Faraway

Harry Woo

2020-5-12

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

## Problem 2.1

*The dataset teengamb concerns a study of teenage gambling in Britain. Fit a regression model with the expenditure on gambling as the response and the sex, status, income and verbal score as predictors. Present the output.*

tg\_lm <- lm(gamble ~ sex + status + income + verbal, data = teengamb)  
tg\_lms <- summary(tg\_lm)  
print(tg\_lms)

##   
## Call:  
## lm(formula = gamble ~ sex + status + income + verbal, data = teengamb)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -51.082 -11.320 -1.451 9.452 94.252   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 22.55565 17.19680 1.312 0.1968   
## sex -22.11833 8.21111 -2.694 0.0101 \*   
## status 0.05223 0.28111 0.186 0.8535   
## income 4.96198 1.02539 4.839 1.79e-05 \*\*\*  
## verbal -2.95949 2.17215 -1.362 0.1803   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22.69 on 42 degrees of freedom  
## Multiple R-squared: 0.5267, Adjusted R-squared: 0.4816   
## F-statistic: 11.69 on 4 and 42 DF, p-value: 1.815e-06

### (a) What percentage of variation in the response is explained by these predictors?

전체 제곱합(SST)에서 회귀 제곱합(SSR)이 설명하는 비중, 즉 모형의 설명력은 결정 계수 R2 이다. 위 Summary 에서와 같이 동 모형의 결정계수 **Multiple R-squared = 0.5267** 이다.

var\_ex <- data.frame(Var\_explained = tg\_lms$r.squared)  
var\_ex %>% gt() %>%   
 fmt\_percent(columns = vars(Var\_explained),  
 decimals = 2)

Var\_explained

52.67%

var\_ex %>% kable()

|  |
| --- |
| Var\_explained |
| 0.5267234 |

### (b) Which observation has the largest (positive) residual? Give the case number.

회귀모형의 residuals 를 데이터프레임으로 변환하여 잔차값 기준으로 내림차순 정렬을 시행해 largest residual의 case number를 추출한 결과, **해당 case number는 24** 이다.

res <- data.frame(case\_no = c(1:47), residual = tg\_lm$residuals)  
res %>%   
 arrange(desc(residual)) %>%   
 slice(1) %>%   
 gt()

case\_no

residual

24

94.25222

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.