9장 보고서

보고서

R의 mtcars 중 자동 트랜스미션 차 19대를 이용하여, 단순회기모형을 이용하여, 마력(x)이 연비(y)를 어떻게 설명할 수 있는지 살펴보자. 이때 유의수준 0.05를 사용한다.

표 1의 분산분석표에서 검정통계량 F= 38.088에 대한 유의확률 이 유의수준

0.05보다 작으므로 , 또는 단순회기모형이 유의하지 않다를 기각한다. 즉, 추정된 단순회기모형이 적합하여 유의하다. 모형의 결정계수가 이므로, 마력은 연비의 총변동 중 69.14%를 설명한다.

표 1. 분산분석표

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 요인 | 제곱합  SS | 자유도  df | 평균제곱합  MS = SS/df | 검정통계량  F | 유의확률  P |
| 회귀 (마력hp) | 182.937 | 1 | 182.937 | F = 38.088 | 1.025e-05 |
| 잔차(residual) | 81.651 | 17 | 4.803 |  |  |
| 총합 | 264.588 | 18 |  |  |  |

표 2의 계수추정표로부터 구한 추정된 희귀직선식은 다음과 같다. (그림 1)

연비 = 26.62485 – 0.05914 x 마력



그림 1. 마력과 연비의 회기분석모형

여기서, 절편에 대한 유의확률 이 유의수준 0.05보다 작으므로, 절편은 유의하다.

기울기에 대한 유의확률 이 유의수준 0.05보다 작으므로, 기울기가 유의하다. 즉, 마력이 연비에 유의하게 영향을 미치며, 마력이 1증가하면, 연비가 0.05914 (마일/갤런) 감소한다.

잔차의 표준오차는 이다.

표 2 계수추정표

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 추정값  (Estimate) | 표준오차  (Std. Error) | 검정통계량  (t value) | 유의확률  Pr(>) |
| 절편(Intercept) | 26.624848 | 1.615883 | 16.477 | 6.92e-12 \*\*\* |
| 마력 | -0.059137 | 0.009582 | -6.172 | 1.02e-05 \*\*\* |

= Residual standard error: 2.192 on 17degrees of freedom

= Multiple R-squared: 0.6914, Adjusted R-squared: 0.6733

그림 2에서 잔차의 QQ-plot을 보면, 잔차가 대체로 정규분포를 벗어나지 않음을 볼 수 있다. 잔차에 대한 샤피로의 정규성검정에서 유의확률 p=0.129이므로, 단순회기모형의 정규성 가정이 성립함을 알 수 있다.

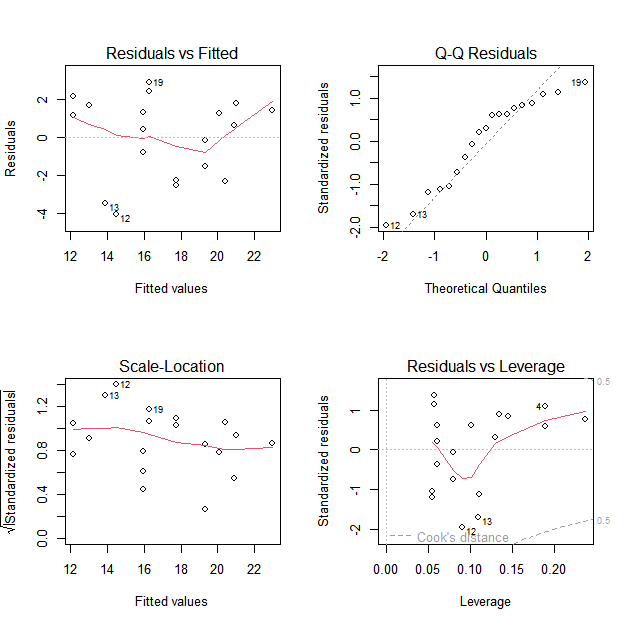


그림 2. 잔차도

|  |
| --- |
| #R 코드 결과  > auto <- subset(mtcars, am==0)  > x<- auto$hp  > y <- auto$mpg  >  > plot(x,y)  > fit <- lm(y~x)  > fit  Call:  lm(formula = y ~ x)  Coefficients:  (Intercept) x  26.62485 -0.05914  > abline(fit)  > anova(fit)  Analysis of Variance Table  Response: y  Df Sum Sq Mean Sq F value Pr(>F)  x 1 182.937 182.937 38.088 1.025e-05 \*\*\*  Residuals 17 81.651 4.803  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  > summary(fit)  Call:  lm(formula = y ~ x)  Residuals:  Min 1Q Median 3Q Max  -4.1018 -1.9026 0.6114 1.5592 2.9241  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 26.624848 1.615883 16.477 6.92e-12 \*\*\*  x -0.059137 0.009582 -6.172 1.02e-05 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 2.192 on 17 degrees of freedom  Multiple R-squared: 0.6914, Adjusted R-squared: 0.6733  F-statistic: 38.09 on 1 and 17 DF, p-value: 1.025e-05  >  > abline(fit)  > par(mfrow=c(2,2))  > plot(fit)  > shapiro.test(fit$resid)  Shapiro-Wilk normality test  data: fit$resid  W = 0.92307, p-value = 0.129 |