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,
```

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

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

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

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

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

```
mdl_between <-plm(session_length~age+session_length_mvavg+previous_duration+is_holiday+absence_time,  da
```

```
##Summaries
```

```
summary(mdl_between)
```

```
## Oneway (individual) effect Between Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg + previous_duration +
##     is_holiday + absence_time, data = panel.data.train, model = "between")
##
## Unbalanced Panel: n = 174, T = 19-4312, N = 169556
## Observations used in estimation: 174
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -4505.846   -33.009    31.877    90.311  2066.097
##
## Coefficients:
##                           Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept)          -7.4925e+01  1.4767e+02 -0.5074    0.6125
## age                  -8.1389e-01  5.0907e+00 -0.1599    0.8732
## session_length_mvavg -9.9072e-02  2.1594e-02 -4.5879 8.722e-06 ***
## previous_duration     1.1104e+00  1.9697e-02 56.3702 < 2.2e-16 ***
## is_holiday            2.0764e+02  1.5944e+02  1.3023    0.1946
## absence_time         -1.1073e-04  2.4847e-04 -0.4457    0.6564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    4366600000
## Residual Sum of Squares: 30800000
## R-Squared:      0.99295
## Adj. R-Squared: 0.99274
## F-statistic: 4730.03 on 5 and 168 DF, p-value: < 2.22e-16
```

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

## MAE and RMSE

mae_between = mean(abs(mdl_between$residuals))
rmse_between = sqrt(mean(abs(mdl_between$residuals)^2))

cat('MAE = ', mae_between, ', RMSE = ', rmse_between)
```

```
## MAE =  152.869 , RMSE =  420.7264
```

```
## Prediction
predval <- prediction(mdl_between, data=test_data)

 mae_pred = mean(abs(predval$session_length-predval$fitted))
```

```r
 rmse_pred = sqrt(mean(abs(predval$session_length-predval$fitted)^2))
#
 cat('MAE = ', mae_pred, ', RMSE = ', rmse_pred)
```

```
## MAE =  5350.997 , RMSE =  10853.87
```