Project Stat139 - Fall

**Topic:**

Create a multiple regression model for debt amount after finishing from college. The model is limited to public institution

**The Fields for the Analysis**

**Response Variable**

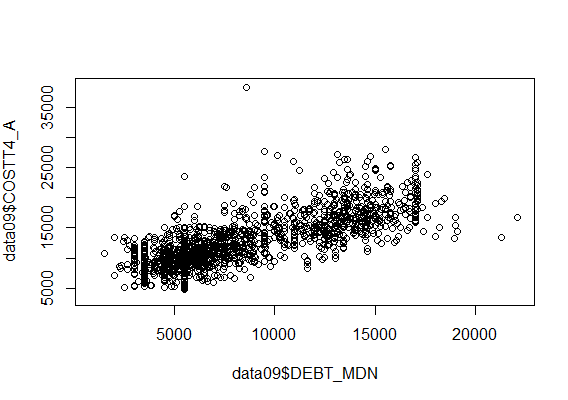
**DEBT\_MDN** - The original amount of the loan principal upon entering repayment

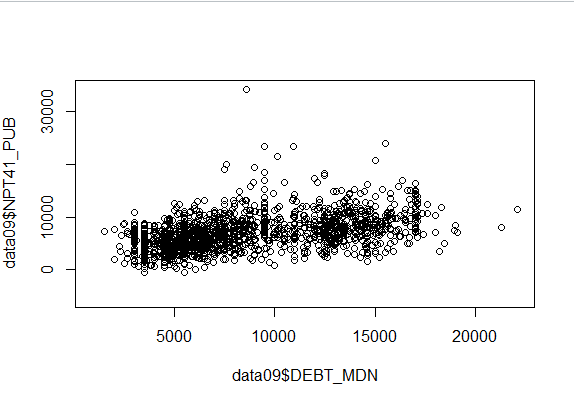
**Predictors**

**NPT41\_PUB** - Average net price for $0-$30,000 family income (public institutions)

**COSTT4\_A** - The average annual total cost of attendance

Tests for linearity between the predictors & the response variable –





Both the predictors are linear to the response variable.

So we can create the multiple regression model using these predictors

‘R’ Code

data2009=read.csv(file.choose(),header=T, stringsAsFactors = FALSE)

# Getting all public institution from the dataset – CONTROL=1

data2009=subset(data2009, data2009$CONTROL==1)

# Keep only the relevant fields in the data frame data09

data09=data2009[1:nrow(data2009),c('INSTNM','PCTFLOAN','CONTROL','C150\_4','CDR2','COSTT4\_A','STABBR','AVGFACSAL','RPY\_1YR\_RT','RPY\_3YR\_RT','RPY\_1YR\_N','GRAD\_DEBT\_MDN','PCIP16','ADM\_RATE','WDRAW\_DEBT\_MDN','md\_earn\_wne\_p6','UGDS\_WHITE','UGDS\_BLACK','UGDS\_HISP','UGDS\_ASIAN','NPT41\_PUB','NPT42\_PUB','NPT43\_PUB','NPT44\_PUB','NPT45\_PUB','SATMTMID','SATVRMID','SATWRMID','faminc','DEBT\_MDN')]

data09$DEBT\_MDN = as.numeric(data09$DEBT\_MDN)

data09$NPT41\_PUB <- as.numeric(data09$NPT41\_PUB)

data09$COSTT4\_A = as.numeric(data09$COSTT4\_A)

sampletest<-data.frame(data09$NPT41\_PUB,data09$COSTT4\_A,data09$DEBT\_MDN)

sampletest <- na.omit(sampletest)

options("scipen"=100, "digits"=4)

data09.lm<-lm(data09.DEBT\_MDN~data09.COSTT4\_A+data09.NPT41\_PUB, data=sampletest)

**Summary of the model**

summary(data09.lm)

Call:

lm(formula = data09.DEBT\_MDN ~ data09.COSTT4\_A + data09.NPT41\_PUB,

data = sampletest)

Residuals:

Min 1Q Median 3Q Max

-13627 -1615 -266 1548 12704

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1176.3218 219.6680 -5.35 0.000000099 \*\*\*

data09.COSTT4\_A 0.9764 0.0253 38.56 < 0.0000000000000002 \*\*\*

data09.NPT41\_PUB -0.4098 0.0326 -12.58 < 0.0000000000000002 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2650 on 1519 degrees of freedom

Multiple R-squared: 0.597, Adjusted R-squared: 0.597

F-statistic: 1.13e+03 on 2 and 1519 DF, p-value: <0.0000000000000002

**Both the predictors are highly significant.**

**As the f-statistic is less than the significance level of .05 we can say this model is significant.**

**The regression Fit equation**

Debt\_Amount = -1176.32 + 0.9764 (Cost per year) - 0.4098 (household-income less than 30k)