PREDICT THE COUNT OF RENTAL BIKES

Aswith Reddy Kovvuri,Yogesh Savirigana, Dilli Prasad

**PERSONS INVOLVED IN THIS GROUP:**

Aswith Reddy Kovvuri (U04778898)

Yogesh Savirigana (U17620025)

Dilli Prasad Sigirala (U42397312)

**DATA SET SOURCE**: The dataset includes the number of public bikes hired at each hour in the Seoul Bike Sharing System, as well as temperature and functioning hours. The dataset was taken from <http://data.seoul.go.kr/> and the data was collected in Seoul from the local bike rentals to maintain enough rental bikes for public in that city.

**ABOUT DATASET:** Many large cities are already offering rental bikes to increase mobility convenience. Delivering a sufficient quantity of rental bikes to the city eventually becomes a big challenge. It is essential to have the rental bike ready and open to the public at the right moment to save waiting time. Predicting the volume of bikes required at each hour is crucial to ensuring a consistent supply of rental bikes.

rm(list=ls())  
library(rio)  
library(moments)  
bikes\_data=import("Bike data.xlsx")  
colnames(bikes\_data)=tolower(make.names(colnames(bikes\_data)))   
attach(bikes\_data)  
names(bikes\_data)

## [1] "date" "rented.bike.count" "hour"   
## [4] "temperature" "humidity" "wind.speed"   
## [7] "visibility" "dew.point.temperature" "solar.radiation"   
## [10] "rainfall" "snowfall" "seasons"   
## [13] "holiday" "functioningday"

**VARIABLES IN DATA SET**: Rented Bike count - Count of bikes rented at each hour (Target variable)  
Hour - Hour of the day (Continuous variable)  
Temperature-Temperature in Celsius (Continuous variable)  
Humidity - % (Continuous variable)  
Windspeed - m/s (Continuous variable)  
Visibility - 10m (Continuous variable)  
Dew point temperature – Celsius (Continuous variable)  
Solar radiation - MJ/m2 (Continuous variable)  
Rainfall – mm (Continuous variable)  
Snowfall – cm (Continuous variable)  
Seasons - Winter, Spring, Summer, Autumn  
Holiday - Holiday/No holiday(binary)  
Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)(binary)

Taking sample of 100 observations

set.seed(123456)  
#taking sample of 100 observations  
bikes\_data\_sample=bikes\_data[sample(1:nrow(bikes\_data),100),]  
attach(bikes\_data\_sample)

## The following objects are masked from bikes\_data:  
##   
## date, dew.point.temperature, functioningday, holiday, hour,  
## humidity, rainfall, rented.bike.count, seasons, snowfall,  
## solar.radiation, temperature, visibility, wind.speed

continous\_factors <- bikes\_data[, c("rented.bike.count","hour","temperature",  
 "humidity","wind.speed","visibility",  
 "dew.point.temperature","solar.radiation","rainfall",  
 "snowfall")]

Checked correlation matrix for the continuous variables and target and picked only two continuous variables among them.

corelation\_data= cor(continous\_factors)  
corelation\_data

## rented.bike.count hour temperature humidity  
## rented.bike.count 1.0000000 0.410257291 0.53855815 -0.1997802  
## hour 0.4102573 1.000000000 0.12411449 -0.2416438  
## temperature 0.5385582 0.124114492 1.00000000 0.1593708  
## humidity -0.1997802 -0.241643787 0.15937080 1.0000000  
## wind.speed 0.1211084 0.285196660 -0.03625170 -0.3366830  
## visibility 0.1992803 0.098753482 0.03479443 -0.5430903  
## dew.point.temperature 0.3797881 0.003054372 0.91279822 0.5368945  
## solar.radiation 0.2618370 0.145130920 0.35350547 -0.4619188  
## rainfall -0.1230740 0.008714642 0.05028186 0.2363967  
## snowfall -0.1418036 -0.021516455 -0.21840486 0.1081835  
## wind.speed visibility dew.point.temperature  
## rented.bike.count 0.121108448 0.19928030 0.379788121  
## hour 0.285196660 0.09875348 0.003054372  
## temperature -0.036251701 0.03479443 0.912798219  
## humidity -0.336683042 -0.54309034 0.536894494  
## wind.speed 1.000000000 0.17150714 -0.176485692  
## visibility 0.171507137 1.00000000 -0.176629730  
## dew.point.temperature -0.176485692 -0.17662973 1.000000000  
## solar.radiation 0.332274246 0.14973803 0.094381345  
## rainfall -0.019674089 -0.16762924 0.125596737  
## snowfall -0.003554186 -0.12169451 -0.150886707  
## solar.radiation rainfall snowfall  
## rented.bike.count 0.26183699 -0.123073960 -0.141803650  
## hour 0.14513092 0.008714642 -0.021516455  
## temperature 0.35350547 0.050281859 -0.218404862  
## humidity -0.46191880 0.236396670 0.108183453  
## wind.speed 0.33227425 -0.019674089 -0.003554186  
## visibility 0.14973803 -0.167629238 -0.121694515  
## dew.point.temperature 0.09438135 0.125596737 -0.150886707  
## solar.radiation 1.00000000 -0.074290110 -0.072300823  
## rainfall -0.07429011 1.000000000 0.008499653  
## snowfall -0.07230082 0.008499653 1.000000000

library(corrplot)

## corrplot 0.92 loaded

corrplot(corelation\_data,method="number")

Chart

Description automatically generated

is.factor(bikes\_data\_sample$functioningday)

## [1] FALSE

bikes\_data\_sample$functioningday=as.factor(bikes\_data\_sample$functioningday)  
is.factor(bikes\_data\_sample$functioningday)

## [1] TRUE

As correlation value with the rented.bike.count(target) is more for hour and temperature, the below variables are considered for linear model hour(hour of the day)- X1  
temperature-X2  
Functional Day-X3  
rented.bike.count-Y(target )

#Continous variables X1, X2 are X1=hour and X2=temperature  
#Binary variable is X3=FunctioningHour  
#Y is the target, Y=RentedBikesCount

Simple Regression model for rented.bike.count(target-Y)- hour(X1):

Y\_X1 =lm(rented.bike.count~hour,data=bikes\_data\_sample)  
summary(Y\_X1)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour, data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1001.4 -377.8 -119.7 333.6 1530.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 268.398 112.327 2.389 0.0188 \*   
## hour 34.904 8.394 4.158 6.88e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 590.9 on 98 degrees of freedom  
## Multiple R-squared: 0.15, Adjusted R-squared: 0.1413   
## F-statistic: 17.29 on 1 and 98 DF, p-value: 6.881e-05

**ANALYSIS:**

• Multiple R-squared value for this model is 0.15 which means that hour of the day explains only 15% of the count of rental bikes  
• P value is 6.881e-05 which is significant (<0.05)  
• The linear equation is **rented.bike.count = 34.904\* hour + 268.398**  
• It means if the hour increases by 1 then the count of rental bikes increases by 34.

Simple Regression model for rented.bike.count(target-Y)- temperature(X2):

Y\_X2=lm(rented.bike.count~temperature,data=bikes\_data\_sample)  
summary(Y\_X2)

##   
## Call:  
## lm(formula = rented.bike.count ~ temperature, data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -920.09 -341.84 -45.03 235.01 1323.81   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 272.273 78.641 3.462 0.000796 \*\*\*  
## temperature 31.212 4.616 6.762 9.9e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 529.2 on 98 degrees of freedom  
## Multiple R-squared: 0.3181, Adjusted R-squared: 0.3112   
## F-statistic: 45.73 on 1 and 98 DF, p-value: 9.904e-10

**ANALYSIS:**

• Multiple R-squared value for this model is 0.31 which means that temperature explains only 31% of the count of rental bikes  
• P value is 9.9e-10 which is significant(<0.05)  
• The linear equation is **rented.bike.count = 31.212\* temperature + 272.273**  
• It means if the temperature increases by 1 then there is a count of rental bikes increases by 31.

Summary of Simple Regression model for rented.bike.count(target-Y)- functioning day(X3):

Y\_X3=lm(rented.bike.count~functioningday,data=bikes\_data\_sample)  
summary(Y\_X3)

##   
## Call:  
## lm(formula = rented.bike.count ~ functioningday, data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -683.6 -510.6 -177.1 266.6 1691.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.002e-12 2.782e+02 0.000 1.0000   
## functioningdayYes 7.006e+02 2.854e+02 2.455 0.0159 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 622.1 on 98 degrees of freedom  
## Multiple R-squared: 0.05792, Adjusted R-squared: 0.04831   
## F-statistic: 6.025 on 1 and 98 DF, p-value: 0.01587

**ANALYSIS:**

#Multiple R-squared value for this model is 0.05 which means that binary variable functioning day explains only 5% of the count of rental bikes  
• P value is 0.01587 which is significant(<0.05)  
• The linear equation is **rented.bike.count = (7.006e+02)\* functioningday + 4.002e-12**  
• It means if the functioningday increases by 1 then there is a count of rental bikes increases by 7.006e+02.  
• R-squared value is very less for this model when compared with the simple linear model of other continuous variables.

Multiple Regression model of Y,X1,X2( rented.bike.count(Y)- hour(X1), temperature(X2)):

Y\_X1\_X2= lm(rented.bike.count~hour+temperature,data=bikes\_data\_sample)  
summary(Y\_X1\_X2)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour + temperature, data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1149.73 -282.99 -38.29 229.80 1194.20   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -22.345 103.244 -0.216 0.829105   
## hour 28.524 7.056 4.042 0.000106 \*\*\*  
## temperature 28.833 4.332 6.656 1.69e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 492.1 on 97 degrees of freedom  
## Multiple R-squared: 0.4165, Adjusted R-squared: 0.4044   
## F-statistic: 34.61 on 2 and 97 DF, p-value: 4.515e-12

**ANALYSIS:**

• Multiple R-squared value for this model is 0.41 which means that continuous variables (hour and temperature)explains only 41% of the count of rental bikes  
• P values of continuous variables, temperature and hour are significant(<0.05)  
• The Multi Regression equation is **rented.bike.count = 28.833\* temperature + 28.524\* hour -22.345**  
• It means if the hour increases by 1 then the count of rental bikes increases by 28.524 and when temperature increases by 1 then count of rental bikes increases by 28.833

Multiple Regression model of Y,X2,X3( rented.bike.count(Y)- temperature(X2),functioningday(X3)):

Y\_X2\_X3= lm(rented.bike.count~temperature+functioningday,data=bikes\_data\_sample)  
summary(Y\_X2\_X3)

##   
## Call:  
## lm(formula = rented.bike.count ~ temperature + functioningday,   
## data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -970.36 -287.44 -45.97 195.72 1269.03   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -518.549 234.538 -2.211 0.029390 \*   
## temperature 32.288 4.374 7.381 5.4e-11 \*\*\*  
## functioningdayYes 818.171 230.135 3.555 0.000586 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 500.4 on 97 degrees of freedom  
## Multiple R-squared: 0.3968, Adjusted R-squared: 0.3843   
## F-statistic: 31.9 on 2 and 97 DF, p-value: 2.259e-11

**ANALYSIS:**

• Multiple R-squared value for this model is 0.39 which means that continuous variable (temperature), binary variable(functioning day)eplains only 39% of the count of rental bikes  
• P values of temperature is 5.4e-11 and p-value of functioning day is 0.000586 which is significant(<0.05)  
• The Multi Regression equation is **rented.bike.count = 32.288 \* temperature + 818.171 \* functioningday -518.549**  
• It means if the temperature increases by 1 then the count of rental bikes increases by 32 and when it is functioning day then count of rental bikes increases by 818

Multiple Regression model of Y,X1,X3( rented.bike.count(Y)- hour(X1),functioningday(X3)):

Y\_X1\_X3= lm(rented.bike.count~hour+functioningday,data=bikes\_data\_sample)  
summary(Y\_X1\_X3)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour + functioningday, data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1031.5 -360.2 -124.6 323.8 1486.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -460.612 275.250 -1.673 0.09746 .   
## hour 35.985 8.107 4.439 2.39e-05 \*\*\*  
## functioningdayYes 754.431 261.845 2.881 0.00488 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 570.1 on 97 degrees of freedom  
## Multiple R-squared: 0.217, Adjusted R-squared: 0.2008   
## F-statistic: 13.44 on 2 and 97 DF, p-value: 7.046e-06

**ANALYSIS:**

• Multiple R-squared value for this model is 0.217 which means that continuous variable (hour), binary variable(functioning day) explains only 21% of the count of rental bikes  
• P values of hour is 2.39e-05 and p-value of functioning day is 0.00488 which are significant(<0.05)  
• The Multi Regression equation is **rented.bike.count = 35.985 \* hour + 754.431\* functioningday -460.612**  
• It means if the hour increases by 1 then the count of rental bikes increases by 35 and when it is functioning day then count of rental bikes increases by 754

Multiple Regression model of Y,X1,X2,X3( rented.bike.count(Y)- hour(X1),temperature(X2), functioningday(X3)):

Y\_X1\_X2\_X3= lm(rented.bike.count~hour+temperature+functioningday,data=bikes\_data\_sample)  
summary(Y\_X1\_X2\_X3)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour + temperature + functioningday,   
## data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1210.17 -260.38 -55.78 238.21 1132.53   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -857.568 227.089 -3.776 0.000276 \*\*\*  
## hour 29.517 6.558 4.501 1.90e-05 \*\*\*  
## temperature 29.873 4.032 7.410 4.93e-11 \*\*\*  
## functioningdayYes 853.498 210.357 4.057 0.000101 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 457 on 96 degrees of freedom  
## Multiple R-squared: 0.5019, Adjusted R-squared: 0.4863   
## F-statistic: 32.24 on 3 and 96 DF, p-value: 1.673e-14

**ANALYSIS:**

• Multiple R-squared value for this model is 0.507 which means that continuous variable (hour, temperature), binary variable(functioning day) explains only 50% of the count of rental bikes  
• P values of hour is 1.90e-05, temperature is 4.93e-11 and p-value of functioning day is 0.000101 which are significant(<0.05)  
• The Multi Regression equation is **rented.bike.count = 29.517\* hour + 29.873*temperature+ 853.498*  functioningday -857.568**  
• It means if the hour increases by 1 then the count of rental bikes increases by 29 and when it is functioning day then count of rental bikes increases by 853 and when temperature increases by 1 then count of rental bikes increases by 29.

Multiple Regression model of Y,X1,X2 and X1.X2( rented.bike.count(Y)- hour(X1),temperature(X2)):

Y\_X1\_X2\_X1.X2= lm(rented.bike.count~hour+temperature+(hour\*temperature),data=bikes\_data\_sample)  
summary(Y\_X1\_X2\_X1.X2)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour + temperature + (hour \*   
## temperature), data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1282.87 -236.59 -46.86 230.54 1077.73   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 184.779 121.519 1.521 0.13166   
## hour 8.944 9.480 0.944 0.34777   
## temperature 8.793 7.951 1.106 0.27153   
## hour:temperature 1.738 0.587 2.960 0.00388 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 473.5 on 96 degrees of freedom  
## Multiple R-squared: 0.4652, Adjusted R-squared: 0.4485   
## F-statistic: 27.84 on 3 and 96 DF, p-value: 4.859e-13

Analysis: • Multiple R-squared value for this model is 0.4652 which means that continuous variable (hour, temperature), (hourtemperature) explains only 46% of the count of rental bikes  
• P values of hour is 0.34777, temperature is 0.27153 which are not significant. So, their values are not in confidence interval and p-value of hourtemperature is 0.00388 which is significant(<0.05)  
• The Multi Regression equation is **rented.bike.count = 8.944\* hour + 8.793*temperature+ 1.738* hour\*temperature - 184.779**  
• It means if the hour increases by 1 then the count of rental bikes increases by around 8 when temperature increases by 1 then count of rental bikes increases by around 8.

Multiple Regression model of Y,X1 and (X1)^2 ( rented.bike.count(Y)- hour(X1),(hour^2)):

Y\_X1\_X1.X1= lm(rented.bike.count~hour+I(hour^2),data=bikes\_data\_sample)  
summary(Y\_X1\_X1.X1)

##   
## Call:  
## lm(formula = rented.bike.count ~ hour + I(hour^2), data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -910.84 -405.74 -56.03 255.44 1485.48   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 69.053 160.133 0.431 0.66727   
## hour 91.703 33.866 2.708 0.00801 \*\*  
## I(hour^2) -2.496 1.443 -1.730 0.08680 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 585 on 97 degrees of freedom  
## Multiple R-squared: 0.1754, Adjusted R-squared: 0.1584   
## F-statistic: 10.32 on 2 and 97 DF, p-value: 8.659e-05

**ANALYSIS:**

• Multiple R-squared value for this model is 0.17 which means that continuous variable (hour, hour^2) explains only 17% of the count of rental bikes  
• P value of hour is 0.00801, hour^2 is 0.08680 which are significant. So, their values are in confidence interval .  
• The Multi Regression equation is **rented.bike.count = 91.703 \* hour -2.496\*(hour^2)+ 69.053**

Multiple Regression model of Y,X1 and (X1)^2 ( rented.bike.count(Y)- hour(X1),(hour^2)):

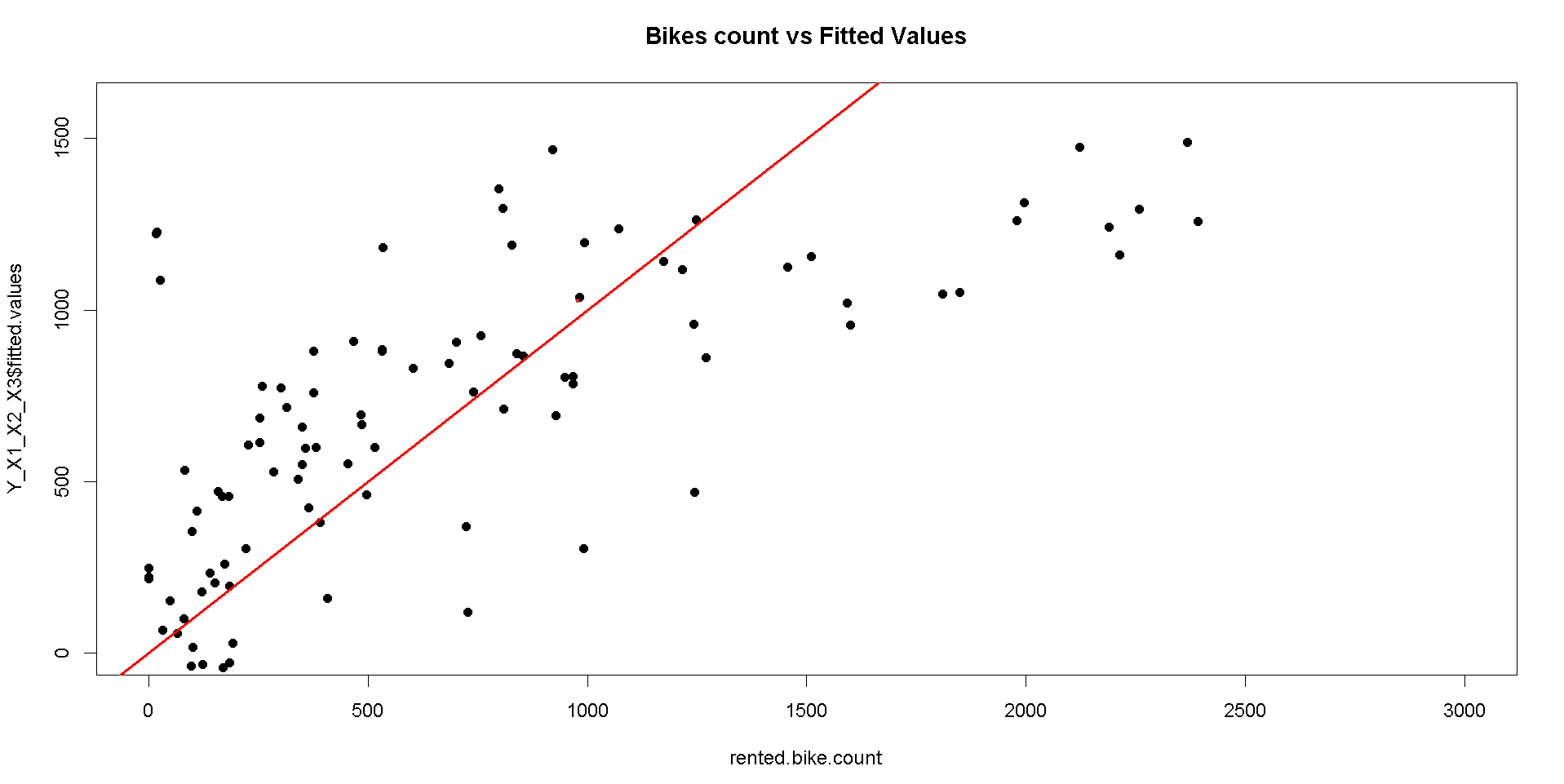
Y\_X2\_X2.X2= lm(rented.bike.count~temperature+I(temperature^2),data=bikes\_data\_sample)  
summary(Y\_X2\_X2.X2)

##   
## Call:  
## lm(formula = rented.bike.count ~ temperature + I(temperature^2),   
## data = bikes\_data\_sample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -908.33 -334.44 -33.15 238.16 1293.14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 267.4258 78.6896 3.398 0.000984 \*\*\*  
## temperature 22.1743 9.4751 2.340 0.021318 \*   
## I(temperature^2) 0.4091 0.3746 1.092 0.277582   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 528.7 on 97 degrees of freedom  
## Multiple R-squared: 0.3264, Adjusted R-squared: 0.3125   
## F-statistic: 23.5 on 2 and 97 DF, p-value: 4.749e-09

**ANALYSIS:**

• Multiple R-squared value for this model is 0.32 which means that continuous variable (temperature, temperature^2) explains only 32% of the count of rental bikes  
• P value of temperature is 0.021318, temperature^2 is 0.277582which are significant. So, their values are in confidence interval.  
• The Multi Regression equation is **rented.bike.count = 22.1743\* temperature - 0.4091\*(temperature^2)+ 267.4258**

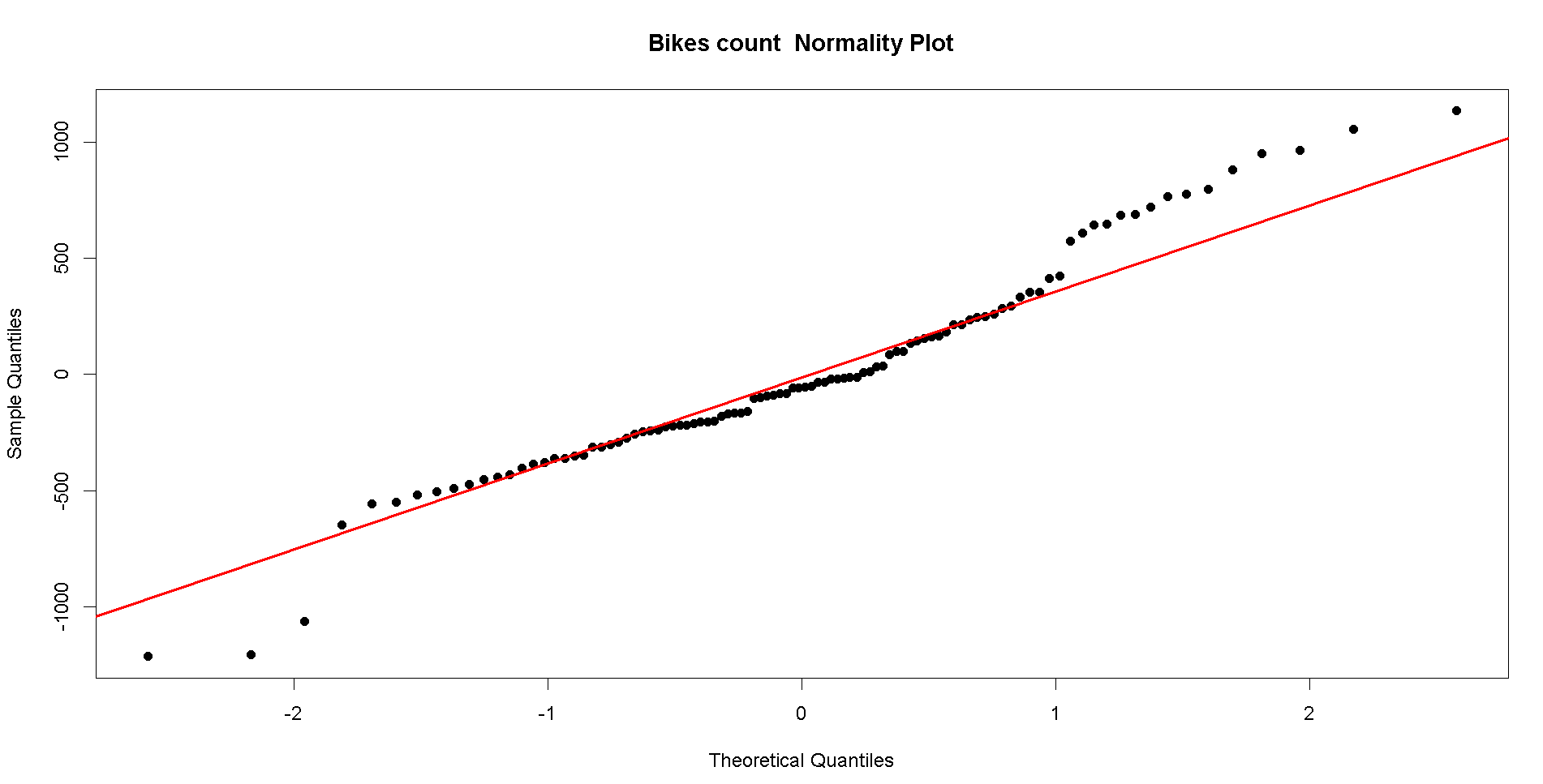
**Good Model:**  
As the P-values of all the factors are less than 0.05 and multiple r-squared value is 0.5019  better for multiple regression model for Y, X1, X2, X3  
After analysing different models, good model is of Y, X1, X2, X3  
The Regression equation of good model is rented.bike.count = 29.517\* hour + 29.873*temperature+ 853.498*  functioningday -857.568  
#Checking Linearity, Normality and Equality for linear model of (Y,X1,X2,X3)

# Linearity  
plot(rented.bike.count,Y\_X1\_X2\_X3$fitted.values,  
 pch=19, xlim=c(0,3000),ylim=c(0,1600),main="Bikes count vs Fitted Values")  
abline(0,1,col="red",lwd=2)  


**ANALYSIS:**

From the above graph, even some data points are away from the red line.  
the Squares of sum of error residuals will be minimum.   So, we can tell that data is linearly distributed

#Normality  
qqnorm(Y\_X1\_X2\_X3$residuals,pch=19,main="Bikes count Normality Plot")  
qqline(Y\_X1\_X2\_X3$residuals,col="red",lwd=2)

  
  
**ANALYSIS:**

From the above graph, even some of the data points are not on red line, the maximum data points fall on the red line.  
So, we can confirm that data is normally distributed

#Equality of variance  
plot(Y\_X1\_X2\_X3$fitted.values,Y\_X1\_X2\_X3$residuals,pch=19,  
 main="cars Residuals")  
abline(0,0,col="red",lwd=2)

Chart, scatter chart

Description automatically generated

**ANALYSIS:**

From the above graph, We can observe that some outliners are far from line and data is asymmetrical, but we can tell that the data is almost equally distributed above and below the line as none of the model is perfect.  
  
  
#prediction 1  
data1 <- data.frame(functioningday=factor("Yes", levels=c("Yes", "No")),   
 hour=12,  
 temperature=8)  
  
predict(Y\_X1\_X2\_X3,data1,interval="predict")  
predict(Y\_X1\_X2\_X3,data1,interval="confidence")

> predict(Y\_X1\_X2\_X3,data1,interval="predict")

fit lwr upr

1 589.1113 -323.6571 1501.88

> predict(Y\_X1\_X2\_X3,data1,interval="confidence")

fit lwr upr

1 589.1113 488.7027 689.5199

**ANALYSIS:**

The number of rental bikes predicted for functionday factor -Yes, hour value 12 and temperature is 8C  
For the above input factors, the number of rental bikes are around 589 and mostly the confidence interval is from 323.6571 to 1501.88.  
Here, the 95% confidence interval is from 488.7027 to 689.5199  
  
#prediction2  
data2 <- data.frame(functioningday=factor("No", levels=c("Yes", "No")),   
 hour=10,  
 temperature=20)  
predict(Y\_X1\_X2\_X3,data2,interval="predict")  
predict(Y\_X1\_X2\_X3,data2,interval="confidence")

> predict(Y\_X1\_X2\_X3,data2,interval="predict")

fit lwr upr

1 35.05158 -960.0895 1030.193

> predict(Y\_X1\_X2\_X3,data2,interval="confidence")

fit lwr upr

1 35.05158 -373.9001 444.0032

**ANALYSIS:**

The number of rental bikes predicted for functionday factor -No, hour = 10 and temperature =20  
For the above input factors, the number of rental bikes are around 35 and mostly the confidence interval is from -960 to 1030.  
Here, the 95% confidence interval is from 0 to 444.