BUAN 6356.002
Problem Set 4
Huyen Nguyen (Htn180001)

Question 1:
Step: AIC = 697.58
BIC(model3) = 951.7885

price ~ bdrms + lotsize + sqrft + colonial + I(lotsize^2) + bdrms:lotsize + bdrms:colonial

	Df	Sum of Sq	RSS	AIC
<none>		162321	697.58	
- bdrms:colonial	1	10358	172678	698.54
- bdrms:lotsize	1	55504	217825	718.98
- I(lotsize^2)	1	84763	247083	730.07
- sqrft	1	95612	257932	733.86

model3 <- lm(price ~ bdrms + lotsize + sqrft + colonial + I(lotsize^2) + bdrms:lotsize + bdrms:colonial, data=hprice1)

```
Call:
lm(formula = price ~ bdrms + lotsize + sqrft + colonial + I(lotsize^2) +
    bdrms:lotsize + bdrms:colonial, data = hprice1)

Residuals:
     Min       1Q   Median       3Q      Max
-112.311  -22.249   -5.863   19.136  204.828

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.473e+02  6.777e+01   3.649 0.000467 ***
bdrms          -6.190e+01  1.968e+01  -3.145 0.002334 **
lotsize        -9.849e-03  4.561e-03  -2.159 0.033817 *
sqrft           8.186e-02  1.192e-02   6.865 1.29e-09 ***
colonial       -9.379e+01  5.219e+01  -1.797 0.076086 .
I(lotsize^2)   -1.221e-07  1.888e-08  -6.463 7.46e-09 ***
bdrms:lotsize   5.657e-03  1.082e-03   5.230 1.33e-06 ***
bdrms:colonial  3.472e+01  1.537e+01   2.259 0.026579 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45.04 on 80 degrees of freedom
Multiple R-squared:  0.8232,  Adjusted R-squared:  0.8077
F-statistic:  53.2 on 7 and 80 DF,  p-value: < 2.2e-16
```

Question 2:
Step: AIC = -5211.47
BIC(model4) = 6720.662

colgpa ~ sat + tothrs + athlete + verbmath + hsize + hsrank + hsperc + female + white + black + I(sat^2) + I(hsrank^2) + I(hsperc^2) + sat:tothrs + sat:hsperc + sat:white + tothrs:hsperc

```
    + athlete:verbmath + athlete:hsize + athlete:hsrank + athlete:hsperc + athlete:black +
    verbmath:white + hsize:hsrank + hsrank:hsperc + hsrank:white + female:white +
    female:black

Residuals:
     Min       1Q   Median       3Q      Max
-2.49634 -0.33448  0.01309  0.37339  2.06413

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.870e+00  4.659e-01   4.013 6.10e-05 ***
sat              5.402e-04  7.491e-04   0.721 0.470870
tothrs           9.020e-03  1.919e-03   4.702 2.66e-06 ***
athlete         -4.512e-01  2.434e-01  -1.854 0.063869 .
verbmath        -5.004e-01  1.961e-01  -2.551 0.010764 *
hsize           -1.012e-02  1.119e-02  -0.905 0.365758
hsrank          -3.655e-03  1.657e-03  -2.205 0.027475 *
hsperc          -1.334e-02  5.377e-03  -2.480 0.013173 *
female          -1.353e-01  1.266e-01  -1.068 0.285378
white            1.601e-01  3.212e-01   0.498 0.618182
black           -4.113e-01  1.001e-01  -4.108 4.06e-05 ***
I(sat^2)         9.725e-07  3.551e-07   2.739 0.006197 **
I(hsrank^2)     -1.217e-05  4.235e-06  -2.874 0.004077 **
I(hsperc^2)      1.601e-04  4.210e-05   3.804 0.000145 ***
sat:tothrs      -8.493e-06  1.763e-06  -4.816 1.51e-06 ***
sat:hsperc      -1.241e-05  4.197e-06  -2.957 0.003129 **
sat:white       -5.708e-04  2.547e-04  -2.241 0.025079 *
tothrs:hsperc    7.487e-05  1.541e-05   4.858 1.23e-06 ***
athlete:verbmath 4.379e-01  2.517e-01   1.740 0.081960 .
athlete:hsize    6.010e-02  3.755e-02   1.601 0.109522
athlete:hsrank  -2.199e-03  9.438e-04  -2.330 0.019846 *
athlete:hsperc   4.916e-03  3.469e-03   1.417 0.156530
athlete:black    2.166e-01  1.166e-01   1.858 0.063226 .
verbmath:white   4.404e-01  2.030e-01   2.170 0.030072 *
hsize:hsrank     6.594e-04  1.992e-04   3.310 0.000942 ***
hsrank:hsperc    7.992e-05  2.974e-05   2.687 0.007233 **
hsrank:white    -1.143e-03  5.703e-04  -2.005 0.045016 *
female:white     2.899e-01  1.279e-01   2.267 0.023433 *
female:black     2.713e-01  1.459e-01   1.860 0.062953 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5308 on 4108 degrees of freedom
Multiple R-squared:  0.3549,  Adjusted R-squared:  0.3505
F-statistic: 80.71 on 28 and 4108 DF,  p-value: < 2.2e-16
```

Question 3:

```
Step:  AIC = 8886.25
BIC(model8) = 10079.29
salary ~ nl + games + I(teamsal^2) + I(years^2) + I(games^2) +
   atbats + runs + hits + doubles + triples + hruns + rbis +
```

```
    bavg + bb + so + sbases + fldperc + frstbase + scndbase + 
    thrdbase + outfield + yrsallst + whitepop + hisppop + hrunsyr + 
    slugavg + blckpb + atbats:runs + atbats:hits + atbats:doubles + 
    atbats:triples + atbats:bb + atbats:sbases + runs:doubles + 
    runs:triples + runs:bavg + runs:fldperc + hits:doubles + 
    hits:triples + hits:bavg + hits:bb + hits:so + hits:fldperc + 
    doubles:bavg + doubles:sbases + doubles:fldperc + triples:hruns + 
    triples:rbis + triples:bb + triples:so + triples:sbases + 
    hruns:rbis + hruns:bavg + hruns:bb + rbis:bavg + rbis:bb + 
    rbis:so + rbis:sbases + bavg:bb + bavg:sbases + bavg:fldperc + 
    bb:so + bb:fldperc + so:sbases + so:fldperc

Residuals:
     Min       1Q   Median       3Q      Max 
-2448836  -288692   -17834   272881  1837852 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -2.797e+06  5.030e+06  -0.556 0.578647    
nl              1.343e+05  8.211e+04   1.635 0.103166    
games           2.786e+03  1.395e+03   1.997 0.046862 *  
I(teamsal^2)    1.739e-10  9.496e-11   1.831 0.068194 .  
I(years^2)     -9.437e+03  2.837e+03  -3.326 0.001005 ** 
I(games^2)     -2.974e+00  7.902e-01  -3.764 0.000206 ***
atbats          9.814e+03  2.669e+03   3.676 0.000287 ***
runs            3.079e+05  9.074e+04   3.393 0.000797 ***
hits           -3.860e+05  5.959e+04  -6.479 4.51e-10 ***
doubles         5.555e+05  2.140e+05   2.596 0.009966 ** 
triples        -4.531e+04  1.492e+04  -3.037 0.002626 ** 
hruns          -2.138e+05  4.729e+04  -4.521 9.32e-06 ***
rbis            9.863e+04  2.251e+04   4.381 1.70e-05 ***
bavg            2.352e+04  1.538e+04   1.529 0.127501    
bb             -3.999e+04  5.335e+04  -0.750 0.454133    
so              4.105e+04  1.928e+04   2.129 0.034158 *  
sbases         -5.381e+04  2.116e+04  -2.543 0.011571 *  
fldperc         3.570e+03  5.369e+03   0.665 0.506672    
frstbase       -1.954e+05  1.530e+05  -1.277 0.202633    
scndbase       -2.137e+05  1.462e+05  -1.461 0.145088    
thrdbase       -2.804e+05  1.598e+05  -1.755 0.080412 .  
outfield       -2.704e+05  1.200e+05  -2.254 0.025019 *  
yrsallst        3.209e+05  4.928e+04   6.512 3.72e-10 ***
whitepop       -3.267e-02  2.260e-02  -1.445 0.149515    
hisppop         7.431e-02  5.454e-02   1.363 0.174145    
hrunsyr        -7.138e+04  2.266e+04  -3.150 0.001823 ** 
slugavg         3.006e+04  8.170e+03   3.679 0.000284 ***
blckpb          8.473e+03  3.928e+03   2.157 0.031896 *  
atbats:runs    -7.036e+00  1.889e+00  -3.726 0.000238 ***
atbats:hits     1.789e+00  9.110e-01   1.964 0.050633 .  
atbats:doubles  3.665e+01  8.842e+00   4.145 4.58e-05 ***
atbats:triples  9.562e+01  2.668e+01   3.584 0.000403 ***
atbats:bb      -1.043e+01  3.561e+00  -2.928 0.003704 ** 
```

```
atbats:sbases    3.714e+00  1.267e+00   2.931 0.003670 **
runs:doubles     1.187e+02  4.103e+01   2.893 0.004137 **
runs:triples     5.508e+02  1.126e+02   4.892 1.74e-06 ***
runs:bavg       -3.147e+02  1.095e+02  -2.875 0.004367 **
runs:fldperc    -2.360e+02  8.733e+01  -2.702 0.007340 **
hits:doubles    -1.789e+02  3.872e+01  -4.621 5.97e-06 ***
hits:triples    -5.362e+02  1.110e+02  -4.832 2.29e-06 ***
hits:bavg        3.567e+02  6.744e+01   5.289 2.58e-07 ***
hits:bb          3.795e+01  1.295e+01   2.931 0.003674 **
hits:so          1.113e+01  2.614e+00   4.259 2.86e-05 ***
hits:fldperc     2.659e+02  5.384e+01   4.939 1.40e-06 ***
doubles:bavg     5.738e+02  1.808e+02   3.174 0.001682 **
doubles:sbases  -8.699e+01  2.790e+01  -3.118 0.002024 **
doubles:fldperc -7.360e+02  2.129e+02  -3.456 0.000638 ***
triples:hruns   -6.127e+02  2.087e+02  -2.936 0.003619 **
triples:rbis     2.875e+02  9.977e+01   2.881 0.004288 **
triples:bb      -1.903e+02  4.300e+01  -4.426 1.40e-05 ***
triples:so      -7.087e+01  3.195e+01  -2.218 0.027378 *
triples:sbases  -1.396e+02  4.805e+01  -2.905 0.003988 **
hruns:rbis       4.224e+01  8.696e+00   4.857 2.05e-06 ***
hruns:bavg       8.991e+02  1.818e+02   4.945 1.35e-06 ***
hruns:bb        -6.783e+01  1.707e+01  -3.974 9.13e-05 ***
rbis:bavg       -3.666e+02  8.621e+01  -4.252 2.94e-05 ***
rbis:bb          1.373e+01  7.357e+00   1.867 0.063022 .
rbis:so         -2.717e+01  4.675e+00  -5.813 1.77e-08 ***
rbis:sbases      1.827e+01  8.239e+00   2.217 0.027468 *
bavg:bb         -1.177e+02  6.211e+01  -1.895 0.059225 .
bavg:sbases      2.387e+02  8.139e+01   2.932 0.003662 **
bavg:fldperc    -3.210e+01  1.715e+01  -1.872 0.062323 .
bb:so            1.467e+01  3.653e+00   4.015 7.74e-05 ***
bb:fldperc       7.144e+01  5.188e+01   1.377 0.169656
so:sbases       -2.306e+01  4.367e+00  -5.281 2.69e-07 ***
so:fldperc      -4.083e+01  1.943e+01  -2.102 0.036499 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 644100 on 264 degrees of freedom
Multiple R-squared:  0.8282,  Adjusted R-squared:  0.7859
F-statistic: 19.57 on 65 and 264 DF,  p-value: < 2.2e-16
```

Question 4:

1. ln(rentit) = - 0.568806 + 0.262227 y90t + 0.040686 ln(popit) + 0.571446 ln(avgincit) + 0.0050436 pctstuit

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

R-Squared: 0.86128

β_1 -hat = 0.262227, p-value = 8.781e-12 < 0.01 ((significant at 1% level)

-> Rental prices in 1990 has increased 26.2227% compare to rents in 1980, other factores remaining constant

β_3 -hat = 0.0050436, p-value = 2.401e-06 < 0.05 (significant at 5% level)

-> 1% increase in student population (as a percentage of
city population during the school year) is associated with 0.50436% point in rental prices, other factores remaining constant

2. The standard errors from part 1 are not valid, unless we thing ai does not really appear in the equation. If ai is in the error term, the errors across the two time periods for each city are positively correlated, and this invalidates the usual OLS standard errors and t statistics.

3. ln(rentit) = 0.3855214 y90t + 0.0722457 ln(popit) + 0.3099604 ln(avgincit) + 0.0112033 pctstuit

β_3 -hat = 0.0112033, p-value = 0.008726 < 0.01 (significant at 1% level)
-> 1% point increase in student population results in 1.12033% point in rental prices, other factores remaining constant
Compared to result in part 1, we can say that the relative size of the student population affects rental prices.

4. Fixed effect model:
plm(formula = log(rent) ~ as.factor(year) + log(pop) + log(avginc) + pctstu, data = pdrental, model = "within")

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

Residuals:
      Min.     1st Qu.      Median     3rd Qu.       Max.
-1.1891e-01 -2.9559e-02  7.8236e-16  2.9559e-02  1.1891e-01

Coefficients:
                Estimate Std. Error t-value  Pr(>|t|)
as.factor(year)90 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:    10.383
Residual Sum of Squares: 0.24368
R-Squared:      0.97653
Adj. R-Squared: 0.95032
```

-> This result is similar in part 3.

Question 5:
1. If past executions of convicted murderers have a deterrent effect, $\beta1 < 0$
If a better economy means less crime in general, this would imply $\beta2 > 0$
2. No evidence of deterrent effect, as the coefficient on exec is actually positive.
3. Fixed effect model:
mrdrte1 = - 0.10010 exec1 - 0.27749 unem1
There's no evidence of deterrent effect as exec1 is not significant at 5% level
4. Heteroskedasticity-robust standard error for the estimation in part 2:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.866888	0.425132	2.0391	0.0441044 *
exec1	-0.131354	0.034322	-3.8271	0.0002272 ***
unem1	-0.044793	0.125158	-0.3579	0.7211874

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5. Texas had by far the largest value of exec = 34. The next highest state was Virginia, with 11. These are three-year totals

6. plm(formula = mrdrte1 ~ exec1 + unem1, data = dataNoTX, model = "pooling")
Coefficients:

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	0.867058	0.436286	1.9874	0.0497 *
exec1	-0.130852	0.288632	-0.4534	0.6513
unem1	-0.044913	0.266750	-0.1684	0.8666

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1847.3
Residual Sum of Squares: 1842.4
R-Squared: 0.0026754
Adj. R-Squared: -0.017888
F-statistic: 0.130103 on 2 and 97 DF, p-value: 0.87816

SE:
t test of coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.867058	0.432720	2.0037	0.04789 *
exec1	-0.130852	0.127042	-1.0300	0.30558
unem1	-0.044913	0.123157	-0.3647	0.71614

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Now the estimated deterrent effect is smaller. Perhaps more importantly, the standard error on Δexeci has increased by a substantial amount. This happens because when we drop Texas, we lose much of the variation in the key explanatory variable, Δexeci.

7.
plm(formula = mrdrte ~ exec + unem + d93 + d90, data = pdmurder1, model = "within")

Coefficients:

	Estimate	Std. Error	t-value	Pr(>|t|)
exec	-0.13832	0.17701	-0.7815	0.43642
unem	0.22132	0.29638	0.7467	0.45701
d93	1.73324	0.70044	2.4745	0.01506 *
d90	1.55621	0.74533	2.0880	0.03939 *

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

The size of the deterrent effect is actually slightly larger than when 1987 is not used. However, the t statistic is only about −0.78. Thus, while the magnitude of the effect is similar, the statistical significance is not. It is somewhat odd that adding another year of data causes the standard error on the exec coefficient to increase nontrivially.

Question 6:

1) log(fare) = 6.2495769 + 0.3526893 bmktshr - 0.8991659 log(dist) + 0.1027463 (log(dist)^2)
R-Squared: 0.39892
If Δconcern = 0.10 then Δlfare = 0.360 (0.10) = 0.036, which means air fare is estimated to be about 3.6% higher

2) 95% CI obtained using the usual OLS standard error is 0.301 to 0.419. It is probably not reliable because the validity of this standard error requires the composite error to have no serial correlation.
The fully robust 95% CI is 0.245 to 0.475, a bit wider than the usual CI.
The wider CI is appropriate, as the neglected serial correlation introduces uncertainty into our parameter estimators.

3) The quadratic has a U-shape, and the turning point is about 0.902 / [2(0.103)] ≈ 4.38. This is the value of log(dist) where the slope becomes positive.
The value of dist is exp(4.38) ≈ 80
The shortest distance in the data set is 95 miles, so the turning point is outside the range of the data.

4) plm(formula = log(fare) ~ bmktshr + log(dist) + I(log(dist)^2),
   data = pairfare, model = "within")

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

Residuals:

```
      Min.     1st Qu.      Median     3rd Qu.        Max.
-0.92015574 -0.05117015  0.00019632  0.05335228  0.97909735
```

Coefficients:

```
          Estimate   Std. Error   t-value    Pr(>|t|)
bmktshr   0.103051   0.031242     3.2985     0.000982 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Total Sum of Squares: 45.169
Residual Sum of Squares: 45.026
R-Squared: 0.0031473

The FE estimate is 0.169

5) Other factors such as population, education levels, types of employers, ... could affect demand for air travel. Each of these can be time-varying, although, over a short stretch of time, they might be roughly constant. These factors could certainly be correlated with concentration.

6) Higher concentration on a route increases airfares. I would go with the FE estimate, 0.169, which allows for concentration to be correlated with all time-constant features that affect costs and demand.

Question 7:

1. 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
```

Probability approve for white 0.9083879
Probability approve for non-white 0.7077922

2. Yes, it is still discriminate because p value of white < 0.05

Question 8:

1. The fraction of men employed is 0.8981877 and the fraction abusing alcohol is about 0.09916514.
2. The simple regression, with heteroskedasticity-robust standard errors:

```
employ =   0.900995   -    0.028305 abuse
          (0.003214)      (0.010206)
R-squared:  0.0007826, n = 9,822
```

The intercept implies that if a man does not abuse alcohol, the chance of being employed is 0.901.
If a man does abuse alcohol, the chance is smaller by 0.028, so 0.87 as the final value.
It is significant model
3. Yes, we get the same sign and statistical significance as in part 2 but the coefficient is smaller.
4. The fitted values from the LPM and probit must be the same. In each case, there are only two fitted values, estimating the probability of being employed with abuse = 0 and abuse = 1. These estimates are the same no matter which model we use (linear, probit, logit, or some other binary response model). The two fitted values were given in the answer to part (2) as 0.901 and 0.873, respectively.

5. Coefficient of abuse is smaller and model still siginificant. With the many controls added, the magnitude of the coefficient on abuse falls some somewhat to about −0.020 (robust t = −1.87). So, the effect on alcohol abuse on the employment probability is smaller but it is still marginally statistically significant (robust two-sided p-value = 0.061).

6. Coefficient of abuse is not the same as linear model. It is very close though.

```
glm(formula = employ ~ abuse + age + I(age^2) + educ + I(educ^2) +
    married + famsize + white + northeast + midwest + south +
    centcity + outercity + qrt1 + qrt2 + qrt3, family = binomial,
    data = alcohol)

Coefficients:
             Estimate  Std. Error  z value  Pr(>|z|)
(Intercept) -3.1140166  0.7082494  -4.397 1.10e-05 ***
abuse       -0.2295500  0.1069989  -2.145   0.0319 *
age          0.1500715  0.0317560   4.726 2.29e-06 ***
I(age^2)    -0.0021809  0.0003757  -5.805 6.42e-09 ***
educ         0.1107005  0.0490033   2.259   0.0239 *
I(educ^2)    0.0017702  0.0021226   0.834   0.4043
married      0.6070615  0.0960064   6.323 2.56e-10 ***
famsize      0.0396923  0.0255054   1.556   0.1197
white        0.8322018  0.0842393   9.879  < 2e-16 ***
northeast    0.2534729  0.1098085   2.308   0.0210 *
midwest      0.1003212  0.1005632   0.998   0.3185
south        0.2104093  0.0969142   2.171   0.0299 *
centcity    -0.1500694  0.0937417  -1.601   0.1094
outercity    0.2054375  0.0949079   2.165   0.0304 *
qrt1        -0.1995331  0.0976319  -2.044   0.0410 *
qrt2        -0.0890819  0.0989791  -0.900   0.3681
qrt3        -0.0150204  0.1014983  -0.148   0.8824
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

7. It is not clear that other health indicators should be controlled for. If they are included, it could be a case of “overcontrolling” because certain health problems may be the result of alcohol abuse. If we hold them fixed when making comparisons we could underestimate the total effect of alcohol abuse on employment. On the other hand, health problems not caused by alcohol abuse could be correlated with both employment and alcohol abuse, in which case we would be attributing too much to alcohol abuse.

8. The indicator of alcohol abuse may be correlated with unobserved factors that affect employment. Certain kinds of health issues were already mentioned in part (7). Depression and low self-esteem, not being motivated are other examples. Indicators for whether one’s parents abused alcohol are reasonable attempts at finding instruments for abuse, but they likely are not completely unrelated to unobserved factors affecting employment.
mothalc and fathalc are significant predictors of abuse, and the signs of the two coefficients make sense.

Question 9:
1. Coeff of y82 is - 0.1926076
Since estimate of $\beta < 0$, a woman at age 82 has the 0.1926076 less number of children born.

```
glm(formula = kids ~ educ + age + I(age^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
-2.91598  -0.67884  -0.04123   0.55625   2.49302

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.0604626  1.2106974  -2.528 0.011476 *
educ        -0.0482027  0.0072302  -6.667 2.61e-11 ***
age          0.2044553  0.0547527   3.734 0.000188 ***
I(age^2)    -0.0022290  0.0006171  -3.612 0.000304 ***
black        0.3603475  0.0610748   5.900 3.63e-09 ***
east         0.0878001  0.0526729   1.667 0.095535 .
northcen     0.1417221  0.0475056   2.983 0.002852 **
west         0.0795427  0.0656991   1.211 0.226006
farm        -0.0148484  0.0575534  -0.258 0.796412
othrural    -0.0572939  0.0691574  -0.828 0.407412
town         0.0306807  0.0485793   0.632 0.527675
smcity       0.0741129  0.0615484   1.204 0.228535
y74          0.0932809  0.0630849   1.479 0.139232
y76         -0.0287888  0.0675828  -0.426 0.670123
y78         -0.0156856  0.0686754  -0.228 0.819334
y80         -0.0196524  0.0689821  -0.285 0.775727
y82         -0.1926076  0.0674991  -2.853 0.004324 **
y84         -0.2143735  0.0694641  -3.086 0.002028 **
---
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: 1184.3  on 1111  degrees of freedom
AIC: 4176.5

Number of Fisher Scoring iterations: 5
```

2. The estimated percentage difference in fertility between a black woman and a nonblack woman, holding other factors fixed is 0.3603475

3. R square is 12%
R square in LM is higher