Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Group Lab 3

U.S. traffic fatalities: 1980-2004

In this lab, you are asked to answer the question **"Do changes in traffic laws affect traffic fatalities?"** To do so, you will conduct the tasks specified below using the data set *driving.Rdata*, which includes 25 years of data that cover changes in various state drunk driving, seat belt, and speed limit laws.

Specifically, this data set contains data for the 48 continental U.S. states from 1980 through 2004. Various driving laws are indicated in the data set, such as the alcohol level at which drivers are considered legally intoxicated. There are also indicators for "per se" laws—where licenses can be revoked without a trial—and seat belt laws. A few economics and demographic variables are also included. The description of the each of the variables in the dataset is come with the dataste.

```
# Import libraries
library(ggplot2)
library(gridExtra)
library(corrplot)
library(gplots)
library(plm)
```

Exercises:

1. (30%) Load the data. Provide a description of the basic structure of the dataset, as we have done throughout the semester. Conduct a very thorough EDA, which should include both graphical and tabular techniques, on the dataset, including both the dependent variable *totfatrte* and the potential explanatory variables. You need to write a detailed narrative of your observations of your EDA. *Reminder: giving an "output dump" (i.e. providing a bunch of graphs and tables without description and hoping your audience will interpret them) will receive a zero in this exercise.*

```
# Clean up the work space before we begin
#rm(list = ls())

load("./driving.Rdata")
drivingdf <- data

desc
```

```
##        variable                                           label
## 1          year                               1980 through 2004
## 2         state              48 continental states, alphabetical
## 3          sl55                               speed limit == 55
## 4          sl65                               speed limit == 65
## 5          sl70                               speed limit == 70
## 6          sl75                               speed limit == 75
## 7        slnone                                  no speed limit
## 8      seatbelt     =0 if none, =1 if primary, =2 if secondary
## 9        minage                           minimum drinking age
## 10      zerotol                             zero tolerance law
## 11          gdl                  graduated drivers license law
## 12        bac10                         blood alcohol limit .10
## 13        bac08                         blood alcohol limit .08
## 14        perse administrative license revocation (per se law)
## 15       totfat                        total traffic fatalities
## 16      nghtfat                      total nighttime fatalities
## 17      wkndfat                        total weekend fatalities
## 18    totfatpvm           total fatalities per 100 million miles
## 19   nghtfatpvm       nighttime fatalities per 100 million miles
## 20   wkndfatpvm         weekend fatalities per 100 million miles
## 21     statepop                                 state population
## 22    totfatrte         total fatalities per 100,000 population
## 23   nghtfatrte     nighttime fatalities per 100,000 population
## 24   wkndfatrte        weekend accidents per 100,000 population
## 25   vehicmiles               vehicle miles traveled, billions
## 26         unem                      unemployment rate, percent
## 27    perc14_24            percent population aged 14 through 24
## 28     sl70plus                            sl70 + sl75 + slnone
## 29       sbprim                      =1 if primary seatbelt law
## 30      sbsecon                    =1 if secondary seatbelt law
## 31          d80                                =1 if year == 1980
## 32          d81
## 33          d82
## 34          d83
## 35          d84
## 36          d85
## 37          d86
## 38          d87
## 39          d88
## 40          d89
## 41          d90
## 42          d91
## 43          d92
## 44          d93
## 45          d94
## 46          d95
## 47          d96
```

```
## 48          d97
## 49          d98
## 50          d99
## 51          d00
## 52          d01
## 53          d02
## 54          d03
## 55          d04                             =1 if year == 2004
## 56 vehicmilespc
```

```
dim(drivingdf)
```

```
## [1] 1200   56
```

```
head(drivingdf)
```

```
##   year state sl55 sl65 sl70 sl75 slnone seatbelt minage zerotol gdl bac10 bac08
## 1 1980     1    1    0    0    0      0        0     18       0   0     1     0
## 2 1981     1    1    0    0    0      0        0     18       0   0     1     0
## 3 1982     1    1    0    0    0      0        0     18       0   0     1     0
## 4 1983     1    1    0    0    0      0        0     18       0   0     1     0
## 5 1984     1    1    0    0    0      0        0     18       0   0     1     0
## 6 1985     1    1    0    0    0      0        0     20       0   0     1     0
##   perse totfat nghtfat wkndfat totfatpvm nghtfatpvm wkndfatpvm statepop
## 1     0    940     422     236      3.20      1.437      0.803  3893888
## 2     0    933     434     248      3.35      1.558      0.890  3918520
## 3     0    839     376     224      2.81      1.259      0.750  3925218
## 4     0    930     397     223      3.00      1.281      0.719  3934109
## 5     0    932     421     237      2.83      1.278      0.720  3951834
## 6     0    882     358     224      2.51      1.019      0.637  3972527
##   totfatrte nghtfatrte wkndfatrte vehicmiles unem perc14_24 sl70plus sbprim
## 1     24.14      10.84       6.06   29.37500  8.8      18.9        0      0
## 2     24.07      11.08       6.33   27.85200 10.7      18.7        0      0
## 3     21.37       9.58       5.71   29.85765 14.4      18.4        0      0
## 4     23.64      10.09       5.67   31.00000 13.7      18.0        0      0
## 5     23.58      10.65       6.00   32.93286 11.1      17.6        0      0
## 6     22.20       9.01       5.64   35.13944  8.9      17.3        0      0
##   sbsecon d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96
## 1       0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2       0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 3       0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 4       0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
## 5       0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0
## 6       0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0
##   d97 d98 d99 d00 d01 d02 d03 d04 vehicmilespc
## 1   0   0   0   0   0   0   0   0     7543.874
```

```
## 2    0    0    0    0    0    0    0    0     7107.785
## 3    0    0    0    0    0    0    0    0     7606.622
## 4    0    0    0    0    0    0    0    0     7879.802
## 5    0    0    0    0    0    0    0    0     8333.562
## 6    0    0    0    0    0    0    0    0     8845.614
```

```
tail(drivingdf)
```

```
##      year state sl55 sl65 sl70 sl75 slnone seatbelt minage zerotol gdl bac10
## 1195 1999    51    0    0    0    1      0        2     21       1   0   1.0
## 1196 2000    51    0    0    0    1      0        2     21       1   0   1.0
## 1197 2001    51    0    0    0    1      0        2     21       1   0   1.0
## 1198 2002    51    0    0    0    1      0        2     21       1   0   0.5
## 1199 2003    51    0    0    0    1      0        2     21       1   0   0.0
## 1200 2004    51    0    0    0    1      0        2     21       1   0   0.0
##      bac08 perse totfat nghtfat wkndfat totfatpvm nghtfatpvm wkndfatpvm
## 1195   0.0     1    189      73      32      2.42      0.935      0.410
## 1196   0.0     1    152      59      37      1.88      0.730      0.458
## 1197   0.0     1    186      76      49      2.16      0.883      0.569
## 1198   0.5     1    176      60      29      1.95      0.665      0.321
## 1199   1.0     1    165      62      32      1.79      0.673      0.347
## 1200   1.0     1    164      67      31      1.77      0.723      0.335
##      statepop totfatrte nghtfatrte wkndfatrte vehicmiles unem perc14_24
## 1195   491780     38.43      14.84   6.510000   7.809920  4.9      16.6
## 1196   493782     30.78      11.95   7.490000   8.085110  3.9      16.1
## 1197   493754     37.67      15.39   9.920000   8.611111  3.9      15.5
## 1198   498830     35.28      12.03   5.809999   9.025640  4.2      15.3
## 1199   501242     32.92      12.37   6.380000   9.217880  4.4      15.1
## 1200   507000     32.35      13.21   6.110000   9.266000  3.7      14.9
##      sl70plus sbprim sbsecon d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91
## 1195        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1196        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1197        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1198        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1199        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1200        1      0       1   0   0   0   0   0   0   0   0   0   0   0   0
##      d92 d93 d94 d95 d96 d97 d98 d99 d00 d01 d02 d03 d04 vehicmilespc
## 1195   0   0   0   0   0   0   0   1   0   0   0   0   0     15880.92
## 1196   0   0   0   0   0   0   0   0   1   0   0   0   0     16373.84
## 1197   0   0   0   0   0   0   0   0   0   1   0   0   0     17440.08
## 1198   0   0   0   0   0   0   0   0   0   0   1   0   0     18093.62
## 1199   0   0   0   0   0   0   0   0   0   0   0   1   0     18390.08
## 1200   0   0   0   0   0   0   0   0   0   0   0   0   1     18276.13
```

2. (15%) How is the our dependent variable of interest *totfatrte* defined? What is the average of this variable in each of the years in the time period covered in this dataset? Estimate a linear regression model of *totfatrte* on a set of dummy variables for the years 1981 through 2004.

> What does this model explain? Describe what you find in this model. Did driving become safer over this period? Please provide a detailed explanation.

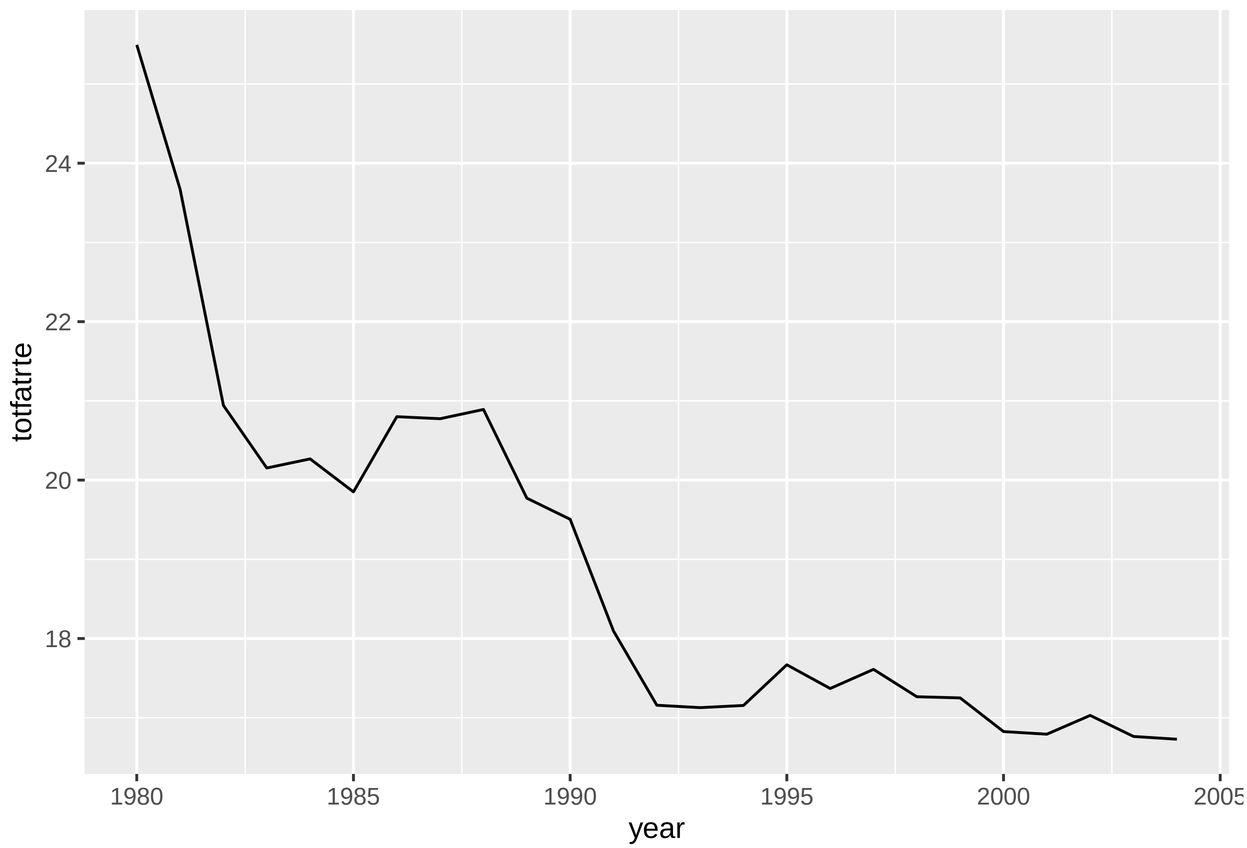
totfatrte defines the total fatalities per 100,000 population.<>

```r
yearlyavg <- aggregate(totfatrte~year, drivingdf, mean)

# Printing the yearly average for total fatality rate
yearlyavg
```

```
##    year totfatrte
## 1  1980  25.49458
## 2  1981  23.67021
## 3  1982  20.94250
## 4  1983  20.15292
## 5  1984  20.26750
## 6  1985  19.85146
## 7  1986  20.80042
## 8  1987  20.77479
## 9  1988  20.89167
## 10 1989  19.77229
## 11 1990  19.50521
## 12 1991  18.09479
## 13 1992  17.15792
## 14 1993  17.12771
## 15 1994  17.15521
## 16 1995  17.66854
## 17 1996  17.36938
## 18 1997  17.61062
## 19 1998  17.26542
## 20 1999  17.25042
## 21 2000  16.82562
## 22 2001  16.79271
## 23 2002  17.02958
## 24 2003  16.76354
## 25 2004  16.72896
```

```r
# Plotting the yearly total fatality rate
ggplot(yearlyavg) +
  geom_line(
    mapping = aes(x = year, y = totfatrte)
)
```




```
lm.fit1 <- lm(totfatrte ~ d81+d82+d83+d84+d85+d86+d87+d88+d89+
                        d90+d91+d92+d93+d94+d95+d96+d97+d98+d99+
                        d00+d01+d02+d03+d04, data=drivingdf)
summary(lm.fit1)
```

```
##
## Call:
## lm(formula = totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 +
##     d87 + d88 + d89 + d90 + d91 + d92 + d93 + d94 + d95 + d96 +
##     d97 + d98 + d99 + d00 + d01 + d02 + d03 + d04, data = drivingdf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9302  -4.3468  -0.7305   3.7488  29.6498
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.4946     0.8671  29.401  < 2e-16 ***
## d81          -1.8244     1.2263  -1.488 0.137094
## d82          -4.5521     1.2263  -3.712 0.000215 ***
## d83          -5.3417     1.2263  -4.356 1.44e-05 ***
## d84          -5.2271     1.2263  -4.263 2.18e-05 ***
```

```
## d85          -5.6431     1.2263  -4.602 4.64e-06 ***
## d86          -4.6942     1.2263  -3.828 0.000136 ***
## d87          -4.7198     1.2263  -3.849 0.000125 ***
## d88          -4.6029     1.2263  -3.754 0.000183 ***
## d89          -5.7223     1.2263  -4.666 3.42e-06 ***
## d90          -5.9894     1.2263  -4.884 1.18e-06 ***
## d91          -7.3998     1.2263  -6.034 2.14e-09 ***
## d92          -8.3367     1.2263  -6.798 1.68e-11 ***
## d93          -8.3669     1.2263  -6.823 1.43e-11 ***
## d94          -8.3394     1.2263  -6.800 1.66e-11 ***
## d95          -7.8260     1.2263  -6.382 2.51e-10 ***
## d96          -8.1252     1.2263  -6.626 5.25e-11 ***
## d97          -7.8840     1.2263  -6.429 1.86e-10 ***
## d98          -8.2292     1.2263  -6.711 3.01e-11 ***
## d99          -8.2442     1.2263  -6.723 2.77e-11 ***
## d00          -8.6690     1.2263  -7.069 2.67e-12 ***
## d01          -8.7019     1.2263  -7.096 2.21e-12 ***
## d02          -8.4650     1.2263  -6.903 8.32e-12 ***
## d03          -8.7310     1.2263  -7.120 1.88e-12 ***
## d04          -8.7656     1.2263  -7.148 1.54e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.008 on 1175 degrees of freedom
## Multiple R-squared:  0.1276, Adjusted R-squared:  0.1098
## F-statistic: 7.164 on 24 and 1175 DF,  p-value: < 2.2e-16
```

3. (15%) Expand your model in *Exercise 2* by adding variables *bac08, bac10, perse, sbprim, sbsecon, sl70plus, gdl, perc14_24, unem, vehicmilespc*, and perhaps *transformations of some or all of these variables*. Please explain carefully your rationale, which should be based on your EDA, behind any transformation you made. If no transformation is made, explain why transformation is not needed. How are the variables *bac8* and *bac10* defined? Interpret the coefficients on *bac8* and *bac10*. Do *per se laws* have a negative effect on the fatality rate? What about having a primary seat belt law? (Note that if a law was enacted sometime within a year the fraction of the year is recorded in place of the zero-one indicator.)

```r
lm.fit2 <- lm(totfatrte ~ d81+d82+d83+d84+d85+d86+d87+d88+d89+
                          d90+d91+d92+d93+d94+d95+d96+d97+d98+d99+
                          d00+d01+d02+d03+d04 + bac08 + bac10 + perse + sbprim + sbsecon + sl70plus
summary(lm.fit2)
```

```
##
## Call:
## lm(formula = totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 +
##     d87 + d88 + d89 + d90 + d91 + d92 + d93 + d94 + d95 + d96 +
##     d97 + d98 + d99 + d00 + d01 + d02 + d03 + d04 + bac08 + bac10 +
##     perse + sbprim + sbsecon + sl70plus + gdl + perc14_24 + unem +
```

```
##     vehicmilespc, data = drivingdf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9160  -2.7384  -0.2778   2.2859  21.4203
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.716e+00  2.476e+00  -1.097 0.272847
## d81          -2.175e+00  8.276e-01  -2.629 0.008686 **
## d82          -6.596e+00  8.534e-01  -7.729 2.33e-14 ***
## d83          -7.397e+00  8.690e-01  -8.512  < 2e-16 ***
## d84          -5.850e+00  8.763e-01  -6.676 3.79e-11 ***
## d85          -6.483e+00  8.948e-01  -7.245 7.82e-13 ***
## d86          -5.853e+00  9.307e-01  -6.289 4.52e-10 ***
## d87          -6.367e+00  9.670e-01  -6.585 6.87e-11 ***
## d88          -6.592e+00  1.014e+00  -6.502 1.17e-10 ***
## d89          -8.071e+00  1.053e+00  -7.667 3.68e-14 ***
## d90          -8.959e+00  1.077e+00  -8.319 2.46e-16 ***
## d91          -1.107e+01  1.101e+00 -10.052  < 2e-16 ***
## d92          -1.288e+01  1.123e+00 -11.473  < 2e-16 ***
## d93          -1.273e+01  1.136e+00 -11.204  < 2e-16 ***
## d94          -1.236e+01  1.157e+00 -10.685  < 2e-16 ***
## d95          -1.195e+01  1.184e+00 -10.098  < 2e-16 ***
## d96          -1.388e+01  1.223e+00 -11.343  < 2e-16 ***
## d97          -1.426e+01  1.250e+00 -11.408  < 2e-16 ***
## d98          -1.504e+01  1.265e+00 -11.886  < 2e-16 ***
## d99          -1.509e+01  1.284e+00 -11.750  < 2e-16 ***
## d00          -1.544e+01  1.305e+00 -11.831  < 2e-16 ***
## d01          -1.618e+01  1.334e+00 -12.131  < 2e-16 ***
## d02          -1.672e+01  1.348e+00 -12.406  < 2e-16 ***
## d03          -1.702e+01  1.359e+00 -12.521  < 2e-16 ***
## d04          -1.671e+01  1.387e+00 -12.049  < 2e-16 ***
## bac08        -2.498e+00  5.375e-01  -4.648 3.73e-06 ***
## bac10        -1.418e+00  3.963e-01  -3.577 0.000362 ***
## perse        -6.201e-01  2.982e-01  -2.079 0.037791 *
## sbprim       -7.533e-02  4.908e-01  -0.153 0.878032
## sbsecon       6.728e-02  4.293e-01   0.157 0.875492
## sl70plus      3.348e+00  4.452e-01   7.521 1.09e-13 ***
## gdl          -4.269e-01  5.269e-01  -0.810 0.417978
## perc14_24     1.416e-01  1.227e-01   1.154 0.248675
## unem          7.571e-01  7.791e-02   9.718  < 2e-16 ***
## vehicmilespc  2.925e-03  9.497e-05  30.804  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.046 on 1165 degrees of freedom
## Multiple R-squared:  0.6078, Adjusted R-squared:  0.5963
```

```
## F-statistic:  53.1 on 34 and 1165 DF,  p-value: < 2.2e-16
```

bac10 is defined as the blood alcohol limit of .10 *bac08* is defined as the blood alcohol limit of .08

Both the variables *bac08* and *bac10* have the negative coefficients of -2.498 and -1.418 respectively. They are statistically significant and it implies that they have a strong negative correlation to the total fatality rate.If we come up with a stricter law and decrease the blood alcohol limit to .10 then the fatalities rate decreases more.

Yes. *perse* variable has a statistically significant negative correlation with the total fatality rate. The coefficient value is -0.6201 which implies a small change in the rate.

TODO write up about primary seatbelt law

4. (15%) Reestimate the model from *Exercise 3* using a fixed effects (at the state level) model. How do the coefficients on *bac08, bac10, perse, and sbprim* compare with the pooled OLS estimates? Which set of estimates do you think is more reliable? What assumptions are needed in each of these models? Are these assumptions reasonable in the current context?

```r
pnldata <- pdata.frame(drivingdf, c("state", "year"))

model.fe <- plm(totfatrte ~ d81+d82+d83+d84+d85+d86+d87+d88+d89+
                      d90+d91+d92+d93+d94+d95+d96+d97+d98+d99+
                      d00+d01+d02+d03+d04 + bac08 + bac10 + perse + sbprim + sbsecon + sl70plus
summary(model.fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 +
##     d87 + d88 + d89 + d90 + d91 + d92 + d93 + d94 + d95 + d96 +
##     d97 + d98 + d99 + d00 + d01 + d02 + d03 + d04 + bac08 + bac10 +
##     perse + sbprim + sbsecon + sl70plus + gdl + perc14_24 + unem +
##     vehicmilespc, data = pnldata, model = "within")
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -8.4273592 -1.0258600 -0.0029547  0.9572345 14.8109310
##
## Coefficients:
##                Estimate  Std. Error  t-value  Pr(>|t|)
## d81          -1.51107133  0.41321486  -3.6569 0.0002672 ***
## d82          -3.02549578  0.44243119  -6.8383 1.316e-11 ***
## d83          -3.50360069  0.45657705  -7.6736 3.628e-14 ***
## d84          -4.25936110  0.46494255  -9.1610 < 2.2e-16 ***
## d85          -4.72679311  0.48547032  -9.7365 < 2.2e-16 ***
```

```
## d86          -3.66118539  0.51769787  -7.0721 2.686e-12 ***
## d87          -4.30578838  0.55532856  -7.7536 2.001e-14 ***
## d88          -4.76712131  0.60155650  -7.9246 5.501e-15 ***
## d89          -6.12997263  0.64019069  -9.5752 < 2.2e-16 ***
## d90          -6.22973766  0.66485076  -9.3701 < 2.2e-16 ***
## d91          -6.91714040  0.68195432 -10.1431 < 2.2e-16 ***
## d92          -7.77417239  0.70288580 -11.0604 < 2.2e-16 ***
## d93          -8.09410864  0.71594741 -11.3055 < 2.2e-16 ***
## d94          -8.50421668  0.73410866 -11.5844 < 2.2e-16 ***
## d95          -8.25540198  0.75623634 -10.9164 < 2.2e-16 ***
## d96          -8.60661913  0.79594975 -10.8130 < 2.2e-16 ***
## d97          -8.70781739  0.81975686 -10.6224 < 2.2e-16 ***
## d98          -9.34924025  0.83373487 -11.2137 < 2.2e-16 ***
## d99          -9.47489124  0.84399083 -11.2263 < 2.2e-16 ***
## d00          -9.99185979  0.85606370 -11.6719 < 2.2e-16 ***
## d01          -9.63121721  0.87255395 -11.0380 < 2.2e-16 ***
## d02          -8.90673015  0.88205263 -10.0977 < 2.2e-16 ***
## d03          -8.93650263  0.88994687 -10.0416 < 2.2e-16 ***
## d04          -9.33936116  0.91107045 -10.2510 < 2.2e-16 ***
## bac08        -1.43722116  0.39421213  -3.6458 0.0002788 ***
## bac10        -1.06266776  0.26883763  -3.9528 8.208e-05 ***
## perse        -1.15161719  0.23398721  -4.9217 9.867e-07 ***
## sbprim       -1.22739974  0.34271485  -3.5814 0.0003564 ***
## sbsecon      -0.34970784  0.25217091  -1.3868 0.1657826
## sl70plus     -0.06253283  0.26931063  -0.2322 0.8164283
## gdl          -0.41177619  0.29257391  -1.4074 0.1595790
## perc14_24     0.18712169  0.09509969   1.9676 0.0493567 *
## unem         -0.57183997  0.06057851  -9.4397 < 2.2e-16 ***
## vehicmilespc  0.00094005  0.00011104   8.4656 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12134
## Residual Sum of Squares: 4535.3
## R-Squared:      0.62624
## Adj. R-Squared: 0.59916
## F-statistic: 55.0943 on 34 and 1118 DF, p-value: < 2.22e-16
```

5. (10%) Would you perfer to use a random effects model instead of the fixed effects model you built in *Exercise 4*? Please explain.

```
model.re <- plm(totfatrte ~ d81+d82+d83+d84+d85+d86+d87+d88+d89+
                      d90+d91+d92+d93+d94+d95+d96+d97+d98+d99+
                      d00+d01+d02+d03+d04 + bac08 + bac10 + perse + sbprim + sbsecon + sl70plus
summary(model.re)
```

```
## Oneway (individual) effect Random Effect Model
```

```
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 +
##     d87 + d88 + d89 + d90 + d91 + d92 + d93 + d94 + d95 + d96 +
##     d97 + d98 + d99 + d00 + d01 + d02 + d03 + d04 + bac08 + bac10 +
##     perse + sbprim + sbsecon + sl70plus + gdl + perc14_24 + unem +
##     vehicmilespc, data = pnldata, model = "random")
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Effects:
##                 var std.dev share
## idiosyncratic 4.057   2.014 0.328
## individual    8.294   2.880 0.672
## theta: 0.8615
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -8.25582 -1.15221 -0.15787  0.93086 16.45691
##
## Coefficients:
##                 Estimate  Std. Error  z-value  Pr(>|z|)
## (Intercept)   1.7149e+01  2.0964e+00   8.1801 2.835e-16 ***
## d81          -1.5489e+00  4.2830e-01  -3.6164 0.0002988 ***
## d82          -3.2433e+00  4.5772e-01  -7.0858 1.383e-12 ***
## d83          -3.7447e+00  4.7212e-01  -7.9318 2.161e-15 ***
## d84          -4.3729e+00  4.8064e-01  -9.0981 < 2.2e-16 ***
## d85          -4.8609e+00  5.0136e-01  -9.6954 < 2.2e-16 ***
## d86          -3.8295e+00  5.3416e-01  -7.1693 7.539e-13 ***
## d87          -4.5014e+00  5.7213e-01  -7.8678 3.610e-15 ***
## d88          -4.9819e+00  6.1887e-01  -8.0500 8.279e-16 ***
## d89          -6.3713e+00  6.5797e-01  -9.6833 < 2.2e-16 ***
## d90          -6.5357e+00  6.8279e-01  -9.5720 < 2.2e-16 ***
## d91          -7.3027e+00  7.0030e-01 -10.4279 < 2.2e-16 ***
## d92          -8.2390e+00  7.2126e-01 -11.4230 < 2.2e-16 ***
## d93          -8.5418e+00  7.3449e-01 -11.6296 < 2.2e-16 ***
## d94          -8.9183e+00  7.5297e-01 -11.8442 < 2.2e-16 ***
## d95          -8.6769e+00  7.7541e-01 -11.1902 < 2.2e-16 ***
## d96          -9.0969e+00  8.1573e-01 -11.1518 < 2.2e-16 ***
## d97          -9.2203e+00  8.3984e-01 -10.9786 < 2.2e-16 ***
## d98          -9.8922e+00  8.5380e-01 -11.5860 < 2.2e-16 ***
## d99          -1.0032e+01  8.6426e-01 -11.6071 < 2.2e-16 ***
## d00          -1.0549e+01  8.7667e-01 -12.0330 < 2.2e-16 ***
## d01          -1.0274e+01  8.9336e-01 -11.5000 < 2.2e-16 ***
## d02          -9.6376e+00  9.0278e-01 -10.6755 < 2.2e-16 ***
## d03          -9.6828e+00  9.1090e-01 -10.6300 < 2.2e-16 ***
## d04          -1.0054e+01  9.3254e-01 -10.7816 < 2.2e-16 ***
```

```
## bac08        -1.5693e+00  4.0384e-01  -3.8860 0.0001019 ***
## bac10        -1.1380e+00  2.7604e-01  -4.1227 3.744e-05 ***
## perse        -1.0933e+00  2.3885e-01  -4.5772 4.712e-06 ***
## sbprim       -1.1761e+00  3.5144e-01  -3.3465 0.0008184 ***
## sbsecon      -3.4758e-01  2.6024e-01  -1.3356 0.1816862
## sl70plus      2.9969e-02  2.7772e-01   0.1079 0.9140655
## gdl          -3.8524e-01  3.0249e-01  -1.2736 0.2028095
## perc14_24     1.9695e-01  9.7213e-02   2.0259 0.0427722 *
## unem         -4.9238e-01  6.1839e-02  -7.9622 1.690e-15 ***
## vehicmilespc  1.1744e-03  1.0983e-04  10.6933 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12834
## Residual Sum of Squares: 5078.6
## R-Squared:      0.60429
## Adj. R-Squared: 0.59274
## Chisq: 1779.05 on 34 DF, p-value: < 2.22e-16
```

```
phtest(model.fe, model.re)
```

```
##
##  Hausman Test
##
## data:  totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 + d87 + d88 + d89 +  ...
## chisq = 148.69, df = 34, p-value = 2.727e-16
## alternative hypothesis: one model is inconsistent
```

Fixed effect model should be used

6. (10%) Suppose that *vehicmilespc*, the number of miles driven per capita, increases by $1,000$. Using the FE estimates, what is the estimated effect on *totfatrte*? Please interpret the estimate.

The coefficient for the *vehicmilespc* variable is 0.00094005 using the FE estimates and it is highly statistically significant. In other words, There will be an increase of 0.94 fatalities per 100k for an increase of 1000 vehicle miles driven per capita.

7. (5%) If there is serial correlation or heteroskedasticity in the idiosyncratic errors of the model, what would be the consequences on the estimators and their standard errors?

There is no serial correlation in the idiosyncratic errors of our model as shown in the p-value below. However if there is Serial correlation then it will not affect the unbiasedness or consistency of OLS estimators, but it does affect their efficiency. With positive serial correlation, the OLS estimates of the standard errors will be smaller than the true standard errors. This will lead to the conclusion that the parameter estimates are more precise than they really are. There will be a tendency to reject the null hypothesis when it should not be rejected.

```r
pbgtest(model.fe)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  totfatrte ~ d81 + d82 + d83 + d84 + d85 + d86 + d87 + d88 + d89 +     d90 + d91 + d9
## chisq = 340.4, df = 25, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```