1) Calories\_consumed-> predict weight gained using calories consumed

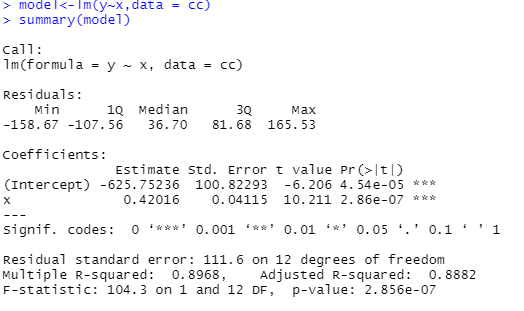
BUSSINESS PROBLEM: To predict relationship between weight gained and calories consumed and build a model

UNIVARIET ANALYSIS

|  |  |
| --- | --- |
| CALORIES CONSUMED  BAR PLOT:    Bimodal plot  HISTOGRAM:    data is right skewed  BOX-PLOT:    No outliers and data is right skewed  QQPLOT    Data is normally distributed and positively skewed | WEIGHT GAINED    Bimodal plot    data is right skewed    No outliers and the data is right skewd    Data is not normally distributed and positively skewed  After transformation data is normally distributed and positively skewd |

BIVARIENT ANLYSIS

|  |
| --- |
| Correlation coefficient  cor(x,y)=0.94  SCATTER PLOT |



Checking for normality of errors

|  |
| --- |
| > shapiro.test(model$residuals)  Shapiro-Wilk normality test  data: model$residuals  W = 0.90951, p-value = 0.155  As the errors are normally distributed and multiple and adjusted r2 values are greater  than 0.85 n close enough hence it’s a right fit model.  Error=103.5 |
|  |
| |  | | --- | |  | |

PYTHON CODE

import pandas as pd

import numpy as np

cc=pd.read\_csv("C:/Users/USER/Desktop/simple\_reg\_assgnment/assignments/cc.csv")

cc.columns="y","x"

import matplotlib.pylab as plt

####one variet analysis#####

plt.hist(cc.x)

plt.hist(cc.y)

plt.boxplot(cc.x)

plt.boxplot(cc.y)

cc.describe()

####bi variet analysis######

import scipy

from scipy import stats

stats.cor(cc.x,cc.y)

np.corrcoef(cc.x,cc.y)

plt.scatter(cc.x,cc.y)

import statsmodels.formula.api as smf

model=smf.ols("y~x",data=cc).fit()

model.summary()

pred=model.predict(pd.DataFrame(cc['x']))

pred

err=pred-cc.y

err\_sq=err\*err

err\_mean=np.mean(err\_sq)

err\_sqrt=np.sqrt(err\_mean)

err\_sqrt