STA 4320 CHAP 3.2.1

He Jiang

Sep 2024

Sec 3.2.1

Advertising dataset

```
fpath = getwd()
Advertising = read.csv(paste0(fpath, "/Advertising.csv"))
y = Advertising$sales
n = nrow(Advertising)
```

See the first 6 locations(data points) as an example

```
head(Advertising)[,-1]
```

```
##      TV radio newspaper sales
## 1 230.1  37.8      69.2  22.1
## 2  44.5  39.3      45.1  10.4
## 3  17.2  45.9      69.3   9.3
## 4 151.5  41.3      58.5  18.5
## 5 180.8  10.8      58.4  12.9
## 6   8.7  48.9      75.0   7.2
```

Matrix forms

```
y = Advertising$sales
X = cbind(1, Advertising$TV, Advertising$radio, Advertising$newspaper)

solve(t(X) %*% X) %*% t(X) %*% y
```

```
##              [,1]
## [1,]  2.938889369
## [2,]  0.045764645
## [3,]  0.188530017
## [4,] -0.001037493
```

```
# note that solve gives inverse in R
# %*% gives matrix multiplication
```

R multiple regression

R multiple regression by specifying the data frame

```
res = summary( lm(sales ~ TV + radio + newspaper, data = Advertising) )
res
```

```
##
```

```
## Call:
## lm(formula = sales ~ TV + radio + newspaper, data = Advertising)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8277 -0.8908  0.2418  1.1893  2.8292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
## TV           0.045765   0.001395  32.809   <2e-16 ***
## radio        0.188530   0.008611  21.893   <2e-16 ***
## newspaper   -0.001037   0.005871  -0.177     0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.686 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

R multiple regression by attaching the data first

```
attach(Advertising) # attach tells R to run the regression on this data
```

```
## The following object is masked _by_ .GlobalEnv:
##
##     X
```

```
res = summary( lm(sales ~ TV + radio + newspaper) )
res
```

```
##
## Call:
## lm(formula = sales ~ TV + radio + newspaper)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8277 -0.8908  0.2418  1.1893  2.8292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
## TV           0.045765   0.001395  32.809   <2e-16 ***
## radio        0.188530   0.008611  21.893   <2e-16 ***
## newspaper   -0.001037   0.005871  -0.177     0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.686 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
# detach(Advertising)
```

Difference between simple and multiple regression

Newspaper only

```
summary( lm(sales ~ newspaper, data = Advertising) )
```

```
##
## Call:
## lm(formula = sales ~ newspaper, data = Advertising)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2272  -3.3873  -0.8392   3.5059  12.7751
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.35141    0.62142   19.88  < 2e-16 ***
## newspaper    0.05469    0.01658    3.30  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.092 on 198 degrees of freedom
## Multiple R-squared:  0.05212,    Adjusted R-squared:  0.04733
## F-statistic: 10.89 on 1 and 198 DF,  p-value: 0.001148
```

Correlation matrix of variables

```
round(cor(Advertising[,-1]), 4)
```

```
##               TV  radio newspaper  sales
## TV        1.0000 0.0548    0.0566 0.7822
## radio     0.0548 1.0000    0.3541 0.5762
## newspaper 0.0566 0.3541    1.0000 0.2283
## sales     0.7822 0.5762    0.2283 1.0000
```