**Problem 1 (Chapter 2 Exercises 4):**

1. *Describe three real-life applications in which classification might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.*
2. Presidential Election campaign winner

Response: Trump/Clinton; Predictors: Political talent, Experience, Air time, Age, Money, Fund raising; Goal: Prediction

1. Starting Bike share system in City

Response : Success/ Failure; Predictors: Road design, New Roads, Income of resident, Age of resident, Transportation system ; Goal: Prediction

1. Professors Success

Response: Successful/Not successful ; Predictors: Years of education, Age, Experience, Researches, Student relations; Goal: Prediction

* *(b) Describe three real-life applications in which regression might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.*

1. What is the average residential rent in downtown Boston over the next 5 years

Response: Average rent price in downtown Boston over the next 5 years car price x1,x2,x3,X4,X5 Predictors: Transportation, Parks, School, Average income of resident, Crime rate, Universities success (Boston is a university city); Goal: inference

1. What is the average car sale price in MA over the next 3 years?

Response: Average car price for next years, x1,x2,x3 Predictors: Gas mileage, efficiency, liability, Average income, Goal: inference

1. What is the average car insurance price in downtown Boston over the next 3 years?

Response: Average car insurance price for next years, x1,x2,x3 Predictors: Crime rate, Accident rate, Resident income, Parking; Goal: inference

*(c) Describe three real-life applications in which cluster analysis might be useful.*

1. Illness clustering, clustering of demographics for an illness (flue…) to see which clusters of community get ill and diagnose illness types
2. Marketing product satisfaction research, research specific product for demographics to find out consumer’s satisfaction rate.
3. Consumer car model recommendations, recommend car model based on consumers who had purchased in the past years, good experience with reliability and most selling car model in the years.

**Problem 2 (Chapter 2 Exercises 8):**

1. ***college = read.csv("/Users/ekinezgi/Documents/Umass Dartmouth/\_MTH-522 Istatistical Learning 2016F/Data/College.csv")***
2. ***fix(college)***

***rownames(college) = college[,1]***

***college = college[,-1]***

***fix(college)***

1. ***(i) summary(college)***

***(ii) pairs(college[,1:10])***

***(iii) plot(college$Private, college$Outstate, xlab = "Private (Yes/No)", ylab ="Tuition USD ($)", main = "Out-of-state Tuition")***

***(iv) Elite = rep("No", nrow(college))***

***Elite[college$Top10perc > 50] = "Yes"***

***Elite = as.factor(Elite)***

***college = data.frame(college, Elite)***

***summary(college$Elite)***

***No Yes***

***699 78***

***plot(college$Elite, college$Outstate, xlab = "Elite University (Yes/No)", ylab ="Tuition USD ($)", main = "Out-of-state Tuition")***

***(v) par(mfrow=c(2,2))***

***hist(college$Apps, col=1)***

***hist(college$perc.alumni, col=4)***

***hist(college$S.F.Ratio, col=5)***

***hist(college$Books, col=6)***

***(vi) par(mfrow=c(1,2))***

***plot(college$Apps, college$Grad.Rate)***

***plot(college$Top10perc, college$Grad.Rate)***

**Problem 3 (Chapter 2 Exercises 9):**

**Auto = read.csv("/Users/ekinezgi/Documents/Umass Dartmouth/\_MTH-522 Istatistical Learning**

**2016F/Data/Auto.csv", na.strings="?", header=T)**

**Auto = na.omit(Auto)**

**str(Auto)**

1. quantitative: mpg, cylinders, displacement, horsepower, weight, acceleration, year

qualitative: name, origin

1. quantitative predictors are first seven columns;

***sapply(Auto[, 1:7], range)***

1. quantitative predictors are first seven columns;

**sapply(Auto[, 1:7], mean)**

**sapply(Auto[, 1:7], sd)**

1. **subsetAuto = Auto[-c(10:85), -c(9)]**

**sapply(subsetAuto, range)**

**sapply(subsetAuto, mean)**

**sapply(subsetAuto, sd)**

1. **pairs(Auto)** #A Matrix of scatterplots is produced for all of the predictors

**plot(Auto$mpg, Auto$cylinders)** # When cylinders increase mpg get less

**plot(Auto$mpg, Auto$year)** # Over years car mpg gets bettter

**plot(Auto$mpg, Auto$origin)** # Origin 3-Japanese car more efficient than 2-Europan car

1. ***pairs(Auto)*** #A Matrix of scatterplots is produced for all of the predictors

As we can see from plots (e) mpg has some correlations with all of the predictors, except the name predictor which has to little observation. Other all predictors be useful to predict mpg.

**Problem 4 (Chapter 2 Exercises 10):**

1. ***library(MASS)***

***?Boston***

***nrow(Boston)***

*[1] 506 #506 rows, records ; Housing info*

***ncol(Boston)***

*[1] 14 #14 column, features*

***dim(Boston)***

*[1] 506 14*

1. ***pairs(Boston)*** #A Matrix of scatterplots is produced for all of the predictors

*lstat has correlation with medv*

*indus has correlation with dis*

*dis has correlation with*

*zn has correlation with nox, age, lstat*

*crime has correlation with age, dis, ptratio, rat*

1. ***plot(Boston)*** :A Matrix of scatterplots is produced for all of the predictors

*plot(Boston$age, Boston$crim) : More age, more crime - For older house we have more crime*

*plot(Boston$tax, Boston$crim) : More tax , more crime*

*plot(Boston$ptratio, Boston$crim) : More ptrati (pupil-teacher ratio by town), more crime*

*plot(Boston$rad, Boston$crim) : Higher rad (index of access to radial highways), more crime*

1. *hist(Boston$crim, breaks=50) : We can see high and low crime rates*

*nrow(Boston[Boston$crim > 20, ])*

*>[1] 18 : 18 suburbs has high crime rate*

*nrow(Boston[Boston$crim <= 20, ])*

*>[1] 488 : 488 town has low crime rate.*

*hist (Boston$tax, breaks=50) : Huge difference between low tax rate town*

*nrow(Boston[Boston$tax = 666, ]) and tax rate = 666*

*>[1] 132*

hist(Boston$ptratio, breaks=50) : Some high *pupil-teacher ratio by town but*

*generaly no it looks average*

1. Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

***nrow(Boston[Boston$chas == 1, ])***

>[1] 35

1. ***median(Boston$ptratio)***

>[1] 19.05

1. Which suburb of Boston has lowest median value of owner- occupied homes? What are the values of the other predictors for that suburb, and how do those values compare to the overall ranges for those predictors?

***row.names(Boston[min(Boston$medv), ])*** #min median value of owner-occupied home

>[1] “5”

***summary(Boston) # For Comparison with min(Boston$medv)))***

***crim zn indus chas …***

***Min. : 0.00632 Min. : 0.00 Min. : 0.46 Min. :0.00000 …***

***1st Qu.: 0.08204 1st Qu.: 0.00 1st Qu.: 5.19 1st Qu.:0.00000 …***

***Median : 0.25651 Median : 0.00 Median : 9.69 Median :0.00000 …***

***Mean : 3.61352 Mean : 11.36 Mean :11.14 Mean :0.06917 …***

***3rd Qu.: 3.67708 3rd Qu.: 12.50 3rd Qu.:18.10 3rd Qu.:0.00000 …***

***Max. :88.97620 Max. :100.00 Max. :27.74 Max. :1.00000 …***

***t(subset(Boston, medv == min(Boston$medv)))***

399 406

crim 38.3518 67.9208 ***# Over 3rd Qu.***

zn 0.0000 0.0000 ***# Minimum***

indus 18.1000 18.1000 ***# 3rd Qu.***

chas 0.0000 0.0000 ***# 0: not bounds with river***

nox 0.6930 0.6930 ***# Over 3rd Qu.***

rm 5.4530 5.6830 ***# Under 1st Qu.***

age 100.0000 100.0000 ***# At Max.***

dis 1.4896 1.4254 ***# Under 1st Qu.***

rad 24.0000 24.0000 ***# At Max.***

tax 666.0000 666.0000 ***# Over 3rd Qu.***

ptratio 20.2000 20.2000 ***# Over 3rd Qu.***

black 396.9000 384.9700 ***# At Max ; Over 1st Qu.***

lstat 30.5900 22.9800 ***# Over 3rd Qu.***

medv 5.0000 5.0000 ***#At Min.***

1. In this data set, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

***dim(subset(Boston, rm > 7))***

>[1] 64 14

***dim(subset(Boston, rm > 8))***

[1] 13 14

***summary(Boston$lstat)***

>Min. 1st Qu. Median Mean 3rd Qu. Max.

>1.73 6.95 11.36 12.65 16.96 37.97

***summary(subset(Boston, rm > 8)$lstat)***

>Min. 1st Qu. Median Mean 3rd Qu. Max.

>2.47 3.32 4.14 4.31 5.12 7.44 ***# Lower lstat***

***> summary(Boston$crim)***

>Min. 1st Qu. Median Mean 3rd Qu. Max.

> 0.00632 0.08204 0.25650 3.61400 3.67700 88.98000

***summary(subset(Boston, rm > 8)$crim)***

> Min. 1st Qu. Median Mean 3rd Qu. Max.

>0.02009 0.33150 0.52010 0.71880 0.57830 3.47400

***# Lower crime***