Exploratory Data Analysis and Descriptive Statistics with R

Mitch Harden

March 12, 2016

Analysis and Report were generated using R version 3.2.4 (2016-03-10) and KNITR via RStudio.

Load R libraries

```
library(foreign)
library(ggplot2)
library(moments)
```

Data acquired from source, uncompressed, and read into R:

```
setwd("~/Maryville/EDL735")
download.file("http://spss5.allenandunwin.com.s3-website-ap-southeast-2.amazonaws.com/Files/sleep5ED.zip
unzip("sleep5ED.zip")
sleep5ed = read.spss("sleep5ED/sleep5ED.sav", to.data.frame=TRUE)
attach(sleep5ed)
```

1. Generate appropriate descriptive statistics to answer the following:

1.1 What percentage of respondents are female? (sex)

```
prop.table(table(sex))
```

```
## sex
##    female      male
## 0.5535055 0.4464945
```

Interpretation

55.35% of participants were female.

1.2 What is the average age of the sample? (age)

```
summary(age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   34.00   44.00   43.87   54.00   84.00      23
```

Interpretation

The average age of participants was 43.87 years

1.3 What is the percentage of the samle indicated that they had a prolem with their sleep (probsleeprec)

```
prop.table(table(probsleeprec))
```

```
## probsleeprec
##        no       yes
## 0.5650558 0.4349442
```

Interpretation

43.49% of participants reported problems with their sleep.

1.4 What is the median number of hours sleep per weeknight? (hourwnit)

```
summary(hourwnit)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   3.000   6.000   7.000   6.972   8.000  10.000       1
```

Interpretation

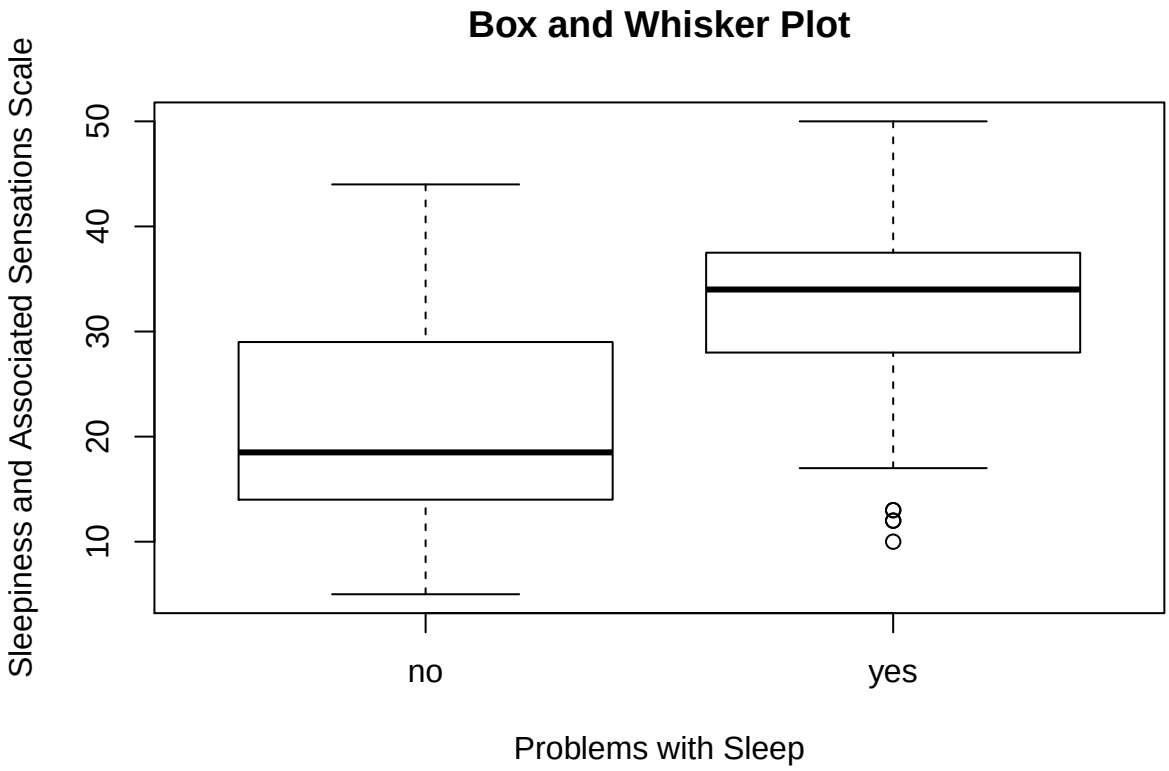
Participants slept a median of 7 hours on week nights.

2. Assess the distribution of scores on the Sleepiness and Associated Sensations Scale (toSAS) for people who feel thatthey do/don't have a sleep problem (probsleeprec).

```
detach(sleep5ed)
# Subset the Data by Sleep Problems
sleep5edclean <- subset(sleep5ed,!(is.na(sleep5ed$totsas)))
noprobs <- subset(sleep5edclean, probsleeprec=="no")
yesprobs <- subset(sleep5edclean, probsleeprec=="yes")
nptotsas <- data.frame(noprobs$totsas)
yptotsas <- data.frame(yesprobs$totsas)
```

2.1 Are there any Outliers on this scale that you would be concerned about?

```
boxplot(totsas ~ probsleeprec, data= sleep5edclean,
        xlab= "Problems with Sleep", ylab = "Sleepiness and Associated Sensations Scale",
        main = "Box and Whisker Plot")
```



Interpretation

There are three cases among those with sleep problems that may constitute outliers scoring far below expeccted ont he Sleepiness and Associated Sensations Scale. Even still, group differences between those with and those without sleep problems are still signifcant.

2.2 Are scores normally distributed for each group?

```
#Skewness and Kurtosis
skewness(noprobs$totsas, na.rm=TRUE)
```

```
## [1] 0.4878863
```

```
kurtosis(noprobs$totsas, na.rm=TRUE) -3
```

```
## [1] -0.8056236
```

```
#Histogram for NoSleep Problems
ggplot(data.frame(nptotsas), aes(x = noprobs$totsas)) +
  geom_histogram(aes(y = ..density..), breaks=seq(5, 50, by = 2), fill = 'red', alpha = 0.5) +
  geom_density(colour = 'blue') + xlab(expression(bold('Sleepiness and Associated Sensations Scale'))) +
  ylab(expression(bold('Density'))) +
  ggtitle("Reports No Sleep Problems")
```

Interpretation

The skewness of the Sleepiness and Associated Sensation Scale totals for the people that feel they do not have sleep problems is 0.488 indicating a skew to the left, while kurtosis is -.806 indicating a platykurtic distribution

```
#Skewness and Kurtosis
skewness(yesprobs$totsas, na.rm=TRUE)
```

```
## [1] -0.6344499
```

```
kurtosis(yesprobs$totsas, na.rm=TRUE) -3
```

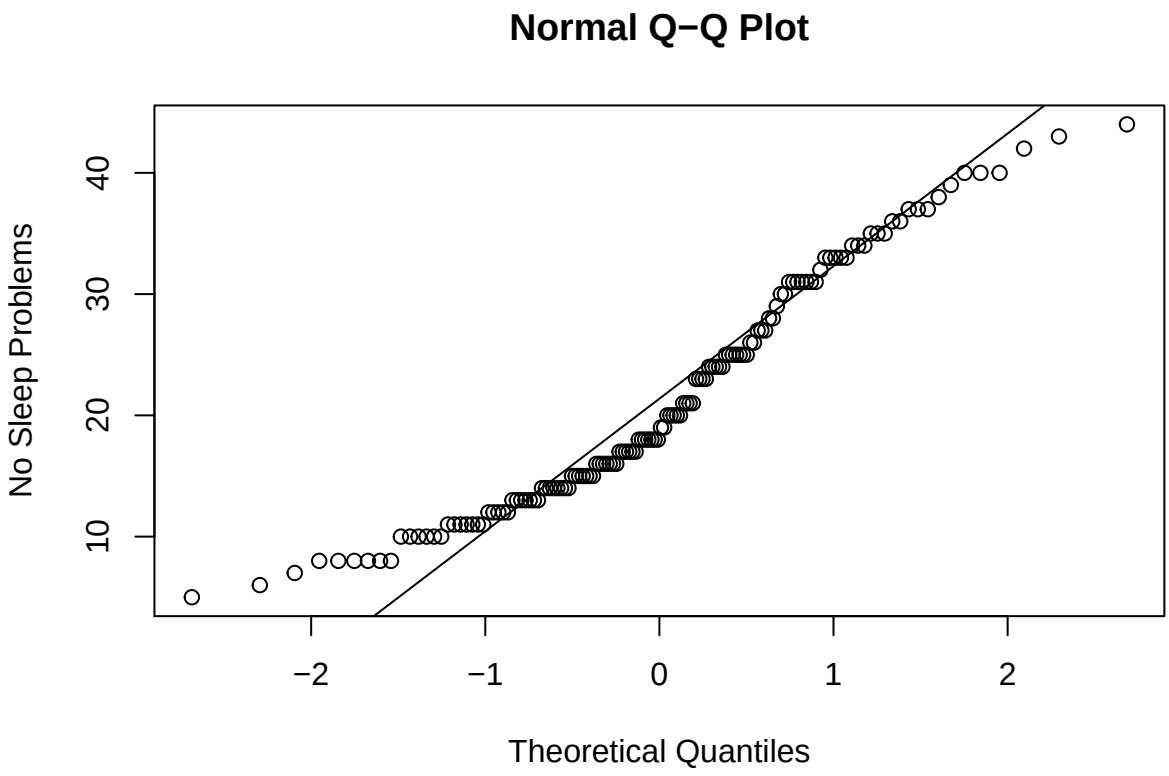
```
## [1] 0.08912234
```

```
#Histogram for Yes Sleep Problems
ggplot(data.frame(yptotsas), aes(x = yesprobs$totsas)) +
  geom_histogram(aes(y = ..density..), breaks=seq(5, 50, by = 2), fill = 'red', alpha = 0.5) +
  geom_density(colour = 'blue') +
  xlab(expression(bold('Sleepiness and Associated Sensations Scale'))) +
  ylab(expression(bold('Density'))) +
  ggtitle("Reports Sleep Problems")
```

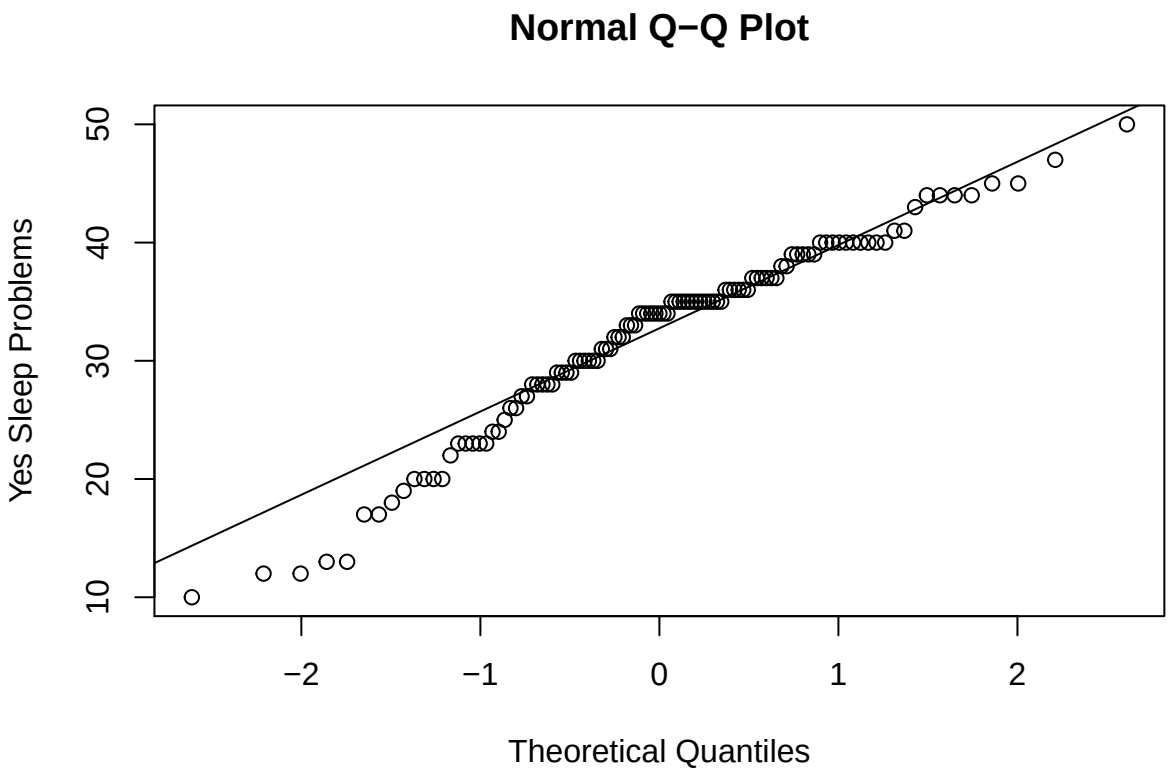
Interpretation

The skewness of the Sleepiness and Associated Sensation Scale totals for the people that feel they do have sleep problems is -0.634 indicating a skew to the right, while kurtosis is .089 indicating a relatively normal distribution

```
#Generate Q-Q plots
qqnorm(noprobs$totsas, ylab = "No Sleep Problems") ; qqline(noprobs$totsas)
```




```
qqnorm(yesprobs$totsas, ylab = "Yes Sleep Problems") ; qqline(yesprobs$totsas)
```



Interpretation

The Q-Q Plots for both groups depict a fairly normal distribution with a light amount of tailing, particularly in among the people reporting sleep problems.