Bangabandhu Sheikh Mujibur Rahman Agricultural University
EDGE_Batch-11
Mid Exam
Marks: 20   Time: 90 minutes
Name: …Ummay Habiba…………………………
Reg. No:…18-05-4603………………Dept.…Agronomy…………………..

Note: Submit the completed file to ***nazmol.stat.bioin@bsmrau.edu.bd*** and ***keyadas57@bsmrau.edu.bd*** with subject ***EDGE11_Mid_Your registration number_ Dept.***

1. Short Questions **(5*1=05)**

1. When comparing the means of two related groups (e.g., pre-test and post-test), the ... **paired t-test**.......... test is used, assuming the data is normally distributed.
2. In regression analysis, the .... **t-test**......... test is used to determine if the slope of a regression line is significantly different from zero, assuming normally distributed residuals.
3. In testing for normality, the ..**Shapiro-Wilk test**........... test is used to check if a data set follows a normal distribution, assuming that the data are parametric.
4. The ..**Kruskal-Wallis test**........... non-parametric test is used when comparing three or more independent groups.
5. The ..**Spearman's rank correlation**........... correlation measures the degree of association between two variables when both are measured at the ordinal level.

2. For the given data set “Reg1”,

a) Present a correlation plot among independent variables using corrplot package.
b) Check the assumptions and fit a multiple linear regression model.
c) Apply forward selection method (stepwise regression) to find best subset of the independent variables.

Answer to the ques no 2

(a)

Here is the code.

```
Data <- read.csv("Reg1.csv")

library("corrplot")

correlations <- cor(Data[,-1])
```

```
corrplot(correlations,

        method = "circle",

        col = colorRampPalette(c("lightgreen", "white", "darkgreen"))(200),

        tl.col = "black",

        mar = c(0, 0, 3, 0))

title("Correlation Plot", col.main = "black", cex.main = 1.2)
```

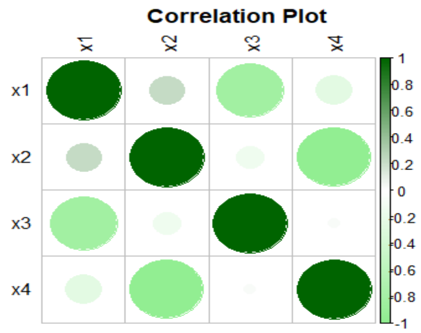


Figure: correlation plot among independent variables using corrplot package

Interpretation: This is a correlation plot where size and color of the circles indicate the strength and direction of the correlation between independent variables (X1, X2, X3, X4). In Darker and larger circles represent stronger correlations, with blue indicating positive correlations and lighter shades represent weaker correlation and no correlations means absence of color.

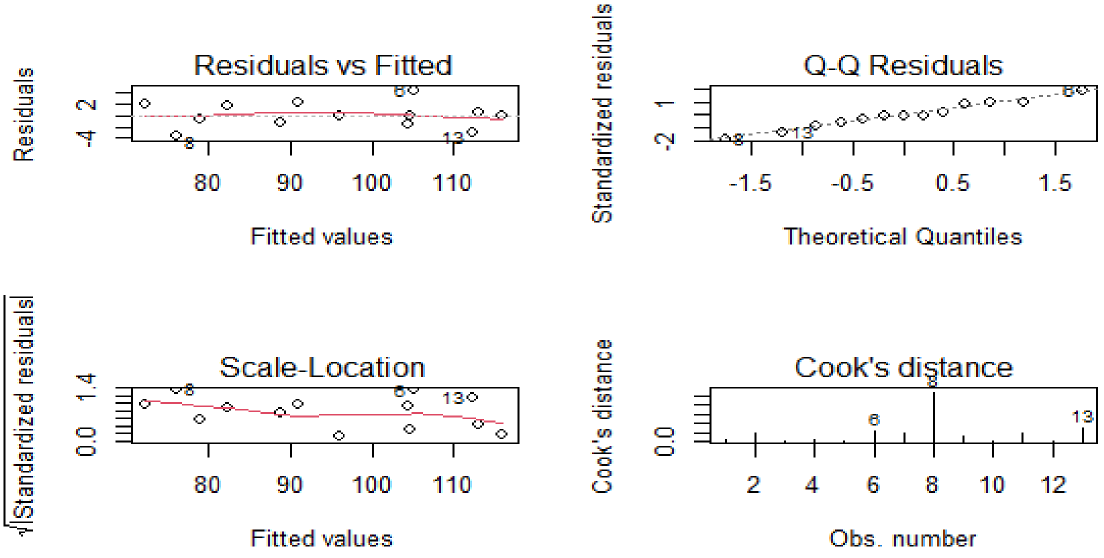
(b)

Here is the code.

```
Data <- read.csv("Reg1.csv", header=T)

head(Data)

model1<-lm(y~.,data=Data)

summary(model1)

AIC(model1)

abline(lm(y~.,data=Data))

model2<-lm(y~x1+x2+x3+x4,data=Data)
```

```
summary(model2)

AIC(model2)

abline(lm(y~x1+x2+x3+x4,data=Data))

model3<-lm(y~0+x1+x2+x3+x4,data=Data)

summary(model3)

AIC(model3)

abline(lm(y~0+x1+x2+x3+x4,data=Data))

par(mfrow = c(2, 2))

plot(model3, which = 1)

plot(model3, which = 2)

plot(model3, which = 3)

plot(model3, which = 4)
```

There are 3 multiple linear regression model (model1, model 2, model 3). In model1 and model2, R-square value is 0.9824 and AIC value is 65.83669. In model3, R-square value is 0.9996 and AIC value is 65.06563. All the model has high R-square value, and $p<0.05$ which is significant. But, in model3, R-square value is higher than any other models, AIC value is lower than any other models. Four plots for model3 are given below:



(c)

Here is the code.

```
library(MASS)
stepwise_model <- stepAIC(lm(y ~ 1, data = Data),
                          scope = list(lower = ~1, upper = ~x1 + x2 + x3 + x4),
                          direction = "forward")
summary(stepwise_model)
```

```
Result:
Residuals:
    Min      1Q      Median     3Q       Max
-3.0919  -1.8016   0.2562   1.2818   3.8982

Coefficients:
             Estimate Std.  Error     t value     Pr(>|t|)
(Intercept)   71.6483     14.1424     5.066     0.000675 ***
x4           -0.2365      0.1733     -1.365     0.205395
x1            1.4519      0.1170     12.410     5.78e-07 ***
x2            0.4161      0.1856      2.242     0.051687 .

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 2.309 on 9 degrees of freedom

Multiple R-squared: 0.9823, Adjusted R-squared: 0.9764

F-statistic: 166.8 on 3 and 9 DF, p-value: 3.323e-08

Interpretation:

Intercept (71.6483): This is predicted value of y when all independent variables are 0.

x4 (-0.2365): For every unit increase in x4, y decreases by 0.2365 units, holding other variables constant. However, p-value(0.205) indicates that it is not statistically significant.

x1 (1.4519): For every unit increase in x1, y increases by 1.4519 units, holding other variables constant. This variable is highly significant (p-value < 0.001).

x2 (0.4161): For every unit increase in x2, y increases by 0.4161 units, holding other variables constant. It is marginally significant (p-value = 0.0517).

3.A randomized complete block design was conducted considering four blocks, seven levels/treatments. Afterward, the yield of certain plant characteristics was observed. The data regarding this experiment were given in the file “RBDdata”. Answer the following question using this data.

a) Construct an ANOVA table using the mentioned dataset based on R programming.
b) Write down the null hypothesis of the treatment effects and interpret the results based on the ANOVA table.
c) Perform a post-hoc test for the treatments and draw a bar diagram with lettering.

Answer to the ques no 3

(a)

Here is the dataset based on r programming.

```
Data.RCBD<-read.csv("RBDdata.CSV")
Data.RCBD<-Data.RCBD[,2:4]
Rep<-c("Rep1","Rep2","Rep3","Rep4")
Treat<-c("Treat1","Treat2","Treat3","Treat4","Treat5","Treat6","Treat7")
r<-length(Rep)
t<-length(Treat)
Block<-gl(r,t,r*t,factor(Rep))
Treat<-gl(t,1,r*t,factor(Treat))
ANOVA.RCBD<-aov(YIELD~Block+Treat,
                data=Data.RCBD)
summary(ANOVA.RCBD)
```

ANOVA Table

```
           Df  Sum Sq  Mean Sq   F value  Pr(>F)
Block       3   1742     580.7      29.61   3.55e-07 ***
Treat       6   12148    2024.6    103.24   5.96e-13 ***
Residuals        18      353          19.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(b)

The null hypothesis based on treatment effects:
H0: µ1=µ2=µ3=⋯=µt & the alternative hypothesis is H1 which is opposite to H0.
H1: µ1 ≠ µ2 ≠ µ3 ≠ ……. ≠ µt

ANOVA Table

```
            Df   Sum Sq  Mean Sq    F value   Pr(>F)
Block        3   1742     580.7        29.61    3.55e-07 ***
Treat        6   12148   2024.6     103.24     5.96e-13 ***
Residuals        18    353             19.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table, p value is less than 0.05. So, the null hypothesis is rejected. It indicates that there is significant differences among treatments effects.

(c)

Here is the r code for post Hoc test.

```
library(agricolae)
PostHoc.Test<-with(Data.RCBD,HSD.test(YIELD,Treat,DFerror=18,MSerror=19.6))
YIELD    groups    Lettering
Treat6   133.25      a
Treat3   127.00     ab
Treat5   125.75     ab
Treat1   125.00     ab
Treat7   121.00      b
Treat4    87.75      c
Treat2    75.25      d
```

From this test, I can say that Treat6, Treat3, Treat5, Treat1 get the same letter which is lettering 'a'. So, these four treatments are better for getting better yield but Treat6 is the best in all of these treatments.

```
Mutplcom.TreatFact<-with(Data.RCBD,HSD.test
                  (YIELD,Treat,DFerror=18,MSerror=19.6))
library(gplots)
Treat.SE.Mat<-Mutplcom.TreatFact$means[,"se"]
Treat.Mean<-Mutplcom.TreatFact$groups
Mean.Mat<-Mutplcom.TreatFact$means
Mean.Mat<-Mean.Mat[order(-Mean.Mat$YIELD)]
Treat.Treat.Mean<-Treat.Mean$YIELD
Treat.SE<-Mean.Mat[, "se"]
Treat.SE.Mat<-Mutplcom.TreatFact$means[order(Mutplcom.TreatFact$means[,"se"])]
```

Here is the code for Barplot.

```
Barplot.Se<-barplot2(Treat.Treat.Mean,
                     names.arg = rownames(Treat.Mean),
                     xlab="Treatment",ylab="Yield",
                     horix=F,plot.ci = T,
                     ci.l=Treat.Treat.Mean-Treat.SE,
                     ci.u=Treat.Treat.Mean-Treat.SE,
                     col="lightblue")
text(Barplot.Se, 7,Treat.Mean$groups, cex=2,
     pos = 3, col= "black")
```

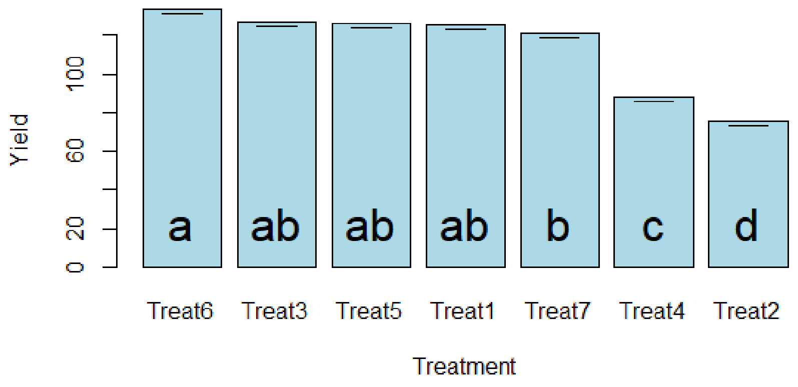


Figure: Barplot with treatment & yield