추세 분석

추세모형 (Trend Model)

- **추세분석**
  - 전통적인 시계열분석 기법으로 관측값 Z_t를 시간의 함수 $f(t)$로서 표현하는 방법

- **다항식 추세모형** (polynomial trend model)

$$Z_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p + \varepsilon_t$$

  - 상수평균모형 (constant mean model) : p=0
  - 선형추세모형 (linear trend model) : p=1
  - 2차 추세모형 (quadratic trend model) : p=2, …
  - 중회귀모형(multiple linear regression model)의 특별한 경우

추세모형 (Trend Model)

- **계절추세모형** (seasonal trend model)
  - 삼각함수나 지시함수를 이용하여 계절성분 설명

- **비선형추세모형** (nonlinear trend model)
  - 시계열이 기하급수적으로 증가하는 양상을 보이거나 비선형적으로 움직일 때 주로 사용됨
  - 일반적으로 성장곡선(growth curve)이 많이 사용됨
  - 예 : $Z_t = \exp(\beta_0 + \beta_1 t)\,\varepsilon_t$

    $\rightarrow \ln(Z_t) = \beta_0 + \beta_1 t + \ln(\varepsilon_t)$

회귀분석 - 회귀모형

- **회귀모형** (seasonal trend model)

$$Z_t = f(X_t; \beta) + \varepsilon_t, \qquad t = 1,2, \dots, n$$

- Z_t : 반응변수 (response variable)

  또는 종속변수(dependent variable)

- X_t : 설명변수(explanatory variable)

  또는 독립변수(independent variable)

- β : (미지의) 회귀계수 또는 모수(regression parameter)

- ε_t : 오차 (error), i.i.d.

$$\varepsilon_t \sim N(0, \sigma^2)$$

회귀분석 - 모수추정

- **최소제곱법** (method of least squares)
  - 오차제곱합 : $S(\beta) = \sum_{t=1}^{n} \{Z_t - f(X_t; \beta)\}^2$
  - 최소제곱추정량(LSE) : $\hat{\beta} = argmin_{\beta}\ S(\beta)$
- **중회귀모형의 경우**
  - 모델 : $Z = X\beta + \varepsilon$
  - LSE : $\hat{\beta} = argmin_{\beta}\ \sum_{t=1}^{n} (Z_t - X_t\beta)^2 = (X'X)^{-1}X'Z$
  - 성질
- **성질**
  - $\mathrm{E}(\hat{\beta}) = \beta$
  - $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

회귀분석 - 중회귀분석

적합값과 잔차

- 적합값 (fitted value) : $\hat{Z} = X\hat{\beta}$
- 잔차 (residual) : $\mathrm{e} = Z - \hat{Z} = (I - X(X'X)^{-1}X')Z$

분산분석표와 결정계수

- 회귀모형의 유의성 검정

ANOVA table

	d.f.	SS	MS	F (ratio)	Sig.
Model	P	SSR	MSR=SSR/p	F=MSR/MSE	P-value
Error	n-p-1	SSE	MSE=SSE/(n-p-1)		
(Corrected) total	n-1	SST			

- 결정계수(coefficient of determination) : $R^2 = SSR/SST$

회귀분석 - 구간추정과 가설검정

- **개별 모수의 신뢰구간 추정**
- **개별모수/회귀모형의 유의성 검정**
- **모수 집합의 유의성 검정**
  - 축소모형 (reduced model) :

$$Z_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_q X_{tq} + \varepsilon_t$$

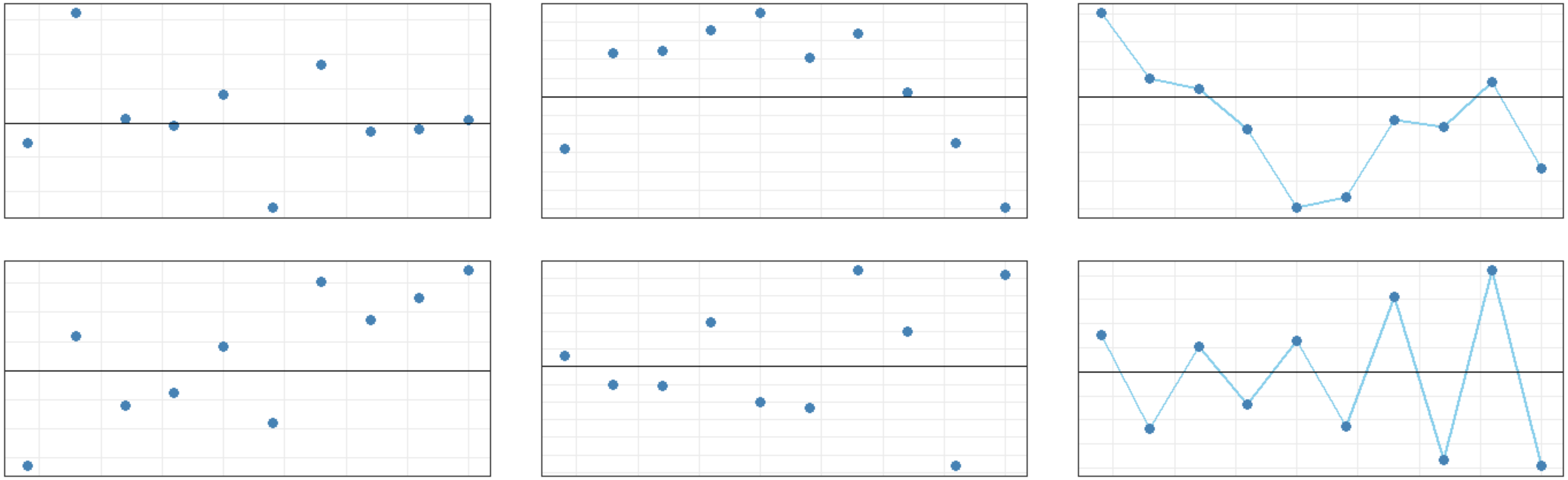
  - 완전모형 (full model) :

$$Z_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_q X_{tq} + \cdots + \beta_p X_{tp} + \varepsilon_t$$

  - 가설 : $H_0 : \beta_{q+1} = \cdots = \beta_p = 0 \;\; vs. \;\; H_1 : not \; H_0$
  - 검정통계량 : $F = \dfrac{\frac{\{SSR(full) - SSR(reduced)\}}{p-q}}{\frac{SSE}{n-p-1}} \sim_{H_0} F(p-q, n-p-1)$
  - F의 값이 크면 귀무가설 기각

회귀분석 - 잔차분석(Residual Analysis)

- **오차항에 대한 가정을 잔차를 통해 검토**
  - 오차항에 대한 가정 : $\varepsilon_t \sim N(0, \sigma^2)$, i.i.d.
  - 잔차의 산점도 : 선형성, 등분산성, 자기상관관계 검토
  - 잔차들의 정규확률 그림 : 정규성 검토
  - 오차항들의 자기상관관계 검정

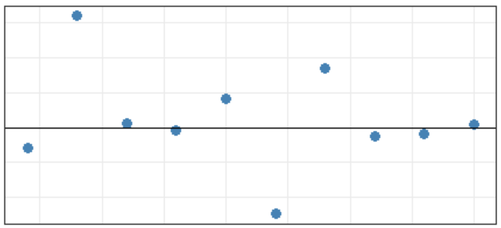


회귀분석 - 잔차분석(Residual Analysis)

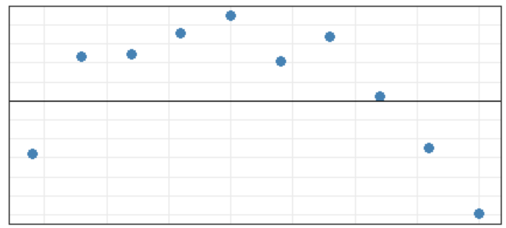
- **Durbin-Watson(DW) 검정**
  - 오차항들 간에 1차 자기상관 검토
  - 검정통계량 : $D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \cong 2(1 - \hat{\rho}_1)$
- $\boldsymbol{H_0} : \varphi = 0$ **에 대한 검정 규칙**
  - 상한값 $= d_U = d_U(\alpha, p, n)$, 하한값=$d_L = d_L(\alpha, p, n)$
  - 양(음)의 상관관계 "$H_1 : \varphi > 0$ ($H_1 : \varphi < 0$)"
    - $D > d_U$ $(4 - D > d_U)$ 이면 유의수준 α에서 귀무가설 기각
    - $D < d_L$ $(4 - D < d_L)$ 이면 유의수준 α에서 귀무가설 기각 못함
    - $d_L < D < d_U$ $(d_L < 4 - D < d_U)$ 이면 검정 유보
  - 자기상관관계 "$H_1 : \varphi \neq 0$"
    - D or $4 - D > d_U$이면 유의수준 2α에서 귀무가설 기각
    - D or $4 - D < d_L$이면 유의수준 2α에서 귀무가설 기각 못함
    - 나머지 경우는 검정 유보

회귀분석 - 잔차분석(Residual Analysis)

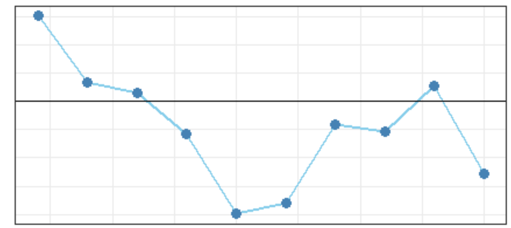
Durbin-Watson(DW) 검정



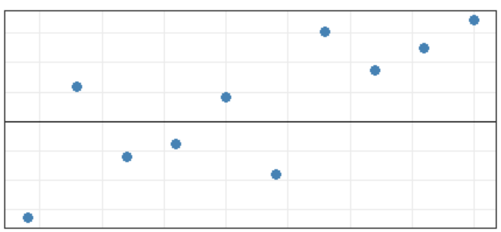
D = 3.013, p-value = 0.921



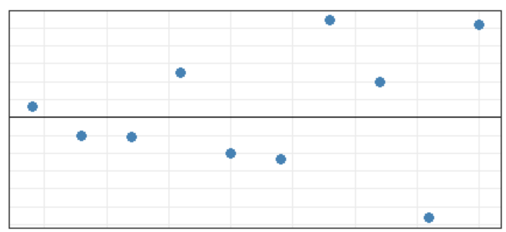
D = 0.734, p-value = 0.001



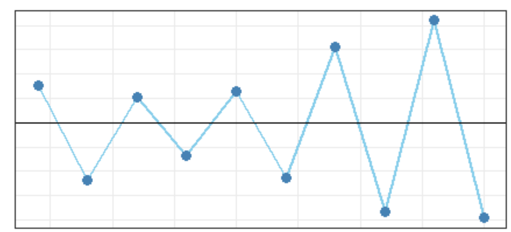
D = 1.066, p-value = 0.015



D = 3.013, p-value = 0.921



D = 2.661 p-value = 0.767



D = 3.507, p-value = 0.996

다항추세모형 - 상수평균모형

- **상수평균 모형 : 불규칙성분만을 갖는 경우**
  - $Z_t = \beta_0 + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \sigma^2)$, i.i.d.
  - LSE : $\hat{\beta}_0 = \bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t$
  - 시점 n에서의 l-시차 후의 예측값 : $\hat{Z}_n(l) = \hat{\beta}_0 = \bar{Z}$
  - 예측오차 : $\hat{e}_n(l) = Z_{n+l} - \hat{Z}_n(l) = \beta_0 + \varepsilon_{n+l} - \bar{Z}$
  - 예측오차의 기대값 : $\mathrm{E}(\hat{e}_n(l)) = 0$
  - 예측 오차의 분산 : $\mathrm{Var}(\hat{e}_n(l)) = \left(1 + \frac{1}{n}\right)\sigma^2$
  - Z_{n+l} 의 100(1-α)% 예측 구간 : $\bar{Z} \pm t_{\frac{\alpha}{2}} \times \sqrt{\left(1 + \frac{1}{n}\right) s^2}$

다항추세모형 - 상수평균모형

- **상수평균 모형 : 불규칙성분만을 갖는 경우**
  - $Z_t = \beta_0 + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \sigma^2)$, i.i.d.
  - 시점 n에서의 l-시차 후의 예측값 : $\hat{Z}_n(l) = \hat{\beta}_0 = \bar{Z}$
  - 예측오차 : $\hat{e}_n(l) = Z_{n+l} - \hat{Z}_n(l) = \beta_0 + \varepsilon_{n+l} - \bar{Z}$

t	1	2	3	4	5	6	7	8	9	10
X	-1	17	2	-9	4	-3	2	15	-19	10
$\hat{X}(1)$										
$e(1)$										

다항추세모형 - 상수평균모형

예측 갱신

- 새로운 관측값 : Z_{n+1}
- 시점 $n+l$에서의 값 예측
  - 시점 n에서의 l-시차 후의 예측값 : $\hat{Z}_n(l) = \hat{\beta}_0 = \bar{Z}$
  - 시점 $n+1$에서의 $(l-1)$-시차 후의 예측값

$$
\begin{aligned}
\hat{Z}_{n+1}(l-1) &= \frac{1}{n+1}(Z_1 + Z_2 + \cdots + Z_n + Z_{n+1}) \\
&= \frac{1}{n+1}\left(\sum_{t=1}^{n} Z_t + Z_{n+1}\right) \\
&= \frac{n}{n+1} \times \frac{1}{n}\sum_{t=1}^{n} Z_t + \frac{1}{n+1} \times Z_{n+1} \\
&= \frac{n}{n+1} \times \hat{Z}_n(l) + \frac{1}{n+1} \times Z_{n+1} \\
&= \hat{Z}_n(l) + \frac{1}{n+1} \times \{Z_{n+1} - \hat{Z}_n(l)\}
\end{aligned}
$$

다항추세모형 - 선형추세모형

선형추세 모형

- 직선형인 추세를 갖고 증가 하는 경우
- $Z_t = \beta_0 + \beta_1 t + \varepsilon_t, \ \ \varepsilon_t \sim N(0, \sigma^2)$, i.i.d.

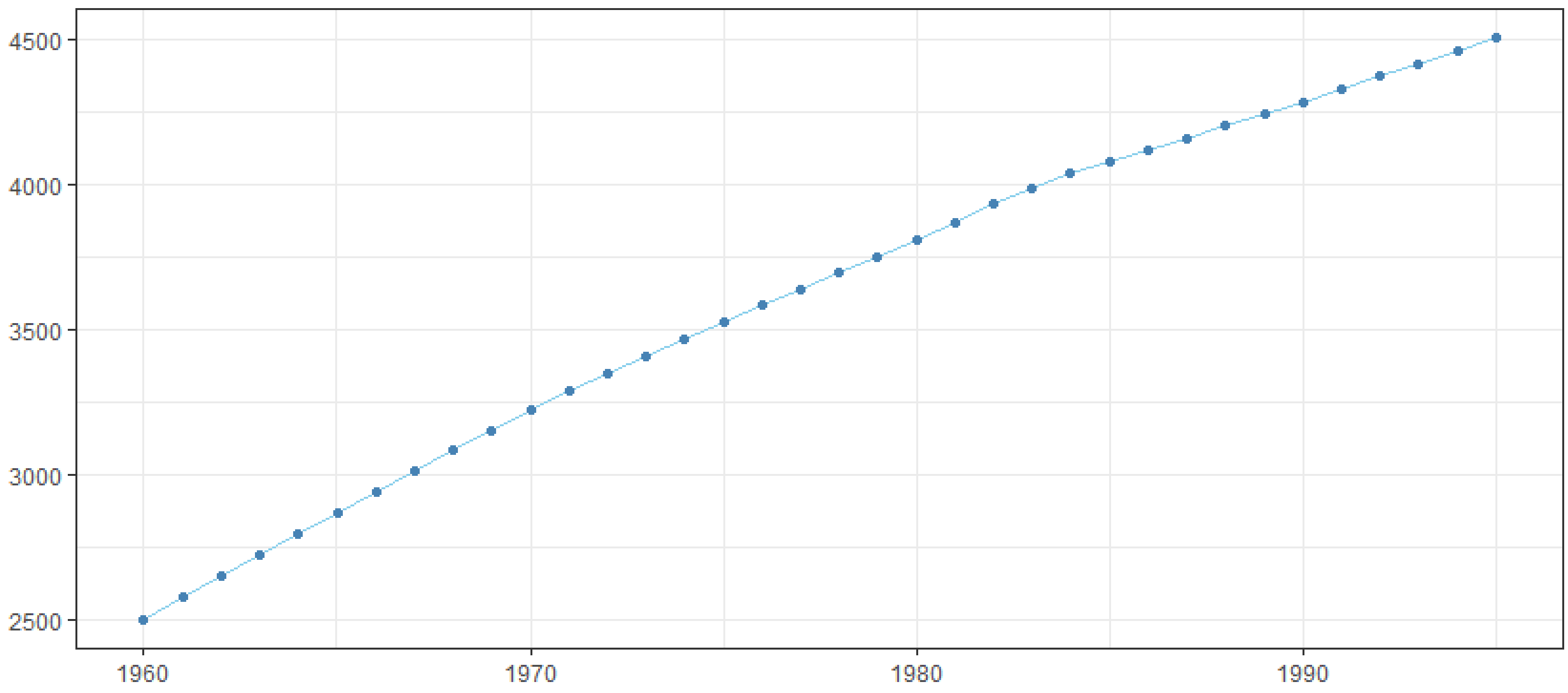
다항추세 모형

- 추세성분이 곡선 형태인 경우
- 2차 추세모형 : $Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t, \ \ \varepsilon_t \sim N(0, \sigma^2)$, i.i.d.
- 다항 추세모형 :

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \mathrm{i.i.d.}$$

다항추세모형 - 선형추세모형

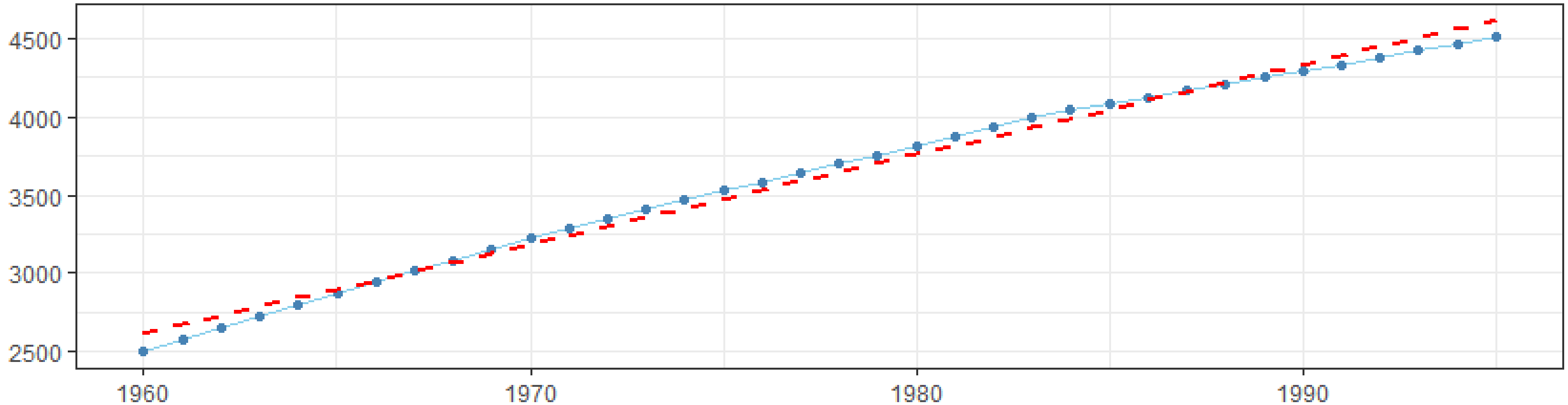
- (예제) 국내인구 총인구



다항추세모형 - 선형추세모형

- **(예제) 국내인구 총인구 :** $Z_t = \beta_0 + \beta_1 t + \varepsilon_t$

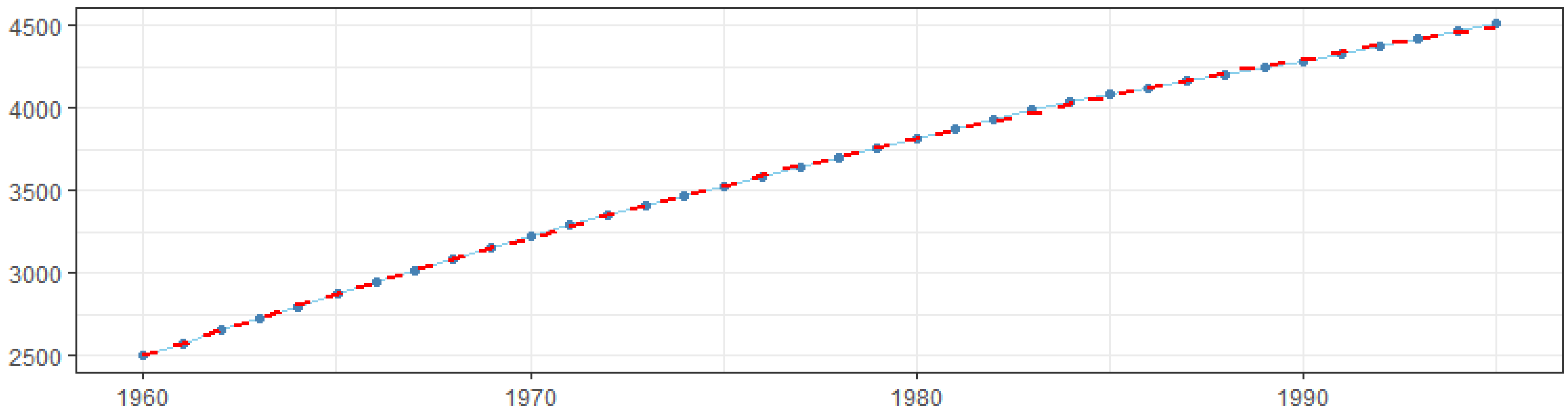
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2559.3889    20.0385  127.72   <2e-16 ***
t             57.0135     0.9444   60.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 58.87 on 34 degrees of freedom
Multiple R-squared:  0.9908,    Adjusted R-squared:  0.9905 
F-statistic:  3644 on 1 and 34 DF,  p-value: < 2.2e-16
```



다항추세모형 - 선형추세모형

- **(예제) 국내인구 총인구 :** $Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2421.49090    4.05820  596.69   <2e-16 ***
t             78.78688    0.50576  155.78   <2e-16 ***
t2            -0.58847    0.01326  -44.38   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.67 on 33 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9998
F-statistic: 1.083e+05 on 2 and 33 DF,  p-value: < 2.2e-16
```



다항추세모형 - 계절추세모형

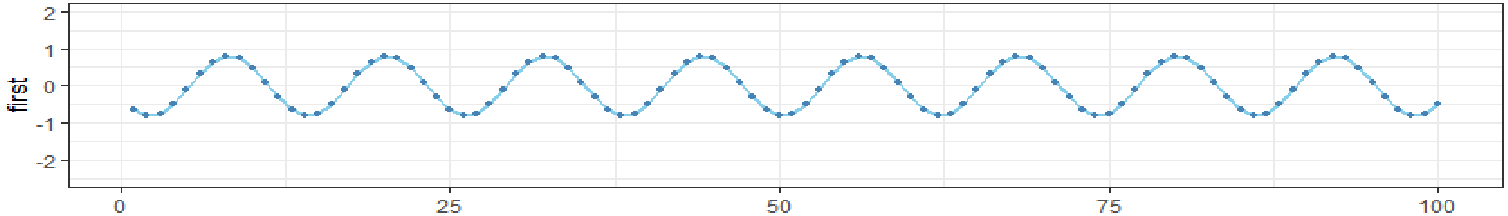
- **계절추세 모형 : 계절성분만을 갖는 경우**
  - $Z_t = \beta_0 + \sum_{i=1}^{m} A_i sin\left(\frac{2\pi i}{s}t + \phi_i\right) + \varepsilon_t$

    $= \beta_0 + \sum_{i=1}^{m}(\beta_{1i}\sin(w_i t) + \beta_{2i}\cos(w_i t)) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$, i.i.d.
  - A_i : 진동폭 (amplitude)
  - ϕ_i : 편각 (phase shift)
  - $w_i = \frac{2\pi i}{s}$: 2π 단위 시간동안 관측되는 i번째 주기항의 빈도 (frequency)
  - $\frac{s}{i}$: i번째 주기항의 주기 (period) ($i \leq m \leq \frac{s}{2}$)

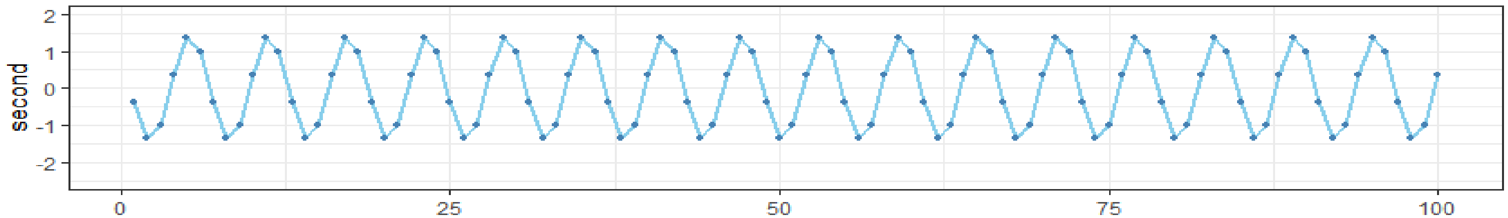
다항추세모형 - 계절추세모형

- (예) $s = 12, m = 2, \beta_0 = 0, A_1 = -0.8, A_2 = 1.4, \phi_1 = \pi/8\,, \phi_2 = 3\pi/4$

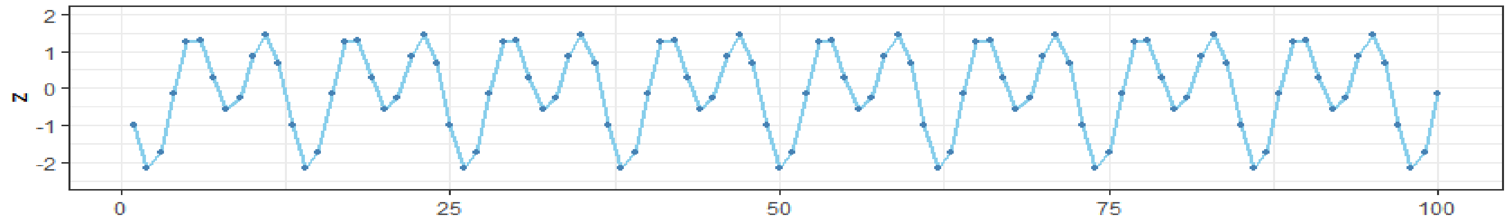
$$A_1 \sin({}^{2\pi t}/_{12} + \pi/8) = \beta_{11} \sin({}^{2\pi t}/_{12}) + \beta_{21} \cos({}^{2\pi t}/_{12})$$



$$A_2 \sin({}^{2\pi t}/_{6} + 3\pi/4) = \beta_{12} \sin({}^{2\pi t}/_{6}) + \beta_{22} \cos({}^{2\pi t}/_{6})$$



$$A_1 \sin({}^{2\pi t}/_{12} + \pi/8) + A_2 \sin({}^{2\pi t}/_{6} + 3\pi/4)$$



다항추세모형 - 선형계절추세모형

- 선형계절추세 모형 : 추세 및 계절성분을 갖는 경우

$$Z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{m} (\beta_{1i} \sin(w_i t) + \beta_{2i} \cos(w_i t)) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \mathrm{i.i.d.}$$

  - 예측

$$\hat{Z}_n(l) = \hat{\beta}_0 + \hat{\beta}_1 (n + l) + \sum_{i=1}^{m} \left(\hat{\beta}_{1i} \sin(w_i (n + l) + \hat{\beta}_{2i} \cos(w_i (n + l)) \right)$$

다항추세모형 - 계절추세모형

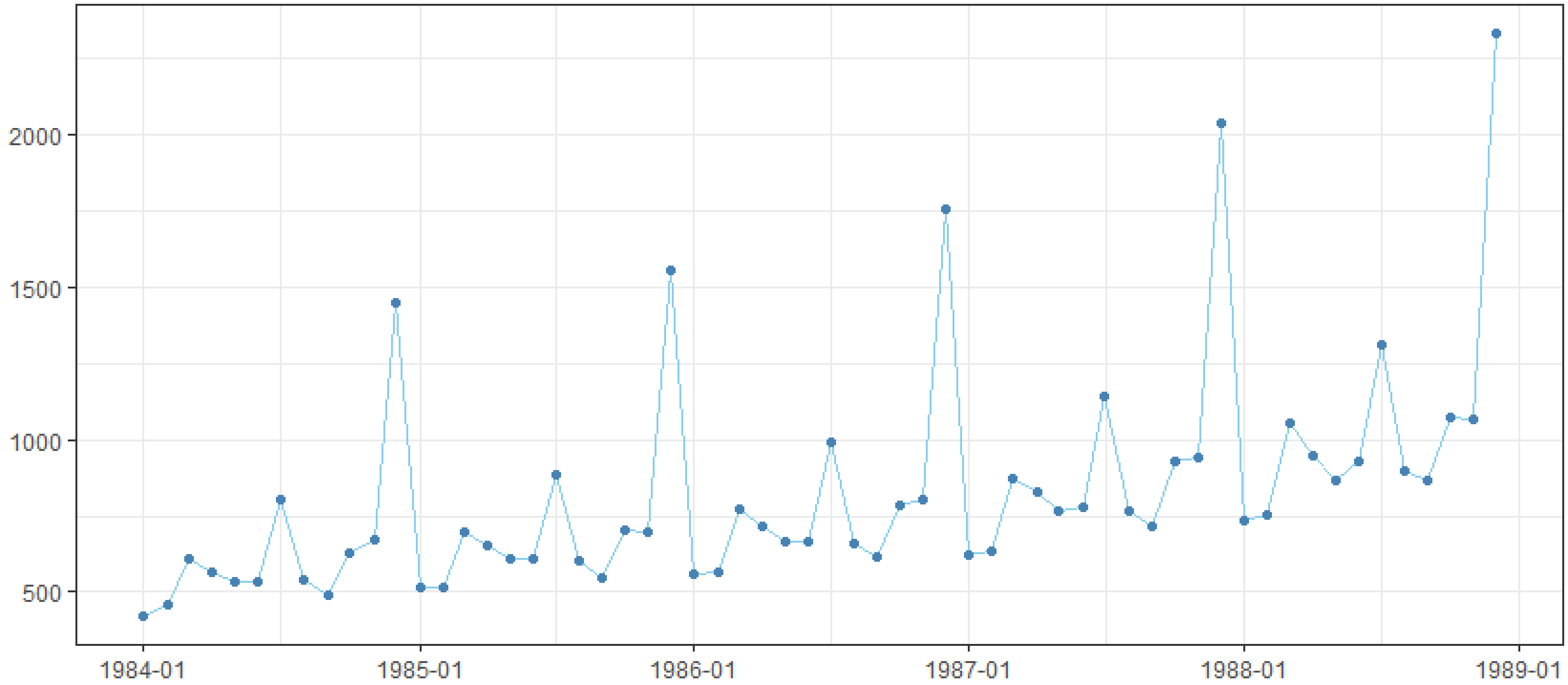
- **계절추세 모형 : 지시함수 사용**
  - $Z_t = \beta_0 + \sum_{i=1}^{s} \beta_i \times IND_{ti} + \varepsilon_t$
  - $IND_{ti} = \begin{cases} 1, & t = i(mod\ s) \\ 0, & etc. \end{cases}$
  - 단) (a) $\beta_0 = 0$ 또는 (b) $\sum_{i=1}^{s} \beta_i = 0$ 또는 (c) $\beta_s = 0$
  - 예) 주기(s)=4 인경우 (1/2/3/4분기), β_0=0

    1분기인 경우- $Z_t = \beta_1 + \varepsilon_t,$ 2분기인 경우 - $Z_t = \beta_2 + \varepsilon_t$

    3분기인 경우- $Z_t = \beta_3 + \varepsilon_t,$ 4분기인 경우 - $Z_t = \beta_4 + \varepsilon_t$

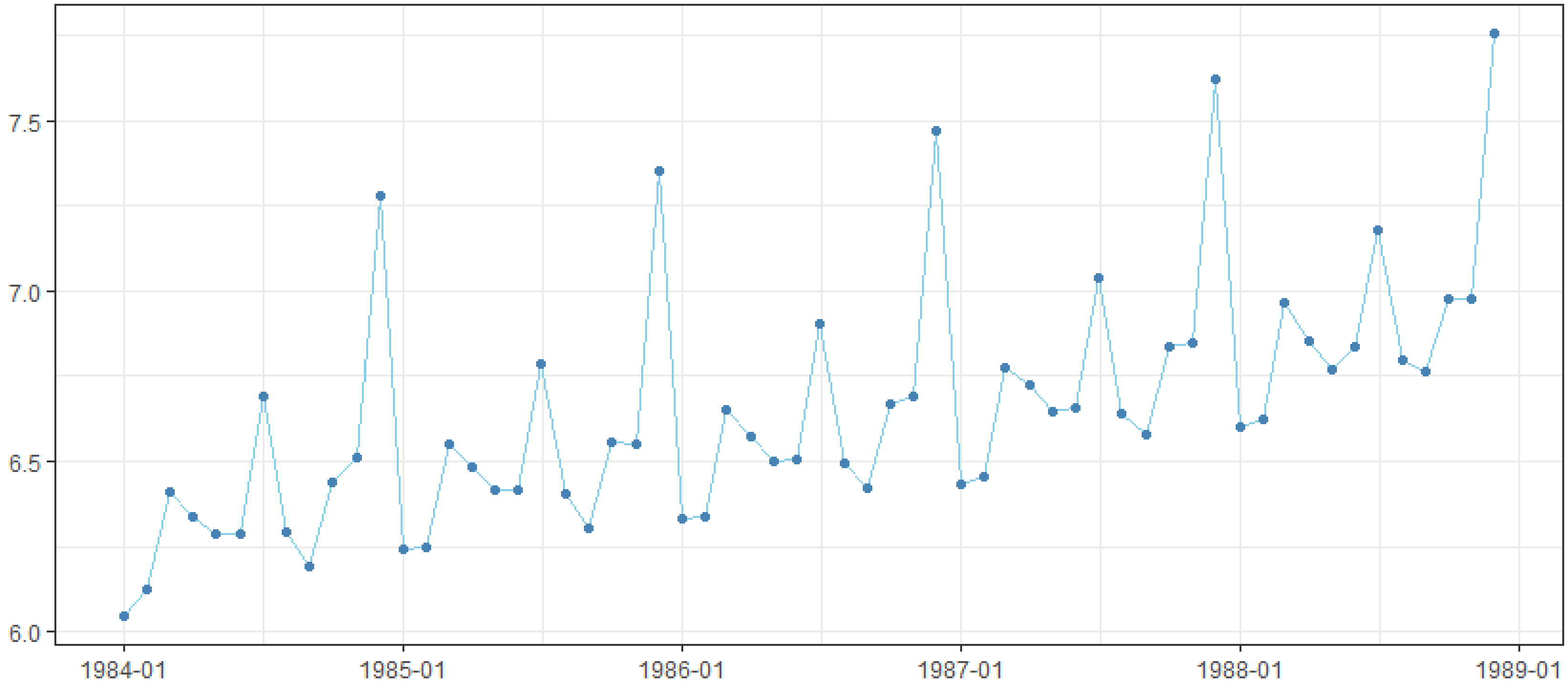
다항추세모형 - 선형계절추세모형

■ **(예제) 백화점 월별 매출액**



다항추세모형 - 선형계절추세모형

- **(예제) 백화점 월별 매출액 (log 변환)**



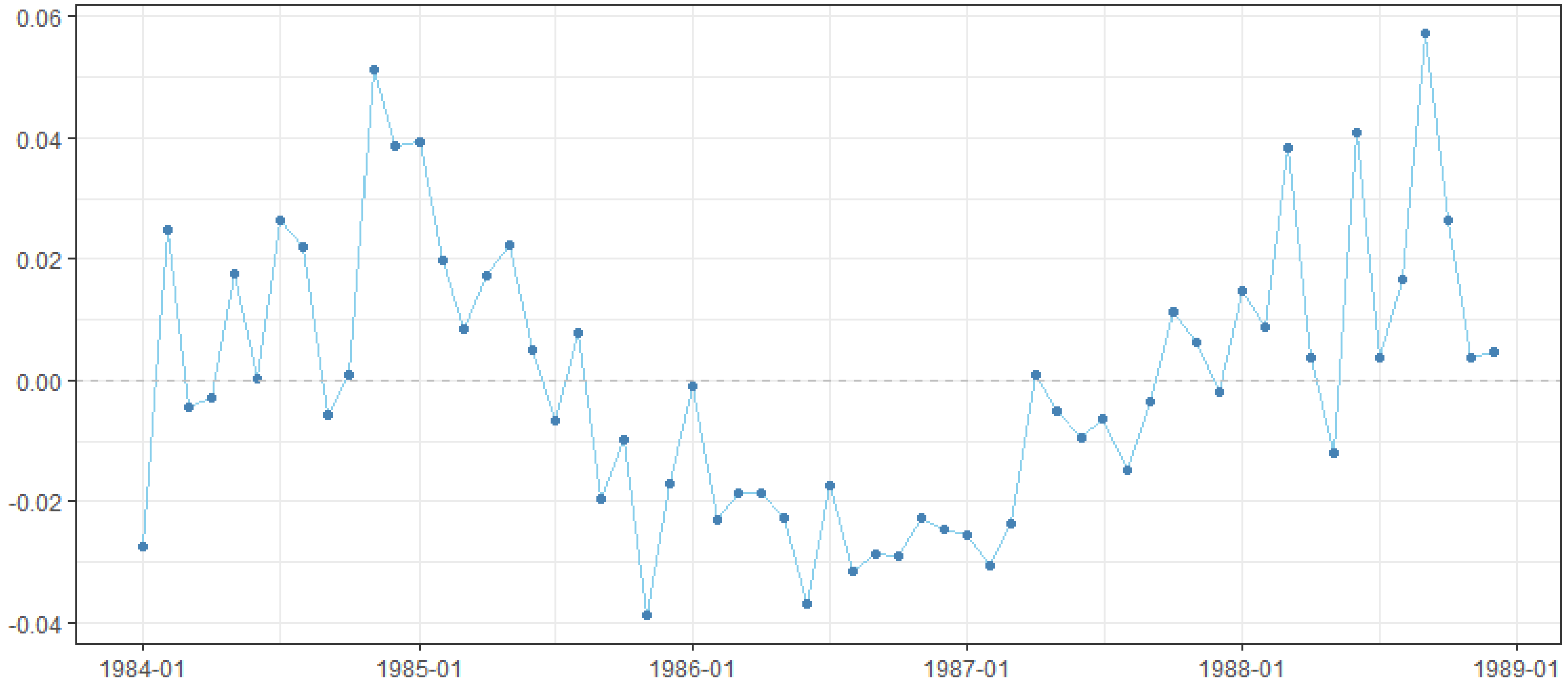
다항추세모형 - 선형계절추세모형

- **(예제) 백화점 월별 매출액 – 중회귀분석 (지시함수)**

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
trend  0.12792    0.00231   55.39   <2e-16 ***
y1     6.07485    0.01222  497.07   <2e-16 ***
y2     6.09146    0.01230  495.43   <2e-16 ***
y3     6.39178    0.01237  516.64   <2e-16 ***
y4     6.30601    0.01245  506.47   <2e-16 ***
y5     6.22390    0.01253  496.62   <2e-16 ***
y6     6.23044    0.01262  493.84   <2e-16 ***
y7     6.59917    0.01270  519.51   <2e-16 ***
y8     6.19494    0.01279  484.31   <2e-16 ***
y9     6.11078    0.01288  474.36   <2e-16 ***
y10    6.34411    0.01298  488.94   <2e-16 ***
y11    6.35237    0.01307  486.00   <2e-16 ***
y12    7.12114    0.01317  540.79   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0253 on 47 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 3.199e+05 on 13 and 47 DF,  p-value: < 2.2e-16

        Durbin-Watson test

data:  reg
DW = 0.82642, p-value = 2.39e-06
alternative hypothesis: true autocorrelation is greater than 0
```

다항추세모형 - 선형계절추세모형

- (예제) 백화점 월별 매출액 – 잔차분석 (지시함수)



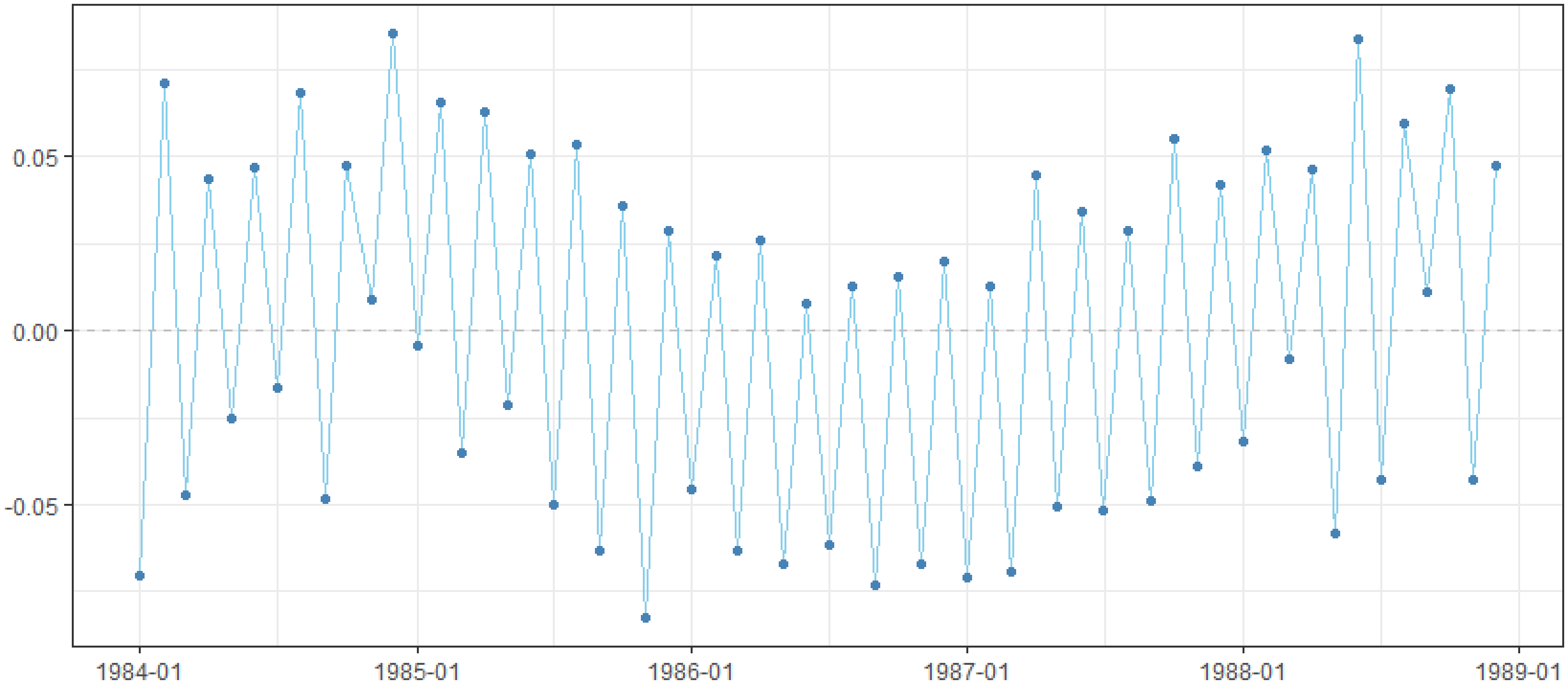
다항추세모형 - 선형계절추세모형

■ (예제) 백화점 월별 매출액 – 중회귀분석 (삼각함수)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.3237250  0.0148062 427.100  < 2e-16 ***
t            0.0107376  0.0004242  25.315  < 2e-16 ***
sin_12      -0.0277129  0.0103066  -2.689 0.009829 ** 
sin_6       -0.0382551  0.0102107  -3.747 0.000481 ***
sin_4       -0.1555546  0.0101931 -15.261  < 2e-16 ***
sin_3        0.0666506  0.0101872   6.543 3.70e-08 ***
sin_2.4      0.0128922  0.0101849   1.266 0.211691    
cos_12       0.0857900  0.0101931   8.416 5.21e-11 ***
cos_6        0.1675743  0.0101931  16.440  < 2e-16 ***
cos_4        0.1592698  0.0101931  15.625  < 2e-16 ***
cos_3        0.1267107  0.0101931  12.431  < 2e-16 ***
cos_2.4      0.2000603  0.0101931  19.627  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05578 on 48 degrees of freedom
Multiple R-squared:  0.9796,	Adjusted R-squared:  0.9749 
F-statistic: 209.5 on 11 and 48 DF,  p-value: < 2.2e-16

	Durbin-Watson test

data:  reg_2
DW = 3.2703, p-value = 1
alternative hypothesis: true autocorrelation is greater than 0
```

다항추세모형 - 선형계절추세모형

■ (예제) 백화점 월별 매출액 – 잔차분석 (삼각함수)



비선형 추세 모형(Nonlinear Trend Model)

모델

- S-Curve : $Z_t = \exp(\beta_0 + \beta_1 X_t)\,\varepsilon_t$

  => $\ln Z_t = \beta_0 + \beta_1 X_t + \ln \varepsilon_t$

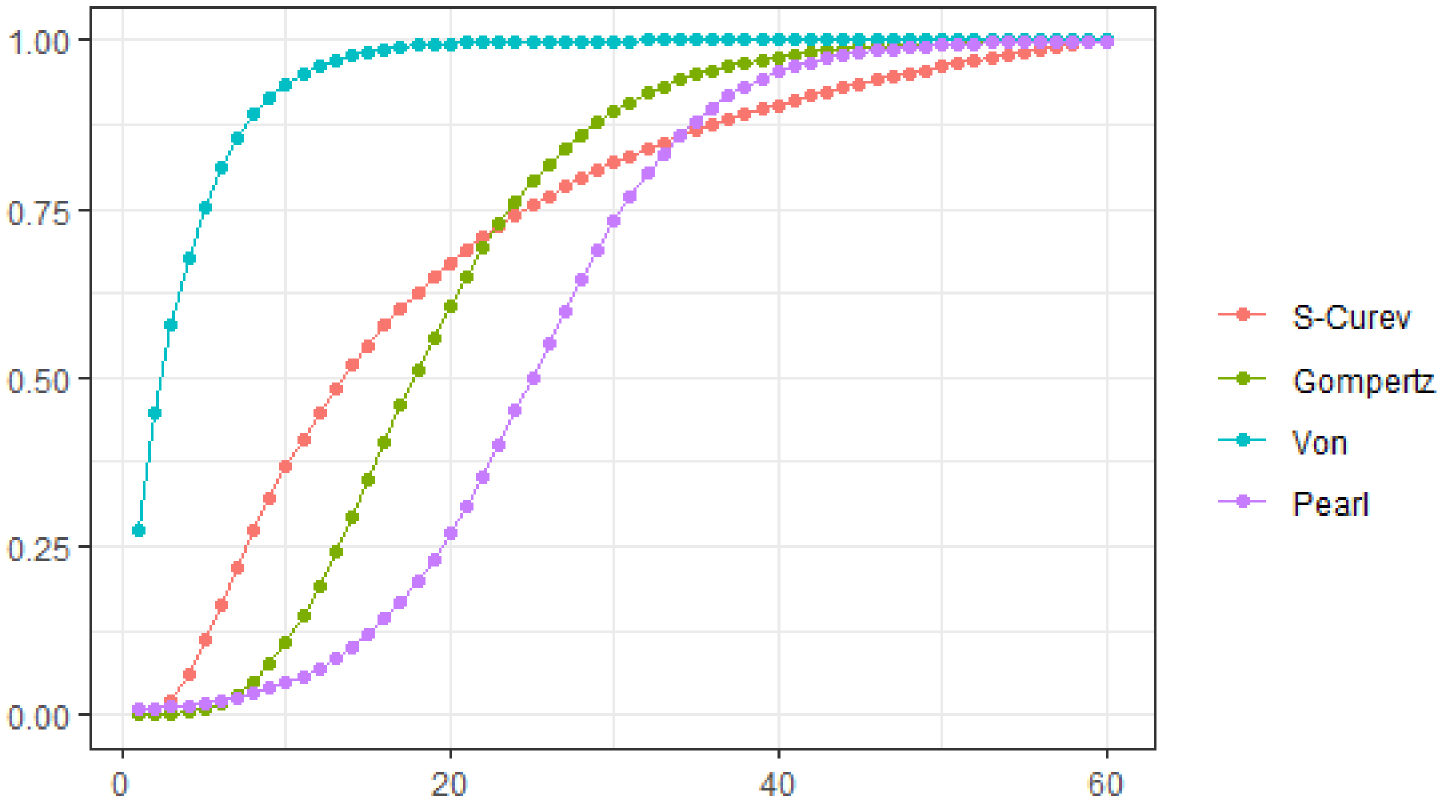
- Gompertz curve : $E(Z_t) = k\exp(-\beta_0\exp(-\beta_1 t))$

- Von Bertalautiff curve : $E(Z_t) = [1 - \beta_0\exp(-\beta_1 t)]^3$

- Pearl curve (or Logistic curve) : $E(Z_t) = \frac{k}{1+\exp(\beta_0+\beta_1 t)}$

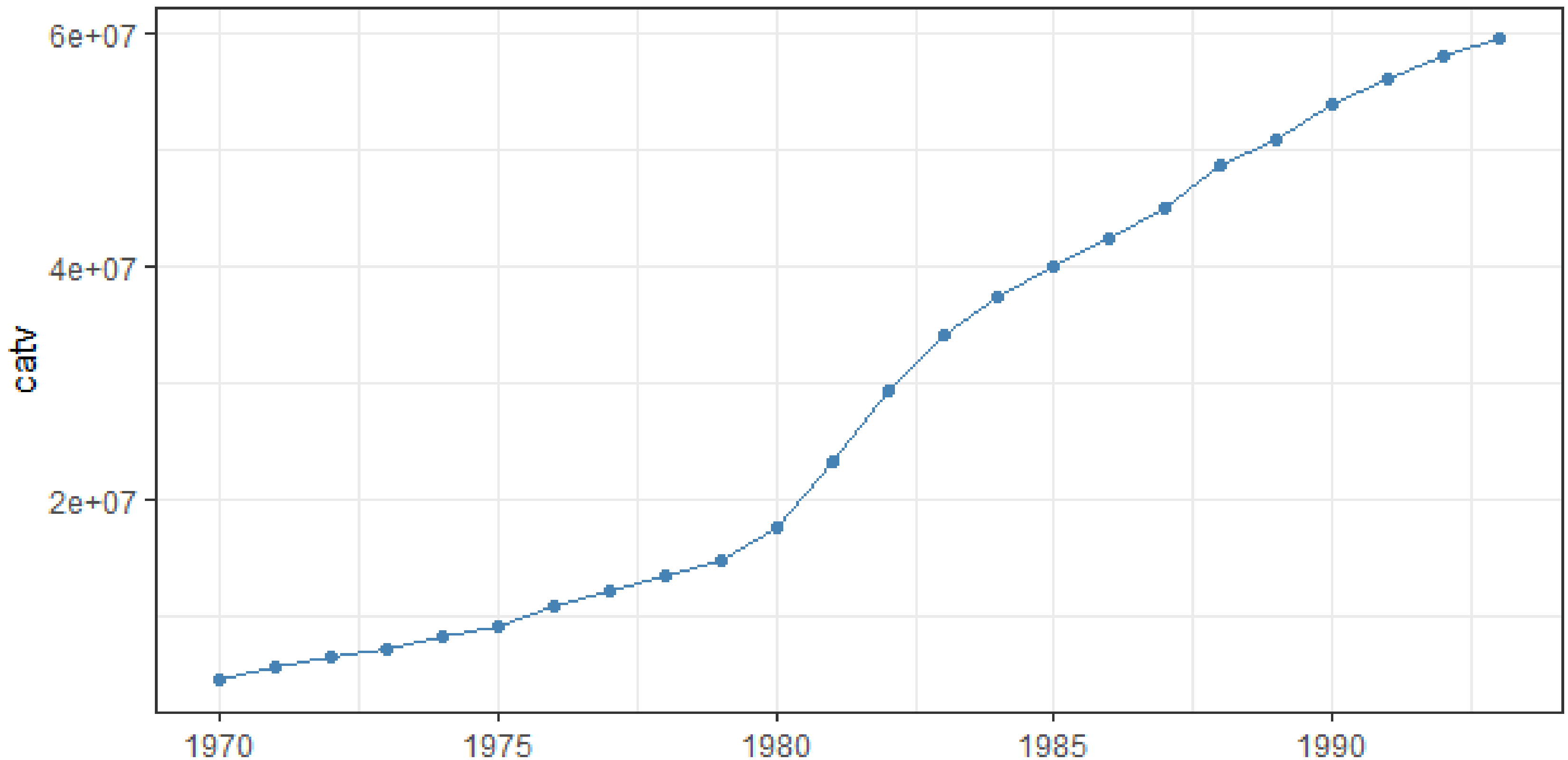
비선형 추세 모형(Nonlinear Trend Model)

■ 모델



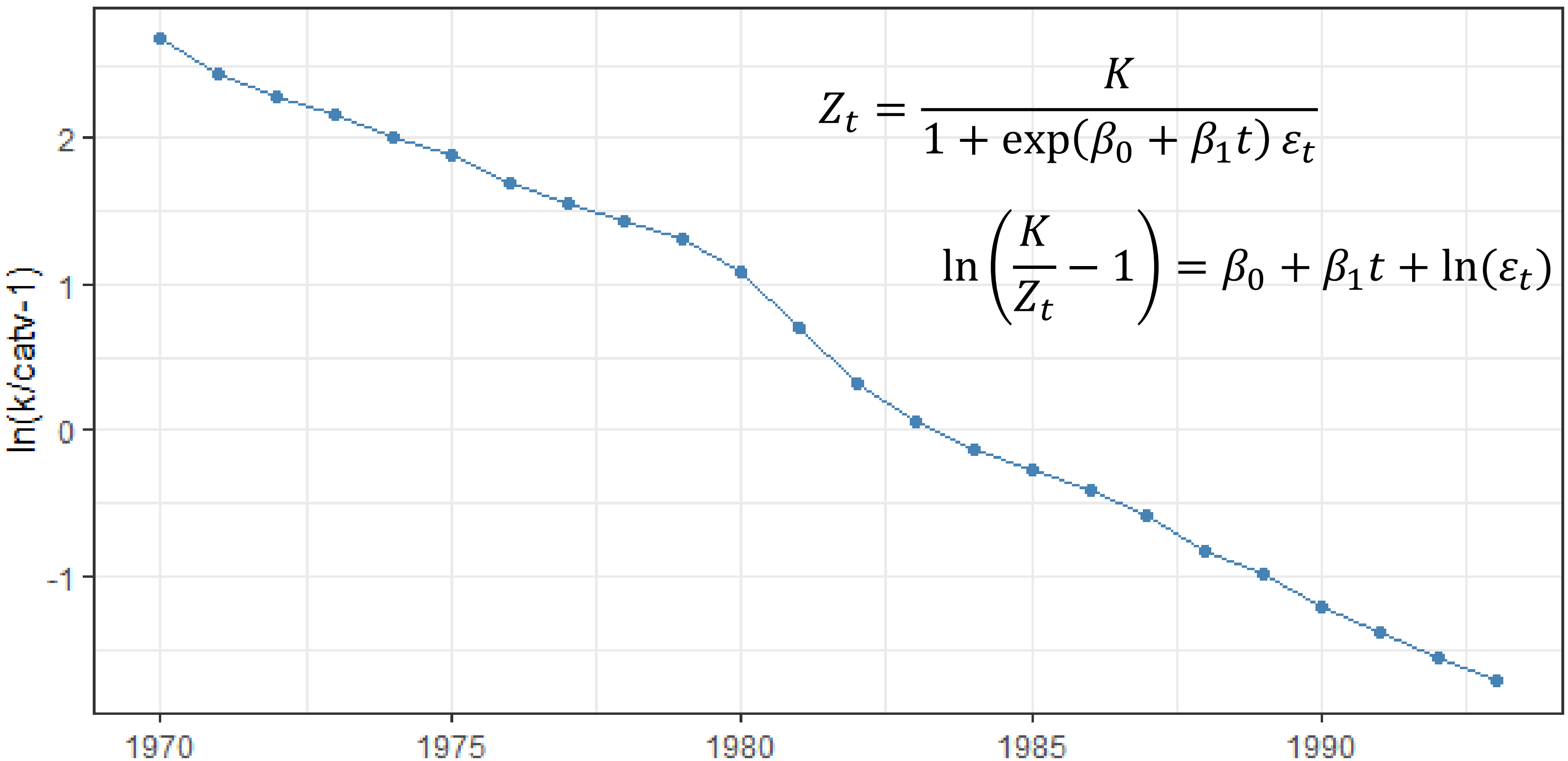
비선형 추세 모형 - 예제

- 예제 : 미국 CATV 누적 가입자 수



비선형 추세 모형 - 예제

- 예제 : 미국 CATV 누적 가입자 수 – 변수 변환(로지스틱)



$$Z_t = \frac{K}{1 + \exp(\beta_0 + \beta_1 t)\,\varepsilon_t}$$

$$\ln\left(\frac{K}{Z_t} - 1\right) = \beta_0 + \beta_1 t + \ln(\varepsilon_t)$$

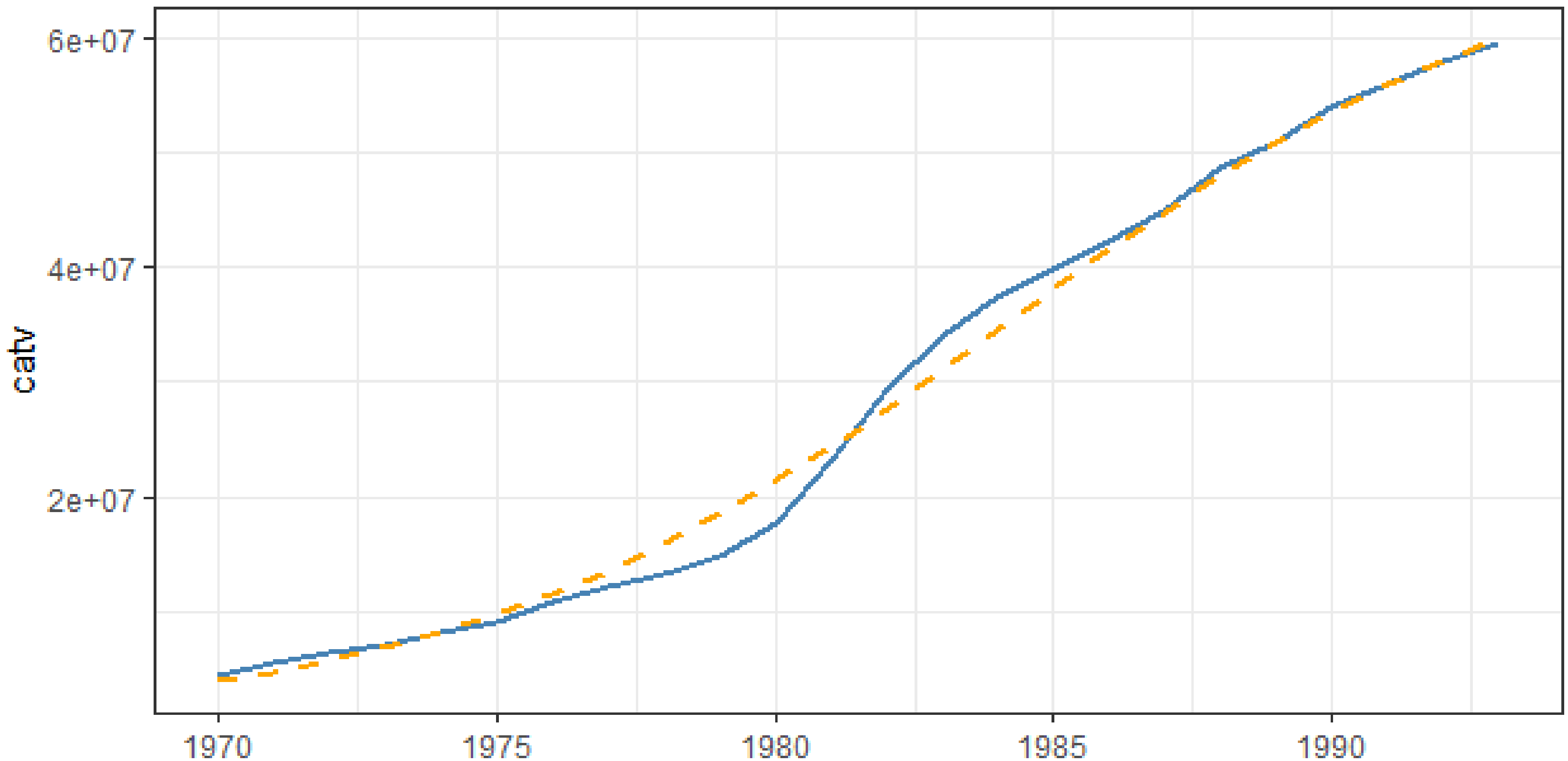
비선형 추세 모형 - 예제

- **예제 : 미국 CATV 누적 가입자 수 – 추세분석 결과**

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.17388 -0.09974 -0.01448  0.07133  0.29312

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.015622   0.056269   53.59   <2e-16 ***
t           -0.199364   0.003938  -50.63   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1335 on 22 degrees of freedom
Multiple R-squared:  0.9915,	Adjusted R-squared:  0.9911
F-statistic:  2563 on 1 and 22 DF,  p-value: < 2.2e-16


	Durbin-Watson test

data:  fit
DW = 0.3014, p-value = 8.43e-10
alternative hypothesis: true autocorrelation is greater than 0
```

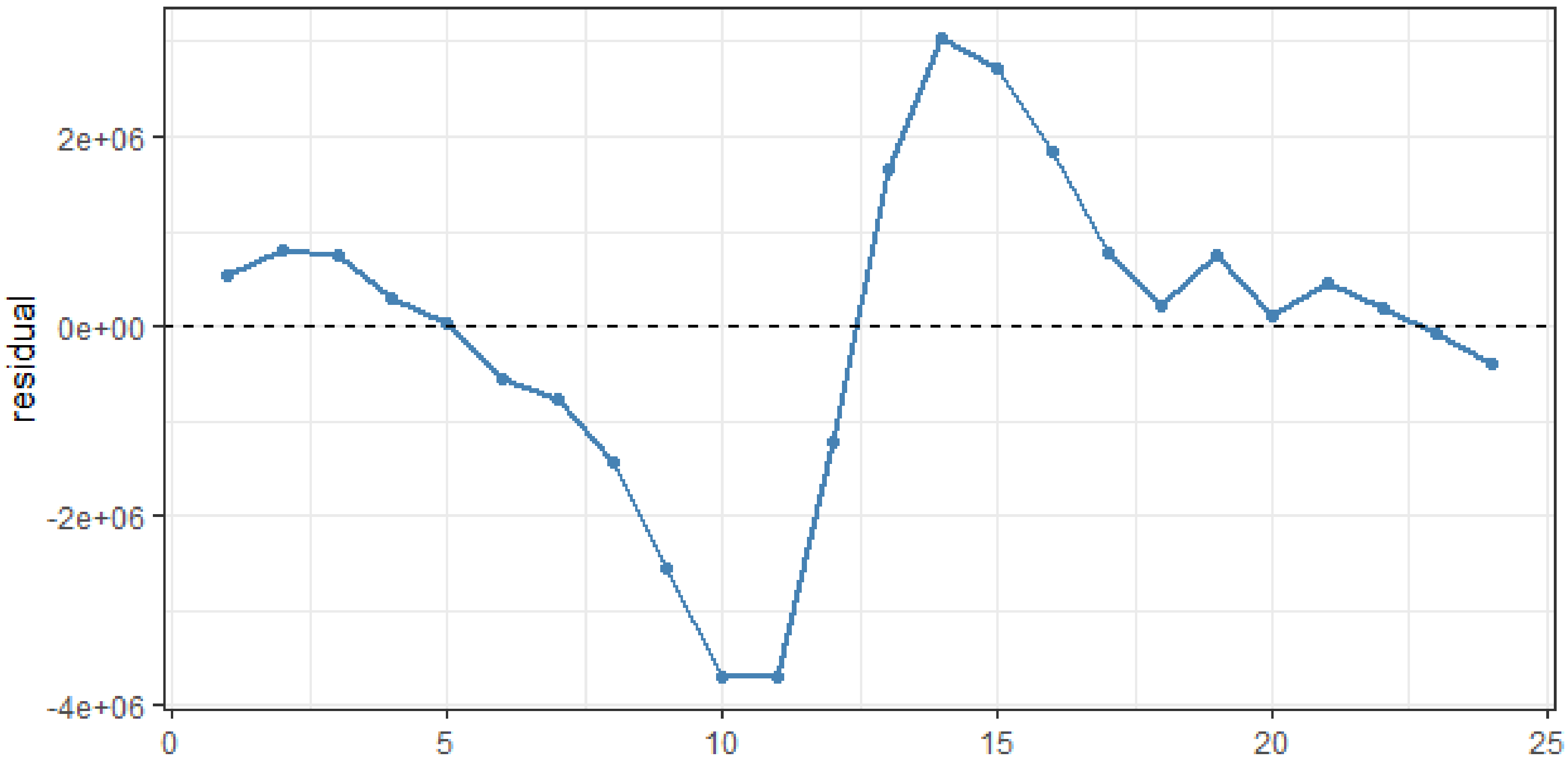
비선형 추세 모형 - 예제

- 예제 : 미국 CATV 누적 가입자 수 – 실제값 vs. 예측값



비선형 추세 모형 - 예제

■ **예제 : 미국 CATV 누적 가입자 수 - 잔차그림**



End of Document