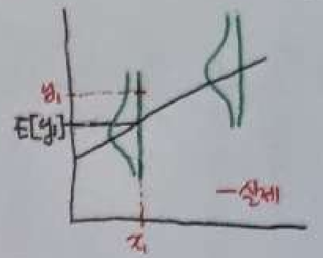
단순 선형 회귀 모형

독립변수 1개 $y = f(x)$



$$\underline{y_i} = \underline{\alpha + \beta x_i} + \underline{\varepsilon_i}$$ = 상수 + 확률변수 = 확률변수

$E[y_i] = \alpha + \beta x_i$: 관찰불가능한 실제선

$\varepsilon_i \sim N[0, \sigma^2]$: (가정) $\varepsilon_1\ \varepsilon_2\ \varepsilon_3 \cdots$ 서로 완전히 무관(독립) / 등분산 / 정규, 독립이 아닌경우 시계열 분석으로 GoGo

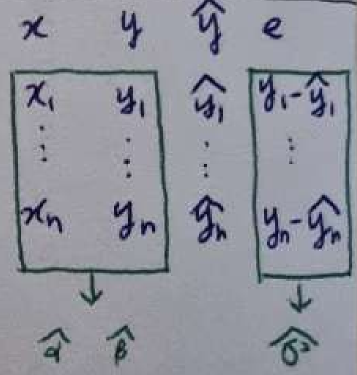
$y_i \sim N[\alpha + \beta x_i, \sigma^2]$

α (절편계수), β (기울기 계수) : 알려지지 않은 모수, 고정된 상수, 회귀계수

σ^2 : 모수

x_i : 주어진 값

$$y_i = \hat{\alpha} + \hat{\beta} x_i + e_i$$



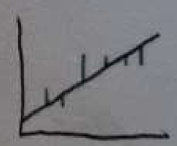
$\widehat{E[y_i]} = \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$: 샘플 자료에 의해 추정된 선

e_i(잔차) : ε_i (오차) 가 관찰불가능 하므로 사용하는 것

$e_i = y_i - \widehat{E[y_i]} = y_i - \hat{y}_i$

$\varepsilon_i = y_i - E[y_i]$

1. α, β 추정 : LSE (Least Square Estimation, 최소제곱추정)
   - (x_i, y_i) n쌍을 보고 추정
   - 수직거리 제곱합인 $SS(\hat{\alpha}, \hat{\beta}) = \Sigma e_i^2 = \Sigma (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$ 이 최소가 되게하는 $\hat{\alpha}, \hat{\beta}$ 찾기



   - 이상치가 있을 경우 적합하지 않음!

2. ε_i의 분산인 σ^2 추정 : e_i의 변동성 이용
   - $\frac{SSE}{n-2} = MSE = \hat{\sigma}^2 = \frac{\Sigma e_i^2}{n-2}$

$\frac{SS\square}{df} = MS\square = S^2$

   - $E[MSE] = \sigma^2$

1

양쪽꼬리 검정

$\hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{S_{xx}}\right]$ $S_{xx} = \sum(x_i - \bar{x})^2$: 넓은범위의 x를 관찰하면 β를 더 잘추정함

$$Z = \frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N[0,1] \quad \oplus \quad \frac{SSE}{\sigma^2} = \frac{(n-2)MSE}{\sigma^2} = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi[n-2]$$

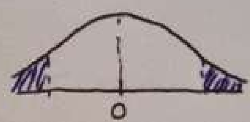
$$\parallel$$

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} = \frac{\hat{\beta}-\beta}{\widehat{S.E.[\hat{\beta}]}} \sim t[n-2]$$

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$: 모형이 유의하다

under H_0,



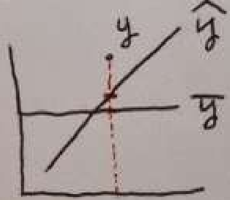
p-value(유의확률) = $P(T > |t_0|) < \alpha$(유의수준) 이면 여기에 속하므로 H_0 기각

2

오른꼬리 검정

$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$

= 모형설명가능 + 모형설명불가능

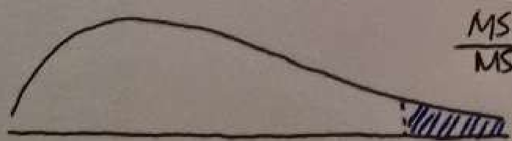


$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$

	SST	=	SSR	+	SSE	... SSR이 크면클수록 모형이 유의함
(자유도)	$(n-1)$		(1)		$(n-2)$	

$H_0 : \beta = 0$ ... $SSR \approx 0$, $E[MSR] = E\left[\frac{SSR}{1}\right] = \sigma^2$, $MSR = MSE$

$H_1 : \beta \neq 0$ ... $SSR \gg 0$, $E[MSR] \gg \sigma^2$, $MSR \gg MSE$ $\because E[MSE] = \sigma^2$



$\frac{MSR}{MSE} \sim F[1, n-2]$

3

적합도 검토

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

$0 \leq R^2 \leq 1$

0 → SST=SSE

1 → SST=SSR

1에 가까울수록 추정된 모형의 적합도 높음