MULTIPLE LINEAR REGRESSION

PROBLEM

The following measurements have been obtained in a study:

Y	1.45	1.93	0.81	0.61	1.55	0.95	0.45	1.14	0.74	0.98	1.41	0.81
X1	0.58	0.86	0.29	0.20	0.56	0.28	0.08	0.41	0.22	0.35	0.59	0.22
X2	0.71	0.13	0.79	0.20	0.56	0.92	0.01	0.60	0.70	0.73	0.13	0.96

Y	0.89	0.68	1.39	1.53	0.91	1.49	1.38	1.73	1.11	1.68	0.66	0.69
X1	0.26	0.12	0.65	0.70	0.30	0.70	0.39	0.72	0.45	0.81	0.04	0.20
X2	0.27	0.21	0.88	0.30	0.15	0.09	0.17	0.25	0.30	0.32	0.82	0.98

Fit a multiple linear regression line and estimate the response variable y when x1 = 0.95 and x2 = 0.00 and carry out residual analysis to check if model assumptions are fulfilled

AIM

To fit a multiple linear regression model on the given data set.

HYPOTHESIS

Null Hypothesis H_0: $\beta 1 = \beta 2 = 0$

Alternate Hypothesis H_a: At least one of these 2 βs is not 0

PROCEDURE

1. Open R studio
2. Read the inputs using data frame()
3. Type the commands to achieve the desired output
4. Get the output

INPUT

1. Enter the independent and dependent variables using R command data frame()
2. Build the model using R command lm()
3. Using summary() function in R, find the co-efficients of the model and p value which is used to using summary()
4. Plot the residuals of the model using qqnorm() and qqline()
5. Predict new values based on previous data using predict()

CODE IN R LANGUAGE

```
#Loading the data using a dataframe
D = data.frame(

x1=c(0.58,0.86,0.29,0.20,0.56,0.28,0.08,0.41,0.22,0.35,0.59,0.22,0.26,0.12,0.65,0.70,0.30,
0.70,0.39,0.72,0.45,0.81,0.04,0.20),

x2=c(0.71,0.13,0.79,0.20,0.56,0.92,0.01,0.60,0.70,0.73,0.13,0.96,0.27,0.21,0.88,0.30,0.15,
0.09,0.17,0.25,0.30,0.32,0.82,0.98),

y=c(1.45,1.93,0.81,0.61,1.55,0.95,0.45,1.14,0.74,0.98,1.41,0.81,0.89,0.68,1.39,1.53,0.91,1.
49,1.38,1.73,1.11,1.68,0.66,0.69)
)

#Building model
fit = lm(y~x1+x2,data = D)
summary(fit)

#Residual Analysis to check if model assumptions are fulfilled
residuals = resid(fit)
qqnorm(residuals)
qqline(residuals)

#Prediction
newdata = data.frame(x1 = 0.95,
            x2 = 0.00)
predict(fit,newdata)
```

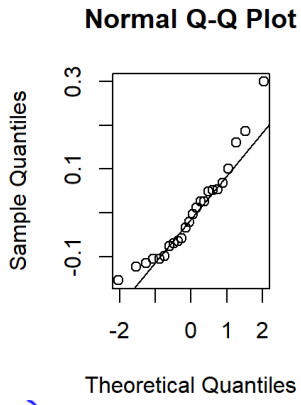
OUTPUT

```
Call:
lm(formula = y ~ x1 + x2, data = D)

Residuals:
     Min       1Q   Median       3Q      Max
-0.15487 -0.08138 -0.01188  0.05192  0.29947

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.43236    0.06764   6.392 2.45e-06 ***
x1           1.66152    0.10388  15.994 3.10e-13 ***
x2           0.00102    0.07752   0.013     0.99
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1152 on 21 degrees of freedom
Multiple R-squared:  0.9294,    Adjusted R-squared:  0.9226
F-statistic: 138.1 on 2 and 21 DF,  p-value: 8.221e-13
```




```
> predict(fit,newdata)
       1
2.010806
```

RESULT

- From the summary, we can see that p value of the F-test is less than 0.05. Hence, we reject the null hypothesis at 5% significance level and conclude that there is a linear relationship between y and independent variables.
- The qq-plot is a straight line and hence the data is normally distributed and the assumptions of the model are fulfilled.
- Based on the multiple linear regression model, the predicted measurement is 2.010806.