STA 108 Project

Jieyi Chen, Qiwen Guan

November 4, 2015

Introduction: The interest of this project is to make inferences about the number of active physicians in terms of the following predictor variables: total population, number of hospital beds and total personal income. Moreover, we also split the dataset based on four different regions and conduct analysis on their relationships. In this report, we use fitting regression models, ANOVA models, confidence intervals, residual plots and Q-Q plots to make analysis on the CDI dataset that provides us information about populous counties in the US.

In this project, we first find the estimated regression functions, then, we plot them on separate graphs and check whether a linear regression relation provides a good fit for each of the three predictor variables. Secondly, we measure and compare the MSE because we want to check which predictor variable has the smallest variability around the fitted regression line. Thirdly, we separate the dataset based on four different regions and find the estimated regression functions for each region. We compare the slope for each estimated regression function and see whether the linear relationship between per capita income and the percentage of individuals in a county having at least a bachelor's degree is positive for all regions. Fourthly, we conduct analysis on MSE and conclude which region has a relatively higher variability around the fitted regression line. Lastly, we also find the 90% confidence interval for B1/ slope for all regions; prepare a residual plot and a normal probability plot to conclude which linear regression model is more appropriate.

Part I: Fitting regression models.

```
The estimated regression functions for Number of active physicians on Total p
opulation is Yihat =  -1.106348e+02 + 2.795425e-03X1i.
The estimated regression functions for Number of active physicians on Number 
of hospital beds is Yihat = -95.9321847 + 0.7431164X2i.
The estimated regression functions for Number of active physicians on Total p
ersonal income is Yihat = -48.3948489 + 0.1317012X3i.

##      (Intercept) Total_population
##    -1.106348e+02     2.795425e-03

## 
## Call:
## lm(formula = Number_of_active_physicians ~ Total_population, 
##     data = CDI)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1969.4  -209.2   -88.0    27.9  3928.7 
```

```
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.106e+02  3.475e+01  -3.184  0.00156 **
## Total_population  2.795e-03  4.837e-05  57.793  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 610.1 on 438 degrees of freedom
## Multiple R-squared:  0.8841, Adjusted R-squared:  0.8838
## F-statistic:  3340 on 1 and 438 DF,  p-value: < 2.2e-16
```

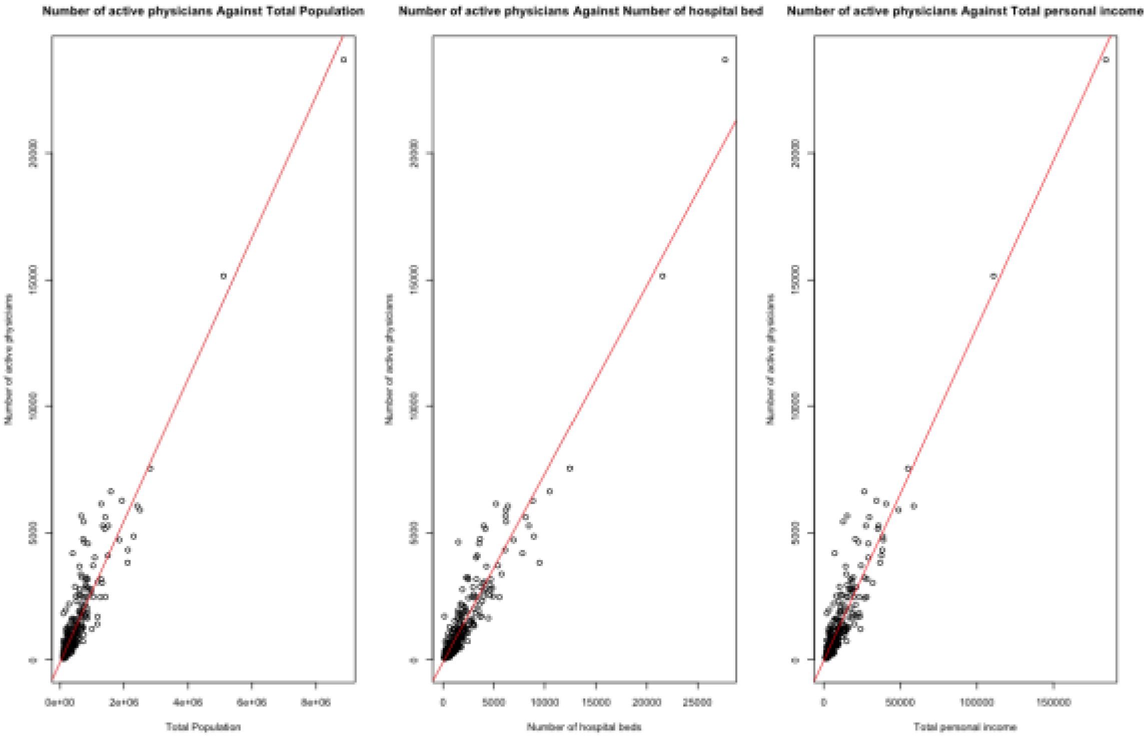
```
##             (Intercept) Number_of_hospital_beds
##             -95.9321847               0.7431164
```

```
##
## Call:
## lm(formula = Number_of_active_physicians ~ Number_of_hospital_beds,
##     data = CDI)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3133.2  -216.8   -32.0    96.2  3611.1
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -95.93218   31.49396  -3.046  0.00246 **
## Number_of_hospital_beds   0.74312    0.01161  63.995  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 556.9 on 438 degrees of freedom
## Multiple R-squared:  0.9034, Adjusted R-squared:  0.9032
## F-statistic:  4095 on 1 and 438 DF,  p-value: < 2.2e-16
```

```
##           (Intercept) Total_personal_income
##           -48.3948489             0.1317012
```

```
##
## Call:
## lm(formula = Number_of_active_physicians ~ Total_personal_income,
##     data = CDI)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1926.6  -194.5   -66.6    44.2  3819.0
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -48.39485   31.83333   -1.52    0.129
## Total_personal_income   0.13170    0.00211   62.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 569.7 on 438 degrees of freedom
## Multiple R-squared:  0.8989, Adjusted R-squared:  0.8987
## F-statistic:  3895 on 1 and 438 DF,  p-value: < 2.2e-16
```

As we can see from the graphs, the linear regression relation appears to provide a good fit for each of the three predictor variables even though there are a few outliers.



```
MSE for Total population is 372203.5.
MSE for Number of hospital beds is 310191.9.
MSE for Total personal income is 324539.4.
Number of hospital beds leads to the smallest variability around the fitted r
egression line because it has the smallest MSE.

## [1] 372203.5

## [1] 372203.5 310191.9 324539.4

The estimated regression functions for NE is Yi = 9223.8156 + 522.1588Xi
The estimated regression functions for NC is Yi = 13581.4052 + 238.6694Xi
The estimated regression functions for S is Yi = 10529.7851 + 330.6117Xi
The estimated regression functions for W is Yi = 8615.0527 + 440.3157Xi

##                (Intercept) Percent_bachelors_degrees
##                  9223.8156                  522.1588

##                (Intercept) Percent_bachelors_degrees
##                 13581.4052                  238.6694
##                (Intercept) Percent_bachelors_degrees
##                 10529.7851                  330.6117
```

```
##                (Intercept) Percent_bachelors_degrees
##                  8615.0527                  440.3157
```

```
As we can see from the regression funcitons, the directions of the relationsh
ip between per capita income in a CDI (Y) and the percentage of individuals i
n a county having at least a bachelor's degree (X) are the same for all regio
ns because their slopes are all positive. Therefore, per capita income increa
ses when the percentage of individuals in a county having at least a bachelor
's degree increases. However, the increase in region NE is the highest and NC
is the slowest because NE has the largest slope value and NC has the smallest
slope value.

The MSE for NC is 4411341.
The MSE for NE is 7335008.
The MSE for S is 7474349.
The MSE for W is 8214318.
The variability around the fitted regression line is relatively higher in W b
ecause it has the largest MSE. It is approximately the same for NE and S beca
use their MSE are relatively close. The MSE for NC is the smallest and theref
ore it has the smallest variability.
```

```
## [1] 4411341
```

```
## [1] 7335008
```

```
## [1] 7474349
```

```
## [1] 8214318
```

Part II: Measuring linear associations.

```
The number of hospital beds accounts for the largest reduction in the variabi
lity in the number of active physicians because its R^2 is closest to 1.
```

```
## [1] 0.8840674
```

```
## [1] 0.9033826
```

```
## [1] 0.8989137
```

Part III. Inference about regression parameters

```
The 90 percent confidence coefficient for NC is (193.4858, 283.853).
The 90 percent confidence coefficient for NE is (460.5177, 583.8).
The 90 percent confidence coefficient for S is (285.7076, 375.5158).
The 90 percent confidence coefficient for W is (364.7585, 515.8729).
No, the regression lines for different regions do not appear to have similar
slopes because their confidence intervals varies.
```

```
##                                5 %    95 %
## Percent_bachelors_degrees 193.4858 283.853
##                                5 %  95 %
## Percent_bachelors_degrees 460.5177 583.8
```

```
##                                5 %     95 %
## Percent_bachelors_degrees 285.7076 375.5158
```

```
##                                5 %     95 %
## Percent_bachelors_degrees 364.7585 515.8729
```

```
## Analysis of Variance Table
## 
## Response: Per_capita_income
##                            Df     Sum Sq    Mean Sq F value    Pr(>F)    
## Percent_bachelors_degrees   1 1450517671 1450517671  197.75 < 2.2e-16 ***
## Residuals                 101  740835765    7335008                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
## 
## Response: Per_capita_income
##                            Df    Sum Sq   Mean Sq F value    Pr(>F)    
## Percent_bachelors_degrees   1 338907694 338907694  76.826 3.344e-14 ***
## Residuals                 106 467602149   4411341                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

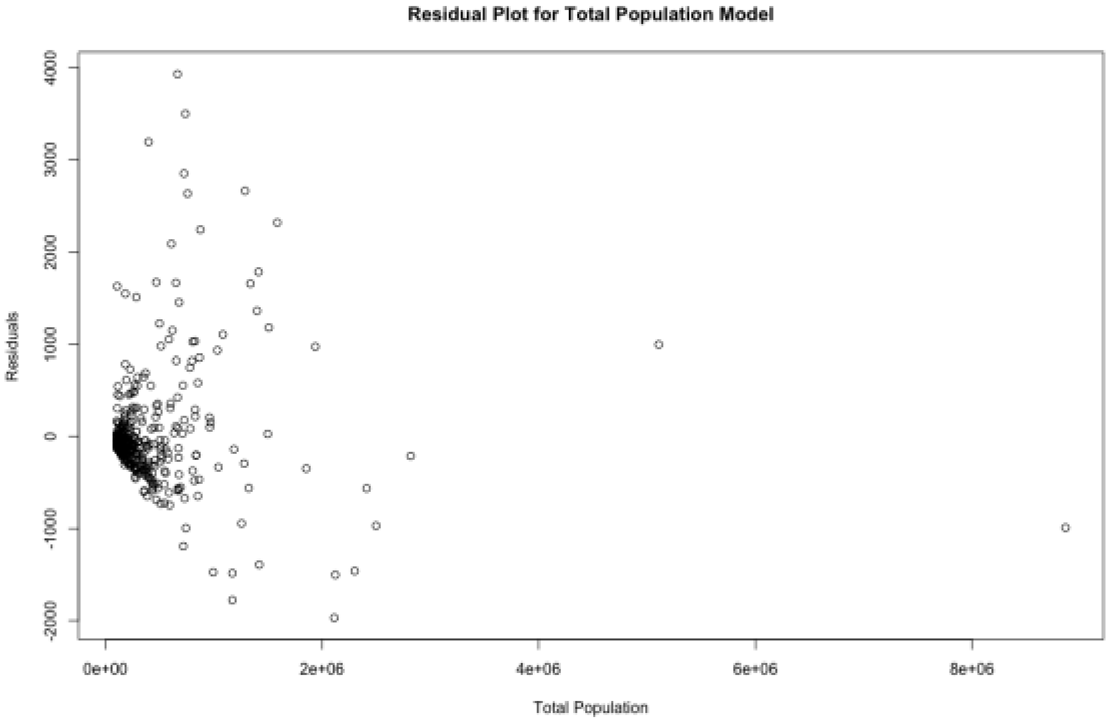
```
## Analysis of Variance Table
## 
## Response: Per_capita_income
##                            Df     Sum Sq    Mean Sq F value    Pr(>F)    
## Percent_bachelors_degrees   1 1109873245 1109873245  148.49 < 2.2e-16 ***
## Residuals                 150 1121152411    7474349                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

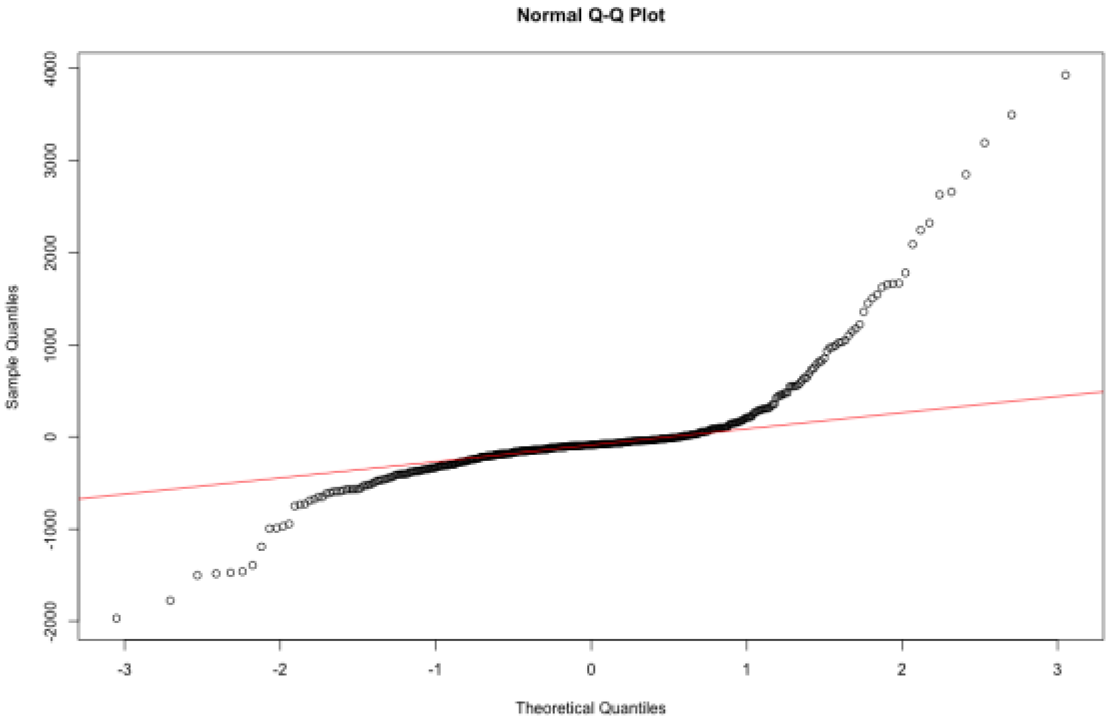
```
## Analysis of Variance Table
## 
## Response: Per_capita_income
##                           Df    Sum Sq   Mean Sq F value    Pr(>F)    
## Percent_bachelors_degrees  1 773745787 773745787  94.195 6.856e-15 ***
## Residuals                 75 616073841   8214318                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part IV: Regression diagnostics.

The residuals for Total Population is

```
##            1            2            3            4            5 
##  -988.674369   992.803480  -214.428820  -967.381296  -565.893436 
##            6            7            8            9           10 
## -1459.698605 -1501.539153 -1969.427598   969.634018  -350.756392
##           11           12           13           14           15
##  2319.273385  1177.037745    25.270774 -1391.339962  1783.103490
##           16           17           18           19           20
##  1359.322538  1655.691994  -563.536734  2658.950151  -296.222410
##           21           22           23           24           25
##  -942.987612  -141.045101 -1776.166841 -1483.300259  1101.534742
##           26           27           28           29           30
##  -336.014719   930.551475 -1472.700907   151.176329    98.009868
##           31           32           33           34           35
##   203.053577  2242.012596   853.159476  -470.264923   580.539388
##           36           37           38           39           40
##  -650.113979  -201.986164  -208.900524   214.015354   289.691337
##           41           42           43           44           45
##  1029.346991  -483.655301  1023.500977  -375.137651   818.236677
##           46           47           48           49           50
##    82.546194   746.218364  2629.422668  -993.058158  3497.162931
##           51           52           53           54           55
##  -674.979133   175.279314  2847.861778 -1189.803031   552.790868
##           56           57           58           59           60
##    28.707764  -555.173826  1452.026418  -416.240419  -232.605099
##           61           62           63           64           65
##  -130.275748  -577.523525  -591.549193  -567.283375    87.421834
##           66           67           68           69           70
##   422.853345  3928.735428   821.345583   108.662481  1664.541007
##           71           72           73           74           75
##    35.482291  1146.409850  2088.091419   356.467270   305.806785
##           76           77           78           79           80
##  -748.166535  -613.051673  1049.422531  -246.667691  -180.405177
##           81           82           83           84           85
##  -135.581813  -399.511928   -46.282453  -379.211552  -726.614349
##           86           87           88           89           90
##  -523.176486  -239.972948  -112.981081  -205.037755   975.776476
##           91           92           93           94           95
##  -732.758724  -138.417176   -39.971942  -281.841575  1221.481929
##           96           97           98           99          100
##    92.200859   264.715211  -566.924691   352.217875  -511.852522
##          101          102          103          104          105
##   333.655986  1670.466150  -689.032454   209.139335  -260.945549
##          106          107          108          109          110
##  -324.860418    94.429032  -519.002735   -86.792066  -563.963352
##          111          112          113          114          115
##  -530.146582  -587.351666  -400.681020  -467.679752  -311.892208
##          116          117          118          119          120
##  -497.332361    79.774118   549.496859  -463.580151   -87.772020
##          121          122          123          124          125
##  -441.678250  -451.188035  3190.731659  -274.742548  -309.160084
##          126          127          128          129          130
##  -400.804812  -134.610660  -650.536458  -110.920196   -97.833285
##          131          132          133          134          135
##  -347.551965  -430.662770  -367.131640  -402.569759   -47.576621
##          136          137          138          139          140
##  -321.921477   687.476236   -47.531139  -590.206872  -368.586035
##          141          142          143          144          145
##   291.107992  -604.898617   643.186514  -184.551734  -359.986555
##          146          147          148          149          150
##  -378.618830  -263.089989  -241.832048   156.073670  -317.148440
##          151          152          153          154          155
##  -212.681101   193.818262  -134.368020  -303.831554  -403.316946
##          156          157          158          159          160
##  -407.291025  -270.369300   -54.173115  -178.329661   642.875417
##          161          162          163          164          165
##   549.710230  -396.172871  -119.239199   -66.197268  -336.130943
##          166          167          168          169          170
##   310.925981    50.440339  1507.705648  -183.224466    51.570959
##          171          172          173          174          175
##  -367.245814  -206.869450   -34.867164  -191.706047  -457.741626
##          176          177          178          179          180
##  -365.893850  -437.983813  -356.277588   561.392801   299.519358
##          181          182          183          184          185
##  -191.259042   -28.610503  -345.458026  -335.211266  -295.148243
##          186          187          188          189          190
##  -101.997290   481.388732  -346.511142   -79.671753  -116.078363
##          191          192          193          194          195
##     1.732310   475.364076   -75.364515  -153.558418  -311.091329
##          196          197          198          199          200
##   314.662165  -121.128687  -162.980530  -233.606452   162.969149
##          201          202          203          204          205
##   221.952883  -109.581552  -158.225771  -298.824501   269.846401
##          206          207          208          209          210
##   468.418694  -333.542170  -102.569362  -152.072285  -301.199351
##          211          212          213          214          215
##  -148.367585  -288.512694   728.641054   -67.283469  -137.454499
##          216          217          218          219          220
##   100.106110   143.272564  -267.472037   -72.672293   452.912204
##          221          222          223          224          225
##   -72.425280    -1.790719  -296.299233  -258.776489   -97.582589
##          226          227          228          229          230
##  -312.530747   -22.442565  -146.844344  -159.176237   -54.846633
##          231          232          233          234          235
##    19.328969    99.251459   -97.593777  -225.264426   -37.454774
##          236          237          238          239          240
##   -60.191498  -199.006238  -144.338132   -14.464944  -269.267740
##          241          242          243          244          245
##    89.462628  -186.743980   612.289565  -117.044362  -186.205735
##          246          247          248          249          250
##  -248.740170  -251.256561   281.973426   -75.005482  -131.694428
##          251          252          253          254          255
##  -212.574225  -241.601673   180.555633  -273.694431  -151.501547
##          256          257          258          259          260
##  -71.468002 -207.277913 1546.328694  782.996801 -181.108663
##         261         262         263         264         265
##  -30.523404 -203.077407  211.970115 -236.824551 -144.659621
##         266         267         268         269         270
##   90.935805 -159.476394  -98.026331    8.956388  -43.403460
##         271         272         273         274         275
## -307.445138  247.796030  -10.713246 -111.371189 -212.909182
##         276         277         278         279         280
##  -38.630658  -93.460137   60.227538   11.054983 -177.455817
##         281         282         283         284         285
##  -44.100545  151.795515 -141.485299  -57.093939  -75.806011
##         286         287         288         289         290
##  111.638462 -151.790256  -65.964844 -162.004998 -108.059891
##         291         292         293         294         295
## -210.214149 -168.905128 -147.560528  -32.546551 -187.306145
##         296         297         298         299         300
## -102.202714   36.661834  -11.138166 -137.260403 -227.201699
##         301         302         303         304         305
## -133.156972  129.413294 -204.226096 -180.996871   96.319774
##         306         307         308         309         310
##  -96.109197 -128.824573  -33.393315  -31.177305   48.914944
##         311         312         313         314         315
##  -35.584675 -184.206531 -153.398653 -171.683024  -96.249733
##         316         317         318         319         320
##  -11.326481 -117.054563 -113.280230   86.222946   50.341116
##         321         322         323         324         325
## -186.950879  -18.849482  -23.402214 -122.022036   55.222690
##         326         327         328         329         330
##  -54.150372  -39.619242 -191.469560 -108.354948   96.086730
##         331         332         333         334         335
##  -94.282013 -137.586462  -97.492689  -97.173248 -100.373504
##         336         337         338         339         340
## -179.833987 -154.824838  -85.860926  438.306800  -11.120138
##         341         342         343         344         345
## -135.955208 -122.406543 -114.640596  -22.442121  -41.285577
##         346         347         348         349         350
##  -79.478209  -79.784943  -85.133609 -147.018235 -107.685579
##         351         352         353         354         355
##  -43.564614  -66.367410 -165.481260  -97.971731  -94.574780
##         356         357         358         359         360
##  140.536018  -57.399687 -137.136917  -92.242381   25.654951
##         361         362         363         364         365
##   11.536271 -106.396638  -60.021289 -150.766906  -91.840858
##         366         367         368         369         370
##  -77.620019  -99.443908  -43.702358   71.389891   12.169815
##         371         372         373         374         375
##   -6.862968  -19.851787  -15.350643  -68.836285  -22.436539
##         376         377         378         379         380
##   10.789890  -87.818751  -41.377073  -46.005791 -153.693975
##         381         382         383         384         385
##   -80.729553    27.633852  -144.642133   -39.957254   -41.518372
##          386          387          388          389          390
##     1.994715   -45.328030    55.580483     1.608437   -78.000203
##          391          392          393          394          395
##   -62.018247   542.487725   -70.380128  -100.678476   -74.015961
##          396          397          398          399          400
##   451.831053   -76.637816   -16.587499   -75.498045   145.999541
##          401          402          403          404          405
##   -84.553191   -34.878732   -28.177080  -112.913548   -22.005035
##          406          407          408          409          410
##    63.112373  -112.042647   -59.866535   304.941343    -6.644934
##          411          412          413          414          415
##   -92.510754   -76.471618   -39.660183   -84.232483   -70.182165
##          416          417          418          419          420
##   -48.388264  -118.206562  1627.005890    -6.979370   -21.641124
##          421          422          423          424          425
##    14.391150  -100.385216   -16.739473   -68.951163   -41.740744
##          426          427          428          429          430
##   -65.745573   -92.582676   -82.079500   166.058238   -88.279246
##          431          432          433          434          435
##  -137.128293   -52.966158   -91.014190   -88.991826  -105.133631
##          436          437          438          439          440
##   -74.024609    21.576407   -83.299832    22.046800   -47.027914
```



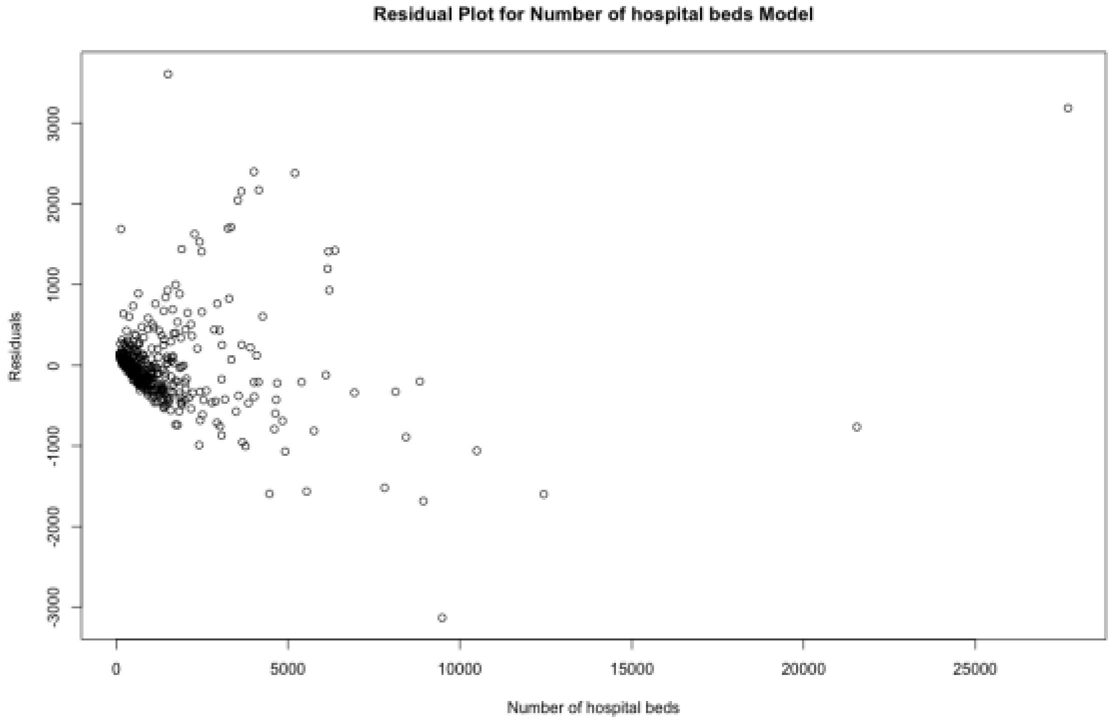


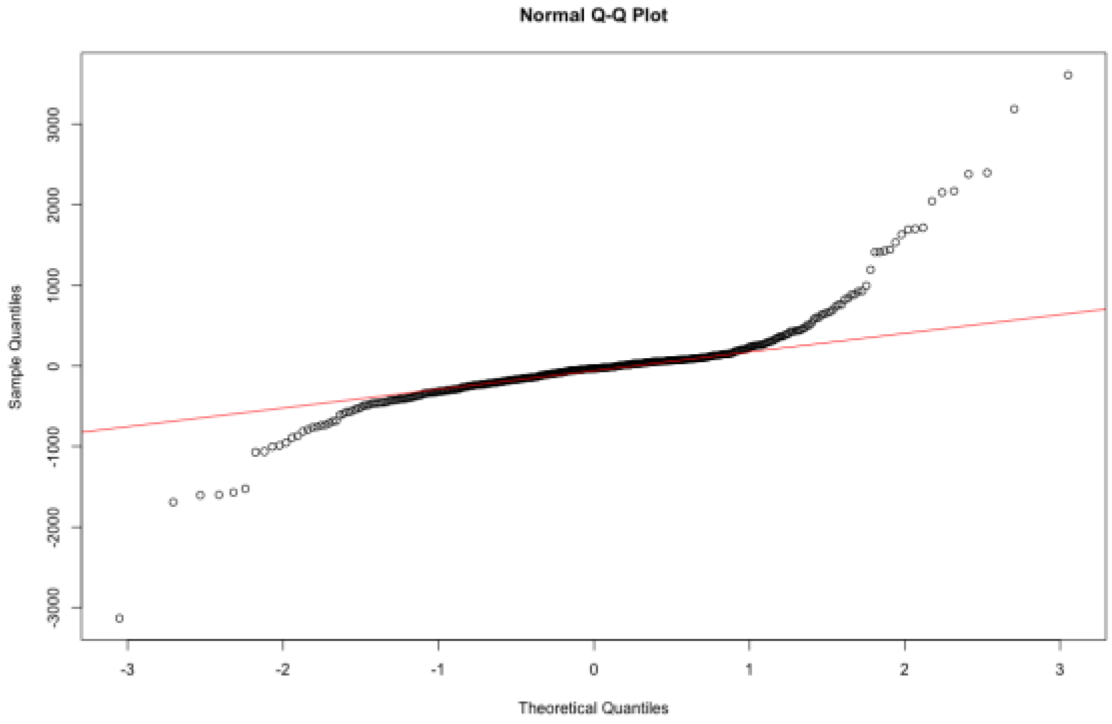
```
As we can see from the residual plot for total population model, there are tw
o outliers (5e+06, 1000),(9e+06, -1000). The residuals are not centered aroun
d the 0 line, which indicates that it is not a good fit.
In addition, the qqline is not linear, which means that the data is not norma
```

```
l. The normality probability plot shows departures from normality in the tail
s.
```

The residuals for Number of hospital beds is

```
##             1             2             3             4             5 
##  3188.6066863  -765.2271832 -1602.1244265  1409.2156773  1425.0235530 
##             6             7             8             9            10 
## -1688.0150574  -120.0505894 -3133.2428687  -199.2171801  -338.8372379 
##            11            12            13            14            15 
## -1061.3317785  2396.7783608  1713.4370289    70.2352138  -327.0907378 
##            16            17            18            19            20 
##  2168.5127093  -891.9981367   215.8055874  2378.7266760   824.5377827 
##            21            22            23            24            25 
## -1567.1622643   121.5583946  -328.5563564  -955.7913976  1697.8312760 
##            26            27            28            29            30 
##   438.3347372  -207.1810306   505.2012199  -598.1831838  -209.7078721 
##            31            32            33            34            35 
##  -207.3130098  2042.2999730  -220.0821225  -422.2882440   762.1698394 
##            36            37            38            39            40 
## -1596.8809226   505.3970355  -171.9490654   659.5997734 -1069.7144868 
##            41            42            43            44            45 
##  1628.3698089  1689.6114648  1409.5447050   533.4417980  -426.5317456 
##            46            47            48            49            50 
##   884.1116949  -690.4945206  3611.0557037  -429.4749841   930.3808827 
##            51            52            53            54            55 
##   366.5322458   647.2224470  2151.9883286   -92.2937809   247.3628866 
##            56            57            58            59            60 
##   359.3604257   888.5945441  1530.8748081  -713.6558797  -315.2622478 
##            61            62            63            64            65 
##   335.8732700  -211.9887512   244.3764236    34.7534424   438.8094337 
##            66            67            68            69            70 
##  -377.8192394  1196.7935884  -791.8896905   689.7349837  -814.1891833 
##            71            72            73            74            75 
## -1004.6718775  1439.7815918   602.7699005  -573.3148554   204.1773769 
##            76            77            78            79            80 
##  -335.2099478    99.8714244   430.8948048   315.5140947  -870.9490654 
##            81            82            83            84            85 
##   -18.9615246  -683.5012879   289.8632717   446.7788234  -109.3488494 
##            86            87            88            89            90 
##   -91.4399124    75.0006352  -342.2450196  -291.9061486  -469.8367753 
##            91            92            93            94            95 
##  -302.5778911  -612.7763226  -165.7684774    17.9821765  -389.9096872 
##            96            97            98            99           100 
##  -430.3266994   760.7794386   -19.0831979   398.3132024  -990.2077946 
##           101           102           103           104           105 
##   390.4139562   249.0709313    98.5579344  -758.0225925  -237.3566945 
##           106           107           108           109           110 
##  -493.9064562   108.2762850   -12.5512797    28.1749161   425.4834844 
##           111           112           113           114           115 
##    67.7785158  -487.1645701  -571.9709078   253.4293388   -34.0644316
##           116           117           118           119           120
##  -252.2753222    54.6889908  -447.9312219  -302.2017951  -341.4020722
##           121           122           123           124           125
##  -147.3491570  -118.5052868 -1521.7797086   -39.6517258   381.0804694
##           126           127           128           129           130
##   127.5397833   342.9252624  -232.1479571  -457.1542642  -464.5028259
##           131           132           133           134           135
##     6.2650563  -411.5318982  -316.5778911  -352.5778911   204.7791310
##           136           137           138           139           140
##  -451.2377897   671.9452591   -52.3025489  -160.4326825   163.5760855
##           141           142           143           144           145
##   581.5494741    40.7040658  -463.6471118  -110.8901507   203.1446134
##           146           147           148           149           150
##  -150.7715535  -110.6520335   -97.3400814  -538.5754303  -288.2937809
##           151           152           153           154           155
##  -421.0268990  -191.2005647  -263.2199462  -369.6238840   -33.5522025
##           156           157           158           159           160
##   230.4559502  -157.7806290   279.8423522  -294.3394662    -0.6035797
##           161           162           163           164           165
##   427.0090955  -248.5240531  -364.8895355  -236.8347747   191.3551965
##           166           167           168           169           170
##   197.9906368  -334.9899817   998.3407366   106.9340304  -243.6692617
##           171           172           173           174           175
##  -295.3403891   301.5570116  -285.9355284  -489.0638164    86.6302311
##           176           177           178           179           180
##   -48.5340514   -28.4151466  -312.3403891    94.2578264   473.3104340
##           181           182           183           184           185
##   298.6211556  -397.6783373   -46.3873048   -17.3231608  -383.5785063
##           186           187           188           189           190
##  -559.1642625   103.8078956  -247.1388816    80.2828997   199.2272161
##           191           192           193           194           195
##   -36.7530948  -401.9884436   -59.3037793  -139.5881971  -461.3582325
##           196           197           198           199           200
##  -394.2093326  -138.3954575  -175.9367588    56.9703325     7.5316306
##           201           202           203           204           205
##   601.0066347  -736.2005647  -198.3954575   129.7403680  -139.0087479
##           206           207           208           209           210
##   734.2913601    43.0157102   235.1167716   101.8601956  -366.4505259
##           211           212           213           214           215
##  -147.3406967   103.2541351   -31.5763531   260.3461209  -260.7530948
##           216           217           218           219           220
##   120.6408447  -409.8157008  -328.6707997  -319.2480956  -137.1091941
##           221           222           223           224           225
##   132.0254010  -236.4133010   178.2353688    62.1533813   -93.2124087
##           226           227           228           229           230
##  -187.5981954   -50.4508336  -455.6057329  -217.4323749  -303.9821365
##           231           232           233           234           235
##  -740.5306678  -332.3300831    49.0163254  -178.5888123  -531.7427888
##           236           237           238           239           240
##  -73.7721687   -4.2133315  -84.7174079 -102.6526487   15.6211556 
##          241          242          243          244          245 
##  183.8695788  147.6117724 -200.0169007  -19.2130239  -58.0021332 
##          246          247          248          249          250 
##   77.6393066   26.3367378  636.3639644  146.5754702   69.9975591 
##          251          252          253          254          255 
## -164.8275448  -19.7640160 -228.9267605  -31.7274062   95.4744089 
##          256          257          258          259          260 
##  -41.5155928   29.8138951  928.2299845  461.4765621  146.7034506 
##          261          262          263          264          265 
## -412.0278218   30.5019431 -292.4126858  -35.5346666  -65.4057635 
##          266          267          268          269          270 
##   26.1727628   75.8780391   68.5113263 -120.2765527  185.6486898 
##          271          272          273          274          275 
##  -13.9845974  130.5582420 -171.3131624 -284.0012104   48.4650257 
##          276          277          278          279          280 
## -158.8087785   47.0250934 -515.5960422 -299.5603552  -32.2868586 
##          281          282          283          284          285 
## -265.9276833 -167.3034717 -216.3319287    7.3739627   93.1533813 
##          286          287          288          289          290 
##   55.7322153 -212.5246683  -63.8735377  -95.2774755 -148.8275448 
##          291          292          293          294          295 
## -311.0475109 -228.2036407 -230.4971341   57.6489974   28.9787929 
##          296          297          298          299          300 
##  151.8777315 -145.8453882 -148.2859358   88.8411217   65.7216017 
##          301          302          303          304          305 
##  -43.8923040 -216.6154237  101.0976977 -202.8735377  -36.8547714 
##          306          307          308          309          310 
## -266.5243607  -75.1307288   15.1536890  -80.4514488  268.5567040 
##          311          312          313          314          315 
##  -28.9745990  140.5654719  -81.3510026  -84.0024409  -43.9289137 
##          316          317          318          319          320 
##   -2.5434346 -104.3416195  -84.2499413 -256.8081633  261.2998204 
##          321          322          323          324          325 
##  -14.8560018 -411.1382663 -185.7721687 -225.3228532  310.0886221 
##          326          327          328          329          330 
##  -55.8003181  -54.4332977   46.4556426  -97.9379893   26.5485512 
##          331          332          333          334          335 
## -119.8919963   73.8777315 -152.6717226   88.5842382   72.4099573 
##          336          337          338          339          340 
##   43.6849919  -80.9382969   59.1164640  366.1264623 -228.7812442 
##          341          342          343          344          345 
##   11.8320462   34.9787929 -250.8090861 -143.9467572 -177.0659696 
##          346          347          348          349          350 
## -170.0202843  -15.7089475  131.3730399   54.2538275  -12.2687076 
##          351          352          353          354          355 
##  139.9052657  -81.2774755    9.7034506 -104.1307288  121.8042044 
##          356          357          358          359          360 
## -205.7900122   73.2907449  -26.7733991  146.5288621  -57.4974417 
##          361          362          363          364          365
##    -0.6720302    52.9878684    58.6026969   -24.9661387    11.9697173
##           366           367           368           369           370
##  -239.3228532    81.9969439   -12.6263449   -18.7543252    59.4653334
##           371           372           373           374           375
##   -54.1673386   -47.1031946    71.8229707   -40.1401120  -196.2677847
##           376           377           378           379           380
##  -168.4695998    69.1621493   -69.8738453  -271.8272372   -72.6907965
##           381           382           383           384           385
##  -132.2224070   -32.3966879    94.8223554    79.1806080     8.8138951
##           386           387           388           389           390
##  -236.2859358   112.2813617   -53.7543252  -179.8366203   131.7857457
##           391           392           393           394           395
##    31.3458133    81.7240625   -52.9658311  -111.4611395   -34.9292213
##           396           397           398           399           400
##   255.9890988    75.5107110  -190.7174079   -29.7549404  -316.3034717
##           401           402           403           404           405
##    28.6117724    78.8411217  -152.2130239    13.7034506    56.1439982
##           406           407           408           409           410
##   262.2719786  -184.4792906   -24.7274062   158.6952979   144.6849919
##           411           412           413           414           415
##   -22.3147005    91.7766701  -140.7452497  -174.5985030    67.5382453
##           416           417           418           419           420
##   -12.5255911    57.1803004   842.0738547  -153.1028870    54.0889298
##           421           422           423           424           425
##   -97.6901812   -29.8835360   -50.1767217    13.4925600    58.8135875
##           426           427           428           429           430
##  -102.9198382    37.0244782    -0.5349743   -38.8916887    24.9052657
##           431           432           433           434           435
##    16.0335537    30.1806080    44.1712248    64.7491359    85.6480746
##           436           437           438           439           440
##   -21.5715840    83.0889298    43.2262933   152.6849919  -126.8738453
```


Residual Plot for Number of hospital beds Model

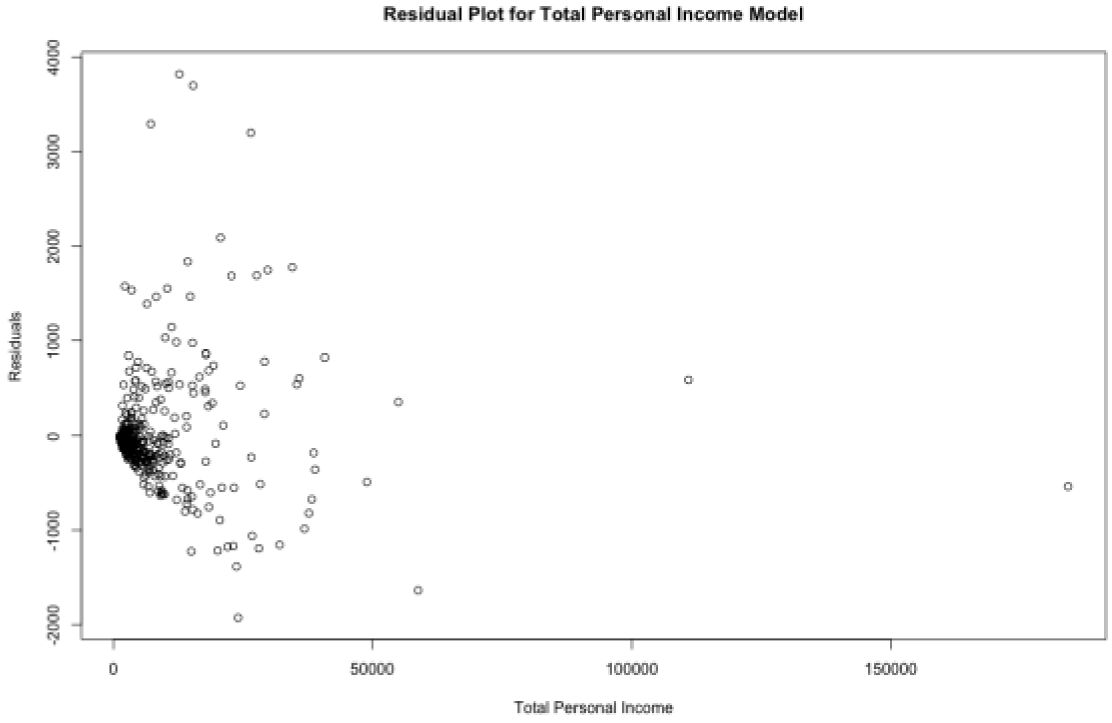

Normal Q-Q Plot

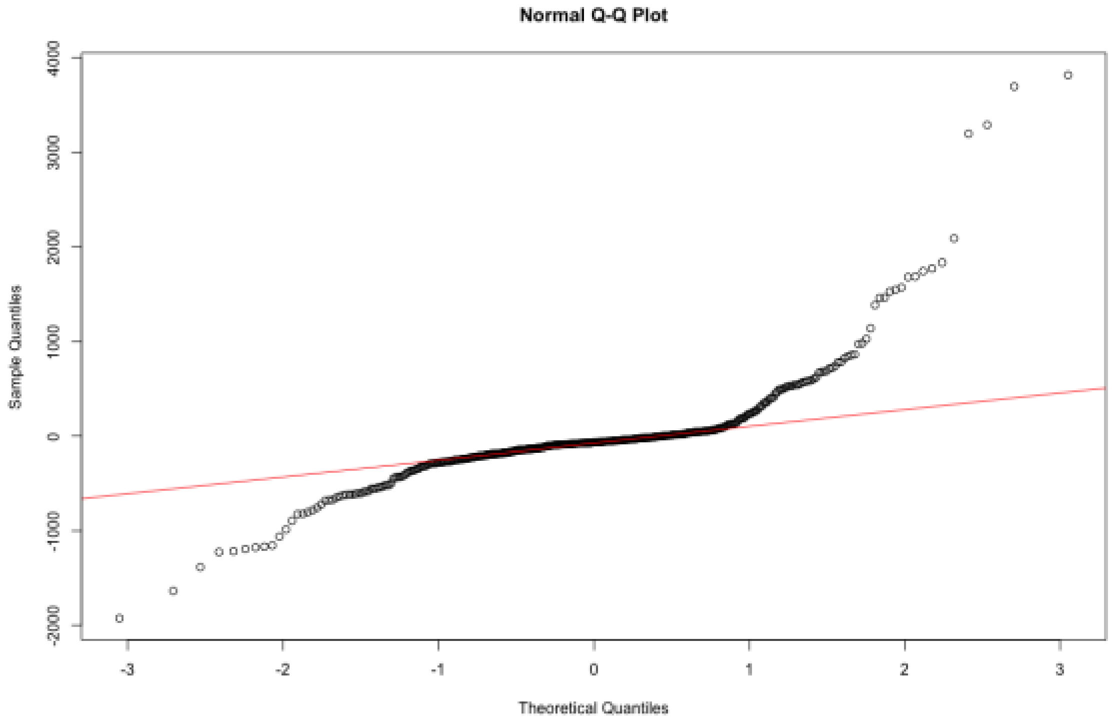
As we can see from the residual plot for number of hospital beds model, there
are two outliers (22000, -1000),(28000, 3000). The residuals are not mainly c
entered around the 0 line, which indicates that it is not a good fit.
In addition, the qqline is not linear(more of a cubic function line), which m

```
eans that the data is not normal. The normality probability plot shows depart
ures from normality in the tails.
```

The residuals for Total Personal Income is

```
##             1             2             3             4             5 
## -5.379152e+02  5.920453e+02  3.574343e+02 -4.908760e+02 -1.636006e+03 
##             6             7             8             9            10 
## -1.819097e+02 -6.740486e+02 -9.846914e+02  1.775411e+03 -3.582301e+02 
##            11            12            13            14            15 
##  3.197733e+03  6.078291e+02 -8.194276e+02 -5.519748e+02  1.746860e+03 
##            16            17            18            19            20 
##  5.444362e+02  1.689306e+03 -1.154394e+03  8.243570e+02 -5.138315e+02 
##            21            22            23            24            25 
## -1.191931e+03  6.893319e+02 -1.215643e+03 -1.175587e+03  2.318075e+02 
##            26            27            28            29            30 
## -8.396239e+01  5.311399e+02 -1.926599e+03  4.614645e+02  3.124729e+02 
##            31            32            33            34            35 
##  4.939178e+02  7.851199e+02  8.645577e+02 -1.166302e+03  1.089101e+02 
##            36            37            38            39            40 
## -7.554338e+02 -2.732569e+02  2.067006e+02 -1.059983e+03  5.317174e+02 
##            41            42            43            44            45 
## -2.291202e+02 -1.383928e+03  8.583969e+02 -8.923233e+02  9.741079e+02 
##            46            47            48            49            50 
## -5.507159e+02  3.460094e+02  1.684295e+03 -8.013986e+02  3.819000e+03 
##            51            52            53            54            55 
## -5.779811e+02  6.701073e+02  2.088975e+03 -1.223572e+03  4.514798e+02 
##            56            57            58            59            60 
##  9.077406e+01 -8.219593e+02  7.435619e+02 -1.799017e+02  2.099451e+01 
##            61            62            63            64            65 
## -5.173075e+02 -7.817650e+02 -6.708832e+02 -2.021772e+02  5.551301e+02 
##            66            67            68            69            70 
##  5.432338e+02  3.698279e+03  9.823326e+02 -6.031831e+02  1.466164e+03 
##            71            72            73            74            75 
##  1.912963e+02  6.230525e+02  1.835775e+03  5.693576e+02  5.073821e+02 
##            76            77            78            79            80 
## -5.612427e+01 -2.868445e+02  1.141839e+03 -8.981842e+01 -2.989651e+01 
##            81            82            83            84            85 
## -6.470099e+02 -2.605377e+02 -2.839258e+02 -4.255593e+02 -4.063181e+02 
##            86            87            88            89            90 
## -6.796231e+02 -7.205966e+02  4.014468e+00 -2.481249e+02  1.029256e+03 
##            91            92            93            94            95 
## -6.146989e+02 -7.268710e+01 -1.638277e+01 -1.758796e+02  1.463440e+03 
##            96            97            98            99           100 
## -2.967160e+02  2.615577e+02 -2.637499e+02  5.246025e+02 -2.672649e+02 
##           101           102           103           104           105 
##  3.840457e+02  1.550971e+03 -5.868449e+02  3.545722e+02 -1.938006e+02 
##           106           107           108           109           110 
## -4.316905e+02  2.757735e+02 -6.229108e+02 -2.105985e+02 -6.130463e+02 
##           111           112           113           114           115 
## -6.080649e+02 -6.352422e+02 -2.136486e+02 -6.183747e+02 -3.424638e+02
##           116           117           118           119           120
## -4.288440e+02 -5.537286e+02  5.748693e+02 -1.846561e+02 -4.637210e+01
##           121           122           123           124           125
## -3.169668e+02 -2.824472e+02  3.291122e+03 -4.297690e+02 -1.887186e+02
##           126           127           128           129           130
## -5.291560e+02 -2.392307e+02 -9.010131e+01 -5.778134e+00  1.808868e+02
##           131           132           133           134           135
## -5.836115e+02 -2.206037e+02 -1.371901e+02 -2.620884e+02 -1.427263e+02
##           136           137           138           139           140
## -1.746012e+02  6.802303e+02  1.215140e+02 -5.377152e+02 -3.483725e+02
##           141           142           143           144           145
## -1.714095e+00 -6.036988e+02  7.190632e+02 -2.067326e+02 -2.447447e+02
##           146           147           148           149           150
## -3.637304e+02 -2.556768e+02 -3.640855e+02  2.637867e+02 -2.757469e+02
##           151           152           153           154           155
## -6.655265e+01 -7.443124e+01 -5.533106e+01 -3.852361e+02 -1.978502e+02
##           156           157           158           159           160
## -4.164695e+02 -2.744020e+02 -1.878305e+02 -1.572821e+02  7.831067e+02
##           161           162           163           164           165
##  4.960910e+02 -2.865704e+02  5.251451e+01 -7.320400e+01 -2.718127e+02
##           166           167           168           169           170
##  4.025858e+02  1.003634e+02  1.387449e+03 -3.962183e+01  1.291616e+02
##           171           172           173           174           175
## -1.598738e+02 -2.219955e+02 -2.332944e+02 -7.626050e+01 -5.163159e+02
##           176           177           178           179           180
## -1.278810e+02 -3.768582e+02 -2.321883e+02  5.194870e+02  4.392130e+01
##           181           182           183           184           185
## -7.510093e+01  8.242123e+01 -4.451935e+02 -5.873785e+01 -1.831254e+02
##           186           187           188           189           190
##  3.766354e+01  5.847919e+02 -1.480916e+01 -1.742821e+02 -1.458418e+02
##           191           192           193           194           195
## -6.387431e+01  5.809957e+02 -4.275228e+01 -5.892617e+01 -2.667582e+02
##           196           197           198           199           200
##  4.122148e+02 -8.934324e+01 -9.459861e+01 -3.134699e+02  9.633887e+01
##           201           202           203           204           205
## -2.185270e+02 -5.179911e+01 -1.501334e+02 -2.220148e+02  2.977826e+02
##           206           207           208           209           210
## -8.824002e+01 -3.139156e+02 -1.890132e+02 -5.776320e+01 -2.110021e+02
##           211           212           213           214           215
## -4.783201e+01 -2.638726e+02  7.229679e+02 -1.353268e+02 -2.827380e+02
##           216           217           218           219           220
##  1.242638e+02  2.506062e+02 -1.966070e+02 -1.200677e+01  4.902917e+02
##           221           222           223           224           225
## -2.016760e+01  3.455506e+01 -3.391731e+02 -2.775450e+02 -6.385988e+01
##           226           227           228           229           230
## -3.054880e+02  3.847593e+00 -8.849051e+01 -1.313867e+02 -1.370960e+01
##           231           232           233           234           235
##  8.297214e+01  1.473533e+02 -9.376195e+01 -1.635209e+02  5.191601e+01
##           236           237           238           239           240
## -7.166992e+01 -7.074965e+01 -1.364660e+02  4.508739e+01 -2.487889e+02
##           241           242           243           244           245
##  2.206839e+02 -1.436049e+02  6.821305e+02 -2.239761e+01 -1.029337e+02
##           246           247           248           249           250
## -1.170970e+02 -2.425792e+02  6.381671e+01 -1.874517e+02 -9.373194e+01
##           251           252           253           254           255
## -1.262659e+02 -3.190684e+02  1.878008e+02 -2.365910e+02 -2.069135e+02
##           256           257           258           259           260
##  1.440189e+01 -1.483335e+02  1.531704e+03  8.443542e+02 -1.867598e+02
##           261           262           263           264           265
##  4.226554e+01 -9.467281e+01  1.878358e+02 -1.858754e+02 -1.423415e+02
##           266           267           268           269           270
##  1.275593e+02 -6.609704e+01 -2.504564e+02  4.398356e+01 -8.158380e+01
##           271           272           273           274           275
## -2.493592e+02 -1.736860e+01  5.918155e+01 -3.509704e+01 -2.067134e+02
##           276           277           278           279           280
##  6.322336e+01 -1.034593e+02  6.613510e+01  4.339724e+01 -1.677286e+02
##           281           282           283           284           285
##  3.858004e+01  2.407991e+02 -9.753057e+01 -3.618317e+01 -1.139219e+02
##           286           287           288           289           290
##  8.479951e+01 -7.883954e+01 -4.717132e+00 -1.325306e+02 -1.001249e+02
##           291           292           293           294           295
## -2.152659e+02 -1.231903e+02 -1.350620e+02 -1.821862e+02 -1.699257e+02
##           296           297           298           299           300
## -7.834061e+01  4.838670e+01  4.777464e+01 -1.743136e+02 -1.906927e+02
##           301           302           303           304           305
## -5.520678e+01  1.397573e+02 -4.871287e+01 -1.352883e+02  4.483328e+01
##           306           307           308           309           310
## -9.161330e+01 -1.027218e+02  2.339032e-01  1.677464e+01  5.173284e+01
##           311           312           313           314           315
##  1.098444e+01 -2.359476e+02 -1.437847e+02 -9.653898e+01 -2.430470e+01
##           316           317           318           319           320
##  4.434406e+00 -5.541657e+01 -7.283025e+01  1.276185e+02 -6.676446e+01
##           321           322           323           324           325
## -1.516741e+02  5.668010e+01 -5.346511e+00 -7.155291e+01  2.087044e+01
##           326           327           328           329           330
## -4.153182e+01 -3.569730e+01 -1.380704e+02 -8.825913e+01  5.502662e+01
##           331           332           333           334           335
## -7.641531e+01 -1.446192e+02 -6.513333e+01  1.631162e+01 -4.845160e+01
##           336           337           338           339           340
## -1.725715e+02 -2.119773e+00 -5.525574e+01  3.984158e+02  4.113346e+01
##           341           342           343           344           345
## -1.200844e+02 -2.395222e+02 -8.725574e+01  7.773387e+00  7.343255e+00
##           346           347           348           349           350
## -1.031414e+02 -7.990834e+01 -1.595293e+02 -1.279679e+02 -1.042452e+02
##           351           352           353           354           355
## -1.793408e+01 -2.133509e+01 -1.251544e+02 -7.653898e+01 -1.239805e+02
##           356           357           358           359           360
##  1.222723e+02 -2.594463e+01 -1.469037e+02 -1.670422e+02  7.731287e+01
##           361           362           363           364           365
##  1.548086e+01 -1.310983e+02  1.557502e+01 -1.104575e+02 -7.518819e+01
##           366           367           368           369           370
## -4.340853e+01 -1.087382e+02 -3.471964e+01  7.422210e+01 -4.017049e+01
##           371           372           373           374           375
##  1.806002e+01 -7.633075e-02  2.265163e+00 -2.987244e+01  1.652266e+01
##           376           377           378           379           380
##  3.371852e+01 -6.067194e+01 -4.390960e+01 -5.875806e+01 -6.524093e+01
##           381           382           383           384           385
## -2.906365e+01  2.783629e+01 -2.057463e+02 -2.746679e+01  1.386290e+01
##           386           387           388           389           390
##  3.405411e+01 -3.680828e+01  5.239222e+01  3.632216e+01 -8.503201e+01
##           391           392           393           394           395
## -6.655417e+01  5.402116e+02 -7.015943e+00 -9.333044e+01 -1.886441e+01
##           396           397           398           399           400
##  2.362799e+02 -4.709742e+01 -7.774511e+00 -4.978506e+01  6.269316e+01
##           401           402           403           404           405
## -6.623842e+01 -7.144570e+01 -2.277451e+01 -7.861494e+01  1.559950e+01
##           406           407           408           409           410
## -6.310508e+01 -5.124557e+01 -6.628148e+01  3.190541e+02 -1.970827e+02
##           411           412           413           414           415
## -8.697966e+01 -3.173270e+01 -3.742837e+01 -7.452630e+01  4.665789e+00
##           416           417           418           419           420
## -8.417300e+01 -8.911262e+01  1.574759e+03  1.733145e+01 -1.136547e+01
##           421           422           423           424           425
##  2.250283e+01 -1.887621e+01 -1.753095e+01 -2.259171e+01 -4.848198e+01
##           426           427           428           429           430
## -8.467658e+01 -8.831186e+01 -8.028738e+01  1.709271e+02 -2.309529e+01
##           431           432           433           434           435
## -1.305706e+02 -7.264745e+01 -5.543553e+01 -8.754614e+01 -1.419493e+02
##           436           437           438           439           440
## -3.890872e+01 -1.190713e+02 -3.884582e+01 -4.174259e+00 -4.651701e+01
```





```
As we can see from the residual plot for total personal income model, there a
re two outliers (110000, 800),(180000, -500). The residuals are not mainly ce
ntered around the 0 line, which indicates that it is not a good fit.
In addition, the qqline is not linear (more of a cubic function line), which
```

```
means that the data is not normal. The normality probability plot shows depar
tures from normality in the tails.

Since these three residual plots and normal Q-Q plots are relatively similar,
we can not conclude which linear regression model is more appropriate in one
case than in the others.
```

Part V: Discussion.

In this report, we find that the number of active physicians has positive linear relationship with total population, number of hospital bed and total personal income. But we can not conclude that the increase in the number of active physicians is caused by the increase in total population, number of hospital bed and total personal income since there are many other factors that can affect the results. In other words, correlation does not prove causation. We also conclude that per capita income and the percentage of individuals in a county having at least a bachelor's degree has a linear relationship because 0 is not within the confidence intervals for all B1 in four different regions. As we can see from the dataset, there are seventeen different variables. However, we only make analysis on a few variables in this report. Therefore, our results might be biased. In addition, we should use transformations in y variables because they have unequal variance and nonnormality.

Appendix:

```
CDI <- read.table("~/Desktop/CDI.txt", quote="\"", comment.char="")
Identification_number = CDI$V1
County = CDI$V2
State = CDI$V3
Land_area = CDI$V4
Total_population = CDI$V5
Percent_of_population_aged_18_34 = CDI$V6
Percent_of_population_65_older = CDI$V7
Number_of_active_physicians = CDI$V8
Number_of_hospital_beds = CDI$V9
Total_serious_crimes = CDI$V10
Percent_high_school_graduates = CDI$V11
Percent_bachelors_degrees = CDI$V12
Percent_below_poverty_level = CDI$V13
Percent_unemployment = CDI$V14
Per_capita_income = CDI$V15
Total_personal_income = CDI$V16
Geographic_region = CDI$V17



names(CDI) = c('Identification_number', 'County', 'State', 'Land_area', 'Tota
l_population',
                 'Percent_of_population_aged_18_34', 'Percent_of_population_6
5_older', 'Number_of_active_physicians', 'Number_of_hospital_beds', 'Total_se
rious_crimes', 'Percent_high_school_graduates', 'Percent_bachelors_degrees',
'Percent_below_poverty_level', 'Percent_unemployment', 'Per_capita_income', '
Total_personal_income', 'Geographic_region' )
```

```
1.43
a.
# Regress Y (Number_of_active_physicians) on Total_population.
fit1 = lm(Number_of_active_physicians~Total_population, data = CDI)
fit1$coefficients
summary(fit1)
# Regress Y (Number_of_active_physicians) on Number_of_hospital_beds.
fit2 = lm(Number_of_active_physicians~Number_of_hospital_beds, data = CDI)
fit2$coefficients
summary(fit2)
# Regress Y (Number_of_active_physicians) on Total_personal_income.
fit3 = lm(Number_of_active_physicians~Total_personal_income, data = CDI)
fit3$coefficients
summary(fit3)

b.
par(mfrow=c(1,3))
plot(CDI$Total_population, CDI$Number_of_active_physicians, xlab = 'Total Pop
ulation', ylab = 'Number of active physicians', main = 'Number of active phys
icians Against Total Population')
abline(fit1, col = 'red')

plot(CDI$Number_of_hospital_beds, CDI$Number_of_active_physicians, xlab = 'Nu
mber of hospital beds', ylab = 'Number of active physicians', main = 'Number
of active physicians Against Number of hospital beds')
abline(fit2, col = 'red')

plot(CDI$Total_personal_income, CDI$Number_of_active_physicians, xlab = 'Tota
l personal income', ylab = 'Number of active physicians', main = 'Number of a
ctive physicians Against Total personal income')
abline(fit3, col = 'red')

c.
# MSE 1
anova.1 = anova(fit1)
SSE1 = anova.1[2,2]
MSE1 = SSE1/(dim(CDI)[1] - 2)

y = function(x){
  fit1$coefficients[1] + fit1$coefficients[2]*x
}

sum((CDI$Number_of_active_physicians - y(CDI$Total_population))^2)/(dim(CDI)[
1] - 2)

# MSE 2
anova.2 = anova(fit2)
SSE2 = anova.2[2,2]
MSE2 = SSE2/(dim(CDI)[1] - 2)
```

```
#MSE 3
anova.3 = anova(fit3)
SSE3 = anova.3[2,2]
MSE3 = SSE3/(dim(CDI)[1] - 2)

c(MSE1,MSE2,MSE3)

1.44
a.
CDI[,17] = factor(CDI[,17], levels = c(1, 2, 3, 4),labels = c('NE', 'NC', 'S'
, 'W'))
data.Geographic_region = split(CDI, CDI$Geographic_region)
level = levels(CDI$Geographic_region)
CDI.NE = data.Geographic_region$NE
CDI.NC = data.Geographic_region$NC
CDI.S = data.Geographic_region$S
CDI.W = data.Geographic_region$W
b.
NE.lm = lm(Per_capita_income ~ Percent_bachelors_degrees, data = CDI.NE)
NE.lm$coefficients

NC.lm = lm(Per_capita_income ~ Percent_bachelors_degrees, data = CDI.NC)
NC.lm$coefficients

S.lm = lm(Per_capita_income ~ Percent_bachelors_degrees, data = CDI.S)
S.lm$coefficients

W.lm = lm(Per_capita_income ~ Percent_bachelors_degrees, data = CDI.W)
W.lm$coefficients

c.
# MSE for region NC
anova(NC.lm)[2,2]/(dim(CDI.NC)[1] - 2)

# MSE for region NE
anova(NE.lm)[2,2]/(dim(CDI.NE)[1] - 2)

# MSE for region S
anova(S.lm)[2,2]/(dim(CDI.S)[1] - 2)

# MSE for region W
anova(W.lm)[2,2]/(dim(CDI.W)[1] - 2)

2.62
# R^2 for Total Population
summary(fit1)$r.squared
```

```
# R^2 for Number of hospital beds
summary(fit2)$r.squared

# R^2 for Total personal income
summary(fit3)$r.squared

2.63
## 90% confidence intervals for each region NC, NE, S, and W, respectively:

confint(NC.lm, 'Percent_bachelors_degrees', level = 0.90)

confint(NE.lm, 'Percent_bachelors_degrees', level = 0.90)

confint(S.lm, 'Percent_bachelors_degrees', level = 0.90)

confint(W.lm, 'Percent_bachelors_degrees', level = 0.90)
anova(NE.lm)
anova(NC.lm)
anova(S.lm)
anova(W.lm)

3.25
par(mar=c(1,1,1,1))
fit.res1 = residuals(fit1)
plot(CDI$Total_population, fit.res1, xlab = 'Total Population', ylab = 'Resid
uals',
     main = 'Residual Plot for Total Population Model')

qqnorm(fit.res1)
qqline(fit.res1, col = 'red')

fit.res2  = residuals(fit2)
plot(CDI$Number_of_hospital_beds, fit.res2, xlab = 'Number of hospital beds',
ylab = 'Residuals',
     main = 'Residual Plot for Number of hospital beds Model')

qqnorm(fit.res2)
qqline(fit.res2, col = 'red')

fit.res3  = residuals(fit3)
plot(CDI$Total_personal_income, fit.res3, xlab = 'Total Personal Income', yla
b = 'Residuals',
     main = 'Residual Plot for Total Personal Income Model')

qqnorm(fit.res3)
qqline(fit.res3, col = 'red')
```