

## 제 3 강

# 단순 선형회귀 모형

(The Simple Linear Regression Model)

### 회귀분석의 목적

1.  $y = f(x)$ 와 같이 이론에서 주어지는 경제적 변수들 간의 관계를 추정(estimate) 및 검정(test).
2. 설명변수의 값을 토대로 종속변수의 값을 예측(forecast)

## 주당 식료품 지출

$y$  = 매주 식료품에 지출된 액수(\$).

$x$  = 소비자의 주당 소득.

$x$ 가 주어졌을 때  $y$ 의 기대값과  $x$ 와의 관계는 선형(linear)일 수 있음

$$E(y|x) = \beta_1 + \beta_2 x$$

## 주당 식료품 지출

$f(y|x=480)$

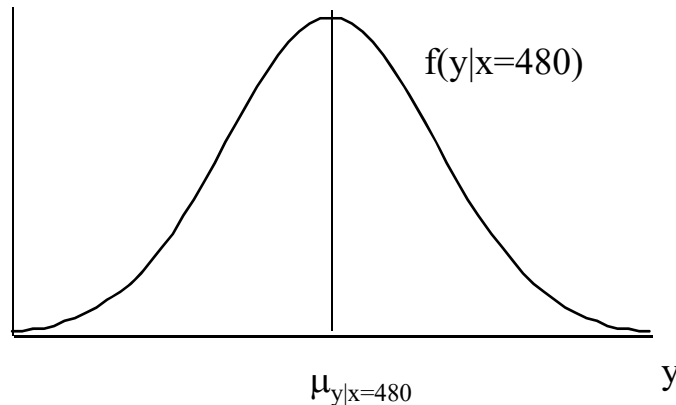


Figure 3.1a Probability Distribution  $f(y|x=480)$  of Food Expenditures if given income  $x=\$480$ .

주당 식료품 지출

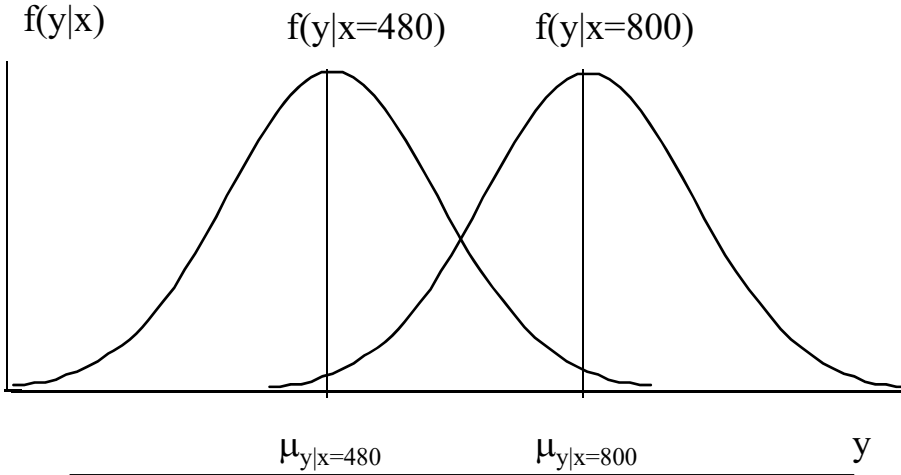


Figure 3.1b Probability Distribution of Food Expenditures if given income  $x=\$480$  and  $x=\$800$ .

주당 식료품 지출

평균지출(average expenditure)

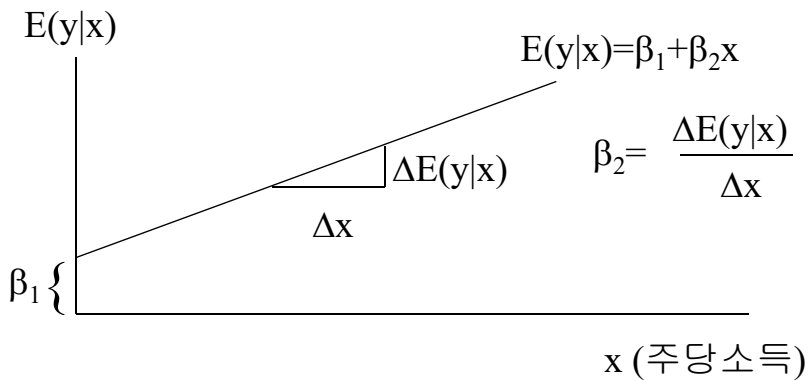


Figure 3.2 The Economic Model: a linear relationship between average expenditure on food and income.

주당 식료품 지출

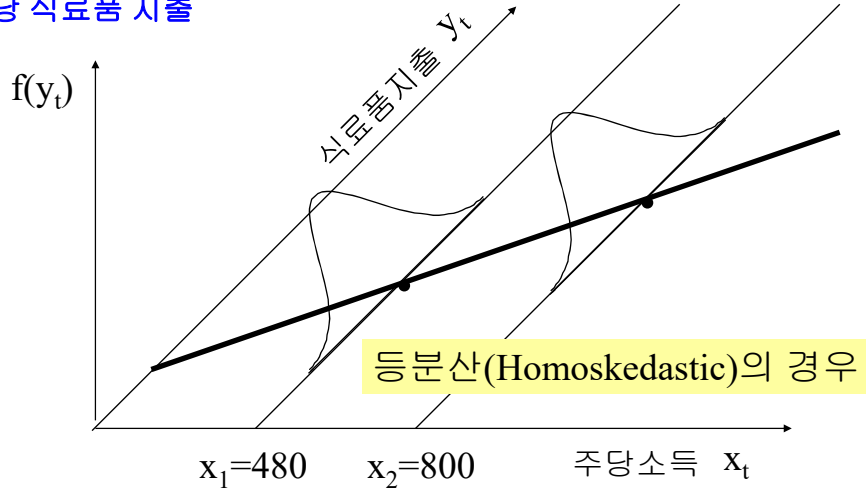


Figure 3.3. The probability density function for  $y_t$  at two levels of household income,  $x_t$

주당 식료품 지출

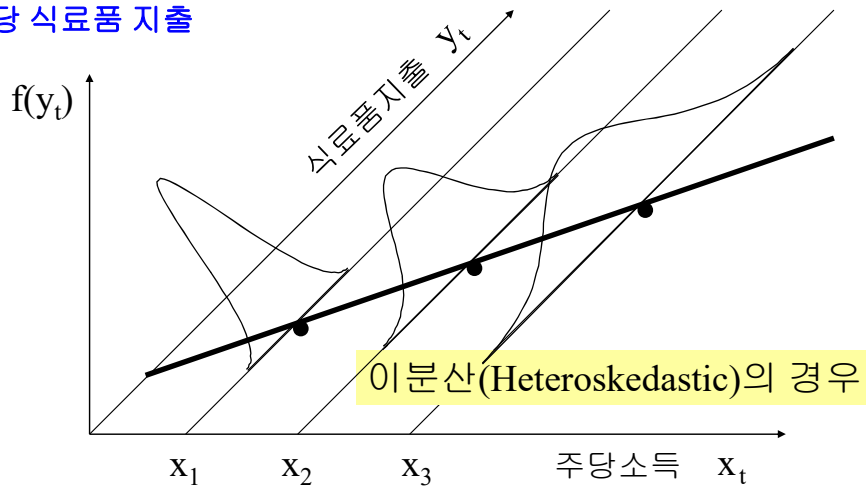


Figure 3.3+. The variance of  $y_t$  increases as household income,  $x_t$ , increases.

1. 주어진  $x$ 에 대해  $y$ 의 평균값은 선형(회귀)식으로 주어진다.

$$E(y) = \beta_1 + \beta_2 x$$

2. 각  $x$ 의 값에 대해  $y$ 의 값은 그 평균값 주변에서 다음의 분산을 갖는 분포를 한다.

$$\text{var}(y) = \sigma^2$$

3.  $y$ 의 값들은 상관되어 있지 않으며, 즉 0의 공분산을 가지며, 따라서 선형관계를 갖지 않는다.

$$\text{cov}(y_i, y_j) = 0$$

4. 변수  $x$ 는 적어도 두개의 다른 값을 가지며 확률변수가 아니다.

실제 종종 사용되기는 하지만 추정(estimation)에 있어서 반드시 필요하지는 않은 추가적 가정

5. (optional)  $y$ 의 값들은 각  $x$ 값에 대해 그들의 평균 주변에서 정규분포를 함

$$y \sim N [(\beta_1 + \beta_2 x), \sigma^2]$$

## 오차항(The Error Term)

계량경제학  
3.11

### 오차항의 도입

$y$ 는 두 부분으로 구성되는 확률변수이다.

I. 체계적 부분:  $E(y) = \beta_1 + \beta_2 x$   
 $y$ 의 평균임.

II. 확률적 부분:  $\varepsilon = y - E(y)$   
 $= y - \beta_1 - \beta_2 x$   
이 부분을 오차항이라고 부름.

$E(y)$ 와  $\varepsilon$  이 함께 모형을 형성함

$$y = \beta_1 + \beta_2 x + \varepsilon$$

## 오차항(The Error Term)

계량경제학  
3.12

### 오차항의 원천

1. 모형에서 구체화 되지 않은 요소들/설명변수들이 오차항에 있을 수 있다.
2.  $y$ 와  $x$ 의 관계가 정확히 선형관계가 아닌 경우 근사(approximation)오차가 오차항에 있게 된다.
3. 관측치에 특정적인 순수하게 예측불가능한 확률적 행태가 오차항에 있게 된다.

오차항의 분포

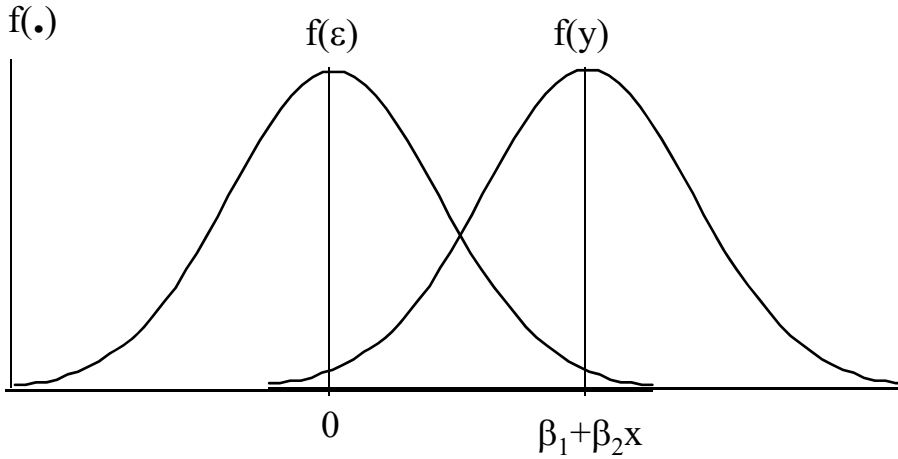


Figure 3.4 Probability density function for  $\varepsilon$  and  $y$

오차항

1.  $x$ 의 각 값에 대해  $y$ 의 값은:  

$$y = \beta_1 + \beta_2 x + \varepsilon$$
2. (주어진  $x$ 에 대해) 오차항  $\varepsilon$ 의 평균적 값은 :  

$$E(\varepsilon) = 0$$
3. (주어진  $x$ 에 대해) 오차항  $\varepsilon$ 의 분산은 :  

$$\text{var}(\varepsilon) = \sigma^2 = \text{var}(y)$$
4. 임의의 두 오차항 쌍(pair)간의 공분산은:  

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \text{cov}(y_i, y_j) = 0$$
5.  $x$ 는 반드시 두 개의 다른 값을 가지며 확률변수가 아니다.
6.  $\varepsilon$ 는 평균은 0이고 분산은  $\text{var}(\varepsilon) = \sigma^2$ 인 정규분포를 한다. (optional) 
$$\varepsilon \sim N(0, \sigma^2)$$

모집단(Population) 회귀값:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

모집단 회귀선(regression line):

$$E(y_t | x_t) = \beta_1 + \beta_2 x_t$$

표본(Sample) 회귀값:

$$y_t = b_1 + b_2 x_t + \hat{\varepsilon}_t$$

표본(Sample) 회귀선:

$$\hat{y}_t = b_1 + b_2 x_t$$

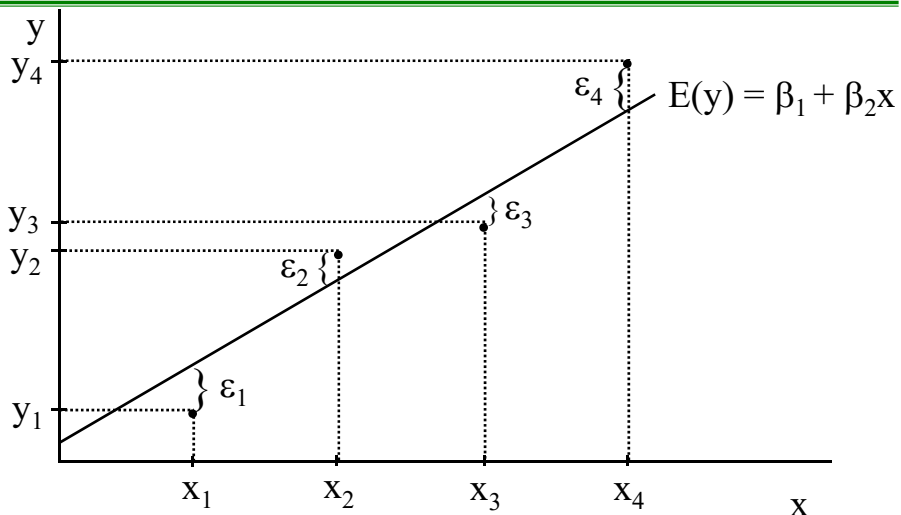


Figure 3.5 The relationship among  $y$ ,  $\varepsilon$  and the true regression line.



모수의 추정 (Estimating the parameters)

계량경제학  
3.17

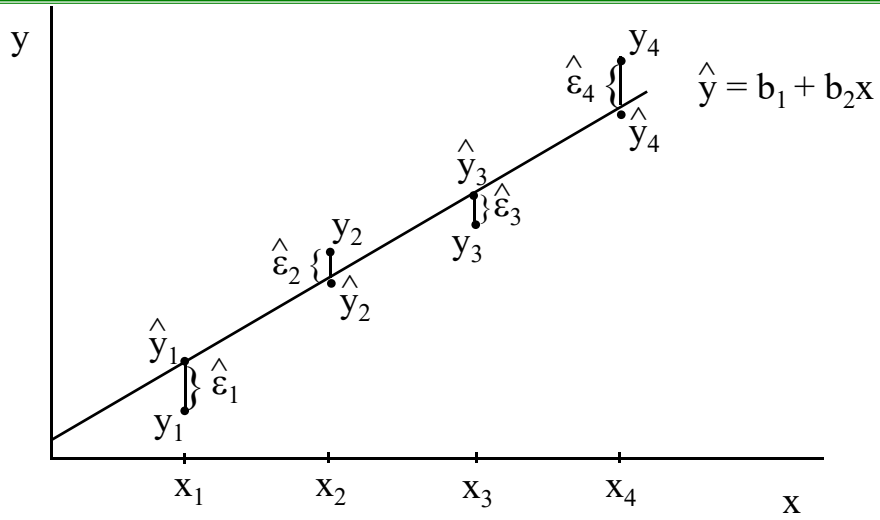
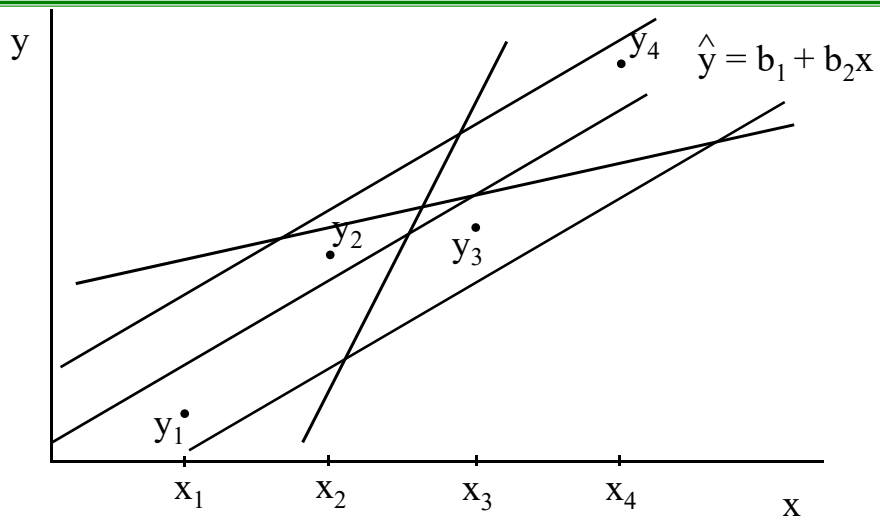


Figure 3.7a The relationship among  $y$ ,  $\hat{e}$  and the fitted regression line.

모수의 추정 (Estimating the parameters)

계량경제학  
3.18



How to find  $b_1$  and  $b_2$

최소제곱의 원칙(The Least Squares Principle)

계량경제학  
3.19

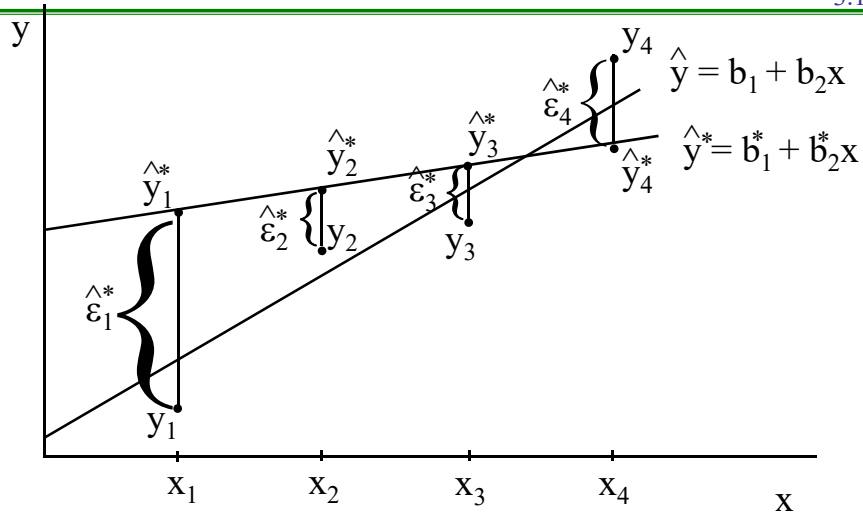


Figure 3.7b The sum of squared residuals from any other line will be larger.

최소제곱의 원칙(The Least Squares Principle)

계량경제학  
3.20

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$$\varepsilon_t = y_t - \beta_1 - \beta_2 x_t$$

편차의 제곱의 합을 최소화 함

$$S(\beta_1, \beta_2) = \sum_{t=1}^T (y_t - \beta_1 - \beta_2 x_t)^2 \quad (3.3.4)$$

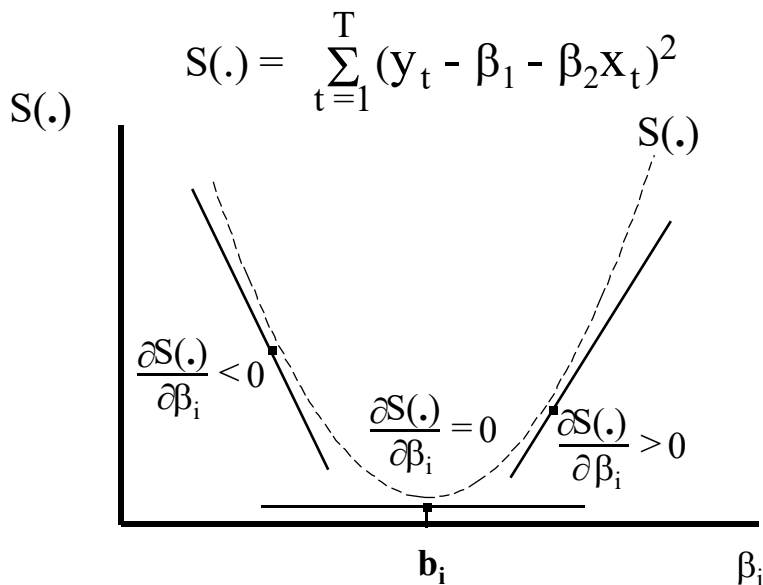
Minimize w.r.t.  $\beta_1$  and  $\beta_2$ :

$$S(\beta_1, \beta_2) = \sum_{t=1}^T (y_t - \beta_1 - \beta_2 x_t)^2 \quad (3.3.4)$$

$$\frac{\partial S(\cdot)}{\partial \beta_1} = -2 \sum (y_t - \beta_1 - \beta_2 x_t)$$

$$\frac{\partial S(\cdot)}{\partial \beta_2} = -2 \sum x_t (y_t - \beta_1 - \beta_2 x_t)$$

두 편미분을 0으로 놓고, 이 두 방정식을 두 미지수  $\beta_1$   $\beta_2$ 에 대해 풀면:



$$\frac{\partial S(.)}{\partial \beta_1} = -2 \sum (y_t - b_1 - b_2 x_t) = 0$$

$$\frac{\partial S(.)}{\partial \beta_2} = -2 \sum x_t (y_t - b_1 - b_2 x_t) = 0$$

이 두 식을 0으로 놓게 되면  $\beta_1$  and  $\beta_2$  는  $b_1$  and  $b_2$  로 표기할 수 있는데 이는 이들이 더 이상 임의의 미지의 값  $\beta_1$  and  $\beta_2$  를 나타내는 것이 아니라  $S(.)$ 의 최소값에 대응되는 어떤 특별한 값을 나타내기 때문임.

$$-2 \sum (y_t - b_1 - b_2 x_t) = 0$$

$$-2 \sum x_t (y_t - b_1 - b_2 x_t) = 0$$

$$\sum y_t - T b_1 - b_2 \sum x_t = 0$$

$$\sum x_t y_t - b_1 \sum x_t - b_2 \sum x_t^2 = 0$$

$$T b_1 + b_2 \sum x_t = \sum y_t$$

$$b_1 \sum x_t + b_2 \sum x_t^2 = \sum x_t y_t$$

최소제곱 추정(The Least Squares Estimation)

계량경제학  
3.25

$$\begin{aligned} T b_1 + b_2 \sum x_t &= \sum y_t \\ b_1 \sum x_t + b_2 \sum x_t^2 &= \sum x_t y_t \end{aligned}$$

$\bar{x}$ 와  $\bar{y}$ 의 정의를 이용하여  $b_1$ 과  $b_2$ 에 대해서 풀

$$b_2 = \frac{T \sum x_t y_t - \sum x_t \sum y_t}{T \sum x_t^2 - (\sum x_t)^2}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

추정값의 활용

계량경제학  
3.26

탄력성

$$\eta = \frac{y \text{의 변화율}}{x \text{의 변화율}} = \frac{\Delta y / y}{\Delta x / x} = \frac{\Delta y}{\Delta x} \frac{x}{y}$$

미적분학을 이용하면 한 점에서의 탄력성을 얻음

$$\eta = \lim_{\Delta x \rightarrow 0} \left[ \frac{\Delta y}{\Delta x} \frac{x}{y} \right] = \frac{\partial y}{\partial x} \frac{x}{y}$$

## 탄력성

$$E(y) = \beta_1 + \beta_2 x$$

$$\frac{\partial E(y)}{\partial x} = \beta_2$$

$$\eta = \frac{\partial E(y)}{\partial x} \frac{x}{E(y)} = \beta_2 \frac{x}{E(y)}$$

 $(\bar{x}, \bar{y})$ 에서의 탄력성

$$\hat{\eta} = \frac{\partial y}{\partial x} \frac{\bar{x}}{\bar{y}} = b_2 \frac{\bar{x}}{\bar{y}}$$

$$\hat{y}_t = b_1 + b_2 x_t = 4 + 1.5 x_t$$

$\bar{x} = 8$  = average number of years of experience  
 $\bar{y} = \$16$  = average wage rate

$$\hat{\eta} = b_2 \frac{\bar{x}}{\bar{y}} = 1.5 \frac{8}{16} = 0.75$$

## 예측

Estimated regression equation:

$$\hat{y}_t = 4 + 1.5 x_t$$

$x_t$  = years of experience

$\hat{y}_t$  = predicted wage rate

If  $x_t = 2$  years, then  $\hat{y}_t = \$7.00$  per hour.

If  $x_t = 3$  years, then  $\hat{y}_t = \$8.50$  per hour.

## Log-log 모형

$$\ln(y) = \beta_1 + \beta_2 \ln(x)$$

$$\frac{\partial \ln(y)}{\partial x} = \beta_2 \frac{\partial \ln(x)}{\partial x}$$

$$\frac{1}{y} \frac{\partial y}{\partial x} = \beta_2 \frac{1}{x} \frac{\partial x}{\partial x}$$

**Log-log 모형**  $\frac{1}{y} \frac{\partial y}{\partial x} = \beta_2 \frac{1}{x} \frac{\partial x}{\partial x}$

$$\frac{x}{y} \frac{\partial y}{\partial x} = \beta_2$$

**x에 대한 y의 탄력성**

$$\eta = \frac{x}{y} \frac{\partial y}{\partial x} = \beta_2$$