

# Panel Data

*Jayashree Raman*

9/18/2018

```
library(plm)

## Loading required package: Formula
library(prediction)
library(Metrics)
library(tseries)
```

## 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 = NULL))

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_fe <- plm(log(session_length) ~ age + session_length_mvavg, data = panel.data.train, model = "within")

##Summaries

summary(mdl_fe)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(session_length) ~ age + session_length_mvavg,
##       data = panel.data.train, model = "within")
##
## Unbalanced Panel: n = 221244, T = 1-2, N = 221496
##
## Residuals:
##      Min. 1st Qu. Median 3rd Qu.   Max.
## -2.0041  0.0000  0.0000  0.0000  2.0041
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## age                  -1.8168e-03  1.3396e-02 -0.1356    0.8922
## session_length_mvavg  1.2594e-04  3.0338e-05  4.1513  4.538e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:     338.39
```

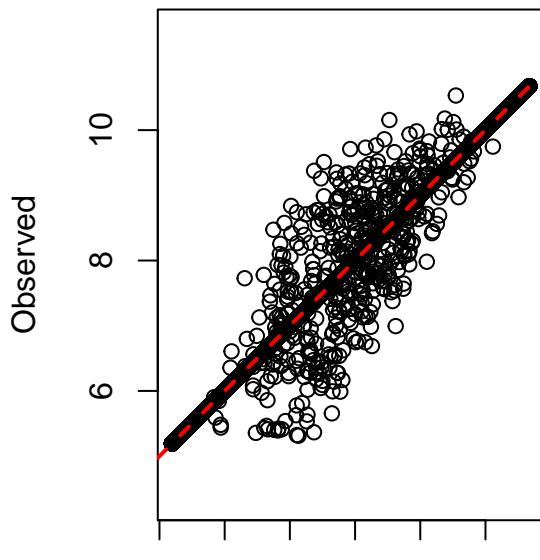
```

## Residual Sum of Squares: 316.26
## R-Squared:      0.065421
## Adj. R-Squared: -827.02
## F-statistic: 8.75001 on 2 and 250 DF, p-value: 0.00021234
#cat(length(panel.data.train$session_length), length(mdl_fe$residuals))
# Fitted vs Observed and Fitted vs Residuals plots
par(mfrow=c(1,2))
plot(log(panel.data.train$session_length)-mdl_fe$residuals, log(panel.data.train$session_length), asp=1,
abline(0,1, col='red', lty='dashed', lwd=2)

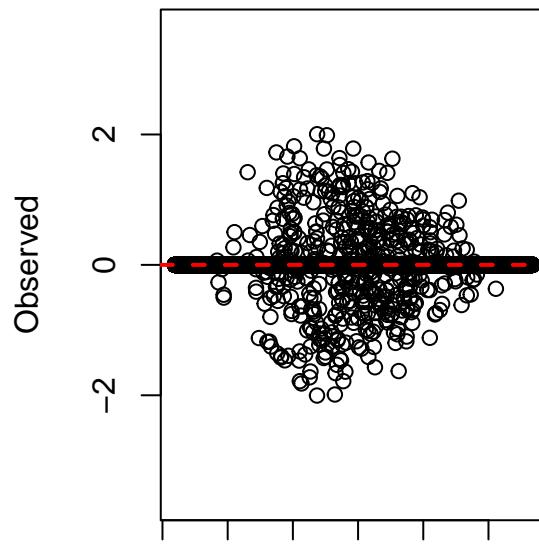
## Fitted vs Residuals plots
plot(log(panel.data.train$session_length)-mdl_fe$residuals,mdl_fe$residuals, asp=1, ylab = "Observed",
abline(0,0, col='red', lty='dashed', lwd=2)

```

**Fitted vs Observed**



**Fitted vs Residuals**



```
## MAE and RMSE
```

```

mae_fe = mean(abs(mdl_fe$residuals))
rmse_fe = sqrt(mean(abs(mdl_fe$residuals)^2))

cat('MAE = ', mae_fe, ', RMSE = ', rmse_fe)

```

```
## MAE = 0.001466625 , RMSE = 0.03778647
```