Project 02

TUAN BUI

11/30/2021

rm(list = ls())  
# Load data  
Project\_data <- read.csv("~/OneDrive - Stony Brook University/SBU/MAT + AMS/Fall 2021/AMS 315/project/02/141951\_project2test.csv", header = T)  
# Check missing data  
sum(is.na(Project\_data))

## [1] 0

# Load packages  
library(MASS)  
library(leaps)  
library(knitr)  
  
cor(Project\_data)

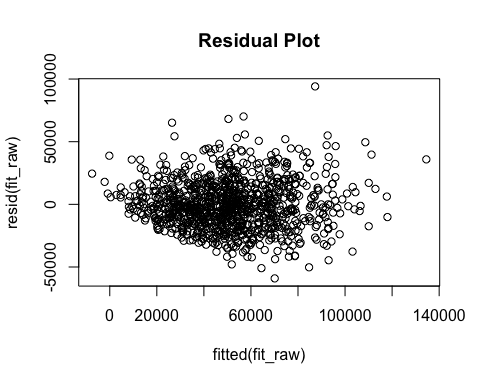
|  |
| --- |
| ## Y E1 E2 E3 E4 ## Y 1.0000000000 0.4390310687 0.024770050 0.008322687 0.437840381 ## E1 0.4390310687 1.0000000000 0.058051184 0.001538038 0.041372493 ## E2 0.0247700501 0.0580511839 1.000000000 0.072539362 0.055223439 ## E3 0.0083226872 0.0015380375 0.072539362 1.000000000 -0.021387270 ## E4 0.4378403814 0.0413724934 0.055223439 -0.021387270 1.000000000 ## G1 -0.0400559407 -0.0339943396 0.055984471 0.032690085 -0.040028169 ## G2 0.0302416763 0.0228116296 0.042849121 -0.045489743 0.004493733 ## G3 0.0145207319 -0.0221407028 -0.074377931 0.021340903 0.007277477 ## G4 -0.0790209781 -0.0416180775 -0.016565021 0.002173811 -0.005623516 ## G5 0.0021127507 0.0049152341 -0.003735070 -0.006086547 -0.009096364 ## G6 0.1308635809 0.0121045877 0.009432646 0.054050025 0.002036620 ## G7 0.0241360546 0.0269375010 -0.009334338 0.011996272 0.034764231 ## G8 0.0255380974 0.0390805046 -0.022943966 -0.006486638 0.008763760 ## G9 -0.0002051762 -0.0004464268 0.060219515 0.002677902 -0.015394130 ## G10 0.0134102999 -0.0196385843 0.020137510 0.029427501 -0.027480272 ## G11 0.0241415553 -0.0127085753 -0.011133452 -0.024048479 0.072549146 ## G12 -0.0270083266 0.0074288219 -0.002888555 0.037998174 0.025582274 ## G13 0.0214825554 0.0451097241 -0.042683632 -0.049029244 0.067986517 ## G14 -0.0446983400 -0.0501156782 0.019357039 -0.009425885 0.036612492 ## G15 -0.0117108280 -0.0602165585 -0.020003097 -0.002091599 0.019972923 ## G16 0.1046744987 0.0073615951 -0.005465025 0.030683000 -0.009361242 ## G17 0.0123041786 -0.0284352439 0.027917543 -0.012847907 0.033753829 ## G18 -0.0110409928 -0.0118239798 0.035899230 0.012127675 -0.012715125 ## G19 0.0078978410 0.0135353568 0.016041480 0.040648136 0.009370941 ## G20 -0.0230328155 -0.0266744600 0.029440663 0.046930906 0.035941337 |

## G1 G2 G3 G4 G5  
## Y -0.040055941 0.030241676 0.014520732 -0.079020978 0.002112751  
## E1 -0.033994340 0.022811630 -0.022140703 -0.041618078 0.004915234  
## E2 0.055984471 0.042849121 -0.074377931 -0.016565021 -0.003735070  
## E3 0.032690085 -0.045489743 0.021340903 0.002173811 -0.006086547  
## E4 -0.040028169 0.004493733 0.007277477 -0.005623516 -0.009096364  
## G1 1.000000000 0.014641277 0.023951021 0.011526578 -0.001457825  
## G2 0.014641277 1.000000000 -0.005505637 0.002287078 0.016766290  
## G3 0.023951021 -0.005505637 1.000000000 0.054926956 -0.005661796  
## G4 0.011526578 0.002287078 0.054926956 1.000000000 0.006069918  
## G5 -0.001457825 0.016766290 -0.005661796 0.006069918 1.000000000  
## G6 -0.003078975 0.006954703 -0.031109557 -0.043187147 -0.037019440  
## G7 0.004485469 0.038724774 -0.039737885 -0.024060611 -0.001321785  
## G8 -0.027390258 0.042770841 -0.023581405 -0.007829064 0.094674887  
## G9 -0.026150747 -0.005587818 0.003344504 -0.035135671 -0.033291893  
## G10 0.005889569 -0.012566760 -0.029723313 0.050866872 0.011556838  
## G11 0.036576259 0.042770841 -0.055577320 -0.047854054 -0.049218164  
## G12 -0.018960981 -0.016971483 -0.022955542 -0.034962978 0.015004152  
## G13 0.003727259 -0.010643482 -0.003948541 0.056533027 0.001457825  
## G14 -0.042838433 -0.020916176 -0.018800918 -0.034743501 -0.032887356  
## G15 -0.005347701 -0.007162878 0.050485650 0.010291853 -0.003218297  
## G16 -0.022318450 -0.040650797 0.017975321 0.034401970 0.039403162  
## G17 -0.047379795 -0.009187426 -0.027580895 -0.015834062 -0.025235988  
## G18 0.035718566 0.005347701 0.012053442 -0.015530085 -0.026528420  
## G19 -0.017232498 0.028857863 -0.045341602 -0.001495332 0.016833431  
## G20 0.029887363 0.007425813 -0.041726959 -0.049238780 0.039548116  
## G6 G7 G8 G9 G10  
## Y 0.1308635809 0.024136055 0.0255380974 -0.0002051762 0.0134102999  
## E1 0.0121045877 0.026937501 0.0390805046 -0.0004464268 -0.0196385843  
## E2 0.0094326461 -0.009334338 -0.0229439662 0.0602195147 0.0201375095  
## E3 0.0540500250 0.011996272 -0.0064866384 0.0026779024 0.0294275014  
## E4 0.0020366199 0.034764231 0.0087637598 -0.0153941302 -0.0274802717  
## G1 -0.0030789753 0.004485469 -0.0273902581 -0.0261507472 0.0058895687  
## G2 0.0069547026 0.038724774 0.0427708414 -0.0055878183 -0.0125667597  
## G3 -0.0311095575 -0.039737885 -0.0235814051 0.0033445040 -0.0297233133  
## G4 -0.0431871465 -0.024060611 -0.0078290639 -0.0351356712 0.0508668724  
## G5 -0.0370194404 -0.001321785 0.0946748870 -0.0332918934 0.0115568384  
## G6 1.0000000000 -0.063005583 0.0009542023 0.0435582052 0.0170911861  
## G7 -0.0630055828 1.000000000 -0.0072795665 -0.0230163041 0.0096237208  
## G8 0.0009542023 -0.007279567 1.0000000000 0.0449006403 -0.0384922700  
## G9 0.0435582052 -0.023016304 0.0449006403 1.0000000000 -0.0046861177  
## G10 0.0170911861 0.009623721 -0.0384922700 -0.0046861177 1.0000000000  
## G11 -0.0350140338 0.036689335 -0.0192098832 -0.0194394022 0.0174884971  
## G12 0.0369662154 0.005014873 0.0169999043 0.0109457428 0.0009665928  
## G13 -0.0009185011 -0.024476363 -0.0045930004 -0.0100496341 -0.0378866284  
## G14 -0.0069781553 -0.022886621 0.0210564981 0.0024583027 -0.0191563410  
## G15 0.0269364975 -0.001245837 0.0347772618 -0.0457992209 -0.0165652742  
## G16 -0.0109176153 -0.006522165 -0.0106192643 0.0444716410 -0.0078230506  
## G17 0.0529083210 0.032692162 -0.0072191787 -0.0154181495 0.0214871233  
## G18 -0.0129109302 0.035496318 -0.0045930004 0.0020171597 0.0061093287  
## G19 0.0589167589 -0.017157460 -0.0011339045 -0.0382011633 0.0112920119  
## G20 -0.0348853515 0.029618209 0.0054926184 -0.0449159965 0.0119131352  
## G11 G12 G13 G14 G15  
## Y 0.024141555 -0.0270083266 0.0214825554 -0.044698340 -0.011710828  
## E1 -0.012708575 0.0074288219 0.0451097241 -0.050115678 -0.060216558  
## E2 -0.011133452 -0.0028885548 -0.0426836323 0.019357039 -0.020003097  
## E3 -0.024048479 0.0379981744 -0.0490292439 -0.009425885 -0.002091599  
## E4 0.072549146 0.0255822737 0.0679865174 0.036612492 0.019972923  
## G1 0.036576259 -0.0189609811 0.0037272589 -0.042838433 -0.005347701  
## G2 0.042770841 -0.0169714827 -0.0106434817 -0.020916176 -0.007162878  
## G3 -0.055577320 -0.0229555417 -0.0039485414 -0.018800918 0.050485650  
## G4 -0.047854054 -0.0349629780 0.0565330272 -0.034743501 0.010291853  
## G5 -0.049218164 0.0150041524 0.0014578250 -0.032887356 -0.003218297  
## G6 -0.035014034 0.0369662154 -0.0009185011 -0.006978155 0.026936498  
## G7 0.036689335 0.0050148730 -0.0244763628 -0.022886621 -0.001245837  
## G8 -0.019209883 0.0169999043 -0.0045930004 0.021056498 0.034777262  
## G9 -0.019439402 0.0109457428 -0.0100496341 0.002458303 -0.045799221  
## G10 0.017488497 0.0009665928 -0.0378866284 -0.019156341 -0.016565274  
## G11 1.000000000 -0.0309575523 -0.0245825370 0.037042508 -0.025174585  
## G12 -0.030957552 1.0000000000 0.0149635207 0.048946694 0.098922465  
## G13 -0.024582537 0.0149635207 1.0000000000 0.006860858 -0.010643482  
## G14 0.037042508 0.0489466940 0.0068608584 1.000000000 -0.012923394  
## G15 -0.025174585 0.0989224649 -0.0106434817 -0.012923394 1.000000000  
## G16 -0.030607763 -0.0629816900 -0.0576556640 -0.035101172 -0.036653209  
## G17 -0.003222277 0.0130034496 0.0113986289 -0.010915523 0.030780472  
## G18 -0.016586722 0.0629330457 0.0202662214 -0.009129175 -0.002647891  
## G19 -0.053089890 -0.0249670646 -0.0027556022 0.047000443 0.004878847  
## G20 -0.014500513 -0.0070297657 0.0061093287 0.008831249 -0.052551904  
## G16 G17 G18 G19 G20  
## Y 0.104674499 0.012304179 -0.011040993 0.007897841 -0.023032816  
## E1 0.007361595 -0.028435244 -0.011823980 0.013535357 -0.026674460  
## E2 -0.005465025 0.027917543 0.035899230 0.016041480 0.029440663  
## E3 0.030683000 -0.012847907 0.012127675 0.040648136 0.046930906  
## E4 -0.009361242 0.033753829 -0.012715125 0.009370941 0.035941337  
## G1 -0.022318450 -0.047379795 0.035718566 -0.017232498 0.029887363  
## G2 -0.040650797 -0.009187426 0.005347701 0.028857863 0.007425813  
## G3 0.017975321 -0.027580895 0.012053442 -0.045341602 -0.041726959  
## G4 0.034401970 -0.015834062 -0.015530085 -0.001495332 -0.049238780  
## G5 0.039403162 -0.025235988 -0.026528420 0.016833431 0.039548116  
## G6 -0.010917615 0.052908321 -0.012910930 0.058916759 -0.034885352  
## G7 -0.006522165 0.032692162 0.035496318 -0.017157460 0.029618209  
## G8 -0.010619264 -0.007219179 -0.004593000 -0.001133905 0.005492618  
## G9 0.044471641 -0.015418150 0.002017160 -0.038201163 -0.044915997  
## G10 -0.007823051 0.021487123 0.006109329 0.011292012 0.011913135  
## G11 -0.030607763 -0.003222277 -0.016586722 -0.053089890 -0.014500513  
## G12 -0.062981690 0.013003450 0.062933046 -0.024967065 -0.007029766  
## G13 -0.057655664 0.011398629 0.020266221 -0.002755602 0.006109329  
## G14 -0.035101172 -0.010915523 -0.009129175 0.047000443 0.008831249  
## G15 -0.036653209 0.030780472 -0.002647891 0.004878847 -0.052551904  
## G16 1.000000000 -0.030607763 0.014321039 0.043196813 -0.011822475  
## G17 -0.030607763 1.000000000 0.011398629 0.042828852 0.001493992  
## G18 0.014321039 0.011398629 1.000000000 -0.018746082 0.042106021  
## G19 0.043196813 0.042828852 -0.018746082 1.000000000 -0.012698021  
## G20 -0.011822475 0.001493992 0.042106021 -0.012698021 1.000000000

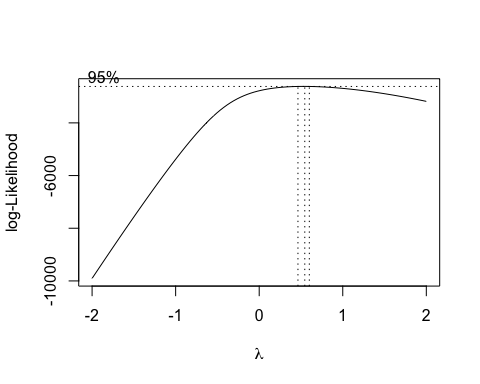
fit\_E <- lm(Y ~ E1 + E2 + E3 + E4, data = Project\_data)  
summary(fit\_E)

##   
## Call:  
## lm(formula = Y ~ E1 + E2 + E3 + E4, data = Project\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -60208 -15708 -2328 12634 124256   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -32038.5 4798.1 -6.677 4.04e-11 \*\*\*  
## E1 4253.9 253.6 16.777 < 2e-16 \*\*\*  
## E2 -240.4 248.9 -0.966 0.334   
## E3 182.7 249.5 0.732 0.464   
## E4 4146.8 247.7 16.740 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22960 on 996 degrees of freedom  
## Multiple R-squared: 0.37, Adjusted R-squared: 0.3675   
## F-statistic: 146.3 on 4 and 996 DF, p-value: < 2.2e-16

fit\_raw <- lm(Y ~ (E1 + E2 + E3 + E4 + G1 + G2 + G3 + G4 + G5 + G6 + G7 + G8 + G9 + G10 + G11 + G12 + G13 + G14 + G15 + G16 + G17 + G18 + G19 + G20)^2, data = Project\_data)  
  
plot(resid(fit\_raw) ~ fitted(fit\_raw), main = 'Residual Plot')



boxcox(fit\_raw)



fit\_trans <- lm( I(Y^.55) ~ (E1 + E2 + E3 + E4 + G1 + G2 + G3 + G4 + G5 + G6 + G7 + G8 + G9 + G10 + G11 + G12 + G13 + G14 + G15 + G16 + G17 + G18 + G19 + G20)^2, data = Project\_data)  
  
summary(fit\_E)$adj.r.square

## [1] 0.3675166

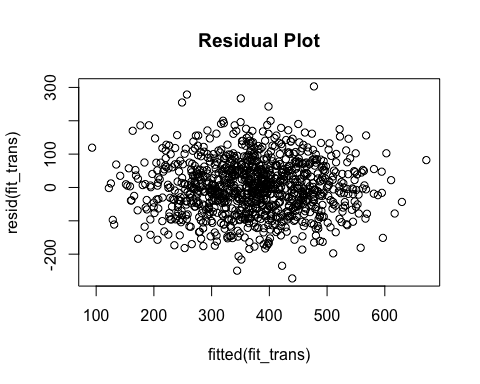
summary(fit\_raw)$adj.r.square

## [1] 0.3651468

summary(fit\_trans)$adj.r.square

## [1] 0.3797804

plot(resid(fit\_trans) ~ fitted(fit\_trans), main = 'Residual Plot')



fit <- regsubsets(model.matrix(fit\_trans)[, -1], I((Project\_data$Y)^.55), nbest = 1, nvmax = 5, method = 'forward', intercept = T)  
temp\_01 <- summary(fit)  
  
Var <- colnames(model.matrix(fit\_trans))  
fit\_select <- apply(temp\_01$which, 1, function(x) paste0(Var[x], collapse = '+'))  
kable(data.frame(cbind(model = fit\_select, adjR2 = temp\_01$adjr2, BIC = temp\_01$bic)), caption = 'Model summary')

Model summary

|  |  |  |
| --- | --- | --- |
| model | adjR2 | BIC |
| (Intercept)+E1:E4 | 0.366199718403503 | -443.661400089459 |
| (Intercept)+E1:E4+G6:G16 | 0.387215331138535 | -471.509161059619 |
| (Intercept)+G4+E1:E4+G6:G16 | 0.389863563271667 | -469.939247148818 |
| (Intercept)+G4+E1:E4+G6:G16+G10:G16 | 0.391916981402857 | -467.409561976871 |
| (Intercept)+G4+E1:E4+E1:G6+G6:G16+G10:G16 | 0.39498745394734 | -466.573615662551 |

fit\_main <- lm( I(Y^.55) ~ E1 + E2 + E3 + E4 + G1 + G2 + G3 + G4 + G5 + G6 + G7 + G8 + G9 + G10 + G11 + G12 + G13 + G14 + G15 + G16 + G17 + G18 + G19 + G20, data=Project\_data)  
temp\_02 <- summary(fit\_main)  
kable(temp\_02$coefficients[abs(temp\_02$coefficients[, 4]) <= 0.001, ], caption='Sig Coefficients')

Sig Coefficients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| E1 | 18.54684 | 1.070737 | 17.321558 | 0.0e+00 |
| E4 | 18.49662 | 1.046815 | 17.669420 | 0.0e+00 |
| G6 | 28.36769 | 6.146802 | 4.615032 | 4.5e-06 |
| G16 | 28.03619 | 6.143072 | 4.563871 | 5.7e-06 |

fit\_2stage <- lm( I(Y^.55) ~ (E1 + E4 + G6 + G16)^2, data = Project\_data)  
temp\_03 <- summary(fit\_2stage)  
temp\_03

##   
## Call:  
## lm(formula = I(Y^0.55) ~ (E1 + E4 + G6 + G16)^2, data = Project\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -267.94 -62.66 1.96 63.95 352.75   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -42.550145 45.193390 -0.942 0.347   
## E1 19.809071 4.066675 4.871 1.29e-06 \*\*\*  
## E4 22.288044 4.049775 5.504 4.74e-08 \*\*\*  
## G6 -20.845642 30.165333 -0.691 0.490   
## G16 47.329715 30.092724 1.573 0.116   
## E1:E4 -0.307254 0.361451 -0.850 0.395   
## E1:G6 3.816977 2.119372 1.801 0.072 .   
## E1:G16 0.003528 2.116790 0.002 0.999   
## E4:G6 0.561730 2.068373 0.272 0.786   
## E4:G16 -2.598639 2.066161 -1.258 0.209   
## G6:G16 12.197590 12.137762 1.005 0.315   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 95.8 on 990 degrees of freedom  
## Multiple R-squared: 0.4152, Adjusted R-squared: 0.4093   
## F-statistic: 70.29 on 10 and 990 DF, p-value: < 2.2e-16

kable(temp\_03$coefficients[ abs(temp\_03$coefficients[, 3]) >= 4, ])

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| E1 | 19.80907 | 4.066675 | 4.871073 | 1.3e-06 |
| E4 | 22.28804 | 4.049775 | 5.503527 | 0.0e+00 |