Case Solution Group 5

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

2025-06-09

Step 1: Specification, hypotheses, and descriptive statistics

```
head(df)
```

```
##       WAGE EDUC AGE RACE SMSA MARRIED REGION QOB REGION1 REGION2 REGION3
## 1 580.1000    9  45    0    0       1      9   3       0       0       0
## 2 642.2115   17  47    0    0       1      3   4       0       0       1
## 3 577.0192   12  42    0    0       1      7   2       0       0       0
## 4 999.1346   10  43    0    0       1      3   2       0       0       1
## 5 307.7885   12  41    0    0       1      6   3       0       0       0
## 6 280.1000   12  40    1    0       1      5   2       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    1    0    0
```

```
summary(df)
```

```
##       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
```

```
describe(df)
```

```
##          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
```

```
round(cov(numeric_data), 3)
```

```
##              WAGE    EDUC    AGE    RACE    SMSA MARRIED REGION
## WAGE    87241.084 315.768  5.257 -10.490 -14.858  10.178  8.855
## EDUC      315.768  10.742 -0.668  -0.138  -0.188   0.022  0.274
## AGE         5.257  -0.668  8.592  -0.004  -0.024   0.021 -0.125
## RACE      -10.490  -0.138 -0.004   0.076  -0.004  -0.011  0.001
## SMSA      -14.858  -0.188 -0.024  -0.004   0.148   0.005  0.033
## MARRIED    10.178   0.022  0.021  -0.011   0.005   0.120 -0.022
## REGION      8.855   0.274 -0.125   0.001   0.033  -0.022  6.064
```

```
round(cor(numeric_data), 3)
```

```
##           WAGE   EDUC    AGE   RACE   SMSA MARRIED REGION
## WAGE     1.000  0.326  0.006 -0.129 -0.131   0.100  0.012
## EDUC     0.326  1.000 -0.070 -0.152 -0.149   0.019  0.034
## AGE      0.006 -0.070  1.000 -0.005 -0.021   0.021 -0.017
## RACE    -0.129 -0.152 -0.005  1.000 -0.036  -0.113  0.002
## SMSA    -0.131 -0.149 -0.021 -0.036  1.000   0.041  0.034
## MARRIED  0.100  0.019  0.021 -0.113  0.041   1.000 -0.026
## REGION   0.012  0.034 -0.017  0.002  0.034  -0.026  1.000
```

```
## # 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    1.0    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    1.0    1.00  5.40
## SMSA       6 549   0.15   0.36   0.00    0.06   0.00  0.00    1.0    1.00  1.96
## MARRIED    7 549   0.85   0.36   1.00    0.93   0.00  0.00    1.0    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    3.0    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   12.0   12.43   1.48  0.00   20.00   20.00
## AGE        4 1949  44.71   2.89   45.0   44.70   2.97 40.00   50.00   10.00
## RACE       5 1949   0.07   0.25    0.0    0.00   0.00  0.00    1.00    1.00
## SMSA       6 1949   0.16   0.37    0.0    0.08   0.00  0.00    1.00    1.00
## MARRIED    7 1949   0.87   0.34    1.0    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    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      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
```

[[1]]

Table 1: REGION 1

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	59622.389	326.291	19.889	-4.301	-19.309	8.220
EDUC	326.291	9.617	-0.350	-0.024	-0.152	-0.054
AGE	19.889	-0.350	8.596	0.001	-0.054	0.074
RACE	-4.301	-0.024	0.001	0.030	-0.005	-0.004
SMSA	-19.309	-0.152	-0.054	-0.005	0.127	-0.005
MARRIED	8.220	-0.054	0.074	-0.004	-0.005	0.129

[[2]]

Table 2: REGION 2

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	126020.884	335.536	29.869	-13.086	-6.225	16.273
EDUC	335.536	10.031	-0.673	-0.158	-0.088	0.091
AGE	29.869	-0.673	8.753	-0.016	0.009	-0.012

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
RACE	-13.086	-0.158	-0.016	0.084	-0.007	-0.014
SMSA	-6.225	-0.088	0.009	-0.007	0.081	0.001
MARRIED	16.273	0.091	-0.012	-0.014	0.001	0.122

[[3]]

Table 3: REGION 3

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	69754.484	247.163	-9.152	-6.535	-13.792	10.302
EDUC	247.163	9.398	-0.690	-0.083	-0.106	-0.007
AGE	-9.152	-0.690	8.344	-0.021	-0.035	-0.018
RACE	-6.535	-0.083	-0.021	0.065	-0.011	-0.006
SMSA	-13.792	-0.106	-0.035	-0.011	0.136	0.006
MARRIED	10.302	-0.007	-0.018	-0.006	0.006	0.113

[[4]]

Table 4: REGION 4

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	69294.730	276.252	-25.048	-4.650	-20.462	12.774
EDUC	276.252	9.903	-0.788	-0.043	-0.246	0.082
AGE	-25.048	-0.788	8.668	0.008	0.015	0.020
RACE	-4.650	-0.043	0.008	0.033	-0.005	-0.004
SMSA	-20.462	-0.246	0.015	-0.005	0.226	0.005
MARRIED	12.774	0.082	0.020	-0.004	0.005	0.101

[[5]]

Table 5: REGION 5

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	68142.878	359.308	-13.139	-15.912	-13.042	10.883
EDUC	359.308	12.144	-0.823	-0.232	-0.257	0.040
AGE	-13.139	-0.823	8.956	0.017	-0.038	0.053
RACE	-15.912	-0.232	0.017	0.119	0.001	-0.019
SMSA	-13.042	-0.257	-0.038	0.001	0.159	0.003
MARRIED	10.883	0.040	0.053	-0.019	0.003	0.118

[[6]]

Table 6: REGION 6

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	108586.360	289.292	-38.510	-10.947	-20.349	3.987
EDUC	289.292	14.004	-1.028	-0.234	-0.204	0.117

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
AGE	-38.510	-1.028	7.682	0.018	0.036	-0.011
RACE	-10.947	-0.234	0.018	0.110	0.006	-0.022
SMSA	-20.349	-0.204	0.036	0.006	0.227	0.011
MARRIED	3.987	0.117	-0.011	-0.022	0.011	0.102

[[7]]

Table 7: REGION 7

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	70637.866	357.447	26.277	-11.392	-12.199	11.726
EDUC	357.447	12.717	-0.614	-0.183	-0.219	0.008
AGE	26.277	-0.614	8.851	0.008	-0.026	0.055
RACE	-11.392	-0.183	0.008	0.082	0.004	-0.010
SMSA	-12.199	-0.219	-0.026	0.004	0.140	0.004
MARRIED	11.726	0.008	0.055	-0.010	0.004	0.114

[[8]]

Table 8: REGION 8

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	55763.738	204.967	8.289	-5.134	-19.118	0.818
EDUC	204.967	8.725	-0.586	-0.044	-0.214	0.034
AGE	8.289	-0.586	8.700	-0.033	-0.049	0.027
RACE	-5.134	-0.044	-0.033	0.030	-0.005	-0.001
SMSA	-19.118	-0.214	-0.049	-0.005	0.211	0.008
MARRIED	0.818	0.034	0.027	-0.001	0.008	0.096

[[9]]

Table 9: REGION 9

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	119854.165	270.881	33.568	-8.149	-7.110	10.765
EDUC	270.881	8.544	-0.359	-0.046	-0.113	-0.022
AGE	33.568	-0.359	8.370	-0.012	-0.016	0.051
RACE	-8.149	-0.046	-0.012	0.058	-0.005	-0.009
SMSA	-7.110	-0.113	-0.016	-0.005	0.093	0.005
MARRIED	10.765	-0.022	0.051	-0.009	0.005	0.156

[[1]]

Table 10: REGION 1

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.431	0.028	-0.102	-0.222	0.094

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
EDUC	0.431	1.000	-0.038	-0.045	-0.137	-0.049
AGE	0.028	-0.038	1.000	0.002	-0.052	0.071
RACE	-0.102	-0.045	0.002	1.000	-0.075	-0.071
SMSA	-0.222	-0.137	-0.052	-0.075	1.000	-0.037
MARRIED	0.094	-0.049	0.071	-0.071	-0.037	1.000

[[2]]

Table 11: REGION 2

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.298	0.028	-0.127	-0.062	0.131
EDUC	0.298	1.000	-0.072	-0.173	-0.098	0.082
AGE	0.028	-0.072	1.000	-0.019	0.011	-0.011
RACE	-0.127	-0.173	-0.019	1.000	-0.084	-0.139
SMSA	-0.062	-0.098	0.011	-0.084	1.000	0.007
MARRIED	0.131	0.082	-0.011	-0.139	0.007	1.000

[[3]]

Table 12: REGION 3

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.305	-0.012	-0.097	-0.142	0.116
EDUC	0.305	1.000	-0.078	-0.107	-0.093	-0.007
AGE	-0.012	-0.078	1.000	-0.029	-0.033	-0.019
RACE	-0.097	-0.107	-0.029	1.000	-0.115	-0.075
SMSA	-0.142	-0.093	-0.033	-0.115	1.000	0.049
MARRIED	0.116	-0.007	-0.019	-0.075	0.049	1.000

[[4]]

Table 13: REGION 4

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.333	-0.032	-0.097	-0.164	0.153
EDUC	0.333	1.000	-0.085	-0.076	-0.164	0.082
AGE	-0.032	-0.085	1.000	0.015	0.010	0.021
RACE	-0.097	-0.076	0.015	1.000	-0.057	-0.075
SMSA	-0.164	-0.164	0.010	-0.057	1.000	0.031
MARRIED	0.153	0.082	0.021	-0.075	0.031	1.000

[[5]]

Table 14: REGION 5

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.395	-0.017	-0.177	-0.125	0.122
EDUC	0.395	1.000	-0.079	-0.193	-0.185	0.034
AGE	-0.017	-0.079	1.000	0.017	-0.032	0.052
RACE	-0.177	-0.193	0.017	1.000	0.008	-0.157
SMSA	-0.125	-0.185	-0.032	0.008	1.000	0.023
MARRIED	0.122	0.034	0.052	-0.157	0.023	1.000

[[6]]

Table 15: REGION 6

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.235	-0.042	-0.100	-0.130	0.038
EDUC	0.235	1.000	-0.099	-0.189	-0.114	0.097
AGE	-0.042	-0.099	1.000	0.019	0.027	-0.012
RACE	-0.100	-0.189	0.019	1.000	0.041	-0.203
SMSA	-0.130	-0.114	0.027	0.041	1.000	0.070
MARRIED	0.038	0.097	-0.012	-0.203	0.070	1.000

[[7]]

Table 16: REGION 7

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.377	0.033	-0.149	-0.122	0.130
EDUC	0.377	1.000	-0.058	-0.179	-0.164	0.007
AGE	0.033	-0.058	1.000	0.010	-0.023	0.055
RACE	-0.149	-0.179	0.010	1.000	0.036	-0.105
SMSA	-0.122	-0.164	-0.023	0.036	1.000	0.033
MARRIED	0.130	0.007	0.055	-0.105	0.033	1.000

[[8]]

Table 17: REGION 8

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.294	0.012	-0.127	-0.176	0.011
EDUC	0.294	1.000	-0.067	-0.087	-0.158	0.037
AGE	0.012	-0.067	1.000	-0.065	-0.036	0.029
RACE	-0.127	-0.087	-0.065	1.000	-0.065	-0.015
SMSA	-0.176	-0.158	-0.036	-0.065	1.000	0.057
MARRIED	0.011	0.037	0.029	-0.015	0.057	1.000

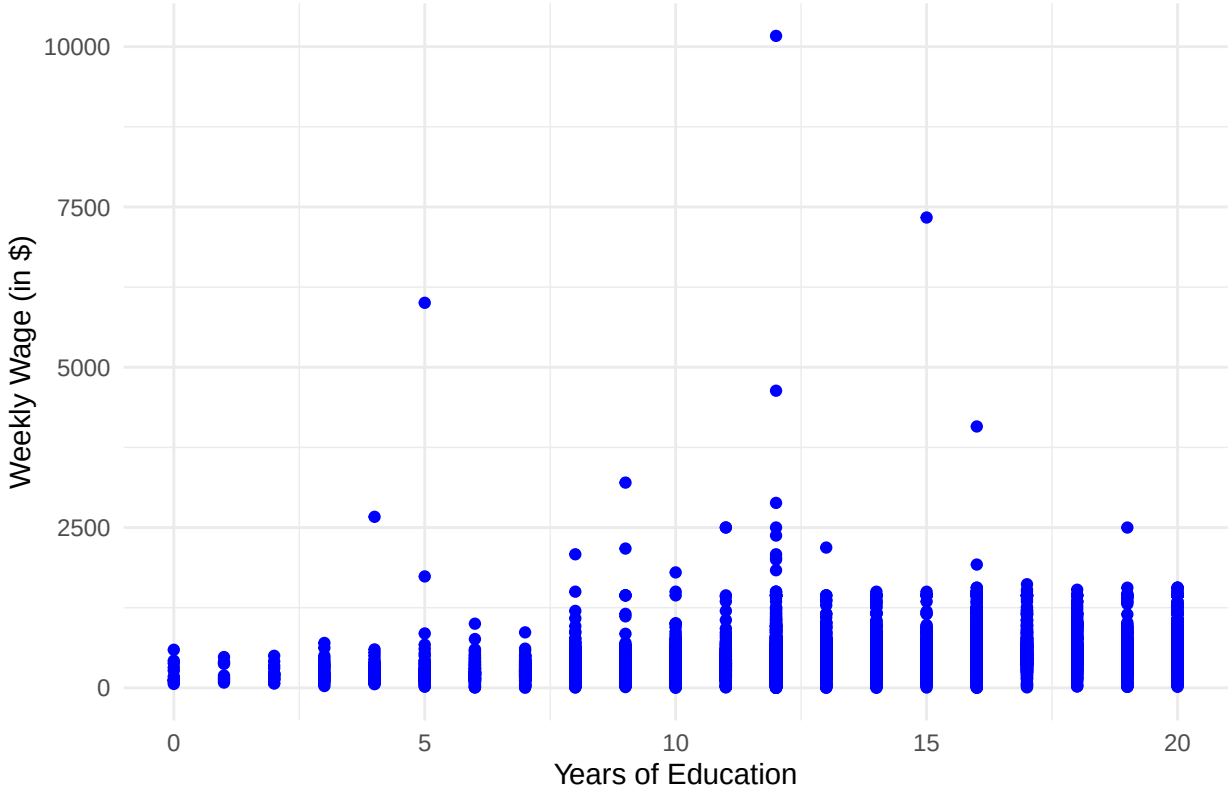
[[9]]

Table 18: REGION 9

	WAGE	EDUC	AGE	RACE	SMSA	MARRIED
WAGE	1.000	0.268	0.034	-0.098	-0.067	0.079
EDUC	0.268	1.000	-0.042	-0.066	-0.127	-0.019
AGE	0.034	-0.042	1.000	-0.017	-0.018	0.044
RACE	-0.098	-0.066	-0.017	1.000	-0.066	-0.096
SMSA	-0.067	-0.127	-0.018	-0.066	1.000	0.044
MARRIED	0.079	-0.019	0.044	-0.096	0.044	1.000

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

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



Step 2: OLS estimation and testing

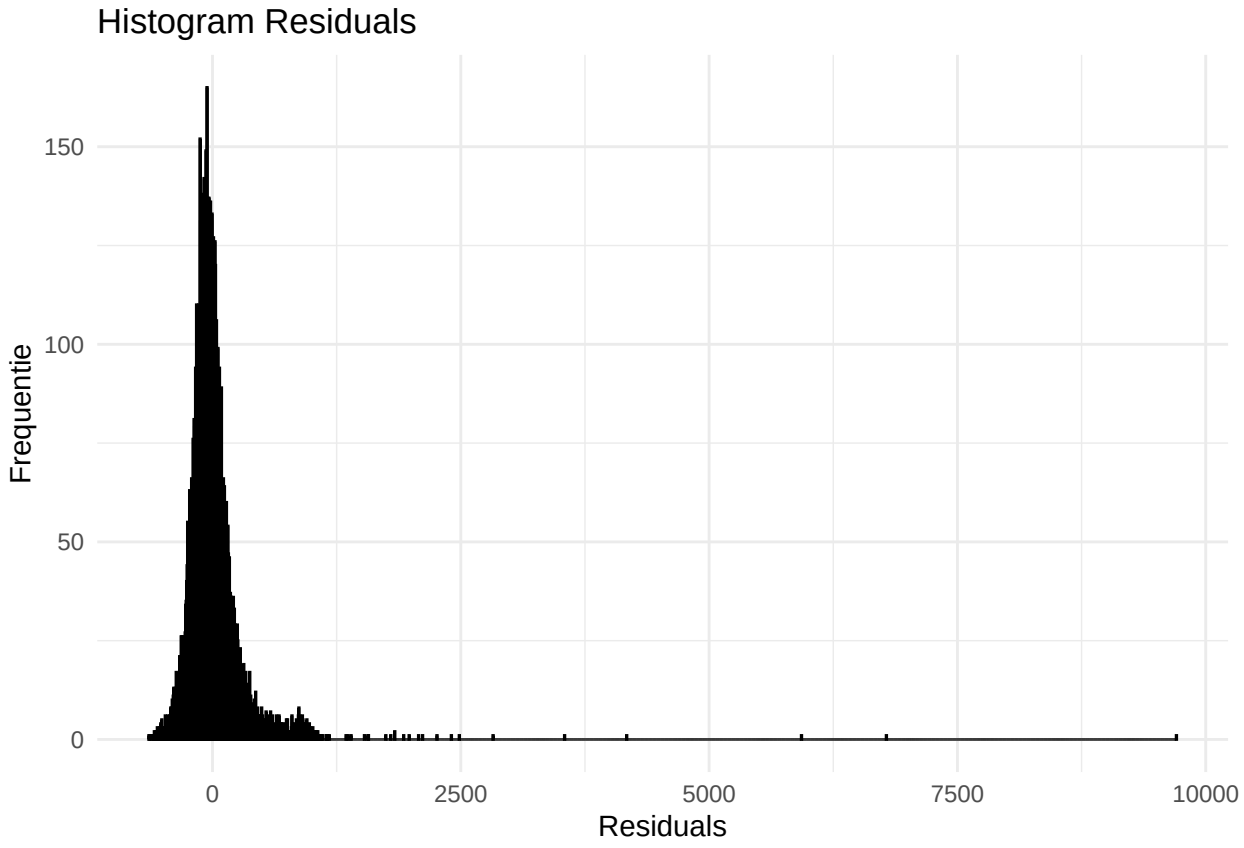
```
linear_model1=lm(WAGE~ EDUC + AGE + RACE + SMSA + MARRIED + REGION2 + REGION3
                 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9  , data = df)
```

```
##
## ===============================================================
##                                 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
```

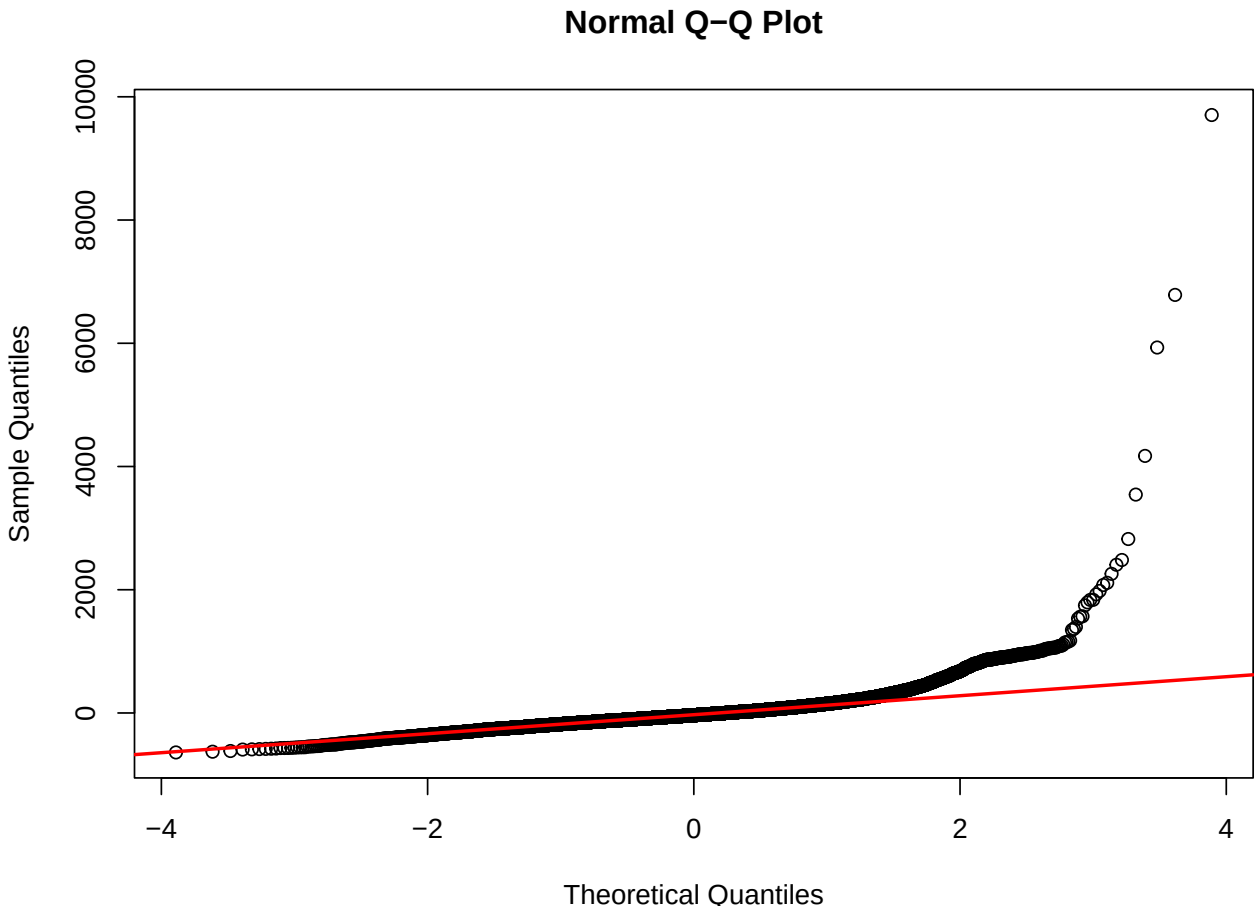
Step 3: Checking GM assumptions and individual remedies



[1] -4.605628e-16

```
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

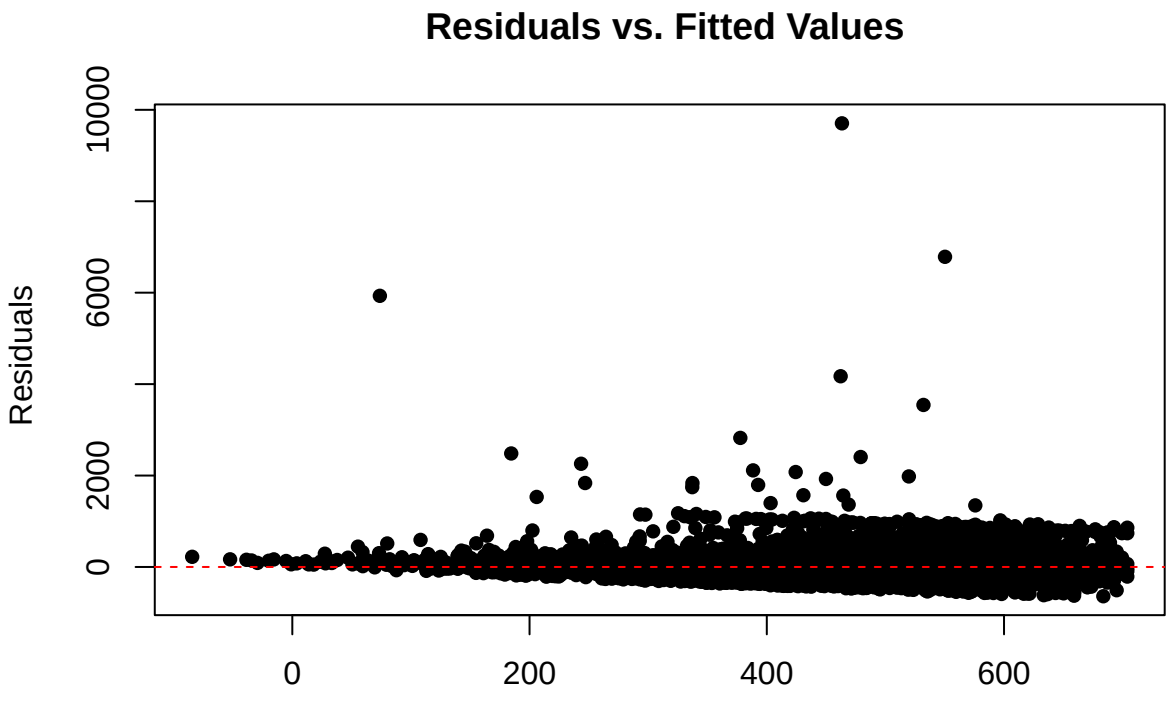


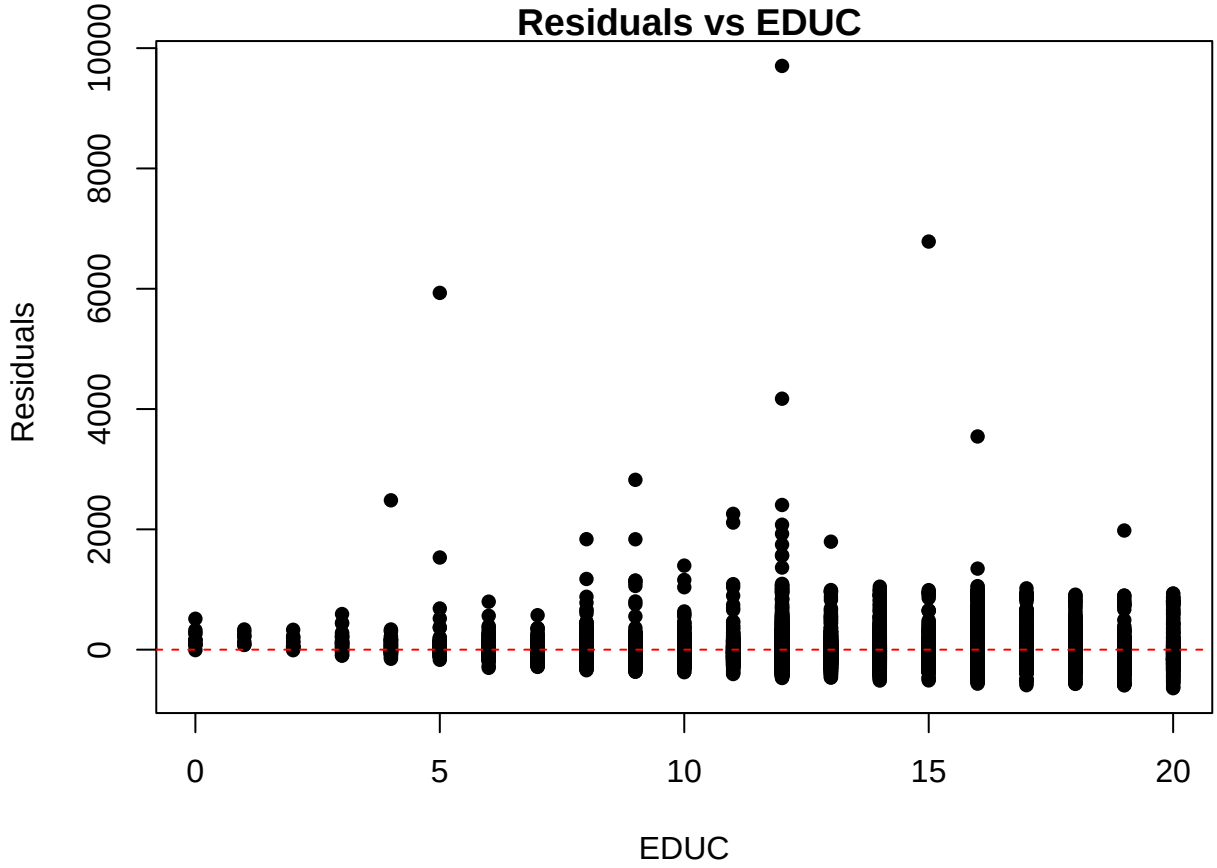

```
Jarque Bera Test
```

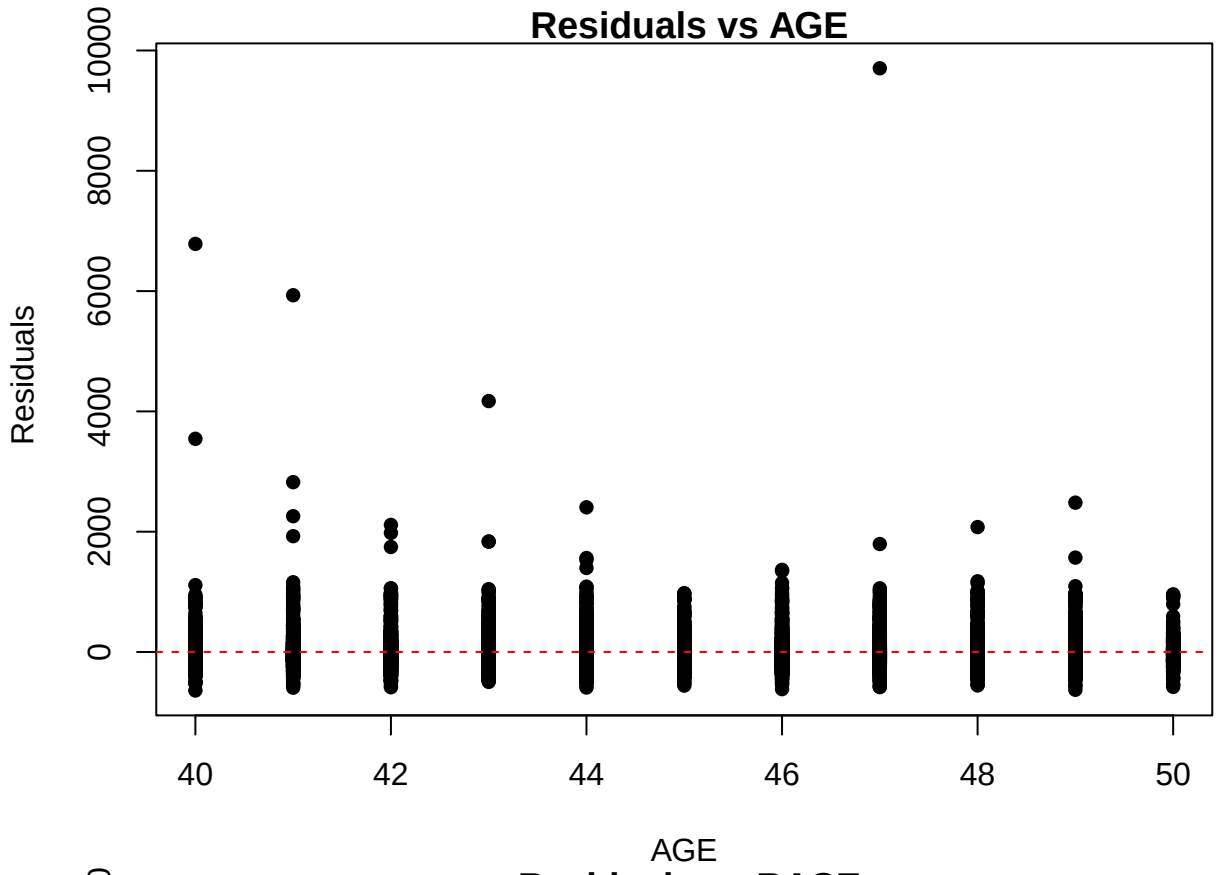
data: residuals X-squared = 21722226, df = 2, p-value < 2.2e-16

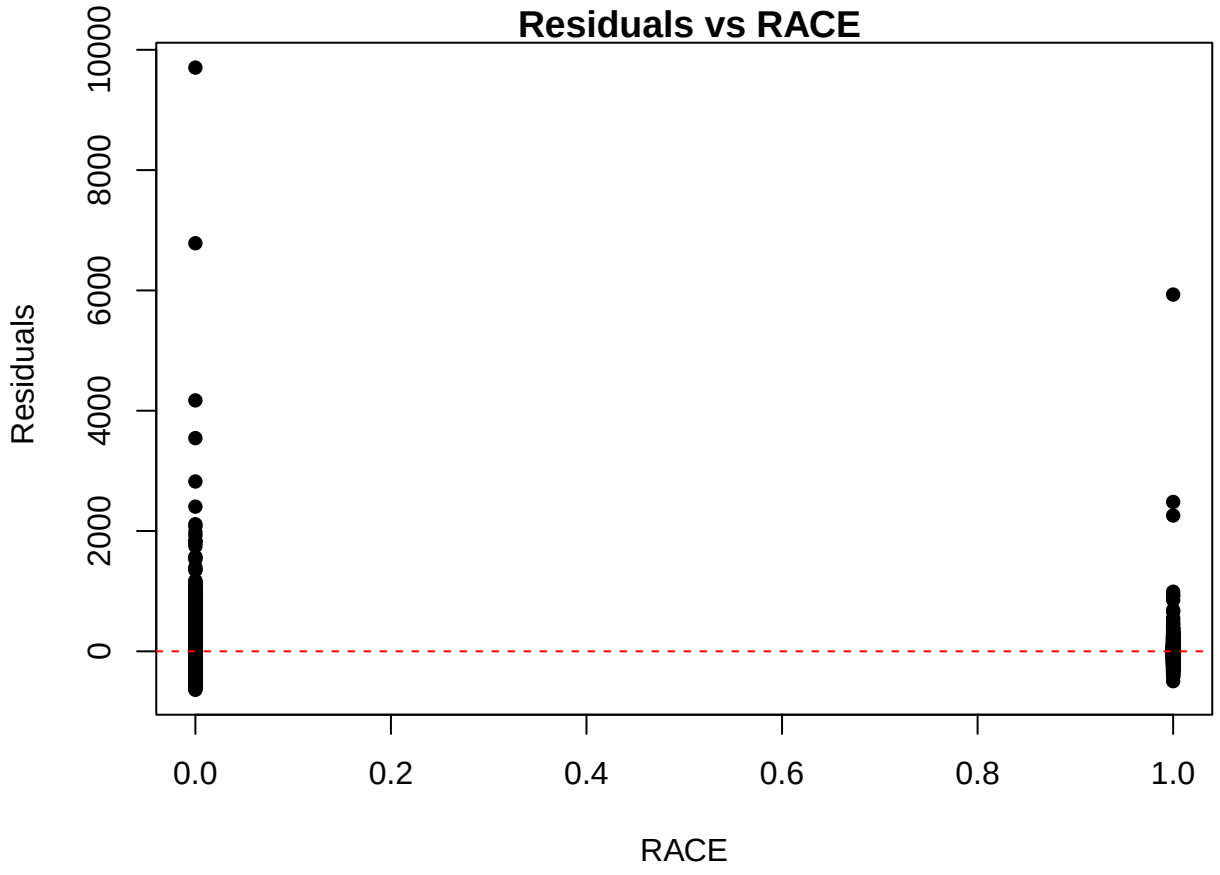
Table 19:

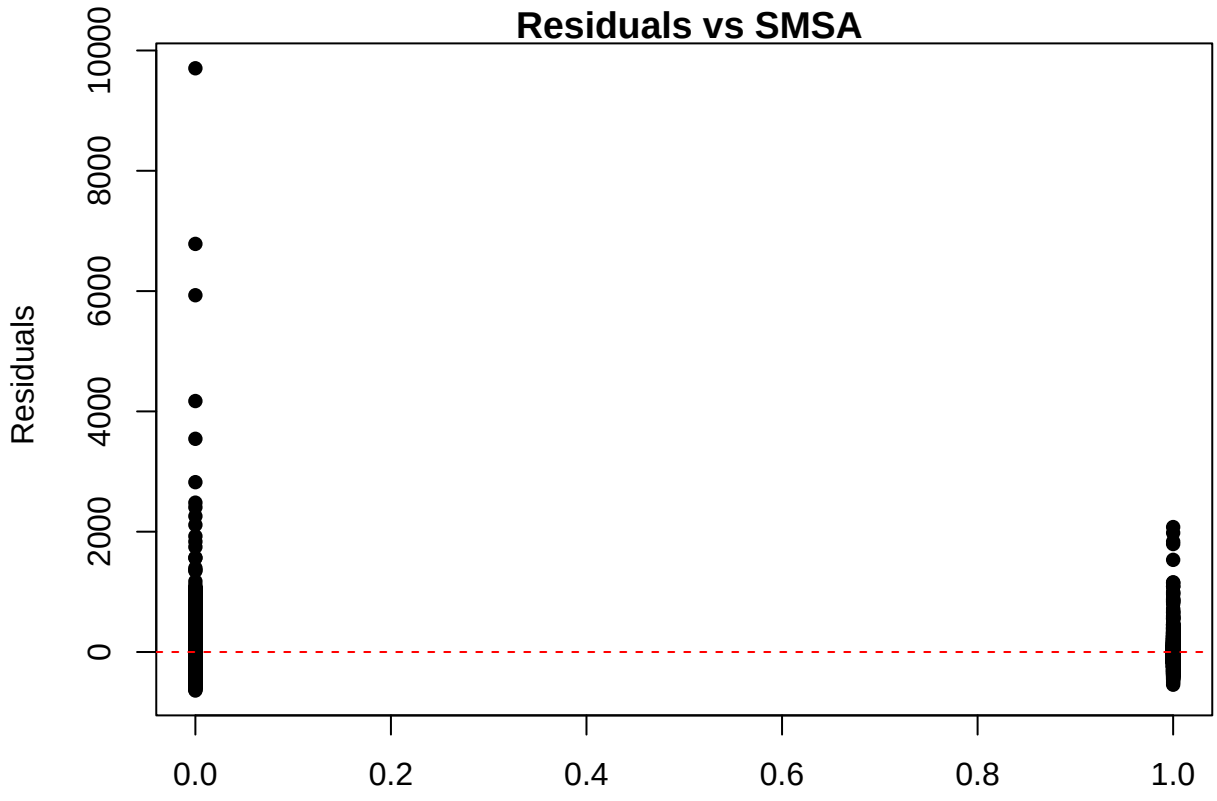
	Variable	VIF
EDUC	EDUC	1.072
AGE	AGE	1.008
RACE	RACE	1.054
SMSA	SMSA	1.073
MARRIED	MARRIED	1.020
REGION2	REGION2	3.281
REGION3	REGION3	3.672
REGION4	REGION4	2.183
REGION5	REGION5	3.510
REGION6	REGION6	2.059
REGION7	REGION7	2.541
REGION8	REGION8	1.817
REGION9	REGION9	2.919

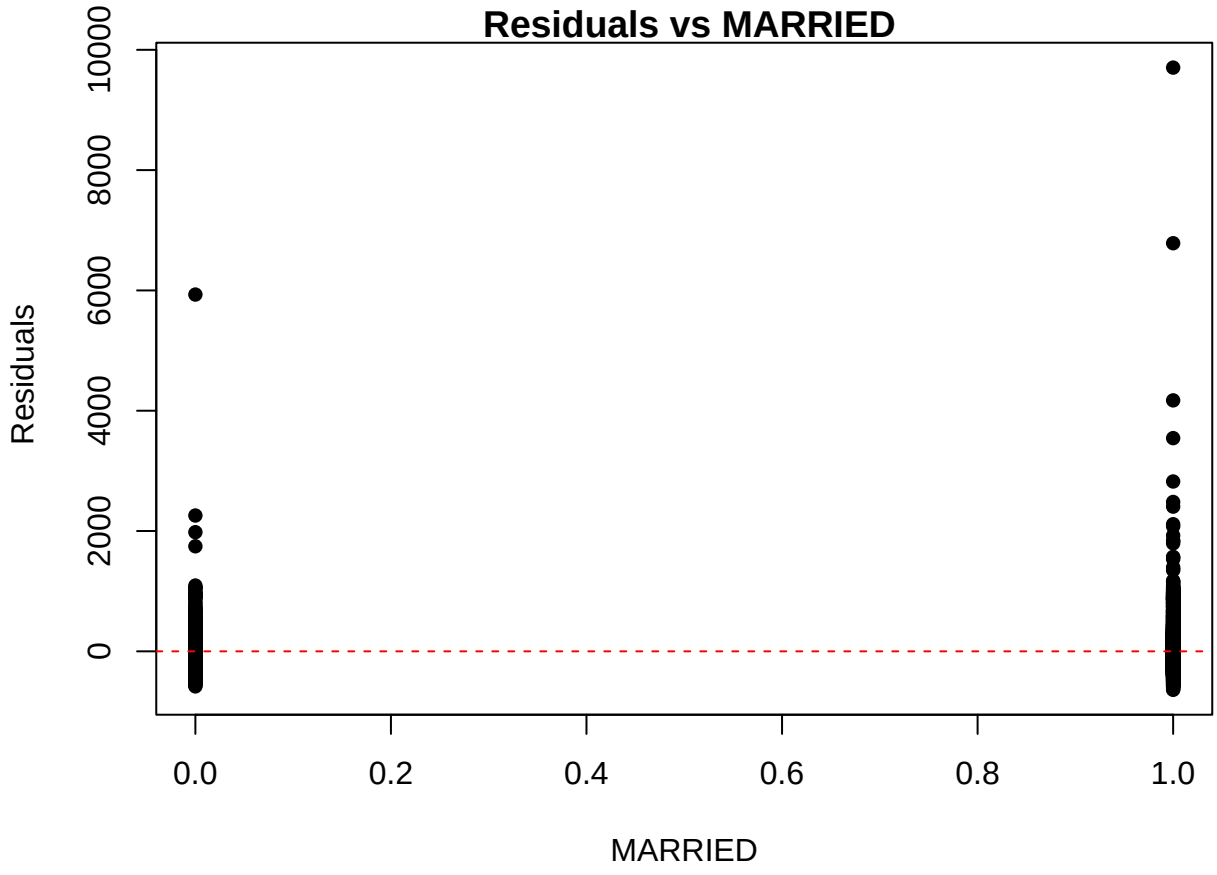

Residuals vs. Fitted Values
Residuals
10000
6000
2000
0
0
200
400
600
Fitted Values

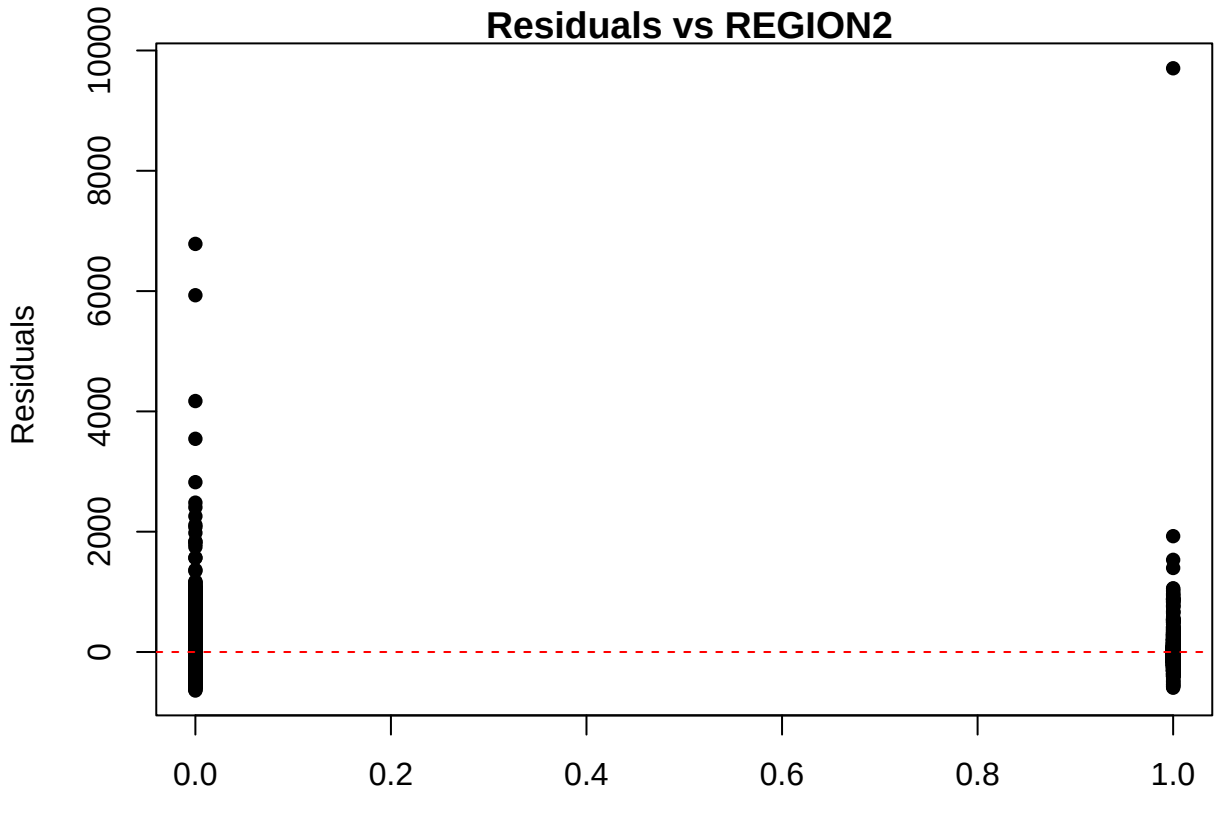

Residuals vs EDUC
Residuals
10000
8000
6000
4000
2000
0
0
5
10
15
20
EDUC

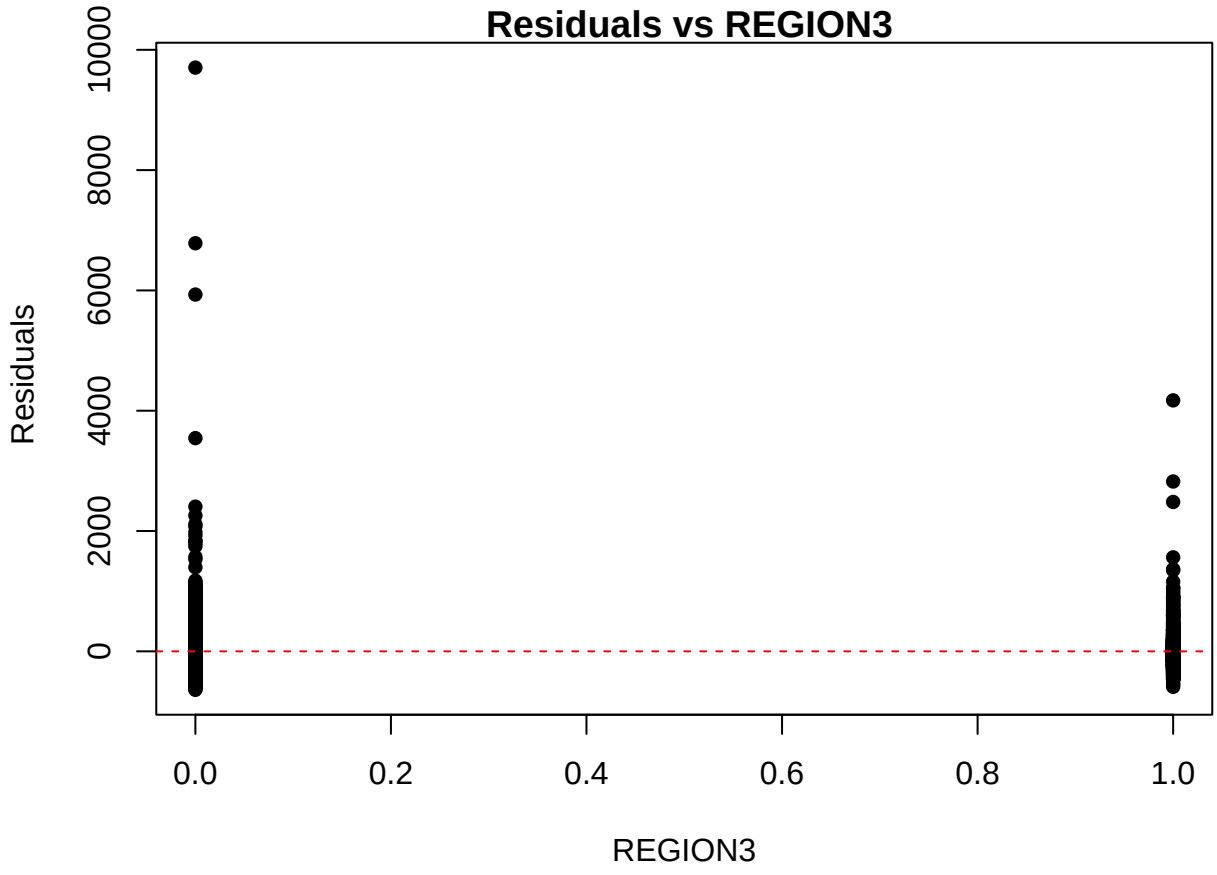

Residuals vs AGE
Residuals
0
2000
4000
6000
8000
10000
40
42
44
46
48
50
AGE

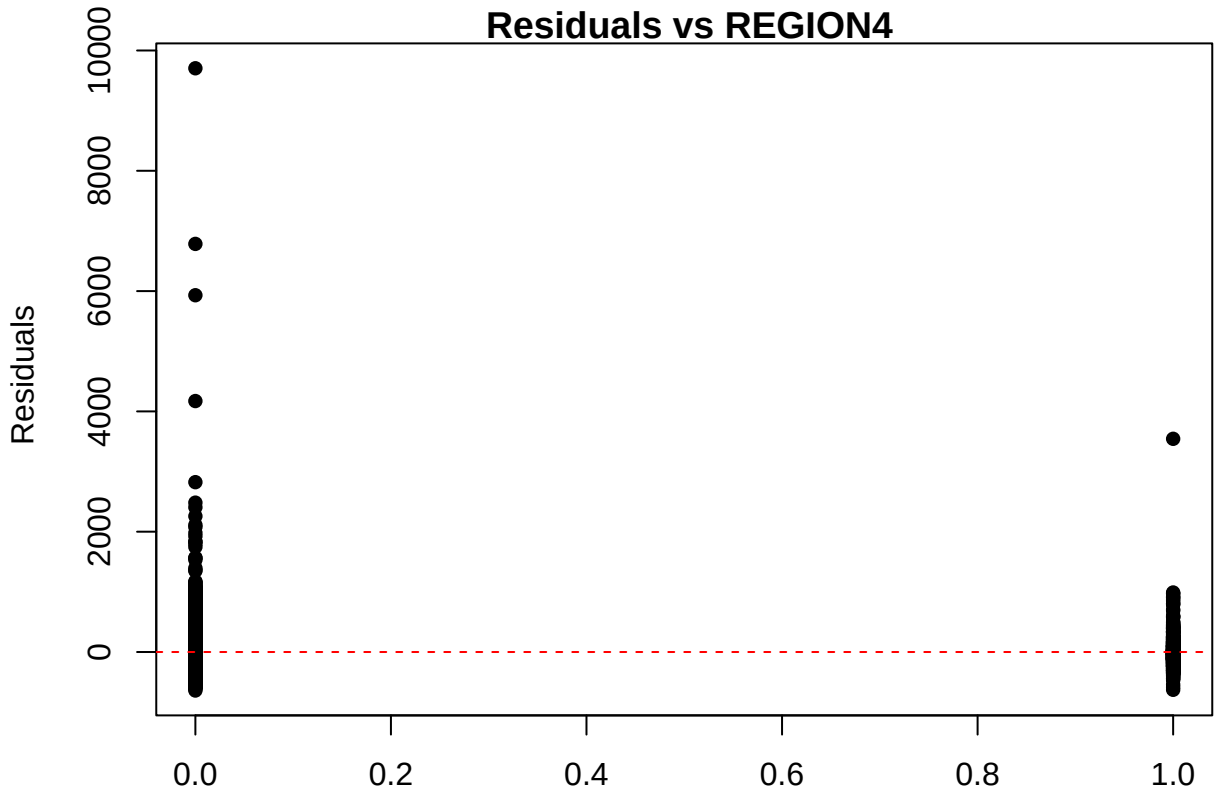

Residuals vs RACE
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
RACE

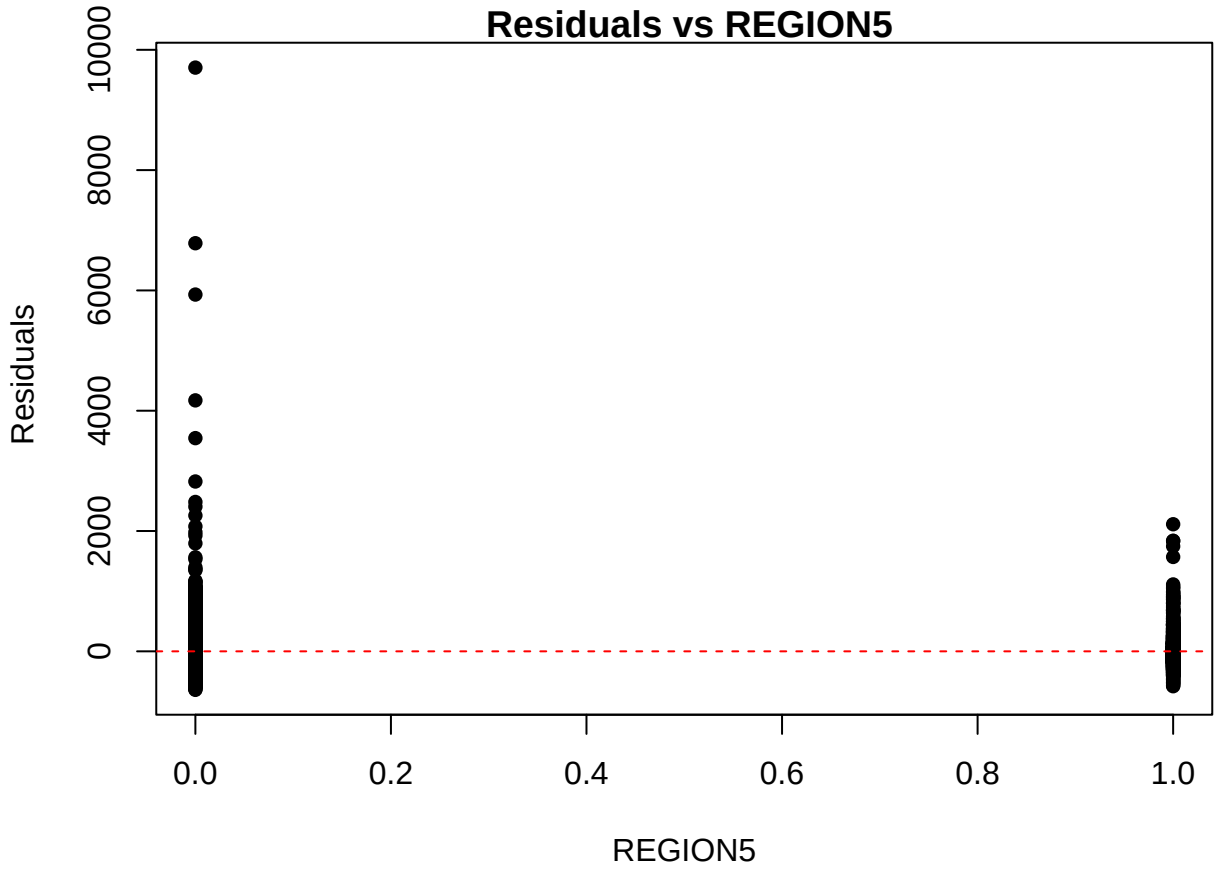

Residuals vs SMSA
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
SMSA

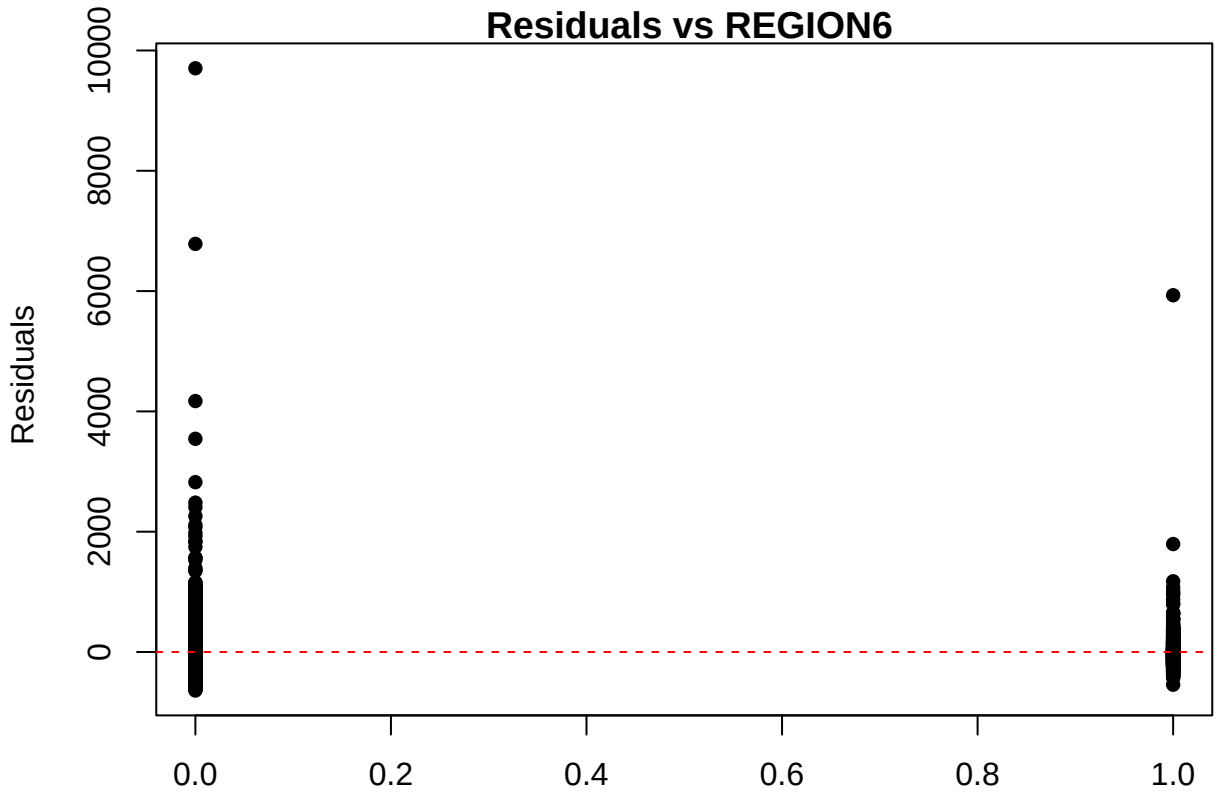

Residuals vs MARRIED
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
MARRIED

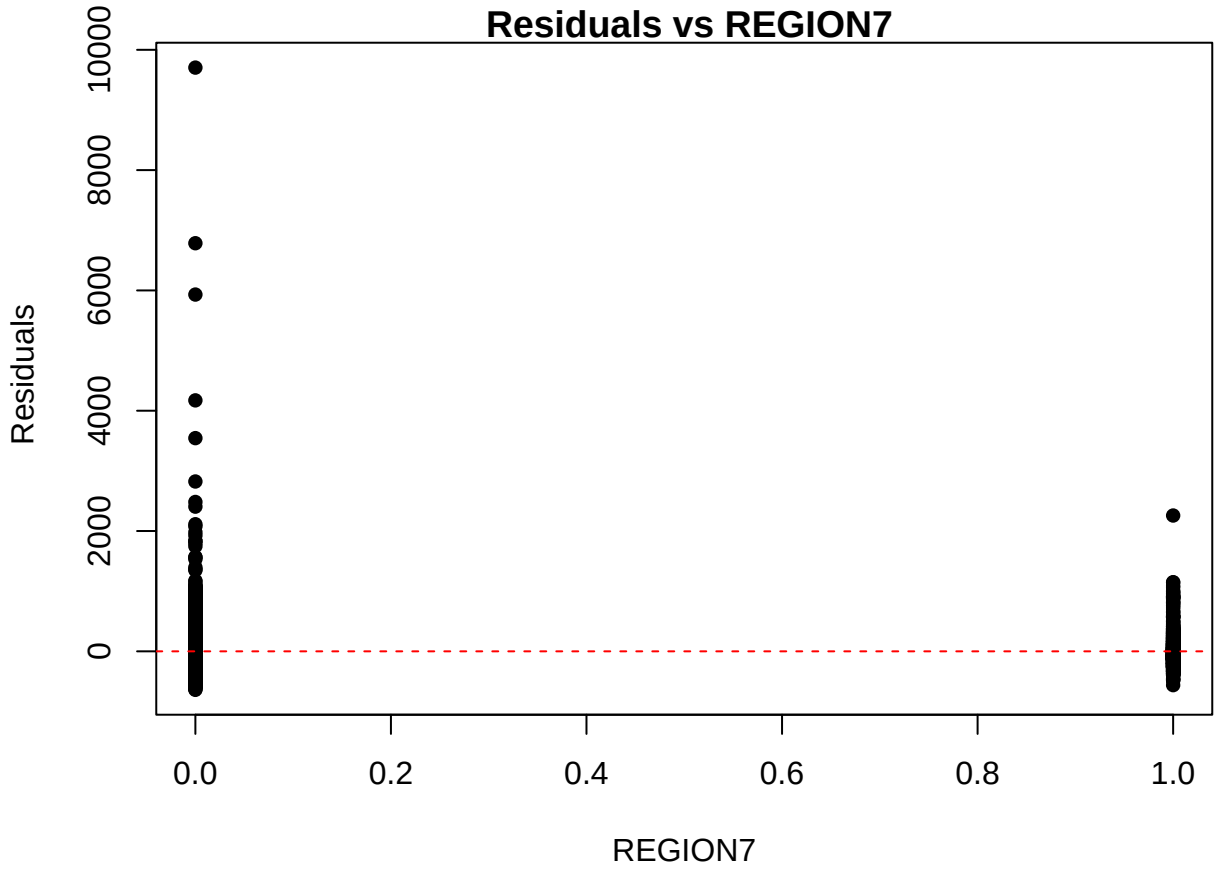

Residuals vs REGION2
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION2

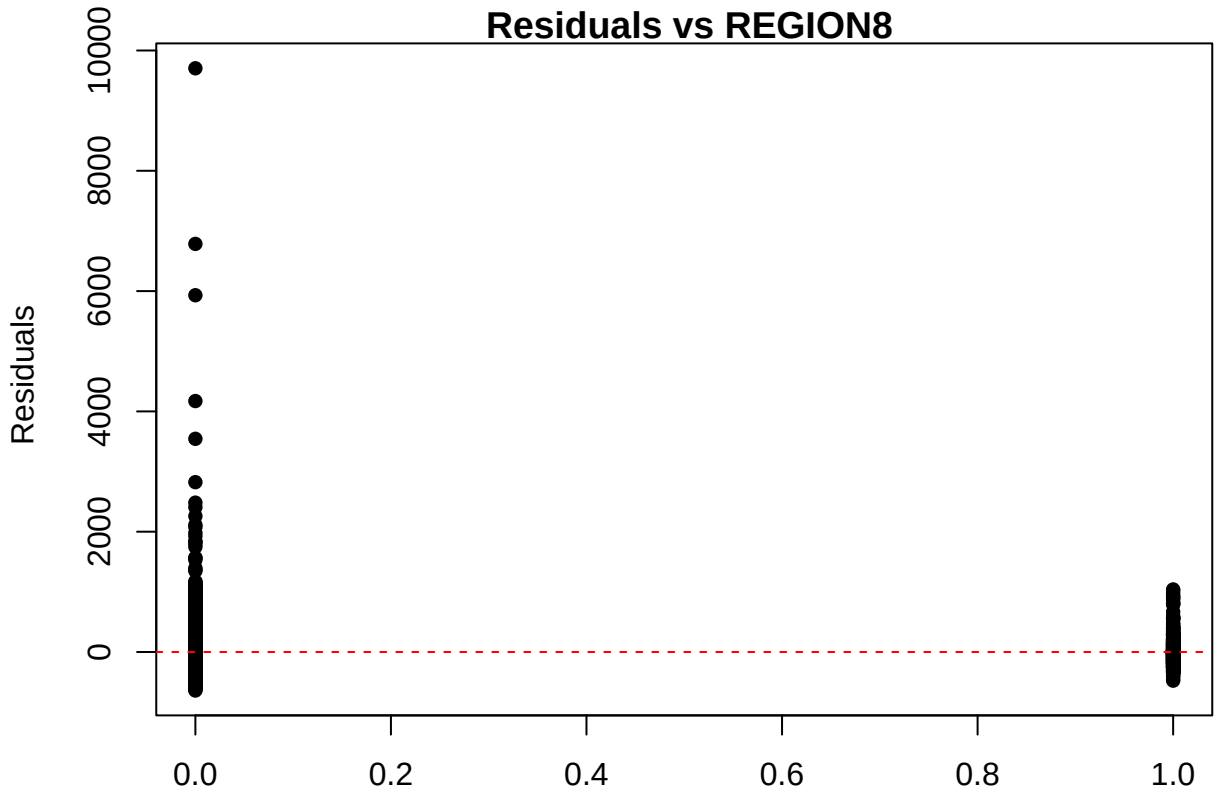

Residuals vs REGION3
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION3

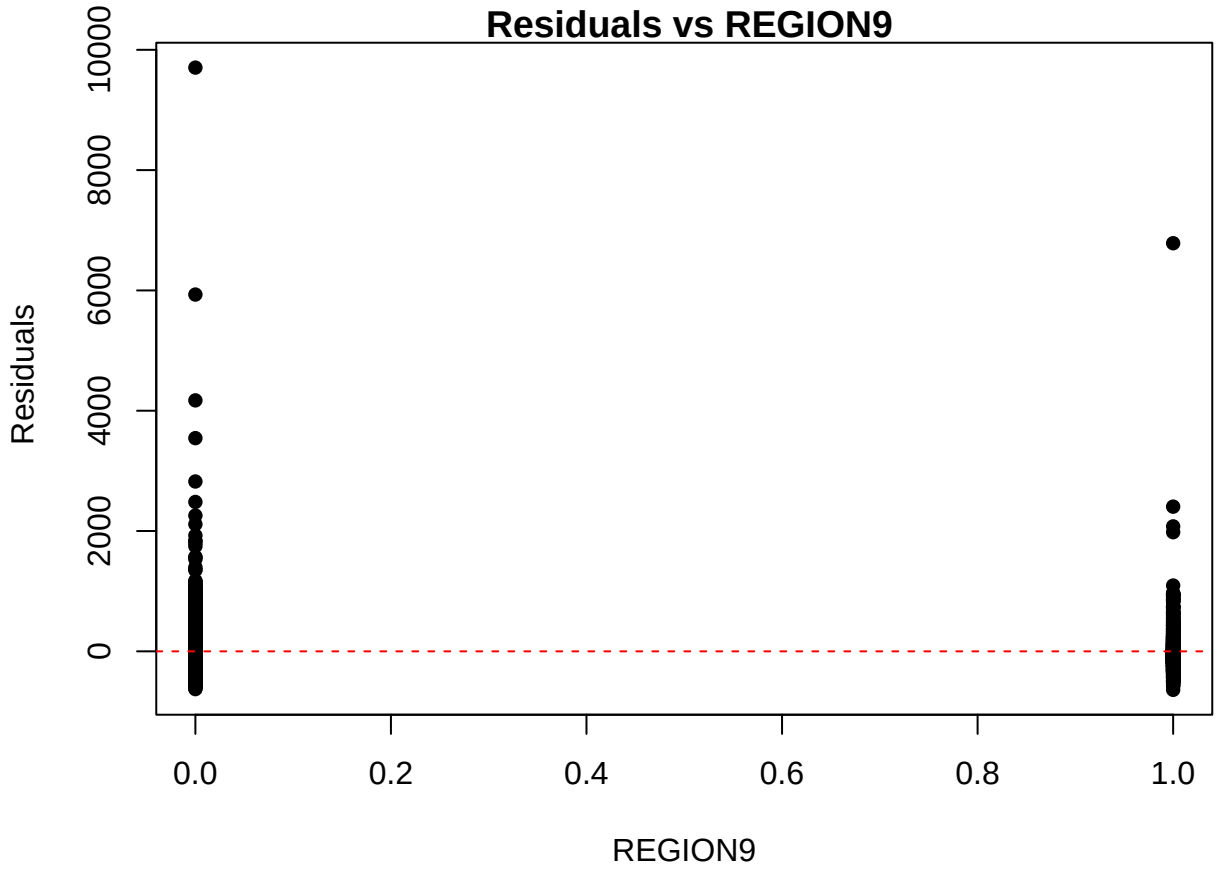

Residuals vs REGION4
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION4


Residuals vs REGION5
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION5


Residuals vs REGION6
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION6


Residuals vs REGION7
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION7


Residuals vs REGION8
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION8


Residuals vs REGION9
Residuals
0
2000
4000
6000
8000
10000
0.0
0.2
0.4
0.6
0.8
1.0
REGION9

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

```
##
##  studentized Breusch-Pagan test
##
## data:  linear_model1
## BP = 7.0287, df = 2, p-value = 0.02977
```

```r
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
```

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

```
##
##  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
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Mon, Jun 09, 2025 - 14:44:50

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Mon, Jun 09, 2025 - 14:44:50

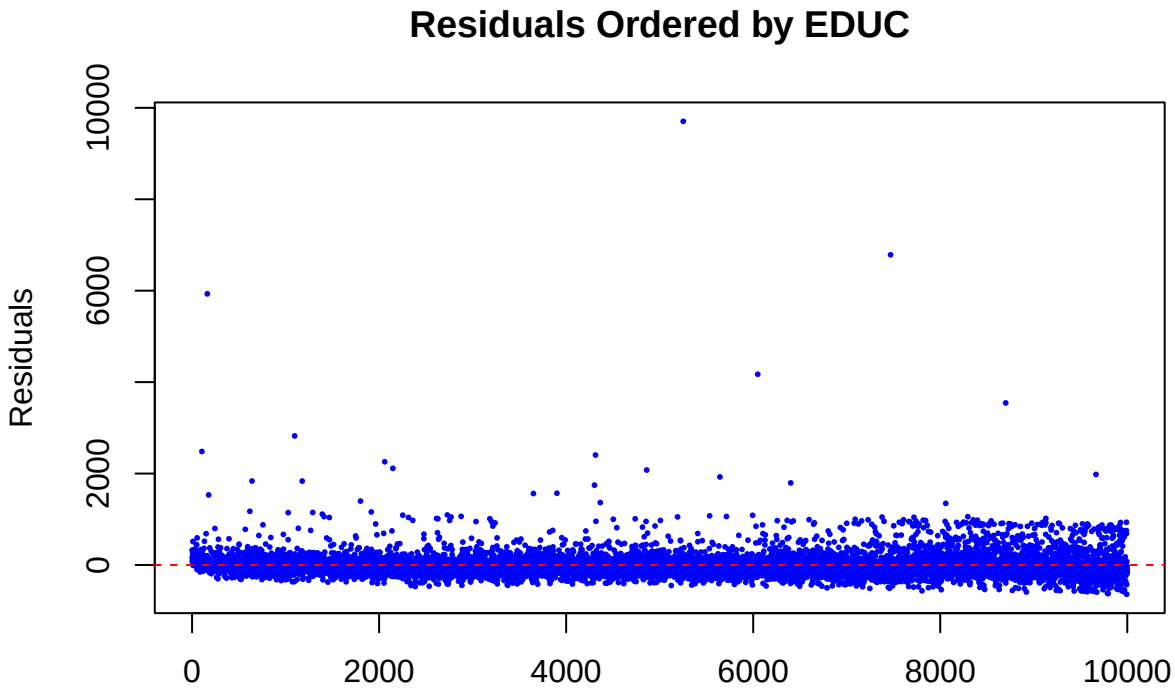
Table 20:

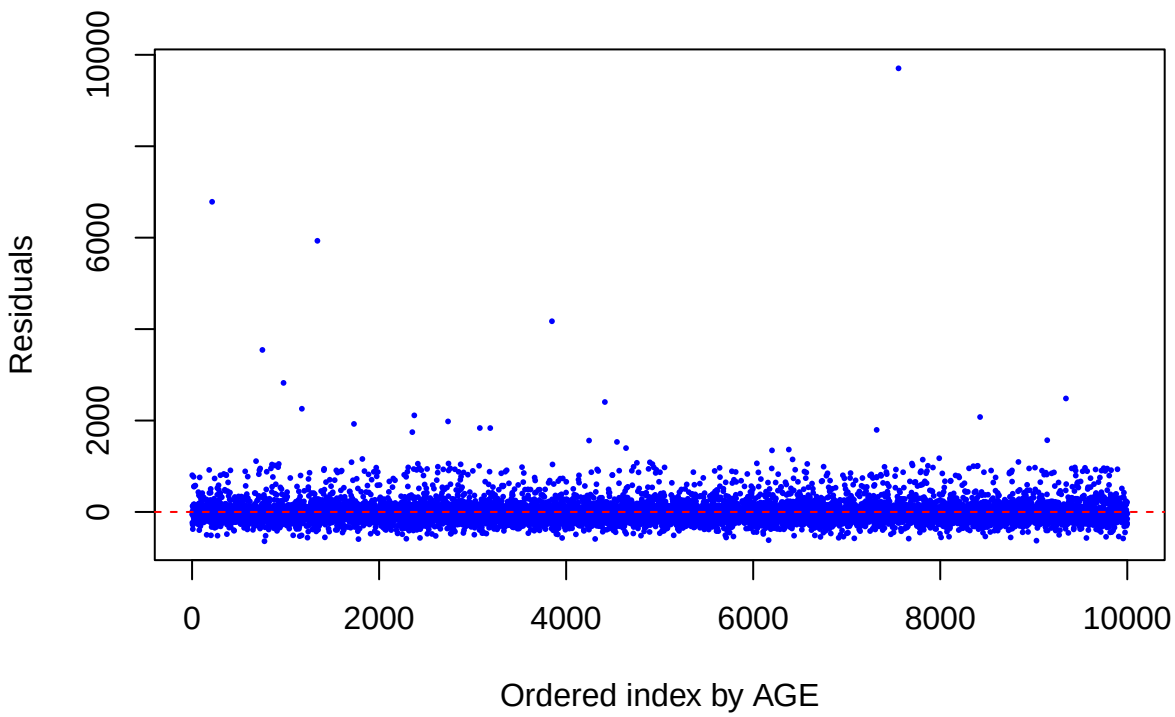
	Dependent variable:
	Wage
Education	26.6959***
	(0.9189)
Age	2.2442**
	(0.9897)
Race	−77.6318***
	(9.3927)
SMSA	−63.7562***
	(6.2889)
Married	78.0013***
	(7.9321)
Region 2	41.1931***
	(12.5732)
Region 3	49.0879***
	(10.7566)
Region 4	18.8453
	(12.9527)
Region 5	4.2049
	(10.8923)
Region 6	7.6044
	(15.3882)
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

Table 21:

	Dependent variable:
	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

	1	2
(Intercept)	−81.669	435.749***
	(49.832)	(51.022)
EDUC	26.696***	17.983***
	(0.919)	(0.769)
AGE	2.244*	−6.545***
	(0.990)	(1.056)
RACE	−77.632***	−69.011***
	(9.393)	(7.888)
SMSA	−63.756***	−78.023***
	(6.289)	(5.579)
MARRIED	78.001***	53.247***
	(7.932)	(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)






```
##
##  Runs Test
##
## data:  residuals_ordered
## statistic = -2.6201, runs = 4870, n1 = 5000, n2 = 5000, n = 10000,
## p-value = 0.00879
```

```
## alternative hypothesis: nonrandomness
##
##  Durbin-Watson test
##
## data:  linear_model_ordered
## DW = 1.9568, p-value = 0.01502
## alternative hypothesis: true autocorrelation is greater than 0
##
##  Breusch-Godfrey test for serial correlation of order up to 6
##
## data:  linear_model_ordered
## LM test = 13.468, df = 6, p-value = 0.03617
```

```
##
## ============================================================
##                                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
```

```
Ramsey_test <- ((ssr_original - ssr_ramsey)/2)/(ssr_ramsey/(n - k - 2))
p_value <- pf(Ramsey_test, 2, n - k - 2, lower.tail = FALSE)
```

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

```
reset_test <- resettest(linear_model1, power = 2:3, type = "fitted")
print(reset_test)
```

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

```
res=linear_model1$residuals
LM_reg=lm(res~EDUC+I(EDUC^2)+I(EDUC^3)+ AGE +I(AGE^2)+ I(AGE^3)+ RACE +
            SMSA + MARRIED + REGION2 + REGION3 + REGION4 + REGION5 + REGION6
          + REGION7 + REGION8 + REGION9, data = df)
```

```
##
## ======================================================
##                            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
## ----------------------
## 76.692            0
## ----------------------
```

```
res=linear_model1$residuals
LM_reg_educ=lm(res~EDUC+I(EDUC^2)+I(EDUC^3)+ AGE + RACE + SMSA + MARRIED + REGION2
               + REGION3 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9, data = df)
```

```
##
## ============================================================
##                              Dependent variable:
##                     -----------------------------------
##                                      res
## ------------------------------------------------------------
## EDUC                               -78.409***
##                                    (12.149)
##                                    t = -6.454
##                                    p = 0.000
## I(EDUC2)                            5.689***
##                                    (1.086)
##                                    t = 5.238
##                                   p = 0.00000
## I(EDUC3)                           -0.127***
##                                    (0.030)
##                                    t = -4.159
##                                   p = 0.00004
```

```
## AGE                                -0.268
##                                   (0.939)
##                                  t = -0.285
##                                  p = 0.776
## RACE                               -4.121
##                                   (10.201)
##                                  t = -0.404
##                                  p = 0.687
## SMSA                                0.401
##                                   (7.373)
##                                  t = 0.054
##                                  p = 0.957
## MARRIED                             2.277
##                                   (8.001)
##                                  t = 0.285
##                                  p = 0.776
## REGION2                            -0.610
##                                   (13.595)
##                                  t = -0.045
##                                  p = 0.965
## REGION3                             0.461
##                                   (13.256)
##                                  t = 0.035
##                                  p = 0.973
## REGION4                            -0.674
##                                   (15.544)
##                                  t = -0.043
##                                  p = 0.966
## REGION5                            -3.686
##                                   (13.450)
##                                  t = -0.274
##                                  p = 0.785
## REGION6                            -6.174
##                                   (16.084)
##                                  t = -0.384
##                                  p = 0.702
## REGION7                            -6.134
##                                   (14.613)
##                                  t = -0.420
##                                  p = 0.675
## REGION8                            -0.951
##                                   (17.050)
##                                  t = -0.056
##                                  p = 0.956
## REGION9                            -1.756
##                                   (14.005)
##                                  t = -0.125
##                                  p = 0.901
## Constant                          339.975***
##                                   (63.486)
##                                  t = 5.355
##                                 p = 0.00000
## ------------------------------------------------------
## Observations                        10,000
```

```
## R2                                0.007
## Adjusted R2                       0.006
## Residual Std. Error         274.032 (df = 9984)
## F Statistic         4.988*** (df = 15; 9984) (p = 0.000)
## ============================================================
## Note:                         *p<0.1; **p<0.05; ***p<0.01
```

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

```
res=linear_model1$residuals
LM_reg_age=lm(res~EDUC+I(AGE^2)+I(AGE^3)+ AGE + RACE + SMSA + MARRIED + REGION2 + REGION3
                + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9, data = df)
```

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                 res
## ------------------------------------------------
## EDUC                            0.017
##                                (0.869)
##                               t = 0.019
##                               p = 0.985
## I(AGE2)                         5.338
##                               (17.875)
##                               t = 0.299
##                               p = 0.766
## I(AGE3)                        -0.036
##                                (0.133)
##                              t = -0.268
##                               p = 0.789
## AGE                           -262.949
##                               (798.583)
##                              t = -0.329
##                               p = 0.742
## RACE                            0.182
##                               (10.226)
##                               t = 0.018
##                               p = 0.986
## SMSA                            0.201
##                                (7.394)
##                               t = 0.027
##                               p = 0.979
## MARRIED                         0.026
##                                (8.026)
##                               t = 0.003
##                               p = 0.998
## REGION2                        -0.069
##                               (13.644)
##                              t = -0.005
```

```
##                                 p = 0.996
## REGION3                          0.149
##                                (13.305)
##                                 t = 0.011
##                                 p = 0.992
## REGION4                         -0.096
##                                (15.600)
##                                t = -0.006
##                                 p = 0.996
## REGION5                         -0.243
##                                (13.493)
##                                t = -0.018
##                                 p = 0.986
## REGION6                          0.510
##                                (16.126)
##                                 t = 0.032
##                                 p = 0.975
## REGION7                         -0.154
##                                (14.646)
##                                t = -0.010
##                                 p = 0.992
## REGION8                         -0.190
##                                (17.107)
##                                t = -0.011
##                                 p = 0.992
## REGION9                          0.110
##                                (14.046)
##                                 t = 0.008
##                                 p = 0.994
## Constant                        4,267.804
##                               (11,866.400)
##                                 t = 0.360
##                                 p = 0.720
## ---------------------------------------------------
## Observations                     10,000
## R2                               0.0002
## Adjusted R2                      -0.001
## Residual Std. Error        275.023 (df = 9984)
## F Statistic         0.165 (df = 15; 9984) (p = 1.000)
## ===================================================
## 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        log(WAGE)
```

```
##                                      Level        Log-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***
## ============================================================
## 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 <- resettest(log_model, power = 2:3, type = "fitted")
print(reset_test)
```

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

```
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
```

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

```
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 + 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(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 + 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(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 + 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(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
```

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

```
##
## ======================================
## Test-statistic Forecast chi sq P-value
## --------------------------------------
## 864.620                         0.999
## --------------------------------------
```

```
reset_test <- resettest(best_mod, power = 2:3, type = "fitted")
print(reset_test)
```

```
##
##  RESET test
##
## data:  best_mod
```

```
## RESET = 2.445, df1 = 2, df2 = 9981, p-value = 0.08678
```

```r
if (!requireNamespace("AER", quietly = TRUE)) install.packages("AER")
library(AER)
```

```
## Warning: package 'AER' was built under R version 4.3.3
```

```
## Loading required package: survival
```

```r
iv_model <- ivreg(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + factor(REGION) |
                    QOB + AGE + RACE + MARRIED + SMSA + factor(REGION),
                  data = df)

summary(iv_model, diagnostics = TRUE)
```

```
##
## 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 *
##
## Diagnostic tests:
##                   df1  df2 statistic p-value
## Weak instruments    1 9986     4.328  0.0375 *
## Wu-Hausman          1 9985     0.007  0.9349
## Sargan              0   NA        NA      NA
## ---
## 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
```

```r
if (!requireNamespace("AER", quietly = TRUE)) install.packages("AER")
library(AER)

iv_modelfactor <- ivreg(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + factor(REGION) |
```

```
                  factor(QOB) + AGE + RACE + MARRIED + SMSA + factor(REGION),
                data = df)

summary(iv_modelfactor, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + factor(REGION) |
##     factor(QOB) + AGE + RACE + MARRIED + SMSA + factor(REGION),
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -513.39 -146.84  -46.46   84.07 9675.05
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      365.66451  435.98983   0.839 0.401658
## EDUC               0.05619   25.81727   0.002 0.998263
## AGE                0.07230    2.32297   0.031 0.975170
## RACE            -124.25131   46.40167  -2.678 0.007424 **
## MARRIED           82.00629    9.24773   8.868  < 2e-16 ***
## SMSA             -96.62067   32.75603  -2.950 0.003188 **
## factor(REGION)2   40.70416   14.27976   2.850 0.004374 **
## factor(REGION)3   37.19148   18.06742   2.058 0.039570 *
## factor(REGION)4   16.05045   16.54083   0.970 0.331894
## factor(REGION)5   -6.18516   17.33281  -0.357 0.721214
## factor(REGION)6  -18.23311   30.17656  -0.604 0.545715
## factor(REGION)7    3.22484   20.36608   0.158 0.874189
## factor(REGION)8   22.22860   18.91487   1.175 0.239946
## factor(REGION)9   79.00354   20.86277   3.787 0.000153 ***
##
## Diagnostic tests:
##                   df1  df2 statistic p-value
## Weak instruments    3 9984     4.130 0.00618 **
## Wu-Hausman          1 9985     1.166 0.28017
## Sargan              2   NA     0.484 0.78508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 287.7 on 9986 degrees of freedom
## Multiple R-Squared: 0.05258, Adjusted R-squared: 0.05135
## Wald test: 42.36 on 13 and 9986 DF,  p-value: < 2.2e-16
```

```
first_stage <- lm(EDUC ~ AGE + RACE + MARRIED + SMSA + factor(REGION)+ QOB, data = df)
summary(first_stage)
```

```
##
## Call:
## lm(formula = EDUC ~ AGE + RACE + MARRIED + SMSA + factor(REGION) +
##     QOB, 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 ***
## 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 ***
## QOB              0.05952    0.02861   2.080 0.037510 *
## ---
## 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
```

```
first_stagefactor <- lm(EDUC ~ AGE + RACE + MARRIED + SMSA + factor(REGION)+ factor(QOB), data = df)
summary(first_stagefactor)
```

```
##
## Call:
## lm(formula = EDUC ~ AGE + RACE + MARRIED + SMSA + factor(REGION) +
##     factor(QOB), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9346  -1.5650  -0.4542   2.1027   9.4079
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     16.59752    0.52615  31.545  < 2e-16 ***
## AGE             -0.07907    0.01094  -7.226 5.34e-13 ***
## RACE            -1.74777    0.11643 -15.011  < 2e-16 ***
## MARRIED          0.15402    0.09239   1.667  0.09555 .
## SMSA            -1.23427    0.08421 -14.657  < 2e-16 ***
## factor(REGION)2 -0.01927    0.15706  -0.123  0.90237
## factor(REGION)3 -0.44465    0.15310  -2.904  0.00369 **
## factor(REGION)4 -0.09355    0.17961  -0.521  0.60249
## factor(REGION)5 -0.38840    0.15528  -2.501  0.01239 *
## factor(REGION)6 -0.96504    0.18536  -5.206 1.96e-07 ***
## factor(REGION)7 -0.51186    0.16856  -3.037  0.00240 **
## factor(REGION)8  0.23877    0.19692   1.212  0.22535
## factor(REGION)9  0.57912    0.16159   3.584  0.00034 ***
## factor(QOB)2     0.11171    0.09141   1.222  0.22170
## factor(QOB)3    -0.03431    0.08853  -0.388  0.69837
## factor(QOB)4     0.25101    0.09056   2.772  0.00559 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.166 on 9984 degrees of freedom
## Multiple R-squared:  0.06836,    Adjusted R-squared:  0.06696
## F-statistic: 48.84 on 15 and 9984 DF,  p-value: < 2.2e-16
```

```
first_stage <- lm(EDUC ~ QOB + AGE + RACE + MARRIED + SMSA + factor(REGION), data = df)
df$first_stage_residuals <- residuals(first_stage)

augmented_modelHausman <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2
                      + REGION3 + REGION4 + REGION5 + REGION6 + REGION7
                      + REGION8 + REGION9 + first_stage_residuals, data = df)
summary(augmented_modelHausman)
```

```
##
## Call:
## lm(formula = WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 +
##     REGION3 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 +
##     REGION9 + 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
## REGION2                  41.131     13.666   3.010  0.00262 **
## REGION3                  47.565     22.901   2.077  0.03782 *
## REGION4                  18.488     16.204   1.141  0.25392
## REGION5                   2.875     21.143   0.136  0.89184
## REGION6                   4.298     43.574   0.099  0.92143
## REGION7                  15.307     26.187   0.585  0.55888
## REGION8                  16.707     19.775   0.845  0.39822
## REGION9                  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
```

```
first_stagefactor <- lm(EDUC ~ factor(QOB) + AGE + RACE + MARRIED + SMSA + factor(REGION), data = df)
df$first_stage_residuals <- residuals(first_stagefactor)

augmented_modelHausmanfactor <- lm(WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 + REGION3 + REGI
summary(augmented_modelHausmanfactor)
```

```
##
## Call:
## lm(formula = WAGE ~ EDUC + AGE + RACE + MARRIED + SMSA + REGION2 +
##     REGION3 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 +
##     REGION9 + first_stage_residuals, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -636.0 -131.9  -35.5   76.3 9702.1
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            365.66451  416.81028   0.877  0.38035
## EDUC                     0.05619   24.68154   0.002  0.99818
## AGE                      0.07230    2.22078   0.033  0.97403
## RACE                  -124.25131   44.36042  -2.801  0.00511 **
## MARRIED                 82.00629    8.84092   9.276  < 2e-16 ***
## SMSA                   -96.62067   31.31506  -3.085  0.00204 **
## REGION2                 40.70416   13.65158   2.982  0.00287 **
## REGION3                 37.19148   17.27261   2.153  0.03133 *
## REGION4                 16.05045   15.81318   1.015  0.31013
## REGION5                 -6.18516   16.57033  -0.373  0.70896
## REGION6                -18.23311   28.84907  -0.632  0.52739
## REGION7                  3.22484   19.47016   0.166  0.86845
## REGION8                 22.22860   18.08279   1.229  0.21900
## REGION9                 79.00354   19.94500   3.961 7.51e-05 ***
## first_stage_residuals   26.67275   24.69685   1.080  0.28017
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 275 on 9985 degrees of freedom
## Multiple R-squared:  0.1342, Adjusted R-squared:  0.133
## F-statistic: 110.5 on 14 and 9985 DF,  p-value: < 2.2e-16
```

Step 4: General remedies and conclusions

```r
library(AER)
library(lmtest)
library(dplyr)


df <- read.csv(file="Data_Sub_AK91.csv", header=TRUE, sep=",")

df <- df %>%
  mutate(
    REGION1 = as.integer(REGION == 1),
    REGION2 = as.integer(REGION == 2),
    REGION3 = as.integer(REGION == 3),
    REGION4 = as.integer(REGION == 4),
    REGION5 = as.integer(REGION == 5),
    REGION6 = as.integer(REGION == 6),
    REGION7 = as.integer(REGION == 7),
    REGION8 = as.integer(REGION == 8),
```

```
    REGION9 = as.integer(REGION == 9)
  )

# Trim extreme wage values
lower_bound <- quantile(log(df$WAGE), 0.005, na.rm = TRUE)
upper_bound <- quantile(log(df$WAGE), 0.995, na.rm = TRUE)
df_trimmed <- df %>% filter(log(WAGE) >= lower_bound & log(WAGE) <= upper_bound)


df_trimmed <- df_trimmed %>%
  mutate(
    QOB_num = as.numeric(as.character(QOB)),  # Convert to numeric first
    QOB = droplevels(factor(QOB))  # Keep as factor for interactions
)


stopifnot(nlevels(df_trimmed$QOB) >= 2)

# 2SLS regression with proper polynomial specification
iv_model <- ivreg(
  log(WAGE) ~ EDUC + I(EDUC^2) + I(EDUC^3) + EDUC:RACE + AGE + RACE + MARRIED + SMSA +
    REGION1 + REGION2 + REGION3 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9 |
    QOB_num + I(QOB_num^2) + I(QOB_num^3) + QOB:RACE + AGE + RACE + MARRIED + SMSA +
    REGION1 + REGION2 + REGION3 + REGION4 + REGION5 + REGION6 + REGION7 + REGION8 + REGION9,
  data = df_trimmed
)

# HAC robust standard errors
robust_se <- vcovHAC(iv_model)
model_results <- coeftest(iv_model, vcov = robust_se)

# Display results
print(model_results)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.1602198  6.8935128  0.6035 0.546191
## EDUC        -0.5547510  1.3424502 -0.4132 0.679441
## I(EDUC^2)    0.1030725  0.1174515  0.8776 0.380196
## I(EDUC^3)   -0.0038967  0.0036941 -1.0548 0.291527
## AGE          0.0110147  0.0141679  0.7774 0.436917
## RACE         0.0834258  2.2917072  0.0364 0.970961
## MARRIED      0.1979697  0.0747194  2.6495 0.008074 **
## SMSA        -0.1458898  0.1244752 -1.1720 0.241210
## REGION1     -0.0460966  0.1066427 -0.4323 0.665567
## REGION2      0.0387098  0.1080106  0.3584 0.720060
## REGION3      0.0610615  0.1171350  0.5213 0.602175
## REGION4     -0.0050486  0.1092356 -0.0462 0.963138
## REGION5     -0.0363396  0.1839483 -0.1976 0.843399
## REGION6      0.0227908  0.3312669  0.0688 0.945151
## REGION7     -0.0344581  0.2008669 -0.1715 0.863797
## REGION8     -0.0505700  0.0537268 -0.9412 0.346604
```

```
## EDUC:RACE   -0.0253100  0.2034994 -0.1244 0.901022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```