volume

<u>Straight line mathematical model</u>

$$y = \beta_0 + \beta_1 x$$

β_0 : y-intercept
β_1 : slope

Linear Regression Model

Data does not fall exactly on the line.

error: $\varepsilon = y - (\beta_0 + \beta_1 x)$

ε : statistical error, random error

y : dependent or response variable

x : independent, predictor, regressor variable

Linear (simple linear) regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Expectation and Variance of Response Variable

Suppose that we can fix x and ε is the random component with mean 0 and variance σ^2
(σ is standard deviation quantifying variability or spread of y)

<u>Expectation of y at x</u> :

$$E(y \mid x) = \mu_{y|x} = E(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x$$

<u>Variance of y at x</u> :

$$\mathrm{Var}(y \mid x) = \mathrm{Var}(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2$$

True Linear Regression Model

In the case that ε is a noise, ε does not contribute anything to the expectation of y for each x

True linear regression model:

$$\mu_{y|x} = \beta_0 + \beta_1 x$$

Thus, for each x , y has mean $(\beta_0 + \beta_1 x)$ and variance σ^2

An Example of Statistical Distribution in Linear Regression

If this noise ε is normally distributed, then for each x ,

y is normally distributed with
mean $(\beta_0 + \beta_1 x)$ and variance σ^2

Notation: $y \mid x \sim N(\beta_0 + \beta_1 x , \sigma^2)$

Statistical Distribution of Y at X in Linear Regression

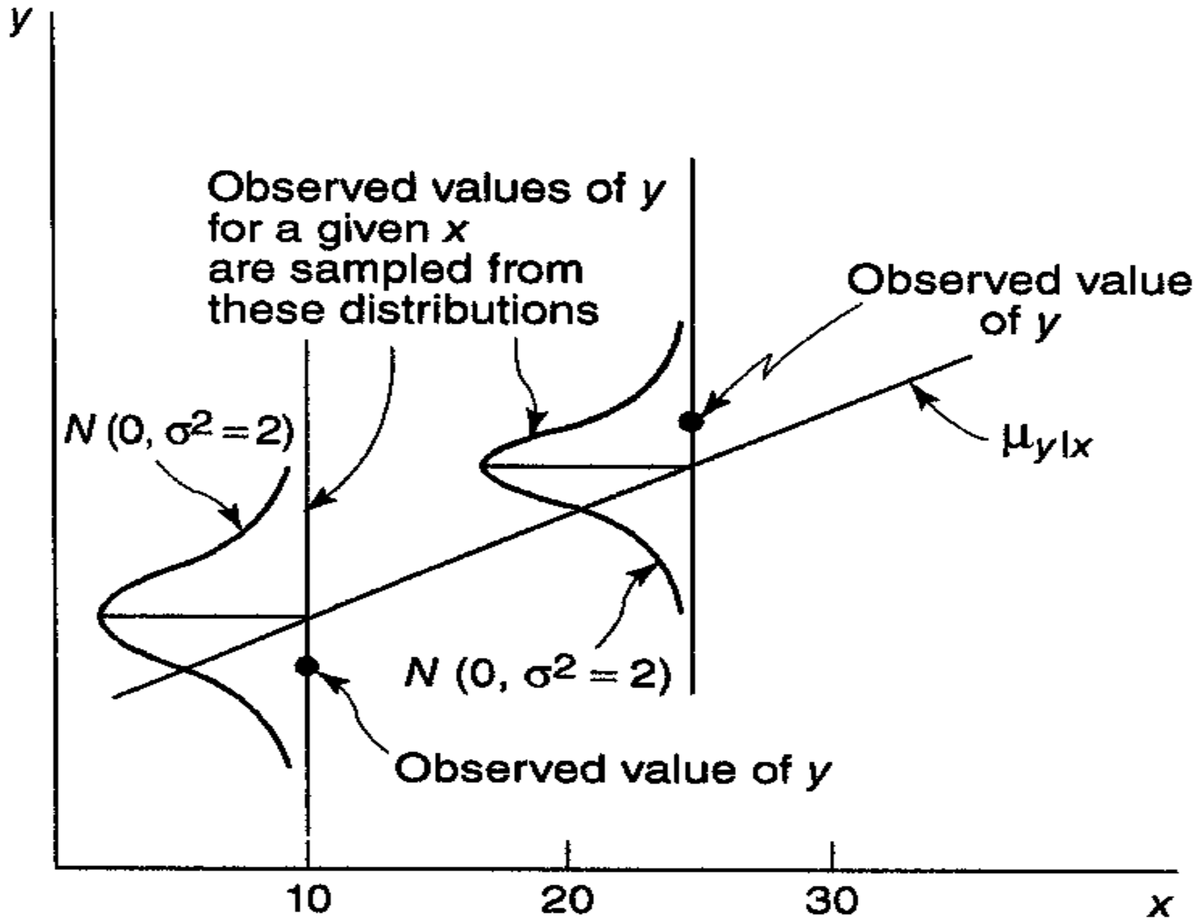


Figure 1.2 How observations are generated in linear regression. **Reference: page 3 of Textbook**

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING