Faculty Name: Prof. José Manuel Magallanes, PhD

Student Name: Samikshya Pandey Class Name: PUBPOL 542 A Wi 21: Computational Thinking For Governance Analytics

File details: Conducting Regression Analysis

File Source: Merged data found in https://github.com/PUBPOL-542-Computational-Thinking/Merge/raw/main/UpdatedTJHemaSamik.csv.

The following documents provides instructions on conducting a OLS regression analysis.

First, we need to import the excel files that consists of the data we need to analyze. To do so, we provide the location for the excel files and tell R to read the csv/excel file using the code read.csv.

The data name final data is created.

```
mergecsv = "https://github.com/PUBPOL-542-Computational-Thinking/Merge/raw/main/UpdatedTJHemaSamik.csv"

finaldata = read.csv(mergecsv)
finaldata
```

```
##                     Country lessthan5_50     Continent        FPI        FDI
## 1                   Albania         33.8        Europe   67629.00 0.19917804
## 2                   Algeria         28.6        Africa  996868.00 0.14482453
## 3                    Angola         89.3        Africa   67381.00 0.15681012
## 4       Antigua and Barbuda           NA North America   10557.00 0.31441262
## 5                 Argentina         12.2 South America  885029.00 0.32153085
## 6                 Australia          0.7       Oceania 2141404.00 0.90036184
## 7                   Austria          0.7        Europe 3867699.00 0.63070452
## 8                Bangladesh         84.5          Asia  764531.00 0.23295973
## 9                  Barbados           NA North America   24951.00 0.44412541
## 10                  Belarus          0.4        Europe  392526.00 0.17650366
## 11                  Belgium          0.3        Europe 6062558.00 0.63214582
## 12                   Belize         53.0 North America   18802.00 0.21108365
## 13                    Benin         90.6        Africa   10084.00 0.12785479
## 14                   Bhutan         38.6          Asia   30902.00 0.19977939
## 15   Bosnia and Herzegovina          3.9        Europe  280006.00 0.25902098
## 16                 Botswana         60.4        Africa   61453.00 0.26245126
## 17                   Brazil         19.8 South America  881521.00 0.60049450
## 18                 Bulgaria          7.5        Europe  479545.00 0.37759098
## 19             Burkina Faso         92.3        Africa   17089.00 0.13078856
## 20                  Burundi         96.8        Africa    4623.00 0.13706477
## 21                 Cambodia           NA          Asia  310203.00 0.15388516
## 22                 Cameroon         68.9        Africa   46430.00 0.10302488
## 23                   Canada          0.7 North America 4999087.00 0.87771189
## 24 Central African Republic         92.8        Africa     968.00 0.05296880
## 25                     Chad         86.2        Africa    1088.00 0.09324554
## 26                    Chile          3.7 South America  901684.00 0.45953962
## 27                 Colombia         27.8 South America  867678.00 0.38143748
## 28                  Comoros         62.3        Africa    1251.00 0.05374534
## 29               Costa Rica         10.9 North America  525710.00 0.27699831
## 30                  Croatia          3.8        Europe  546794.00 0.49692753
## 31                   Cyprus          0.1        Europe  122586.00 0.47735870
## 32                  Denmark          0.2        Europe 2090845.00 0.64490145
## 33                 Djibouti         70.6        Africa   56471.00 0.15629709
## 34                 Dominica           NA North America    6894.00 0.22356336
```

```
## 35        Dominican Republic          13.8 North America   393367.00 0.17905858
## 36                   Ecuador          24.2 South America   399225.00 0.17488688
## 37                     Egypt          70.4        Africa  2613098.00 0.31059951
## 38               El Salvador          25.7 North America   297002.00 0.22879329
## 39         Equatorial Guinea            NA        Africa     3354.00 0.10593032
## 40                   Eritrea            NA        Africa     3380.00 0.08361098
## 41                   Estonia           1.0        Europe   688950.00 0.28488263
## 42                  Ethiopia          85.0        Africa   103584.00 0.13464746
## 43                      Fiji          48.6       Oceania    44183.00 0.21740668
## 44                   Finland           0.1        Europe  1415408.00 0.73641503
## 45                    France           0.2        Europe  8256757.00 0.77707875
## 46                     Gabon          32.2        Africa     9771.00 0.12545852
## 47                   Georgia          42.9        Europe   158067.00 0.30651662
## 48                   Germany           0.2        Europe 16442309.00 0.71374440
## 49                     Ghana          56.9        Africa   224596.00 0.14493361
## 50                    Greece           4.7        Europe  1109189.00 0.53508729
## 51                   Grenada            NA North America     7699.10 0.29496527
## 52                 Guatemala          48.8 North America   498742.00 0.23107600
## 53                    Guinea          92.3        Africa     8464.00 0.10549801
## 54             Guinea-Bissau          93.4        Africa      890.00 0.10380552
## 55                    Guyana          56.4 South America     9376.00 0.16092123
## 56                     Haiti          78.9 North America    44007.66 0.11254802
## 57                  Honduras          50.3 North America   217467.00 0.20701154
## 58                   Hungary           3.0        Europe  1426364.00 0.44075400
## 59                   Iceland           0.2        Europe   123843.00 0.50920200
## 60                     India          87.4          Asia  6799109.00 0.41823876
## 61                 Indonesia          53.2          Asia  3028095.00 0.36588052
## 62                   Ireland           0.7        Europe   839558.00 0.67312020
## 63                    Israel           2.7          Asia   954216.00 0.55996156
## 64                     Italy           3.1        Europe  9492899.00 0.76360083
## 65                   Jamaica          29.7 North America   103057.00 0.27481329
## 66                     Japan           1.0          Asia 10581454.00 0.85917377
## 67                    Jordan          18.1          Asia   462557.00 0.38103896
## 68                Kazakhstan           8.6          Asia   525109.00 0.33864823
## 69                     Kenya          87.0        Africa   424480.00 0.18637806
## 70                  Kiribati          69.4       Oceania      925.00 0.09353153
## 71                    Kuwait            NA          Asia   397252.00 0.51866102
## 72                    Latvia           4.0        Europe   627882.00 0.25796631
## 73                   Lebanon           1.9          Asia   430885.00 0.29978487
## 74                   Lesotho          89.9        Africa    17446.00 0.14479280
## 75                   Liberia          92.2        Africa     8662.00 0.12948039
## 76                     Libya            NA        Africa   108965.00 0.14495303
## 77                 Lithuania           3.8        Europe   917520.00 0.21916281
## 78                Luxembourg           0.5        Europe   364141.00 0.75357300
## 79                Madagascar          97.3        Africa    30716.00 0.10033925
## 80                    Malawi          96.7        Africa    18001.00 0.08255562
## 81                  Malaysia           2.7          Asia  2429761.00 0.64984596
## 82                  Maldives          54.3          Asia    37737.00 0.17961434
## 83                      Mali          94.9        Africa    12038.00 0.14128189
## 84                     Malta           0.2        Europe   103526.00 0.53520727
## 85                Mauritania          58.8        Africa    10822.00 0.11040888
## 86                 Mauritius          12.7        Africa    88239.00 0.41907090
## 87                    Mexico          23.0 North America  5819335.00 0.41442460
## 88                  Mongolia          28.9          Asia    51227.00 0.38211742
```

```
## 89                   Morocco         31.3        Africa   984641.00 0.36102104
## 90                Mozambique         91.8        Africa    52976.00 0.14452003
## 91                   Myanmar         67.2          Asia   317722.00 0.13075243
## 92                   Namibia         50.1        Africa    85359.00 0.40106532
## 93               Netherlands          0.5        Europe  6005014.00 0.71647942
## 94               New Zealand           NA       Oceania   566797.00 0.59054077
## 95                 Nicaragua         34.8 North America    65974.00 0.14759944
## 96                     Niger         93.4        Africa     9704.00 0.12854712
## 97                   Nigeria         92.0        Africa   634434.00 0.22954693
## 98                    Norway          0.5        Europe  1062346.00 0.66496569
## 99                      Oman           NA          Asia   237201.00 0.39146367
## 100                 Pakistan         76.0          Asia   961296.00 0.23933755
## 101                   Panama         12.7 North America   125155.00 0.33859685
## 102         Papua New Guinea         86.9       Oceania    18520.00 0.21659438
## 103                 Paraguay         17.0 South America   199529.00 0.12279055
## 104                     Peru         22.1 South America   872279.00 0.37356547
## 105              Philippines         30.8          Asia  1913429.00 0.36652714
## 106                   Poland          2.1        Europe  5628459.00 0.47386059
## 107                 Portugal          1.8        Europe  1666766.00 0.72150308
## 108                    Qatar           NA          Asia   231493.00 0.53062397
## 109                  Romania         15.6        Europe  1264442.00 0.30938628
## 110                   Rwanda         91.6        Africa    20173.00 0.09173958
## 111                    Samoa         33.9       Oceania    11724.00 0.20346199
## 112             Saudi Arabia           NA          Asia  2194895.00 0.50915480
## 113                  Senegal         88.1        Africa   119772.00 0.11101273
## 114               Seychelles          6.6        Africa    18017.00 0.30775678
## 115             Sierra Leone         92.1        Africa     9946.00 0.06779619
## 116                Singapore           NA          Asia  1733921.00 0.71351969
## 117                 Slovenia          0.1        Europe  1266490.00 0.38186246
## 118          Solomon Islands         84.7       Oceania     3388.00 0.09186842
## 119             South Africa         57.1        Africa  1226400.00 0.63099885
## 120              South Sudan         84.8        Africa      196.00 0.05980796
## 121                    Spain          2.2        Europe  4522109.00 0.87340266
## 122                Sri Lanka         40.5          Asia   457259.00 0.27259490
## 123                    Sudan         73.2        Africa   117169.00 0.10871744
## 124                 Suriname         55.7 South America    10655.00 0.20779303
## 125                   Sweden          0.5        Europe  2587012.00 0.77644378
## 126              Switzerland          0.0        Europe  1529491.00 0.96395850
## 127               Tajikistan         54.2          Asia    91942.00 0.09288083
## 128                 Thailand          8.6          Asia  2380747.00 0.75304329
## 129                     Togo         90.1        Africa    22137.00 0.17467067
## 130                    Tonga         27.5          Asia     1225.00 0.22962116
## 131      Trinidad and Tobago         32.9 South America   102583.00 0.36635539
## 132                  Tunisia         18.3        Africa   393146.00 0.24119623
## 133                   Turkey          3.0  Asia, Europe  3611252.00 0.50876945
## 134             Turkmenistan         92.5          Asia   146302.00 0.11770969
## 135                   Uganda         87.8        Africa   120881.00 0.10177443
## 136                  Ukraine          4.0        Europe   949042.00 0.21754150
## 137     United Arab Emirates           NA          Asia  1998983.00 0.46262041
## 138                  Uruguay          2.9 South America   115791.00 0.24814370
## 139               Uzbekistan         96.4          Asia   781532.00 0.18048804
## 140                  Vanuatu         72.3       Oceania     1917.00 0.19699873
## 141                   Zambia         87.2        Africa    78041.00 0.11706393
##           FIEI
```

```
## 1   0.6068800
## 2   0.7490481
## 3   0.5739094
## 4   0.8255330
## 5   0.6067996
## 6   0.8023205
## 7   0.7859553
## 8   0.7101875
## 9   0.7649003
## 10  0.7150669
## 11  0.7738822
## 12  0.5258123
## 13  0.6839244
## 14  0.6884582
## 15  0.6857747
## 16  0.6644014
## 17  0.4965359
## 18  0.7194204
## 19  0.7077064
## 20  0.4978104
## 21  0.5789372
## 22  0.5428803
## 23  0.7923923
## 24  0.3007314
## 25  0.4743631
## 26  0.5292681
## 27  0.6193316
## 28  0.2159101
## 29  0.5844653
## 30  0.7002185
## 31  0.6024292
## 32  0.6407437
## 33  0.6965424
## 34  0.6408365
## 35  0.5340516
## 36  0.5414830
## 37  0.7811567
## 38  0.5733249
## 39  0.5422373
## 40  0.5244487
## 41  0.6603419
## 42  0.7604579
## 43  0.6107052
## 44  0.7914551
## 45  0.8021559
## 46  0.5130172
## 47  0.7209770
## 48  0.6674078
## 49  0.4231702
## 50  0.7212464
## 51  0.6146590
## 52  0.6406029
## 53  0.5508823
## 54  0.6115757
```

```
## 55  0.6155873
## 56  0.6356684
## 57  0.5335459
## 58  0.6824715
## 59  0.3833873
## 60  0.5957979
## 61  0.6680098
## 62  0.6429925
## 63  0.7457603
## 64  0.5277964
## 65  0.5394618
## 66  0.8389234
## 67  0.7146253
## 68  0.5675504
## 69  0.5957358
## 70  0.4594176
## 71  0.7854659
## 72  0.6486675
## 73  0.7823648
## 74  0.5072089
## 75  0.3555835
## 76  0.6509023
## 77  0.6431199
## 78  0.7782964
## 79  0.5547301
## 80  0.3243614
## 81  0.8093556
## 82  0.7286226
## 83  0.7491804
## 84  0.7788513
## 85  0.5000512
## 86  0.7233781
## 87  0.6097551
## 88  0.6786496
## 89  0.6636147
## 90  0.5102428
## 91  0.7557051
## 92  0.7072083
## 93  0.8272909
## 94  0.8061357
## 95  0.5411191
## 96  0.7327073
## 97  0.5812165
## 98  0.6477642
## 99  0.7759047
## 100 0.7519640
## 101 0.7307849
## 102 0.6428054
## 103 0.2435544
## 104 0.5987251
## 105 0.7394740
## 106 0.7595357
## 107 0.6869475
## 108 0.8261111
```

```
## 109 0.7117417
## 110 0.3784485
## 111 0.5927287
## 112 0.3698774
## 113 0.4590433
## 114 0.7524770
## 115 0.3570445
## 116 0.7922577
## 117 0.7584446
## 118 0.4311946
## 119 0.7421878
## 120 0.3738774
## 121 0.7522793
## 122 0.7554075
## 123 0.6141043
## 124 0.6455230
## 125 0.7961733
## 126 0.7441428
## 127 0.3894490
## 128 0.7742626
## 129 0.6713957
## 130 0.7640336
## 131 0.6548195
## 132 0.6164912
## 133 0.6125297
## 134 0.7120350
## 135 0.4874557
## 136 0.4890976
## 137 0.3977882
## 138 0.5264210
## 139 0.4847626
## 140 0.5880392
## 141 0.4521352
```

```
row.names(finaldata)= NULL
```

Once we have imported the data set, we need to ensure the data structure is fit for analysis. To learn more about the data, we call the functions str to learn details of data frame.

```
### veryfying data structure

str(finaldata, width = 50, strict.width = 'cut')
```

```
## 'data.frame':    141 obs. of  6 variables:
##  $ Country     : chr  "Albania" "Algeria" "Ango"..
##  $ lessthan5_50: num  33.8 28.6 89.3 NA 12.2 0.7..
##  $ Continent   : chr  "Europe" "Africa" "Africa"..
##  $ FPI         : num  67629 996868 67381 10557 8..
##  $ FDI         : num  0.199 0.145 0.157 0.314 0...
##  $ FIEI        : num  0.607 0.749 0.574 0.826 0...
```

Now to conduct a regression analysis, the first process is to develop hypothesis against which the anlysis will be conducted. For the purpose of this anlysis, we have developed 3 set of hypothesis

Hypothesis 1 FDI decreases as percentage of population earning less than $5.50 increases

Hypothesis 2 FPI decreases as percentage of population earning less than $5.50 increases

Hypothesis 3 Percentage of population earning less than $5.50 decreases as FPI and FIEI advances

```
## hypothesis 1 : FDI decreases as percentage of population earning less than $5.50 increases

hypo1 = formula(FDI~ lessthan5_50)

#hypothesis 2:  FPI decreases as percentage of population earning less than $5.50 increases

hypo2 = formula(FPI~ lessthan5_50)

#hypothesis 3: Percentage of population earning less than $5.50 decreases as FPI and FIEI advances

hypo3 =formula(lessthan5_50~ FPI*FIEI )
```

After explaining the hypothesis, we need to get the results. Since the dependent variables are not a binary outcome, we can use OLS regression for analysis.

The regression analysis required uses the code glm. This code fits the generalized linear models.We can observe results below:

```
### Getting results

Result1 = glm(hypo1,
              data = finaldata,
              family = 'gaussian')

Result2 = glm(hypo2,
              data = finaldata,
              family = 'gaussian')

Result3 = glm(hypo3,
              data = finaldata,
              family = 'gaussian')
```

Reading results: We call the functions summary to obtain results for each of our hypotheses.

Interpreting from the summary of result, we can learn that:

For the first hypothesis: Can we observe a decrease in FDI when the poverty rate (percentage of people earning less than 5.50 falls) Interpretation for hypothesis 1. We can observe an indirect relationship between poverty and FDI like we had initially hypothesize.

Similar reading can be done for hypothesis 2 and hypothesis 3.

```
### Seeing results

# For the first hypothesis: Can we observe a decrease in FDI when the poverty rate (percentage of peopl

summary(Result1)
```

```
##
## Call:
```

```
## glm(formula = hypo1, family = "gaussian", data = finaldata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.33148  -0.10082  -0.00274   0.07349   0.45416
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5097954  0.0217208   23.47   <2e-16 ***
## lessthan5_50 -0.0045418  0.0004019  -11.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02599656)
##
##     Null deviance: 6.5442  on 125  degrees of freedom
## Residual deviance: 3.2236  on 124  degrees of freedom
##   (15 observations deleted due to missingness)
## AIC: -98.317
##
## Number of Fisher Scoring iterations: 2
```

```
### Interpretation for hypothesis 1. We can observe an indirect relationship between poverty and FDI li
```

```
### Results for hypothesis 2

summary(Result2)
```

```
##
## Call:
## glm(formula = hypo2, family = "gaussian", data = finaldata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2055528  -1189794   -476872    114913  14283255
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2163918     298266   7.255 3.83e-11 ***
## lessthan5_50   -24318       5518  -4.407 2.24e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 4.901977e+12)
##
##     Null deviance: 7.0305e+14  on 125  degrees of freedom
## Residual deviance: 6.0785e+14  on 124  degrees of freedom
##   (15 observations deleted due to missingness)
## AIC: 4043.4
##
## Number of Fisher Scoring iterations: 2
```

```
# The results also show an indirect relationship
```

```
## results for hypothesis 3

summary(Result3)
```

```
##
## Call:
## glm(formula = hypo3, family = "gaussian", data = finaldata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -67.421  -26.629   -5.349   25.517   68.242
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.047e+02  1.489e+01   7.029 1.29e-10 ***
## FPI         -5.681e-06  8.152e-06  -0.697 0.487165
## FIEI        -9.518e+01  2.405e+01  -3.958 0.000128 ***
## FPI:FIEI     2.418e-06  1.162e-05   0.208 0.835550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 998.722)
##
##     Null deviance: 160982  on 125  degrees of freedom
## Residual deviance: 121844  on 122  degrees of freedom
##   (15 observations deleted due to missingness)
## AIC: 1233.7
##
## Number of Fisher Scoring iterations: 2
```

Now after obtaining the regression values, we need to test for the better model. The first way to do this by using a chi-square distribution test.

```
### Searching for a better model

anova(Result1, Result2, test = "Chisq")
```

```
## Warning in anova.glmlist(c(list(object), dotargs), dispersion = dispersion, :
## models with response '"FPI"' removed because response differs from model 1
```

```
## Analysis of Deviance Table
##
## Model: gaussian, link: identity
##
## Response: FDI
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
```

```
## NULL                           125     6.5442
## lessthan5_50  1   3.3207       124     3.2236 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Result2, Result3, test = "Chisq")
```

```
## Warning in anova.glmlist(c(list(object), dotargs), dispersion = dispersion, :
## models with response '"lessthan5_50"' removed because response differs from
## model 1
```

```
## Analysis of Deviance Table
##
## Model: gaussian, link: identity
##
## Response: FPI
##
## Terms added sequentially (first to last)
##
##
##              Df   Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                             125 7.0305e+14
## lessthan5_50  1 9.5202e+13       124 6.0785e+14 1.048e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now,since Result 2 is better and Result 1 and Result 3 is better than result 2, we can conclude that Hypothesis 3, i.e Result 3 fits the model best.

To confirm the results, we can also test the residual value for all three hypothesis. No test the residual, we call the package rsq for library.

```
## Checking the Rsquare to understand the residual value for all 3 hypothesis
library(rsq)
rsq(Result1, adj = T)
```

```
## Warning in cbind(y, yfit): number of rows of result is not a multiple of vector
## length (arg 1)
```

```
## Warning in wt * vresidual(y, f0$fitted.values, family = family(f0))^2: longer
## object length is not a multiple of shorter object length
```

```
## [1] 0.5562347
```

```
rsq(Result2, adj = T)
```

```
## Warning in cbind(y, yfit): number of rows of result is not a multiple of vector
## length (arg 1)
```

```
## Warning in cbind(y, yfit): longer object length is not a multiple of shorter
## object length
```

```
## [1] 0.1776667
```

```
rsq(Result3, adj = T)
```

```
## [1] 0.2245085
```

We can now proceed to visualize the data. For this, we need to call two package from library, dotwhisker and ggplot.

Using the code dwplot, we plot the predicted value of Result 3 against 2 standard deviation.

```
# summary plots to visualize the data

library(dotwhisker)
```

```
## Warning: package 'dotwhisker' was built under R version 4.0.4
```

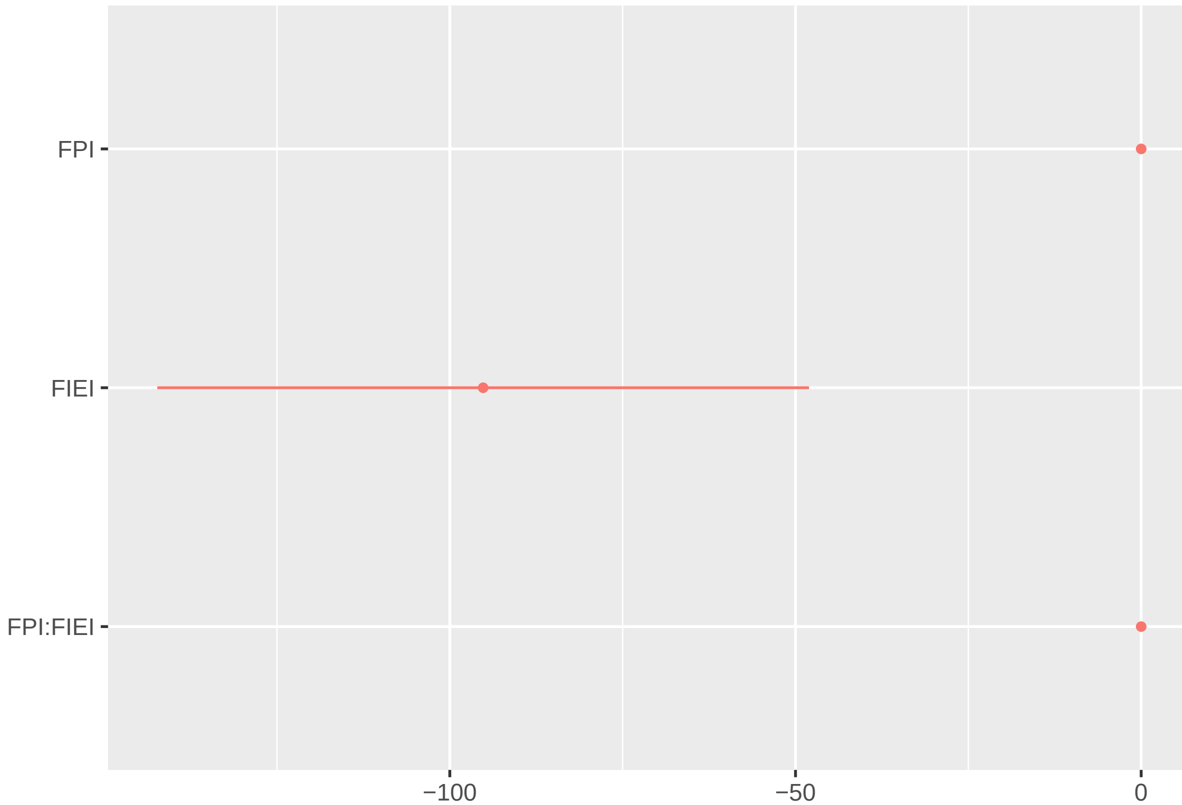
```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.4
```

```
## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.
## TMB was built with Matrix version 1.3.2
## Current Matrix version is 1.2.18
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a
```

```
## Registered S3 method overwritten by 'broom.mixed':
##   method      from
##   tidy.gamlss broom
```

```
library(ggplot2)

dwplot(Result3,by_2sd = F)
```



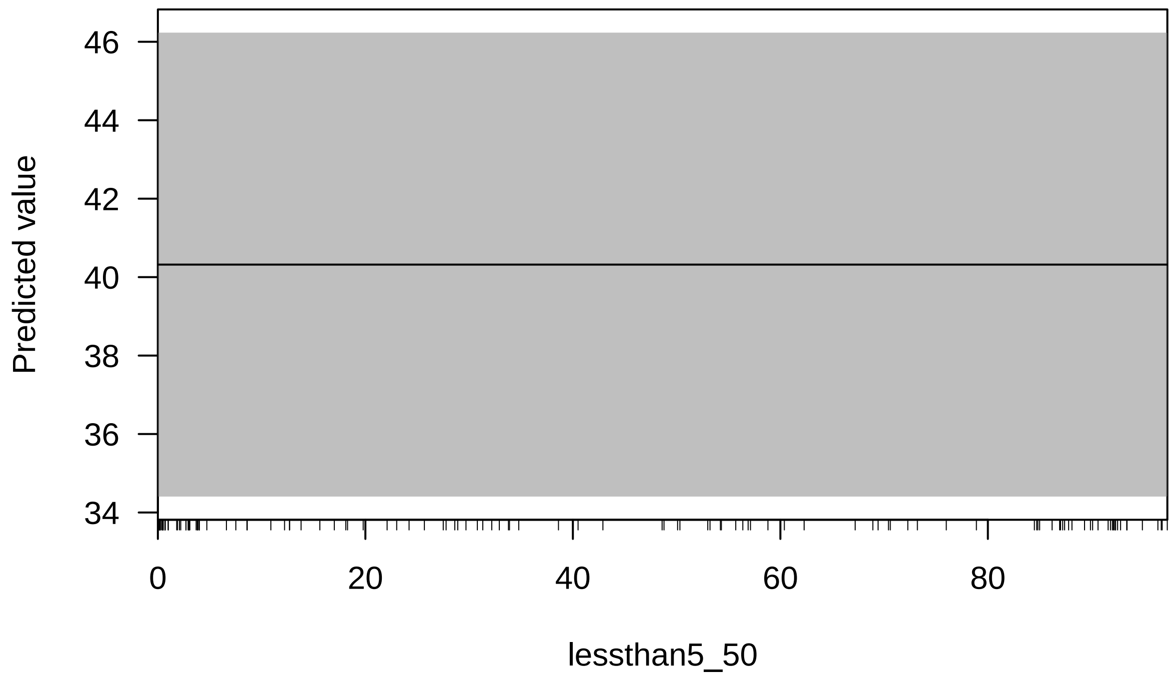
To capture the result of predicted value againsts given value, we can also use margins. For this, we must call margins from library.

The two graph uses data value collected for Result 3 and result 1.

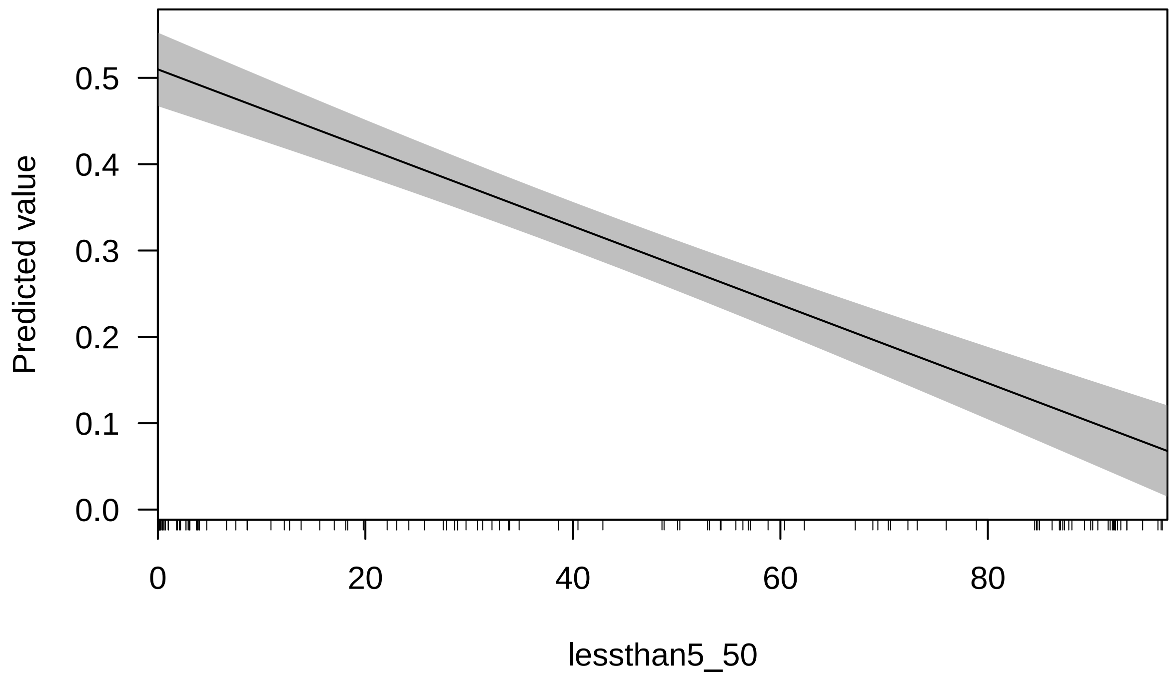
```
# Using the margins library

library(margins)
```

```
## Warning: package 'margins' was built under R version 4.0.4
```

```
cplot(Result3,'lessthan5_50')
```

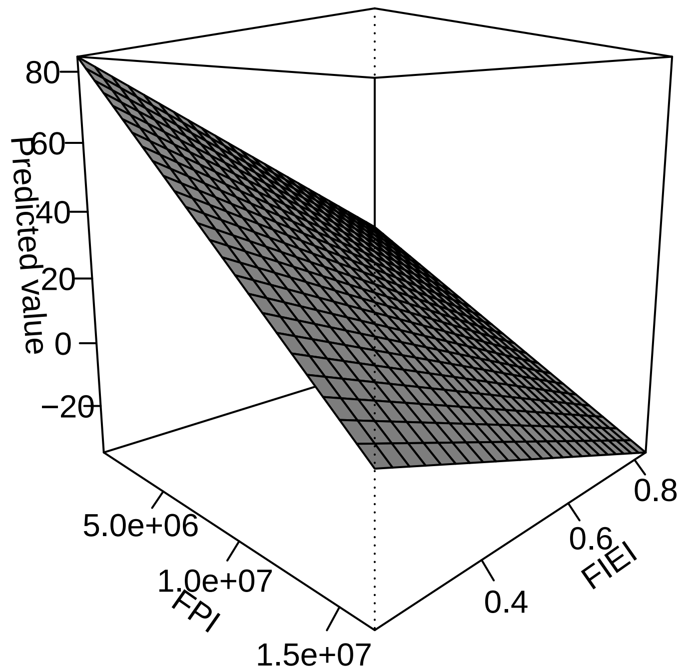



```
cplot(Result1,'lessthan5_50')
```



Another way to visaluze the result is looking for interactions between our values. To do this, we can call for perspe code, as can be oberved below.

```
## Looking into the interactions
persp(Result3)
```



The above codes above, therefore, shows a tutorial in conducting regression analysis in R.

Thank you!