Lab 5: Pareto and the 1 percent

Stat 597A

Friday, 16 October 2015

Part I

1. Loading Data

```
setwd("/Users/ankitashankhdhar/Documents/Grad 2nd yr/Comp Stats/Lab 5")
data <- read.csv(file = "wtid-report.csv", header = TRUE, sep = ",")

data <- data[1:100, ]
year <- data$Year
P99 <- data$P99.income.threshold
P99.5 <- data$P99.5.income.threshold
P99.9 <- data$P99.9.income.threshold
income <- data.frame(year, P99, P99.5, P99.9)

yr1972 <- which(income$year == 1972)
income$P99[yr1972]
```

```
## [1] 209076.6
```

```
yr1942 <- which(income$year == 1942)
income$P99.5[yr1942]
```

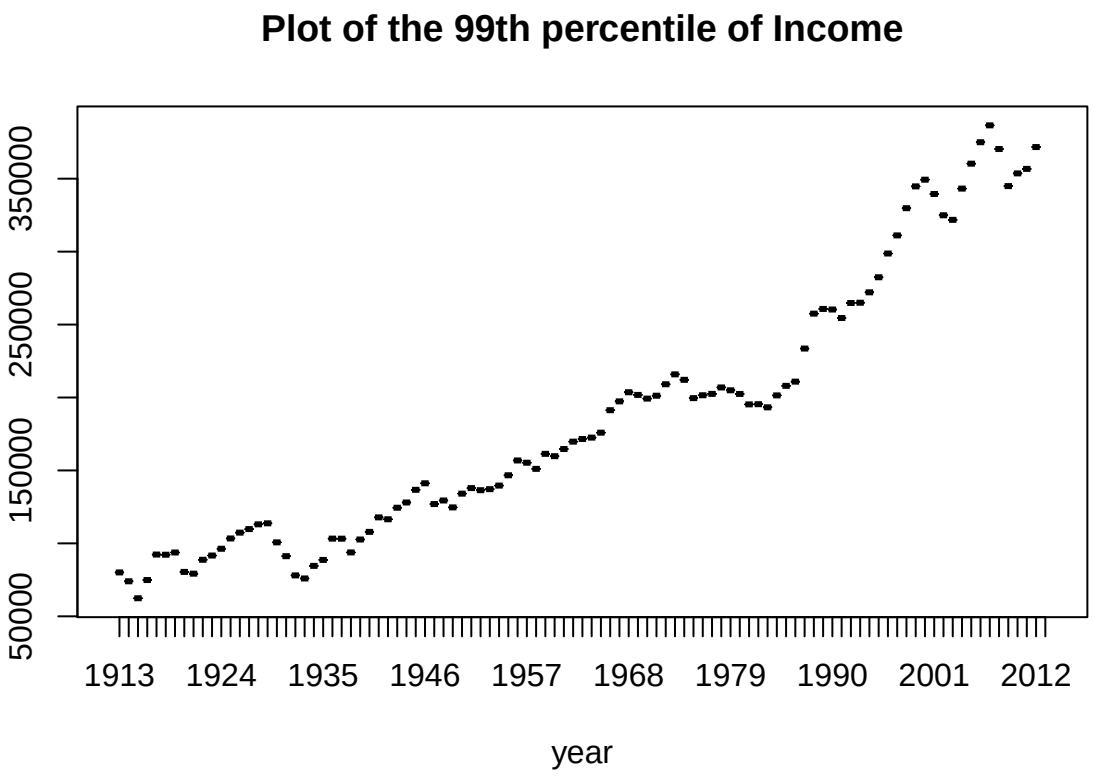
```
## [1] 183217
```

```
yr1922 <- which(income$year == 1922)
income$P99.5[yr1922]
```

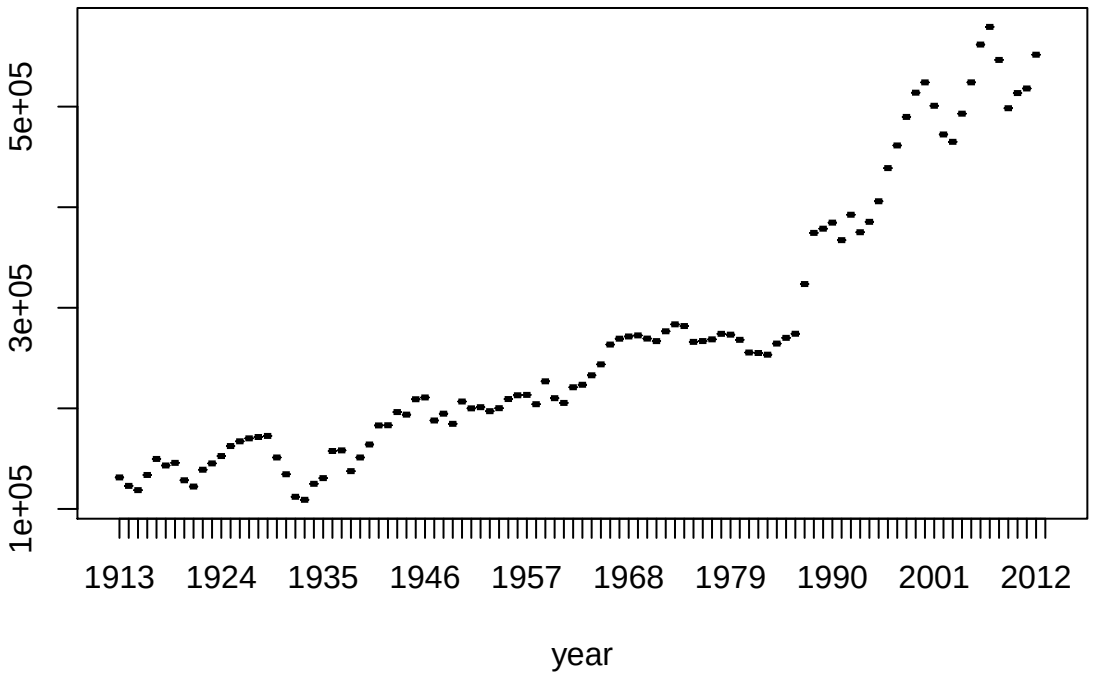
```
## [1] 139003.1
```

2.

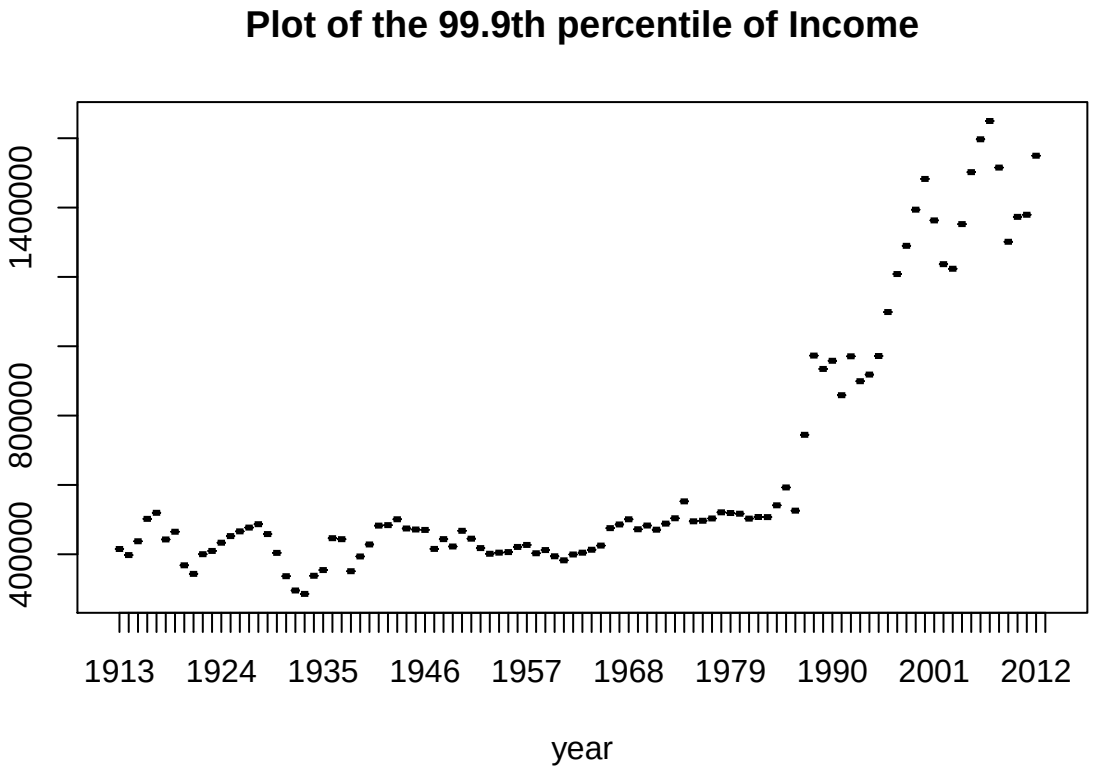
```
plot(year,P99,main='Plot of the 99th percentile of Income',
     xlab='year')
```




```
plot(year,P99.5,main='Plot of the 99.5th percentile of Income',
     xlab='year')
```



```
plot(year,P99.9,main='Plot of the 99.9th percentile of Income',
     xlab='year')
```



3.

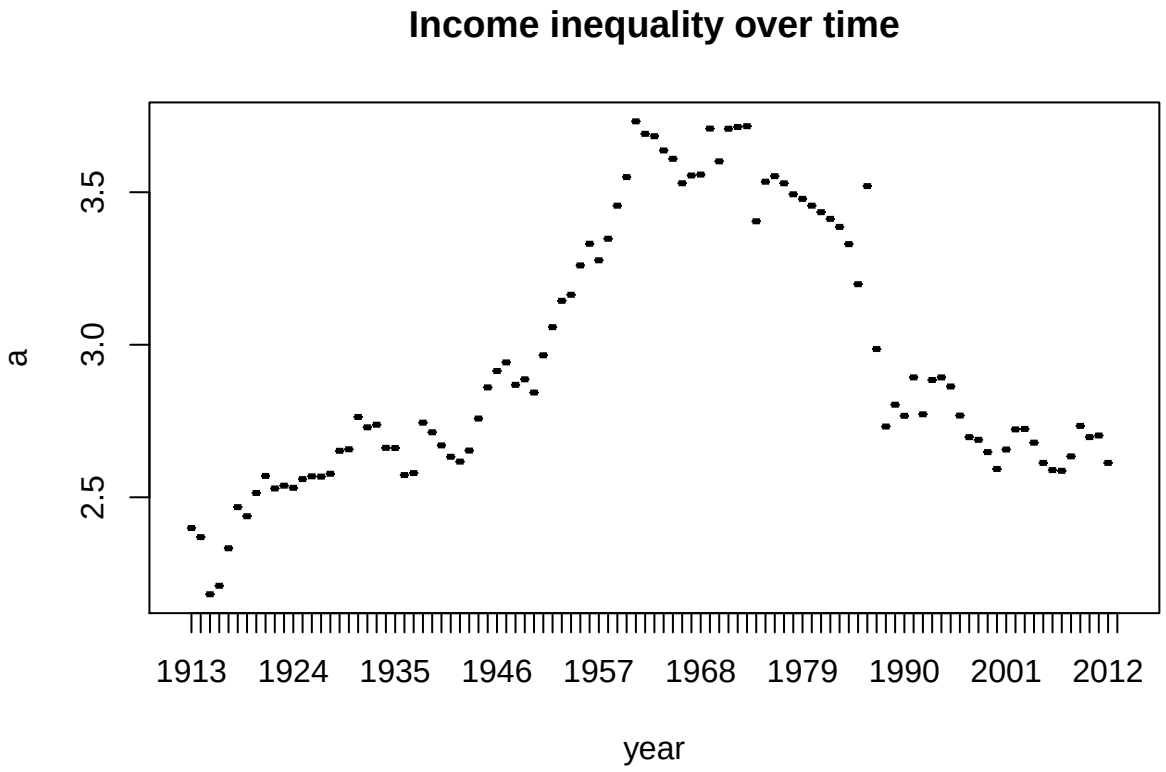
```
exponent.est_ratio <-function(p1, p2){
  a=1-(log(10)/log(p1/p2))
  return (a)
}

exponent.est_ratio(1e6,1e7)
```

```
## [1] 2
```

Part II

4.

```
estimate <- exponent.est_ratio(P99,P99.9)
income <- cbind(income,estimate)
plot(year, estimate, main='Income inequality over time',
     xlab='year',ylab='a')
```



Yes the results look reasonable. They tend to suggest that income inequality was smallest in the years 1955 to 1983 and after that started to increase as a started to decrease. However, we are not using all of the data to estimate a, so the results could be misleading. There was a boost in the economy around the 2000 which led to more income disparity.

5. We have to calculate xmin here

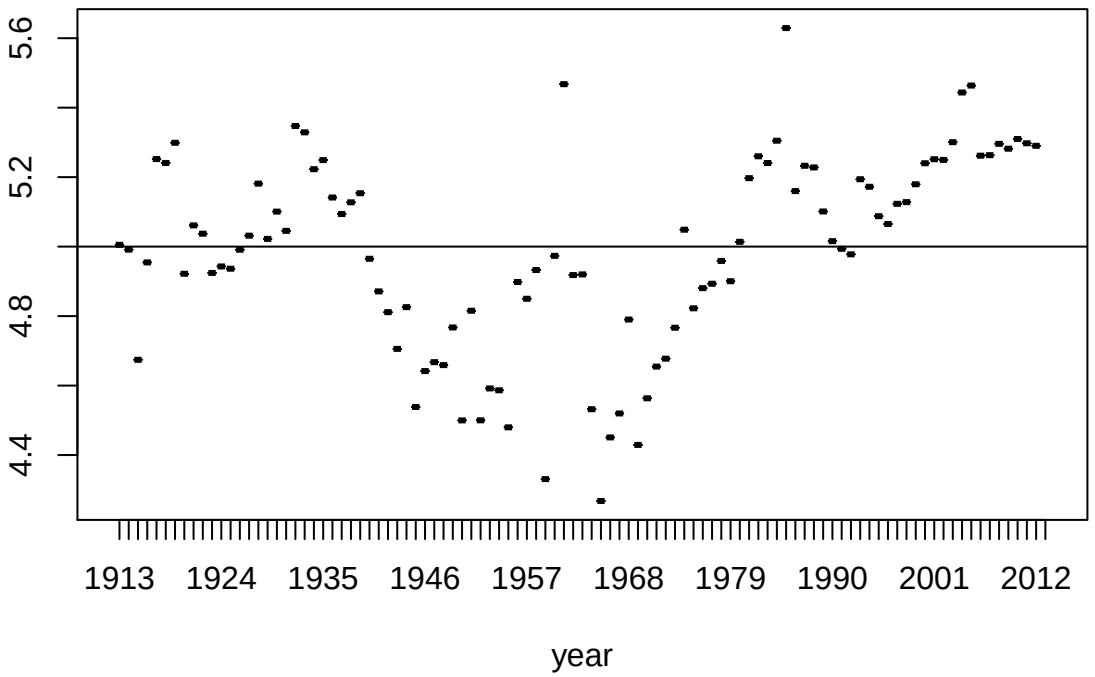
```
# method of moments
prob <- function(p1,a,w,househ.in){
  xmin<- p1/((0.01)^(1/(-a+1)))
  proppeeps<-(w/xmin)^(-a+1)
  households<-househ.in*proppeeps
  return(households)
}
a<-income$estimate[which(income$year==2012)]
p1<-income$P99[which(income$year==2012)]
households<-prob(p1,a,50000000,160681*1000)
cat(round(households,0),"household had an income of over 50 million")
```

```
## 592 household had an income of over 50 million
```

6.

```
propeople <-function(p1, p2,a){
  l=(p1/p2)^(-a+1)
  return (l)
}
prop<- propeople(income$P99.5,income$P99.9,income$estimate)
income <- cbind(income,prop)
```

```
plot(year, income$prop, xlab='year')
abline(h=5)
```



This is not a good fit because our value of a doesn't even depend on P99.5. So a better way to compute a would be if it depended on all of the data and that in return would give better results.