Regression Analysis : HW03

CH 04

1. (1) β_0의 90% 신뢰구간을 구하시오

$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$ 에서 $MSE = 0.11$, $S_{xx} = 3.26$, $\bar{x} = 1.6$, $n = 9$, $\hat{\beta}_0 = -0.11$ 이므로,

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{0.11 \times \left(\frac{1}{9} + \frac{1.6^2}{3.26}\right)} = 0.314.$$

즉, 귀무가설 $H_0 : \beta_0 = 0$ 하에서 $\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t(7)$ 이므로,

$$P\left(-t_{0.05}(7) < \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} < t_{0.05}(7)\right) = P\left(\hat{\beta}_0 - t_{0.05}(7)\cdot\hat{\sigma}_{\hat{\beta}_0} < \beta_0 < \hat{\beta}_0 + t_{0.05}(7)\cdot\hat{\sigma}_{\hat{\beta}_0}\right) = 0.9$$

∴ β_0의 90% CI : $(-0.705,\ 0.485)$.

(2) 다음의 가설 검정을 수행하시오.

$$H_0 : \beta_1 = 1 \quad vs. \quad H_1 : \beta_1 > 1.$$

i) $\alpha = 0.05$로 설정

ii) $T_0 = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(7)$ as H_0

iii) $t_0 > t_{0.05}(7) = 1.895$ 이면 귀무가설은 기각 가능 (단측 검정)

iv) $t_0 = \frac{2.16 - 1}{0.18} = 6.44 > t_{0.05}(7)$ 이므로 유의수준 $\alpha = 0.05$에서 귀무가설은 기각하고 대립가설을 수용한다 즉, β_1은 1보다 크다.

(3) 무게가 3,000 kg이 되는 차량의 평균 에너지 소모량을 예측하시오. 이것은 무게가 1,000 kg이 되는 차량의 에너지 소모량의 몇 배인가?

i) $\widehat{E(Y|X=3)} = -0.11 + 2.16 \times 3 = 6.37.$

ii) $\widehat{E(Y|X=1)} = -0.11 + 2.16 \times 1 = 2.05.$

∴ 대략 3.1배를 소모한다고 예측할 수 있다.

1. (4) 무게가 3.000 kg이 되는 차량의 평균 에너지 소모량과 하나의 개별 y값의 90% 신뢰구간을 각각 구하시오.

i) $\mu_0 = E(Y|x_0)$, $\hat{\mu}_0 = \widehat{E(Y|x_0)} = \hat{\beta}_0 + \hat{\beta}_1 x_0$, $\widehat{Var}(\hat{\mu}_0) = MSE\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)$.

$x_0 = 3$일 때, $\hat{\mu}_0 = 6.37$, $\widehat{Var}(\hat{\mu}_0) = 0.11 \times \left(\frac{1}{9} + \frac{(3-1.6)^2}{3.26}\right)$, $\widehat{s.e.}(\hat{\mu}_0) = \sqrt{\widehat{Var}(\hat{\mu}_0)} \doteqdot 0.28$

$\therefore$ μ_0의 90% CI : $6.37 \pm t_{0.05}(7) \cdot 0.28 = (5.839,\ 6.901)$ (Since $t_{0.05}(7) = 1.895$).

ii) $y_0 = \beta_0 + \beta_1 x_0 + \epsilon_0$, $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$, $\widehat{Var}(\hat{y}_0) = MSE\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)$.

$x_0 = 3$일 때, $\hat{y}_0 = 6.37$, $\widehat{Var}(\hat{y}_0) = 0.11 \times \left(1 + \frac{1}{9} + \frac{(3-1.6)^2}{3.26}\right)$, $\widehat{s.e.}(\hat{y}_0) = \sqrt{\widehat{Var}(\hat{y}_0)} \doteqdot 0.434$.

$\therefore$ y_0의 90% CI : $6.37 \pm t_{0.05}(7) \cdot 0.434 = (5.548,\ 7.192)$.

(5) 원점을 지나는 회귀직선에서 회귀계수(기울기)에 대한 90% 신뢰구간을 구하시오.

적합된 원점을 지나는 회귀직선 : $\hat{y} = 2.10x$.

$\hat{\beta}_1 = 2.10$, $\widehat{Var}(\hat{\beta}_1) = \frac{MSE}{S_{xx}} = \frac{0.096}{3.26}$, $\widehat{s.e.}(\hat{\beta}_1) = 0.172$. (Since SST = 116.67, SSR = 115.9, df = 8).

$\therefore$ 원점을 지나는 회귀직선의 회귀계수 β_1의 90% CI : $2.10 \pm t_{0.05}(8) \cdot 0.172 = (1.78,\ 2.42)$.

2. 기울기 β_1의 $100(1-\alpha)$% 신뢰구간이 0을 그 구간 속에 포함하고 있으면, 가설검정 $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$ 에서 유의수준 α로 귀무가설이 채택되고, 만약 신뢰구간이 0을 포함하고 있지 않으면, 대립가설이 채택된다. 이것은 옳은 주장인가?

→ 주장은 옳은 주장이다. 신뢰구간에 0이 포함되었다는 것은 아래와 같다.

$\hat{\beta}_1 - t_{\frac{\alpha}{2}}(n-2) \cdot \widehat{s.e.}(\hat{\beta}_1) < 0 < \hat{\beta}_1 + t_{\frac{\alpha}{2}}(n-2) \cdot \widehat{s.e.}(\hat{\beta}_1)$. 해당 부등식을 다르게 표현하면,

$-t_{\frac{\alpha}{2}}(n-2) < \frac{\hat{\beta}_1}{\widehat{s.e.}(\hat{\beta}_1)} < t_{\frac{\alpha}{2}}(n-2) \Rightarrow \left|\frac{\hat{\beta}_1 - \beta_1}{\widehat{s.e.}(\hat{\beta}_1)}\right| < t_{\frac{\alpha}{2}}(n-2)$ ($\because \beta_1 = 0$ as H_0) 으로, 가설검정에서 귀무가설을 기각하지 못하는 조건과 동일하다.

반대로, 신뢰구간에 0이 포함되어 있지 않다는 것도 비슷하게,

$\hat{\beta}_1 - t_{\frac{\alpha}{2}}(n-2) \cdot \widehat{s.e.}(\hat{\beta}_1) > 0$ or $\hat{\beta}_1 + t_{\frac{\alpha}{2}}(n-2) \cdot \widehat{s.e.}(\hat{\beta}_1) < 0 \Rightarrow \frac{\hat{\beta}_1}{\widehat{s.e.}(\hat{\beta}_1)} > t_{\frac{\alpha}{2}}(n-2)$ or $-\frac{\hat{\beta}_1}{\widehat{s.e.}(\hat{\beta}_1)} > t_{\frac{\alpha}{2}}(n-2)$

$\Rightarrow \left|\frac{\hat{\beta}_1}{\widehat{s.e.}(\hat{\beta}_1)}\right| > t_{\frac{\alpha}{2}}(n-2)$ 로서, 가설검정에서 귀무가설 기각 및 대립가설 수용의 조건과 동일하다.

따라서 문제의 주장은 옳다.

3. 단순 선형회귀모형에서 기울기 $\beta_1 = 0$ 에 대한 검정통계량은 다음과 같이 표본 크기 (n)와 표본 상관계수(r)의 함수임을 보이시오

$$t = \frac{\hat{\beta}_1}{\sqrt{MSE / S_{(xx)}}} = \sqrt{n-2}\,\frac{r}{\sqrt{1-r^2}}.$$

i) $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{r \cdot \sqrt{S_{yy}}}{\sqrt{S_{xx}}}.$

ii) $SSE = SST - SSR. = SST\left(\dfrac{SST}{SST} - \dfrac{SSR}{SST}\right) = SST(1-R^2)$

$\therefore MSE = \dfrac{SST(1-R^2)}{n-2.} = \dfrac{S_{yy}(1-R^2)}{n-2}.$

iii) $R^2 = \dfrac{SSR}{SST} = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \dfrac{\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \dfrac{\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. = \hat{\beta}_1^2 \dfrac{S_{xx}}{S_{yy}}.$

$r^2 = \dfrac{\{S_{xy}\}^2}{S_{xx}S_{yy}} = \dfrac{\{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}.$ $\quad \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x})^2.$

$\therefore r^2 = \dfrac{\{\hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x})^2\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2} = \hat{\beta}_1^2 \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \hat{\beta}_1^2 \dfrac{S_{xx}}{S_{yy}}$ 이므로, $R^2 = r^2.$

i) + ii) + iii) 에 따라 준식을 다시 쓰면,

$$t = \frac{\dfrac{r \cdot \sqrt{S_{yy}}}{\sqrt{S_{xx}}}}{\sqrt{\dfrac{\dfrac{S_{yy}(1-r^2)}{n-2}}{S_{xx}}}} = \sqrt{n-2} \cdot \frac{r}{\sqrt{1-r^2}}$$ 이다.