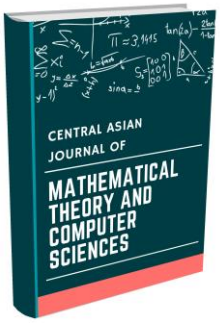


CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

http://cajmtcs.centralasianstudies.org/index.php

Volume: 01 Issue: 01 | Sep 2020

Fundamental types of regression analysis on various research data using R Programming

Yagyanath Rimal

[1]Faculty of Science and Technology, Pokhara University, Nepal
Corresponding E-mail: rimal.yagya@pu.edu.np

ABSTRACT

The goal of this analytical review paper is for discussing the relationship between various types of regression analysis methods whose output were sufficiently analyzed using R programming. The regression analysis is calculated with three different case studies of different datasets for explaining linear and multiple regressions. Similarly, polynomial regression analysis is calculated with 2955 observation and 8 attributes of Florida date set whose residual standard error is calculated with 11520 on 2949 degrees of freedom, multiple R-squared is 0.07828, the adjusted R-squared is 0.07672 and F-statists is 50.09 on 5. Likewise, the quintal regression analysis is carried out through binary data sets of 20 observations of 4 attributes whose, AIC value is fit between two or more models at 26 percentages and 75 percent accuracy. The primary purpose of this paper is to explain the relationship of linear, multiple, quantile and polynomial regression models to achieve final conclusion with different data sets. Therefore, this paper presents easiest way of fundamental types of regression analysis commands and R programming strengths for of data analysis.

ARTICLE INFO

Article history:
Received 10 July 2020
Received in revised form 24 July 2020
Accepted 31 August 2020

Keywords:
Data Analytics, Quantile Polynomial, Big Data, Regression Equation.

1.INTRODUCTION

Regression analysis is a statistical process that allows you to investigate the linear relationship between the search variables is widely used for data set research predictions. The data analysis process always requires data analysis in modern data science (Manyika, 2011). Linear regression simply summarizes the association of related variables (Fiona, 2018). For example, sales and relationships with the company, sales records depend on factors in simple words, regression analysis is used to model the relationship between dependent and independent variables in the research field. However, the researcher can make independent decisions and employees accordingly. The independent variable is labeled as variable X, and the variable is the variable Y, which can be displayed on a graph, with X and Y independent variable (Astrid Schneider, 2010). X and Y can be expressed algebraically as Y = a + bX, where Y interception is called, and it is the slope of the regression line that determines the relationship between the search data. Likewise, the slope line refers to the slope of the line, as the line rises or falls sharply. The linear regression prediction model can be applied to individual or multiple independent or dependent variables. It is assumed that the relationship between these variables is of a uniform nature. The linear regression equation for multiple Y = B2X2 + B3X3 + B1 + ... + € BK, where y is the dependent variable to be estimated, and X is the independent variable and ε is the error term. B are the regression coefficients. Although, every regression has some assumptions that we must satisfy before performing the analysis. B1 is the term of the coefficient that is used to calculate the proportion of the independent variable. Therefore, it is that changing a unit into an independent variable decreases the value of the dependent variable if all other factors are constant. Regression analysis predicts the value of a variable based on one or more independent variables. The coefficient explains the impact of changes in the independent variable in the dependent variable. Regression analysis is widely used for prediction. Multivariate mode in which there are more than two variables involved, but there were two other sub-categories, for example linear and non-linear models of each type. In the linear model, the line is inserted into a line that has two simpler and multiple models. The model is known as a univariate multiple model. For the other variable, which is variable for the other, it is predictive of the variable. The simple linear regression line is included in the group of data represented. The distance from the average gives the residual value that is the discrepancy between the real

values and the predicates. Therefore, the procedure for finding the best solution is called the least squares method. The linear regression model represents an independent variable. After having built the model, it is foreseen by modifying the independent coefficient of the model.

However, there was some previous hypothesis that the predictive variable is not random; the error term is random due to the fact that the data must be independent of each other. Likewise, the coefficient of determination is a measure of the goodness of adaptation. When $R^2 = 0$ means that there is no relation, if R^2 = -ve there is a relation -ve the same way when R2 is + ve there is a positive relation. Therefore, the R^2 value can be explained, as the model explains the data and the percentage model. Differences in terms or residues. It assumed the houses of fair R^2 is 0.75, this means 75% of the variation in the values of the dependent variable explained by the model and the remaining 25% is not explained or residual error conditions. In a linear expression, sometimes the dependent variable is explained by a single variable. Multiple regressions, which attempts to explain the variable using more than one independent variable. Multiple regressions may be linear and non-linear (Blokhin, 2018). Polynomial regression is a technique for regulating a non-linear equation using a polynomial relationship of an independent variable. The polynomial relationship of dependent and independent variable should be in nonlinear. Polynomial regression is one of several line methods as: f (x) = c0 + c1 + c2 x 2 …cn xn where n is the degree of a polynomial and c is a set of regression coefficients which creates unnecessary additional features of fitting large data sets. Polynomial models are an excellent tool for determining input factors for response variables. A second order quadratic polynomial model for two explanatory variables has the form of the following equation as : $Y = \alpha 0 + \alpha 1x1 + \alpha 2x2 + \alpha 11x21 + \alpha 22x22\ \alpha 12x1x2 + \varepsilon$. Generally polynomial regression transforms the linear model to better fit with non-linear data. Linear regression is an automatic learning technique that allows researcher to be associated with a variable or an independent with dependent variable. Similarly, quintile regression is the degree of use of linear regression use outliers, high skewedness' and heteroscedasticity in the variable datasets. Although linear regression predicts the average of a relationship of dependent and independent variables which predicts the quintile for the given independent variables. Which try to estimate the quintile of the dependent variable given the values of X, where the dependent variable must be continuous. The QR is more appropriate when the conditions of normality and homoscedasticity are violated. Although it has recently increased its popularity in educational statistics (Chen, 2013), quintile regression provides an alternative to ordinary least squares (OLS) which typically assumes that the associations between independent and dependent variables are equal for levels. Quantum regression is not an estimated regression in a quintile, or sub-sample of data as the name suggests. Quantum methods allow the analyst to relax the intake of the common regression slope. In the OLS regression, the goal is to reduce the distances between the values predicted by the regression line and the observed values. In contrast, quintile regression reflects values differentially and thus seeks to reduce weighted distances (Cook, 2013). The main advantage of the quintile regression method is that the method makes it possible to understand the relationships between the external variables and the mean of the data, so it is useful for understanding the results that are not normally distributed and having non-linear predictive variables. Suppose that the regression equation for the 25th regression quartile is: y = 5.2333 + 700.823 x. It means that for an increase of units in x the estimated increase of the 25th quintile of e in 700,823 units. The advantages of quintile on linear regression are quite advantageous when the heteroscedasticity is present in the data, is robust for the anomalous values, the distribution of the dependent variable can be described through different quintiles and is more useful than linear regression when the data is asymmetric. The main advantage of the quintile regression methodology is that the method allows to understand the relationships between variables outside the data mean, so it is useful to understand the results that are not normally distributed and that have non-linear relationships with the predictor variables. The coefficients we obtain in the quintile regression for a particular quintile must differ significantly from those we obtain from linear regression. This can be done by looking at the confidence intervals of the regression coefficients of the estimates obtained from both regressions. The programming language R is a statistical language of open source programming that is free and has the support of a large community, developed by Ross Ihaka and Robert Gentleman. Software R is both a software and a programming language in which the user can develop many programs (Rogers, 1973). It has the ability to write many lines of code and produce output in the console command. R is available for all platforms; now there are more than 10,000 R packages available for download (Smith, 2018), which largely support data analysis.

Case 1: How to find linear regression equation of Table No 1 data sets.

Here the relationship between age and glucose can easily calculated mathematically after calculating xy, x2, y2 and

Table No: 1 Calculate Linear Regression

S.N	Age(X)	Glucose(Y)	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Sum	247	486	20485	11409	40022

sum of each value sets mathematically.

a= (Σy)*(Σx2)- (Σx)*(Σxy)/n (Σx2)-(Σx)2

=65.1416 (Intercept)

b= n(Σxy)-(Σx)* (Σy)/n(Σx2)-(Σx)2

b=0.3852 (Coefficient)

y=a+bx (Equation)

y=65.14+0.38x
From this equation we can easily predict y' value by substituting x values.

Using R Programming

The lm command is used to calculate simple linear regression in R programming which fits linear models. It can be used to carry out regression, single analysis of variance and analysis of covariance the above table could be analysis as follows.

```
> library(readxl)
> gulcose <- read_excel("C:/Users/Yagya
/Desktop /gulcose.xlsx")
> View(gulcose)
> str(gulcose)
> reg=lm(gulcose$Gulcose~gulcose$Age,
data=gulcose)
> reglm(formula = gulcose$Gulcose ~ gulcose$
Age, data = gulcose)
Coefficients:
(Intercept)  gulcose$Age
  65.1416      0.3852
> plot(gulcose$Age,gulcose$Gulcose,xlab
="Age",ylab="Gulcose")
> abline(lm(gulcose$Gulcose~gulcose$Age))
```

The figure demonstrates positive relationship between glucose and age relationship of above table data sets .

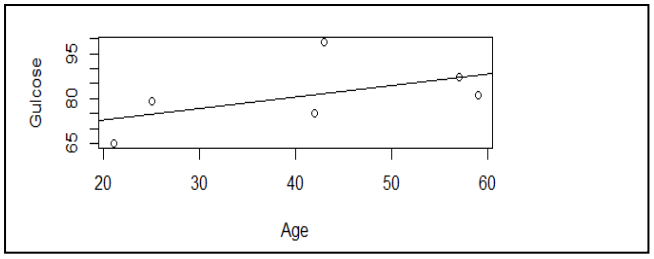
Case 2: How to find the regression of Table No Similarly from the Table No 2 the consumption rate of ice-cream uses of XYZ company having income, price and temperature variables data can be analyzed whether the income price and temperature increase or decrease the consumption of ice-cream linear relationship for future prediction could analyzed. Here is consumption is dependent variable depends on various independent variables where we can easily find out the pattern of consumption rate based on income price and temperature relationship.

Table No 2: Find the regression

S.N	Consume	Income	Price	Temp
1	0.38	77	0.2	37
2	0.37	75	0.22	38
3	0.40	82	0.24	39
4	0.40	85	0.26	39
5	0.41	84	0.28	40
6	0.42	84	0.3	40
7	0.43	85	0.32	41
8	0.44	86	0.34	42
9	0.45	87	0.36	42
10	0.46	88	0.38	43
11	0.47	91	0.4	43

Using R Programming

```
>onee=read.csv("c:/income.csv",header=True)
> View(onee)
> onee
> reg=lm(Consum~Income+Price+ Temp,data=onee)
> summary(reg)
lm(formula = onee$Consum ~ Income + Price + Temp,
data = onee)
```




```
Residuals:
Min       1Q   Median       3Q      Max
-0.0067851 -0.0015990  0.0004569  0.0015990
Coefficients:
Estimate Std. Error t value Pr(>|t|)
Income    1.675e-03  8.141e-04   2.058   0.0786
Price     3.934e-01  1.607e-01   2.448   0.0442 *
Temp      1.035e-16  5.011e-03   0.000   1.0000
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

Residual standard error: 0.004872 on 7 degrees of freedom, Multiple R-squared: 0.9853, Adjusted R-squared: 0.9789 F-statistic: 155.9 on 3 and 7 DF, p-value: 9.017e-07 From the above coefficient we can easily predicated equation of regression line consumption= intercept+(1.675e-03)*income+(3.934e-01)*price+(1.035e-16)*temp + e (residual value: the lowest will be taken when regression will high or vice versa).

The price variable followed by * indicates a more significant relationship means that there are more ice cream sales when customer prices have risen. If the intersection indicates -ve with consumption, there was -ve with the dependent variable. The residual value is much lower (0.00.487 thousand indicates no much impact on residual values, similarly, the multiple square R value and R2 adjusted value suggested that 0.98% trust that the model satisfied the data and the value p (9.017 e-07) If <0.05 indicates that we can reject the null hypothesis, then we conclude that there is no relation between the dependent and independent variables, the value of t is greater than the independent variable means that there was a positive correlation between the variables, therefore, it is concluded that the salary increases 001.675 when a unitary change in income in a similar way, the ice cream will increase by 0.3934.

Logistic regression is a statistical method used to analyze one or more independent variables. In this type, the dependent variables in the binary data are encoded as 1 for TRUE and 0 for FALSE (dichotomous characteristics). The goal of logistic regression is to find the most appropriate model to describe the relationship between dichotomous characteristics and the set of independent variables. Logistic regression generates formula coefficient to predict a logit transformation of interest probability with formula logit (p) = B0 + b1x1 + B2X2 + B3X3 + BnXn where p is the probability of the presence of the characteristics. Logit transformation is defined as registered probabilities: Quota = (p / p-1) and logit (p) = ln (p / 1-p). However, the process requires data preparation, identification of derived variables,

classification and continuous diagnosis of variable model. Logistic regression allows you to determine the probability that an event is acquired in binary format, which gives an S line shape while linear regression produces a straight line.

The Table No 3 stores the record of students who qualify college, GPA, range admit two categorical variables (yes for admission and zero for student records similar column interval not supported includes four categories 1,2,3,4 description of categories of factor type.

Table No 3: Find the Regression

admit	grade	Gpa	Rank
0	380	3.6	3
1	660	3.67	3
1	800	4	1
1	640	3.19	4
0	800	2.93	4
1	640	2.93	2
1	520	3.8	1
0	760	3.39	2
1	700	3.92	3
0	400	3.6	2
0	700	3.92	3
1	440	3.22	3
1	800	4	1
1	700	4	1
1	700	3.92	3
0	400	3.6	2
0	700	3.92	3
1	440	3.22	3
1	800	4	1
1	700	4	1

Using R Programming

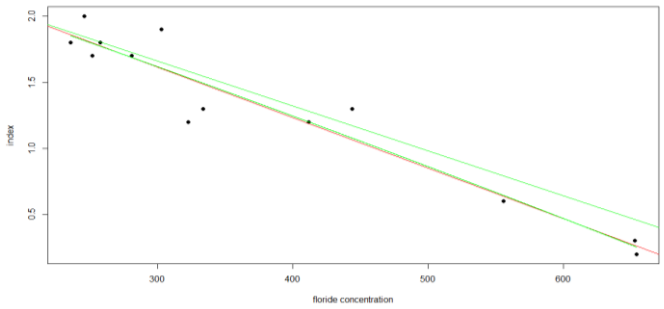
```
> last <- read_excel("C:/Users/Yagya/ Desktop/last.xlsx")
> View(last)
> head(last)
# A tibble: 6 x 4
  admit grade   Gpa  Rank
  <dbl> <dbl> <dbl> <dbl>
1    0.  380.  3.60    3.
2    1.  660.  3.67    3.
3    1.  800.  4.00    1.
4    1.  640.  3.19    4.
5    0.  800.  2.93    4.
6    1.  640.  2.93    2.
> summary(last)
     admit         grade          Gpa           Rank
 Min.   :0.00   Min.   :380   Min.   :2.930   Min.   :1.0
 1st Qu.:0.00   1st Qu.:500   1st Qu.:3.348   1st Qu.:1.0
 Median :1.00   Median :700   Median :3.735   Median :2.5
 Mean   :0.65   Mean   :634   Mean   :3.642
> str(last)
Classes 'tbl_df', 'tbl' and 'data.frame':          20 obs. of  4
variables:
 $ admit: num  0 1 1 1 0 1 1 0 1 0 ...
 $ grade: num  380 660 800 640 800 640 520 760
 $ Gpa  : num  3.6 3.67 4 3.19 2.93 2.93 3.8 3.39
 $ Rank : num  3 3 1 4 4 2 1 2 3 2 ...
> last$admit=as.factor(last$admit)# converting
dicitonomous
> last$Rank=as.factor(last$Rank) # converting multilevel
categorical variable
> summary(last)
 admit     grade         Gpa       Rank
 0: 7   Min.   :380   Min.   :2.930   1:6
 > xtab(~admit+Rank,data=last)
Error in xtab(~admit + Rank, data = last) :
  could not find function "xtab"
> xtabs(~admit+Rank,data=last)# which display the two
way table of admit and rank data
     Rank
admit 1 2 3 4
    0 0 3 3 1
    1 6 1 5 1
> fit= glm(admit ~ grade + Gpa + Rank, data=last,family =
binomial)
> summary(fit)
Call:
glm(formula = admit ~ grade + Gpa + Rank, family =
binomial,  data = last)
Deviance Residuals:
 Min       1Q   Median       3Q      Max
-1.56462  -0.40073   0.00008   0.53338   1.56462
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)3.571e+01  4.375e+03  .008  0.993
grade    3.348e-03  5.969e-03   0.561    0.575
Gpa      -4.677e+00  3.207e+00  -1.458    0.145
Rank2   -2.346e+01  4.375e+03  -0.005    0.996
Rank3   -1.983e+01  4.375e+03  -0.005    0.996
Rank4   -2.381e+01  4.375e+03  -0.005    0.996
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 25.898  on 19  degrees of freedom Residual
deviance: 14.793  on 14  degrees of freedom AIC: 26.793
```

Number of Fisher Scoring iterations: 18. The deviance residual displays minimum, maximum, median and quartiles values, the logistic coefficient displays the data of intercepts of 3.571e+01 the grade has positive impact on when each unit changes of grade changes by log odd 3.348e-03 but Gpa grade had –ve relation when one unit change by -4.677e+00 in every changes similarly the three Rank categories have –ve of log odd with rank 1 category respectively. The AIC value is fit between two or more models. If you don't set Rank variable as factor it will display its category too.

```
> head(fit$fitted)
> fitt=round(fit$fitted)
> comparee=ftable(fitt,last$admit)
> accuracy=sum(diag(comparee,last$admit)
> acuracy=sum(diag(comparee))
> accuracy
> acuracy/20*100
[1] 75
```

Polynomial Regression Using R Programming

```
> floride <-
read_excel("C:/Users/Yagya/Desktop/floride.xlsx")
> View(floride)
> x=floride$Dmfper100
> y=floride$FlouridePPM
>xsq=x^2
>xcub=x^3
>xqua=x^4
> install.packages("ggplot")
> plot(x,y, pch=19,xlab="floride
concentration",ylab="index")
>fit1=lm(y~x)
> anova(fit1)
Analysis of Variance Table
Response: y
Df Sum Sq Mean Sq F value   Pr(>F)
x         1 8.1047  8.1047  288.69 6.99e-15 ***
Residuals 24 0.6738  0.0281
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' Null
hypothesis has no relationship
fit2=lm(y~x+xsq)
> anova(fit2)
Analysis of Variance Table
Response: y
Df Sum Sq Mean Sq  F value    Pr(>F)
x         1 8.1047  8.1047 277.2516 2.518e-14 ***
xsq       1 0.0014  0.0014   0.0492    0.8264
Residuals 23 0.6723  0.0292
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
> abline(fit1, col="red")
> abline(fit2, col="green")
> abline(fit1, col="red")
> xv=seq(min(x),max(x),0.01)
> yv=predict(fit2,list(x=xv,xsq=xv^2))
> lines(xv,xy,col="black")
```

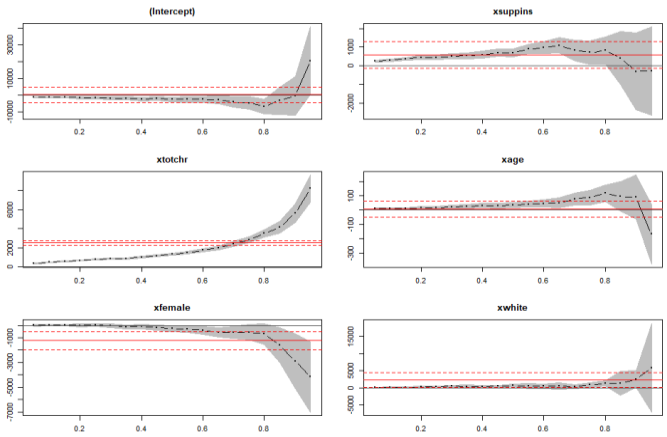


Quantile Regression in R

The quantile regression needs install **quantreg** package in order to carry out quantile regression. install.packages("quantreg") library(quantreg). The **rq** function try to predict the estimate the 25th quantile of Quantile_health data sets having 2955 obs. of 8 variables where **tau = 0.25 is** median regression. Similarly, it can run quantile regression for multiple quantiles in a single plot using sequence parameter. We can check whether our quantile regression results differ from the OLS results using plots.

```
> str(mydata)
Classes 'tbl_df', 'tbl' and 'data.frame':          2955 obs. of  8
variables:
$ dupersid: num  93193020 72072017 25296013
 ……..
 $ white   : num  1 1 1 1 1 1 1 1 1 1 ...
> y=cbind(totexp)
> attach(mydata)
> y=cbind(totexp)
> x=cbind(suppins,totchr,age,female,white)
> qreg=lm(y~x,data=mydata)
> summary(qreg)
Call:
lm(formula = y ~ x, data = mydata)
Residuals:
Min     1Q Median     3Q    Max
-16146  -5372  -2804    457 115461
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   461.492   2777.453   0.166  0.86805
xsuppins      585.984    436.309   1.343  0.17936
xtotchr      2528.079    164.834  15.337  < 2e-16 ***
xage            6.711     33.768   0.199  0.84248
xfemale     -1239.866    433.110  -2.863  0.00423 **
Residual standard error: 11520 on 2949 degrees of
freedom. Multiple R-squared:  0.07828,       Adjusted R-
squared:  0.07672   F-statistic: 50.09 on 5 and 2949 DF,  p-
value: < 2.2e-16
> install.packages("quatrile")
> install.packages("quantreg")
> install.packages("quantreg")
> summary(rq (y ~ x, data=mydata,tau=0.25,method='fn'),
se='ker')
Call: rq(formula = y ~ x, tau = 0.25, data = mydata, method
= "fn")
tau: [1] 0.25
Coefficients:
Value      Std. Error  t value     Pr(>|t|)
```

```
(Intercept) -1412.88 709.49206 -1.9914   .0465
xsuppins     453.44  113.929    3.98004  .0007
……..
xwhite       338.08  283.93954    1.199    0.233
> summary(rq (y ~ x, data=mydata,tau=0.5,method='fn'),
se='ker')
Call: rq(formula = y ~ x, tau = 0.5, data = mydata, method =
"fn")
tau: [1] 0.5
Coefficients:
Value      Std. Error  t value    Pr(>|t|)
(Intercept) -2182.33238   993.25180    -2.19716
xsuppins     687.22222   159.39319    4.31149
……..
xwhite       562.66571   418.05964    1.34590
> summary(rq (y ~ x, data=mydata,tau=0.75,method='fn'),
se='ker')
Call: rq(formula = y ~ x, tau = 0.75, data = mydata, method
= "fn")
tau: [1] 0.75
Coefficients:
Value      Std. Error  t value    Pr(>|t|)
(Intercept) -4512.04545  2120.58771    -2.12773
xsuppins     708.40909   336.40195    2.10584
……..
xwhite       801.68182   625.02743    1.28263
> q1=summary(rq (y ~ x,
data=mydata,tau=0.25,method='fn'), se='ker')
> q2=summary(rq (y ~ x,
data=mydata,tau=0.5,method='fn'), se='ker')
> q3=summary(rq (y ~ x,
data=mydata,tau=0.75,method='fn'), se='ker')
> regall=rq(y~x,tau=seq(0.05,0.95, by= .05),data=mydata)
> regall
rq(formula = y ~ x, tau = seq(0.05, 0.95, by = 0.05), data =
mydata)
Coefficients:
tau= 0.05 tau= 0.10  tau= 0.15    tau= 0.20   (Intercept)   -
1008.0   -932.75 -989.10345 -1315.133333 -1412.88889 -
1702.375
xsuppins      247.4    301.15  371.81034   446.600000
453.44444   506.125
……..
(Intercept) -1707.65766 -2155.72941     -1953 -
xsuppins     541.73874   596.94118      702
……..
tau= 0.65   tau= 0.70   tau= 0.75  tau= 0. (Intercept) -
2605.22222 -3906.57143 -4512.04545 -6584.4118 -
390.5366
xsuppins    1094.23333   833.28571   708.40909
……..
xwhite       464.94444   340.14286   801.68182
Degrees of freedom: 2955 total; 2949 residual
>regall=summary(rq(y~x,tau=seq(0.05,0.95, by=
.05),data=mydata))
>plot(regall)
```



Numerous quantiles are represented by the X axis. The red center line indicates the estimates of the OLS coefficients and the lines of the red dot are the confidence intervals around those OLS coefficients for different quantiles. The black point line is the quantile regression estimate and the gray area is the confidence interval for them for different quantiles. We can see that, for the whole variable, it is the regression match estimated for most of the quantiles. Therefore, our use of quantile regression is not justifiable for these quantiles. In other words, we want both the red and the gray lines to overlap as little as possible to justify the use of quantile regression.

CONCLUSIONS

In statistics, linear regression models the relationship between a dependent variable and one or more explanatory variables using a linear function. If two or more explanatory variables have a linear relation to the dependent variable, the regression is called multiple linear regression. Multiple regression, on the other hand, is a larger class of regressions that includes linear and nonlinear regressions with multiple explanatory variables. Regression analysis is a common way to discover a relationship between dependent and explanatory variables. However, the statistical relationship does not mean that the explanatory variables cause the dependent variable; It can be a significant association in the data. Linear regression attempts to draw a line closer to the data by identifying the slope and intersection that define the line and minimizing regression errors. However, many relationships in the data do not follow a straight line, so statisticians use non-linear regression instead. The relationship between the search variables is considered a correlation, it is a number that can be used to describe the degree of association between them. The correlation is between -1 and 1 and expresses the relative variation between the search variables. Multiple correlation and partial correlation are classified as correlated variations between three or more variables. Two variables are correlated only when they vary in such a way that the higher and lower values of a variable correspond to the higher and lower

values of the other variable. We may also know if they are related when the highest value of a variable corresponds to the lowest value of the other. The research data could be easily analyzed using the R programming instead of the mathematical calculation.

REFERENCES

[1] J. Manyika, "Big data: The next frontier for innovation, competition," 2011.

[2] G. Fiona, "Overfitting regression analysis," www.bmj.com, 2018.

[3] D. M. Astrid Schneider, "Linear regression analysis. medicine," University of Multiplier, 2010.

[4] Y. Rimal, "Cross- validation method for overtting research data using r programming," ICC, 2018.

[5] A. Blokhin, "What is the difference between linear regression and multiple regression," University of Multiplier, 2018.

[6] F. Chen, "Principles of quantile regression and an application," University of Multiplier, 2013.

[7] W. Rogers, "The implementation of model formula by ross ihaka was based s," University of Multiplier, 1973.

[8] B. L. Cook, "Thinking beyond the mean: a practical guide for using quantile regression methods for health services research," Shanghai Arch Psychiatry, 2013.

[9] B. Magnensi, "Pls reglm algorithm insight," 2014.

[10] D. Smith, "Revolutionary analytics," Blog, Revolutionary Analytics, 2018.

[11] B. HelegMevik, "Introduction to pls packages," 2