Statistics-III

Trishan Mondal

Assignment-8

Problem 1

```
corn = read.table("corn.txt")
# Getting the correlation matrix
A = matrix(data=NA ,nrow = 10, ncol=10)
for(i in 1:10){
for(j in 1:10){
A[i,j] = cor(corn[,i],corn[,j])}
}
```

Now we need to check which variable has the maximum correlation with Y.

```
x = c()
for(i in 1:9){
x[i] = A[10,i]
}
#Finding the maximum
which.max(abs(x))

## [1] 1

max(abs(x))

## [1] 0.7517696
```

We got X_1 has a maximum correlation with Y.

```
lm.corn1 = lm(corn[,10]~corn[,1])
summary(lm.corn1)

##
## Call:
## lm(formula = corn[, 10] ~ corn[, 1])
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.501  -5.445   0.861   6.727  14.040
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   32.494      3.156  10.295 1.59e-11 ***
## corn[, 1]      1.028      0.162   6.347 4.58e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.86 on 31 degrees of freedom
## Multiple R-squared:  0.5652,Adjusted R-squared:  0.5511
## F-statistic: 40.29 on 1 and 31 DF,  p-value: 4.583e-07
```

In this case, $R^2 = 56.5\%$. Now we will look at the partial correlation coefficient $r_{iy.1}$.

```
#The following vector will give us the values of r_{iy.1} for different i
y = c()
partial.cor = function(x,y,z){
return((x-y*z)/(sqrt((1-y^2)*(1-z^2))))
}

for(i in 2:9){
y[i] = partial.cor(A[i,10],A[i,1],A[1,10])
}
y

## [1]          NA  0.24183786 -0.10890392 -0.13792990  0.07878379  0.46162737
## [7] -0.30711144  0.24886731 -0.42650034

which.max(abs(y))

## [1] 6

summary(lm(corn[,10]~corn[,6]+corn[,1]))

##
## Call:
## lm(formula = corn[, 10] ~ corn[, 6] + corn[, 1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.789  -6.174   2.843   6.197  10.169
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.4560     4.0065   6.104 1.04e-06 ***
## corn[, 6]     3.1446     1.1032   2.850  0.00782 **
## corn[, 1]     0.8453     0.1595   5.299 1.00e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.99 on 30 degrees of freedom
## Multiple R-squared:  0.6578,Adjusted R-squared:  0.635
## F-statistic: 28.84 on 2 and 30 DF,  p-value: 1.032e-07
```

In this case, We will choose X_6 and R^2 value is 65.78%. X_6 will also have an important contribution. Now,we need to compute $r_{iy.16}$. The values are computed as following,

$$
\begin{aligned}
r_{2y.16} &= 0.300038611141932 \\
r_{3y.16} &= 0.179398775787862 \\
r_{4y.16} &= 0.0128078684982457 \\
r_{5y.16} &= 0.0759728637870637 \\
r_{7y.16} &= 0.158816380491146 \\
r_{8y.16} &= 0.284062595649432 \\
r_{9y.16} &= 0.287927197687353
\end{aligned}
$$

We will take X_2. The R^2 value is 68.8%. So as of now $Y \sim X_1 + X_6 + X_2$. Now will look at $r_{iy.162}$.

$$
\begin{aligned}
r_{3y.162} &= 0.152797944791066 \\
r_{4y.162} &= 0.12459244201669 \\
r_{5y.162} &= 0.251774564688212 \\
r_{7y.162} &= 0.0499070701252337 \\
r_{8y.162} &= 0.241092329779549 \\
r_{9y.162} &= 0.327243940162817
\end{aligned}
$$

We can take X_9. In this case R^2 value 72.19% which is $> 70\%$. So, we need $Y \sim X_1 + X_6 + X_2 + X_9$ corresponding terms appear in regression.

Problem 2

We will assign μ_i to the Temperature and we will assign y_{ij} to the density of j-th brick at μ_i temperature. We fit a linear model corresponding this by $y_{ij} = \mu_i + \epsilon_{ij}$ where $i = 1, \cdots 4$ and $j = 1, \cdots 4$.

We will assume the null hypothesis as $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. Now,

$$RSS_{H_0} = \sum_i \sum_i (y_{ij} - \bar{y})^2 = RSS = 0.2975$$
$$RSS = \sum_i \sum_j (y_{ij} - \bar{y_i})^2 = 0.484375$$

For this case $n = 16$ and $k = 4$ So, we will use the formula $\frac{(RSS_{H_0} - RSS)/(k-1)}{RSS/(\sum_i n_i - k)} \sim F_{k-1,n-k} = F_{3,12}$. In this case F- value is, 2.5126.

```
qf(0.90,3,12)

## [1] 2.605525
```

we can accept the null hypothesis. So, firing temperature does not affect the density of bricks.