ECON 4101 Econometrics
CM03 Homework

Pranav Singh

Jan 20, 2017

Problems 1-6

```
fph <- read.csv("http://evansresearch.us/DSC/Spring2017/ECMT/Data/fphB752.csv", header = T)
fph.orig <- fph
fph <- fph$fph
summary(fph)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3385    7796    8186   10380    8622  151700
```

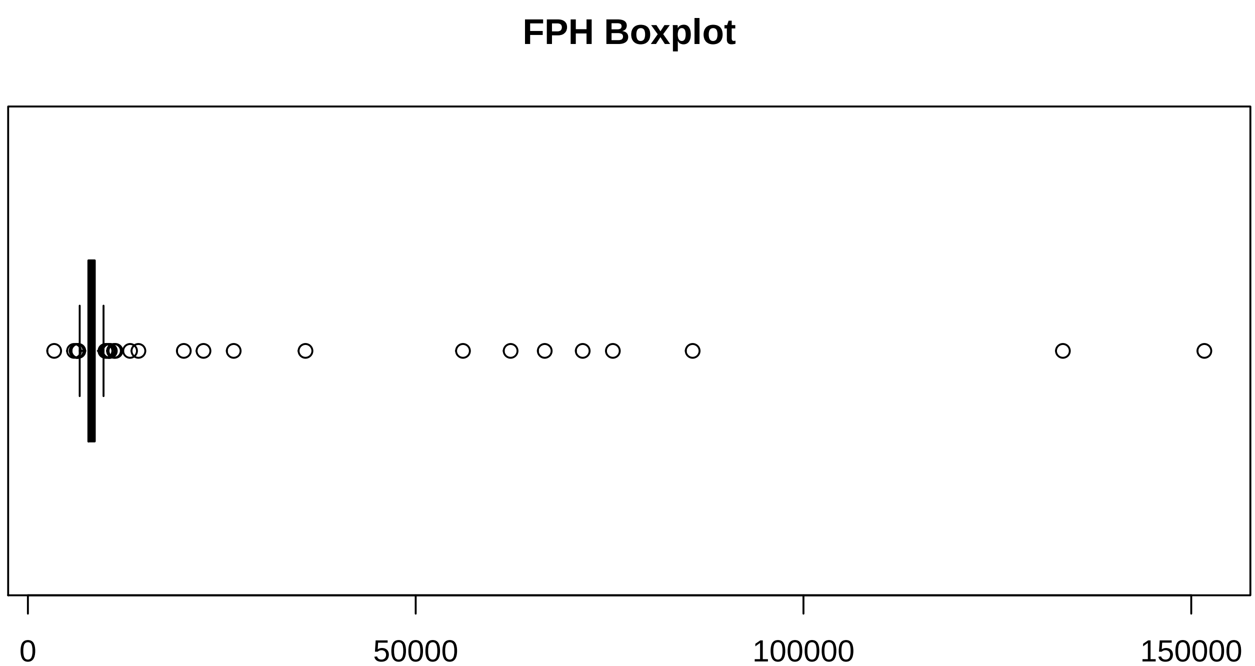
```
print(paste0("Standard deviation: ", sd(fph)))
```

```
## [1] "Standard deviation: 13434.9325447197"
```

```
print(paste0("Coefficient of variation: ", sd(fph)/mean(fph)))
```

```
## [1] "Coefficient of variation: 1.29413208531688"
```

```
fph.boxplot <- boxplot(fph, horizontal = T, main = "FPH Boxplot")
```



Problem 7

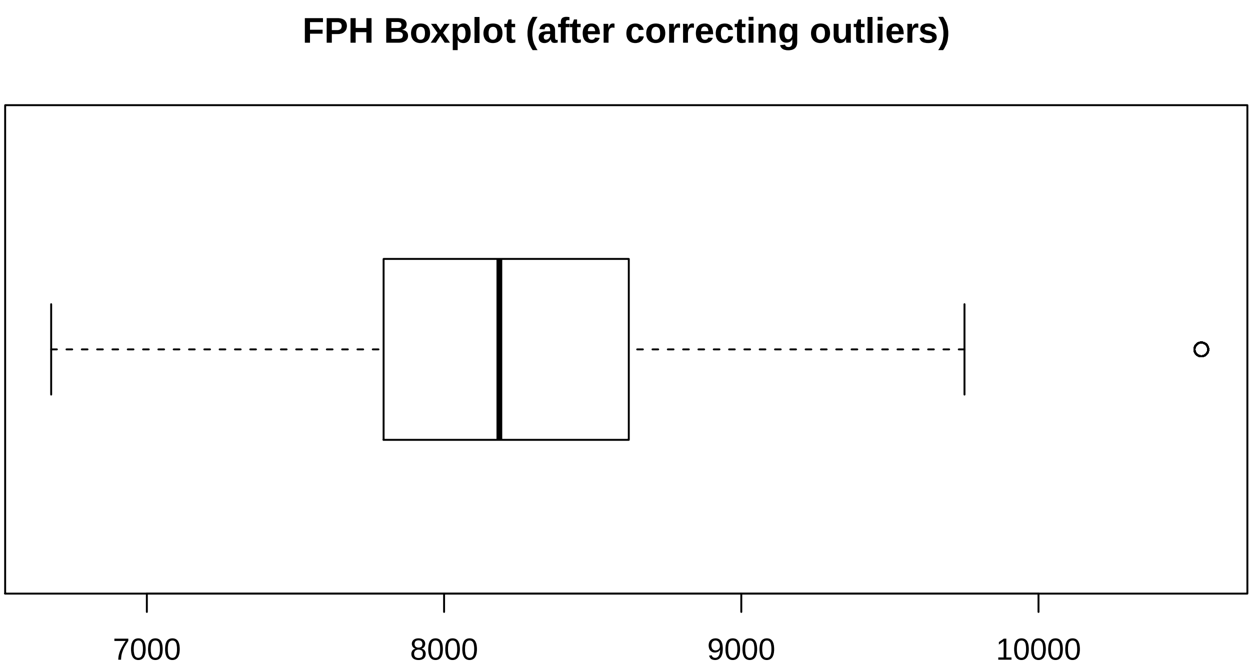
```r
q1 <- quantile(fph, 0.25)
q3 <- quantile(fph, 0.75)
iqr = q3 - q1
low.whisker <- q1 - 1.5 * iqr
high.whisker <- q3 + 1.5 * iqr

q.05 <- quantile(fph, 0.05)
q.95 <- quantile(fph, 0.95)
fph[fph < low.whisker] <- q.05
fph[fph > high.whisker] <- q.95

summary(fph)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6678    7796    8186    8330    8622   10550
```

```r
print(paste0("Standard deviation: ", sd(fph)))
```

```
## [1] "Standard deviation: 809.151139066041"
```

```r
fph.boxplot <- boxplot(fph, horizontal = T, main = "FPH Boxplot (after correcting outliers)")
```



Problem 8

```r
invisible(library(data.table))
# convert a copy of mtcars data.frame into a data.table for convenience :)
mtcars <- setDT(copy(mtcars))
mtcars <- mtcars[cyl %in% c(4, 6), ]
setorder(mtcars, cyl)
```

```r
mtcars[, .(.N, mean.mpg = mean(mpg)), by = .(cyl)]
```

```
##    cyl  N mean.mpg
## 1:   4 11 26.66364
## 2:   6  7 19.74286
```

```r
var.test(mpg ~ cyl, data = mtcars)
```

```
##
##  F test to compare two variances
##
## data:  mpg by cyl
## F = 9.6261, num df = 10, denom df = 6, p-value = 0.01182
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   1.762592 39.198688
## sample estimates:
## ratio of variances
##           9.626086
```

```r
t.test(mpg ~ cyl, data = mtcars, paired = F, var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by cyl
## t = 4.7191, df = 12.956, p-value = 0.0004048
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.751376 10.090182
## sample estimates:
## mean in group 4 mean in group 6
##        26.66364        19.74286
```

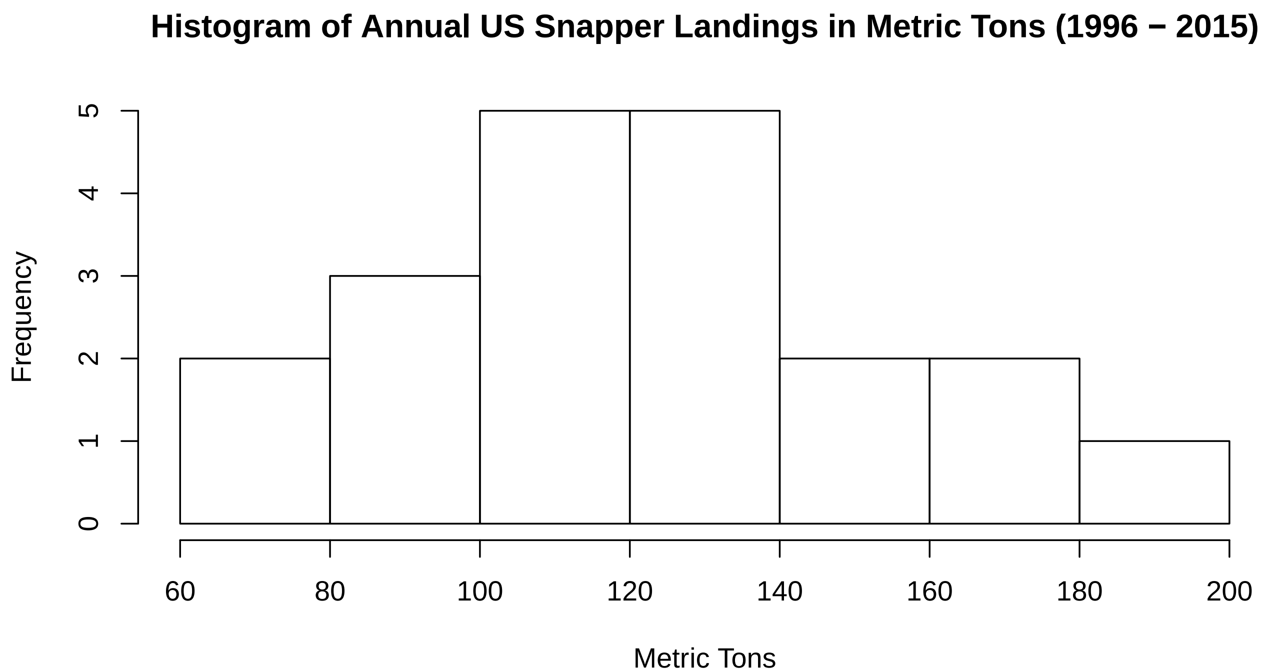
```r
# for kicks and giggles, let's verify:
x <- mtcars[cyl == 4, ]$mpg
y <- mtcars[cyl == 6, ]$mpg
x.n <- length(x)
x.var <- var(x)
x.mean <- mean(x)
y.n <- length(y)
y.var <- var(y)
y.mean <- mean(y)

t.statistic <- -1 * abs((x.mean - y.mean))/sqrt(x.var/x.n + y.var/y.n)
t.degf <- (x.var/x.n + y.var/y.n)^2/(x.var^2/x.n^2/(x.n - 1) + y.var^2/y.n^2/(y.n -
    1))
t.pvalue <- 2 * pt(t.statistic, t.degf)
print(paste0("manual p-value calculation for Welch's t-test: ", t.pvalue))
```

```
## [1] "manual p-value calculation for Welch's t-test: 0.000404849534170228"
```

The tiny p-value of approximately 0.000405 from Welch's t-test suggests that we have sufficent evidence at the 95% confidence level to reject the hypothesis that the mean fuel economy (mpg) of 6 cylinder cars is the same as 4 cylinder cars.

Problem 9

```r
# http://www.st.nmfs.noaa.gov/commercial-fisheries/commercial-landings/annual-landings/index
# Species: snappers Years: 1996-2015 Geographical area: All States
snappers <- setDT(read.csv("~/Downloads/MF_ANNUAL_LANDINGS.RESULTS", skip = 4))
title <- "Histogram of Annual US Snapper Landings in Metric Tons (1996 - 2015)"
hist(snappers$Metric.Tons, main = title, xlab = "Metric Tons")
```



Problem 10

```r
women <- setDT(copy(women))
print(paste0("Sample covariance: ", cov(women$height, women$weight)))
```

```
## [1] "Sample covariance: 69"
```

```r
print(paste0("Population covariance: ", sum((women$height - mean(women$height)) *
    (women$weight - mean(women$weight)))/nrow(women)))
```

```
## [1] "Population covariance: 64.4"
```

```r
print(paste0("Population covariance ( = (n-1)/n * Sample covariance ): ", (nrow(women) -
    1)/(nrow(women)) * cov(women$height, women$weight)))
```

```
## [1] "Population covariance ( = (n-1)/n * Sample covariance ): 64.4"
```