BUAN 6356.501 - Business Analytics with R (Spring 2019)
Problem Set 4

Question 1 <– hprice1

```
> AIC(best_model1) # 910.4328
[1] 910.4328
> BIC(best_model1) # 925.2968
[1] 925.2968
> summary(best_model1)

Call:
lm(formula = price ~ assess + sqrft + colonial + assess:sqrft,
    data = hprice1)

Residuals:
     Min       1Q   Median       3Q      Max
-103.086  -20.747   -2.724   17.914  198.887

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.806e+02  5.907e+01   3.057 0.003009 **
assess        2.793e-01  2.081e-01   1.342 0.183236
sqrft        -7.932e-02  2.749e-02  -2.885 0.004986 **
colonial      2.161e+01  9.752e+00   2.216 0.029461 *
assess:sqrft  2.483e-04  6.876e-05   3.611 0.000522 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.06 on 83 degrees of freedom
Multiple R-squared:  0.8475,    Adjusted R-squared:  0.8402
F-statistic: 115.4 on 4 and 83 DF,  p-value: < 2.2e-16
```

Question 2 <– gpa2

```
> AIC(best_model2) # 6590.858
[1] 6590.858
> BIC(best_model2) # 6679.446
[1] 6679.446
> summary(best_model2)

Call:
lm(formula = colgpa ~ sat + tothrs + hsize + hsrank + hsperc +
    female + black + sat:tothrs + sat:hsperc + tothrs:hsperc +
    hsize:hsrank + hsrank:hsperc, data = gpa2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.49445 -0.34247  0.01724  0.37326  2.06664

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.353e-01  1.502e-01   2.898 0.003781 **
sat            2.314e-03  1.368e-04  16.913  < 2e-16 ***
tothrs         9.961e-03  1.917e-03   5.195 2.14e-07 ***
hsize          3.031e-02  9.388e-03   3.229 0.001253 **
hsrank        -5.543e-03  7.103e-04  -7.803 7.59e-15 ***
hsperc         1.276e-02  3.777e-03   3.378 0.000738 ***
female         1.452e-01  1.740e-02   8.346  < 2e-16 ***
black         -3.177e-01  3.814e-02  -8.329  < 2e-16 ***
sat:tothrs    -9.260e-06  1.761e-06  -5.259 1.53e-07 ***
sat:hsperc    -2.483e-05  3.557e-06  -6.980 3.42e-12 ***
tothrs:hsperc  6.972e-05  1.544e-05   4.515 6.50e-06 ***
hsize:hsrank   2.158e-04  7.241e-05   2.980 0.002897 **
hsrank:hsperc  5.236e-05  8.367e-06   6.258 4.29e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5357 on 4124 degrees of freedom
Multiple R-squared:  0.3404,    Adjusted R-squared:  0.3385
F-statistic: 177.3 on 12 and 4124 DF,  p-value: < 2.2e-16
```

Question 3 <– mlb1

```
> AIC(best_model3) # 732.2394
[1] 732.2394
> BIC(best_model3) # 767.0376
[1] 767.0376
> summary(best_model3)

Call:
lm(formula = log(salary) ~ years + yrsallst + allstar + gamesyr +
    hrunsyr + games + runs, data = mlb1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.10387 -0.41272 -0.01853  0.43534  2.70470

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.0449913  0.1638187  67.422  < 2e-16 ***
years        0.2113811  0.0371704   5.687 2.76e-08 ***
yrsallst    -0.2045959  0.0575265  -3.557 0.000428 ***
allstar      0.0278327  0.0053722   5.181 3.77e-07 ***
gamesyr      0.0175337  0.0020831   8.417 1.05e-15 ***
hrunsyr      0.0279053  0.0070762   3.944 9.72e-05 ***
games       -0.0018508  0.0004367  -4.238 2.89e-05 ***
runs         0.0015752  0.0003950   3.988 8.15e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6731 on 345 degrees of freedom
Multiple R-squared:  0.6824,    Adjusted R-squared:  0.6759
F-statistic: 105.9 on 7 and 345 DF,  p-value: < 2.2e-16
```

Question 4 <– rental

1. The average rent in 1990 increased by 26% when compared to the average rent in 1980, $\beta_3 = 0.005$ For every 1 unit increase in $pctstu$ (relative student size), there is nearly 0.5% increase in the rent.

```
Call:
plm(formula = log(rent) ~ y90 + log(pop) + log(avginc) + pctstu,
    data = pdrental, model = "pooling")

Balanced Panel: n = 64, T = 2, N = 128

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.242332 -0.078236 -0.016416  0.043890  0.480817

Coefficients:
              Estimate Std. Error t-value  Pr(>|t|)
(Intercept) -0.5688064  0.5348806 -1.0634    0.2897
y90          0.2622267  0.0347632  7.5432 8.781e-12 ***
log(pop)     0.0406863  0.0225154  1.8070    0.0732 .
log(avginc)  0.5714460  0.0530980 10.7621 < 2.2e-16 ***
pctstu       0.0050436  0.0010192  4.9486 2.401e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    14.058
Residual Sum of Squares: 1.9501
R-Squared:      0.86128
Adj. R-Squared: 0.85677
F-statistic: 190.922 on 4 and 123 DF, p-value: < 2.22e-16
```

2. The above standard errors are not valid because we did not use fixed-difference model.
3. We get $\beta_3 = 0.0112$. For 1 point increase in $pctstu$, there is nearly 1.12% increase in rent.

```
Call:
plm(formula = log(rent) ~ y90 + log(pop) + log(avginc) + pctstu,
    data = pdrental, model = "fd")

Balanced Panel: n = 64, T = 2, N = 128
Observations used in estimation: 64

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.186972 -0.062160 -0.014383  0.055183  0.237830

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 0.3855214  0.0368245 10.4692 3.661e-15 ***
log(pop)    0.0722457  0.0883426  0.8178  0.416713
log(avginc) 0.3099604  0.0664771  4.6627 1.788e-05 ***
pctstu      0.0112033  0.0041319  2.7114  0.008726 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    0.7191
Residual Sum of Squares: 0.48736
R-Squared:      0.32226
Adj. R-Squared: 0.28837
F-statistic: 9.50992 on 3 and 60 DF, p-value: 3.1362e-05
```

4. Yes, we get same estimates and same errors as we got in fixed-difference model.

Question 5 <– murder

1. β_1 should be negative for deterrent effect. β_2 could be positive as unemployed might do crimes.
2. For 1990-93, $\beta_1 = 0.12773$ implying no deterrent effect but it is statistically insignificant.

```
call:
plm(formula = mrdrte ~ as.factor(year) + exec + unem, data = murder_9093,
    model = "pooling")

Balanced Panel: n = 51, T = 2, N = 102

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-13.0666  -3.3556  -1.6472   1.6071  66.3873

Coefficients:
                 Estimate Std. Error t-value Pr(>|t|)
(Intercept)       -5.27800    4.42780 -1.1920 0.236134
as.factor(year)93 -2.06742    2.14463 -0.9640 0.337421
exec               0.12773    0.26324  0.4852 0.628599
unem               2.52889    0.78172  3.2350 0.001659 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    11401
Residual Sum of Squares: 10243
R-Squared:      0.10161
Adj. R-Squared: 0.07411
F-statistic: 3.69475 on 3 and 98 DF, p-value: 0.0144
```

3. $\beta_1 = -0.10384$ implying a deterrent effect. With every execution, murder rate drop by 10.38%

```
Call:
plm(formula = mrdrte ~ as.factor(year) + exec + unem, data = murder_9093,
    model = "within")

Balanced Panel: n = 51, T = 2, N = 102

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.2795e+00 -3.4897e-01  2.4980e-16  3.4897e-01  1.2795e+00

Coefficients:
                 Estimate Std. Error t-value Pr(>|t|)
as.factor(year)93  0.413267   0.209385  1.9737  0.05418 .
exec              -0.103840   0.043414 -2.3918  0.02073 *
unem              -0.066591   0.158686 -0.4196  0.67662
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    33.47
Residual Sum of Squares: 27.936
R-Squared:      0.16533
Adj. R-Squared: -0.75627
F-statistic: 3.16936 on 3 and 48 DF, p-value: 0.032614
```

4. The robust standard error for $exec$ is 0.164
5. Texas had 34 executions. Next highest is 11 implying Texas has a lead of 23 executions.
6. The usual and robust errors are close to each other when compare to pooled model.

```
> tidy(model5a)
# A tibble: 3 x 5
  term         estimate std.error statistic p.value
  <chr>           <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   0.413       0.211     1.95    0.0569
2 exec         -0.0675      0.105    -0.643   0.523
3 unem         -0.0700      0.160    -0.437   0.664
> tidyhac(model5a)
# A tibble: 3 x 5
  term         estimate std.error statistic p.value
  <chr>           <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   0.413       0.194     2.12    0.0338
2 exec         -0.0675      0.0767   -0.880   0.379
3 unem         -0.0700      0.142    -0.494   0.621
```

7. For all three years model, $\beta_1 = -0.13832$ but the coefficients from both models are insignificant.

```
call:
plm(formula = mrdrte ~ as.factor(year) + exec + unem, data = pdmurder,
    model = "within")

Balanced Panel: n = 51, T = 3, N = 153

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-26.685751  -0.658371  -0.065721   0.674717  13.394112

Coefficients:
                 Estimate Std. Error t-value Pr(>|t|)
as.factor(year)90  1.55621    0.74533  2.0880  0.03939 *
as.factor(year)93  1.73324    0.70044  2.4745  0.01506 *
exec              -0.13832    0.17701 -0.7815  0.43642
unem               0.22132    0.29638  0.7467  0.45701
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1311.5
Residual Sum of Squares: 1215.2
R-Squared:      0.073367
Adj. R-Squared: -0.43723
F-statistic: 1.93981 on 4 and 98 DF, p-value: 0.10984
```

Question 6 <– airfare

1. If $\Delta bmktshr = 0.10$ then $fare$ increases by 3.6%

```
Call:
plm(formula = log(fare) ~ as.factor(year) + bmktshr + log(dist) +
    I(log(dist)^2), data = pdairfare, model = "pooling")

Balanced Panel: n = 1149, T = 4, N = 4596

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.307955 -0.253740 -0.026729  0.247446  0.958113

Coefficients:
                     Estimate Std. Error t-value  Pr(>|t|)
(Intercept)          6.2092571  0.4206247 14.7620 < 2.2e-16 ***
as.factor(year)1998  0.0211244  0.0140419  1.5044  0.132553
as.factor(year)1999  0.0378496  0.0140413  2.6956  0.007052 **
as.factor(year)2000  0.0998700  0.0140432  7.1116 1.324e-12 ***
bmktshr              0.3601204  0.0300691 11.9764 < 2.2e-16 ***
log(dist)           -0.9016003  0.1282730 -7.0288 2.391e-12 ***
I(log(dist)^2)       0.1030196  0.0097255 10.5927 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    875.09
Residual Sum of Squares: 519.64
R-Squared:      0.40619
Adj. R-Squared: 0.40541
F-statistic: 523.175 on 6 and 4589 DF, p-value: < 2.22e-16
```

2. For β_1, usual confidence interval is $(0.301, 0.419)$ and robust confidence interval is $(0.298, 0.423)$

```
> tidy(model6a,conf.int=TRUE)
# A tibble: 7 x 7
  term                estimate std.error statistic  p.value conf.low conf.high
  <chr>                  <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)            6.21     0.421      14.8   3.24e-48  5.38       7.03
2 as.factor(year)1998    0.0211   0.0140      1.50  1.33e- 1 -0.00640    0.0487
3 as.factor(year)1999    0.0378   0.0140      2.70  7.05e- 3  0.0103     0.0654
4 as.factor(year)2000    0.0999   0.0140      7.11  1.32e-12  0.0723     0.127
5 bmktshr                0.360    0.0301     12.0   1.44e-32  0.301      0.419
6 log(dist)             -0.902    0.128      -7.03  2.39e-12 -1.15      -0.650
7 I(log(dist)^2)         0.103    0.00973    10.6   6.40e-26  0.0840     0.122
> tidyw(model6a,conf.int=TRUE)
# A tibble: 7 x 7
  term                estimate std.error statistic  p.value conf.low conf.high
  <chr>                  <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)            6.21      0.473     13.1   2.23e-39  5.28       7.14
2 as.factor(year)1998    0.0211    0.0142     1.49  1.36e- 1 -0.00668    0.0489
3 as.factor(year)1999    0.0378    0.0144     2.63  8.64e- 3  0.00960    0.0661
4 as.factor(year)2000    0.0999    0.0144     6.94  3.98e-12  0.0717     0.128
5 bmktshr                0.360     0.0318    11.3   1.21e-29  0.298      0.423
6 log(dist)             -0.902     0.141     -6.39  1.70e-10 -1.18      -0.625
7 I(log(dist)^2)         0.103     0.0105     9.83  8.23e-23  0.0825     0.124
```

3. Coefficient of $[ln(dist)]^2$ shows increasing marginal effect. Relation between $ln(fare)$ and $dist$ becomes positive when the first derivative of $dist$ with respect to $ln(dist)$ equals zero and this happens when $dist = 79.50877$ which does not lie in our data range of $(95, 2724)$
4. For fixed effects, coefficient estimate of $\beta_1 = 0.17$

```
Call:
plm(formula = log(fare) ~ as.factor(year) + bmktshr + log(dist) +
    I(log(dist)^2), data = pdairfare, model = "within")

Balanced Panel: n = 1149, T = 4, N = 4596

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.8812810 -0.0397341  0.0023906  0.0436350  0.9188103

Coefficients:
                     Estimate Std. Error t-value  Pr(>|t|)
as.factor(year)1998 0.0228328  0.0044515  5.1292 3.071e-07 ***
as.factor(year)1999 0.0363819  0.0044495  8.1766 4.061e-16 ***
as.factor(year)2000 0.0977716  0.0044555 21.9441 < 2.2e-16 ***
bmktshr             0.1688590  0.0294101  5.7415 1.020e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    45.169
Residual Sum of Squares: 39.06
R-Squared:      0.13524
Adj. R-Squared: -0.1541
F-statistic: 134.611 on 4 and 3443 DF, p-value: < 2.22e-16
```

5. Avg flights/route & Avg passengers/flight can be in α_i and might be correlated to $bmktshr$.
6. Yes, for both linear and pooled models the $bmktshr$ coefficient estimate is positive.

Question 7 <– loanapp

1. For both logit and linear models, approval rate for whites is 90.84% and for nonwhites is 70.78%

```
Call:
glm(formula = approve ~ white, family = binomial(), data = loanapp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1864   0.4384   0.4384   0.4384   0.8314

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.8847     0.1253   7.061 1.65e-12 ***
white         1.4094     0.1512   9.325  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1480.7  on 1988  degrees of freedom
Residual deviance: 1401.8  on 1987  degrees of freedom
AIC: 1405.8

Number of Fisher Scoring iterations: 5
```

2. Yes, there is significant evidence of discrimination against nonwhites.

```
Call:
glm(formula = approve ~ white + hrat + obrat + loanprc + unem +
    male + married + dep + sch + cosign + chist + pubrec + mortlat1 +
    mortlat2 + vr, family = binomial(), data = loanapp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9549   0.2545   0.3458   0.4768   2.0827

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.80171    0.59467   6.393 1.63e-10 ***
white        0.93776    0.17290   5.424 5.84e-08 ***
hrat         0.01326    0.01288   1.030  0.30313
obrat       -0.05303    0.01128  -4.702 2.58e-06 ***
loanprc     -1.90495    0.46041  -4.138 3.51e-05 ***
unem        -0.06658    0.03281  -2.029  0.04242 *
male        -0.06639    0.20642  -0.322  0.74776
married      0.50328    0.17799   2.828  0.00469 **
dep         -0.09073    0.07333  -1.237  0.21598
sch          0.04123    0.17840   0.231  0.81723
cosign       0.13206    0.44608   0.296  0.76720
chist        1.06658    0.17121   6.230 4.67e-10 ***
pubrec      -1.34067    0.21736  -6.168 6.92e-10 ***
mortlat1    -0.30988    0.46351  -0.669  0.50378
mortlat2    -0.89468    0.56857  -1.574  0.11559
vr          -0.34983    0.15372  -2.276  0.02286 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1476  on 1970  degrees of freedom
Residual deviance: 1201  on 1955  degrees of freedom
  (18 observations deleted due to missingness)
AIC: 1233

Number of Fisher Scoring iterations: 5
```

Question 8 <– alcohol

1. 89.82% of all people are employed and 9.92% of all people have abused alcohol.
2. Alcohol abuse lowers employability by 2.83% and it is significant at 5% which is as expected.

```
> tidyw(model8)
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   0.901    0.00318     284.    0
2 abuse        -0.0283   0.0112       -2.54  0.0112
```

3. For logit model, the sign remains same, but the significance level improved from 5% to 1%
   Logit model suggests that alcohol abuse lowers employability by 28.3%, 10 times of linear model.

```
Call:
glm(formula = employ ~ abuse, family = binomial(), data = alcohol)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1506   0.4566   0.4566   0.4566   0.5219

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.20832    0.03559  62.046   <2e-16 ***
abuse       -0.28337    0.10251  -2.764   0.0057 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6463.8  on 9821  degrees of freedom
Residual deviance: 6456.6  on 9820  degrees of freedom
AIC: 6460.6

Number of Fisher Scoring iterations: 4
```

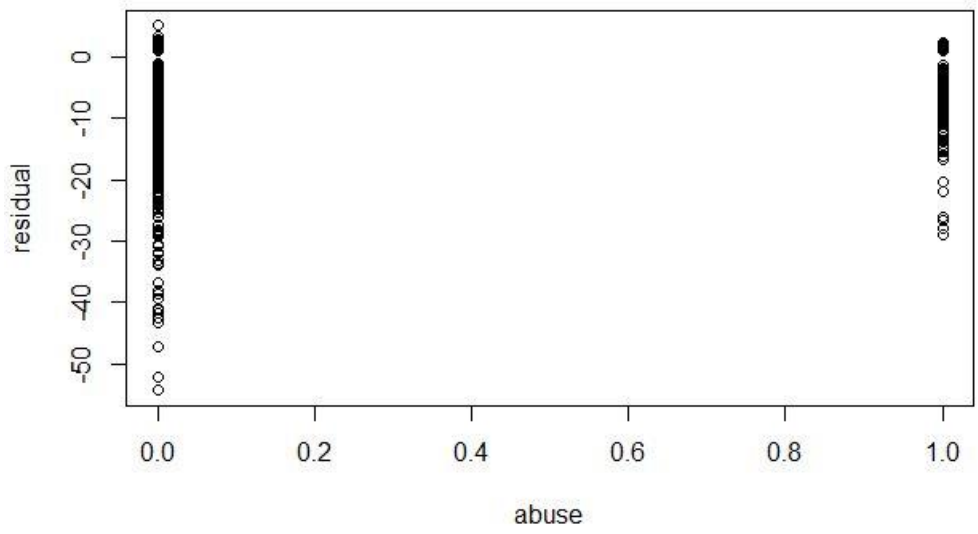
4. Fitted values are 0.901 for $abuse = 0$ and 0.873 for $abuse = 1$. They are same for both models.
5. Coefficient is almost is similar and it is still significant at 5% level.

```
> tidy(model8b)
# A tibble: 17 x 5
   term         estimate std.error statistic  p.value
   <chr>           <dbl>     <dbl>     <dbl>    <dbl>
 1 (Intercept)  0.179     0.0661       2.71   6.69e- 3
 2 abuse       -0.0202    0.00990     -2.04   4.09e- 2
 3 age          0.0160    0.00284      5.63   1.90e- 8
 4 I(age^2)    -0.000231  0.0000342   -6.76   1.46e-11
 5 educ         0.0369    0.00540      6.84   8.34e-12
 6 I(educ^2)   -0.000875  0.000210    -4.17   3.12e- 5
 7 married      0.0574    0.00884      6.48   9.33e-11
 8 famsize      0.00299   0.00218      1.37   1.70e- 1
 9 white        0.0986    0.00866     11.4    7.88e-30
10 northeast    0.0166    0.00917      1.81   7.02e- 2
11 midwest      0.00471   0.00864      0.546  5.85e- 1
12 south        0.0152    0.00834      1.82   6.84e- 2
13 centcity    -0.0155    0.00829     -1.87   6.20e- 2
14 outercity    0.0149    0.00775      1.92   5.51e- 2
15 qrt1        -0.0187    0.00830     -2.25   2.43e- 2
16 qrt2        -0.00674   0.00832     -0.811  4.17e- 1
17 qrt3        -0.00159   0.00840     -0.189  8.50e- 1
```

6. Marginal effect of abuse is still 23% (10 times of linear model) and its t-statistic is -2.14535
7. Considering their significance, it is obvious to include health related variables in our model.

```
> tidy(model8d)
# A tibble: 21 x 5
   term         estimate std.error statistic   p.value
   <chr>           <dbl>     <dbl>     <dbl>     <dbl>
 1 (Intercept) -0.196     0.0646      -3.04  2.38e-  3
 2 exhealth     0.574     0.0195      29.4   9.04e-182
 3 vghealth     0.570     0.0196      29.1   3.85e-178
 4 goodhealth   0.535     0.0197      27.1   2.02e-156
 5 fairhealth   0.393     0.0221      17.8   5.55e- 70
 6 abuse       -0.0157    0.00945     -1.66  9.68e-  2
 7 age          0.0162    0.00270      6.00  2.02e-  9
 8 I(age^2)    -0.000215  0.0000326   -6.59  4.56e- 11
 9 educ         0.0168    0.00519      3.23  1.23e-  3
10 I(educ^2)   -0.000353  0.000201    -1.75  7.96e-  2
# ... with 11 more rows
```

8. As $abuse$ goes from 0 to 1, $residuals$ decrease implying that $abuse$ and $residual$ are correlated. $mothalc$ & $fathalc$ can't be used as instrument variables as they have low correlation with $abuse$.



Question 9 <– fertil1

1. Fertility dropped by 19.9% in the year 1982

```
Call:
glm(formula = kids ~ educ + age + I(educ^2) + black + east +
    northcen + west + farm + othrural + town + smcity + y74 +
    y76 + y78 + y80 + y82 + y84, family = poisson(), data = fertil1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.83050  -0.68862  -0.04099   0.55084   2.77002

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.529391   0.293052   1.806  0.07084 .
educ         0.073904   0.038775   1.906  0.05665 .
age          0.007318   0.003131   2.337  0.01942 *
I(educ^2)   -0.004974   0.001528  -3.255  0.00114 **
black        0.349381   0.061034   5.724 1.04e-08 ***
east         0.069049   0.052838   1.307  0.19128
northcen     0.114356   0.047852   2.390  0.01686 *
west         0.053123   0.065857   0.807  0.41987
farm        -0.015461   0.057470  -0.269  0.78791
othrural    -0.048372   0.068889  -0.702  0.48257
town         0.031044   0.048547   0.639  0.52252
smcity       0.078378   0.061528   1.274  0.20271
y74          0.082107   0.062997   1.303  0.19246
y76         -0.039072   0.067420  -0.580  0.56223
y78         -0.029722   0.068552  -0.434  0.66461
y80         -0.015803   0.069034  -0.229  0.81893
y82         -0.199045   0.067367  -2.955  0.00313 **
y84         -0.222174   0.069405  -3.201  0.00137 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1331.1  on 1128  degrees of freedom
Residual deviance: 1185.9  on 1111  degrees of freedom
AIC: 4178

Number of Fisher Scoring iterations: 5
```

2. Estimated percentage difference in fertility for black woman nonblack woman is 34.94%
3. For poisson model, $R^2 = 0.11898$ and for linear model, $R^2 = 0.1248$ both of which are similar.