

# 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_pooled <- plm(session_length~age+session_length_mvavg+previous_duration+is_holiday+absence_time, da

##Summaries

summary(mdl_pooled)

## Pooling Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg + previous_duration +
```

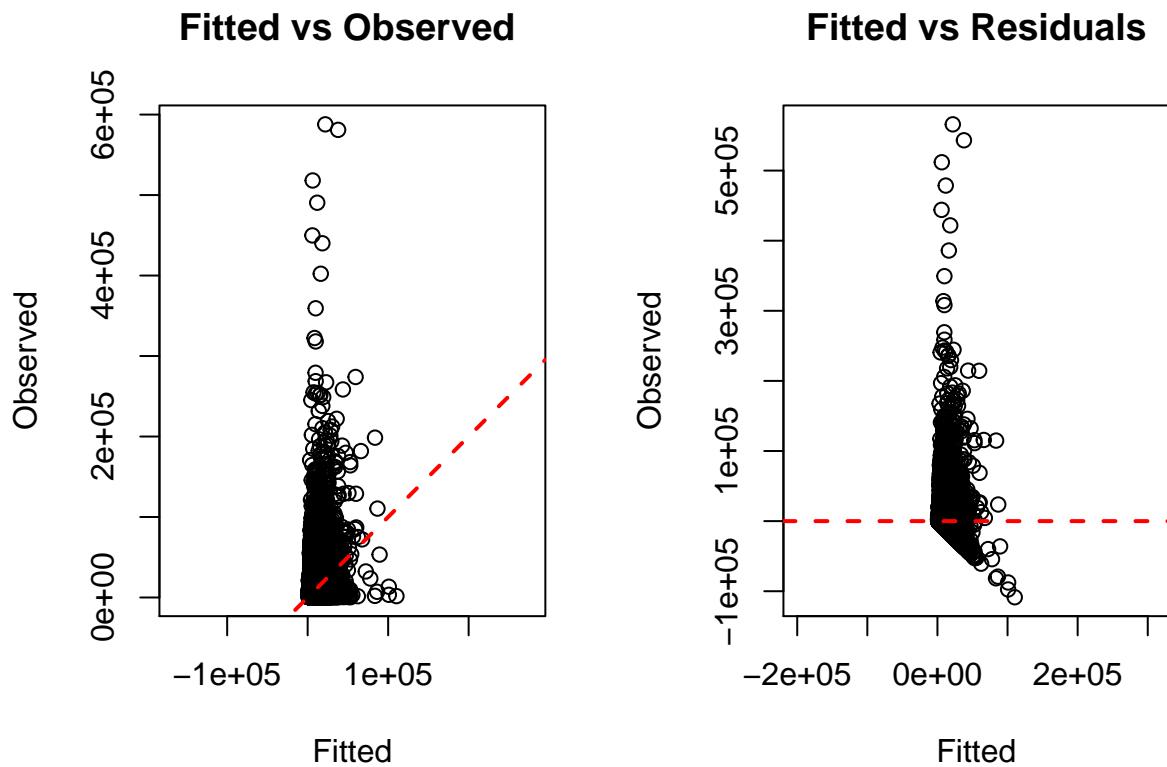
```

##      is_holiday + absence_time, data = panel.data.train, model = "pooling")
##
## Unbalanced Panel: n = 174, T = 19-4312, N = 169556
##
## Residuals:
##      Min.    1st Qu.     Median    3rd Qu.     Max.
## -108569.3   -3340.8   -1545.0    1475.2  565949.6
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## (Intercept)            6.3855e+02  1.0102e+02   6.3208 2.609e-10 ***
## age                    3.0953e-01  3.5901e+00   0.0862   0.9313
## session_length_mvavg  6.9506e-01  6.4212e-03 108.2438 < 2.2e-16 ***
## previous_duration      1.6008e-01  2.3733e-03  67.4502 < 2.2e-16 ***
## is_holiday             -2.6509e+01  5.1160e+01  -0.5182   0.6043
## absence_time           -1.4716e-05  4.6387e-05  -0.3173   0.7511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:  1.4832e+13
## Residual Sum of Squares: 1.3006e+13
## R-Squared:      0.1231
## Adj. R-Squared: 0.12307
## F-statistic: 4760.3 on 5 and 169550 DF, p-value: < 2.22e-16

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

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

```



```
## MAE and RMSE

mae_pooled = mean(abs(mdl_pooled$residuals))
rmse_pooled = sqrt(mean(abs(mdl_pooled$residuals)^2))

cat('MAE = ', mae_pooled, ', RMSE = ', rmse_pooled)

## MAE = 4355.626 , RMSE = 8758.281
```