GR5223 - HW2

MJS2364

15 February 2018

Problem 8

The 10 largest U.S. industrial corporations in the 1990 Fortune 500 yielded the following data (in millions of dollars):

```
# Loads dataset from the data CSV file
data <- read.csv("Companies.csv", as.is = TRUE)
#data <- as.matrix(data)

print(data)
```

```
##     Sales Profits Assets
## 1  126974    4224 173297
## 2   96933    3835 160893
## 3   86656    3510  83219
## 4   63438    3758  77734
## 5   55265    3939 128344
## 6   50976    1809  39080
## 7   39069    2946  38528
## 8   36156     359  51038
## 9   35209    2480  34715
## 10  32416    2413  25636
```

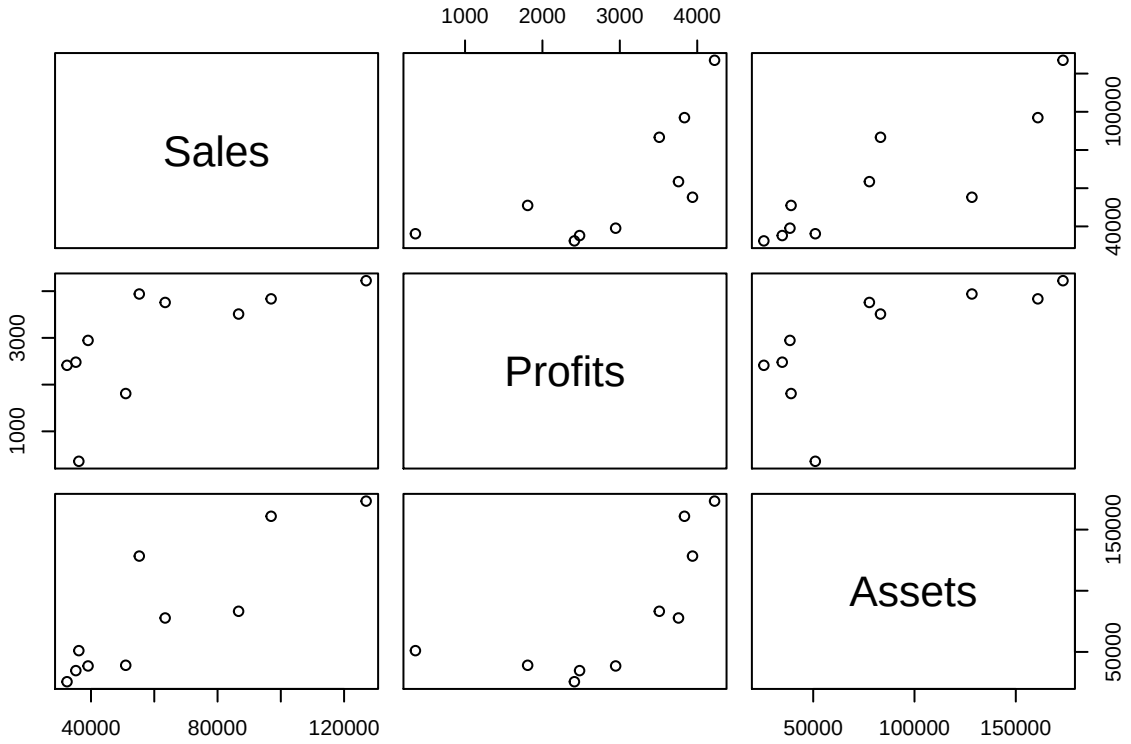
(a) Study the univariate distributions by constructing boxplots and/or histograms.

```
par(mfrow=c(2,3))

hist(data$Sales,6, xlab="Sales", main="Histogram of Sales")
hist(data$Profits,6, xlab = "Profits", main="Histogram of Profits")
hist(data$Assets,6, xlab = "Assets", main="Histogram of Assets")

boxplot(data$Sales, xlab="Sales", main="Boxplot of Sales")
boxplot(data$Profits, xlab = "Profits", main="Boxplot of Profits")
boxplot(data$Assets, xlab = "Assets", main="Boxplot of Assets")
```

Looking at the histograms we observe that the univariate distributions do not follow any obvious, known distribution. Sales distribution is skewed to the right while Profits are skewed to the left. This skewness is also clearly visible on the boxplots. Assets can be seen as bimodal with highest frequency towards the left end (0~40,000) and right end of the range (160,000+), and lower frequency for profits in between.

(b) Construct a scatterplot matrix and describe the relationships among the three variables.

```
library(graphics)
pairs(data)
```



We observe that Assets seem to have a positive linear association with Sales. The larger the assets, the larger the sales (and vice versa). However, Profits and Assets seem to have a positive curvilinear association. Sales and Profits also seem to have a rather nonlinear association.

(c) Calculate summary statistics: means, standard deviations, and correlations.

```
ave_sales <- mean(data[,1])
ave_profits <- mean(data[,2])
ave_assets <- mean(data[,3])

stdev_sales <- sd(data[,1])
stdev_profits <- sd(data[,2])
stdev_assets <- sd(data[,3])

corr_matrix <- cor(data)

print(ave_sales)
```

```
## [1] 62309.2
```

```
print(ave_profits)
```

```
## [1] 2927.3
```

```
print(ave_assets)
```

```
## [1] 81248.4
```

```r
print(stdev_sales)
```

```
## [1] 31630.8
```

```r
print(stdev_profits)
```

```
## [1] 1195.834
```

```r
print(stdev_assets)
```

```
## [1] 54593.86
```

```r
print(corr_matrix)
```

```
##             Sales   Profits    Assets
## Sales   1.0000000 0.6761554 0.8754869
## Profits 0.6761554 1.0000000 0.6993102
## Assets  0.8754869 0.6993102 1.0000000
```

The Sales variable has a mean of 62309.2 and a standard deviation of 31630.8 The Profits variable has a mean of 2927.3 and a standard deviation of 1195.834 The Assets variable has a mean of 81248.4 and a standard deviation of 54593.86

Correlation between Sales and Profits is 0.676, between Sales and Assets it is 0.875, and between Profits and Assets it is 0.699.