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-char-sec-train.csv", header=TRUE, row.names = NU

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

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

```
mdl_fd <-plm(session_length~age+session_length_mvavg,  data = panel.data.train, model = "fd")

##Summaries

summary(mdl_fd)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg, data = panel.data.train,
```

```
##     model = "fd")
##
## Unbalanced Panel: n = 221244, T = 1-2, N = 221496
## Observations used in estimation: 252
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -17815   -3479    -106     389    4142   29717
##
## Coefficients:
##                       Estimate Std. Error t-value  Pr(>|t|)
## age                  -78.79559   56.92435 -1.3842    0.1675
## session_length_mvavg   0.73620    0.12892  5.7107 3.182e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.2778e+10
## Residual Sum of Squares: 1.1421e+10
## R-Squared:      0.10924
## Adj. R-Squared: 0.10568
## F-statistic: 29.7026 on 1 and 250 DF, p-value: 1.2063e-07
```

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

```
## 221496 252
```

```
## 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 =  4956.749 , RMSE =  6732.097
```