

Case Solution Group 5

Balint Keller, Ling-Fei Chen , Elisabeth Bonte, Lieke Vanhaverbeke, Stef Desender, Seppe Van Campe

2025-05-15

Step 1: Specification, hypotheses, and descriptive statistic

```
##          WAGE EDUC AGE RACE SMSA MARRIED REGION QOB REGION1 REGION2 REGION3
## 1 580.1000   9  45    0    0     1     9   3     0     0     0     0
## 2 642.2115  17  47    0    0     1     3   4     0     0     0     1
## 3 577.0192  12  42    0    0     1     7   2     0     0     0     0
## 4 999.1346  10  43    0    0     1     3   2     0     0     0     1
## 5 307.7885  12  41    0    0     1     6   3     0     0     0     0
## 6 280.1000  12  40    1    0     1     5   2     0     0     0     0
##          REGION4 REGION5 REGION6 REGION7 REGION8 REGION9 QOB1 QOB2 QOB3 QOB4
## 1          0      0      0      0      0     1     0     0     1     0
## 2          0      0      0      0      0     0     0     0     0     1
## 3          0      0      0      1      0     0     0     1     0     0
## 4          0      0      0      0      0     0     0     1     0     0
## 5          0      0      1      0      0     0     0     0     1     0
## 6          0      1      0      0      0     0     0     0     1     0
##          WAGE          EDUC          AGE          RACE
## Min. : 0.096  Min. : 0.00  Min. :40.00  Min. :0.0000
## 1st Qu.: 278.558 1st Qu.:12.00 1st Qu.:42.00 1st Qu.:0.0000
## Median : 384.712 Median :12.00 Median :45.00 Median :0.0000
## Mean   : 436.524 Mean  :12.71 Mean  :44.68 Mean  :0.0832
## 3rd Qu.: 520.100 3rd Qu.:15.00 3rd Qu.:47.00 3rd Qu.:0.0000
## Max.   :10167.500 Max.  :20.00 Max.  :50.00 Max.  :1.0000
##          SMSA          MARRIED          REGION          QOB
## Min. :0.0000  Min. :0.0000  Min. :1.000  Min. :1.000
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:3.000 1st Qu.:1.000
## Median :0.0000  Median :1.0000  Median :5.000  Median :3.000
## Mean   :0.1813  Mean  :0.8609  Mean  :4.767  Mean  :2.502
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:7.000 3rd Qu.:3.000
## Max.   :1.0000  Max.  :1.0000  Max.  :9.000  Max.  :4.000
##          REGION1          REGION2          REGION3          REGION4
## Min. :0.0000  Min. :0.0000  Min. :0.0000  Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000
## Mean   :0.0549  Mean  :0.1584  Mean  :0.1949  Mean  :0.0732
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max.   :1.0000  Max.  :1.0000  Max.  :1.0000  Max.  :1.0000
##          REGION5          REGION6          REGION7          REGION8
## Min. :0.0000  Min. :0.000  Min. :0.0000  Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000
```

```

## Median :0.0000  Median :0.000  Median :0.0000  Median :0.0000
## Mean   :0.1773  Mean   :0.064  Mean   :0.0995  Mean   :0.0494
## 3rd Qu.:0.0000 3rd Qu.:0.000 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max.   :1.0000  Max.   :1.000  Max.   :1.0000  Max.   :1.0000
##      REGION9          QOB1          QOB2          QOB3
## Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000
## Mean   :0.1284  Mean   :0.2533  Mean   :0.2354  Mean   :0.2674
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000
##      QOB4
## Min.   :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean   :0.2439
## 3rd Qu.:0.0000
## Max.   :1.0000

##      vars     n   mean     sd median trimmed    mad   min    max range
## WAGE      1 10000 436.52 295.37 384.71  401.88 173.27  0.1 10167.5 10167.4
## EDUC      2 10000 12.71   3.28  12.00   12.72   2.97  0.0   20.0   20.0
## AGE       3 10000 44.68   2.93  45.00   44.67   4.45  40.0   50.0   10.0
## RACE      4 10000  0.08   0.28   0.00    0.00   0.00  0.0    1.0    1.0
## SMSA      5 10000  0.18   0.39   0.00    0.10   0.00  0.0    1.0    1.0
## MARRIED   6 10000  0.86   0.35   1.00    0.95   0.00  0.0    1.0    1.0
## REGION    7 10000  4.77   2.46   5.00    4.65   2.97  1.0    9.0    8.0
## QOB       8 10000  2.50   1.12   3.00    2.50   1.48  1.0    4.0    3.0
## REGION1   9 10000  0.05   0.23   0.00    0.00   0.00  0.0    1.0    1.0
## REGION2  10 10000  0.16   0.37   0.00    0.07   0.00  0.0    1.0    1.0
## REGION3  11 10000  0.19   0.40   0.00    0.12   0.00  0.0    1.0    1.0
## REGION4  12 10000  0.07   0.26   0.00    0.00   0.00  0.0    1.0    1.0
## REGION5  13 10000  0.18   0.38   0.00    0.10   0.00  0.0    1.0    1.0
## REGION6  14 10000  0.06   0.24   0.00    0.00   0.00  0.0    1.0    1.0
## REGION7  15 10000  0.10   0.30   0.00    0.00   0.00  0.0    1.0    1.0
## REGION8  16 10000  0.05   0.22   0.00    0.00   0.00  0.0    1.0    1.0
## REGION9  17 10000  0.13   0.33   0.00    0.04   0.00  0.0    1.0    1.0
## QOB1     18 10000  0.25   0.43   0.00    0.19   0.00  0.0    1.0    1.0
## QOB2     19 10000  0.24   0.42   0.00    0.17   0.00  0.0    1.0    1.0
## QOB3     20 10000  0.27   0.44   0.00    0.21   0.00  0.0    1.0    1.0
## QOB4     21 10000  0.24   0.43   0.00    0.18   0.00  0.0    1.0    1.0

##      skew kurtosis   se
## WAGE    7.39  170.46 2.95
## EDUC   -0.07   0.55 0.03
## AGE     0.05  -1.18 0.03
## RACE    3.02   7.11 0.00
## SMSA    1.65   0.74 0.00
## MARRIED -2.09   2.35 0.00
## REGION   0.35  -1.06 0.02
## QOB    -0.03  -1.35 0.01
## REGION1  3.91  13.27 0.00
## REGION2  1.87   1.50 0.00
## REGION3  1.54   0.37 0.00
## REGION4  3.28   8.74 0.00

```

```

## REGION5 1.69      0.85 0.00
## REGION6 3.56      10.69 0.00
## REGION7 2.68      5.16 0.00
## REGION8 4.16      15.29 0.00
## REGION9 2.22      2.93 0.00
## QOB1     1.13      -0.71 0.00
## QOB2     1.25      -0.44 0.00
## QOB3     1.05      -0.90 0.00
## QOB4     1.19      -0.58 0.00

##          WAGE    EDUC   AGE   RACE   SMSA MARRIED REGION   QOB REGION1
## WAGE    87241.084 315.768 5.257 -10.490 -14.858 10.178 8.855 2.397 -0.854
## EDUC     315.768 10.742 -0.668 -0.138 -0.188 0.022 0.274 0.117 0.016
## AGE      5.257 -0.668 8.592 -0.004 -0.024 0.021 -0.125 -0.387 0.016
## RACE    -10.490 -0.138 -0.004 0.076 -0.004 -0.011 0.001 -0.004 -0.003
## SMSA    -14.858 -0.188 -0.024 -0.004 0.148 0.005 0.033 -0.005 -0.002
## MARRIED 10.178 0.022 0.021 -0.011 0.005 0.120 -0.022 -0.003 -0.001
## REGION   8.855 0.274 -0.125 0.001 0.033 -0.022 6.064 -0.002 -0.207
## QOB     2.397 0.117 -0.387 -0.004 -0.005 -0.003 -0.002 1.245 0.001
## REGION1 -0.854 0.016 0.016 -0.003 -0.002 -0.001 -0.207 0.001 0.052
## REGION2  3.820 0.043 0.009 0.001 -0.015 0.000 -0.438 0.005 -0.009
## REGION3  3.379 -0.040 0.006 -0.003 -0.004 0.002 -0.344 -0.003 -0.011
## REGION4 -1.138 0.000 -0.011 -0.004 0.012 0.002 -0.056 -0.006 -0.004
## REGION5 -6.815 -0.054 -0.014 0.010 0.003 0.001 0.041 0.002 -0.010
## REGION6 -3.960 -0.068 0.006 0.003 0.011 0.002 0.079 -0.003 -0.004
## REGION7 -1.995 -0.032 0.003 0.001 -0.001 0.001 0.222 0.004 -0.005
## REGION8 -0.217 0.020 -0.017 -0.003 0.006 0.002 0.160 -0.001 -0.003
## REGION9  7.781 0.115 0.001 -0.003 -0.010 -0.007 0.544 0.000 -0.007
## QOB1    -0.819 -0.041 0.193 0.000 0.003 -0.001 0.005 -0.380 -0.001
## QOB2    -0.478 0.013 -0.064 0.003 -0.001 0.002 -0.014 -0.118 0.001
## QOB3     1.016 -0.021 -0.065 -0.001 -0.002 0.001 0.016 0.133 0.000
## QOB4     0.281 0.049 -0.064 -0.002 0.000 -0.002 -0.007 0.365 0.000
##          REGION2 REGION3 REGION4 REGION5 REGION6 REGION7 REGION8 REGION9 QOB1
## WAGE     3.820  3.379 -1.138 -6.815 -3.960 -1.995 -0.217 7.781 -0.819
## EDUC     0.043 -0.040  0.000 -0.054 -0.068 -0.032  0.020 0.115 -0.041
## AGE      0.009  0.006 -0.011 -0.014  0.006  0.003 -0.017 0.001 0.193
## RACE     0.001 -0.003 -0.004  0.010  0.003  0.001 -0.003 -0.003 0.000
## SMSA    -0.015 -0.004  0.012  0.003  0.011 -0.001  0.006 -0.010 0.003
## MARRIED 0.000  0.002  0.002  0.001  0.002  0.001  0.002 -0.007 -0.001
## REGION   -0.438 -0.344 -0.056  0.041  0.079  0.222  0.160 0.544 0.005
## QOB      0.005 -0.003 -0.006  0.002 -0.003  0.004 -0.001 0.000 -0.380
## REGION1 -0.009 -0.011 -0.004 -0.010 -0.004 -0.005 -0.003 -0.007 -0.001
## REGION2  0.133 -0.031 -0.012 -0.028 -0.010 -0.016 -0.008 -0.020 -0.003
## REGION3 -0.031  0.157 -0.014 -0.035 -0.012 -0.019 -0.010 -0.025 0.001
## REGION4 -0.012 -0.014  0.068 -0.013 -0.005 -0.007 -0.004 -0.009 0.002
## REGION5 -0.028 -0.035 -0.013  0.146 -0.011 -0.018 -0.009 -0.023 0.001
## REGION6 -0.010 -0.012 -0.005 -0.011  0.060 -0.006 -0.003 -0.008 0.001
## REGION7 -0.016 -0.019 -0.007 -0.018 -0.006  0.090 -0.005 -0.013 -0.001
## REGION8 -0.008 -0.010 -0.004 -0.009 -0.003 -0.005  0.047 -0.006 0.001
## REGION9 -0.020 -0.025 -0.009 -0.023 -0.008 -0.013 -0.006 0.112 -0.001
## QOB1    -0.003  0.001  0.002  0.001  0.001 -0.001  0.001 -0.001 0.189
## QOB2     0.002  0.000  0.001 -0.002  0.000 -0.003  0.000  0.001 -0.060
## QOB3    -0.001 -0.002  0.000 -0.001 -0.001  0.004 -0.001  0.001 0.001
## QOB4     0.001  0.000 -0.003  0.002  0.000  0.000  0.000 -0.001 -0.062

```

```

##          QOB2    QOB3    QOB4
## WAGE      -0.478   1.016   0.281
## EDUC       0.013  -0.021   0.049
## AGE       -0.064  -0.065  -0.064
## RACE       0.003  -0.001  -0.002
## SMSA      -0.001  -0.002   0.000
## MARRIED    0.002   0.001  -0.002
## REGION     -0.014   0.016  -0.007
## QOB       -0.118   0.133   0.365
## REGION1    0.001   0.000   0.000
## REGION2    0.002  -0.001   0.001
## REGION3    0.000  -0.002   0.000
## REGION4    0.001   0.000  -0.003
## REGION5   -0.002  -0.001   0.002
## REGION6    0.000  -0.001   0.000
## REGION7   -0.003   0.004   0.000
## REGION8    0.000  -0.001   0.000
## REGION9    0.001   0.001  -0.001
## QOB1      -0.060  -0.068  -0.062
## QOB2       0.180  -0.063  -0.057
## QOB3      -0.063   0.196  -0.065
## QOB4      -0.057  -0.065   0.184

## # A tibble: 9 x 11
##   REGION mean_wage mean_educ mean_age sd_wage sd_educ sd_age var_wage var_educ
##   <int>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1       1      421.     13.0      45.0     244.     3.10     2.93     59622.     9.62
## 2       2      461.     13.0      44.7     355.     3.17     2.96    126021.    10.0
## 3       3      454.     12.5      44.7     264.     3.07     2.89     69754.     9.40
## 4       4      421.     12.7      44.5     263.     3.15     2.94     69295.     9.90
## 5       5      398.     12.4      44.6     261.     3.48     2.99     68143.    12.1
## 6       6      375.     11.7      44.8     330.     3.74     2.77    108586.    14.0
## 7       7      416.     12.4      44.7     266.     3.57     2.98     70638.    12.7
## 8       8      432.     13.1      44.3     236.     2.95     2.95     55764.     8.73
## 9       9      497.     13.6      44.7     346.     2.92     2.89    119854.     8.54
## # i 2 more variables: var_age <dbl>, count <int>

## 
## Descriptive statistics by group
## REGION: 1
##          vars   n   mean      sd median trimmed     mad     min     max   range skew
## REGION     1 549  1.00    0.00    1.00    1.00    0.00  1.00  1.00    0.00  NaN
## WAGE       2 549 420.98 244.18 384.71  389.65 171.07  3.94 1562.5 1558.56  1.74
## EDUC       3 549 13.01   3.10   12.00   12.91   2.97  5.00  20.0   15.00  0.33
## AGE        4 549 44.97   2.93   45.00   45.02   4.45 40.00  50.0   10.00 -0.05
## RACE       5 549  0.03   0.17   0.00    0.00    0.00  0.00  1.00   1.00  5.40
## SMSA       6 549  0.15   0.36   0.00    0.06    0.00  0.00  1.00   1.00  1.96
## MARRIED    7 549  0.85   0.36   1.00    0.93    0.00  0.00  1.00   1.00 -1.94
##          kurtosis     se
## REGION      NaN  0.00
## WAGE        4.51 10.42
## EDUC       -0.16  0.13
## AGE        -1.22  0.13
## RACE        27.22  0.01

```

```

## SMSA      1.85  0.02
## MARRIED   1.78  0.02
##
## REGION: 2
##          vars   n   mean     sd median trimmed    mad   min   max range
## REGION    1 1584  2.00  0.00  2.00   2.00  0.00  2.0  2.0  0.0
## WAGE      2 1584 460.64 354.99 403.94  422.27 171.07  2.4 10167.5 10165.1
## EDUC      3 1584 12.98  3.17 12.00  12.90  2.97  0.0 20.0 20.0
## AGE       4 1584 44.74  2.96 45.00  44.73  4.45 40.0 50.0 10.0
## RACE      5 1584  0.09  0.29  0.00  0.00  0.00  0.0  1.0  1.0
## SMSA      6 1584  0.09  0.28  0.00  0.00  0.00  0.0  1.0  1.0
## MARRIED   7 1584  0.86  0.35  1.00  0.95  0.00  0.0  1.0  1.0
##          skew kurtosis   se
## REGION    NaN        NaN 0.00
## WAGE      13.60    352.35 8.92
## EDUC      0.18     0.42 0.08
## AGE       0.04    -1.22 0.07
## RACE      2.82     5.94 0.01
## SMSA      2.88     6.32 0.01
## MARRIED   -2.05    2.20 0.01
##
## REGION: 3
##          vars   n   mean     sd median trimmed    mad   min   max range
## REGION    1 1949  3.00  0.00  3.00   3.00  0.00  3.00 3.00 0.00
## WAGE      2 1949 453.86 264.11 409.2  425.41 164.62  0.15 4634.17 4634.02
## EDUC      3 1949 12.51  3.07 12.00  12.43  1.48  0.00 20.00 20.00
## AGE       4 1949 44.71  2.89 45.00  44.70  2.97 40.00 50.00 10.00
## RACE      5 1949  0.07  0.25  0.00  0.00  0.00  0.00  1.00  1.00
## SMSA      6 1949  0.16  0.37  0.00  0.08  0.00  0.00  1.00  1.00
## MARRIED   7 1949  0.87  0.34  1.00  0.96  0.00  0.00  1.00  1.00
##          skew kurtosis   se
## REGION    NaN        NaN 0.00
## WAGE      4.11     43.18 5.98
## EDUC      0.22     0.56 0.07
## AGE       0.05    -1.12 0.07
## RACE      3.37     9.39 0.01
## SMSA      1.83     1.36 0.01
## MARRIED   -2.21    2.88 0.01
##
## REGION: 4
##          vars   n   mean     sd median trimmed    mad   min   max range skew
## REGION    1 732   4.00  0.00  4.00   4.00  0.00  4.00 4.00  0  NaN
## WAGE      2 732 420.98 263.24 384.71  390.86 146.34 10.25 4076.25 4066 4.76
## EDUC      3 732 12.71  3.15 12.00  12.63  1.48  0.00 20.00 20  0.16
## AGE       4 732 44.52  2.94 44.00  44.48  4.45 40.00 50.00 10  0.11
## RACE      5 732  0.03  0.18  0.00  0.00  0.00  0.00  1.00  1  5.12
## SMSA      6 732  0.34  0.48  0.00  0.30  0.00  0.00  1.00  1  0.66
## MARRIED   7 732  0.89  0.32  1.00  0.98  0.00  0.00  1.00  1 -2.43
##          kurtosis   se
## REGION    NaN 0.00
## WAGE      52.28 9.73
## EDUC      0.58 0.12
## AGE       -1.19 0.11
## RACE      24.24 0.01

```

```

## SMSA      -1.57 0.02
## MARRIED    3.93 0.01
##
## REGION: 5
##          vars   n   mean     sd median trimmed    mad   min   max range
## REGION    1 1773  5.00  0.00  5.00    5.00  0.00  5.0  5.00  0.00
## WAGE      2 1773 398.09 261.04 346.25  362.95 189.49  0.1 2501.25 2501.15
## EDUC      3 1773 12.41  3.48 12.00   12.46  2.97  0.0  20.00 20.00
## AGE       4 1773 44.60  2.99 45.00   44.59  4.45 40.0  50.00 10.00
## RACE      5 1773  0.14  0.34  0.00    0.05  0.00  0.0  1.00  1.00
## SMSA      6 1773  0.20  0.40  0.00    0.12  0.00  0.0  1.00  1.00
## MARRIED   7 1773  0.86  0.34  1.00    0.95  0.00  0.0  1.00  1.00
##          skew kurtosis   se
## REGION    NaN      NaN 0.00
## WAGE      2.23     8.76 6.20
## EDUC     -0.19     0.27 0.08
## AGE       0.06    -1.22 0.07
## RACE      2.10     2.42 0.01
## SMSA      1.52     0.31 0.01
## MARRIED   -2.12    2.51 0.01
##
## REGION: 6
##          vars   n   mean     sd median trimmed    mad   min   max range skew
## REGION    1 640   6.00  0.00  6.00    6.00  0.00  6.00  6.00  0.00  NaN
## WAGE      2 640 374.65 329.52 334.13  335.99 181.61  1.98 6005 6003.02 8.62
## EDUC      3 640  11.66  3.74 12.00   11.65  2.97  0.00  20  20.00 -0.03
## AGE       4 640  44.77  2.77 45.00   44.74  2.97 40.00  50  10.00  0.06
## RACE      5 640   0.12  0.33  0.00    0.03  0.00  0.00  1  1.00  2.26
## SMSA      6 640   0.35  0.48  0.00    0.31  0.00  0.00  1  1.00  0.63
## MARRIED   7 640   0.88  0.32  1.00    0.98  0.00  0.00  1  1.00 -2.40
##          kurtosis   se
## REGION    NaN  0.00
## WAGE      133.54 13.03
## EDUC      0.08  0.15
## AGE      -1.08  0.11
## RACE      3.12  0.01
## SMSA     -1.60  0.02
## MARRIED   3.76  0.01
##
## REGION: 7
##          vars   n   mean     sd median trimmed    mad   min   max range
## REGION    1 995   7.00  0.00  7.00    7.00  0.00  7.00  7.00  0.00
## WAGE      2 995 416.47 265.78 373.17  381.59 191.03  4.69 2501.25 2496.56
## EDUC      3 995  12.39  3.57 12.00   12.56  2.97  0.00  20.00 20.00
## AGE       4 995  44.71  2.98 45.00   44.71  4.45 40.00  50.00 10.00
## RACE      5 995   0.09  0.29  0.00    0.00  0.00  0.00  1.00  1.00
## SMSA      6 995   0.17  0.37  0.00    0.09  0.00  0.00  1.00  1.00
## MARRIED   7 995   0.87  0.34  1.00    0.96  0.00  0.00  1.00  1.00
##          skew kurtosis   se
## REGION    NaN  0.00
## WAGE      2.01  7.02 8.43
## EDUC     -0.47  0.56 0.11
## AGE       0.02 -1.19 0.09
## RACE      2.85  6.14 0.01

```

```

## SMSA      1.77    1.12 0.01
## MARRIED -2.18    2.74 0.01
##
## REGION: 8
##          vars   n   mean     sd median trimmed    mad   min   max range
## REGION     1 494    8.00    0.00    8.00    8.00    0.00    8.00    8.00    0.00
## WAGE       2 494  432.13  236.14  384.71  401.01 145.41   1.06 1469.49 1468.43
## EDUC       3 494   13.12    2.95   12.00   13.09   2.97   1.00   20.00   19.00
## AGE        4 494   44.34    2.95   44.00   44.27   2.97  40.00   50.00   10.00
## RACE       5 494    0.03    0.17    0.00    0.00    0.00    0.00    1.00    1.00
## SMSA       6 494    0.30    0.46    0.00    0.25    0.00    0.00    1.00    1.00
## MARRIED    7 494    0.89    0.31    1.00    0.99    0.00    0.00    1.00    1.00
##          skew kurtosis   se
## REGION     NaN      NaN  0.00
## WAGE       1.84     4.94 10.62
## EDUC       0.06     0.95  0.13
## AGE        0.17    -1.16  0.13
## RACE       5.46    27.84  0.01
## SMSA       0.86    -1.26  0.02
## MARRIED   -2.53     4.41  0.01
##
## REGION: 9
##          vars   n   mean     sd median trimmed    mad   min   max range skew
## REGION     1 1284    9.00    0.00    9.00    9.00    0.00    9.0    9  0.0  NaN
## WAGE       2 1284  497.11 346.20  440.48  454.64 196.73   7.1 7335 7327.9  7.21
## EDUC       3 1284   13.61    2.92   13.00   13.56   1.48   0.0   20  20.0  0.01
## AGE        4 1284   44.69    2.89   45.00   44.68   2.97  40.0   50  10.0  0.03
## RACE       5 1284    0.06    0.24    0.00    0.00    0.00    0.0    1   1.0  3.65
## SMSA       6 1284    0.10    0.30    0.00    0.00    0.00    0.0    1   1.0  2.60
## MARRIED    7 1284    0.81    0.40    1.00    0.88    0.00    0.0    1   1.0 -1.55
##          kurtosis   se
## REGION     NaN  0.00
## WAGE       121.20  9.66
## EDUC       0.85  0.08
## AGE        -1.18  0.08
## RACE       11.30  0.01
## SMSA       4.76  0.01
## MARRIED   0.39  0.01
##
##          Descriptive statistics by group
## AGE_GROUP: 1
##          vars   n   mean     sd median trimmed    mad   min   max range
## WAGE       1 1835  439.82 340.56  384.71  401.20 171.07   1.06 7335 7333.94
## EDUC       2 1835   12.98    3.11   12.00   12.93   2.97   0.00   20  20.00
## AGE        3 1835   40.56    0.50   41.00   40.58   0.00  40.00   41  1.00
## RACE       4 1835    0.09    0.28    0.00    0.00    0.00    0.00    1   1.00
## SMSA       5 1835    0.18    0.39    0.00    0.10    0.00    0.00    1   1.00
## MARRIED    6 1835    0.85    0.36    1.00    0.94    0.00    0.00    1   1.00
## AGE_GROUP  7 1835    1.00    0.00    1.00    1.00    0.00    1.00    1   0.00
##          skew kurtosis   se
## WAGE       8.61   139.95  7.95
## EDUC       0.07    0.54  0.07
## AGE        -0.25   -1.94  0.01

```

```

## RACE      2.95    6.70 0.01
## SMSA     1.65    0.71 0.01
## MARRIED   -1.95   1.82 0.01
## AGE_GROUP NaN     NaN 0.00
##
## -----
## AGE_GROUP: 2
##          vars   n   mean     sd median trimmed   mad   min   max range
## WAGE      1 2029 433.85 281.52 384.71 397.82 178.77 2.4 4634.17 4631.76
## EDUC      2 2029 12.90   3.22 12.00 12.90   2.97 0.0 20.00 20.00
## AGE       3 2029 42.50   0.50 43.00 42.50   0.00 42.0 43.00 1.00
## RACE      4 2029 0.08   0.27 0.00 0.00   0.00 0.0 1.00 1.00
## SMSA      5 2029 0.19   0.39 0.00 0.12   0.00 0.0 1.00 1.00
## MARRIED   6 2029 0.86   0.35 1.00 0.94   0.00 0.0 1.00 1.00
## AGE_GROUP 7 2029 2.00   0.00 2.00 2.00   0.00 2.0 2.00 0.00
##          skew kurtosis   se
## WAGE      3.45    30.53 6.25
## EDUC      -0.05    0.55 0.07
## AGE       -0.01   -2.00 0.01
## RACE      3.07    7.45 0.01
## SMSA      1.55    0.41 0.01
## MARRIED   -2.02   2.07 0.01
## AGE_GROUP NaN     NaN 0.00
##
## -----
## AGE_GROUP: 3
##          vars   n   mean     sd median trimmed   mad   min   max range
## WAGE      1 2035 426.73 245.77 384.71 398.04 171.07 1.98 2884.62 2882.64
## EDUC      2 2035 12.75   3.26 12.00 12.75   2.97 0.00 20.00 20.00
## AGE       3 2035 44.48   0.50 44.00 44.47   0.00 44.00 45.00 1.00
## RACE      4 2035 0.08   0.28 0.00 0.00   0.00 0.00 1.00 1.00
## SMSA      5 2035 0.19   0.40 0.00 0.12   0.00 0.00 1.00 1.00
## MARRIED   6 2035 0.87   0.34 1.00 0.96   0.00 0.00 1.00 1.00
## AGE_GROUP 7 2035 3.00   0.00 3.00 3.00   0.00 3.00 3.00 0.00
##          skew kurtosis   se
## WAGE      2.12    9.53 5.45
## EDUC      -0.01    0.38 0.07
## AGE       0.09   -1.99 0.01
## RACE      3.03    7.19 0.01
## SMSA      1.55    0.39 0.01
## MARRIED   -2.16   2.66 0.01
## AGE_GROUP NaN     NaN 0.00
##
## -----
## AGE_GROUP: 4
##          vars   n   mean     sd median trimmed   mad   min   max range
## WAGE      1 1876 444.74 341.11 391.25 408.69 191.22 0.15 10167.5 10167.35
## EDUC      2 1876 12.62   3.38 12.00 12.61   2.97 0.00 20.0 20.00
## AGE       3 1876 46.51   0.50 47.00 46.52   0.00 46.00 47.0 1.00
## RACE      4 1876 0.09   0.28 0.00 0.00   0.00 0.00 1.0 1.00
## SMSA      5 1876 0.17   0.38 0.00 0.09   0.00 0.00 1.0 1.00
## MARRIED   6 1876 0.86   0.35 1.00 0.95   0.00 0.00 1.0 1.00
## AGE_GROUP 7 1876 4.00   0.00 4.00 4.00   0.00 4.00 4.0 0.00
##          skew kurtosis   se
## WAGE      13.04   351.30 7.88
## EDUC      -0.06    0.47 0.08
## AGE       -0.05   -2.00 0.01

```

```

## RACE      2.96    6.74 0.01
## SMSA     1.74    1.03 0.01
## MARRIED   -2.05   2.19 0.01
## AGE_GROUP NaN     NaN 0.00
##
## -----
## AGE_GROUP: 5
##          vars   n   mean      sd median trimmed    mad   min   max range
## WAGE        1 2225 438.28 266.11 384.71  404.37 173.64  0.1 2667.5 2667.4
## EDUC        2 2225 12.37   3.37  12.00   12.40   2.97  0.0  20.0  20.0
## AGE         3 2225 48.69   0.67  49.00   48.62   1.48  48.0  50.0  2.0
## RACE        4 2225  0.08   0.27  0.00    0.00  0.00  0.0  1.0  1.0
## SMSA        5 2225  0.17   0.37  0.00    0.08  0.00  0.0  1.0  1.0
## MARRIED     6 2225  0.87   0.33  1.00    0.97  0.00  0.0  1.0  1.0
## AGE_GROUP   7 2225  5.00   0.00  5.00    5.00  0.00  5.0  5.0  0.0
##          skew kurtosis   se
## WAGE       2.05    7.54 5.64
## EDUC      -0.18    0.63 0.07
## AGE        0.45   -0.79 0.01
## RACE       3.06    7.37 0.01
## SMSA       1.79    1.21 0.01
## MARRIED   -2.24    3.00 0.01
## AGE_GROUP NaN     NaN 0.00

```

2.3 Covariance matrices per regio

[[1]]

Table 1: Covariance Matrix for REGION: 1

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	59622.389	326.291	19.889	-	-	8.220	-	4.997	-	0.758	0.763
EDUC	326.291	9.617	-	4.301	19.309	-	2.715	-	6.518	-	0.064
AGE	19.889	-0.350	8.596	0.001	-0.054	0.074	-	0.149	-	-	-
RACE	-4.301	-0.024	0.001	0.030	-0.005	-0.004	-	0.002	0.001	-	0.000
SMSA	-19.309	-0.152	-	-	0.127	-0.005	-	0.001	0.006	-	-
MARRIED	8.220	-0.054	0.074	-	-0.005	0.129	-	0.001	0.000	0.002	-
QOB	-2.715	0.029	-	-	-0.010	-0.005	1.217	-	-	0.127	0.365
QOB1	4.997	0.053	0.149	0.002	0.001	0.001	-	0.181	-	-	-
QOB2	-6.518	-0.071	-	0.001	0.006	0.000	-	-	0.189	-	-
QOB3	0.758	-0.046	-	-	-0.005	0.002	0.127	-	-	0.195	-
QOB4	0.763	0.064	-	0.000	-0.002	-0.003	0.365	-	-	-	0.186
				0.033				0.058	0.062	0.065	

[[2]]

Table 2: Covariance Matrix for REGION: 2

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	126020.884	335.536	29.869	-	-	16.273	-	0.966	1.492	-	-
EDUC	335.536	10.031	-	13.086	6.225	-	5.375	-	0.026	0.025	0.507 1.951
AGE	-	-	0.673	-0.158	-	0.091	-	0.105	-	0.024	0.027
RACE	29.869	-0.673	8.753	-0.016	0.009	-0.012	-	0.195	-	-	-
SMSA	-	-	-	0.084	-	-0.014	0.391	-	0.045	0.106	0.045
MARRIED	-	-	-	0.016	0.007	0.001	-	0.002	0.003	0.001	0.002
QOB	-	-	-	-	-	0.001	-	0.003	-	-	0.001
QOB1	-	-	-	-	-	-	-	0.180	-	-	-
QOB2	-	-	-	-	-	-	-	-	0.187	-	-
QOB3	-	-	-	-	-	-	-	-	-	0.194	-
QOB4	-	-	-	-	-	-	-	-	-	-	0.189

[[3]]

Table 3: Covariance Matrix for REGION: 3

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	69754.484	247.163	-	-	-	10.302	-	4.382	1.401	-	-
EDUC	247.163	9.398	9.152	6.535	13.792	-	13.651	-	-	2.296	3.487
AGE	-	-	-	-	-	-	-	-	-	-	-
RACE	-	-	0.690	0.083	-	-0.035	-0.018	-0.386	0.188	-	-
SMSA	-	-	-	-	-	0.021	-	-	-	0.069	0.040 0.079
MARRIED	-	-	-	-	-	-0.016	-	-0.004	0.002	-	-
QOB	-	-	-	-	-	-	-	-	-	0.001	0.001
QOB1	-	-	-	-	-	-	-	-	-	-	-
QOB2	-	-	-	-	-	-	-	-	-	-	-

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
QOB3	-2.296	-0.031	-	0.000	0.000	0.000	0.131	-	-	0.191	-
QOB4	-3.487	0.067	-	-	0.000	-0.010	0.373	-	-	-	0.186
			0.079	0.001				0.064	0.058	0.063	

[[4]]

Table 4: Covariance Matrix for REGION: 4

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	69294.730	276.252	-	-	-	12.774	13.151	-	-	3.178	2.367
EDUC	276.252	9.903	-0.788	-	-0.246	0.082	0.221	-	0.038	-	0.089
AGE	-25.048	-0.788	8.668	0.008	0.015	0.020	-	0.157	-	-	-
RACE	-4.650	-0.043	0.008	0.033	-0.005	-0.004	-	0.004	-	0.000	-
SMSA	-20.462	-0.246	0.015	-	0.226	0.005	-	0.003	0.004	-	0.005
MARRIED	12.774	0.082	0.020	-	0.005	0.101	-	-	0.003	0.005	-
QOB	13.151	0.221	-0.339	-	-0.005	-0.006	1.211	-	-	0.159	0.332
QOB1	-5.239	-0.085	0.157	0.004	0.003	-0.002	-	0.389	0.102	-	-
QOB2	-0.306	0.038	-0.039	-	0.004	0.003	-	-	0.186	-	-
QOB3	3.178	-0.041	-0.056	0.000	-0.012	0.005	0.159	-	-	0.198	-
QOB4	2.367	0.089	-0.063	-	0.005	-0.006	0.332	-	-	-	0.166
				0.002				0.057	0.051	0.057	

[[5]]

Table 5: Covariance Matrix for REGION: 5

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	68142.878	359.308	-	-	-	10.883	10.164	-	-	0.067	3.441
EDUC	359.308	12.144	13.139	15.912	13.042			3.215	0.293		
AGE	-13.139	-0.823	8.956	0.017	-0.038	0.053	-	0.181	-	-	-
RACE	-15.912	-0.232	0.017	0.119	0.001	-0.019	-	0.001	0.004	0.003	-
SMSA	-13.042	-0.257	-0.038	0.001	0.159	0.003	-	0.004	-	0.005	-
MARRIED	10.883	0.040	0.053	-0.019	0.003	0.118	-	0.000	-	0.002	-
							0.001	0.001	0.001	0.001	

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
QOB	10.164	0.124	-0.347	-0.015	-0.009	-0.001	1.276	-	-	0.128	0.379
QOB1	-3.215	-0.070	0.181	0.001	0.004	0.000	-	0.391	0.115	-	-
QOB2	-0.293	0.056	-0.043	0.004	-0.005	-0.001	-	0.191	-	-	-
QOB3	0.067	-0.025	-0.110	0.003	0.005	0.002	0.115	0.058	-	0.059	0.057
QOB4	3.441	0.040	-0.028	-0.008	-0.005	-0.001	0.379	-	-	0.194	-
								0.068	0.059	-	0.067
								-	-	-	0.190
								0.066	0.057	0.067	

[[6]]

Table 6: Covariance Matrix for REGION: 6

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	108586.360	289.292	-	-	-	3.987	32.889	-	-	1.185	11.524
EDUC	289.292	14.004	-1.028	-0.234	-0.204	0.117	0.345	8.655	4.054	-	0.071
AGE	-38.510	-1.028	7.682	0.018	0.036	-0.011	-	-	0.130	0.013	-
RACE	-10.947	-0.234	0.018	0.110	0.006	-0.022	0.010	-	-	0.085	0.042
SMSA	-20.349	-0.204	0.036	0.006	0.227	0.011	-	0.341	-	0.004	0.057
MARRIED	3.987	0.117	-0.011	-0.022	0.011	0.102	0.003	-	0.003	0.003	-
QOB	32.889	0.345	-0.341	0.010	-0.028	0.003	1.272	-	-	0.004	0.002
QOB1	-8.655	-0.130	0.185	-0.004	0.005	-0.004	-	0.400	0.106	-	0.140
QOB2	-4.054	-0.013	-0.085	0.003	0.007	0.003	-	-	0.199	-	0.366
QOB3	1.185	0.071	-0.042	-0.001	-0.001	0.003	0.140	-	-	0.063	-
QOB4	11.524	0.072	-0.057	0.003	-0.011	-0.002	0.366	-	-	-	0.192
								0.071	0.060	-	-
								-	-	-	0.062
								0.065	0.055	0.062	-
								-	-	-	0.181

[[7]]

Table 7: Covariance Matrix for REGION: 7

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	70637.866	357.447	26.277	-	-	11.726	8.040	-	-	8.554	-
EDUC	357.447	12.717	-	11.392	12.199	-	-	2.625	4.360	-	1.569
AGE	26.277	-0.614	8.851	-0.183	-0.219	0.008	0.193	-	-	0.012	0.076

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
RACE	-11.392	-0.183	0.008	0.082	0.004	-0.010	0.001	-	0.006	-	0.000
SMSA	-12.199	-0.219	-	0.004	0.140	0.004	0.016	-	-	0.003	0.008
MARRIED	11.726	0.008	0.055	-0.010	0.004	0.114	-	0.009	0.002	0.005	-
QOB	8.040	0.193	-	0.001	0.016	-0.009	1.224	-	-	0.140	0.352
QOB1	-2.625	-0.053	0.236	-0.003	-0.001	-0.002	-	0.185	-	-	-
QOB2	-4.360	-0.012	-	0.006	-0.005	0.005	-	0.380	0.112	-	-
QOB3	8.554	-0.012	-	-0.003	-0.002	0.005	0.140	-	0.163	0.063	0.050
QOB4	-1.569	0.076	-	0.000	0.008	-0.008	0.352	-	0.063	-	0.075
				0.145				0.059	0.050	0.075	

[[8]]

Table 8: Covariance Matrix for REGION: 8

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	55763.738	204.967	8.289	-	-	0.818	7.934	-	3.356	-	2.893
EDUC	204.967	8.725	-	5.134	19.118	-	-	4.199	-	2.050	-
AGE	8.289	-0.586	8.700	0.586	0.044	-0.049	0.027	-	0.084	-	0.004
RACE	-5.134	-0.044	-	0.033	-0.030	-0.005	-0.001	0.014	-	0.089	0.139
SMSA	-19.118	-0.214	-	0.049	0.005	0.211	0.008	0.024	-	0.000	-
MARRIED	0.818	0.034	0.027	-	0.008	0.096	0.003	0.000	0.001	-	0.005
QOB	7.934	0.208	-	0.014	0.024	0.003	1.293	-	-	0.127	0.384
QOB1	-4.199	-0.084	0.259	-	-0.008	0.000	-	0.397	0.114	-	-
QOB2	3.356	0.024	-	0.008	0.003	0.000	0.001	-	0.063	0.066	0.068
QOB3	-2.050	-0.004	-	0.089	-	-0.001	-0.006	0.127	-	0.180	-
QOB4	2.893	0.064	-	0.139	0.005	-0.001	-0.006	0.114	0.063	-	0.058
				0.032	0.000	0.009	0.005	0.384	-	0.066	0.062

[[9]]

Table 9: Covariance Matrix for REGION: 9

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	119854.165	270.881	33.568	-	-	10.765	-	0.768	-	3.349	-
EDUC	270.881	8.544	-	8.149	7.110	-	1.088	-	2.282	-	1.834
AGE	33.568	-0.359	8.370	-	-	-0.022	0.145	-	0.022	-	0.064
RACE	-8.149	-0.046	-	0.012	0.016	0.051	-	0.207	-	-	-
SMSA	-7.110	-0.113	-	0.058	-	-0.009	0.413	-	0.087	0.035	0.086
MARRIED	10.765	-0.022	0.051	-	0.005	0.156	0.008	-	0.002	0.004	-0.000
QOB	-1.088	0.145	-	-	-	0.009	0.008	1.207	-	-	0.138
QOB1	0.768	-0.052	0.207	0.002	0.003	-0.002	-	0.184	-	-	-
QOB2	-2.282	0.022	-	0.004	0.001	-0.002	-	0.367	0.367	0.123	-
QOB3	3.349	-0.034	-	-	-	-	0.123	0.060	-	0.068	0.057
QOB4	-1.834	0.064	-	0.000	0.004	0.003	0.351	-	0.068	-	0.180
						0.086	-	0.057	0.057	0.065	-

2.4 Correlation matrices per region

[[1]]

Table 10: Correlation Matrix for REGION: 1

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED	QOB	QOB1	QOB2	QOB3	QOB4
WAGE	1.000	0.431	0.028	-0.102	-0.222	0.094	-0.010	0.048	-0.061	0.007	0.007
EDUC	0.431	1.000	-0.038	-0.045	-0.137	-0.049	0.008	0.040	-0.053	-0.033	0.048
AGE	0.028	-0.038	1.000	0.002	-0.052	0.071	-0.071	0.119	-0.080	-0.010	-0.026
RACE	-0.102	-0.045	0.002	1.000	-0.075	-0.071	-0.027	0.024	0.017	-0.036	-0.004
SMSA	-0.222	-0.137	-0.052	-0.075	1.000	-0.037	-0.026	0.007	0.038	-0.031	-0.014
MARRIED	0.094	-0.049	0.071	-0.071	-0.037	1.000	-0.013	0.008	0.000	0.011	-0.019
QOB	-0.010	0.008	-0.071	-0.027	-0.026	-0.013	1.000	-0.768	-0.274	0.261	0.767
QOB1	0.048	0.040	0.119	0.024	0.007	0.008	-0.768	1.000	-0.324	-0.334	-0.318
QOB2	-0.061	-0.053	-0.080	0.017	0.038	0.000	-0.274	-0.324	1.000	-0.349	-0.332
QOB3	0.007	-0.033	-0.010	-0.036	-0.031	0.011	0.261	-0.334	-0.349	1.000	-0.342
QOB4	0.007	0.048	-0.026	-0.004	-0.014	-0.019	0.767	-0.318	-0.332	-0.342	1.000
						0.086	-	0.057	0.057	0.065	-

[[2]]

Table 11: Correlation Matrix for REGION: 2

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4	
WAGE	1.000	0.298	0.028	-0.127	-0.062	0.131	-0.014	0.006	0.010	-0.003	-0.013
EDUC	0.298	1.000	-0.072	-0.173	-0.098	0.082	-0.030	0.020	0.018	-0.017	-0.020
AGE	0.028	-0.072	1.000	-0.019	0.011	-0.011	-0.119	0.156	-0.035	-0.081	-0.035
RACE	-0.127	-0.173	-0.019	1.000	-0.084	-0.139	-0.005	-0.028	0.053	-0.012	-0.014
SMSA	-0.062	-0.098	0.011	-0.084	1.000	0.007	-0.010	0.025	-0.016	-0.015	0.007
MARRIED	0.131	0.082	-0.011	-0.139	0.007	1.000	0.036	-0.017	-0.025	0.009	0.033
QOB	-0.014	-0.030	-0.119	-0.005	-0.010	0.036	1.000	-0.768	-0.277	0.252	0.771
QOB1	0.006	0.020	0.156	-0.028	0.025	-0.017	-0.768	1.000	-0.320	-0.331	-0.323
QOB2	0.010	0.018	-0.035	0.053	-0.016	-0.025	-0.277	-0.320	1.000	-0.344	-0.335
QOB3	-0.003	-0.017	-0.081	-0.012	-0.015	0.009	0.252	-0.331	-0.344	1.000	-0.347
QOB4	-0.013	-0.020	-0.035	-0.014	0.007	0.033	0.771	-0.323	-0.335	-0.347	1.000

[[3]]

Table 12: Correlation Matrix for REGION: 3

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4	
WAGE	1.000	0.305	-0.012	-0.097	-0.142	0.116	-0.046	0.038	0.012	-0.020	-0.031
EDUC	0.305	1.000	-0.078	-0.107	-0.093	-0.007	0.032	-0.006	-0.022	-0.023	0.051
AGE	-0.012	-0.078	1.000	-0.029	-0.033	-0.019	-0.119	0.149	-0.057	-0.032	-0.064
RACE	-0.097	-0.107	-0.029	1.000	-0.115	-0.075	-0.015	0.021	-0.011	-0.004	-0.007
SMSA	-0.142	-0.093	-0.033	-0.115	1.000	0.049	-0.005	0.008	-0.006	0.000	-0.003
MARRIED	0.116	-0.007	-0.019	-0.075	0.049	1.000	-0.057	0.013	0.057	-0.002	-0.067
QOB	-0.046	0.032	-0.119	-0.015	-0.005	-0.057	1.000	-0.786	-0.242	0.267	0.769
QOB1	0.038	-0.006	0.149	0.021	0.008	0.013	-0.786	1.000	-0.331	-0.348	-0.339
QOB2	0.012	-0.022	-0.057	-0.011	-0.006	0.057	-0.242	-0.331	1.000	-0.327	-0.319
QOB3	-0.020	-0.023	-0.032	-0.004	0.000	-0.002	0.267	-0.348	-0.327	1.000	-0.335
QOB4	-0.031	0.051	-0.064	-0.007	-0.003	-0.067	0.769	-0.339	-0.319	-0.335	1.000

[[4]]

Table 13: Correlation Matrix for REGION: 4

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4	
WAGE	1.000	0.333	-0.032	-0.097	-0.164	0.153	0.045	-0.045	-0.003	0.027	0.022
EDUC	0.333	1.000	-0.085	-0.076	-0.164	0.082	0.064	-0.061	0.028	-0.029	0.069
AGE	-0.032	-0.085	1.000	0.015	0.010	0.021	-0.104	0.119	-0.030	-0.042	-0.053
RACE	-0.097	-0.076	0.015	1.000	-0.057	-0.075	-0.037	0.053	-0.038	0.004	-0.023
SMSA	-0.164	-0.164	0.010	-0.057	1.000	0.031	-0.010	0.013	0.022	-0.058	0.025
MARRIED	0.153	0.082	0.021	-0.075	0.031	1.000	-0.018	-0.012	0.024	0.033	-0.049
QOB	0.045	0.064	-0.104	-0.037	-0.010	-0.018	1.000	-0.791	-0.215	0.324	0.741
QOB1	-0.045	-0.061	0.119	0.053	0.013	-0.012	-0.791	1.000	-0.351	-0.375	-0.316
QOB2	-0.003	0.028	-0.030	-0.038	0.022	0.024	-0.215	-0.351	1.000	-0.348	-0.294
QOB3	0.027	-0.029	-0.042	0.004	-0.058	0.033	0.324	-0.375	-0.348	1.000	-0.313
QOB4	0.022	0.069	-0.053	-0.023	0.025	-0.049	0.741	-0.316	-0.294	-0.313	1.000

[[5]]

Table 14: Correlation Matrix for REGION: 5

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4
WAGE	1.000	0.395	-0.017	-0.177	-0.125	0.122	0.034	-0.028	-0.003	0.001
EDUC	0.395	1.000	-0.079	-0.193	-0.185	0.034	0.032	-0.046	0.039	-0.017
AGE	-0.017	-0.079	1.000	0.017	-0.032	0.052	-0.103	0.138	-0.034	-0.083
RACE	-0.177	-0.193	0.017	1.000	0.008	-0.157	-0.037	0.008	0.029	0.018
SMSA	-0.125	-0.185	-0.032	0.008	1.000	0.023	-0.019	0.025	-0.028	0.028
MARRIED0.122	0.034	0.052	-0.157	0.023	1.000	-0.003	0.000	-0.004	0.013	-0.010
QOB	0.034	0.032	-0.103	-0.037	-0.019	-0.003	1.000	-0.791	-0.245	0.257
QOB1	-0.028	-0.046	0.138	0.008	0.025	0.000	-0.791	1.000	-0.317	-0.352
QOB2	-0.003	0.039	-0.034	0.029	-0.028	-0.004	-0.245	-0.317	1.000	-0.321
QOB3	0.001	-0.017	-0.083	0.018	0.028	0.013	0.257	-0.352	-0.321	1.000
QOB4	0.030	0.026	-0.022	-0.053	-0.027	-0.010	0.769	-0.345	-0.314	-0.350
										1.000

[[6]]

Table 15: Correlation Matrix for REGION: 6

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4
WAGE	1.000	0.235	-0.042	-0.100	-0.130	0.038	0.088	-0.059	-0.029	0.008
EDUC	0.235	1.000	-0.099	-0.189	-0.114	0.097	0.082	-0.078	-0.008	0.043
AGE	-0.042	-0.099	1.000	0.019	0.027	-0.012	-0.109	0.149	-0.073	-0.035
RACE	-0.100	-0.189	0.019	1.000	0.041	-0.203	0.026	-0.030	0.018	-0.008
SMSA	-0.130	-0.114	0.027	0.041	1.000	0.070	-0.052	0.022	0.037	-0.006
MARRIED0.038	0.097	-0.012	-0.203	0.070	1.000	0.009	-0.030	0.023	0.024	-0.016
QOB	0.088	0.082	-0.109	0.026	-0.052	0.009	1.000	-0.795	-0.223	0.283
QOB1	-0.059	-0.078	0.149	-0.030	0.022	-0.030	-0.795	1.000	-0.335	-0.363
QOB2	-0.029	-0.008	-0.073	0.018	0.037	0.023	-0.223	-0.335	1.000	-0.323
QOB3	0.008	0.043	-0.035	-0.008	-0.006	0.024	0.283	-0.363	-0.323	1.000
QOB4	0.082	0.045	-0.048	0.022	-0.054	-0.016	0.762	-0.342	-0.305	-0.330
										1.000

[[7]]

Table 16: Correlation Matrix for REGION: 7

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4
WAGE	1.000	0.377	0.033	-0.149	-0.122	0.130	0.027	-0.023	-0.041	0.070
EDUC	0.377	1.000	-0.058	-0.179	-0.164	0.007	0.049	-0.034	-0.008	-0.007
AGE	0.033	-0.058	1.000	0.010	-0.023	0.055	-0.168	0.184	-0.055	-0.018
RACE	-0.149	-0.179	0.010	1.000	0.036	-0.105	0.003	-0.025	0.048	-0.020
SMSA	-0.122	-0.164	-0.023	0.036	1.000	0.033	0.039	-0.007	-0.036	-0.010
MARRIED0.130	0.007	0.055	-0.105	0.033	1.000	-0.025	-0.013	0.036	0.034	-0.057
QOB	0.027	0.049	-0.168	0.003	0.039	-0.025	1.000	-0.797	-0.251	0.273
QOB1	-0.023	-0.034	0.184	-0.025	-0.007	-0.013	-0.797	1.000	-0.289	-0.380
QOB2	-0.041	-0.008	-0.055	0.048	-0.036	0.036	-0.251	-0.289	1.000	-0.338
QOB3	0.070	-0.007	-0.018	-0.020	-0.010	0.034	0.273	-0.380	-0.338	1.000
QOB4	-0.014	0.050	-0.114	0.002	0.052	-0.057	0.743	-0.322	-0.287	-0.377
										1.000

[[8]]

Table 17: Correlation Matrix for REGION: 8

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4	
WAGE	1.000	0.294	0.012	-0.127	-0.176	0.011	0.030	-0.040	0.033	-0.020	0.028
EDUC	0.294	1.000	-0.067	-0.087	-0.158	0.037	0.062	-0.064	0.019	-0.003	0.050
AGE	0.012	-0.067	1.000	-0.065	-0.036	0.029	-0.138	0.198	-0.071	-0.109	-0.025
RACE	-0.127	-0.087	-0.065	1.000	-0.065	-0.015	0.070	-0.107	0.041	0.064	0.006
SMSA	-0.176	-0.158	-0.036	-0.065	1.000	0.057	0.046	-0.038	0.000	-0.005	0.044
MARRIED	0.011	0.037	0.029	-0.015	0.057	1.000	0.009	0.002	0.007	-0.046	0.036
QOB	0.030	0.062	-0.138	0.070	0.046	0.009	1.000	-0.789	-0.236	0.259	0.777
QOB1	-0.040	-0.064	0.198	-0.107	-0.038	0.002	-0.789	1.000	-0.335	-0.344	-0.351
QOB2	0.033	0.019	-0.071	0.041	0.000	0.007	-0.236	-0.335	1.000	-0.316	-0.322
QOB3	-0.020	-0.003	-0.109	0.064	-0.005	-0.046	0.259	-0.344	-0.316	1.000	-0.331
QOB4	0.028	0.050	-0.025	0.006	0.044	0.036	0.777	-0.351	-0.322	-0.331	1.000

[[9]]

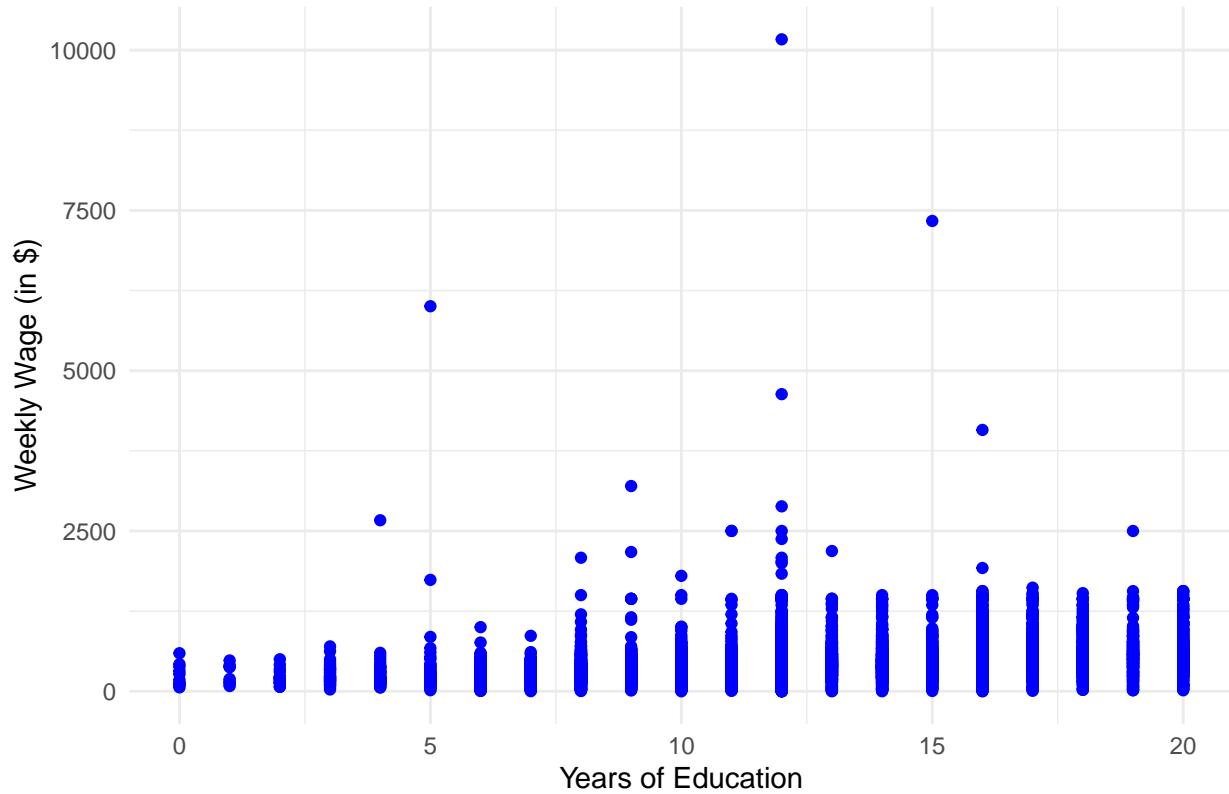
Table 18: Correlation Matrix for REGION: 9

	WAGE	EDUC	AGE	RACE	SMSA	MARRIEDQOB	QOB1	QOB2	QOB3	QOB4	
WAGE	1.000	0.268	0.034	-0.098	-0.067	0.079	-0.003	0.005	-0.015	0.022	-0.013
EDUC	0.268	1.000	-0.042	-0.066	-0.127	-0.019	0.045	-0.041	0.017	-0.026	0.052
AGE	0.034	-0.042	1.000	-0.017	-0.018	0.044	-0.130	0.167	-0.070	-0.027	-0.070
RACE	-0.098	-0.066	-0.017	1.000	-0.066	-0.096	-0.032	0.021	0.035	-0.050	-0.004
SMSA	-0.067	-0.127	-0.018	-0.066	1.000	0.044	-0.009	0.021	0.009	-0.056	0.029
MARRIED	0.079	-0.019	0.044	-0.096	0.044	1.000	0.018	-0.011	-0.010	0.005	0.016
QOB	-0.003	0.045	-0.130	-0.032	-0.009	0.018	1.000	-0.777	-0.260	0.281	0.755
QOB1	0.005	-0.041	0.167	0.021	0.021	-0.011	-0.777	1.000	-0.323	-0.352	-0.314
QOB2	-0.015	0.017	-0.070	0.035	0.009	-0.010	-0.260	-0.323	1.000	-0.352	-0.315
QOB3	0.022	-0.026	-0.027	-0.050	-0.056	0.005	0.281	-0.352	-0.352	1.000	-0.343
QOB4	-0.013	0.052	-0.070	-0.004	0.029	0.016	0.755	-0.314	-0.315	-0.343	1.000

```
## Warning: Use of 'df$EDUC' is discouraged.
## i Use 'EDUC' instead.
```

```
## Warning: Use of 'df$WAGE' is discouraged.
## i Use 'WAGE' instead.
```

Relationship Between Education and Wage



3. Estimate the baseline Specification

```
#linear model
linear_model1=lm(WAGE~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3
+ REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 , data = df)

##
## =====
##             Dependent variable:
##             -----
##                   WAGE
## -----
## EDUC           26.696*** 
##                  (0.869)
## t = 30.724
## p = 0.000
## AGE            2.244** 
##                  (0.942)
## t = 2.382
## p = 0.018
## RACE          -77.632*** 
##                  (10.225)
## t = -7.592
```

```

##          p = 0.000
##  SMSA      -63.756***  

##                      (7.393)
##          t = -8.624
##          p = 0.000
##  MARRIED    78.001***  

##                      (8.026)
##          t = 9.719
##          p = 0.000
##  REGION2    41.193***  

##                      (13.644)
##          t = 3.019
##          p = 0.003
##  REGION3    49.088***  

##                      (13.305)
##          t = 3.690
##          p = 0.0003
##  REGION4    18.845  

##                      (15.600)
##          t = 1.208
##          p = 0.228
##  REGION5    4.205  

##                      (13.492)
##          t = 0.312
##          p = 0.756
##  REGION6    7.604  

##                      (16.123)
##          t = 0.472
##          p = 0.638
##  REGION7    17.081  

##                      (14.646)
##          t = 1.166
##          p = 0.244
##  REGION8    15.897  

##                      (17.106)
##          t = 0.929
##          p = 0.353
##  REGION9    63.710***  

##                      (14.046)
##          t = 4.536
##          p = 0.00001
##  Constant   -81.669*  

##                      (46.614)
##          t = -1.752
##          p = 0.080
##  -----
##  Observations           10,000
##  R2                    0.134
##  Adjusted R2            0.133
##  Residual Std. Error     275.029 (df = 9986)
##  F Statistic        118.953*** (df = 13; 9986) (p = 0.000)
##  =====
##  Note:                  *p<0.1; **p<0.05; ***p<0.01

```

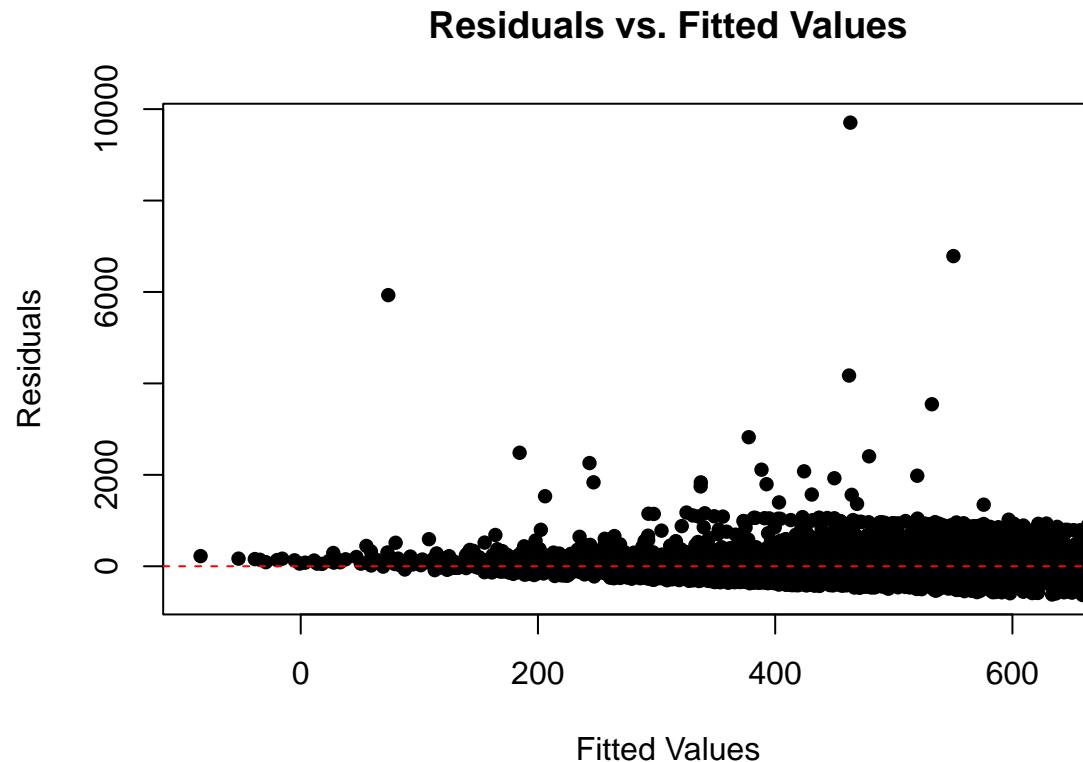
```

##
## Linear hypothesis test:
## AGE = 0
## RACE = 0
## MARRIED = 0
## SMSA = 0
##
## Model 1: restricted model
## Model 2: WAGE ~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
##
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1   9990  773577022
## 2   9986  755352847  4  18224175 60.232 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Linear hypothesis test:
## REGION2 = 0
## REGION3 = 0
## REGION4 = 0
## REGION5 = 0
## REGION6 = 0
## REGION7 = 0
## REGION8 = 0
## REGION9 = 0
##
## Model 1: restricted model
## Model 2: WAGE ~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
##
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1   9994  759957235
## 2   9986  755352847  8  4604388 7.6089 3.389e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

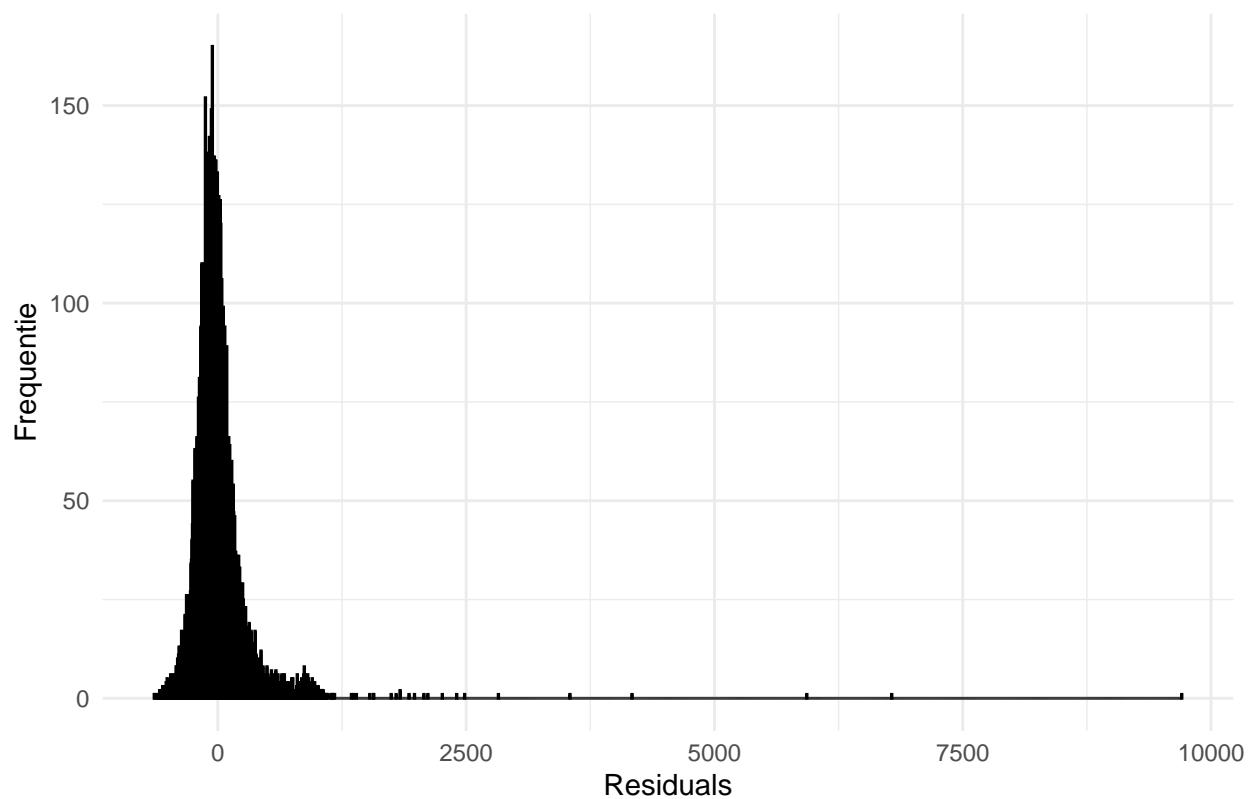
```

4. Evaluating Gauss-Markov Assumptions and Applying Remedial Measures

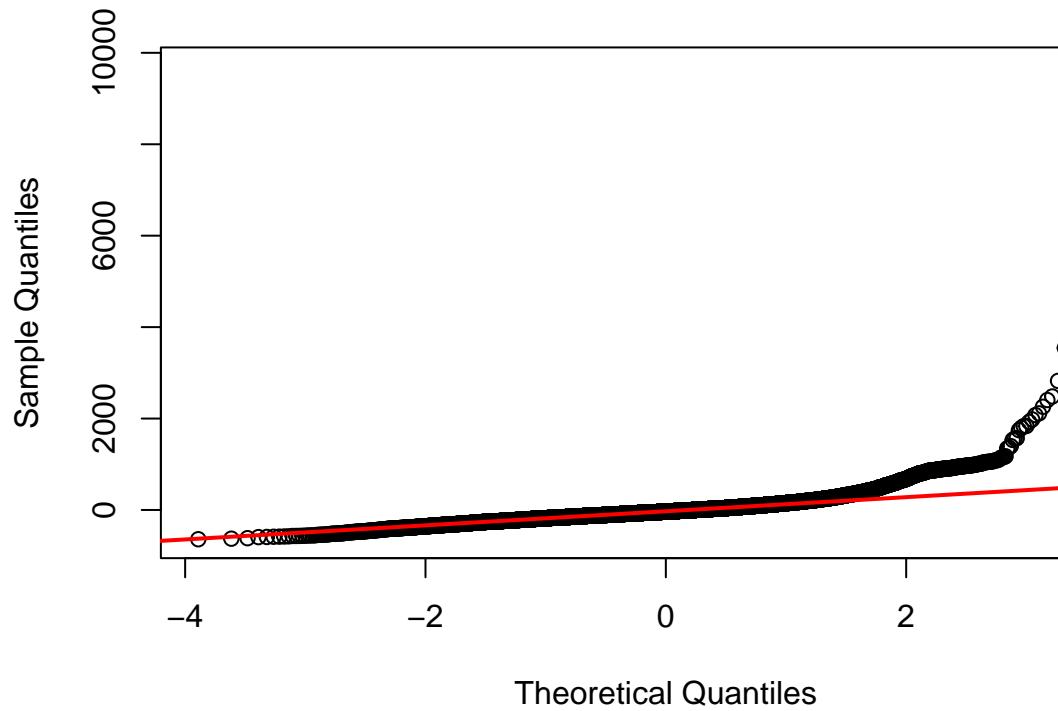


4a. Stochastic Regressors

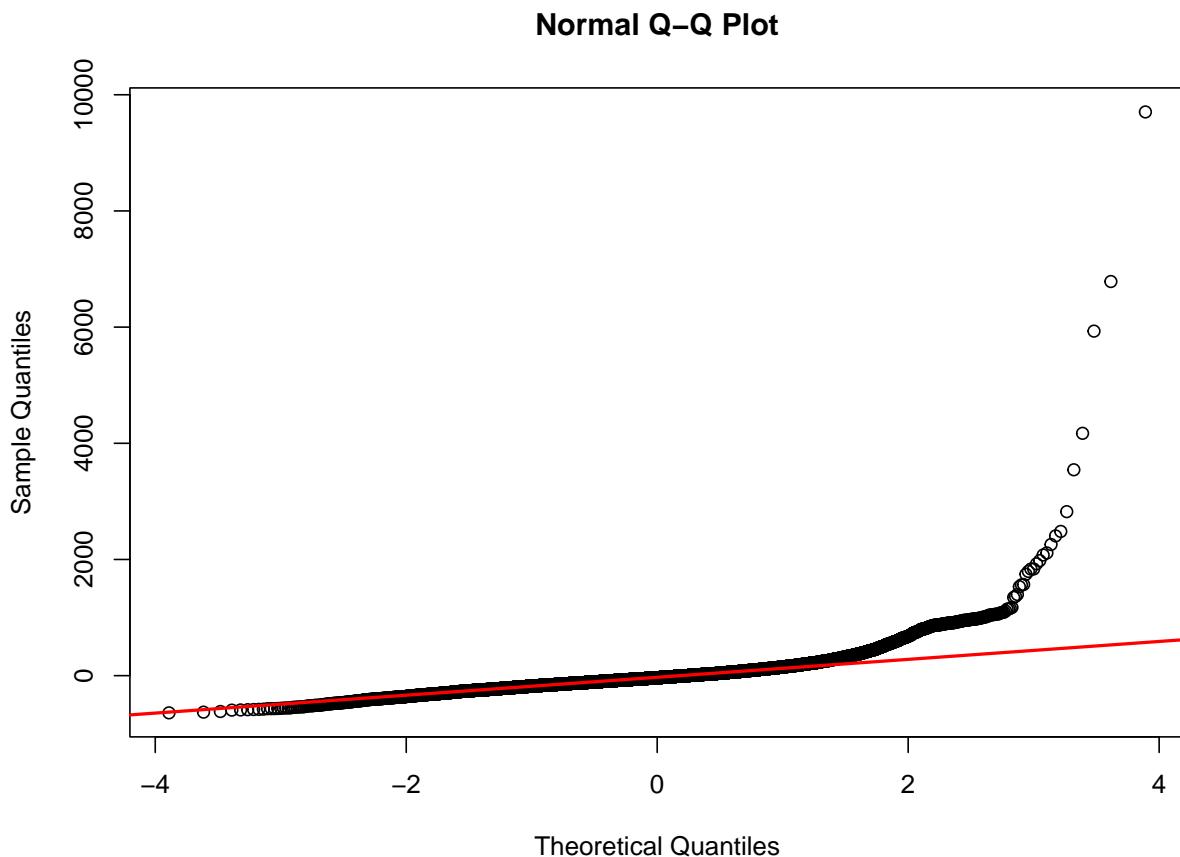
Histogram Residuals



Normal Q–Q Plot



4b. Normality Error Terms



Jarque Bera Test

```
data: residuals X-squared = 21722226, df = 2, p-value < 2.2e-16
```

```
print(mean(residuals))

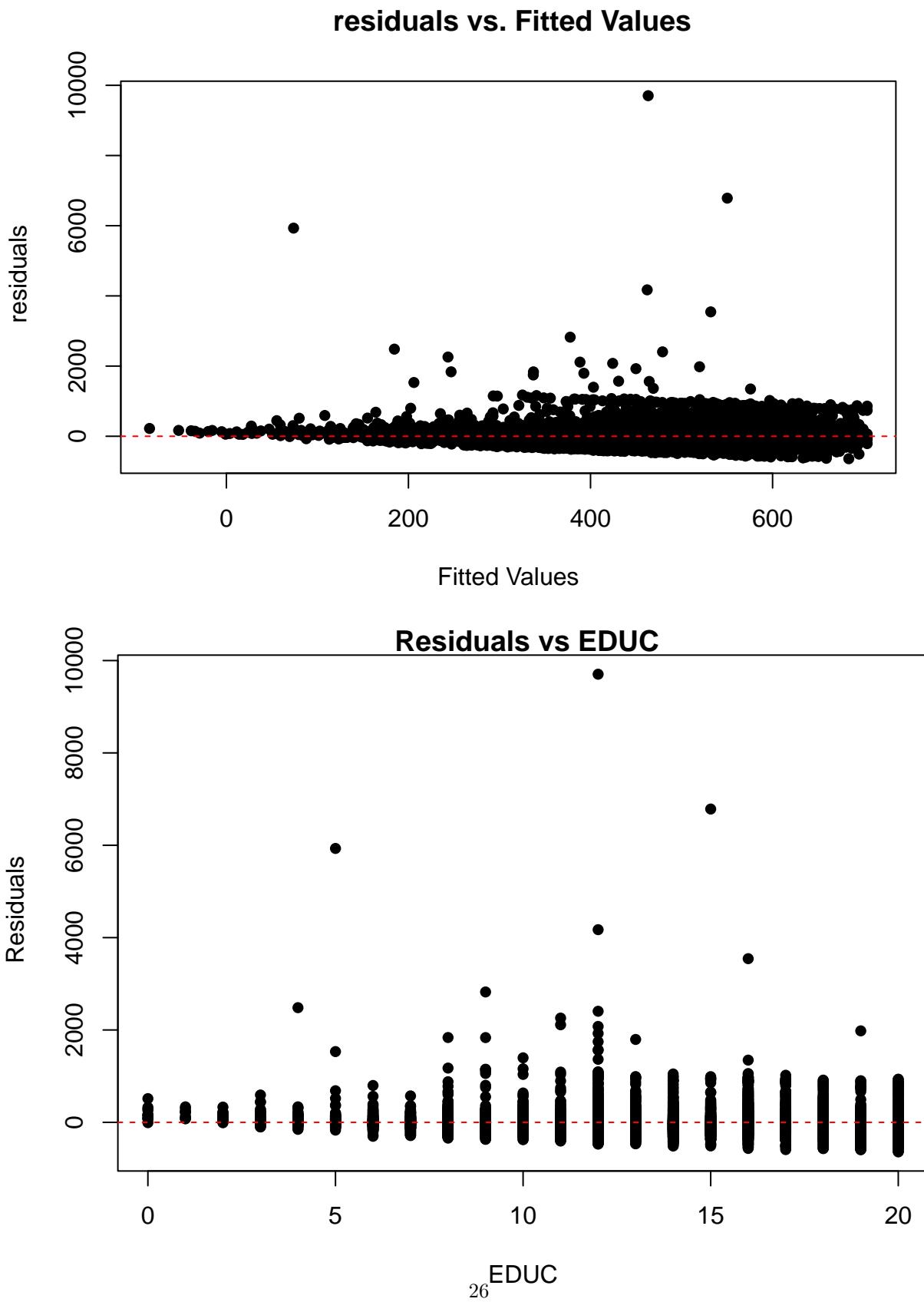
## [1] -4.605628e-16

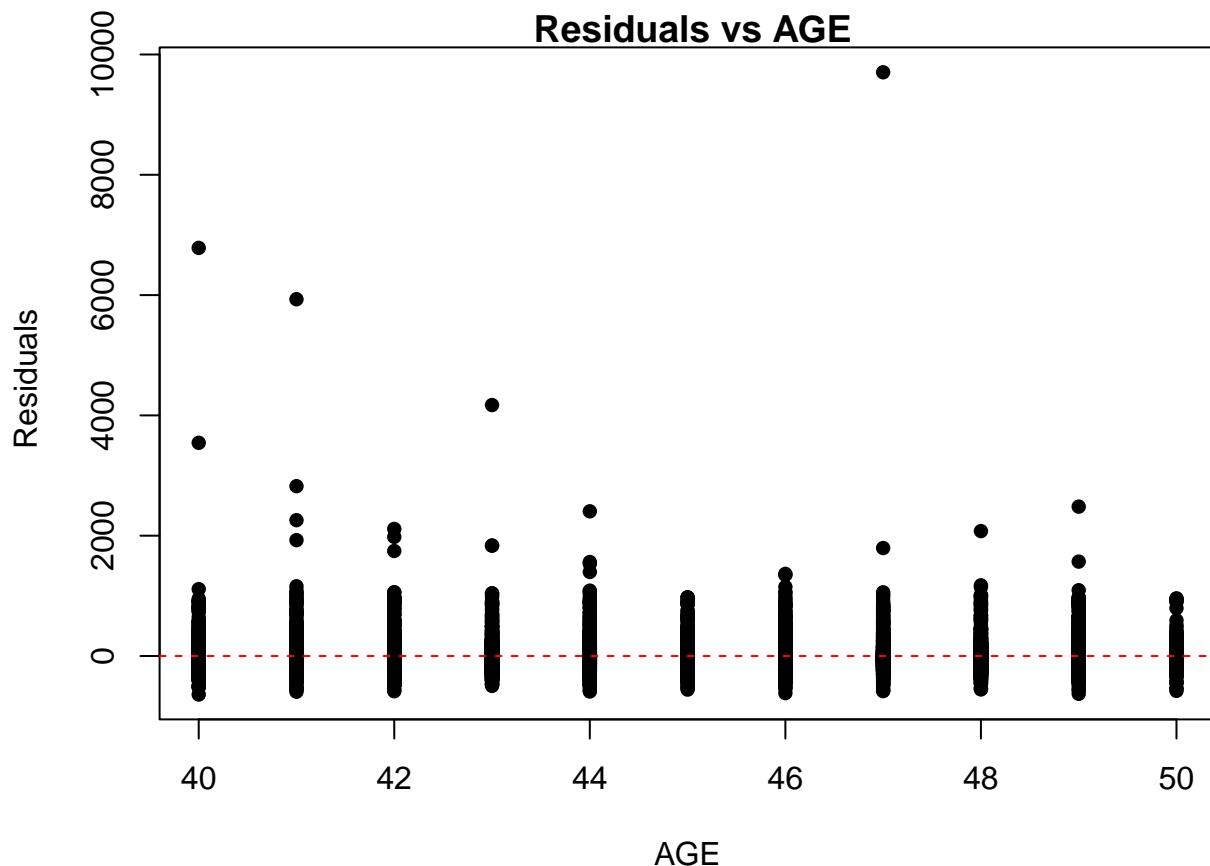
print(t.test(residuals, mu = 0))

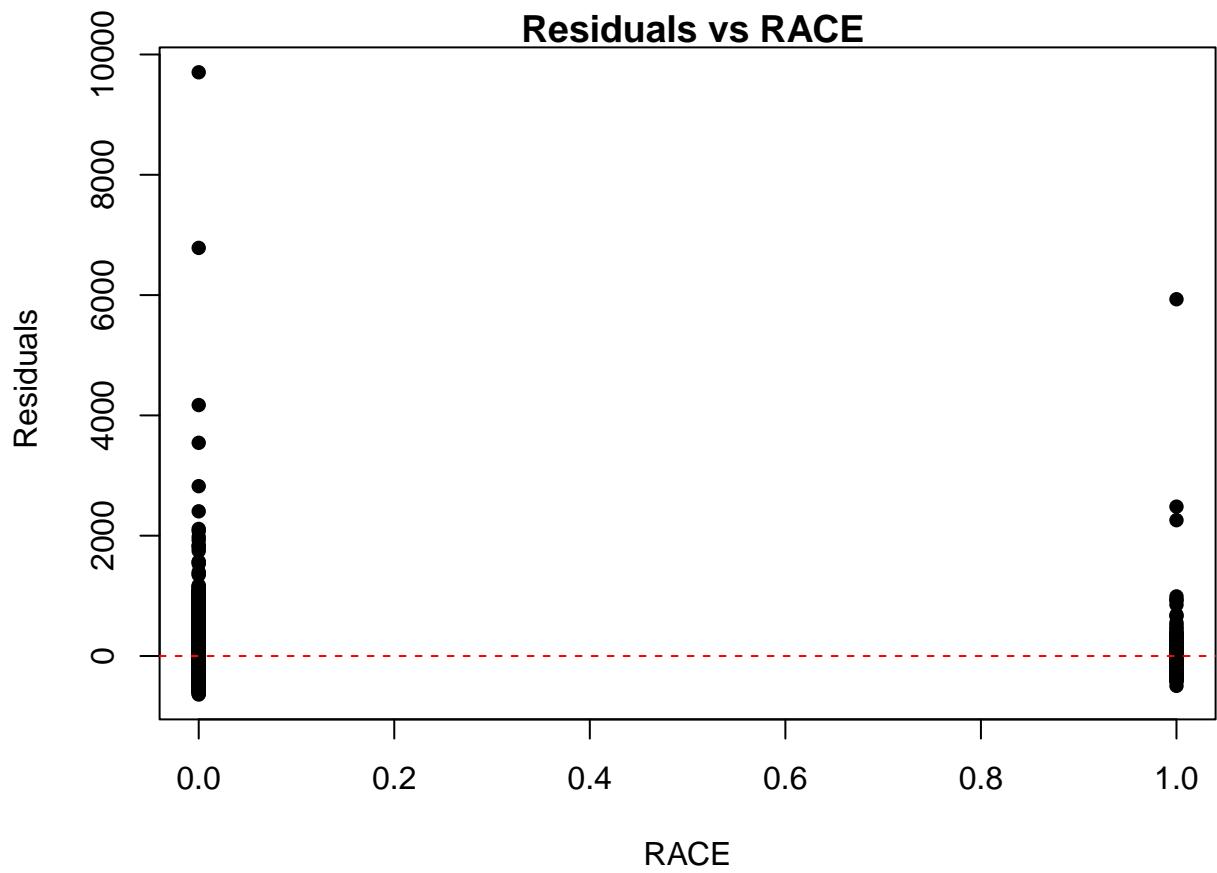
##
## One Sample t-test
##
## data: residuals
## t = -1.6757e-16, df = 9999, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -5.387624 5.387624
## sample estimates:
##      mean of x
## -4.605628e-16
```

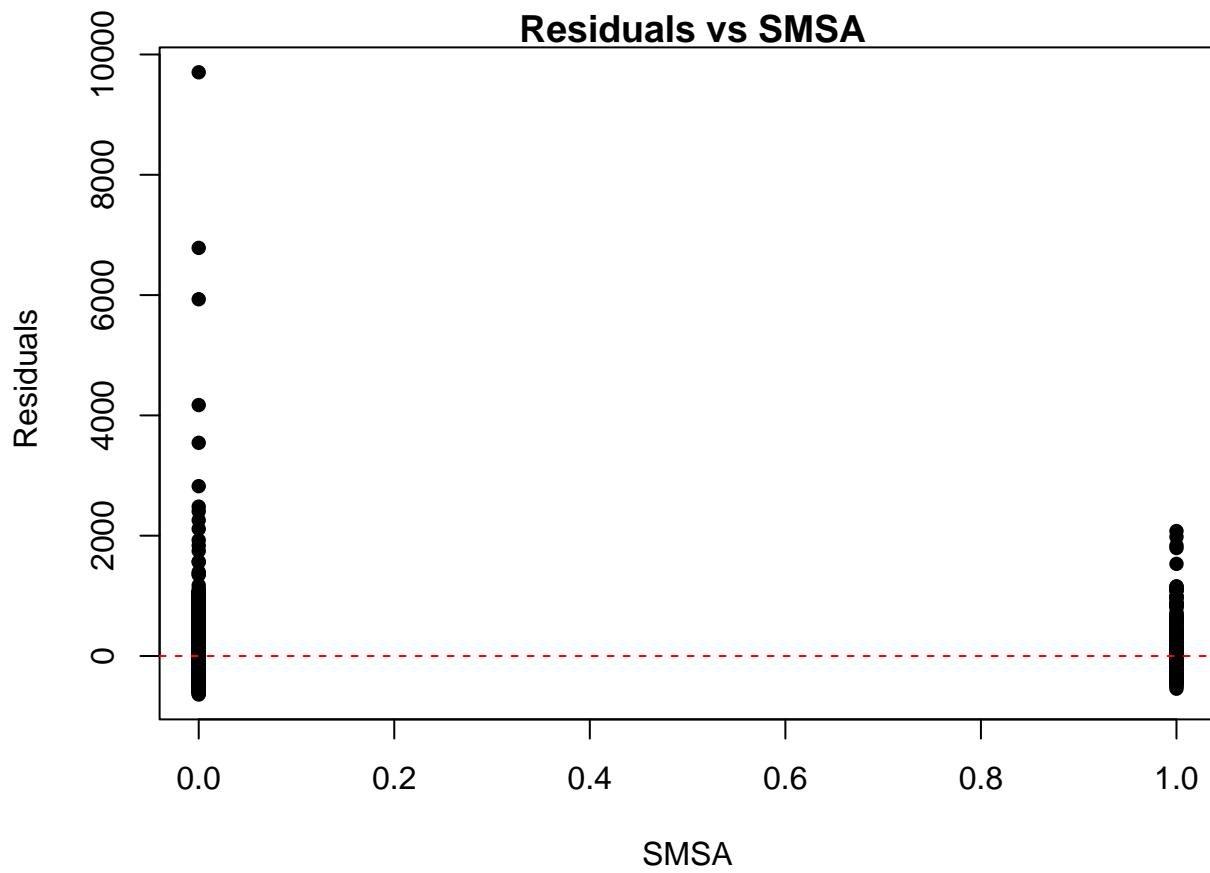
```
if (show_interpretation) {  
  cat("Assumption 3: The expected V value of the error terms is zero: CHECK")  
}
```

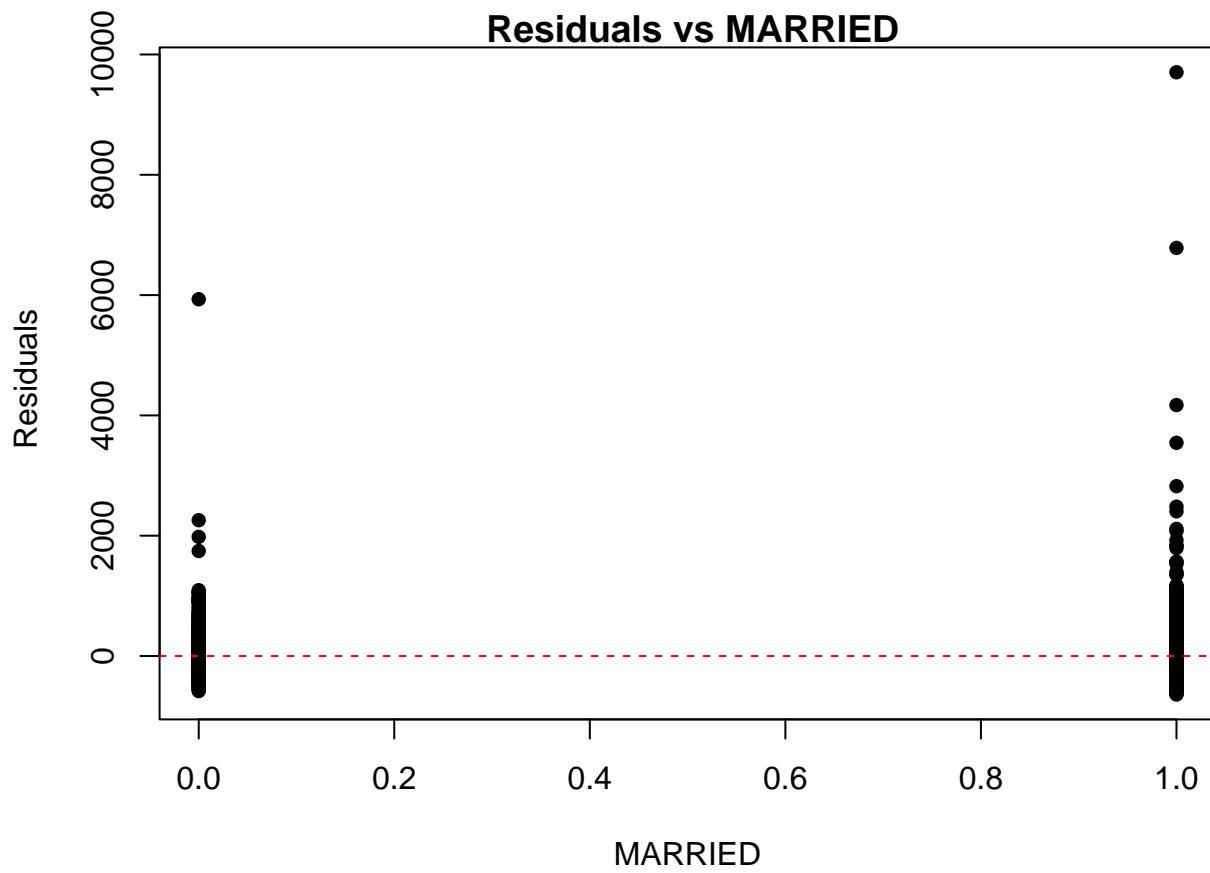
4(d) Heteroskedasticity

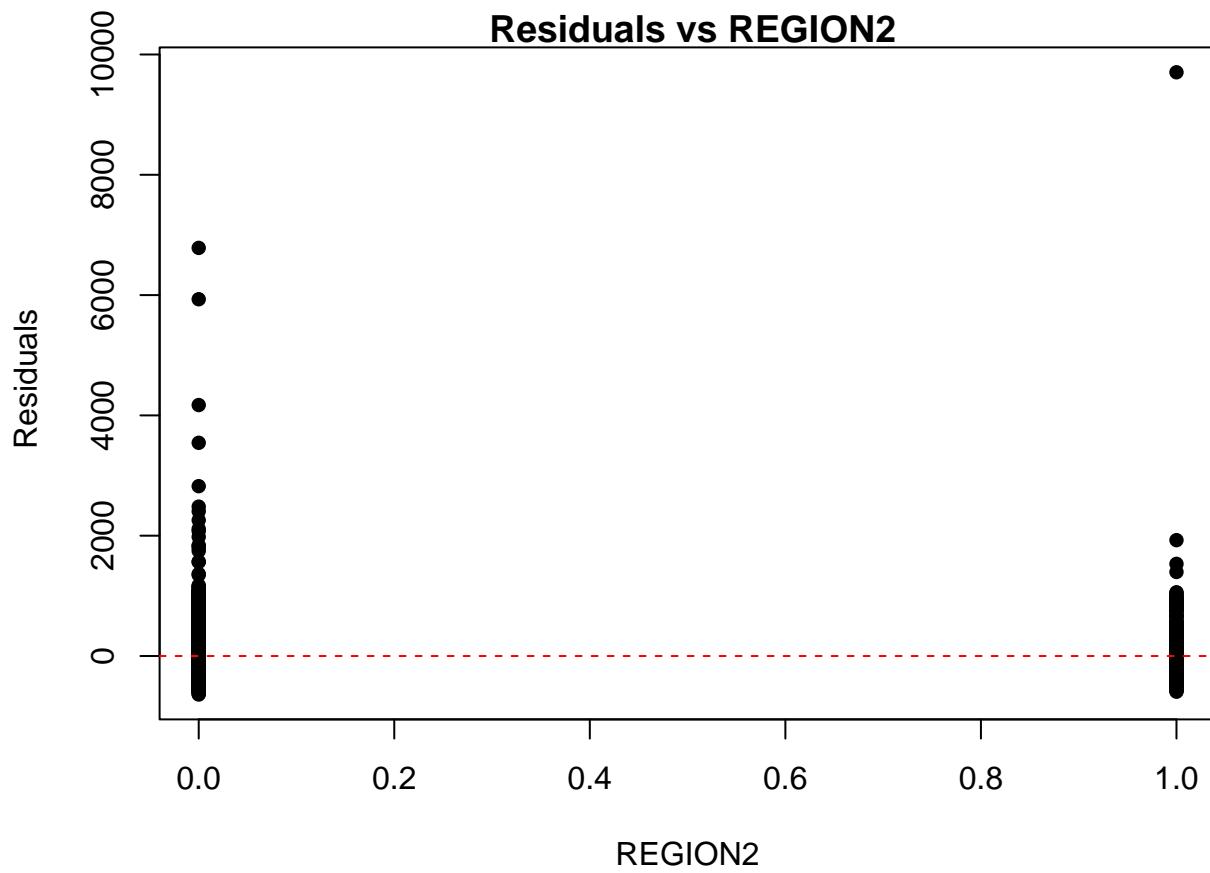


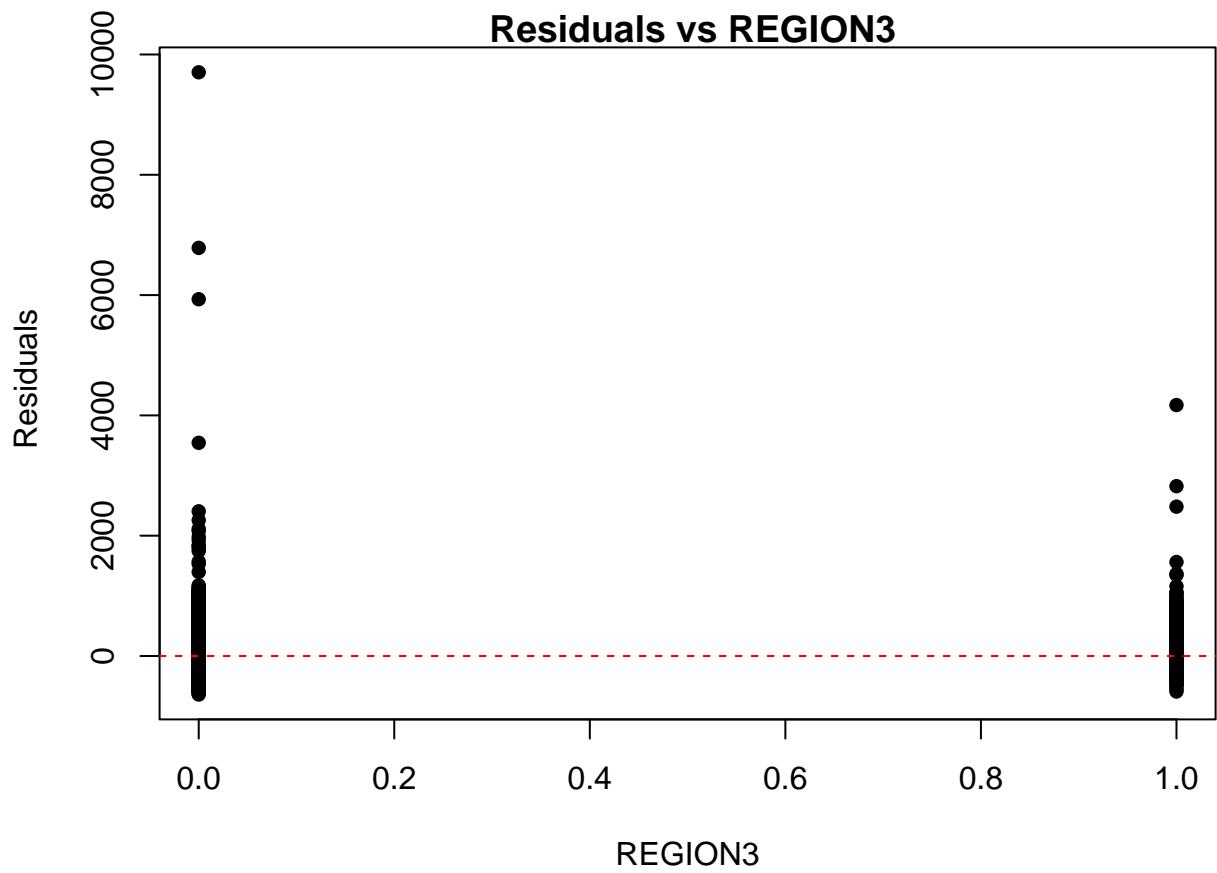


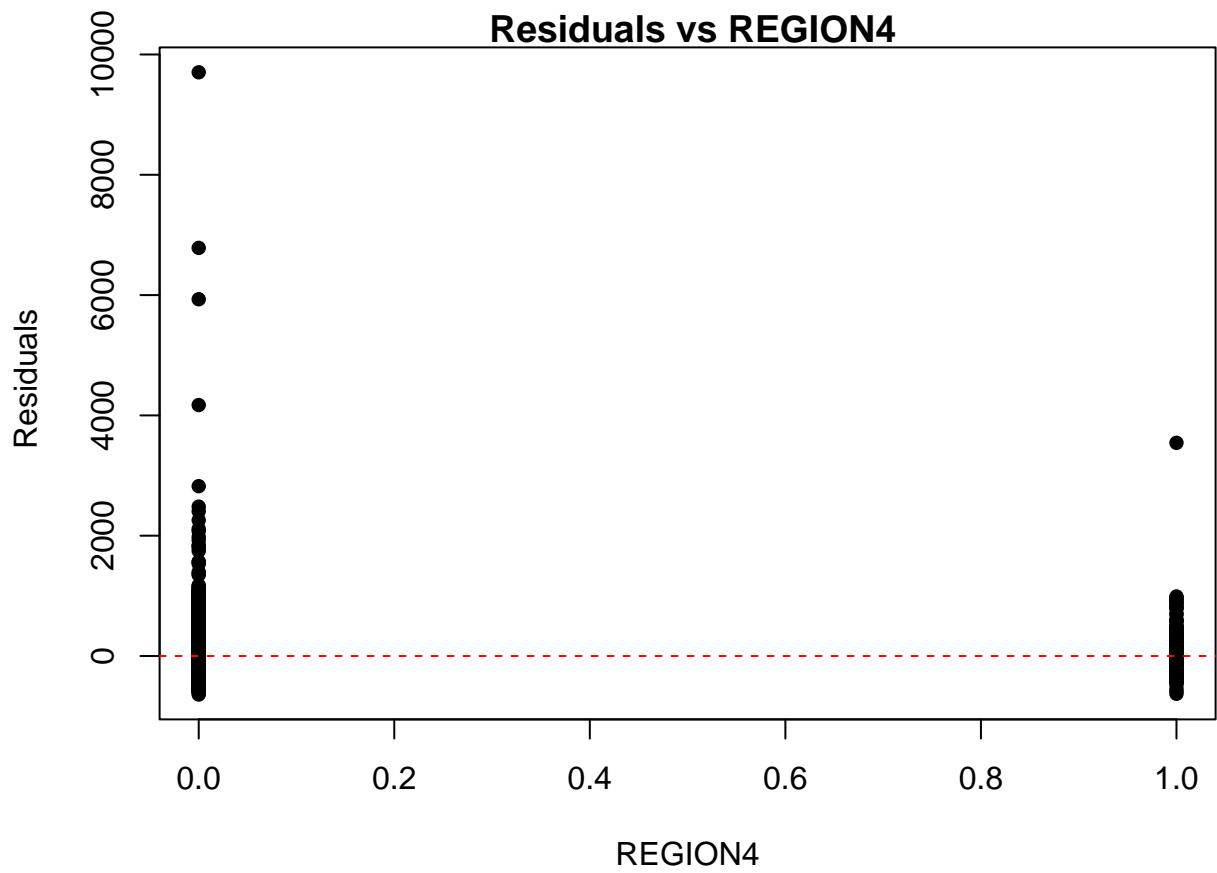


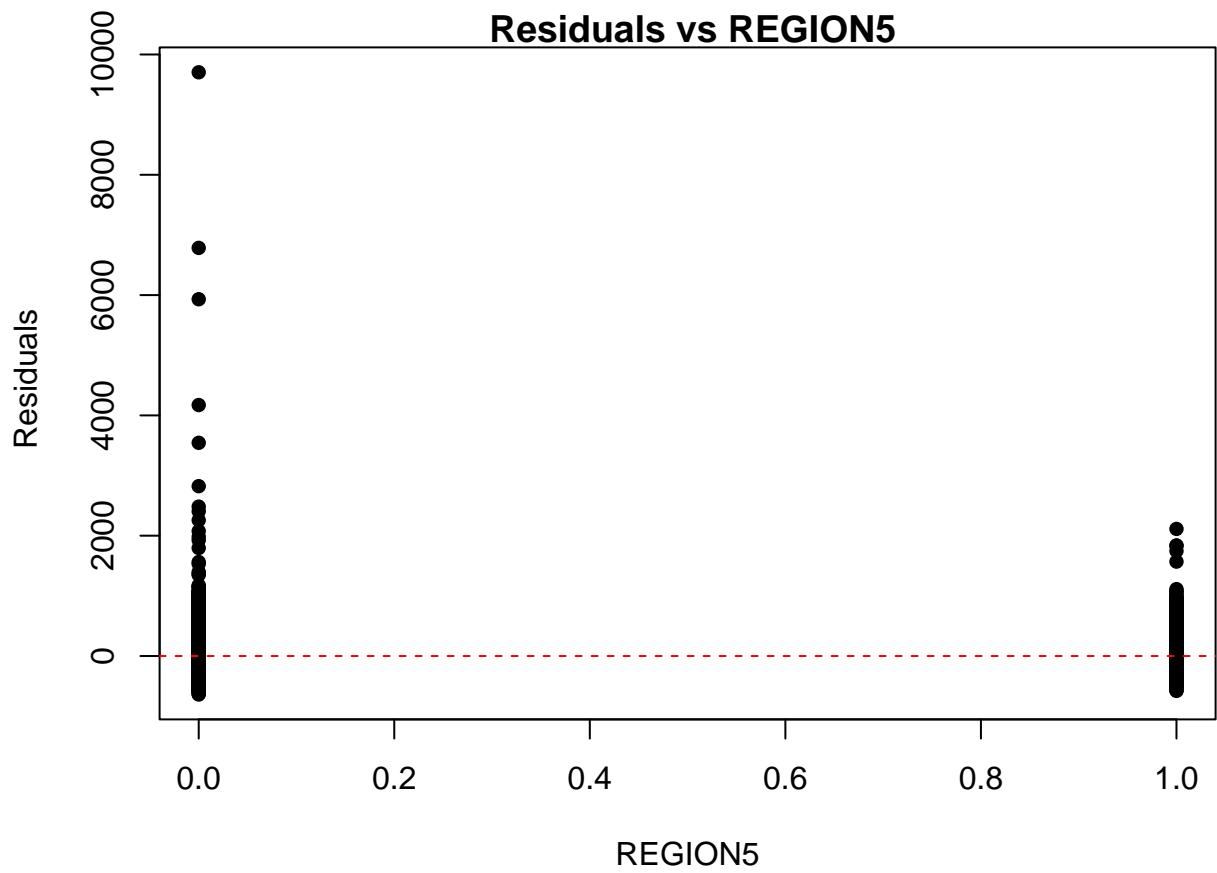


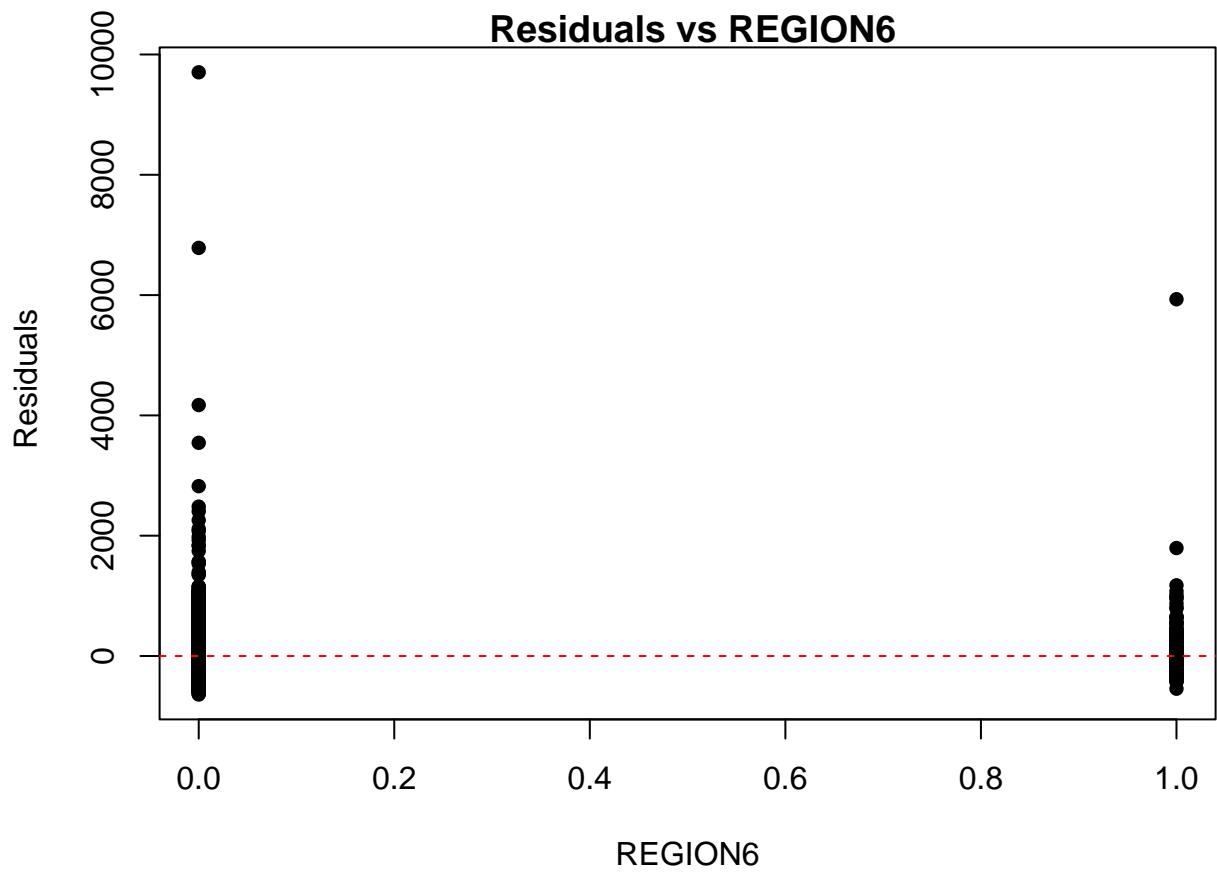


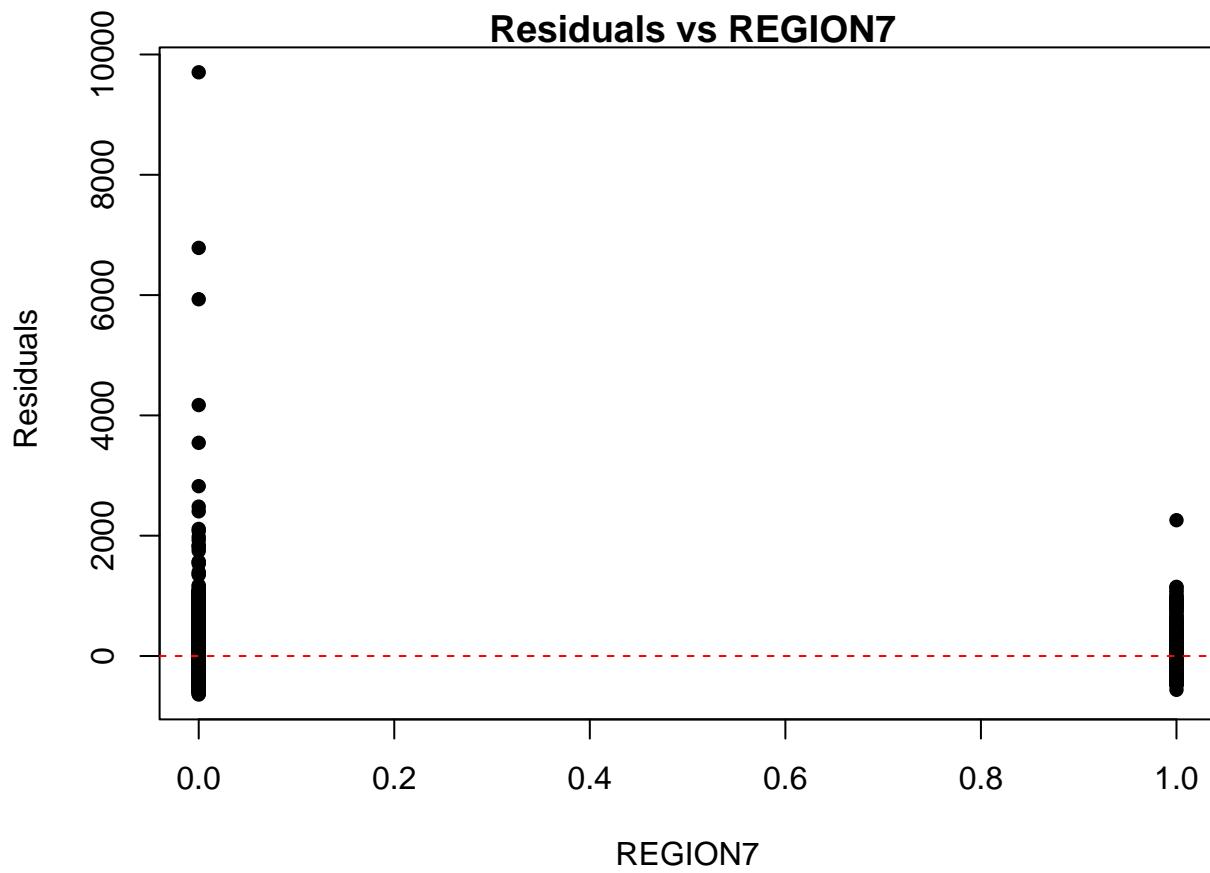


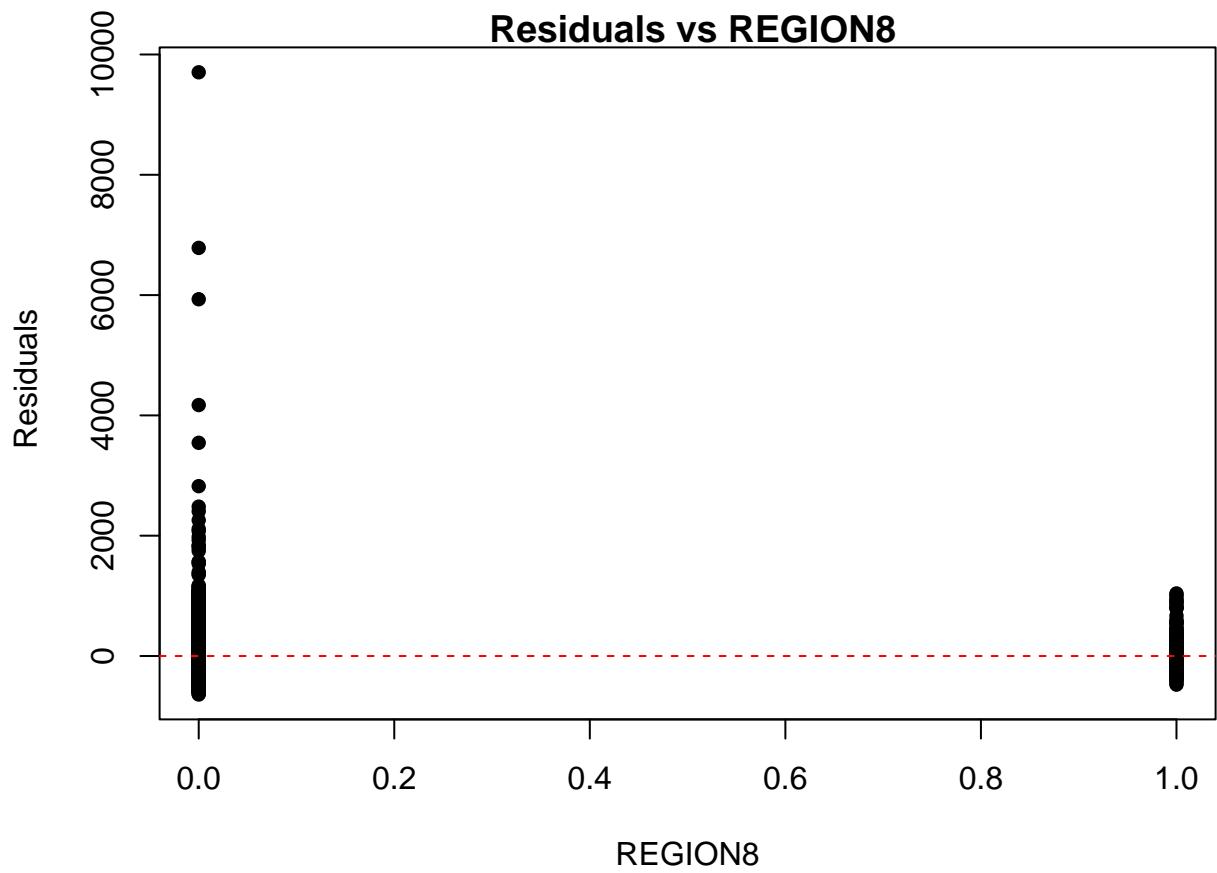


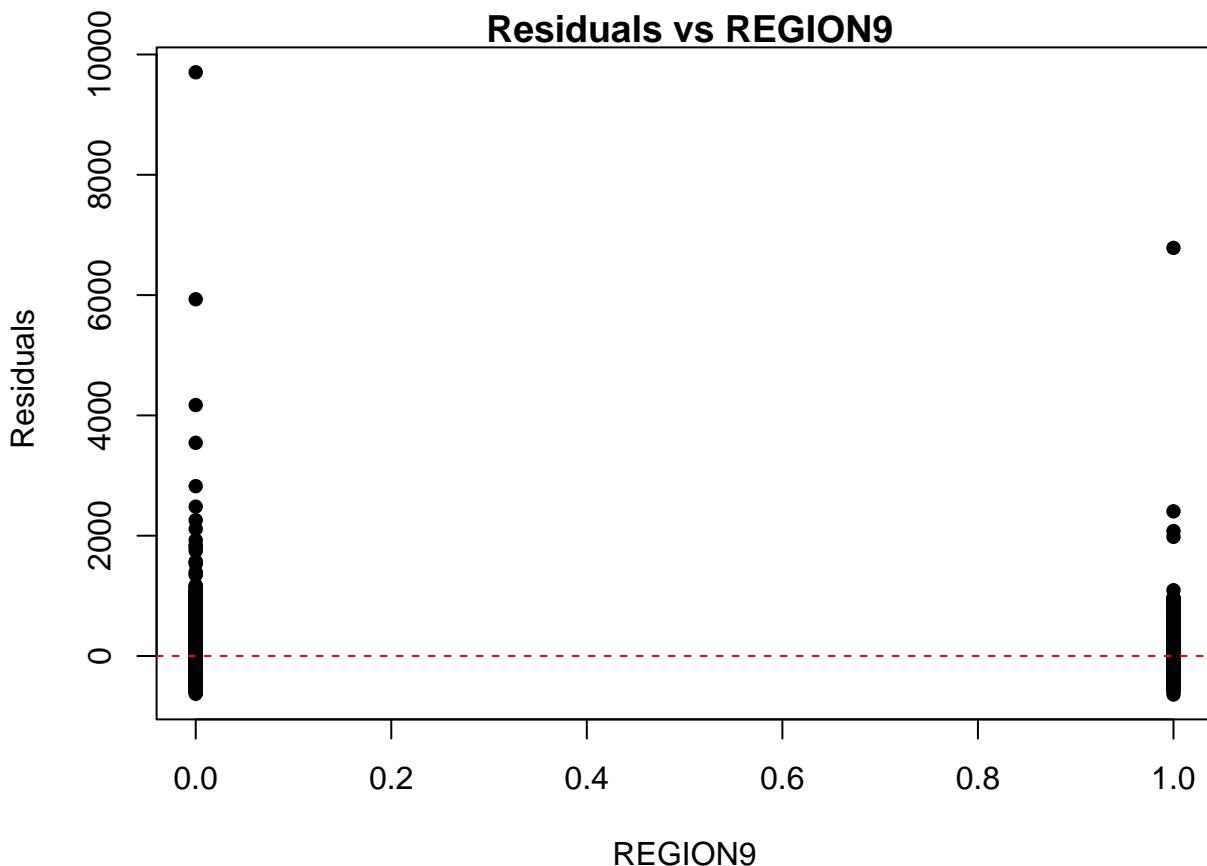












4(d).1 Heteroskedasticity test **

```
white_test <- bptest(linear_model1, ~ fitted(linear_model1) + I(fitted(linear_model1)^2))
print(white_test)
```

white's general test

studentized Breusch-Pagan test

data: linear_model1 BP = 7.0287, df = 2, p-value = 0.02977

```
# Step 1: Define a trimmed version of your data (removing outliers based on WAGE, for example)
# Adjust variable and trimming logic as needed
trimmed_data <- df %>%
  filter(between(WAGE, quantile(WAGE, 0.01, na.rm = TRUE),
                quantile(WAGE, 0.99, na.rm = TRUE)))

# Step 2: Refit your linear model on the trimmed dataset
linear_model_trimmed <- lm(WAGE ~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3
                           + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 , data = trimmed_data)

# Step 3: White test on the new model
```

```

white_test_trimmed <- bptest(linear_model_trimmed,
                                ~ fitted(linear_model_trimmed) + I(fitted(linear_model_trimmed)^2))
print(white_test_trimmed)

```

studentized Breusch-Pagan test

data: linear_model_trimmed BP = 361.55, df = 2, p-value < 2.2e-16

```

gqtest_result <- gqtest(linear_model1, order.by =df$EDUC ,fraction = 1000)
print(gqtest_result)

```

goldfeld-quant test

Goldfeld-Quandt test

data: linear_model1 GQ = 1.8257, df1 = 4486, df2 = 4486, p-value < 2.2e-16 alternative hypothesis: variance increases from segment 1 to 2

4(d).2 addressing heteroskedasticity

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, mei 15, 2025 - 1:09:37

EGLS

```

# Step 1: Get design matrix and response variable
X <- model.matrix(linear_model1) # Includes intercept & region dummies
Y <- df$WAGE

# Step 2: Transform X and Y
y_hat <- fitted(linear_model1)
X_star <- X / y_hat
Y_star <- Y / y_hat # Element-wise division

# Step 3: EGLS estimation (no intercept needed as it's in X_star)
egls_model <- lm(Y_star ~ X_star - 1) # -1 removes additional intercept

stargazer(egls_model,
          type = "latex",
          title = "EGLS Estimation Results",
          dep.var.labels = "Transformed WAGE",
          covariate.labels = c("Intercept", "EDUC", "AGE", "RACE", "SMSA", "MARRIED",
                               "REGION2", "REGION3", "REGION4", "REGION5",
                               "REGION6", "REGION7", "REGION8", "REGION9"),
          notes = "Variables transformed by dividing by y hat)",
          notes.append = TRUE,
          digits = 4)

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, mei 15, 2025 - 1:09:37

Table 19: OLS Regression with Robust Standard Errors

<i>Dependent variable:</i>	
	Weekly Wage
Education	26.6959*** (0.9189)
Age	2.2442** (0.9897)
Race	-77.6318*** (9.3927)
SMSA	-63.7562*** (6.2889)
Married	78.0013*** (7.9321)
Region 2	41.1931*** (12.5732)
Region 3	49.0879*** (10.7566)
Region 4	18.8453 (12.9527)
Region 5	4.2049 (10.8923)
Region 6	7.6044 (15.3882)
Region 7	17.0805 (12.0005)
Region 8	15.8968 (13.7257)
Region 9	63.7105*** (13.0328)
Constant	-81.6693 (49.8318)
Observations	10,000
R ²	0.1341
Adjusted R ²	0.1330
Residual Std. Error	275.0294 (df = 9986)
F Statistic	118.9531*** (df = 13; 9986)

Note:

*p<0.1; **p<0.05; ***p<0.01

Heteroskedasticity-consistent standard errors in parentheses

Table 20: EGLS Estimation Results

<i>Dependent variable:</i>	
	Transformed WAGE
Intercept	435.7491*** (51.0221)
EDUC	17.9831*** (0.7695)
AGE	-6.5454*** (1.0558)
RACE	-69.0106*** (7.8876)
SMSA	-78.0229*** (5.5790)
MARRIED	53.2469*** (5.7072)
REGION2	37.1832* (19.6988)
REGION3	60.2545*** (19.3016)
REGION4	25.2715 (18.7175)
REGION5	9.2088 (18.0923)
REGION6	12.2196 (18.5587)
REGION7	37.1148* (19.4049)
REGION8	40.5698* (24.2491)
REGION9	65.6674*** (21.2749)
Observations	10,000
R ²	0.6063
Adjusted R ²	0.6058
Residual Std. Error	1.0188 (df = 9986)
F Statistic	1,098.5050*** (df = 14; 9986)

Note: *p<0.1; **p<0.05; ***p<0.01
Variables transformed by dividing by y hat)

```

# Show table in basic format
modelsummary(
  list("Baseline (White SEs)" = linear_model1, "EGLS" = egls_model),
  coef_rename = c(
    "X_starEDUC" = "EDUC",
    "X_starAGE" = "AGE",
    "X_starRACE" = "RACE",
    "X_starMARRIED" = "MARRIED",
    "X_starSMSA" = "SMSA",
    "X_star(Intercept)" = "(Intercept)",
    "X_starREGION_factor2" = "REGION_factor2",
    "X_starREGION_factor3" = "REGION_factor3",
    "X_starREGION_factor4" = "REGION_factor4",
    "X_starREGION_factor5" = "REGION_factor5",
    "X_starREGION_factor6" = "REGION_factor6",
    "X_starREGION_factor7" = "REGION_factor7",
    "X_starREGION_factor8" = "REGION_factor8",
    "X_starREGION_factor9" = "REGION_factor9"
  ),
  vcov=list(vcovHC(linear_model1, type="HCO"),NULL),
  stars=TRUE
)

```

```

if (show_interpretation) {
  cat("\n##### **Comparison with Robust Standard Errors:**\n")
  cat("- EGLS provides more efficient estimates than OLS with robust standard errors\n")
  cat("- Standard errors are typically larger than OLS with incorrect variance formula\n")
  cat("- The efficiency gain comes from properly modeling the heteroskedasticity structure\n")
  cat("- Interpretation of coefficients remains the same as OLS\n")
}

```

```

correlation_matrix <- cor(df[, c("EDUC", "AGE", "RACE", "SMSA", "MARRIED",
                                "REGION2", "REGION3", "REGION4", "REGION5",
                                "REGION6", "REGION7", "REGION8", "REGION9")])

```

```

kable(correlation_matrix, format = "latex") %>%
  kable_styling(latex_options = c("striped", "hold_position"))

```

```

print(correlation_matrix)

```

	EDUC	AGE	RACE	SMSA	MARRIED
## EDUC	1.0000000000	-0.069506286	-0.151952924	-0.14907033	0.019198364
## AGE	-0.069506286	1.0000000000	-0.005148653	-0.02085355	0.020809977
## RACE	-0.151952924	-0.005148653	1.0000000000	-0.035563889	-0.113282921
## SMSA	-0.149070326	-0.020853549	-0.035563888	1.0000000000	0.040644736
## MARRIED	0.019198364	0.020809977	-0.113282921	0.04064474	1.0000000000
## REGION2	0.035717362	0.008497716	0.014092931	-0.10391873	-0.003692641
## REGION3	-0.030839760	0.005358420	-0.023908637	-0.02447606	0.013937881
## REGION4	-0.000578807	-0.014907585	-0.049908666	0.11787788	0.020881363

	Baseline (White SEs)	EGLS
(Intercept)	-81.669 (49.832)	435.749*** (51.022)
EDUC	26.696*** (0.919)	17.983*** (0.769)
AGE	2.244* (0.990)	-6.545*** (1.056)
RACE	-77.632*** (9.393)	-69.011*** (7.888)
SMSA	-63.756*** (6.289)	-78.023*** (5.579)
MARRIED	78.001*** (7.932)	53.247*** (5.707)
REGION2	41.193** (12.573)	
REGION3	49.088*** (10.757)	
REGION4	18.845 (12.953)	
REGION5	4.205 (10.892)	
REGION6	7.604 (15.388)	
REGION7	17.081 (12.001)	
REGION8	15.897 (13.726)	
REGION9	63.710*** (13.033)	
X_starREGION2		37.183+ (19.699)
X_starREGION3		60.255** (19.302)
X_starREGION4		25.272 (18.718)
X_starREGION5		9.209 (18.092)
X_starREGION6		12.220 (18.559)
X_starREGION7		37.115+ (19.405)
X_starREGION8		40.570+ (24.249)
X_starREGION9		65.667** (21.275)

	EDUC	AGE	RACE	SMSA	MARRIED	REGION2	REGION3	REGION4	I
EDUC	1.0000000	-0.0695063	-0.1519529	-0.1490703	0.0191984	0.0357174	-0.0308398	-0.0005788	-
AGE	-0.0695063	1.0000000	-0.0051487	-0.0208535	0.0208100	0.0084977	0.0053584	-0.0149076	-
RACE	-0.1519529	-0.0051487	1.0000000	-0.0355639	-0.1132829	0.0140929	-0.0239086	-0.0499087	-
SMSA	-0.1490703	-0.0208535	-0.0355639	1.0000000	0.0406447	-0.1039187	-0.0244761	0.1178779	-
MARRIED	0.0191984	0.0208100	-0.1132829	0.0406447	1.0000000	-0.0036926	0.0139379	0.0208814	-
REGION2	0.0357174	0.0084977	0.0140929	-0.1039187	-0.0036926	1.0000000	-0.2134547	-0.1219235	-
REGION3	-0.0308398	0.0053584	-0.0239086	-0.0244761	0.0139379	-0.2134547	1.0000000	-0.1382750	-
REGION4	-0.0005788	-0.0149076	-0.0499087	0.1178779	0.0208814	-0.1219235	-0.1382750	1.0000000	-
REGION5	-0.0429531	-0.0121791	0.0914728	0.0194065	0.0042555	-0.2013995	-0.2284097	-0.1304656	-
REGION6	-0.0844276	0.0083125	0.0395758	0.1134396	0.0177385	-0.1134428	-0.1286569	-0.0734877	-
REGION7	-0.0328062	0.0031565	0.0087286	-0.0107468	0.0071483	-0.1442097	-0.1635501	-0.0934184	-
REGION8	0.0284551	-0.0261790	-0.0436106	0.0711933	0.0209567	-0.0988984	-0.1121620	-0.0640659	-
REGION9	0.1046830	0.0014838	-0.0301200	-0.0774250	-0.0608085	-0.1665132	-0.1888447	-0.1078665	-

```

## REGION5 -0.042953090 -0.012179069 0.091472810 0.01940650 0.004255528
## REGION6 -0.084427585 0.008312525 0.039575771 0.11343959 0.017738535
## REGION7 -0.032806167 0.003156501 0.008728591 -0.01074680 0.007148294
## REGION8 0.028455078 -0.026179022 -0.043610641 0.07119329 0.020956712
## REGION9 0.104683025 0.001483814 -0.030119960 -0.07742498 -0.060808526
##          REGION2      REGION3      REGION4      REGION5      REGION6
## EDUC      0.035717362 -0.03083976 -0.000578807 -0.042953090 -0.084427585
## AGE       0.008497716 0.00535842 -0.014907585 -0.012179069 0.008312525
## RACE      0.014092931 -0.02390864 -0.049908666 0.091472810 0.039575771
## SMSA     -0.103918730 -0.02447606 0.117877877 0.019406502 0.113439595
## MARRIED   -0.003692641 0.01393788 0.020881363 0.004255528 0.017738535
## REGION2   1.000000000 -0.21345470 -0.121923453 -0.201399480 -0.113442761
## REGION3   -0.213454705 1.000000000 -0.138274958 -0.228409743 -0.128656896
## REGION4   -0.121923453 -0.13827496 1.000000000 -0.130465639 -0.073487689
## REGION5   -0.201399480 -0.22840974 -0.130465639 1.000000000 -0.121390774
## REGION6   -0.113442761 -0.12865690 -0.073487689 -0.121390774 1.000000000
## REGION7   -0.144209678 -0.16355005 -0.093418354 -0.154313279 -0.086920406
## REGION8   -0.098898407 -0.11216196 -0.064065924 -0.105827415 -0.059609658
## REGION9   -0.166513184 -0.18884475 -0.107866461 -0.178179412 -0.100363538
##          REGION7      REGION8      REGION9
## EDUC      -0.032806167 0.02845508 0.104683025
## AGE       0.003156501 -0.02617902 0.001483814
## RACE      0.008728591 -0.04361064 -0.030119960
## SMSA     -0.010746803 0.07119329 -0.077424983
## MARRIED   0.007148294 0.02095671 -0.060808526
## REGION2   -0.144209678 -0.09889841 -0.166513184
## REGION3   -0.163550053 -0.11216196 -0.188844746
## REGION4   -0.093418354 -0.06406592 -0.107866461
## REGION5   -0.154313279 -0.10582742 -0.178179412
## REGION6   -0.086920406 -0.05960966 -0.100363538
## REGION7   1.000000000 -0.07577645 -0.127583227
## REGION8   -0.075776449 1.000000000 -0.087496055
## REGION9   -0.127583227 -0.08749605 1.000000000

```

```
print(max(abs(correlation_matrix)))
```

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

```

print('VIF values')

## [1] "VIF values"

vif_values = vif(linear_model1)
print(vif_values)

##      EDUC      AGE      RACE      SMSA    MARRIED   REGION2   REGION3   REGION4
## 1.072047 1.008200 1.054291 1.072522 1.019728 3.280936 3.672041 2.182705
##  REGION5   REGION6   REGION7   REGION8   REGION9
## 3.510183 2.058605 2.540830 1.816650 2.919060

```

4(e) Autocorrelation

```

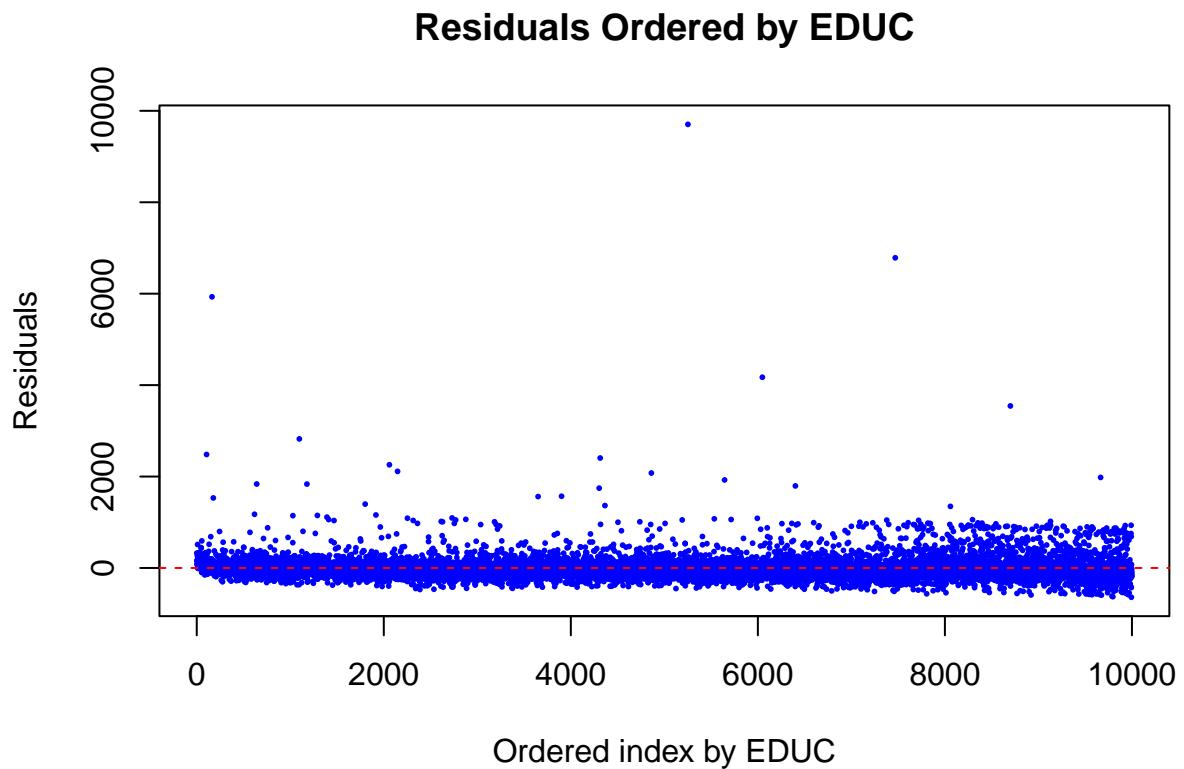
df_ordered_educ <- df %>% arrange(EDUC)

# Re-run baseline linear model on ordered data
linear_model_ordered <- lm(
  WAGE ~ EDUC + AGE + RACE + SMSA + MARRIED +
  REGION2 + REGION3 + REGION4 + REGION5 +
  REGION6 + REGION7 + REGION8 + REGION9,
  data = df_ordered_educ
)

# Store residuals
residuals_ordered <- residuals(linear_model_ordered)

plot(residuals_ordered,
  type = "p", # "p" for points instead of "l" for lines
  pch = 16, # Solid circle points (other options: 1-25)
  col = "blue", # Color of points
  cex = 0.4, # Size of points
  main = "Residuals Ordered by EDUC",
  ylab = "Residuals",
  xlab = "Ordered index by EDUC")
abline(h = 0, lty = 2, col = "red") # Reference line

```



Runs Test

```
data: residuals_ordered statistic = -2.6201, runs = 4870, n1 = 5000, n2 = 5000, n = 10000, p-value =
0.00879 alternative hypothesis: nonrandomness
=====
Test-statistic P-value ----- 1.957 0.015 -----
Test-statistic P-value ----- 13.468 0.036 -----
--
```

4(F). MODEL SPECIFICATION

```
if (show_interpretation) {
  cat("\nIn earlier tests we both saw indications of heteroskedasticity and autocorrelation, these two
}

=====
Dependent variable: WAGE
----- EDUC -23.274**
(9.053)
t = -2.571
p = 0.011
```

AGE -2.364*
(1.220)
t = -1.937
p = 0.053
RACE 44.302
(27.709)
t = 1.599
p = 0.110
SMSA 48.280**
(23.258)
t = 2.076
p = 0.038
MARRIED -60.243**
(27.647)
t = -2.179
p = 0.030
I(Yfit2) 0.003***
(0.001)
t = 3.419
p = 0.001
I(Yfit3) -0.00000*
(0.00000)
t = -1.742
p = 0.082
REGION2 -38.952**
(19.736)
t = -1.974
p = 0.049
REGION3 -43.479**
(21.499)
t = -2.022
p = 0.044
REGION4 -17.440
(16.867)
t = -1.034
p = 0.302
REGION5 -9.229
(13.555)
t = -0.681
p = 0.496
REGION6 -16.522
(16.341)
t = -1.011
p = 0.313
REGION7 -20.320
(15.930)
t = -1.276
p = 0.203
REGION8 -13.612
(17.950)
t = -0.758
p = 0.449
REGION9 -61.727**
(26.282)

t = -2.349
p = 0.019
Constant 410.901***
(82.840)
t = 4.960
p = 0.00000

----- Observations 10,000

R2 0.143

Adjusted R2 0.142

Residual Std. Error 273.622 (df = 9984)

F Statistic 111.158*** (df = 15; 9984) (p = 0.000) =====

Note: p<0.1; p<0.05; p<0.01

===== Test-statistic P-value ----- 52.514 0

RESET test

data: linear_model1 RESET = 52.514, df1 = 2, df2 = 9984, p-value < 2.2e-16

===== Dependent variable:

res	EDUC
(12.150)	-78.358***
t = -6.449	
p = 0.000	
I(EDUC2) 5.689***	
(1.086)	
t = 5.237	
p = 0.00000	
I(EDUC3) -0.127***	
(0.030)	
t = -4.160	
p = 0.00004	
AGE -293.290	
(795.744)	
t = -0.369	
p = 0.713	
I(AGE2) 6.035	
(17.811)	
t = 0.339	
p = 0.735	
I(AGE3) -0.041	
(0.133)	
t = -0.309	
p = 0.758	
RACE -3.948	
(10.202)	
t = -0.387	
p = 0.699	
SMSA 0.600	
(7.374)	
t = 0.081	
p = 0.936	

MARRIED 2.297

(8.001)

t = 0.287

p = 0.775

REGION2 -0.676

(13.595)

t = -0.050

p = 0.961

REGION3 0.602

(13.257)

t = 0.045

p = 0.964

REGION4 -0.770

(15.544)

t = -0.050

p = 0.961

REGION5 -3.922

(13.451)

t = -0.292

p = 0.771

REGION6 -5.673

(16.087)

t = -0.353

p = 0.725

REGION7 -6.282

(14.614)

t = -0.430

p = 0.668

REGION8 -1.142

(17.051)

t = -0.067

p = 0.947

REGION9 -1.657

(14.006)

t = -0.118

p = 0.906

Constant 5,038.892

(11,824.640)

t = 0.426

p = 0.671

Observations 10,000

R2 0.008

Adjusted R2 0.006

Residual Std. Error 274.028 (df = 9982)

F Statistic 4.538*** (df = 17; 9982) (p = 0.000) =====

Note: *p*<0.1; ***p*<0.05**; *p*<0.01

===== Test-statistic P-value ----- 76.692 0

===== Dependent variable:

===== res

===== EDUC -78.358***

(12.150)

t = -6.449
p = 0.000
I(EDUC2) 5.689***
(1.086)
t = 5.237
p = 0.00000
I(EDUC3) -0.127***
(0.030)
t = -4.160
p = 0.00004
AGE -293.290
(795.744)
t = -0.369
p = 0.713
I(AGE2) 6.035
(17.811)
t = 0.339
p = 0.735
I(AGE3) -0.041
(0.133)
t = -0.309
p = 0.758
RACE -3.948
(10.202)
t = -0.387
p = 0.699
SMSA 0.600
(7.374)
t = 0.081
p = 0.936
MARRIED 2.297
(8.001)
t = 0.287
p = 0.775
REGION2 -0.676
(13.595)
t = -0.050
p = 0.961
REGION3 0.602
(13.257)
t = 0.045
p = 0.964
REGION4 -0.770
(15.544)
t = -0.050
p = 0.961
REGION5 -3.922
(13.451)
t = -0.292
p = 0.771
REGION6 -5.673
(16.087)
t = -0.353
p = 0.725

REGION7 -6.282

(14.614)

t = -0.430

p = 0.668

REGION8 -1.142

(17.051)

t = -0.067

p = 0.947

REGION9 -1.657

(14.006)

t = -0.118

p = 0.906

Constant 5,038.892

(11,824.640)

t = 0.426

p = 0.671

Observations 10,000

R2 0.008

Adjusted R2 0.006

Residual Std. Error 274.028 (df = 9982)

F Statistic 4.538*** (df = 17; 9982) (p = 0.000) =====

Note: *p*<0.1; ***p*<0.05**; *p*<0.01

===== Test-statistic P-value ----- 74.385 0

===== Dependent variable:

----- res

----- EDUC -78.358***

(12.150)

t = -6.449

p = 0.000

I(EDUC2) 5.689***

(1.086)

t = 5.237

p = 0.00000

I(EDUC3) -0.127***

(0.030)

t = -4.160

p = 0.00004

AGE -293.290

(795.744)

t = -0.369

p = 0.713

I(AGE2) 6.035

(17.811)

t = 0.339

p = 0.735

I(AGE3) -0.041

(0.133)

t = -0.309

p = 0.758

RACE -3.948

(10.202)

t = -0.387
 p = 0.699
 SMSA 0.600
 (7.374)
 t = 0.081
 p = 0.936
 MARRIED 2.297
 (8.001)
 t = 0.287
 p = 0.775
 REGION2 -0.676
 (13.595)
 t = -0.050
 p = 0.961
 REGION3 0.602
 (13.257)
 t = 0.045
 p = 0.964
 REGION4 -0.770
 (15.544)
 t = -0.050
 p = 0.961
 REGION5 -3.922
 (13.451)
 t = -0.292
 p = 0.771
 REGION6 -5.673
 (16.087)
 t = -0.353
 p = 0.725
 REGION7 -6.282
 (14.614)
 t = -0.430
 p = 0.668
 REGION8 -1.142
 (17.051)
 t = -0.067
 p = 0.947
 REGION9 -1.657
 (14.006)
 t = -0.118
 p = 0.906
 Constant 5,038.892
 (11,824.640)
 t = 0.426
 p = 0.671

Observations 10,000

R2 0.008

Adjusted R2 0.006

Residual Std. Error 274.028 (df = 9982)

F Statistic 4.538*** (df = 17; 9982) (p = 0.000) =====

Note: $p < 0.1$; $p < 0.05$; $p < 0.01$

Test-statistic P-value ----- 2.479 0.290 -----

Baseline vs Log-Linear Model with Robust SEs

Dependent variable:	
	WAGE
	Level
	(1)
	(2)
EDUC	26.696*** 0.061 (0.920) (0.002) AGE 2.244 0.005 (0.990) (0.002) RACE -77.632 -0.270 (9.399) (0.027) SMSA -63.756 -0.182 (6.293) (0.018) MARRIED 78.001 0.259 (7.938) (0.021) REGION2 41.193 0.102 (12.582) (0.031) REGION3 49.088 0.139 (10.764) (0.031) REGION4 18.845 0.066 (12.962) (0.035) REGION5 4.205 -0.016 (10.900) (0.032) REGION6 7.604 -0.037 (15.399) (0.038) REGION7 17.081 0.021 (12.009) (0.034) REGION8 15.897 0.072 (13.735) (0.038) REGION9 63.710 0.147 (13.042) (0.032) Constant -81.669 4.657 (49.867) (0.108)
Observations	10,000 10,000
R2	0.134 0.166
Adjusted R2	0.133 0.165
Residual Std. Error	(df = 9986) 275.029 0.622
F Statistic	(df = 13; 9986) 118.953*** 152.546
<hr/>	
Note: p<0.1; p<0.05; ***p<0.01	
Baseline vs Log-Linear Model with Robust SEs ===== TRUE	

```
df      AIC
linear_model1 15 140732.32 log_model 15 18897.87 df BIC linear_model1 15 140840.48 log_model 15
19006.02 Linear model r-squared:[1] 0.134091 Log model r-squared:[1] 0.1656844
```

RESET test

data: log_model RESET = 3.482, df1 = 2, df2 = 9984, p-value = 0.03078

Reset test still shows misspecification but model has improved compared to baseline.

```
educ_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
+ REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(EDUC^2), data = df)
anova(linear_model1, educ_reg)
```

```
## Analysis of Variance Table
##
## Model 1: WAGE ~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
## Model 2: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(EDUC^2)
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1   9986  755352847
## 2   9985  751033088  1   4319759 57.431 3.811e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
stargazer(educ_reg,type="text")
```

```
##  
## -----  
##          Dependent variable:  
## -----  
##          WAGE  
## -----  
## EDUC           -4.259  
##                  (4.176)  
##  
## AGE            1.966**  
##                  (0.940)  
##  
## RACE           -81.460***  
##                  (10.209)  
##  
## MARRIED        79.888***  
##                  (8.007)  
##  
## SMSA           -64.560***  
##                  (7.373)  
##  
## REGION2        40.801***  
##                  (13.606)  
##  
## REGION3        49.505***  
##                  (13.267)  
##  
## REGION4        18.609  
##                  (15.556)  
##  
## REGION5        1.490  
##                  (13.459)  
##  
## REGION6        1.554  
##                  (16.097)  
##  
## REGION7        13.477  
##                  (14.612)  
##  
## REGION8        16.740  
##                  (17.058)  
##  
## REGION9        64.151***  
##                  (14.007)  
##  
## I(EDUC2)       1.222***  
##                  (0.161)  
##  
## Constant       113.617**  
##                  (53.148)  
##
```

```

## -----
## Observations          10,000
## R2                  0.139
## Adjusted R2          0.138
## Residual Std. Error  274.256 (df = 9985)
## F Statistic          115.183*** (df = 14; 9985)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01

educpolythree_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
                         + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(EDUC^2)+I(EDUC^3), data = df)
anova(educ_reg, educpolythree_reg)

## Analysis of Variance Table
##
## Model 1: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(EDUC^2)
## Model 2: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(EDUC^2) + I(EDUC^3)
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1  9985  751033088
## 2  9984  749734167  1   1298920 17.297 3.223e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#educpolyfour_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
#                         + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(EDUC^2)+I(EDUC^3)+I(EDUC^4), data = df)
#anova(educpolythree_reg, educpolyfour_reg)
#stargazer(educpolyfour_reg, type="text")
#
#educpolyfive_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
#                         + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(EDUC^2)+I(EDUC^3)+I(EDUC^4)+I(EDUC^5), data = df)
#anova(educpolyfour_reg, educpolyfive_reg)
#stargazer(educpolyfive_reg, type="text")

age_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
               + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(AGE^2), data = df)
anova(linear_model1, age_reg)

## Analysis of Variance Table
##
## Model 1: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
## Model 2: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(AGE^2)
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1  9986  755352847
## 2  9985  755171041  1   181806 2.4039 0.1211

```

```

ageinteraction_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
                         + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(AGE*EDUC), data = df)
anova(linear_model1,ageinteraction_reg)

## Analysis of Variance Table
##
## Model 1: WAGE ~ EDUC + AGE + RACE + MARRIED + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
## Model 2: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(AGE * EDUC)
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1    9986 755352847
## 2    9985 755352738  1     109.36 0.0014 0.9697

raceinteraction_reg <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3
                           + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 + I(RACE*EDUC), data = df)
anova(linear_model1,raceinteraction_reg)

## Analysis of Variance Table
##
## Model 1: WAGE ~ EDUC + AGE + RACE + MARRIED + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9
## Model 2: WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 +
##           REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 +
##           I(RACE * EDUC)
##   Res.Df      RSS Df Sum of Sq      F     Pr(>F)
## 1    9986 755352847
## 2    9985 752115125  1    3237722 42.984 5.793e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

===== Test-statistic P-value ----- 864.620 0.999 -----
-- 

RESET test

data: best_mod RESET = 2.445, df1 = 2, df2 = 9981, p-value = 0.08678

## Loading required package: survival

##
## Call:
## ivreg(formula = WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + factor(REGION) |
##        QOB + AGE + RACE + MARRIED + SMSA + factor(REGION), data = df)
##
## Residuals:
##     Min      1Q      Median      3Q      Max 
## -620.56 -132.12   -37.68    73.95  9700.42 
##
## Coefficients:
```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24.421    702.921 -0.035  0.97229
## EDUC         23.287     41.777  0.557  0.57727
## AGE          1.966      3.533  0.556  0.57790
## RACE        -83.598    73.807 -1.133  0.25739
## MARRIED      78.514    10.195  7.701 1.48e-14 ***
## SMSA         -67.962    52.056 -1.306  0.19174
## factor(REGION)2 41.131    13.676  3.007  0.00264 **
## factor(REGION)3 47.565    22.917  2.076  0.03796 *
## factor(REGION)4 18.488    16.215  1.140  0.25426
## factor(REGION)5  2.875    21.159  0.136  0.89191
## factor(REGION)6  4.298    43.605  0.099  0.92149
## factor(REGION)7 15.307    26.206  0.584  0.55916
## factor(REGION)8 16.707    19.790  0.844  0.39856
## factor(REGION)9 65.668    27.795  2.363  0.01817 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 275.2 on 9986 degrees of freedom
## Multiple R-Squared: 0.1328, Adjusted R-squared: 0.1316
## Wald test: 46.29 on 13 and 9986 DF, p-value: < 2.2e-16

##
## Call:
## lm(formula = EDUC ~ QOB + AGE + RACE + MARRIED + SMSA + factor(REGION),
##      data = df)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -13.9228 -1.5791 -0.4719  2.1043  9.4704 
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.51952    0.52632 31.387 < 2e-16 ***
## QOB          0.05952    0.02861  2.080 0.037510 * 
## AGE          -0.07883   0.01090 -7.235 4.98e-13 ***
## RACE         -1.74628   0.11644 -14.998 < 2e-16 ***
## MARRIED      0.15138    0.09240  1.638 0.101382  
## SMSA         -1.23208   0.08423 -14.628 < 2e-16 ***
## factor(REGION)2 -0.01863   0.15710 -0.119 0.905587  
## factor(REGION)3 -0.44419   0.15313 -2.901 0.003731 ** 
## factor(REGION)4 -0.09780   0.17965 -0.544 0.586198  
## factor(REGION)5 -0.38929   0.15530 -2.507 0.012202 * 
## factor(REGION)6 -0.96658   0.18540 -5.214 1.89e-07 ***
## factor(REGION)7 -0.52131   0.16856 -3.093 0.001988 ** 
## factor(REGION)8  0.24120   0.19696  1.225 0.220746  
## factor(REGION)9  0.57584   0.16163  3.563 0.000369 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.167 on 9986 degrees of freedom
## Multiple R-squared: 0.06761, Adjusted R-squared: 0.0664
## F-statistic: 55.7 on 13 and 9986 DF, p-value: < 2.2e-16

```

```

## 
## Call:
## lm(formula = WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + factor(REGION) +
##     first_stage_residuals, data = df)
## 
## Residuals:
##    Min      1Q Median      3Q     Max 
## -639.9 -131.9 -35.8   76.0 9703.9 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -24.421    702.414 -0.035  0.97227    
## EDUC          23.287    41.747  0.558  0.57699    
## AGE           1.966    3.531  0.557  0.57763    
## RACE          -83.598   73.754 -1.133  0.25704    
## MARRIED       78.514   10.188  7.707 1.41e-14 ***  
## SMSA          -67.962   52.019 -1.306  0.19142    
## factor(REGION)2 41.131   13.666  3.010  0.00262 **  
## factor(REGION)3 47.565   22.901  2.077  0.03782 *   
## factor(REGION)4 18.488   16.204  1.141  0.25392    
## factor(REGION)5  2.875   21.143  0.136  0.89184    
## factor(REGION)6  4.298   43.574  0.099  0.92143    
## factor(REGION)7 15.307   26.187  0.585  0.55888    
## factor(REGION)8 16.707   19.775  0.845  0.39822    
## factor(REGION)9 65.668   27.775  2.364  0.01808 *  
## first_stage_residuals 3.411   41.756  0.082  0.93490  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 275 on 9985 degrees of freedom
## Multiple R-squared:  0.1341, Adjusted R-squared:  0.1329 
## F-statistic: 110.4 on 14 and 9985 DF,  p-value: < 2.2e-16

```

5. GENERAL REMEDIATION AND CONCLUSION

```

lower_bound <- quantile(df$WAGE, 0.005, na.rm = TRUE)
upper_bound <- quantile(df$WAGE, 0.995, na.rm = TRUE)
dataset_trimmed <- df %>% filter(WAGE >= lower_bound, WAGE <= upper_bound)

reg_final <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 + REGION4 + REGION5 + REGI

reg_egls_final <- gls(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 + REGION4 + REGION5 + REGI
stargazer(reg_egls_final, type = "text", digits = 4, style = "all")

## 
## =====
##                               Dependent variable:
##                               -----
##                               WAGE
## -----
## EDUC                      39.0205***
```

```

##                               (11.4756)
##                               t = 3.4003
##                               p = 0.0007
## AGE                           2.2541***  

##                                         (0.6555)
##                                         t = 3.4388
##                                         p = 0.0006
## RACE                          -3.6910  

##                                         (17.6096)
##                                         t = -0.2096
##                                         p = 0.8340
## MARRIED                      65.8495***  

##                                         (4.9193)
##                                         t = 13.3858
##                                         p = 0.0000
## SMSA                           -49.2194***  

##                                         (4.6154)
##                                         t = -10.6641
##                                         p = 0.0000
## REGION2                      32.4728***  

##                                         (9.5043)
##                                         t = 3.4167
##                                         p = 0.0007
## REGION3                      49.9037***  

##                                         (9.2812)
##                                         t = 5.3768
##                                         p = 0.000000
## REGION4                      21.1348**  

##                                         (10.7030)
##                                         t = 1.9747
##                                         p = 0.0484
## REGION5                      0.2373  

##                                         (9.1248)
##                                         t = 0.0260
##                                         p = 0.9793
## REGION6                      -11.1097  

##                                         (10.3433)
##                                         t = -1.0741
##                                         p = 0.2828
## REGION7                      11.5234  

##                                         (9.9738)
##                                         t = 1.1554
##                                         p = 0.2480
## REGION8                      29.2010**  

##                                         (12.0969)
##                                         t = 2.4139
##                                         p = 0.0158
## REGION9                      56.2612***  

##                                         (10.1235)
##                                         t = 5.5575
##                                         p = 0.000000
## I(EDUC2)                     -7.7340***  

##                                         (2.0267)
##                                         t = -3.8160

```

```

##          p = 0.0002
## I(EDUC3)      0.6891***  

##                (0.1425)
##          t = 4.8361
##          p = 0.000002
## I(EDUC4)      -0.0174***  

##                (0.0034)
##          t = -5.1043
##          p = 0.000001
## I(RACE * EDUC) -6.8060***  

##                (1.6348)
##          t = -4.1631
##          p = 0.00004
## Constant       43.2820  

##                (38.9031)
##          t = 1.1126
##          p = 0.2659
## -----
## Observations      9,900
## Log Likelihood   -66,492.7100
## Akaike Inf. Crit. 133,025.4000
## Bayesian Inf. Crit. 133,169.4000
## -----
## Note:           *p<0.1; **p<0.05; ***p<0.01

```