Chapter 10 - Basic Regression Analysis with Time Series Data

Thalles Quinaglia Liduares

22/03/2022

Exercise 10.11

Upload packages

```
library(lmreg)
library(wooldridge)
library(tseries)
library(car)
```

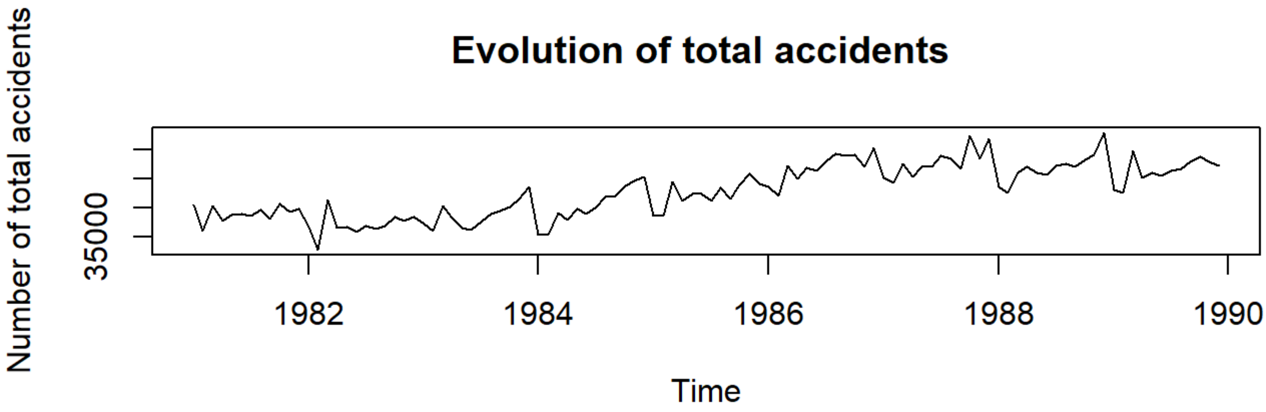
Upload database

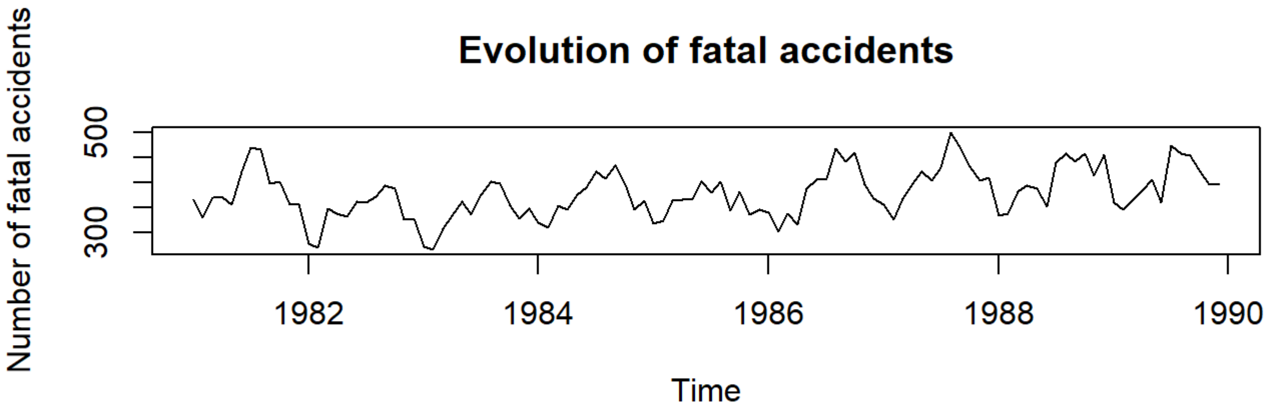
```
data<-wooldridge::traffic2

attach(data)
```

The file TRAFFIC2.RAW contains 108 monthly observations on automobile accidents, traffic laws, and some other variables for California from January 1981 through December 1989. Use this data set to answer the following questions.

(i) During what month and year did California's seat belt law take effect? When did the highway speed limit increase to 65 miles per hour?

```
totacc.ts<-ts(totacc, start=c(1981,1), frequency = 12)

fatacc.ts<-ts(fatacc, start=c(1981,1), frequency = 12)

par(mfrow=c(2,1))

plot.ts(totacc.ts,
        ylab="Number of total accidents",
        main="Evolution of total accidents ")

plot.ts(fatacc.ts,
        ylab="Number of fatal accidents",
        main="Evolution of fatal accidents")
```





The number of fatal accidents fall sharply during 1981, but the lowest value of serie was reached in the beggining of 1982.

(ii) Regress the variable `log(totacc)` on a linear time trend and 11 monthly dummy variables, using January as the base month. Interpret the coefficient estimate on the time trend. Would you say there is seasonality in total accidents?

```
lm1<-lm(ltotacc~feb+mar+apr+may+jun+jul+aug
        +sep+oct+nov+dec+t)

summary(lm1)
```

```
##
## Call:
## lm(formula = ltotacc ~ feb + mar + apr + may + jun + jul + aug +
##     sep + oct + nov + dec + t)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.100216 -0.039778 -0.002915  0.038747  0.138016
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.4685665  0.0190028 550.895  < 2e-16 ***
## feb         -0.0426865  0.0244475  -1.746 0.084035 .
## mar          0.0798245  0.0244491   3.265 0.001523 **
## apr          0.0184849  0.0244517   0.756 0.451536
## may          0.0320981  0.0244554   1.313 0.192509
## jun          0.0201918  0.0244602   0.825 0.411158
## jul          0.0375826  0.0244660   1.536 0.127835
## aug          0.0539830  0.0244729   2.206 0.029808 *
## sep          0.0423610  0.0244809   1.730 0.086811 .
## oct          0.0821135  0.0244899   3.353 0.001149 **
## nov          0.0712785  0.0244999   2.909 0.004511 **
## dec          0.0961572  0.0245111   3.923 0.000165 ***
## t            0.0027471  0.0001611  17.057  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05186 on 95 degrees of freedom
## Multiple R-squared:  0.7969, Adjusted R-squared:  0.7712
## F-statistic: 31.06 on 12 and 95 DF,  p-value: < 2.2e-16
```

The coefficient associated to `t` is equal to 0.002. Hence, there's an increase in `ltotacc` of approximately 0.2% as time passes by, which suggests a presence of tendency in the serie.

Now, testing the presence of seasonality in the serie.

```
hyp<-c("feb=0", "mar=0", "apr=0", "may=0",
       "jun=0", "jul=0", "aug=0","sep=0", "oct=0",
       "nov=0", "dec=0")

linearHypothesis(lm1,hyp)
```

```
## Linear hypothesis test
##
## Hypothesis:
## feb = 0
## mar = 0
## apr = 0
## may = 0
## jun = 0
## jul = 0
## aug = 0
## sep = 0
## oct = 0
## nov = 0
## dec = 0
##
## Model 1: restricted model
## Model 2: ltotacc ~ feb + mar + apr + may + jun + jul + aug + sep + oct +
##     nov + dec + t
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    106 0.40786
## 2     95 0.25550 11   0.15236 5.1501 2.712e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If all coefficients associated to the month dummies are equal to 0, then is there no evidence of seasonality in the serie. But, via F-Test, it's possible to reject the null hypothesis that all coefficients are jointly equal to 0 at the 1% significance level. Hence, there's evidence of seasonality in the serie.

(iii) Add to the regression from part (ii) the variables wkends, unem, spdlaw, and beltlaw. Discuss the coefficient on the unemployment variable. Does its sign and magnitude make sense to you?

```
lm2<-lm(ltotacc~feb+mar+apr+may+jun+jul+aug
        +sep+oct+nov+dec+t+wkends+unem+spdlaw
        +beltlaw)

summary(lm2)
```

```
##
## Call:
## lm(formula = ltotacc ~ feb + mar + apr + may + jun + jul + aug +
##     sep + oct + nov + dec + t + wkends + unem + spdlaw + beltlaw)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.084488 -0.023998 -0.002176  0.024586  0.089315
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.6398628  0.0630860 168.657  < 2e-16 ***
## feb         -0.0338346  0.0177683  -1.904  0.06004 .
## mar          0.0769530  0.0167941   4.582 1.46e-05 ***
## apr          0.0104561  0.0170469   0.613  0.54116
## may          0.0237074  0.0169389   1.400  0.16504
## jun          0.0219334  0.0172149   1.274  0.20588
## jul          0.0499293  0.0167036   2.989  0.00360 **
## aug          0.0559526  0.0168173   3.327  0.00127 **
## sep          0.0420693  0.0172819   2.434  0.01687 *
## oct          0.0817171  0.0169554   4.820 5.73e-06 ***
## nov          0.0768721  0.0172455   4.458 2.36e-05 ***
## dec          0.0990863  0.0170705   5.805 9.32e-08 ***
## t            0.0011011  0.0002579   4.270 4.79e-05 ***
## wkends       0.0033333  0.0037761   0.883  0.37970
## unem        -0.0212173  0.0033974  -6.245 1.33e-08 ***
## spdlaw      -0.0537593  0.0126036  -4.265 4.87e-05 ***
## beltlaw      0.0954528  0.0142351   6.705 1.65e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03524 on 91 degrees of freedom
## Multiple R-squared:  0.9101, Adjusted R-squared:  0.8943
## F-statistic: 57.61 on 16 and 91 DF,  p-value: < 2.2e-16
```

The coefficient associated to variable `unem` is equal to -0.021 and have statistical significance at the 1% level. As `unem` increases, `ltotacc` decreases which suggests some sense, since individuals in situation of unemployment might have no access to a vehicle.

(iv) In the regression from part (iii), interpret the coefficients on `spdlaw` and `beltlaw`. Are the estimated effects what you expected? Explain.

The coefficient associated to `spdlaw` has the expected sign, as limits of velocity is imposed, there's a tendency in decrease the number of total accidents. In the other hand, the coefficient sign of `beltlaw` suggests to the contrary. A possible explanation for this result, is that, as individuals use the belt, they might incur in high speed velocities, cause the belt causes a sensation of safety.

(v) The variable `prcfat` is the percentage of accidents resulting in at least one fatality. Note that this variable is a percentage, not a proportion. What is the average of `prcfat` over this period? Does the magnitude seem about right?

```
round(mean(prcfat),2)*100
```

```
## [1] 89
```

The mean percentual value of `prcfat` is equal to 89%. It seems, a very high percentage of fatalities in the total number of car accidents.

(vi) Run the regression in part (iii) but use prcfat as the dependent variable in place of log(totacc). Discuss the estimated effects and significance of the speed and seat belt law variables.

```
options(scipen=999)

lm3<-lm(prcfat~feb+mar+apr+may+jun+jul+aug
        +sep+oct+nov+dec+t+wkends+unem+spdlaw
        +beltlaw)

summary(lm3)
```

```
##
## Call:
## lm(formula = prcfat ~ feb + mar + apr + may + jun + jul + aug +
##     sep + oct + nov + dec + t + wkends + unem + spdlaw + beltlaw)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.132085 -0.032573  0.000621  0.038226  0.132080
##
## Coefficients:
##                Estimate  Std. Error t value            Pr(>|t|)
## (Intercept)  1.02979855  0.10295235  10.003 0.000000000000000252 ***
## feb          0.00086075  0.02899674   0.030            0.976384
## mar          0.00009226  0.02740688   0.003            0.997321
## apr          0.05822009  0.02781947   2.093            0.039152 *
## may          0.07163924  0.02764324   2.592            0.011129 *
## jun          0.10126183  0.02809371   3.604            0.000510 ***
## jul          0.17661205  0.02725922   6.479 0.000000004640051325 ***
## aug          0.19261166  0.02744483   7.018 0.000000000390766401 ***
## sep          0.16001639  0.02820296   5.674 0.000000164166411598 ***
## oct          0.10103565  0.02767019   3.651            0.000435 ***
## nov          0.01394901  0.02814359   0.496            0.621345
## dec          0.00920049  0.02785799   0.330            0.741960
## t           -0.00223523  0.00042083  -5.312 0.000000765623917802 ***
## wkends       0.00062587  0.00616243   0.102            0.919328
## unem        -0.01542594  0.00554440  -2.782            0.006563 **
## spdlaw       0.06708765  0.02056831   3.262            0.001560 **
## beltlaw     -0.02950528  0.02323073  -1.270            0.207288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05751 on 91 degrees of freedom
## Multiple R-squared:  0.7174, Adjusted R-squared:  0.6677
## F-statistic: 14.44 on 16 and 91 DF,  p-value: < 0.00000000000000022
```

In this case, the coefficient sign of `beltlaw` is as expected, but without statistical significance. The percentual of fatalities increases, as speed law increases in this case, which is not an expected result in empirical terms.