Math189 HW4

Yuetong Lyu

5/12/2019

Problem 1

```
boot.t.test = function(x, y, conf = 0.95, B = 9999){
  m = length(x)
  n = length (y)
  sehat = sqrt((var(x)/m) + (var(y)/n))
  diffmean = mean(x) - mean(y)
  counts = 0
  D = diffmean/sehat
  t_ratio = numeric(B)
  for(b in 1:B){
    xboot = sample(x, m, replace = TRUE)
    yboot = sample(y, n, replace = TRUE)
    t_ratio[b] = (mean(xboot) - mean(yboot) - diffmean)/sqrt((var(xboot)/m + (var(yboot)/n)))
    if(t_ratio[b]>D){counts = counts + 1}
  }
  talphaboot = as.numeric(quantile(t_ratio, conf))
  pval = (counts + 1)/10000
  cat("The one-sided confidence interval is:", (diffmean - talphaboot*sehat), "to infinity, and the p-va
}
setwd("~/Desktop/math185_files")
dat = read.table("memory.txt", header = TRUE, sep = "\t", fill = FALSE)
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =
## dec, : number of items read is not a multiple of the number of columns
```

```
gingko = dat$Gingko
placebo = as.numeric(na.omit(dat$Placebo))
boot.t.test(gingko, placebo, conf = 0.95, B = 9999)
```

```
## The one-sided confidence interval is: -2.076033 to infinity, and the p-value is: 0.9418
```

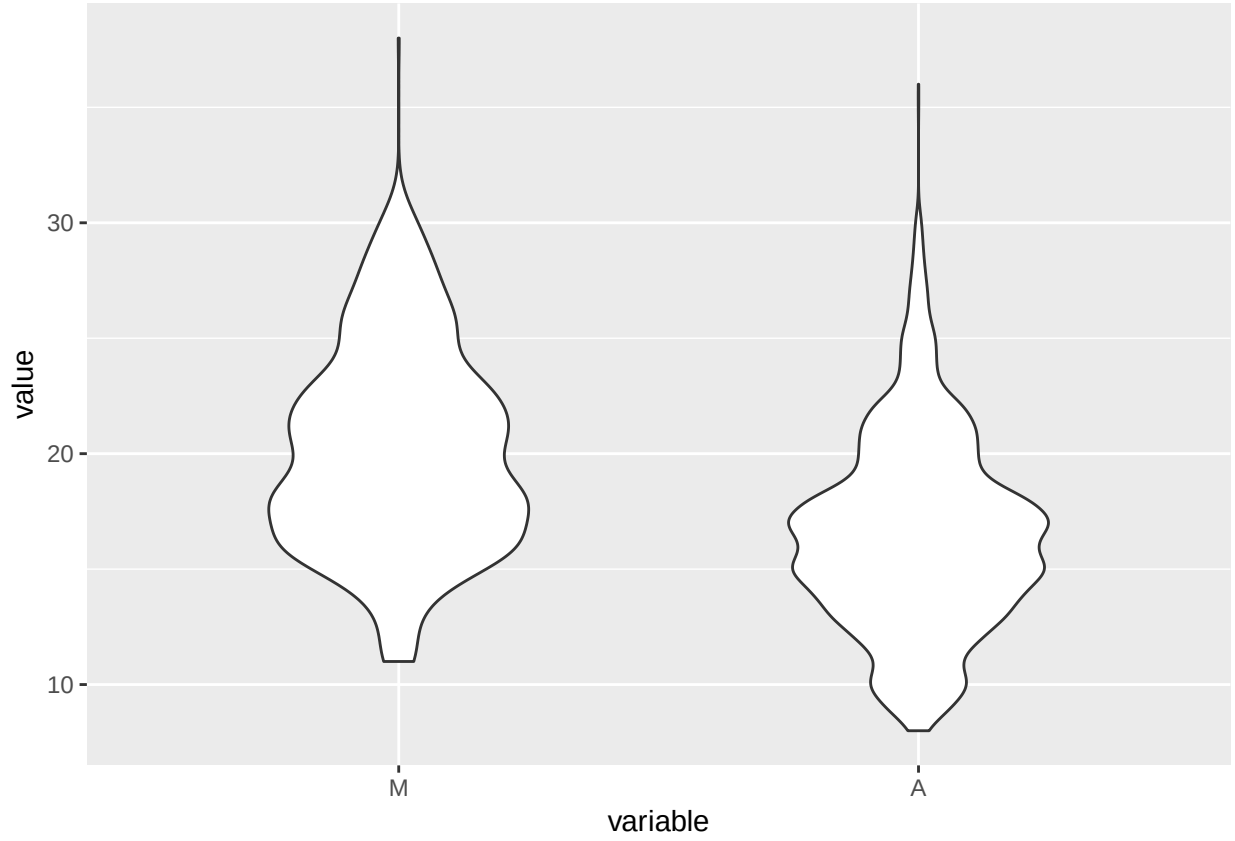
Problem 2A

```
library(ggplot2)
setwd("~/Desktop/math185_files")
car = read.csv("cars.csv", header = TRUE, sep = ",", fill = FALSE)

manual = car[which(car$Classification == "Manual transmission"),]
automatic = car[which(car$Classification == "Automatic transmission"),]

manualcitympg = manual$City.mpg
autocitympg = automatic$City.mpg
```

```
c1 = data.frame(value = manualcitympg, variable = "M")
c2 = data.frame(value = autocitympg,variable = "A")
dat = rbind(c1,c2)
head(dat)
```

```
##   value variable
## 1    21        M
## 2    16        M
## 3    22        M
## 4    18        M
## 5    18        M
## 6    18        M
```

```
ggplot(dat, aes(x = variable, y = value)) + geom_violin(scale = "width", adjust = 1,width = 0.5)
```



H0: True difference in means is equal to 0. H1: True difference in means is not equal to 0. Since p-value is small, we reject our null hypothesis.

```
t.test(manualcitympg,autocitympg)
```

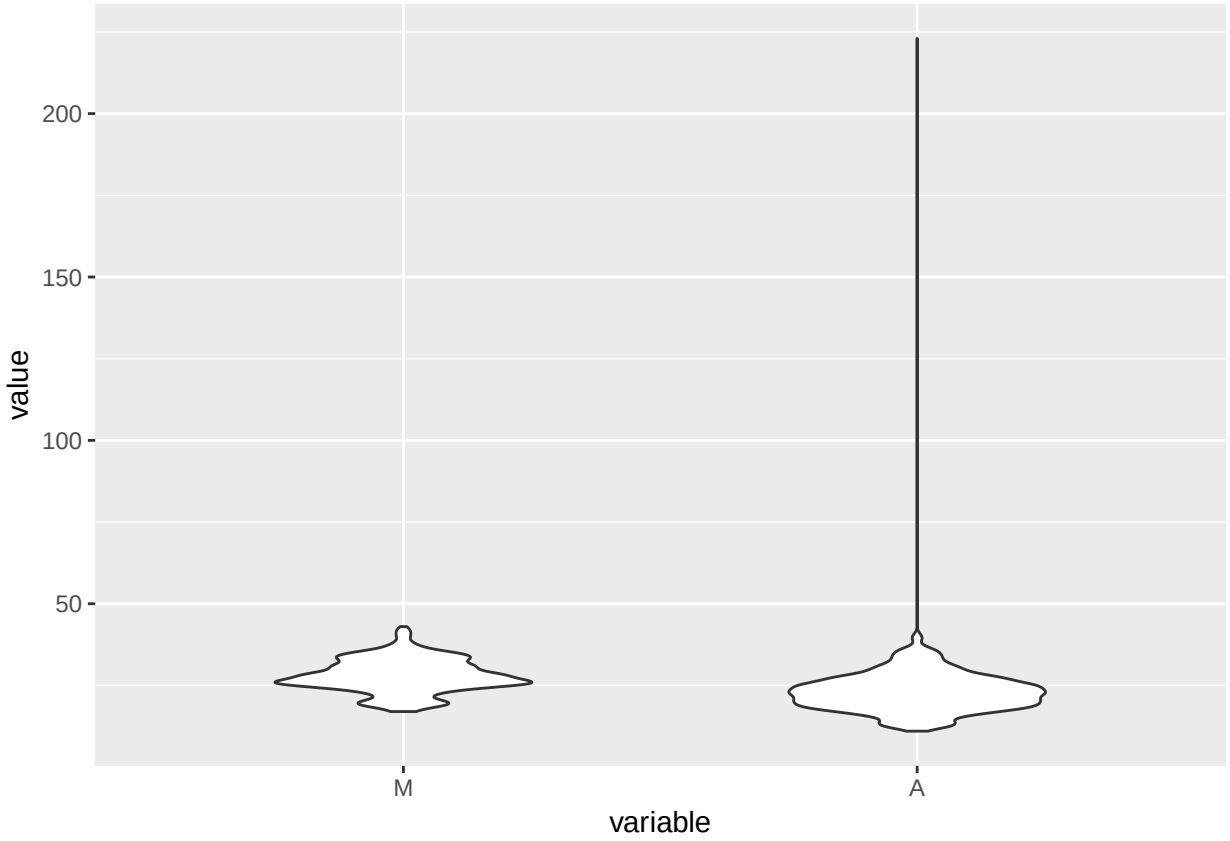
```
##
##  Welch Two Sample t-test
##
## data:  manualcitympg and autocitympg
## t = 25.234, df = 1732.9, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.423028 3.999982
## sample estimates:
## mean of x mean of y
##  20.16327  16.45176
```

```
manualhighwaympg = manual$Highway.mpg
autohighwaympg = automatic$Highway.mpg
c1 = data.frame(value = manualhighwaympg, variable = "M")
c2 = data.frame(value = autohighwaympg, variable = "A")
dat = rbind(c1,c2)
head(dat)
```

```
##   value variable
## 1    30        M
## 2    27        M
## 3    30        M
## 4    25        M
## 5    25        M
## 6    25        M
```

```
ggplot(dat, aes(x = variable, y = value)) + geom_violin(scale = "width", adjust = 1,width = 0.5)
```



H0: The true difference in means is equal to 0. H1:The true difference in means is not equal to 0. Given the p-value is small, we reject our null hypothesis.

```r
t.test(manualhighwaympg,autohighwaympg)
```

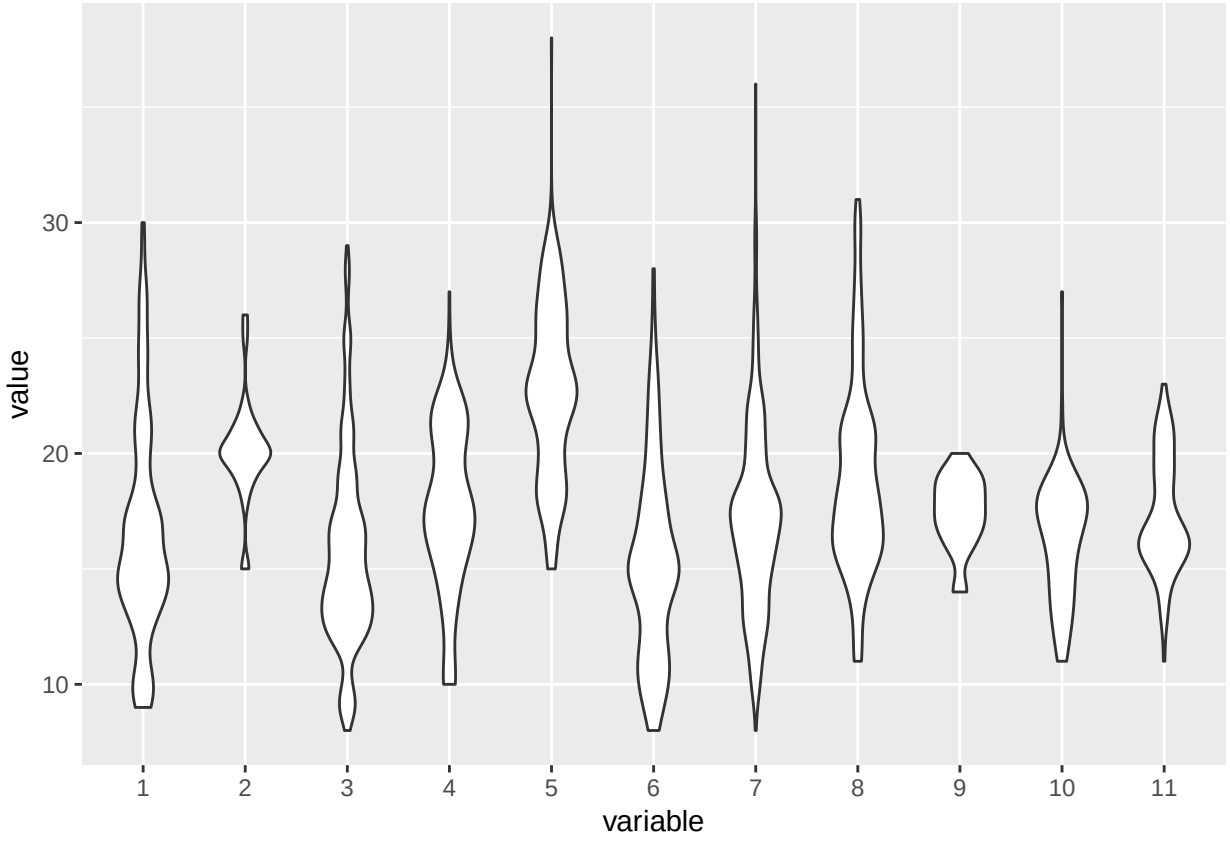
```
##
##  Welch Two Sample t-test
##
## data:  manualhighwaympg and autohighwaympg
## t = 26.676, df = 2242.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.533825 5.253309
## sample estimates:
## mean of x mean of y
##  27.93256  23.03900
```

Problem 2B

For City.mpg

```r
lev = levels(car$Transmission)

lev1 = car[which(car$Transmission == lev[1]),]
lev2 = car[which(car$Transmission == lev[2]),]
lev3 = car[which(car$Transmission == lev[3]),]
lev4 = car[which(car$Transmission == lev[4]),]
lev5 = car[which(car$Transmission == lev[5]),]
lev6 = car[which(car$Transmission == lev[6]),]
lev7 = car[which(car$Transmission == lev[7]),]
lev8 = car[which(car$Transmission == lev[8]),]
lev9 = car[which(car$Transmission == lev[9]),]
lev10 = car[which(car$Transmission == lev[10]),]
lev11 = car[which(car$Transmission == lev[11]),]

lev1citympg = lev1$City.mpg
lev2citympg = lev2$City.mpg
lev3citympg = lev3$City.mpg
lev4citympg = lev4$City.mpg
lev5citympg = lev5$City.mpg
lev6citympg = lev6$City.mpg
lev7citympg = lev7$City.mpg
lev8citympg = lev8$City.mpg
lev9citympg = lev9$City.mpg
lev10citympg = lev10$City.mpg
lev11citympg = lev11$City.mpg

c1 = data.frame(value = lev1citympg, variable = "1")
c2 = data.frame(value = lev2citympg, variable = "2")
c3 = data.frame(value = lev3citympg, variable = "3")
c4 = data.frame(value = lev4citympg, variable = "4")
c5 = data.frame(value = lev5citympg, variable = "5")
c6 = data.frame(value = lev6citympg, variable = "6")
c7 = data.frame(value = lev7citympg, variable = "7")
c8 = data.frame(value = lev8citympg, variable = "8")
c9 = data.frame(value = lev9citympg, variable = "9")
```

```
c10 = data.frame(value = lev10citympg, variable = "10")
c11 = data.frame(value = lev11citympg, variable = "11")

dat = rbind(c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,c11)
ggplot(dat, aes(x = variable, y = value)) + geom_violin(scale = "width", adjust = 1,width = 0.5)
```




```
oneway.test(value~variable, data = dat, var.equal = F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and variable
## F = 137.05, num df = 10.00, denom df = 632.63, p-value < 2.2e-16
```

For Highway.mpg

```
lev1highwaympg = lev1$Highway.mpg
lev2highwaympg = lev2$Highway.mpg
lev3highwaympg = lev3$Highway.mpg
lev4highwaympg = lev4$Highway.mpg
lev5highwaympg = lev5$Highway.mpg
lev6highwaympg = lev6$Highway.mpg
lev7highwaympg = lev7$Highway.mpg
lev8highwaympg = lev8$Highway.mpg
```

```
lev9highwaympg = lev9$Highway.mpg
lev10highwaympg = lev10$Highway.mpg
lev11highwaympg = lev11$Highway.mpg

c1 = data.frame(value = lev1highwaympg, variable = "1")
c2 = data.frame(value = lev2highwaympg, variable = "2")
c3 = data.frame(value = lev3highwaympg, variable = "3")
c4 = data.frame(value = lev4highwaympg, variable = "4")
c5 = data.frame(value = lev5highwaympg, variable = "5")
c6 = data.frame(value = lev6highwaympg, variable = "6")
c7 = data.frame(value = lev7highwaympg, variable = "7")
c8 = data.frame(value = lev8highwaympg, variable = "8")
c9 = data.frame(value = lev9highwaympg, variable = "9")
c10 = data.frame(value = lev10highwaympg, variable = "10")
c11 = data.frame(value = lev11highwaympg, variable = "11")

dat = rbind(c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,c11)
ggplot(dat, aes(x = variable, y = value)) + geom_violin(scale = "width", adjust = 1,width = 0.5)
```


```
oneway.test(value~variable, data = dat, var.equal = F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and variable
```

```
## F = 108.06, num df = 10.00, denom df = 628.61, p-value < 2.2e-16
```