REGRESSION

* BASICS OF REGRESSION
* TESTS FOR SIGNIFICANCE
  - ADEQUACY OF MODEL
* MULTI~~VARIATE~~ PLE REGRESSION
* CHOICE OF VARIBLES

$$Y = \beta_0 + \beta_1 x + \underset{\substack{\uparrow \\ \text{ERROR}}}{\varepsilon}$$

$$E(Y) = E(\beta_0 + \beta_1 x + \varepsilon)$$

$$E(Y) = \beta_0 + \beta_1 x \Rightarrow \text{REGRESSION EQUATION}$$

$$\boxed{\text{LINEAR REGRESSION EQUATION}}$$

Dep

X	Y

$$\boxed{y = ax + c}$$

$$\cancel{Y = ax_1 +}$$

$$\boxed{Y = \beta_0 + \beta_1 x_1}$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_3$$

6-1 Introduction To Empirical Models

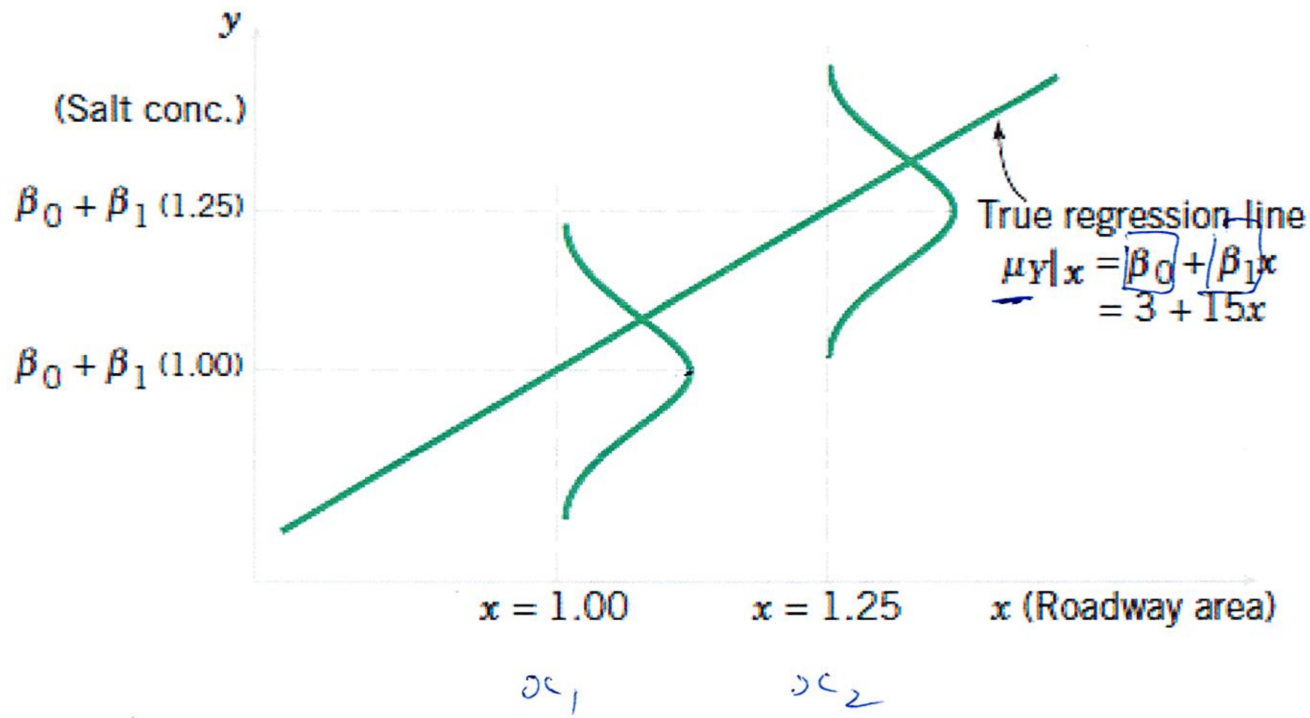
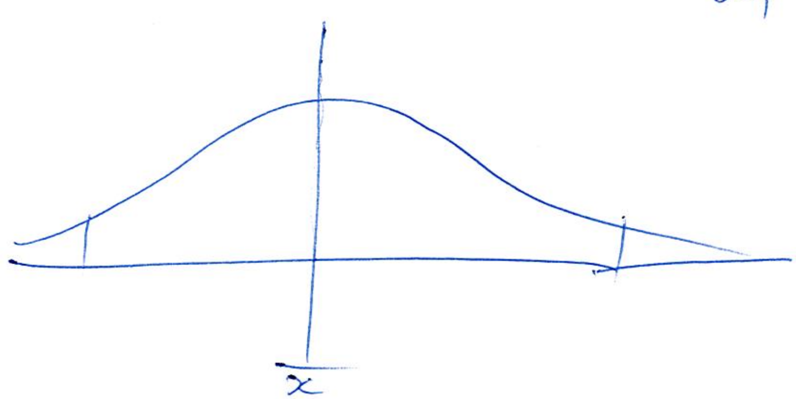


Figure 6-2 The distribution of Y for a given value of x for the salt concentration–roadway area data.



6-1 Introduction To Empirical Models

We think of the regression model as an empirical model.

Suppose that the mean and variance of ε are 0 and σ^2, respectively, then

$$E(Y|x) = E(\beta_0 + \beta_1 x + \epsilon) = \beta_0 + \beta_1 x + E(\epsilon) = \beta_0 + \beta_1 x$$

The variance of Y given x is

$$V(Y|x) = V(\beta_0 + \beta_1 x + \epsilon) = V(\beta_0 + \beta_1 x) + V(\epsilon) = 0 + \sigma^2 = \sigma^2$$

6-2 Simple Linear Regression

6-2.1 Least Squares Estimation

- Suppose that we have *n* pairs of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

- The **method of least squares** is used to estimate the parameters, β0 and β1 by minimizing the sum of the squares of the vertical deviations in Figure 6-6.

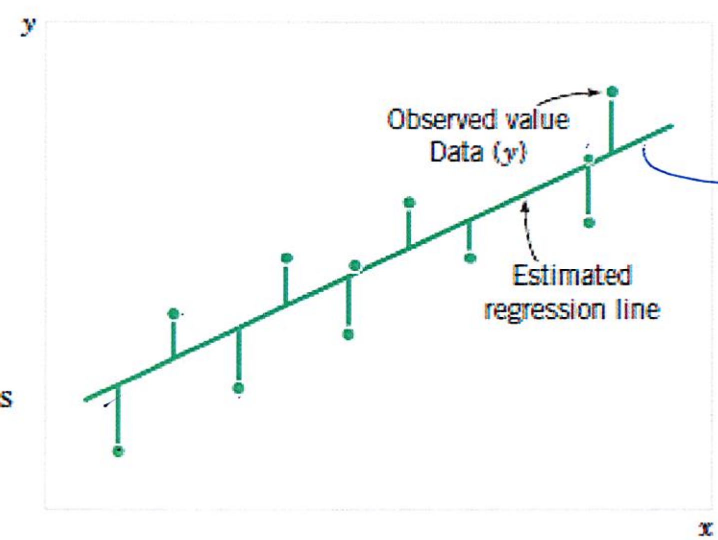
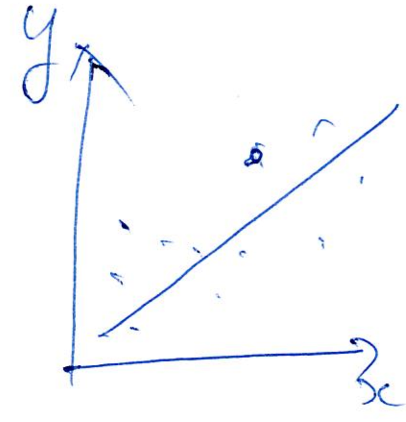


Figure 6-6 Deviations of the data from the estimated regression model.

$\rightarrow E(Y) = \beta_0 + \beta_1 x$

6-2 Simple Linear Regression

6-2.1 Least Squares Estimation



- Using Equation 6-8, the n observations in the sample can be expressed as

$$y_i = \overbrace{\beta_0 + \beta_1 x_i}^{\hat{y}_i} + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

- The sum of the squares of the deviations of the observations from the true regression line is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \overbrace{\beta_0 - \beta_1 x_i}^{\hat{y}_i})^2$$

$$\epsilon_i^2 = (y_i - \hat{y}_i)^2$$

6-2 Simple Linear Regression

6-2.1 Least Squares Estimation

The **fitted** or **estimated regression line** is therefore

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (6\text{-}15)$$

Note that each pair of observations satisfies the relationship

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \boxed{e_i,} \qquad i = 1, 2, \ldots, n$$

where $e_i = y_i - \hat{y}_i$ is called the **residual.** The residual describes the error in the fit of the model to the ith observation y_i. Subsequently we will use the residuals to provide information about the **adequacy** of the fitted model.

6-2 Simple Linear Regression

6-2.1 Least Squares Estimation

EXAMPLE 6-1

Therefore, the least squares estimates of the slope and intercept are

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{64.4082}{3.67068} = 17.5467$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 17.135 - (17.5467)0.824 = 2.6765$$

The fitted simple linear regression model is

$$\boxed{\hat{y} = \underset{\hat{\beta}_0}{\underset{\uparrow}{2.6765}} + \underset{\hat{\beta}_1}{\underset{\uparrow}{17.5467}}x}$$

$\Rightarrow$ Josephi@purdue.edu

6-2 Simple Linear Regression

6-2.1 Least Squares Estimation

Table 6-2 Minitab Regression Analysis Output for Salt Concentration and Roadway Data

Regression Analysis: Salt conc (y) versus Roadway area (x)

The regression equation is

Salt conc (y) = 2.68 + 17.5 Roadway area (x)

t-test H_0 H_1

Predictor	Coef	SE Coef	T	P
Constant	2.6765 $\leftarrow \hat{\beta}_0$	0.8680	3.08	0.006
Roadway area	17.5467 $\leftarrow \hat{\beta}_1$	0.9346	18.77	0.000

S = 1.791 $\leftarrow \hat{\sigma}$ R-Sq = 95.1% R-Sq(adj) = 94.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1130.1 $\leftarrow SS_R$	1130.1	352.46	0.000
Residual Error	$(n-2)$ 18	57.7 $\leftarrow SS_E$	3.2 $\leftarrow \hat{\sigma}^2$		
Total	$(n-2)+1 = 19$	1187.9 $\leftarrow SS_T$			

2

$(n-3)$

$(n-3)+2$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$n = 20$

n.

20

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

eq 1.

eq 2