# Homework 5

zhang zhuohan

**Background**: .....

1. We estimate $a$ by minimizing

$$\left[\left(\frac{P99}{P99.9}\right)^{-a+1} - 10\right]^2 + \left[\left(\frac{P99.5}{P99.9}\right)^{-a+1} - 5\right]^2 + \left[\left(\frac{P99}{P99.5}\right)^{-a+1} - 2\right]^2$$

   Write a function, `percentile_ratio_discrepancies`, which takes as inputs $P99$, $P99.5$, $P99.9$ and $a$, and returns the value of the expression above. Check that when $P99 = 1e6$, $P99.5 = 2e6$, $P99.9 = 1e7$ and $a = 2$, your function returns 0.

```
percentile_ratio_discrepancies <- function(P99, P99.5, P99.9, a){
    return (((P99/P99.9)^(-a+1)-10)^2 +
            ((P99/P99.9)^(-a+1)-10)^2 +
            ((P99/P99.5)^(-a+1)-2)^2)
}
percentile_ratio_discrepancies(1e6, 2e6, 1e7, 2)
```

```
## [1] 0
```

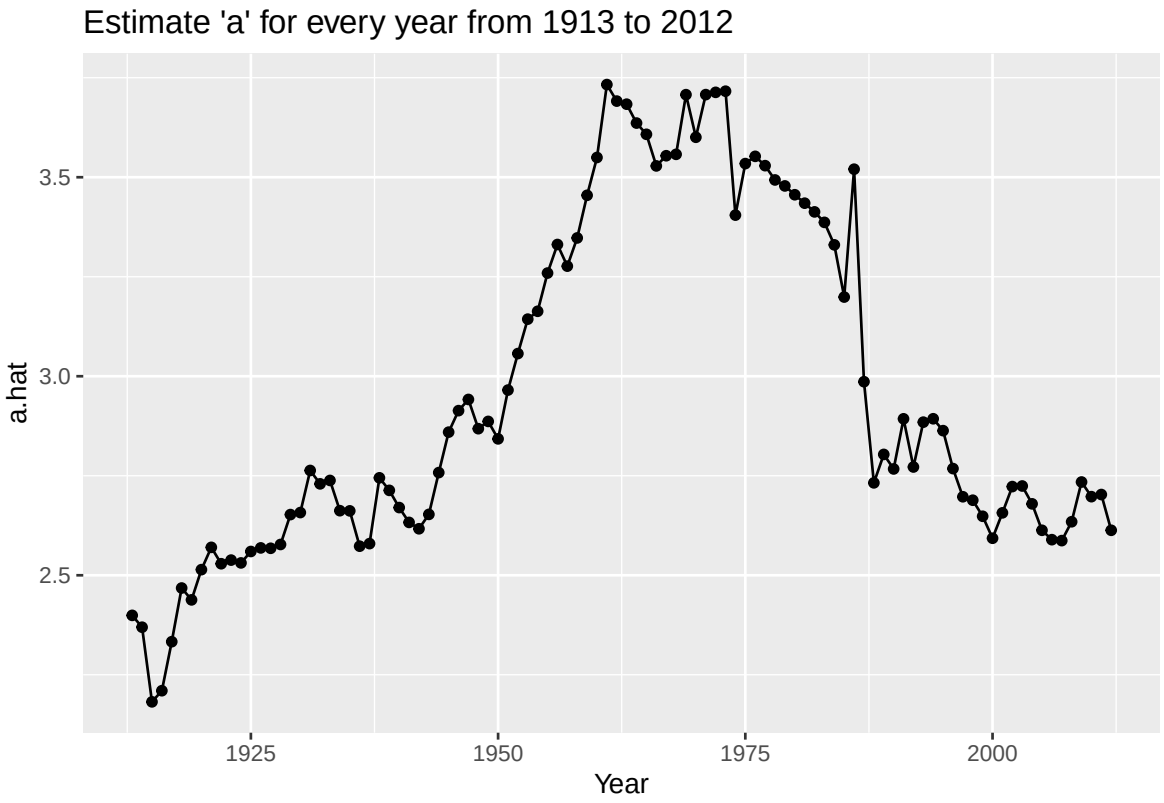
2. Write a function, `exponent.multi_ratios_est`, which takes as inputs $P99$, $P99.5$, $P99.9$, and estimates $a$. It should minimize your `percentile_ratio_discrepancies` function. The starting value for the minimization should come from (4). Check that when $P99 = 1e6$, $P99.5 = 2e6$ and $P99.9 = 1e7$, your function returns an $a$ of 2.

```
exponent.multi_ratios_est <- function(P99, P99.5, P99.9){
    a = 1 - log(10)/log(P99/P99.9)
    a.adj <- function(a){
        return(percentile_ratio_discrepancies(P99, P99.5, P99.9, a))
    }
    return(nlm(a.adj, a)$estimate)
}
exponent.multi_ratios_est(1e6, 2e6, 1e7)
```

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

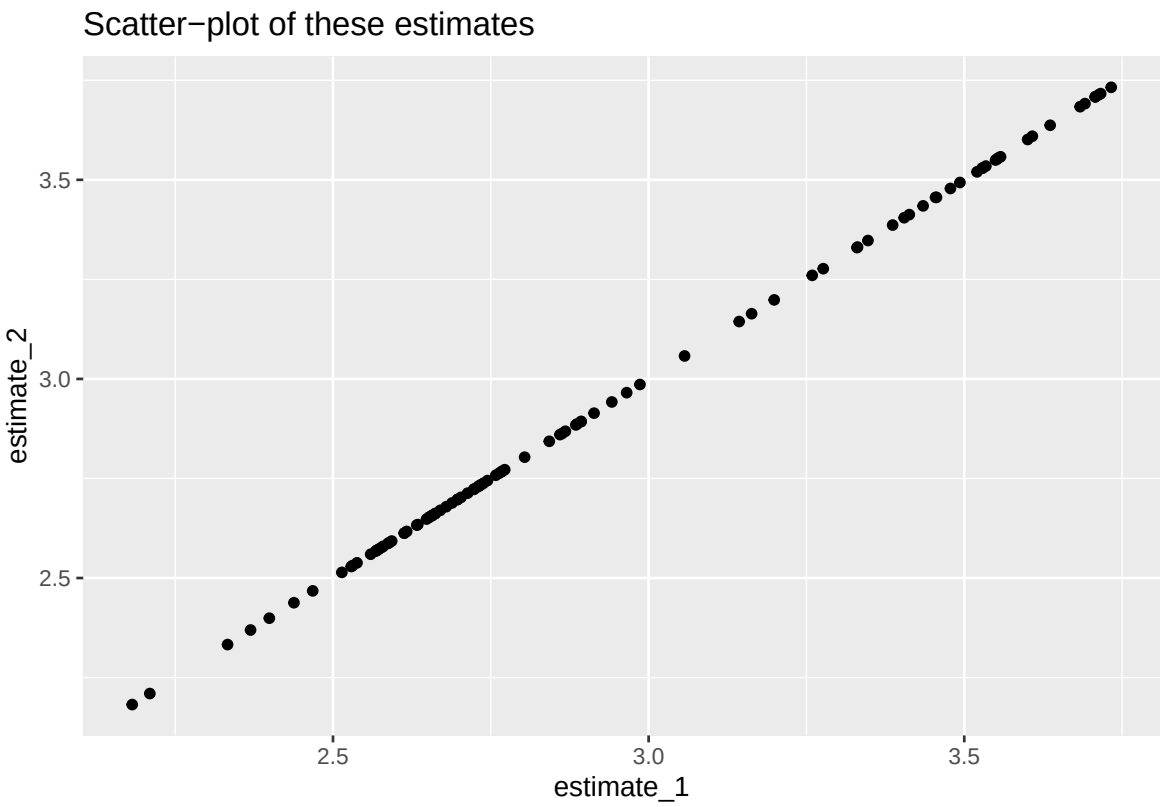
3. Write a function which uses `exponent.multi_ratios_est` to estimate $a$ for the US for every year from 1913 to 2012. (There are many ways you could do thi, including loops.) Plot the estimates; make sure the labels of the plot are appropriate.

```
library(dplyr)
library(ggplot2)
wtid.report <- read.csv("../data/wtid-report.csv", header=T)
temp <- tbl_df(wtid.report[,c(2,5:7)])
temp <- temp %>% mutate(a.hat = apply(temp, 1, function(x){
            return (exponent.multi_ratios_est(x[2], x[3], x[4]))
        }))
ggplot(temp) + geom_point(aes(Year, a.hat)) +
    geom_line(aes(Year, a.hat)) +
    labs(x = "Year", y = "a.hat",
         title = "Estimate 'a' for every year from 1913 to 2012")
```




4. Use (4) to estimate $a$ for the US for every year. Make a scatter-plot of these estimates against those from problem 3. If they are identical or completely independent, something is wrong with at least one part of your code. Otherwise, can you say anything about how the two estimates compare?

```
temp <- temp %>% mutate(estimate2 =
        1-log(10)/log(P99.income.threshold/P99.9.income.threshold))
ggplot(temp) + geom_point(aes(a.hat, estimate2)) +
labs(x = "estimate_1", y = "estimate_2",
     title = "Scatter-plot of these estimates")
```


Scatter−plot of these estimates
estimate_2
3.5
3.0
2.5
2.5
3.0
3.5
estimate_1