Panel Data

Jayashree Raman

9/18/2018

```
library(plm)
```

```
## Warning: package 'plm' was built under R version 3.5.1
```

```
## Loading required package: Formula
```

```
library(prediction)
```

```
## Warning: package 'prediction' was built under R version 3.5.1
```

```
library(Metrics)
```

```
## Warning: package 'Metrics' was built under R version 3.5.1
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 3.5.1
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
train_data <- na.omit(read.csv(file="usersessions-with-genre-train.csv", header=TRUE, row.names = NULL,

panel.data.train <- plm.data(train_data, index = c("userid","session_start"))
```

```
## Warning: use of 'plm.data' is discouraged, better use 'pdata.frame' instead
```

```
mdl_fd <-plm(session_length~age+session_length_mvavg+previous_duration+is_holiday+absence_time,  data =

##Summaries

summary(mdl_fd)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg + previous_duration +
```

```
##     is_holiday + absence_time, data = panel.data.train, model = "fd")
##
## Unbalanced Panel: n = 174, T = 19-4312, N = 169556
## Observations used in estimation: 169382
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -387974   -3009    -461      -3    2399  518665
##
## Coefficients:
##                          Estimate  Std. Error   t-value Pr(>|t|)
## age                   -4.2003e+02  5.2674e+02   -0.7974   0.4252
## session_length_mvavg -3.7435e+00  1.0185e-01  -36.7541   <2e-16 ***
## previous_duration     -4.5480e-01  2.1233e-03 -214.1979   <2e-16 ***
## is_holiday            -1.9193e+01  7.5960e+01   -0.2527   0.8005
## absence_time          -1.8753e-05  3.9531e-05   -0.4744   0.6352
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.2189e+13
## Residual Sum of Squares: 1.6988e+13
## R-Squared:      0.23441
## Adj. R-Squared: 0.23439
## F-statistic: 12965.2 on 4 and 169377 DF, p-value: < 2.22e-16
```

```
cat(length(panel.data.train$session_length), length(mdl_fd$residuals))
```

```
## 169556 169382
```

```
# Fitted vs Observed and Fitted vs Residuals plots
# par(mfrow=c(1,2))
# plot(panel.data.train$session_length-mdl_fd$residuals, panel.data.train$session_length, asp=1, ylab =
# abline(0,1, col='red', lty='dashed', lwd=2)
#
# ## Fitted vs Residuals plots
# plot(panel.data.train$session_length-mdl_fd$residuals,mdl_fd$residuals, asp=1, ylab = "Observed", xla
# abline(0,0, col='red', lty='dashed', lwd=2)

## MAE and RMSE

mae_fd = mean(abs(mdl_fd$residuals))
rmse_fd = sqrt(mean(abs(mdl_fd$residuals)^2))

cat('MAE = ', mae_fd, ', RMSE = ', rmse_fd)
```

```
## MAE =  4919.198 , RMSE =  10014.66
```