Exercises03

Dennis Perrett

1/14/2022

Exercise 15. A psychiatric clinic collects information on the variables

Y Overall clinical impression (score, the higher the better)

X_1 Does of an antipsychotic (in mg)

X_2 Days since hospitalisation

and the follwing values are observed for four patients:

```
A <- data.frame(matrix(c(40,45,50,65,1,2,3,4,36,33,37,37),4,3))
row.names(A) <- c(1,2,3,4)
col.names <- c("Y","X1","X2")
A
```

```
##   X1 X2 X3
## 1 40  1 36
## 2 45  2 33
## 3 50  3 37
## 4 65  4 37
```

To predict the overall clinical impression, a regression model of the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \qquad (i = 1, ..., 4)$$

is set up with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d.. Perform all calculations by hand and using R.

a. Compute the estimated values for $\beta = (\beta_0, \beta_2, \beta_2)'$ and σ^2 according to the ordinary least squares method. Note: we have

$$(X'X)^{-1} = \frac{1}{166} \begin{pmatrix} 23369 & 393 & -680 \\ 393 & 43 & -14 \\ -680 & -14 & 20 \end{pmatrix}$$

and

$$(X'X)^{-1}X' = \frac{1}{166} \begin{pmatrix} -718 & 1715 & -612 & -219 \\ -68 & 17 & 4 & 47 \\ 26 & -48 & 18 & 4 \end{pmatrix}$$

$$
\begin{aligned}
(X'X)^{-1}X'Y =& \beta \\
=& \frac{1}{166}\begin{pmatrix} -718 & 1715 & -612 & -219 \\ -68 & 17 & 4 & 47 \\ 26 & -48 & 18 & 4 \end{pmatrix} \cdot \begin{pmatrix} 40 \\ 45 \\ 50 \\ 60 \end{pmatrix} \\
=& \frac{1}{166}\begin{pmatrix} -718*40+1715*45+-612*50+-219*60 \\ -68*40+17*45+4*50+47*60 \\ 26*40+-48*45+18*50+4*60 \end{pmatrix} \\
=& \frac{1}{166}\begin{pmatrix} 3640 \\ 1300 \\ 40 \end{pmatrix} \\
=& \begin{pmatrix} 21.81 \\ 7.81 \\ 0.24 \end{pmatrix}
\end{aligned}
$$

```r
X <- cbind(1,as.matrix(A[,2:3]))
y <- as.matrix(A[,1])

beta <- solve((t(X)%*%X))%*%t(X)%*%y
beta
```

```
##           [,1]
##     21.8072289
## X2   7.8313253
## X3   0.2409639
```

b. For a Type I error of $\alpha = 0.05$, test whether the predictors contribute at all to a prediction of the overall clinical impression.

$$
F = \frac{\frac{R^2}{p}}{\frac{1-R^2}{n-p-1}} = \frac{\frac{SS_{reg}}{p}}{\frac{SS_{res}}{n-p-1}}
$$

where $p =$ number of non-intercept betas. And $n =$ the number of samples.

and

$$
\begin{aligned}
SS_{tot} \quad =& \quad SS_{reg} \quad + \quad SS_{res} \\
(y-\bar{y})'(y-\bar{y}) =& (\hat{y}-\bar{y})'(\hat{y}-\bar{y}) + (y-\hat{y})'(y-\hat{y}) \\
\sum_{i=1}^{n}(y_i-\bar{y})^2 =& \sum_{i=1}^{n}(\hat{y}-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\hat{y})^2
\end{aligned}
$$

therefore:

$$
\begin{aligned}
SS_{reg} =& (38.313-50)^2 + (45.42169-50)^2 + (54.24-50)^2 + (62.048-50)^2 \\
=& (-11.68)^2 + (-4.58)^2 + (4.22)^2 + (12.04)^2 \\
=& 136.58 + 20.96 + 17.78 + 145.156 \\
=& 320.4819 \\
SS_{res} =& (38.313-40)^2 + (45.42169-45)^2 + (54.24-50)^2 + (62.048-65)^2 \\
=& (-1.6867470)^2 + (0.4216867)^2 + (4.2168675)^2 + (-2.9518072)^2 \\
=& 2.8451154 + 0.1778197 + 17.7819713 + 8.7131659 \\
=& 29.51807
\end{aligned}
$$

and

$$F = \frac{\frac{SS_{reg}}{p}}{\frac{SS_{res}}{n-p-1}} = \frac{\frac{320.4819}{2}}{\frac{29.51807}{4-2-1}} = 5.428571 < F_{crit:0.95} = 199.5$$

We cannot reject the null that the linear combination of the betas and X has no significant predictive power. Not significant. Do not reject.

```r
# By Hand-R
y.hat = X%*%beta
mean(y)
```

```
## [1] 50
```

```r
n <- 4
p <- length(beta)-1
ss.res<- sum((y.hat-y)^2)
ss.reg<-sum((y.hat-mean(y))^2)
((ss.reg)/(p))/((ss.res)/(n-p-1))
```

```
## [1] 5.428571
```

```r
qf(0.95,2,1)
```

```
## [1] 199.5
```

With R:

```r
A
```

```
##   X1 X2 X3
## 1 40  1 36
## 2 45  2 33
## 3 50  3 37
## 4 65  4 37
```

```r
lreg <- lm(X1 ~ X2 + X3, A)
summary(lreg)
```

```
##
## Call:
## lm(formula = X1 ~ X2 + X3, data = A)
##
## Residuals:
##       1       2       3       4
##  1.6867 -0.4217 -4.2169  2.9518
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   21.807     64.463   0.338    0.792
## X2             7.831      2.765   2.832    0.216
## X3             0.241      1.886   0.128    0.919
##
## Residual standard error: 5.433 on 1 degrees of freedom
## Multiple R-squared:  0.9157, Adjusted R-squared:  0.747
## F-statistic: 5.429 on 2 and 1 DF,  p-value: 0.2904
```

F-statistic: 5.429, with p-value of 0.2904. Confirms above. Estimates are the betas.