Exercise 1

Yutong Wang

2024-10-07

(a) Regression, inference, $n = 50$, $p = 6$.
(b) Classification, prediction, $n = 20$, $p = 15$.
(c) Regression, prediction, $n = 365/7 \approx 52$, $p = 3$.

```
# getwd()
forbes2000 = read.csv("Forbes2000.csv", row.names = 1, skip = 3)

summary(forbes2000)
```

```
##       rank            name              country            category
##  Min.   :   1.0   Length:2000        Length:2000        Length:2000
##  1st Qu.: 500.8   Class :character   Class :character   Class :character
##  Median :1000.5   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1000.5
##  3rd Qu.:1500.2
##  Max.   :2000.0
##
##      sales             profits            assets          marketvalue
##  Min.   :  0.010   Min.   :-25.8300   Min.   :   0.270   Min.   :  0.02
##  1st Qu.:  2.018   1st Qu.:  0.0800   1st Qu.:   4.025   1st Qu.:  2.72
##  Median :  4.365   Median :  0.2000   Median :   9.345   Median :  5.15
##  Mean   :  9.697   Mean   :  0.3811   Mean   :  34.042   Mean   : 11.88
##  3rd Qu.:  9.547   3rd Qu.:  0.4400   3rd Qu.:  22.793   3rd Qu.: 10.60
##  Max.   :256.330   Max.   : 20.9600   Max.   :1264.030   Max.   :328.54
##                    NA's   :5
```

```
class(forbes2000$name)
```

```
## [1] "character"
```

```
forbes2000$name = as.factor(forbes2000$name)
class(forbes2000$name)
```

```
## [1] "factor"
```

```
attach(forbes2000)
# country
class(country)
```

```
## [1] "character"
```

```
country = as.factor(country)
class(country)
```

```
## [1] "factor"
```

```
# levels(country)

length(levels(country))
```

```
## [1] 61
```

```
category = as.factor(category)

length(levels(category))
```

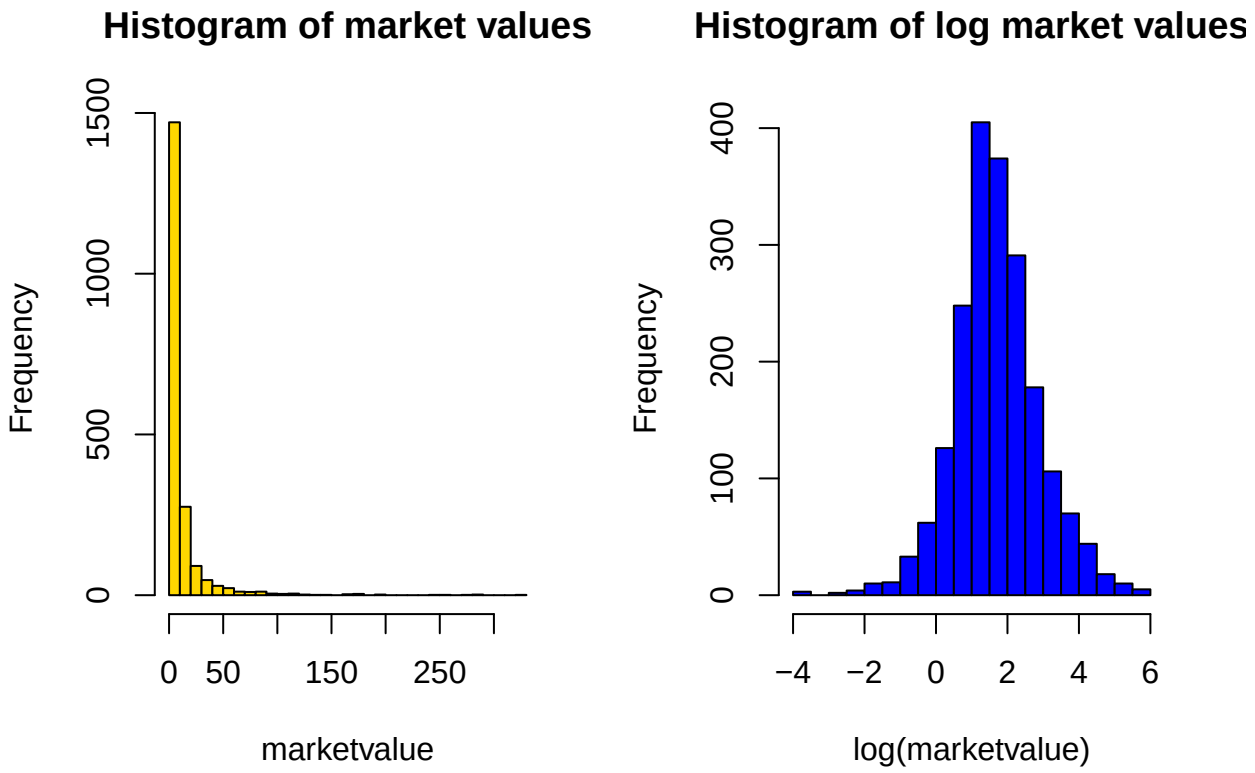
```
## [1] 27
```

```
name[rank<=20]
```

```
##  [1] Citigroup               General Electric        American Intl Group
##  [4] ExxonMobil              BP                      Bank of America
##  [7] HSBC Group              Toyota Motor            Fannie Mae
## [10] Wal-Mart Stores         UBS                     ING Group
## [13] Royal Dutch/Shell Group Berkshire Hathaway      JP Morgan Chase
## [16] IBM                     Total                   BNP Paribas
## [19] Royal Bank of Scotland  Freddie Mac
## 2000 Levels: Aareal Bank ABB Group Abbey National ... Zurich Financial Services
```

```
# rank

# rank <= 10
```

```
par(mfrow=c(1,2))
hist(marketvalue, nclass=25, main="Histogram of market values", col="gold")
hist(log(marketvalue), nclass=25, main="Histogram of log market values", col="blue")
```



```r
mean(profits)
```

```
## [1] NA
```

```r
mean(profits, na.rm = T)
```

```
## [1] 0.3811328
```

```r
median(profits[country=="United States"])
```

```
## [1] NA
```

```r
median(profits[country=="United States"], na.rm=T)
```

```
## [1] 0.24
```

```r
median(profits[country=="United Kingdom"])
```

```
## [1] NA
```

```r
median(profits[country=="United Kingdom"], na.rm=T)
```

```
## [1] 0.205
```

```r
a = 3

a == 2
```

```
## [1] FALSE
```

```r
# basic logics
TRUE & FALSE
```

```
## [1] FALSE
```

```r
TRUE | FALSE
```

```
## [1] TRUE
```

```r
name[( (country=="Germany")&(profits<0) )]
```

```
##  [1] Allianz Worldwide       Deutsche Telekom        E.ON
##  [4] HVB-HypoVereinsbank     Commerzbank             Infineon Technologies
##  [7] BHW Holding             Bankgesellschaft Berlin W&W-Wustenrot
## [10] mg technologies         Nurnberger Beteiligungs SPAR Handels
## [13] Mobilcom
## 2000 Levels: Aareal Bank ABB Group Abbey National ... Zurich Financial Services
```

```r
table(category[country=="Bermuda"])
```

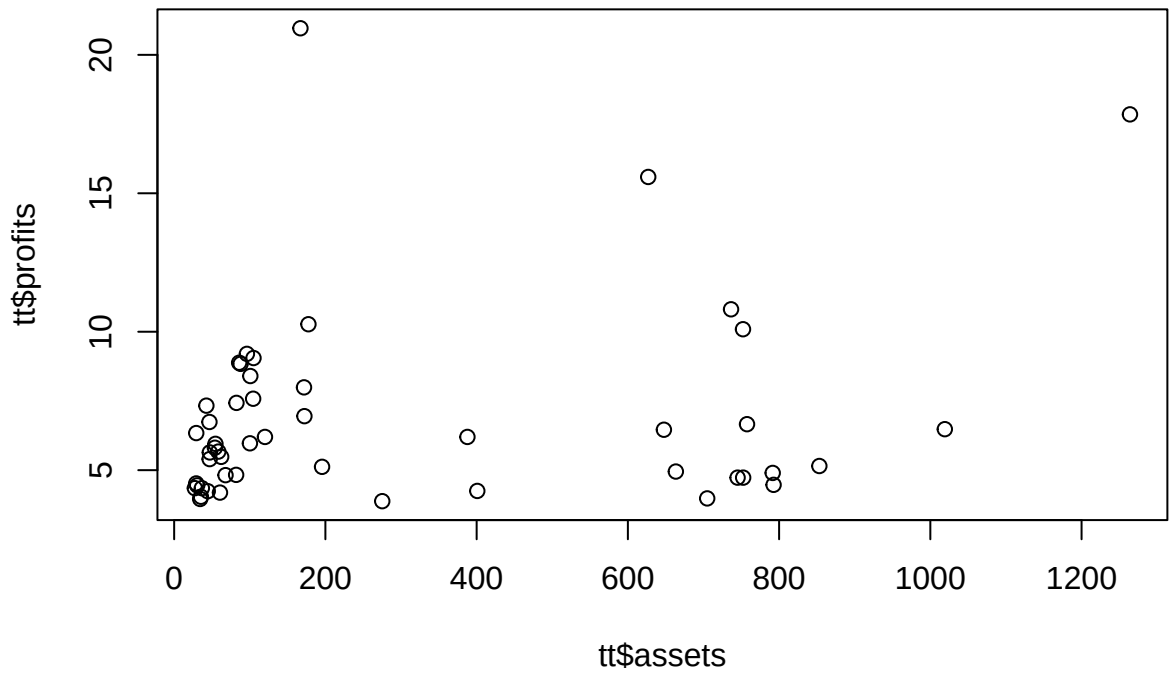
```
##
##              Aerospace & defense                          Banking
##                                0                                1
##     Business services & supplies                   Capital goods
##                                0                                1
##                         Chemicals                   Conglomerates
##                                0                                2
##                      Construction               Consumer durables
##                                0                                0
##           Diversified financials           Drugs & biotechnology
##                                0                                0
##             Food drink & tobacco                     Food markets
```

```
##                                1                                1
## Health care equipment & services     Hotels restaurants & leisure
##                                0                                0
##    Household & personal products                        Insurance
##                                0                               10
##                         Materials                            Media
##                                0                                1
##             Oil & gas operations                        Retailing
##                                2                                0
##                   Semiconductors             Software & services
##                                0                                1
##  Technology hardware & equipment      Telecommunications services
##                                0                                0
##                 Trading companies                   Transportation
##                                0                                0
##                         Utilities
##                                0
```

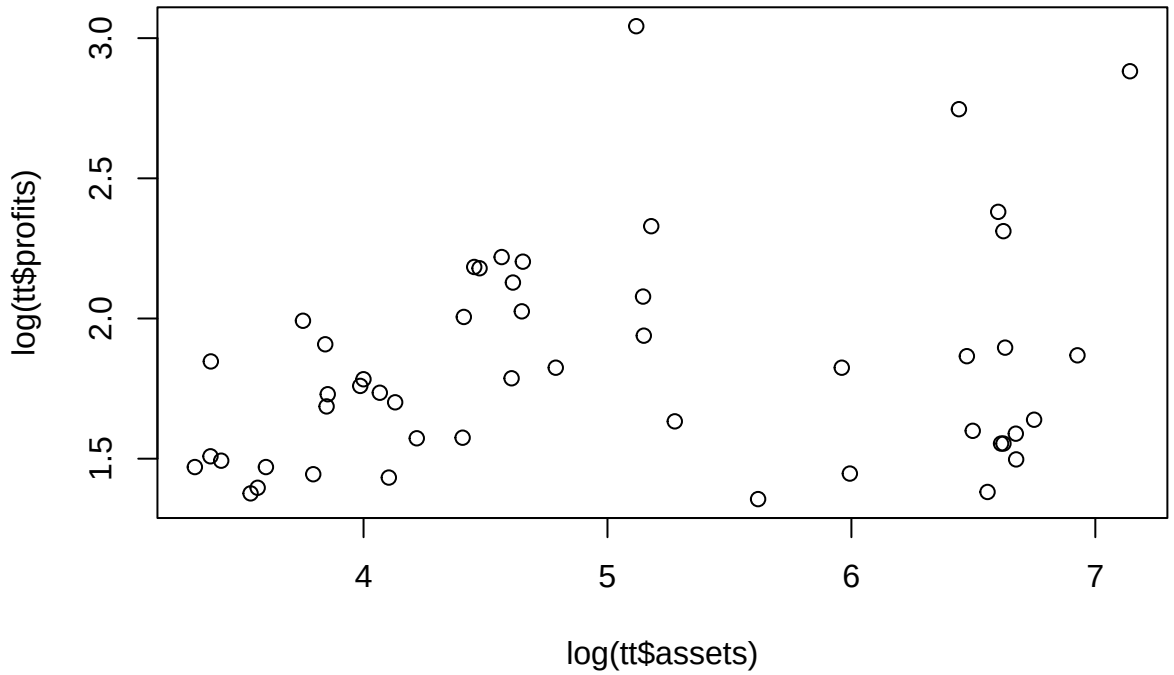
```
profits.sort = sort(profits, decreasing = T)
profit_ori = profits
# sort in descending order


tt = subset(forbes2000, profits>=profits.sort[50])
# subset can create a sub-dataframe

dim(tt)
```

```
## [1] 50  8
```

```
# par(mfrow=c(1,1))

plot(tt$assets, tt$profits)
```



```
plot(log(tt$assets), log(tt$profits))
```




```
plot(log(tt$assets), log(tt$profits), xlab="log Assets", ylab="log Profits", cex.lab=2, pch=20, col="da
```