

# SS 3859 Assignment

*Jihye Park*

**2018-11-05**

a) In order to produce summary statistics of the response variable and the explanatory variables, I had to download the NAICExpense data from Jed Frees' website. I set my working directory to the folder that csv file in. And the summary, I will use the codes below.

```
setwd("/users/caroline/Downloads")  
NAICExpense <- read.csv("NAICExpense.csv")  
summary (NAICExpense)
```

```

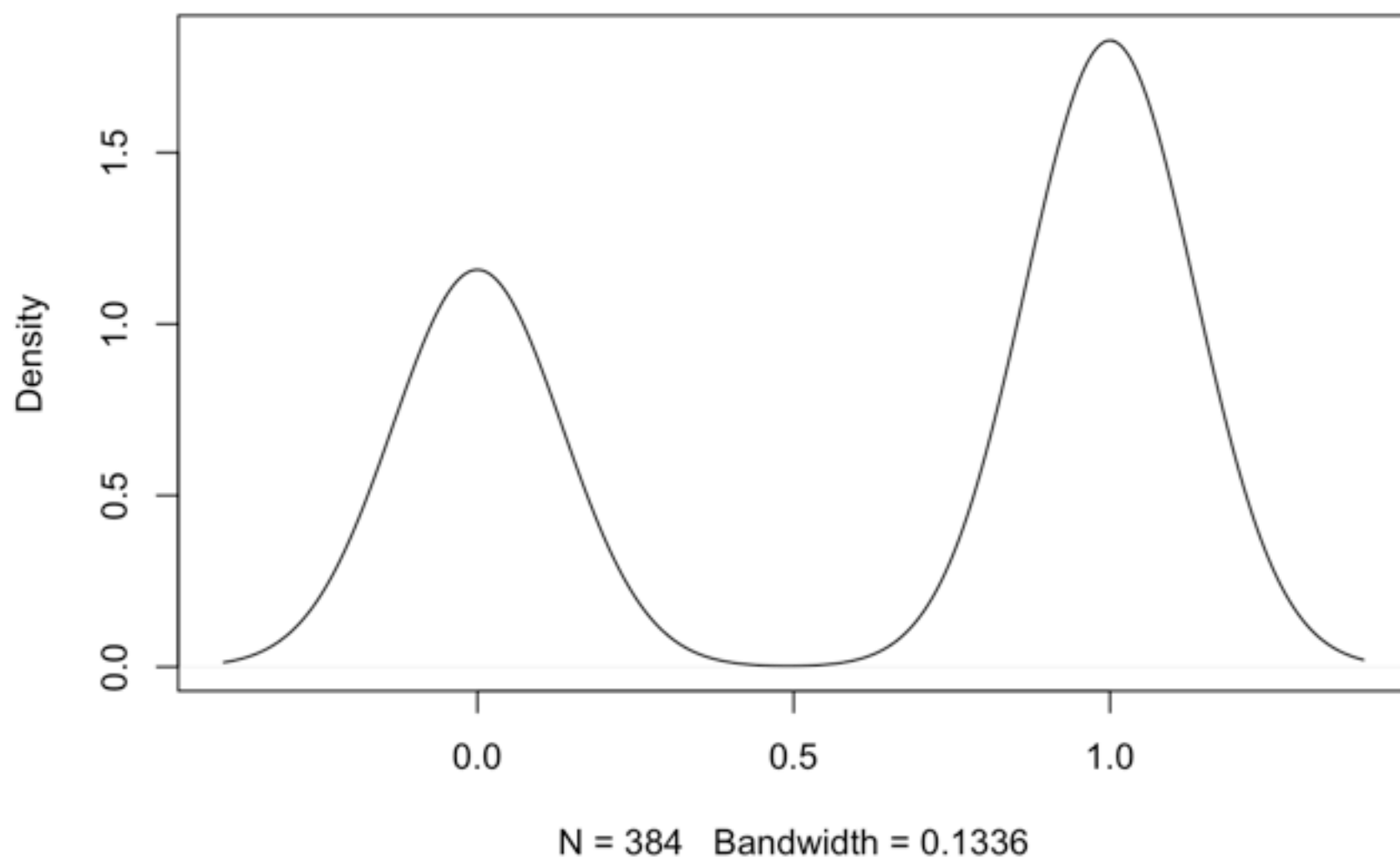
##          COMPANY_NAME      GROUP      MUTUAL
## AAA Mid-Atlantic Ins Co      : 1      Min.      :0.000      Min.      :0.0000
## Acceptance Ind Ins Co      : 1      1st Qu.:0.000      1st Qu.:0.0000
## Accredited Surety & Cas Co Inc: 1      Median :1.000      Median :0.0000
## Ace Ins Co                  : 1      Mean      :0.612      Mean      :0.1875
## Admiral Ind Co              : 1      3rd Qu.:1.000      3rd Qu.:0.0000
## Adriatic Ins Co            : 1      Max.      :1.000      Max.      :1.0000
## (Other)                     :378
##          STOCK          RBC          EXPENSES          STAFFWAGE
## Min.      :0.0000      Min.      :0.000e+00      Min.      : -0.002038      Min.      : 51.73
## 1st Qu.:0.0000      1st Qu.:6.257e+08      1st Qu.: 0.001584      1st Qu.: 80.06
## Median :1.0000      Median :2.753e+09      Median : 0.008504      Median : 84.38
## Mean      :0.6823      Mean      :2.247e+10      Mean      : 0.043190      Mean      : 87.18
## 3rd Qu.:1.0000      3rd Qu.:1.118e+10      3rd Qu.: 0.029826      3rd Qu.: 93.82
## Max.      :1.0000      Max.      :8.388e+11      Max.      : 1.236946      Max.      :137.48
##
##          AGENTWAGE          LONGLOSS          SHORTLOSS
## Min.      : 47.47      Min.      : -0.070623      Min.      : -0.0031685
## 1st Qu.: 74.81      1st Qu.: 0.000000      1st Qu.: 0.0002369
## Median : 78.77      Median : 0.001784      Median : 0.0040240
## Mean      : 80.15      Mean      : 0.024926      Mean      : 0.0373586
## 3rd Qu.: 85.44      3rd Qu.: 0.011280      3rd Qu.: 0.0217943
## Max.      :126.17      Max.      : 0.853915      Max.      : 1.1710587
## NA's      :19
##          GPWPERSONAL          GPWCOMM          ASSETS
## Min.      : -0.0037514      Min.      : -0.000648      Min.      :0.000321
## 1st Qu.: 0.0000000      1st Qu.: 0.003838      1st Qu.:0.012758
## Median : 0.0003125      Median : 0.023807      Median :0.056746
## Mean      : 0.0531127      Mean      : 0.122657      Mean      :0.356543
## 3rd Qu.: 0.0272581      3rd Qu.: 0.086440      3rd Qu.:0.197437
## Max.      : 1.8224858      Max.      : 4.189401      Max.      :8.705380
##
##          CASH          LIQUIDRATIO
## Min.      :0.000018      Min.      : 1.788
## 1st Qu.:0.011377      1st Qu.: 87.403
## Median :0.050469      Median : 96.027
## Mean      :0.332871      Mean      : 92.597
## 3rd Qu.:0.184971      3rd Qu.:103.861
## Max.      :8.823477      Max.      :127.858
##

```

In this code, the summary give that the mean and median and quartile for each variables. To test the skewness we can use the code below.

```
plot(density(NAICExpense$GROUP))
```

**density.default(x = NAICExpense\$GROUP)**



As the following graph, this variable group is bi-modal since this graph has two peaks.

b. We have to use the attach code.

```
attach(NAICExpense)
NAICExpense$LNEXPENSES <- log(1+EXPENSES)
NAICExpense$LNLONGLOSS <- log(1+LONGLOSS)
NAICExpense$LNSHORTLOSS <- log(1+SHORTLOSS)
NAICExpense$LNGPWPERSONAL <- log(1+GPWPERSONAL)
NAICExpense$LNGPWCOMM <- log(1+GPWCOMM)
NAICExpense$LNASSETS <- log(1+ASSETS)
NAICExpense$LNCASH <- log(1+CASH)
NAICExpense$LNSTAFFWAGE <- log(1+STAFFWAGE)
NAICExpense$LNAGENTWAGE <- log(1+AGENTWAGE)
```

c. The following code is producing a table of correlations for the non-binary variables.

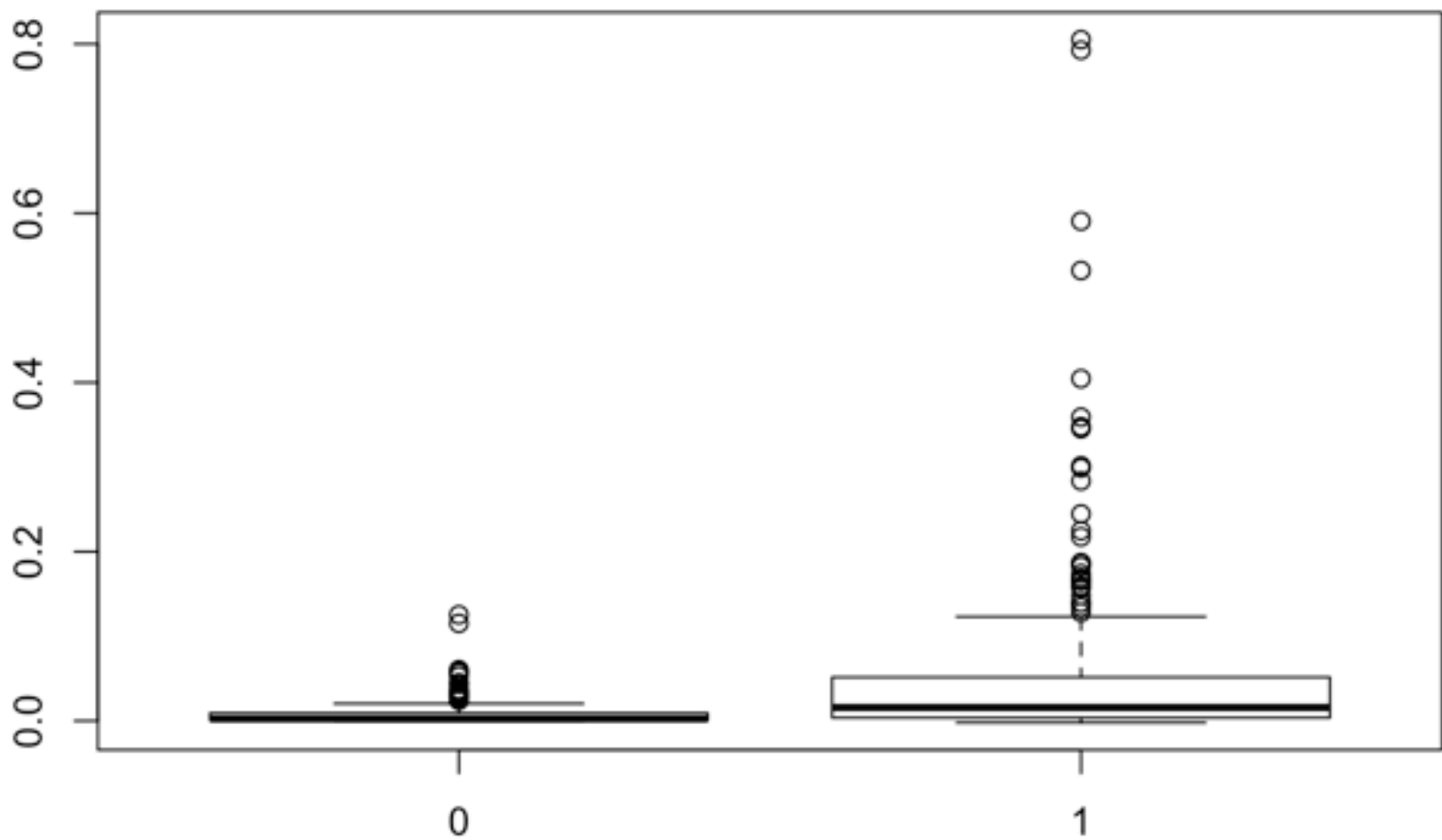
```
cor(NAICExpense[c("LNEXPENSES", "LNLONGLOSS", "LNSHORTLOSS", "LNKPWPERSONAL", "LNKPWCO  
MM", "LNASSETS", "LNCASH", "LNSTAFFWAGE", "LNAGENTWAGE")], use="complete.obs")
```

##	LNEXPENSES	LNLONGLOSS	LNSHORTLOSS	LNGPWPPERSONAL	LNGPWCOMM
##	LNEXPENSES	1.00000000	0.91831035	0.9299562	0.80914739 0.86407992
##	LNLONGLOSS	0.91831035	1.00000000	0.8672083	0.85407685 0.75195971
##	LNSHORTLOSS	0.92995618	0.86720834	1.0000000	0.75839391 0.84052628
##	LNGPWPPERSONAL	0.80914739	0.85407685	0.7583939	1.00000000 0.61720746
##	LNGPWCOMM	0.86407992	0.75195971	0.8405263	0.61720746 1.00000000
##	LNASSETS	0.86292823	0.75029418	0.8168381	0.64465690 0.89056855
##	LNCASH	0.86035271	0.74717924	0.8130868	0.64447389 0.87766591
##	LNSTAFFWAGE	0.10119232	0.07584390	0.0880556	0.04193561 0.13997749
##	LNAGENTWAGE	0.08683784	0.08956618	0.0606250	0.04677582 0.09621657
##	LNASSETS	LNCASH	LNSTAFFWAGE	LNAGENTWAGE	
##	LNEXPENSES	0.86292823	0.86035271	0.10119232	0.08683784
##	LNLONGLOSS	0.75029418	0.74717924	0.07584390	0.08956618
##	LNSHORTLOSS	0.81683807	0.81308682	0.08805560	0.06062500
##	LNGPWPPERSONAL	0.64465690	0.64447389	0.04193561	0.04677582
##	LNGPWCOMM	0.89056855	0.87766591	0.13997749	0.09621657
##	LNASSETS	1.00000000	0.99719848	0.12239853	0.09837449
##	LNCASH	0.99719848	1.00000000	0.12472843	0.09797375
##	LNSTAFFWAGE	0.12239853	0.12472843	1.00000000	0.78368834
##	LNAGENTWAGE	0.09837449	0.09797375	0.78368834	1.00000000

The three variable that are most higholy correlated with LNEXPENSES are LNLONGLOSS of 0.91831035 value, LNSHORTLOSS of 0.9299562 value and LNGPWCOMM have 0.86407992 value.

d.

```
boxplot(LNEXPENSES ~ NAICExpense$GROUP, data=NAICExpense)
```



This shows the box plot and the result is that the Group 1, affiliate companies has higher expense level.

e.

```
linearmodel <- lm(LNEXPENSES ~ LNLONGLOSS+LNSHORTLOSS +LNGPWPERSONAL+LNGPWCOMM+LNASSET  
TS+LNCASH+GROUP +STOCK+MUTUAL+LNSTAFFWAGE+LNAGENTWAGE, data=NAICExpense)  
  
summary(linearmodel)
```

```
##
## Call:
## lm(formula = LNEXPENSES ~ LNLONGLOSS + LNSHORTLOSS + LNGPWPERSOAL +
##      LNGPWCOMM + LNASSETS + LNCASH + GROUP + STOCK + MUTUAL +
##      LNSTAFFWAGE + LNAGENTWAGE, data = NAICExpense)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.133582	-0.003153	-0.000142	0.002596	0.125525

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0072596	0.0457795	-0.159	0.87409
LNLONGLOSS	0.4575804	0.0454308	10.072	< 2e-16 ***
LNSHORTLOSS	0.3090791	0.0325909	9.484	< 2e-16 ***
LNGPWPERSOAL	0.0617514	0.0193494	3.191	0.00154 **
LNGPWCOMM	0.0799024	0.0168573	4.740	3.11e-06 ***
LNASSETS	-0.0517077	0.0475079	-1.088	0.27716
LNCASH	0.0957329	0.0468678	2.043	0.04183 *
GROUP	-0.0001869	0.0028205	-0.066	0.94721
STOCK	-0.0032614	0.0037975	-0.859	0.39102
MUTUAL	-0.0020573	0.0042684	-0.482	0.63011
LNSTAFFWAGE	0.0005959	0.0143888	0.041	0.96699
LNAGENTWAGE	0.0016827	0.0166109	0.101	0.91937

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02235 on 353 degrees of freedom
## (19 observations deleted due to missingness)
## Multiple R-squared:  0.9422, Adjusted R-squared:  0.9404
## F-statistic: 523.1 on 11 and 353 DF, p-value: < 2.2e-16
```

The result shows that the residual standard error is 0.2235, the  $R^2$  value of 0.9422 and adjusted r squared value( $R_a^2$ ) of 0.9404.

f.

```
linearmodel2 <- lm(LNEXPENSES ~ LNLONGLOSS+LNSHORTLOSS +LNGPWPERSOAL+LNGPWCOMM+LNASS
ETS+GROUP+LNSTAFFWAGE +LNAGENTWAGE, data=NAICExpense)
```

i.

```
summary(linearmodel2)
```

```
##
## Call:
## lm(formula = LNEXPENSES ~ LNLONGLOSS + LNSHORTLOSS + LNGPWPERSOAL +
##      LNGPWCOMM + LNASSETS + GROUP + LNSTAFFWAGE + LNAGENTWAGE,
##      data = NAICExpense)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -0.135640 -0.002442 -0.000204  0.002372  0.129172
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0123692   0.0457302   -0.270   0.78695
## LNLONGLOSS    0.4582237   0.0449961   10.184 < 2e-16 ***
## LNSHORTLOSS   0.3132434   0.0325236    9.631 < 2e-16 ***
## LNGPWPERSOAL  0.0631744   0.0191221    3.304  0.00105 **
## LNGPWCOMM     0.0688575   0.0160603    4.287 2.33e-05 ***
## LNASSETS      0.0442429   0.0073793    5.996 4.98e-09 ***
## GROUP        -0.0014443   0.0026199   -0.551  0.58179
## LNSTAFFWAGE    0.0027486   0.0143276    0.192  0.84798
## LNAGENTWAGE    0.0001926   0.0165764    0.012  0.99074
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02242 on 356 degrees of freedom
## (19 observations deleted due to missingness)
## Multiple R-squared:  0.9414, Adjusted R-squared:  0.94
## F-statistic: 714.4 on 8 and 356 DF,  p-value: < 2.2e-16
```

This gives the result that the residual standard error value of 0.02242,  $R^2$  value of 0.9414, adjusted r square value( $R_a^2$ ) of 0.94.

- ii. LNGPWCOMM coefficient is 0.0688575 and that is when LNGPWCOMM increases by 1, LNEXPENSES increases by 0.0688575.
- iii. This means that when GPWCOMM increases by \$2.72 then EXPENSES increases by \$1.07. Also, when GPWCOMM increases by \$1, EXPENSES increases by \$0.39.

g.

```
attach(NAICExpense)
```

```
## The following objects are masked from NAICExpense (pos = 3):
##
## AGENTWAGE, ASSETS, CASH, COMPANY_NAME, EXPENSES, GPWCOMM,
## GPWPERSOAL, GROUP, LIQUIDRATIO, LONGLOSS, MUTUAL, RBC,
## SHORTLOSS, STAFFWAGE, STOCK
```

```
NAICExpense$LNLONGLOSS_sq <- LNLONGLOSS^2
NAICExpense$LNSHORTLOSS_sq <- LNSHORTLOSS^2
NAICExpense$LNGPWPERSONAL_sq <- LNGPWPERSONAL^2
NAICExpense$LNGPWCOMM_sq <- LNGPWCOMM^2

linearmodel3 <- lm(LNEXPENSES ~ LNLONGLOSS+LNSHORTLOSS +LNGPWPERSONAL+LNGPWCOMM+LNASS
ETS+GROUP +LNSTAFFWAGE +LNAGENTWAGE+LNLONGLOSS_sq+LNSHORTLOSS_sq
+LNGPWPERSONAL_sq+LNGPWCOMM_sq, data=NAICExpense)
```

i.

```
summary(linearmodel3)
```

```
##
## Call:
## lm(formula = LNEXPENSES ~ LNLONGLOSS + LNSHORTLOSS + LNGPWPERSONAL +
##     LNGPWCOMM + LNASSETS + GROUP + LNSTAFFWAGE + LNAGENTWAGE +
##     LNLONGLOSS_sq + LNSHORTLOSS_sq + LNGPWPERSONAL_sq + LNGPWCOMM_sq,
##     data = NAICExpense)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.127422 -0.003482 -0.001427  0.002448  0.136426
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0170444   0.0428783   -0.398  0.691235
## LNLONGLOSS    0.4043201   0.0638821    6.329 7.50e-10 ***
## LNSHORTLOSS   0.3912107   0.0485005    8.066 1.15e-14 ***
## LNGPWPERSONAL 0.1258984   0.0362399    3.474 0.000577 ***
## LNGPWCOMM    -0.0299266   0.0224901   -1.331 0.184163
## LNASSETS      0.0512294   0.0072425    7.073 8.20e-12 ***
## GROUP         0.0006227   0.0025308    0.246 0.805780
## LNSTAFFWAGE   0.0013179   0.0134359    0.098 0.921916
## LNAGENTWAGE   0.0029188   0.0155519    0.188 0.851235
## LNLONGLOSS_sq 0.5300803   0.2523430    2.101 0.036384 *
## LNSHORTLOSS_sq -0.3690037   0.0869848   -4.242 2.83e-05 ***
## LNGPWPERSONAL_sq -0.1517765   0.0855151   -1.775 0.076787 .
## LNGPWCOMM_sq  0.1190789   0.0207438    5.740 2.04e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02099 on 352 degrees of freedom
## (19 observations deleted due to missingness)
## Multiple R-squared:  0.9492, Adjusted R-squared:  0.9474
## F-statistic: 547.7 on 12 and 352 DF,  p-value: < 2.2e-16
```



In this code, we can see that the residual standard error is that 0.02099 and the  $R^2$  value is 0.9492, and the adjusted R squared value is 0.9474.

ii. Yes, since s decreased (residual standard error) decreased and  $R^2$  increased.

h.

```
linearmodel4 <- lm(LNEXPENSES ~ LNLONGLOSS+LNSHORTLOSS +LNGPWPERSOAL+LNGPWCOMM+LNASS
ETS+GROUP +LNLONGLOSS_sq +LNSHORTLOSS_sq+LNGPWPERSOAL_sq+LNGPWCOMM_sq,
data=NAICExpense)
```

```
summary(linearmodel4)
```

```
##
## Call:
## lm(formula = LNEXPENSES ~ LNLONGLOSS + LNSHORTLOSS + LNGPWPERSOAL +
##      LNGPWCOMM + LNASSETS + GROUP + LNLONGLOSS_sq + LNSHORTLOSS_sq +
##      LNGPWPERSOAL_sq + LNGPWCOMM_sq, data = NAICExpense)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.129951 -0.003032 -0.001484  0.002596  0.140608
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0017045  0.0018137   0.940  0.347926
## LNLONGLOSS     0.3720507  0.0652027   5.706 2.36e-08 ***
## LNSHORTLOSS    0.3755400  0.0497915   7.542 3.56e-13 ***
## LNGPWPERSOAL   0.1630821  0.0367811   4.434 1.22e-05 ***
## LNGPWCOMM     -0.0199110  0.0221787  -0.898  0.369896
## LNASSETS       0.0446285  0.0067875   6.575 1.64e-10 ***
## GROUP         0.0000212  0.0025371   0.008  0.993336
## LNLONGLOSS_sq  0.6570863  0.2595077   2.532  0.011749 *
## LNSHORTLOSS_sq -0.3306514  0.0896817  -3.687  0.000261 ***
## LNGPWPERSOAL_sq -0.2071748  0.0877140  -2.362  0.018693 *
## LNGPWCOMM_sq   0.1092892  0.0212770   5.136 4.52e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02177 on 373 degrees of freedom
## Multiple R-squared:  0.9424, Adjusted R-squared:  0.9409
## F-statistic: 610.5 on 10 and 373 DF, p-value: < 2.2e-16
```

In this code, we can see that the residual standard error is that 0.02177 and the  $R^2$  value is 0.9424, and the adjusted R squared value is 0.9409.

i.

ii.

```
attach(NAICExpense)
```

```
## The following objects are masked from NAICExpense (pos = 3):  
##  
## AGENTWAGE, ASSETS, CASH, COMPANY_NAME, EXPENSES, GPWCOMM,  
## GPWPERSONAL, GROUP, LIQUIDRATIO, LNAGENTWAGE, LNASSETS,  
## LNCASH, LNXEXPENSES, LNGPWCOMM, LNGPWPERSONAL, LNLONGLOSS,  
## LNSHORTLOSS, LNSTAFFWAGE, LONGLOSS, MUTUAL, RBC, SHORTLOSS,  
## STAFFWAGE, STOCK
```

```
## The following objects are masked from NAICExpense (pos = 4):  
##  
## AGENTWAGE, ASSETS, CASH, COMPANY_NAME, EXPENSES, GPWCOMM,  
## GPWPERSONAL, GROUP, LIQUIDRATIO, LONGLOSS, MUTUAL, RBC,  
## SHORTLOSS, STAFFWAGE, STOCK
```

```
GROUPc <- GROUP - mean(GROUP)  
LNASSETSc <- LNASSETS - mean(LNASSETS)  
LNLONGLOSSc <- LNLONGLOSS - mean(LNLONGLOSS)  
LNSHORTLOSSc <- LNSHORTLOSS - mean(LNSHORTLOSS)  
LNGPWPERSONALc <- LNGPWPERSONAL - mean(LNGPWPERSONAL)  
LNGPWCOMMc <- LNGPWCOMM - mean(LNGPWCOMM)
```

```
NAICExpense$GROUPintLNASSETS <- GROUPc*LNASSETSc  
NAICExpense$GROUPintLNLONGLOSS <- GROUPc*LNLONGLOSSc  
NAICExpense$GROUPintLNSHORTLOSS <- GROUPc*LNSHORTLOSSc  
NAICExpense$GROUPintLNGPWPERSONAL <- GROUPc*LNGPWPERSONALc  
NAICExpense$GROUPintLNGPWCOMM <- GROUPc*LNGPWCOMMc
```

```
linearmodel5 <- lm(LNXEXPENSES ~ LNLONGLOSS+LNSHORTLOSS +LNGPWPERSONAL+LNGPWCOMM+LNASS  
ETS+GROUP+GROUPintLNASSETS +GROUPintLNLONGLOSS+GROUPintLNSHORTLOSS +GROUPintLNGPWPERS  
ONAL+GROUPintLNGPWCOMM, data=NAICExpense)
```

```
summary(linearmodel5)
```

```
##
## Call:
## lm(formula = LNEXPENSES ~ LNLONGLOSS + LNSHORTLOSS + LNGPWPERSONAL +
##      LNGPWCOMM + LNASSETS + GROUP + GROUPintLNASSETS + GROUPintLNLONGLOSS +
##      GROUPintLNSHORTLOSS + GROUPintLNGPWPERSONAL + GROUPintLNGPWCOMM,
##      data = NAICExpense)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.139143 -0.002150 -0.000102  0.002448  0.135855
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.002741   0.003563   0.769   0.4422
## LNLONGLOSS      0.443612   0.099162   4.474 1.02e-05 ***
## LNSHORTLOSS     0.370598   0.147159   2.518   0.0122 *
## LNGPWPERSONAL   0.071013   0.074156   0.958   0.3389
## LNGPWCOMM       0.078926   0.055137   1.431   0.1531
## LNASSETS        0.034929   0.019994   1.747   0.0815 .
## GROUP          -0.004968   0.006110  -0.813   0.4167
## GROUPintLNASSETS  0.009855   0.050795   0.194   0.8463
## GROUPintLNLONGLOSS 0.029238   0.248038   0.118   0.9062
## GROUPintLNSHORTLOSS -0.142808   0.376797  -0.379   0.7049
## GROUPintLNGPWPERSONAL 0.013999   0.189519   0.074   0.9412
## GROUPintLNGPWCOMM -0.027210   0.140788  -0.193   0.8469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02323 on 372 degrees of freedom
## Multiple R-squared:  0.9347, Adjusted R-squared:  0.9327
## F-statistic: 483.7 on 11 and 372 DF, p-value: < 2.2e-16
```

This show sthat the residual standard error is 0.02323 and the r squared value is 0.9347, and Adjusted R-squared value is 0.9327.

- ii. In this case that if GROUP=0, there is no regression line. So, EXPENSES will not increase when GPWCOMM increases by \$1.00.
- iii. Suppose that GPWCOMM increases by \$1.00, if the  $\ln[\text{GPWCOMM}]$  increases by \$1.00 and the  $\ln[\text{EXPENSES}]$ , GPWCOMM increases \$e, and EXPENSES increases by  $3^{(0.078926)}$ , so the EXPENSES increases 0.398092161 \$ when GPWCOMM increases by \$1.00.