

MA5013: Assignment 2

B Aditi \|\| MM19B022

Tables

Table B1

```{r}
tb1 <- read.csv('Assignment 2/Table_B1.csv',header=TRUE,
                      stringsAsFactors=FALSE)
print(tb1)
```

Table B3

```{r}
tb3 <- read.csv('Assignment 2/Table_B3.csv',header=TRUE,
                      stringsAsFactors=FALSE)
print(tb3)
```

Table B4

```{r}
tb4 <- read.csv('Assignment 2/Table_B4.csv',header=TRUE,
                      stringsAsFactors=FALSE)
print(tb4)
```

Table B6

```{r}
tb6 <- read.csv('Assignment 2/Table_B6.csv',header=TRUE,
                      stringsAsFactors=FALSE)
print(tb6)
```

Problems

Question 3.1

```{r}
#(a) Fitting the model using x2, x7 and x8
model <- lm(Y ~ X2 + X7 + X8, data = tb1)
summary(model)
```

```{r}
#(b) anova test
anova(model)
```

```{r}
#(c) t statistics for hypothesis
summary(model)$coef[, c( "t value", "Pr(>|t|)")]
```

```{r}
#(d) R^2 and R_adj^2 calculation
print(paste("R-squared error is",summary(model)$r.squared))
print(paste("Adjusted R-squared error is",summary(model)$adj.r.squared))
```

```{r}
#(e) Partial F-statistics for contribution of X7
red_model <- lm(Y ~ X2 + X8, data = tb1)
summary(red_model)
```

```{r}
#(e) contd..
anova(red_model, model)
```

Thus, from this result we can extract that the partial F score for the variable 'X7' is 4.8324.

```{r}
#(e) relationship between partial f and t tests
t_val <- summary(model)$coef[3,3]
f_val <- t_val^2
print(paste("The t-value of X7 is found to be", t_val, "while the square of it is",
f_val, "which is the same as its partial F-score"))
```

Question 3.3

```{r}
#(a) 95% CI on X7
cof <- confint(model)["X7",]
print(paste("Thus the 95% confidence interval for coefficient of X7 is
between",cof[1],"and",cof[2]))
print(cof)
```

```{r}
#(b) Prediction for given dataset
new_data <- data.frame(X2 = 2300, X7 = 56.0, X8 = 2100)
pred_interval <- predict(model, newdata = new_data, interval = "confidence", level =
0.95)
pred_interval
```

```{r}
print(paste("Thus, the 5% confidance interval for the predicted value of the given data
is found to be between",pred_interval[2],"and",pred_interval[3] ))
```

Problem 3.4

```{r}
#(a) Significance of the model
#In order to test the significance first we must get the summary of the model
new_model <- lm(Y ~ X7 + X8, data = tb1)
summary(new_model)
```

The value of F-statistics is found to be 15.13 and the p-value is obtained to be 0.00005 which implies that the model is significant.

```{r}
#In addition, we can also perform anova to test the significance
anova(new_model)
```

Since the p-values of both variables are less than 0.05, we can conclude that this is a valid model.

```{r}
#(b) R^2 and R_adj^2 calculation
```

```
print(paste("R-squared error is",summary(new_model)$r.squared))
print(paste("Adjusted R-squared error is",summary(new_model)$adj.r.squared))
```

We noticed that botht the R-squared and adjusted R-squared values reduce significantly.

```{r}
#(c) 95% CI for X7
cof1 <- confint(new_model)["X7",]
print(paste("Thus the 95% confidence interval for coefficient of X7 is between",cof1[1],"and",cof1[2]))
print(cof1)
```

```{r}
#(c) 95% confidance interval in prediction
new_data1 <- data.frame(X7 = 56.0, X8 = 2100)
pred_interval1 <- predict(new_model, newdata = new_data1, interval = "confidence", level = 0.95)
pred_interval1
```

```{r}
print(paste("Thus, the 5% confidance interval for the predicted value of the given data is found to be between",pred_interval1[2],"and",pred_interval1[3] ))
```

From this, we can clearly see that the interval size for both coefficient of X7 and the prediction has increased significantly

(d) Overall, we observe that omitting an important regressor can increase the range for a given confidance interval while decreasing the R-squared and Adjusted R-Squared values. It significantly affects the different parameters of the model and its significance.

Problem 3.5

```{r}
#(a) Fitting multiple linear regression
model1 = lm(y ~ X1 + X6, data = tb3)
summary(model1)
```

```{r}
#(b) constructung anova table
a1 = anova(model1)
a1
```

```{r}
#(c) R-squared and adjusted R-squared values
print(paste("R-squared error is",summary(model1)$r.squared))
print(paste("Adjusted R-squared error is",summary(model1)$adj.r.squared))
```

```{r}
#(d) 95% CI for X1
cof3 <- confint(model1)["X1",]
print(paste("Thus the 95% confidence interval for coefficient of X7 is between",cof3[1],"and",cof3[2]))
print(cof3)
```

```{r}
#(e) t statistics for X1 and X6
summary(model1)$coef[, c( "t value", "Pr(>|t|)")]
```

We notice that based on the t-statistics and p-values, we understand that X1 is significant while X6 is not so significant.

```{r}
#(f) 95% CI on new data with prediction
new_data1 <- data.frame(X1 = 275, X6 = 2)
pred_interval1 <- predict(model1, newdata = new_data1, interval = "confidence", level = 0.95)
pred_interval1
```

```{r}
#(g)  95% CI on new data with prediction
new_data11 <- data.frame(X1 = 257, X6 = 2)
pred_interval11 <- predict(model1, newdata = new_data11, interval = "confidence", level = 0.95)
pred_interval11
```

Problem 3.7

```{r}
model4 = lm(X ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9, data = tb4)
summary(model4)
```

(b) From the above summary we can see that the F-statistics value is 10.94 with a p-value of 0.00006, implying its a significant model.

```{r}
#(c) t-test for each variable
summary(model4)$coef[, c( "t value", "Pr(>|t|)")]
```

From these values, we observe that none of the t-statistic values are significant. We also notice that there is a multicolinearity problem

```{r}
#(d) significance of Lot size(x3) and Living space(X4)
red_model4 <- lm(X ~ X1 + X2 + X5 + X6 + X7 + X8 + X9, data = tb4)
anova(red_model4, model4)
```

Since the F-value of the 2 variables come to 0.2559, we can say they are not significant variables.

(e) As stated in the test statistic analysis, there is a multicollinearity problem.

Problem 3.9

```{r}
#(a) Fitting multiple regression
model6 <- lm(y ~ X1 + X4, data = tb6)
summary(model6)
```

(b) Since the F-statistics value is 24.44 and p-value is 0.0000012, the model is significant. For further analysis refer to the anova table

```{r}
anova(model6)
```

```{r}
#(c) R-squared and adjusted R-squared values
print(paste("R-squared error is",summary(model6)$r.squared))
print(paste("Adjusted R-squared error is",summary(model6)$adj.r.squared))
```

```{r}
#(d) t-test on X1 and X4
summary(model6)$coef[, c( "t value", "Pr(>|t|)")]
```

Thus, based on the t and p values, we can conclude that X1 is signific

(e) Based on the results obtained, multicolinearity does not seems to be a problem.