myHomework5

chenpeng

2024-07-06

目录

1.

```
percentile_ratio_discrepancies <- function(P99, P99.5, P99.9,
    a) {
    y1 <- (P99/P99.9)^(-a + 1) - 10
    y2 <- (P99.5/P99.9)^(-a + 1) - 5
    y3 <- (P99/P99.5)^(-a + 1) - 2
    return(y1^2 + y2^2 + y3^2)
}

# P99=1e6; P99.5=2e6; P99.9=1e7; a=2
percentile_ratio_discrepancies(1e+06, 2e+06, 1e+07, 2)
```

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

2.

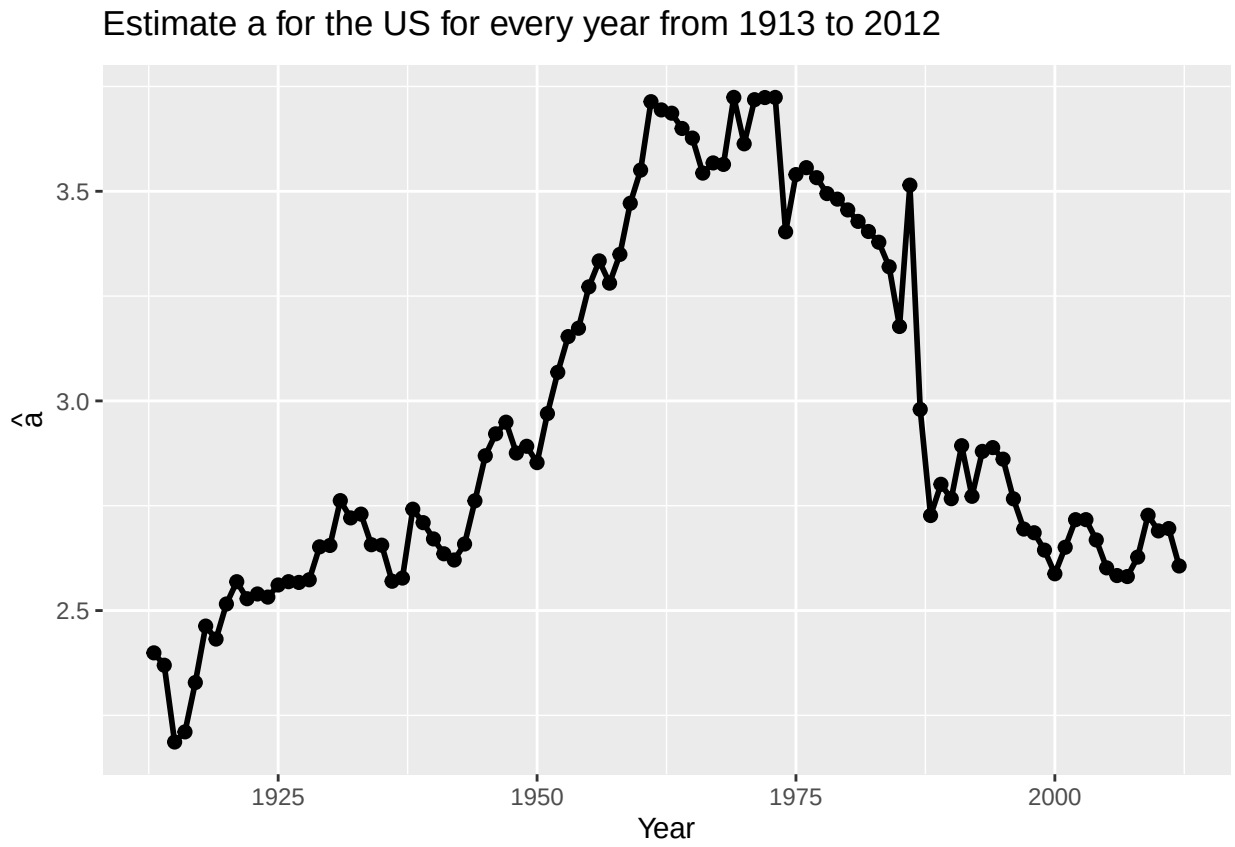
```
exponent.multi_ratios_est <- function(P99, P99.5, P99.9) {
    P99 = as.numeric(P99)
    P99.5 = as.numeric(P99.5)
    P99.9 = as.numeric(P99.9)
    a <- 1 - log(10)/(log(P99) - log(P99.9))
    a.est <- function(a) {
        return(percentile_ratio_discrepancies(P99, P99.5,
            P99.9, a))
    }
    return(nlm(a.est, a)$estimate)
}
exponent.multi_ratios_est(1e+06, 2e+06, 1e+07)
```

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

3.

```
# 读取数据
wtid <- read.csv("../../data/wtid-report.csv", header = TRUE)

# 选择需要的列，并转换为 tibble 格式
t <- tibble(wtid[, c(2, 5:7)])

# 计算 a.hat 列
a.hat <- apply(t, 1, function(x) {
  P99 <- x[2]; P99.5 <- x[3]; P99.9 <- x[4]
  return(exponent.multi_ratios_est(P99, P99.5, P99.9))
})

# 将 a.hat 添加到 t 中
t <- mutate(t, a.hat = a.hat)

# 绘制 ggplot 图形
ggplot(t, aes(x = Year, y = a.hat)) +
  geom_point(size = 2) +
```

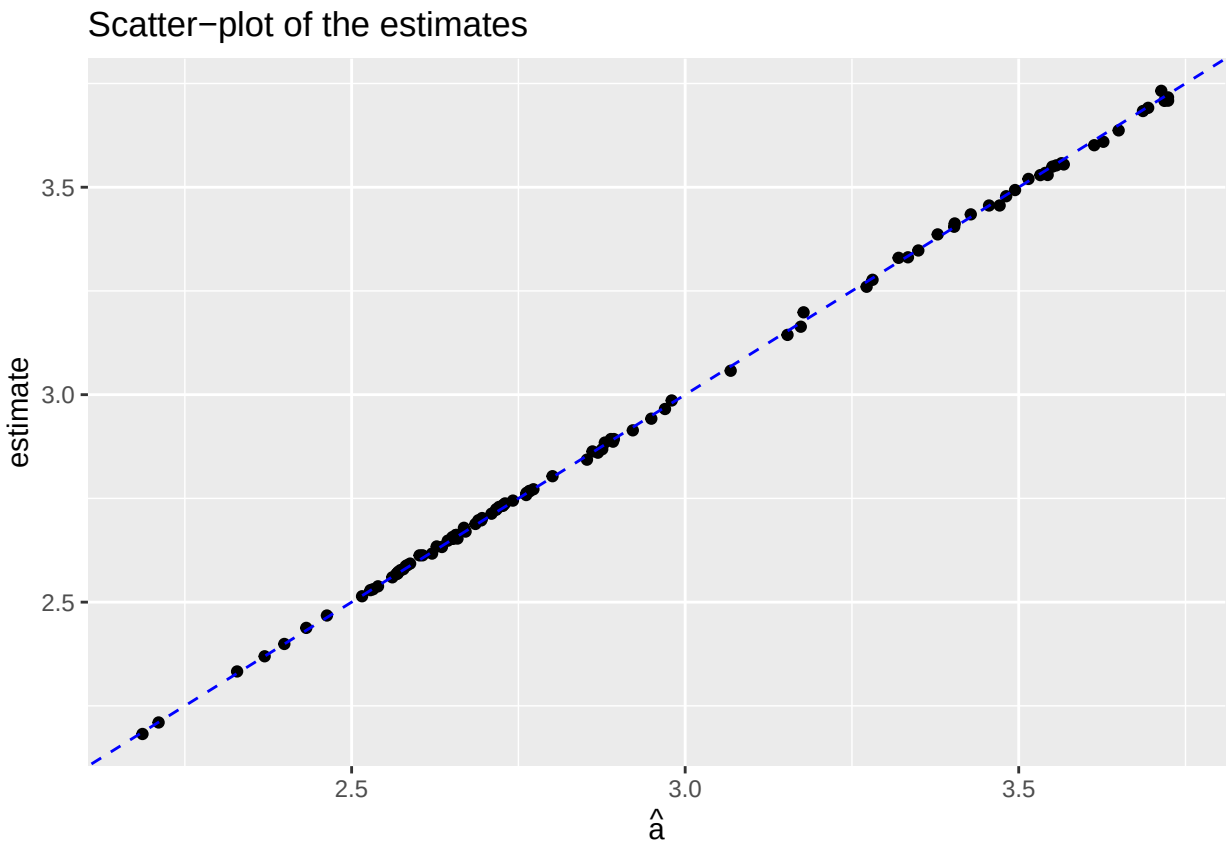
```
  geom_line(size = 1) +
  labs(
    y = expression(hat(a)),  # 使用 expression() 函数来渲染 TeX 表达式
    title = expression(paste("Estimate ", a, " for the US for every year from 1913 to 2012"))
  )
```



4.

```
t <- mutate(t, estimate = 1 - log(10)/log(P99.income.threshold/P99.9.income.threshold))
# 计算参考线的范围
range <- range(c(t$a.hat, t$estimate))

# 绘制 ggplot 图形，并添加参考线 y = x
ggplot(t, aes(x = a.hat, y = estimate)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "blue") +  # 添加 y = x 参考线
  labs(x = expression(hat(a)), y = "estimate",
```

```
       title = "Scatter-plot of the estimates") +
  xlim(range) + ylim(range)  # 设置 x 和 y 轴的范围相同，保证图形比例正确
```



在 R 中，使用 cor() 函数可以计算两个向量之间的相关系数。相关系数衡量了两个变量之间线性关系的强度和方向。

```
cor(t$a.hat, t$estimate)
```

```
## [1] 0.999883
```

二者有很强的正线性相关关系