제 6 강

단순 회귀모형 - 기타
(Simple Regression Model - etc)

y_t의 변동에 대한 설명

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

설명된 부분: $\hat{y}_t = b_1 + b_2 x_t$

설명되지 않은 부분:

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

y_t의 변동에 대한 설명

$$y_t = \hat{y}_t + \hat{\varepsilon}_t$$

$\bar{y}$ 를 기준으로 삼음

$$y_t - \bar{y} = \hat{y}_t - \bar{y} + \hat{\varepsilon}_t$$

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}\hat{\varepsilon}_t^2$$

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$$

cross product term drops out

y_t의 변동에 대한 설명

총변동

SST = total sum of squares

SST y_t의 $\bar{y}$주변에서의 총 변동을 측정

$$\mathrm{SST} = \sum_{t=1}^{T}(y_t - \bar{y})^2$$

y_t의 변동에 대한 설명

설명되는 변동

SSR = regression sum of squares

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

SSR $\hat{y}_t$ 의 $\bar{y}$주변에서의 변동을 측정함

$$SSR = \sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2$$

y_t의 변동에 대한 설명

설명 안되는 변동

SSE = error sum of squares

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

SSE y_t 의 $\hat{y}_t$ 주변에서의변동을 측정함

$$SSE = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$$

y_t의 변동에 대한 설명

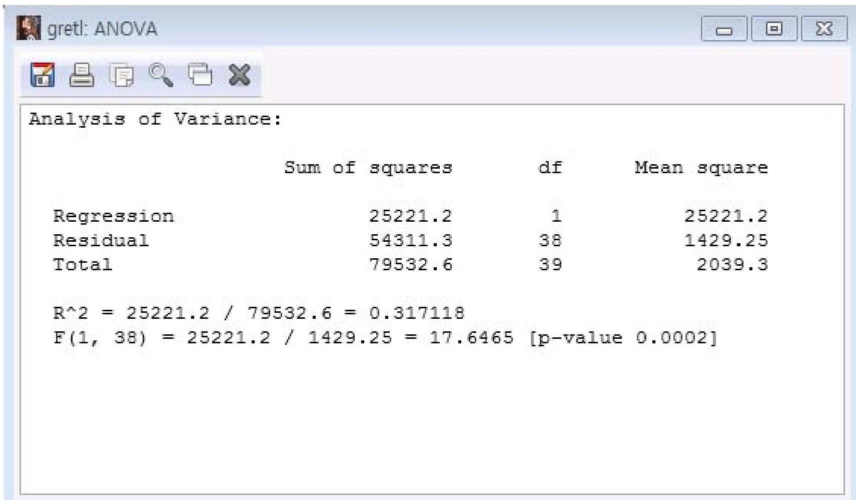
분산분석표

Table 6.1 Analysis of Variance Table

Source of Variation	DF	Sum of Squares	Mean Square
Explained	1	SSR	SSR/1
Unexplained	T-2	SSE	SSE/(T-2) [= $\hat{\sigma}^2$]
Total	T-1	SST	

y_t의 변동에 대한 설명

분산분석표




```
Analysis of Variance:

                     Sum of squares       df      Mean square

  Regression               25221.2         1          25221.2
  Residual                 54311.3        38          1429.25
  Total                    79532.6        39           2039.3

  R^2 = 25221.2 / 79532.6 = 0.317118
  F(1, 38) = 25221.2 / 1429.25 = 17.6465 [p-value 0.0002]
```

y_t의 변동에 대한 설명

결정계수

y_t 의 변동 중 얼만큼의 부분이 설명되고 있는가?

$$R^2 = \frac{SSR}{SST}$$

$$0 \leq R^2 \leq 1$$

y_t의 변동에 대한 설명

결정계수

$$SST = SSR + SSE$$

Dividing by SST

$$\frac{SST}{SST} = \frac{SSR}{SST} + \frac{SSE}{SST}$$

$$1 = \frac{SSR}{SST} + \frac{SSE}{SST}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

결정계수

R^2 는 단지 기술적(descriptive) 척도임

R^2 가 회귀모형의 질적 수준을 나타내는 것은 아님.

R^2 를 극대화하는 것에 초점을 맞추는 것은 좋은 생각이 아님

상관분석

수학적 상관계수

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\,\sqrt{\operatorname{var}(Y)}}$$

표본(경험적) 상관계수

$$r = \frac{\widehat{\operatorname{cov}}(X,Y)}{\sqrt{\widehat{\operatorname{var}}(X)}\,\sqrt{\widehat{\operatorname{var}}(Y)}}$$

상관분석

$$\widehat{\text{var}}(X) = \sum_{t=1}^{T}(x_t - \bar{x})^2/(T-1)$$

$$\widehat{\text{var}}(Y) = \sum_{t=1}^{T}(y_t - \bar{y})^2/(T-1)$$

$$\widehat{\text{cov}}(X,Y) = \sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})/(T-1)$$

상관분석

표본상관계수

$$r = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

상관분석과 결정계수

단순회귀분석에 있어서:

$$\boxed{r^2 = R^2}$$

R^2 는 또한 y_t 와 $\hat{y}_t$ 간의 상관계수의 제곱을 나타냄
: 적합도(goodness of fit)의 척도?

상관분석과 결정계수 (참고)

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2} = \frac{\sum(b_1 + b_2 x_t - b_1 - b_2\bar{x})^2}{\sum(y_t - \bar{y})^2} = \frac{b_2^{\,2}\sum(x_t - \bar{x})^2}{\sum(y_t - \bar{y})^2}$$

$$= \left(\frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2}\right)^2 \frac{\sum(x_t - \bar{x})^2}{\sum(y_t - \bar{y})^2}$$

$$= \frac{\left(\sum(x_t - \bar{x})(y_t - \bar{y})\right)^2}{\sum(x_t - \bar{x})^2 \sum(y_t - \bar{y})^2}$$

$$= r^2$$

y_t의 변동에 대한 설명

상관분석과 결정계수 (참고)

$$r^2 = \frac{\left(\sum(\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y})\right)^2}{\sum(\hat{y}_t - \bar{\hat{y}})^2 \sum(y_t - \bar{y})^2}$$

$$= \frac{\left(\sum(b_1 + b_2 x_t - b_1 - b_2 \bar{x})(y_t - \bar{y})\right)^2}{\sum(b_1 + b_2 x_t - b_1 - b_2 \bar{x})^2 \sum(y_t - \bar{y})^2}$$

$$= \frac{b_2^{\,2}\left(\sum(x_t - \bar{x})(y_t - \bar{y})\right)^2}{b_2^{\,2}\sum(x_t - \bar{x})^2 \sum(y_t - \bar{y})^2} = R^2$$

회귀분석결과의 발표

추정결과

통계패키지의 전형적인 회귀분석 결과 보고 내용

Computer Generated Least Squares Results

(1)	(2)	(3)	(4)	(5)
Variable	Parameter Estimate	Standard Error	t for H0: Parameter=0	Prob>|t|
INTERCEPT	40.7676	22.1387	1.841	0.0734
X	0.1283	0.0305	4.201	0.0002

회귀분석결과의 발표

추정결과

$$b_1 = 40.7676 \qquad b_2 = 0.1283$$

$$se(b_1) = \sqrt{\hat{var}(b_1)} = \sqrt{490.12} = 22.1287$$

$$se(b_2) = \sqrt{\hat{var}(b_2)} = \sqrt{0.0009326} = 0.0305$$

$$t = \frac{b_1}{se(b_1)} = \frac{40.7676}{22.1287} = 1.84$$

$$t = \frac{b_2}{se(b_2)} = \frac{0.1283}{0.0305} = 4.20$$

회귀분석결과의 발표

분산분석

종속변수의 변동의 원천들:

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Explained	1	25221.2229	25221.2229
Unexplained	38	54311.3314	1429.2455
Total	39	79532.5544	

R-square: 0.3171

분산분석

$$SST = \sum(y_t - \bar{y})^2 = 79532$$

$$SSR = \sum(\hat{y}_t - \bar{y})^2 = 25221$$

$$SSE = \sum \hat{e}_t^{\,2} = 54311$$

$$SSE/(T\text{-}2) = \hat{\sigma}^2 = 1429.2455$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 0.317$$

$$\hat{y}_t = 40.7676 + 0.1283x_t$$
$$(\text{s.e.})\ (22.1387)\ (0.0305)$$

$$\hat{y}_t = 40.7676 + 0.1283x_t$$
$$(t)\quad (1.84)\quad (4.20)$$

회귀분석결과의 발표

$$R^2 = 0.317$$

이 R^2 값은 낮아보일 수 있으나 개별적 미시적 수준 에서 분석된 횡단면 **(cross-sectional)** 자료와 관련된 연구에서 전형적인 경우임

집계적 거시적 수준에서 분석된 시계열 **(time-series)** 자료와 관련된 연구에서는 매우 높은 수준의 R^2값이 전형적으로 나타남

관측치의 단위를 변경하는 효과

X의 단위를 변경

해당모수의 추정값과 표준오차는 변화 하지만 t값과 R^2 값은 불변.

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

$$y_t = \beta_1 + (c\beta_2)(x_t/c) + \varepsilon_t$$

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

where

$$\beta_2^* = c\beta_2 \quad \text{and} \quad x_t^* = x_t/c$$

X의 단위를 변경

```
gretl: model 1
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 1: OLS, using observations 1-40
Dependent variable: y

                    coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const               40.7676      22.1387       1.841    0.0734  *
  x_divided_by_100    12.8289       3.05393      4.201    0.0002  ***

Mean dependent var   130.3130   S.D. dependent var   45.15857
Sum squared resid    54311.33   S.E. of regression   37.80536
R-squared            0.317118   Adjusted R-squared   0.299148
F(1, 38)             17.64653   P-value(F)           0.000155
Log-likelihood      -201.0297   Akaike criterion     406.0594
Schwarz criterion    409.4372   Hannan-Quinn         407.2807
```


Y의 단위를 변경

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

$$y_t/c = (\beta_1/c) + (\beta_2/c)x_t + \varepsilon_t/c$$

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

where $y_t^* = y_t/c \qquad \varepsilon_t^* = \varepsilon_t/c$

$\beta_1^* = \beta_1/c \quad$ and $\quad \beta_2^* = \beta_2/c$

모든 모수의 추정값이 변화하지만 역시 t값과 R^2 값은 불변

관측치의 단위를 변경하는 효과

Y의 단위를 변경

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-40
Dependent variable: y_divided_by_100

             coefficient   std. error    t-ratio   p-value
  --------------------------------------------------------
  const      0.407676      0.221387       1.841    0.0734   *
  x          0.00128289    0.000305393    4.201    0.0002   ***

Mean dependent var   1.303130   S.D. dependent var   0.451586
Sum squared resid    5.431133   S.E. of regression   0.378054
R-squared            0.317118   Adjusted R-squared   0.299148
F(1, 38)             17.64653   P-value(F)           0.000155
Log-likelihood      -16.82291   Akaike criterion     37.64582
Schwarz criterion    41.02358   Hannan-Quinn         38.86711
```

관측치의 단위를 변경하는 효과

x와 y의 단위를 같이 변경

기울기에 대한 추정값외에 모든 모수에 대한 추정값이 변화하나 t값이나 R^2값은 불변

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

$$y_t/c = (\beta_1/c) + (c\beta_2/c)x_t/c + \varepsilon_t/c$$

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

where $y_t^* = y_t/c \qquad \varepsilon_t^* = \varepsilon_t/c$

$\beta_1^* = \beta_1/c$ and $x_t^* = x_t/c$

관측치의 단위를 변경하는 효과

x와 y의 단위를 같이 변경

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
Model 3: OLS, using observations 1-40
Dependent variable: y_divided_by_100

                    coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------------
  const              0.407676     0.221387      1.841    0.0734   *
  x_divided_by_100   0.128289     0.0305393     4.201    0.0002   ***

Mean dependent var   1.303130   S.D. dependent var   0.451586
Sum squared resid    5.431133   S.E. of regression   0.378054
R-squared            0.317118   Adjusted R-squared   0.299148
F(1, 38)             17.64653   P-value(F)           0.000155
Log-likelihood      -16.82291   Akaike criterion     37.64582
Schwarz criterion    41.02358   Hannan-Quinn         38.86711
```

함수의 형태

선형 vs. 비선형

단순 ***선형***회귀모형에 있어서 ***선형(linear)***은 변수들간의 관계가 선형임을 의미하는 것이 아니라 모수들이 모형에 선형의 방식으로 들어감을 의미함

함수의 형태

선형 vs. 비선형

선형모형:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \qquad y_t = \beta_1 + \beta_2 \ln(x_t) + \varepsilon_t$$

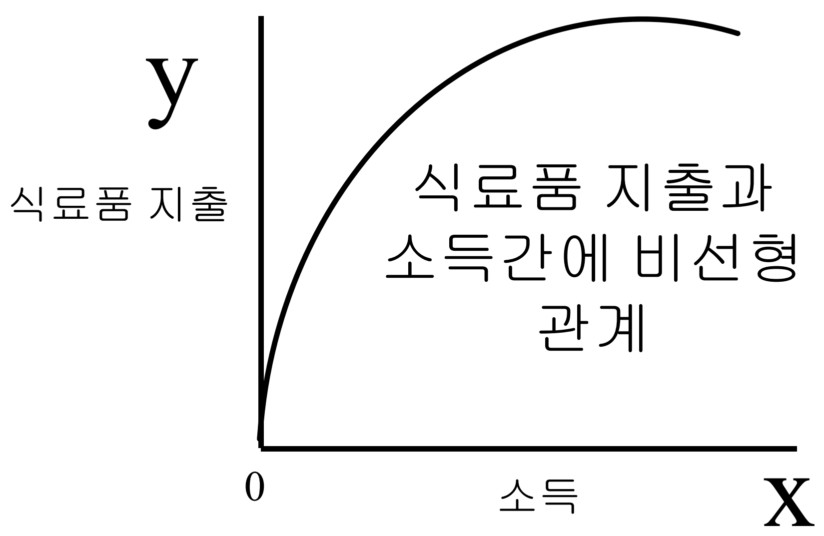
$$\ln(y_t) = \beta_1 + \beta_2 x_t + \varepsilon_t \qquad y_t = \beta_1 + \beta_2 x_t^2 + \varepsilon_t$$

비선형모형:

$$y_t = \beta_1 + \beta_2 x_t^{\beta_3} + \varepsilon_t \qquad y_t^{\beta_3} = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$$y_t = \beta_1 + \beta_2 x_t + \exp(\beta_3 x_t) + \varepsilon_t$$

함수의 형태



유용한 함수의 형태

각 형태를 파악하고 특히 그 기울기와 탄력성을 관찰

1. Linear
2. Reciprocal
3. Log-Log
4. Log-Linear
5. Linear-Log
6. Log-Inverse

선형

Linear

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

slope: β_2 elasticity: $\beta_2 \dfrac{x_t}{y_t}$

함수의 형태

역수

Reciprocal

$$y_t = \beta_1 + \beta_2 \frac{1}{x_t} + \varepsilon_t$$

slope:

$$-\beta_2 \frac{1}{x_t^2}$$

elasticity:

$$-\beta_2 \frac{1}{x_t\, y_t}$$

함수의 형태

로그-로그

Log-Log

$$\ln(y_t) = \beta_1 + \beta_2 \ln(x_t) + \varepsilon_t$$

slope: $\beta_2 \frac{y_t}{x_t}$ elasticity: β_2

로그-선형

Log-Linear

$$\ln(y_t) = \beta_1 + \beta_2 x_t + \varepsilon_t$$

slope: $\beta_2 y_t$ elasticity: $\beta_2 x_t$

선형-로그

Linear-Log

$$y_t = \beta_1 + \beta_2 \ln(x_t) + \varepsilon_t$$

slope: $\beta_2 \frac{1}{x_t}$ elasticity: $\beta_2 \frac{1}{y_t}$

로그-역수

Log-Inverse

$$\ln(y_t) = \beta_1 + \beta_2 \frac{1}{x_t} + \varepsilon_t$$

$$\text{slope: } -\beta_2 \frac{y_t}{x_t^2} \qquad \text{elasticity: } -\beta_2 \frac{1}{x_t}$$

선택의 기준 – 선형모형의 가정

1. $E(\varepsilon_t) = 0$
2. $\mathrm{var}(\varepsilon_t) = \sigma^2$
3. $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$
4. $\varepsilon_t \sim N(0, \sigma^2)$

수요모형

1. Demand Models

* quantity demanded (y_d) and price (x)

* constant elasticity

$$\ln(y_t^d) = \beta_1 + \beta_2 \ln(x)_t + \varepsilon_t$$

공급모형

2. Supply Models

* quality supplied (y^s) and price (x)

* constant elasticity

$$\ln(y_t^S) = \beta_1 + \beta_2 \ln(x_t) + \varepsilon_t$$

생산함수

3. Production Functions

* output (y) and input (x)

* constant elasticity

Cobb-Douglas Production Function:

$$\ln(y_t) = \beta_1 + \beta_2 \ln(x_t) + \varepsilon_t$$

비용함수

4a. Cost Functions

* total cost (y) and output (x)

$$y_t = \beta_1 + \beta_2 x^2_t + \varepsilon_t$$

경제모형과 가능한 함수형태의 선택

비용함수

4b. Cost Functions

* average cost (y/x) and output (x)

$$(y_t/x_t) = \beta_1/x_t + \beta_2 x_t + \varepsilon_t/x_t$$

잔차의 정규분포?

가설검정과 구간 추정에 있어서 중요한 전제

Skewness(비대칭도) – 어떤 확률변수의 분포가 평균을 중심으로 얼마나 비대칭인가를 나타내는 척도 (정규분포의 Skewness = 0)

: $s \equiv E[(X-\mu_X)^3]/\sigma_X^3$,

Kurtosis(첨예도)-어떤 확률변수의 분포가 얼마나 뾰쪽한가를 나타내는 척도(정규분포의 Kurtosis=3)

: $k \equiv E[(X-\mu_X)^4] / \sigma_X^4$,

Jacque-Bera (JB) Normality test (자크베라의 정규성 검정)

귀무가설과 대립가설 :

H_0: $X \sim N(\mu_X, \sigma_X^2)$, H_1: H_0 is not true

NOTE: 정규분포임을 보이는 목적으로 사용하기에는 부적절한 가설설정임

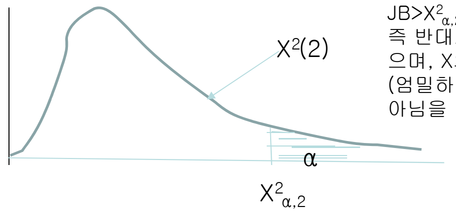
잔차의 정규분포?

검정통계량

$$JB \equiv \frac{T}{6}\left(\hat{s}^2 + \frac{\left(\hat{k}-3\right)^2}{4}\right) \sim \chi^2(2) : \text{under } H_0$$

X가 정규분포라면 그 경험적 비대칭도와 경험적 첨예도는 각각 0과 3에 가까운 값일 것이므로 JB의 값은 0에 가까운 값이 됨. 따라서 JB가 0보다 충분히 클 때, 즉 오른쪽 꼬리의 값이 나올 때 귀무가설을 기각함

기각역



JB>$X^2_{\alpha,2}$ 이면 유의수준 α에서 H_0를 기각함. 즉 반대로 JB<$X^2_{\alpha,2}$ 이면 H_0를 기각할 수 없으며, X의 분포가 정규분포라고 볼 수 있음 (엄밀하게는 X의 분포가 정규분포가 아님을 입증하지 못한 것임)