Chapter 7 Workshop

Table of contents

Dataset `Prestige`

We will continue to use dataset `Prestige` from the `car` R package.

Exercise 7.1

Obtain the matrix plot of the numerical variables `education`, `income`, `women`, and `prestige`.

Obtain their correlation matrix.

Fit a (full) multiple regression of `prestige` on `education`, `income`, & `women`.

Obtain the plots for residual diagnostics.

```r
library(car)
library(tidyverse)
library(GGally)

Prestige |>
  select(prestige, education, income, women) |>
  ggpairs(aes(colour=Prestige$type))

# Old style pairs plot
Prestige |>
  select(prestige, education, income, women) |>
  pairs()

Prestige |>
  select(prestige, education, income, women) |>
  cor()
```

Regression outputs

```r
full.reg <- lm(prestige ~ education + income + women,
               data = Prestige)

summary(full.reg)

anova(full.reg)

extractAIC(full.reg)
```

Residual plots

```r
library(ggfortify)

autoplot(full.reg, 1:6)

# Old style plots
plot(full.reg, 1) # the argument 1 can be changed up to 6

# or just use
par(mfrow=c(2,2))
plot(full.reg)
```

Exercise 7.2

Perform stepwise regression analysis of `prestige` on `education`, `income`, & `women`.

```
full.reg = lm(prestige ~ education + income + women,
              data = Prestige)

step(full.reg)

step(full.reg, direction="backward")

step(full.reg, direction="both")
```

You can also use the `MASS` package.

```
library("MASS")

stepAIC(full.reg, direction="backward")

stepAIC(full.reg, direction="both")
```

The function `update()` will be handy. For example, see try the following codes:

```
m1 = update(full.reg,.~.-women)

summary(m1)
```

Note that `.~.-women` means that the model is fitted without the `women` variable.

Further options are available in `leaps` and `HH` packages (installation commands are given below).

```
install.packages("leaps", repos = "https://cran.r-project.org") install.packages("HH",
repos = "https://cran.r-project.org")
```

```
library(leaps)
```

```
model = regsubsets(prestige ~ education + income + women,
                    data=Prestige)

library(HH)

summaryHH(model)

plot(summaryHH(model))
```

Exercise 7.3

Perform a polynomial regression of `prestige` on `income`.

```r
# Cubic fit
p.model <- lm(prestige ~ poly(income,3),
              data = Prestige)

summary(p.model)

extractAIC(p.model)

plot(p.model)

autoplot(p.model)
```

- More R code examples are here