ReadMe

# lmTutorial

## Excercise for the understanding lm on the states dataset

The dataset compares the data from 6 states , containing information about the region, population, are, density , csat, vsat, percent, expense

The Excercise involves changing the lm equation to include a new interation to see how the csat value is affected

setwd("C:/Users/s/Downloads/linear\_regression/linear\_regression")  
getwd() # where am I?

## [1] "C:/Users/s/Downloads/linear\_regression/linear\_regression"

list.files("dataSets") # files in the dataSets folder

## [1] "Exam.rds" "states.dta" "states.rds"

## Load the states data  
## ââââââââââââââââââââââââ  
  
# read the states data  
states.data <- readRDS("dataSets/states.rds")

## 1. Add on to the regression equation that you created in exercise 1 by

## generating an interaction term and testing the interaction.

sat.expense.metro.by.percent <- lm(csat ~ expense\*metro,  
 data=states.data)   
  
print(coef(summary(sat.expense.metro.by.percent)))

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.161950e+03 9.822358e+01 11.8296468 1.495599e-15  
## expense -3.108410e-02 1.877408e-02 -1.6556924 1.045915e-01  
## metro -1.999615e+00 1.431661e+00 -1.3967101 1.692035e-01  
## expense:metro 2.135553e-04 2.558244e-04 0.8347732 4.081600e-01

Then we add region to the model

sat.expense.region.metro.by.percent <- lm(csat ~ expense\*metro\*region,  
 data=states.data)

Looking at the summary it seems like there is significant difference between the regions

# compare both using the anova() function

anova(sat.expense.metro.by.percent, sat.expense.region.metro.by.percent)

## Analysis of Variance Table  
##   
## Model 1: csat ~ expense \* metro  
## Model 2: csat ~ expense \* metro \* region  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 46 157583   
## 2 34 70237 12 87345 3.5235 0.001886 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1