Week 2

1. 파라미터 추정.

Estimator 추정량 : a function of samples → 알려지지 않은 파라미터 추정

- 점추정 $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \sum (X_i - \bar{X})(Y_i - \bar{Y}) / \sum (X_i - \bar{X})^2$ $\hat{\sigma}^2 = \left(\frac{1}{n-2}\right) \sum e_i^2$

<mark>BLUE | Best Linear Unbiased Estimator</mark>
  → 불편추정량 & 최소 추정량간의 최소의 분산.

- 구간추정 : $\hat{\theta} - \alpha \cdot std(\hat{\theta}) \leq \theta \leq \hat{\theta} + \alpha \cdot std(\hat{\theta})$

  $\hat{\beta}_0 - t_{\alpha/2, n-2} \cdot sd(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} \cdot sd(\hat{\beta}_0)$

  $\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot sd(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot sd(\hat{\beta}_1)$

2. 가설검정

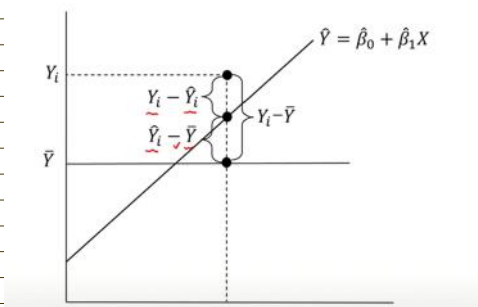β_1 : $H_0 : \beta_1 = 0$ / $H_1 : \beta_1 \neq 0$ $p val = 2P(T > |t'|)$, $T \sim t(n-2)$ <mark>[귀무가설기각시 유의수준]</mark>

  $\quad t' = \hat{\beta}_1 - 0 / sd(\hat{\beta}_1)$

  $\alpha = 0.05$ 하 귀무.

3. 결정계수 R^2

SSE $= \sum (Y_i - \hat{Y}_i)^2$ X로 설명못하는 부분.

SSR $= \sum (\hat{Y}_i - \bar{Y})^2$ X로 의영의 설명

SST $= \sum (Y_i - \bar{Y})^2$ Y들의 변동

SST = SSE + SSR $\dfrac{SSR}{SST} = R^2 \ (0 \leq R^2 \leq 1)$

  → 현재 X가 Y의 반응을 크게나 설명하는가

$R^2 = 1 - \dfrac{SSE}{SST}$ → 유의하지 않은 변수가 추가되면 증가.

$adj. R^2 = 1 - \left[\dfrac{n-1}{n-(p+1)} \right] \dfrac{SSE}{SST}$

  → 유의하지 않은 변수 추가되는 경우 R^2 커짐 방지.

4. 분산분석 ANOVA

$\dfrac{SSE}{SSR} \begin{cases} >1 & → X가 Y 설명가능, 기울기 \neq 0 \\ \\ 0 \leq \ \leq 1 & → X변화가 Y 설명X, 기울기 = 0 \end{cases}$ → 1 보다 커야 큰관인가?

$SSR \sim \chi^2_{(1)}$ $SSE \sim \chi^2_{(n-2)}$ $F^* = \dfrac{SSR/1}{SSE/n-2} \sim F(1,n-2)$

Source	DF	SS	MS	F	P
Model	1	SSR	MSR	F*	P-value
Error	$n-2$	SSE	MSE		
Total	$n-1$	SST			

$MSR = \dfrac{SSR}{1}$

$MSE = \dfrac{SSE}{n-2}$

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$

$F^* = \dfrac{MSR}{MSE} \rightarrow$ $pval = P(Y \geq F^*)$, $Y \sim F(1, n-2)$

==다중회귀 검정==

$H_0 : \beta_1 = \beta_2 \cdots \beta_p = 0$ $H_1 : $ 적어도 하나의 $\beta \neq 0$

$pval$ 유의 $\rightarrow$ 적어도 하나의 $\beta \neq 0$ $\sim$ ==회귀 직선의 유의성 검정==