

Lab 1: Exploratory Analysis of CEO Salary Data

Carmen Easterwood & Andrew Kabatznick

January 29, 2016

Introduction

Aaa

```
setwd("~/Desktop/MIDS/Statistics/stats_lab1")
ceosal <- load("ceo_w203.RData", ceo.env <- new.env())
ceo.df <- ceo.env[["CEO"]]
```

Univariate Analysis of Key Variables

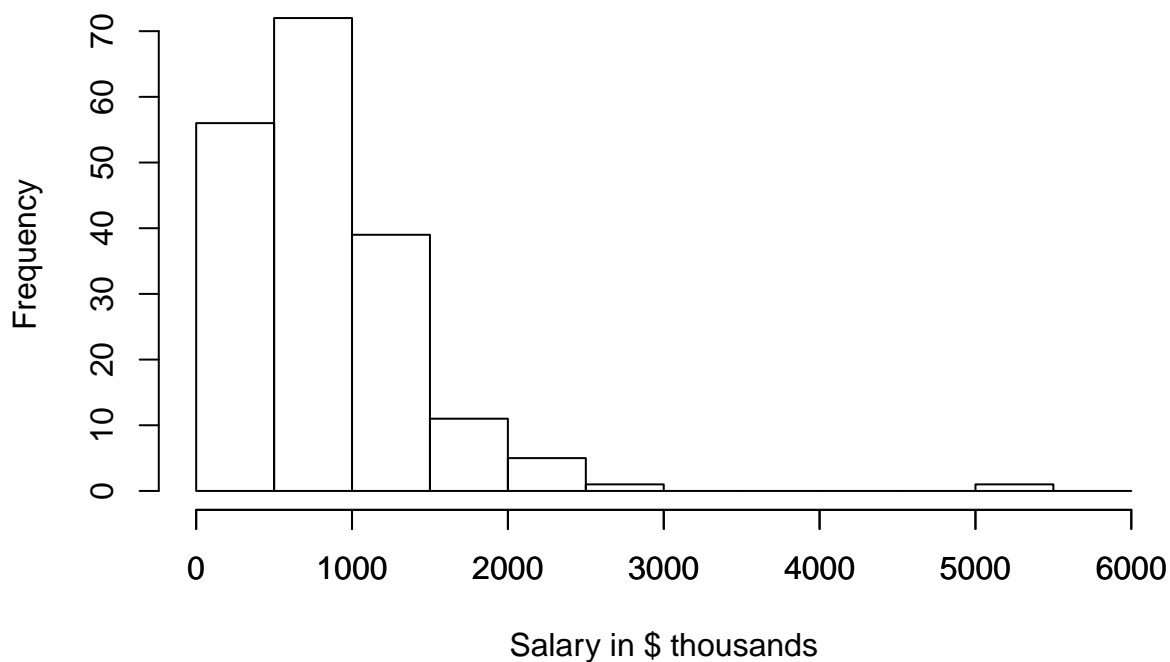
In the following section we will perform a univariate analysis of each of the variables in this dataset. Salary is our outcome variable, and profits and mktval are the key measures of market performance.

Salary

CEO salary distribution is strongly skewed right.

```
hist(ceo.df$salary, main = "Histogram of CEO Salary in 1990",
     xlab = "Salary in $ thousands", breaks = seq(0, 6000, by = 500))
axis(1, at = seq(0, 6000, by = 1000))
```

Histogram of CEO Salary in 1990



Median salary is \$697 thousand, and there is one extreme outlier at \$5.3 million.

```
summary(ceo.df$salary)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    100.0   467.0   697.0   852.9  1101.0  5299.0
```

Age

CEO age peaks between 50 and 65 years old, but ranges all the way from 21 to 86.

```
hist(ceo.df$age, breaks = 14, main = "Histogram of CEO Age", xlab = "Age")
```



```
summary(ceo.df$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    21.00   51.00   57.00   55.78   61.00   86.00
```

The variance is large:

```
var(ceo.df$age)
```

```
## [1] 85.37996
```

College

College is a dummy variable that takes a value of 1 if the CEO is a college graduate and 0 otherwise.

```
pct.college <- (sum(ceo.df$college) / length(ceo.df$college))
```

96.2% of the CEOs in this dataset are college graduates.

Grad

Grad is a dummy variable that takes a value of 1 if the CEO holds an advanced degree and 0 otherwise.

```
pct.grad <- (sum(ceo.df$grad) / length(ceo.df$grad))
```

55.1% of the CEOs in this dataset are college graduates.

Note: Should we mention there are 2 CEOs with an advanced degree but no college degree?
(Can that be correct?)

Key Bivariate Relationships

Aaa

Relationship 1

Aaa

Possible Secondary Variables

Aaa

Potential Confounding Effects

Aaa

Variable Coding Issues and Missing Values

Aaa

Conclusion

Aaa