# Appendix section

Code:

# Part 1

### Dataset

Democrat<-c(170, 448, 369, 182, 394, 206, 258, 433, 503, 426, 409, 421, 355, 516, 226, 535, 489,

337, 464, 508, 325, 521, 533, 533, 237, 476, 312, 493, 411, 464, 192, 689, 398)

Republicant<-c(377, 373, 284, 249, 402, 336, 377, 263, 352, 341, 290, 365, 416, 299, 334, 353, 363,346, 285, 371, 387, 333, 431, 349, 307, 294, 374, 254, 366, 416, 274, 343, 415)

## 1a Descriptive statistics

sprintf("The Mean,Std Deviation of Democrat is :%f,%f",mean(Democrat),sd(Democrat))

sprintf("The Mean,Std Deviation of Republicant is :%f,%f",mean(Republicant),sd(Republicant))

## 1b Difference in means and confidence interval

Md=mean(Democrat)

Mr=mean(Republicant)

nd=length(Democrat)

nr=length(Republicant)

df=nr+nd-2

S\_square=(sum((Democrat-Md)^2)+sum((Republicant-Mr)^2))/(nd+nr-2)

Mean\_diff=Md-Mr

std\_mean\_diff=sqrt(S\_square/nd+S\_square/nr)

q\_alpha\_95=qt(1-0.05/2,df=df)

q\_alpha\_99=qt(1-0.01/2,df=df)

lower\_95=Mean\_diff-q\_alpha\_95\*std\_mean\_diff

upper\_95=Mean\_diff+q\_alpha\_95\*std\_mean\_diff

lower\_99=Mean\_diff-q\_alpha\_99\*std\_mean\_diff

upper\_99=Mean\_diff+q\_alpha\_99\*std\_mean\_diff

## 1c Summary

t.test(Democrat,Republicant,alternative = "greater",conf.level = 0.95)

# Part 2

### Read Data

pmdata <- read.csv("./pmgb2012\_2014.csv",header = TRUE)

pmdata<-na.omit(pmdata) # remove NA value if has

pmdata<-pmdata[!duplicated(pmdata),] # remove duplicate data if has

str(pmdata)

edu\_level3<-pmdata$edu\_level3

edu\_level4<-pmdata$edu\_level4

pop\_density<-pmdata$pop\_density

describe<-function(variable,name){

Mean<-mean(variable)

Med<-median(variable)

Range<-range(variable)

Variance<-var(variable)

sprintf("The variable %s --- mean:%f,median:%f,min:%f,max:%f,variance:%f",name,

Mean,Med,Range[1],Range[2],Variance)

}

describe(edu\_level3,name="edu\_level3")

describe(edu\_level4,name="edu\_level4")

describe(pop\_density,name="pop\_density")

plot(density(edu\_level3),col='red',main="edu\_level3")

boxplot(edu\_level3,main="edu\_level3")

plot(density(edu\_level4),col='red',main="edu\_level4")

boxplot(edu\_level4,main="edu\_level4")

prop.table(table(pop\_density))

boxplot(pop\_density,main="pop\_density")

## 2b Visualization

### Scatterplot

plot(x=pmdata$edu\_level4,y=pmdata$pmdeaths\_total,xlab=" degree-level education",ylab ="premature mortality" )

### box plot

library(ggplot2)

library(ggpubr)

country=pmdata$country

country=ifelse(country==1,"England",ifelse(country==2,"Scotland","Wales"))

df=data.frame(pmdeaths\_total=pmdata$pmdeaths\_total,country=as.factor(country))

ggplot(data=df,aes(x=country,y=pmdeaths\_total,color=country))+geom\_boxplot()+

stat\_compare\_means(comparisons = list(c("England","Scotland"),c("England","Wales"),c("Scotland","Wales")),

aes(label=..p.signif..))

## 2c Difference in Means

mean\_delta=mean(pmdata$pmdeaths\_male)-mean(pmdata$pmdeaths\_female)

cat("The mean difference between premature mortality among men and women in Great Britain is :",mean\_delta,"\n")

cat("Conduct t-test: \n")

t.test(pmdata$pmdeaths\_male,pmdata$pmdeaths\_female,conf.level = 0.95,alternative = "greater")

## 2d Linear Regression

model<-lm(pmdeaths\_total~mean\_income,data = pmdata)

summary(model)