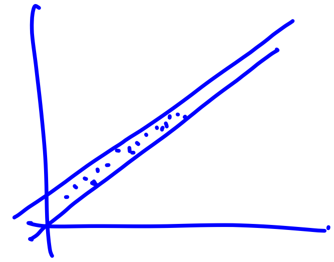
635



추세분석 : 회귀모형

2023. 3. 14

시계열 자료의 전통적 분석 방법

"모수"

관측값. 오차.

t에 대한 다항식.

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t$$

$$p = 0 \Rightarrow Z_t = \beta_0 + \varepsilon_t \quad \textbf{상수평균모형}$$

$$p = 1 \Rightarrow Z_t = \beta_0 + \beta_1 t + \varepsilon_t \quad \textbf{선형 추세모형}$$

$$p = 2 \Rightarrow Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t \quad \textbf{2차 추세모형}$$

$$\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \varepsilon_t$$

t $\quad$ t^2

2^x $\quad$ 3^x

비선형 추세모형

$$Z_t = \varepsilon_t \exp\left(\beta_0 + \beta_1 t\right)$$

모수.

e $\quad$ e^x $\quad$ exp( )

기하급수적 증가 양상.

비선형성

"성장곡선"

$$Z_t = \varepsilon_t e^{\beta_0 + \beta_1 t} = \varepsilon_t e^{\beta_0} e^{\beta_1 t}$$

ln $\quad$ log $\quad$ $\log_e$

회귀모형

두 개 이상의 변수들 $Z, T_1, T_2, \cdots, X_p$ 사이의 상호관련성을 다음과 같은 꼴로 표현한 것.

모수. $\beta = \begin{cases} (\beta_1, \cdots, \beta_p) \text{ or} \\ (\beta_0, \beta_1, \cdots, \beta_p). \end{cases}$

$$Z_t = f(X_t; \beta) + \varepsilon_t \quad (t = 1, 2, \cdots, p)$$

종속변수 (반응변수)

독립변수 p개. (설명변수)

t : 시간. 관측된 순서.

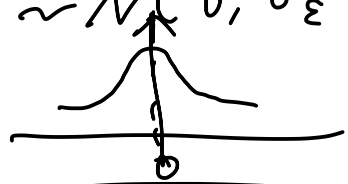
ε_t : 오차.

→ 가정들

평균 0. 분산 σ_ε^2 $(E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma_\varepsilon^2)$

$Cov(\varepsilon_{t_1}, \varepsilon_{t_2}) = 0 \ (t_1 \neq t_2)$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$



⇒ ε_t : 서로 독립.
각자 $N(0, \sigma_\varepsilon^2)$.

⇒ ε_t : "확률오차" random error

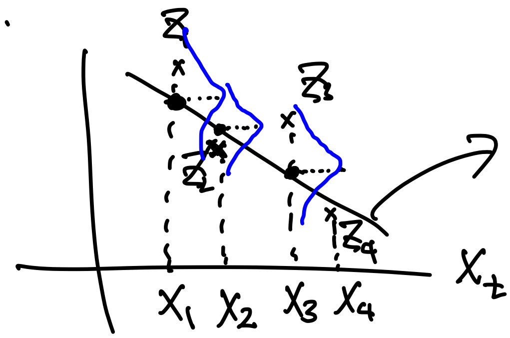
확률오차 ε_t

회귀모형은 확률모형

설명변수 ~~$X_{t1}, \dots, X_{tp}$~~ 가 주어지면 Z_t들은 서로 독립이고 평균이 $f(X_t;\beta)$, 분산이 σ_ε^2 인 정규분포.

$X_{t1}, \dots, X_{tp}$

$$\boxed{Z_t = f(X_t;\beta) + \varepsilon_t}$$



$f(X_t;\beta)$

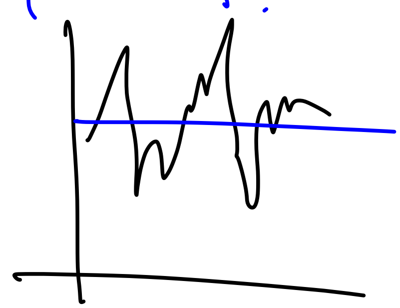
$$X_1 \longmapsto Z_1 = f(X_1;\beta) + \varepsilon_1$$

$$\sim N(\underbrace{f(X_1;\beta)}_{E(Z_1|X_1)}, \sigma_\varepsilon^2)$$

Outlier
이상점

이 값 중심으로 확률 오차 ε_t 만큼 차이로 랜덤하게 산포되어 있다.

$$\boxed{Z_t = \beta_0 + \varepsilon_t}$$



$$Z_t = \beta_0 + \beta_1 X_t + \varepsilon_t.$$

다중 선형 회귀 모형

$$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_p X_{tp} + \varepsilon_t.$$

$$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t1}^2 + \beta_4 X_{t2}^2 + \beta_5 X_{t1} X_{t2}$$

회귀모형은 확률모형

설명변수 $X_t1, \cdots, X_tp$ 가 주어지면 Z_t들은 서로 독립이고 평균이 $f(X_t;\beta)$, 분산이 σ_ε^2 인 정규분포.

$$Z_t = \varepsilon_t \exp(\beta_0 + \beta_1 X_t)$$

$$\Rightarrow \ln Z_t = \ln(\varepsilon_t e^{\beta_0} e^{\beta_1 X_t})$$

$\log_e = \ln$

$$= \beta_0 + \beta_1 X_t + \ln(\varepsilon_t)$$

$$\boxed{Z_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t^*}$$

변환 후 선형화!

2^x 3^x $\boxed{e}$

$$\boxed{e^x} = \exp(x) \qquad e^{f(x)} = \exp(f(x))$$

$$e^{2x} = \exp(2x)$$

<기초지식> $\sum_{n=1}^{N} a_n = a_1 + a_2 + a_3 + \cdots + a_N$

$$\sum_{n=1}^{10} n = 1+2+3+4+\cdots+10$$

$$\sum_{n=1}^{10} n^2 = 1^2+2^2+3^2+\cdots+10^2$$

$$\left(\sum_{n=1}^{3} a_n\right)^2 = (a_1+a_2+a_3)^2$$
$$= a_1^2+a_2^2+a_3^2$$
$$+2a_1a_2+2a_1a_3+2a_2a_3$$

$$\sum_{n=1}^{N}(a_n \pm b_n) = \sum_{n=1}^{N} a_n \pm \sum_{n=1}^{N} b_n$$

$$\sum_{n=1}^{N}(a_n \pm b_k) = \sum_{n=1}^{N} a_n \pm \underbrace{\sum_{n=1}^{N} b_k}_{= N b_k}$$

ex) $\sum_{n=1}^{10} 1 = 1+1+\cdots+1 = 10.$

$$\boxed{\sum_{n=1}^{N} c = Nc}$$

$$\left(\sum_{n=1}^{N} a_n\right)\left(\sum_{n=1}^{N} b_n\right) \neq \sum_{n=1}^{N} a_n b_n$$

$$\sum_{i,j=1}^{N} a_i b_j = a_1 b_1 + a_1 b_2 + \cdots + a_1 b_N$$
$$+ a_2 b_1 + a_2 b_2 + \cdots + a_2 b_N$$
$$+ \cdots$$
$$+ a_N b_1 + a_N b_2 + \cdots + a_N b_N$$
$$= \left(\sum_{n=1}^{N} a_n\right)\left(\sum_{n=1}^{N} b_n\right)$$

평균 : $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} = E(X) = E(x)$.

$$\frac{1}{n}\sum_{t=1}^{n} z_t = \bar{z}$$

분산 : $Var(X)$ or $Var(x) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^2$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$$

$$\boxed{E(X^2) - E(X)^2}$$

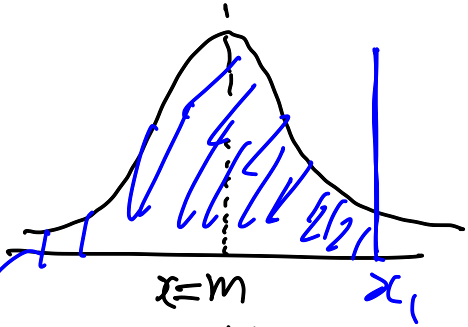
$N(m, \sigma^2)$

$X \sim N(m, \sigma^2)$

σ: 표준편차

m: 평균

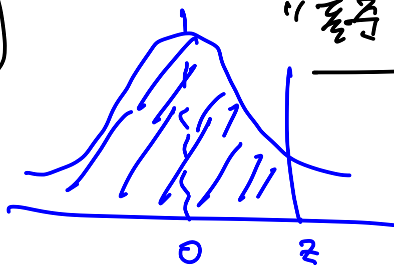
표준편차^2 = 분산.



평균

$\boxed{Z = \frac{X-m}{\sigma}} \Rightarrow Z \sim N(0, 1^2)$

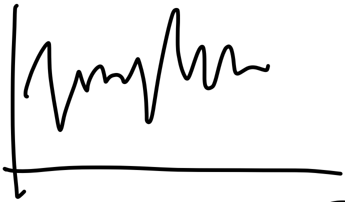
"표준 정규분포"

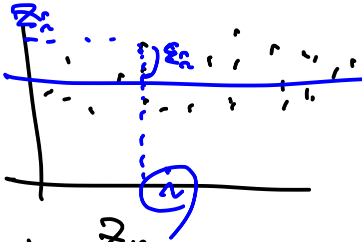


$P(X \le x_1) = P(Z \le z)$

$z = \frac{x_1 - m}{\sigma}$

<최소제곱법> → 행렬/편미분





$$\boxed{Z_t = \beta_0 + \epsilon_t}$$

$$f(X_t;\beta) = \beta_0$$

$Z_1, \dots, Z_n$

$Z_i = \beta_0 + \epsilon_i$

→ $\epsilon_i = Z_i - \beta_0$

$$S(\beta_0) = \sum_{i=1}^{n} (Z_i - \beta_0)^2$$ ⇒ 최소가 되게하는 $\beta_0 = \hat{\beta}_0$ 찾으려!

$$= \sum_{i=1}^{n} (\beta_0^2 - 2\beta_0 Z_i + Z_i^2)$$

$$\frac{dS}{d\beta_0} = \sum_{i=1}^{n} (2\beta_0 - 2Z_i)$$

$$= 2 \sum_{i=1}^{n} (\beta_0 - Z_i)$$

$$= 2\left(\sum_{i=1}^{n} \beta_0 - \sum_{i=1}^{n} Z_i\right)$$

$$= 2\left(n\beta_0 - \sum_{i=1}^{n} Z_i\right)$$

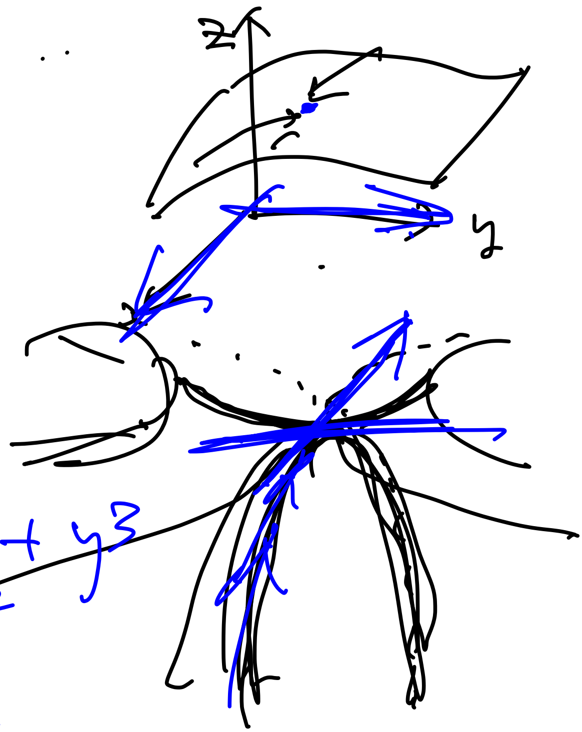
$$= 2n\left(\beta_0 - \frac{1}{n}\sum_{i=1}^{n} Z_i\right)$$

$$= 2n(\beta_0 - \bar{Z}) = 0$$

$\beta_0 = \hat{\beta}_0$

$$\boxed{\hat{\beta}_0 = \bar{Z}}$$

$f(x,y) = x^2 + y^2$



$$\frac{\partial f}{\partial x} = 2x$$

$$\frac{\partial f}{\partial y} = 2y$$

$$f(x,y) = x^3 + 2x^2y + y^3$$

$(2y)x^2$

$$\frac{\partial f}{\partial x} = 3x^2 + 4xy$$

$$\frac{\partial f}{\partial y} = 0 + 2x^2 + 3y^2$$

$$f(x,y) = x^2y - y^3 + 2xy$$

$$\frac{\partial f}{\partial x} = y \times 2x + 0 + 2y = 2xy + 2y$$

$$\frac{\partial f}{\partial y} = x^2 - 3y^2 + 2x$$

$\beta_0 + \beta_1 x_i$