Undergrad Biostatistics - R Training - Week V

Ege Ulgen

One-sample t Test

We will use the AIDS data again (this time reading from a URL:

```
URL <- "https://raw.githubusercontent.com/egeulgen/MED131_22_23/main/data/aids_dataset.txt"
aids_df <- read.delim(URL, sep = " ")
```

Let's explore the data:

```
# print first 6 rows (default)
head(aids_df)
```

```
##   id treatment   age gender week_1 cd4_1 week_2 cd4_2
## 1  1      trt2 36.43   male      0    23   7.57    21
## 2  2      trt4 47.85   male      0    21   8.00    49
## 3  4      trt3 36.60   male      0    61   7.14    61
## 4  5      trt1 35.95   male      0    36   8.00    31
## 5  6      trt2 38.40   male      0    11   7.29    11
## 6  7      trt2 45.08   male      0    11   9.00    41
```

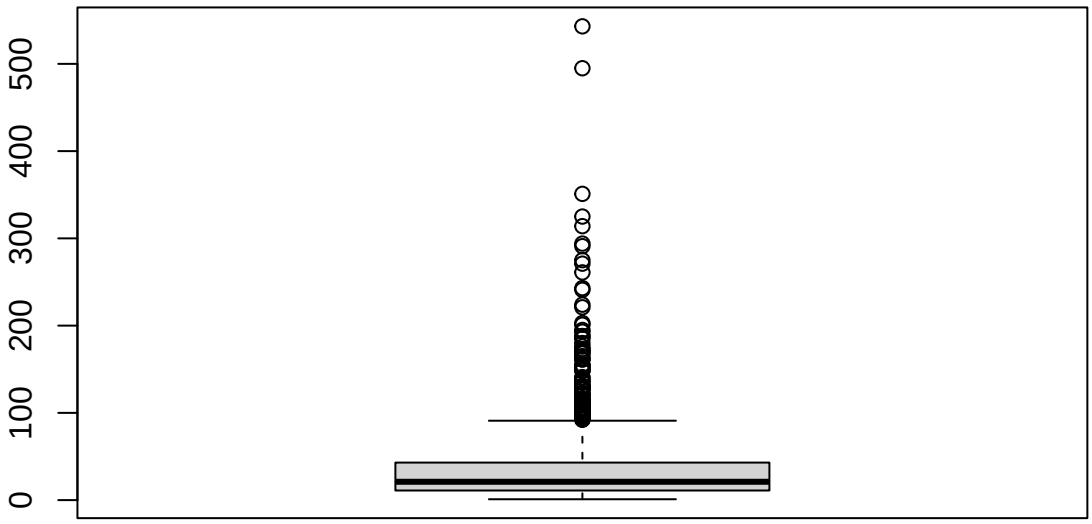
```
# how many patients are there in each treatment group?
table(aids_df$treatment)
```

```
##
## trt1 trt2 trt3 trt4
##  289  288  293  308
```

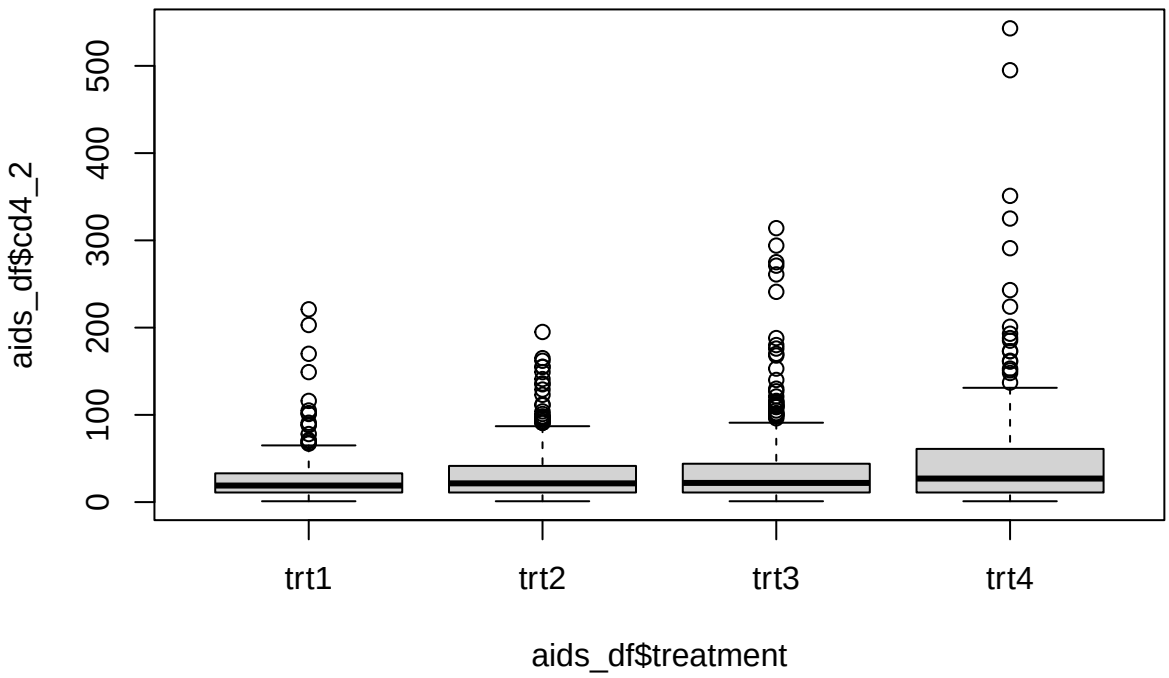
```
# summary of all variables
summary(aids_df)
```

```
##        id        treatment              age          gender
##  Min.   :   1   Length:1178        Min.   :14.9   Length:1178
##  1st Qu.: 331   Class :character   1st Qu.:31.8   Class :character
##  Median : 650   Mode  :character   Median :36.8   Mode  :character
##  Mean   : 659                      Mean   :37.7
##  3rd Qu.: 993                      3rd Qu.:42.5
##  Max.   :1313                      Max.   :74.2
##      week_1      cd4_1           week_2          cd4_2
##  Min.   :0   Min.   :  1.0   Min.   : 2.14   Min.   :  1.0
##  1st Qu.:0   1st Qu.: 11.0   1st Qu.: 7.86   1st Qu.: 11.0
##  Median :0   Median : 21.0   Median : 8.14   Median : 21.0
##  Mean   :0   Mean   : 26.5   Mean   :10.12   Mean   : 36.7
##  3rd Qu.:0   3rd Qu.: 36.0   3rd Qu.:10.54   3rd Qu.: 43.0
##  Max.   :0   Max.   :181.0   Max.   :38.00   Max.   :543.0
```

```
boxplot(aids_df$cd4_2)
```




```
boxplot(aids_df$cd4_2~aids_df$treatment)
```

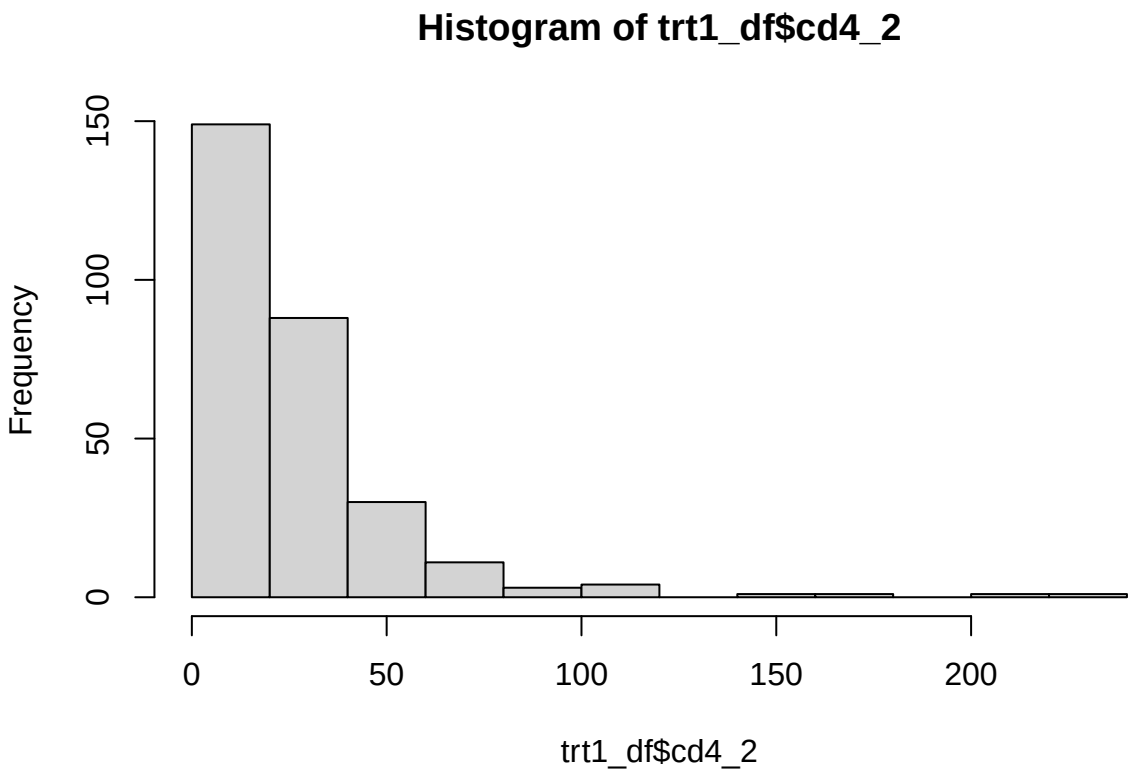


We'll use the "trt1" subset and perform hypothesis tests for CD4 at second measurement:

```
# subset for trt1
trt1_df <- subset(aids_df, treatment == "trt1")

summary(trt1_df$cd4_2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0    11.0    19.0    26.3    33.0   221.0
```

```
hist(trt1_df$cd4_2)
```




```
?t.test

# is the CD4 level at week 2 for trt1 different than 0?
t.test(trt1_df$cd4_2)
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = 16.2, df = 288, p-value <2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  23.115 29.508
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 different than 20?
t.test(trt1_df$cd4_2, mu = 20)
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = 3.89, df = 288, p-value = 0.00013
## alternative hypothesis: true mean is not equal to 20
## 95 percent confidence interval:
```

```
##  23.115 29.508
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 different than 25?
t.test(trt1_df$cd4_2, mu = 25)
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = 0.808, df = 288, p-value = 0.42
## alternative hypothesis: true mean is not equal to 25
## 95 percent confidence interval:
##  23.115 29.508
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 different than 20? (alpha = 0.1, conf. level = 90%)
t.test(trt1_df$cd4_2, mu = 20, conf.level = 0.9)
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = 3.89, df = 288, p-value = 0.00013
## alternative hypothesis: true mean is not equal to 20
## 90 percent confidence interval:
##  23.632 28.991
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 different than 20? (alpha = 0.01, conf. level = 99%)
t.test(trt1_df$cd4_2, mu = 20, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = 3.89, df = 288, p-value = 0.00013
## alternative hypothesis: true mean is not equal to 20
## 99 percent confidence interval:
##  22.100 30.522
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 larger than 25?
t.test(trt1_df$cd4_2, mu = 25, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
```

```
## t = 0.808, df = 288, p-value = 0.21
## alternative hypothesis: true mean is greater than 25
## 95 percent confidence interval:
##  23.632    Inf
## sample estimates:
## mean of x
##    26.311
```

```
# is the CD4 level at week 2 for trt1 smaller than 60?
t.test(trt1_df$cd4_2, mu = 60, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  trt1_df$cd4_2
## t = -20.7, df = 288, p-value <2e-16
## alternative hypothesis: true mean is less than 60
## 95 percent confidence interval:
##    -Inf 28.991
## sample estimates:
## mean of x
##    26.311
```

Two-sample t Test

We'll use the `genderweight` data from the `datarium` package for this exercise. First, let's install the package (you'll need to this only once):

```
# install.packages("datarium")
```

Let's load and explore the data:

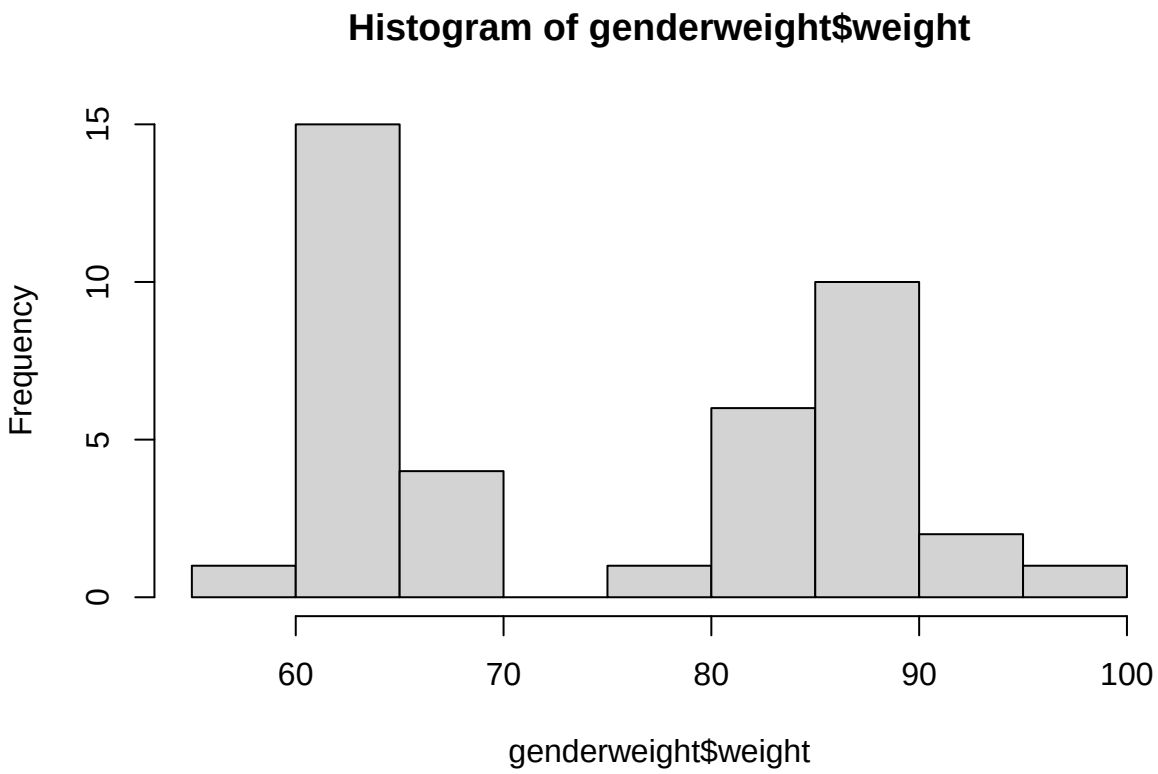
```
data("genderweight", package = "datarium")

head(genderweight)
```

```
##   id group weight
## 1  1     F 61.586
## 2  2     F 64.555
## 3  3     F 66.169
## 4  4     F 59.309
## 5  5     F 64.858
## 6  6     F 65.012
```

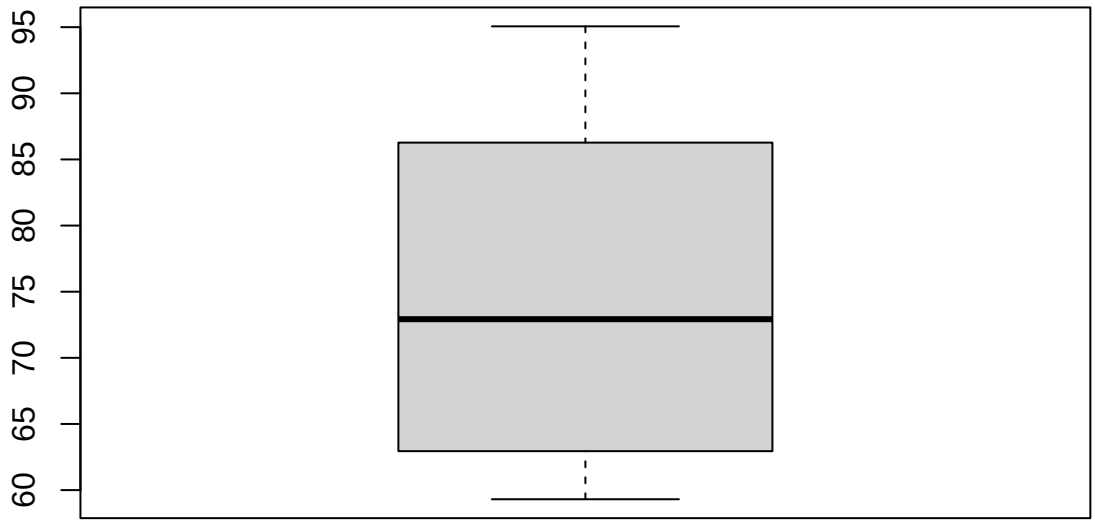
```
# frequency table of gender (group)
table(genderweight$group)
```

```
##
##  F  M
## 20 20
```

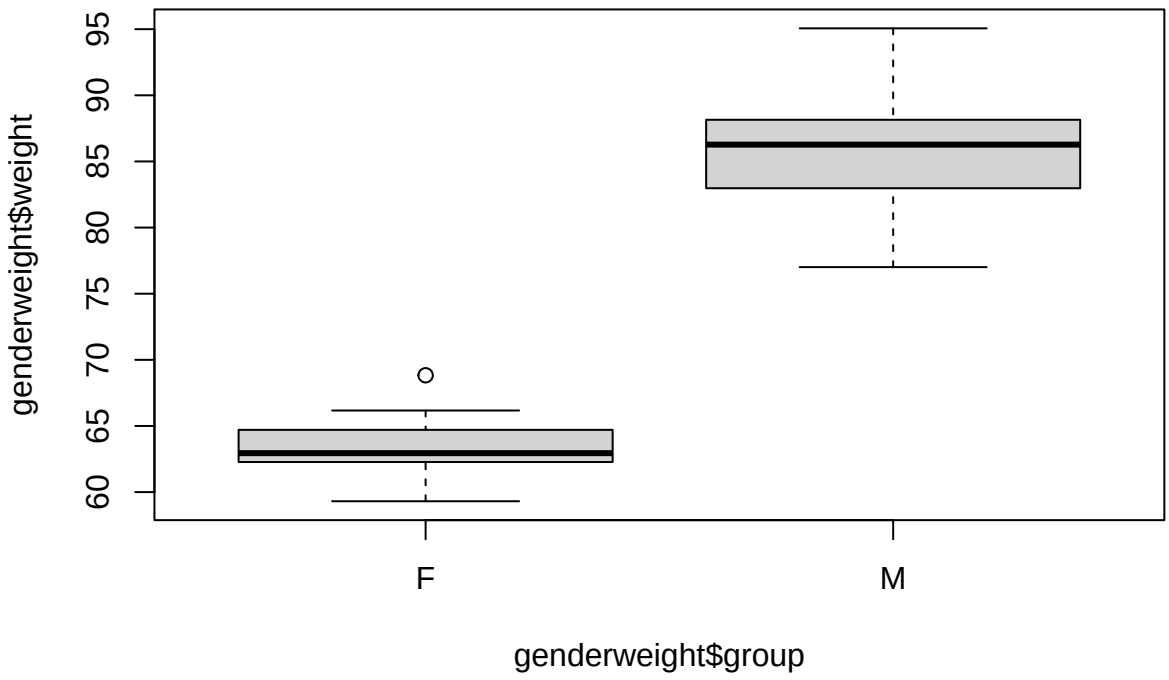
```
# histogram of weight (overall)
hist(genderweight$weight)
```




```
# boxplot of weight (overall)
boxplot(genderweight$weight)
```




```
# boxplot of weight by gender
boxplot(genderweight$weight~genderweight$group)
```




```
# mean weight of men
mean(genderweight$weight[genderweight$group == "M"])
```

```
## [1] 85.826
```

```
# mean weight of women
mean(genderweight$weight[genderweight$group == "F"])
```

```
## [1] 63.499
```

```
# sd of weight of men
sd(genderweight$weight[genderweight$group == "M"])
```

```
## [1] 4.3535
```

```
# sd of weight of women
sd(genderweight$weight[genderweight$group == "F"])
```

```
## [1] 2.0276
```

Is the mean weight of men significantly different than the mean weight of women?

```
t.test(weight~group, data = genderweight)
```

```
##
##  Welch Two Sample t-test
##
## data:  weight by group
## t = -20.8, df = 26.9, p-value <2e-16
## alternative hypothesis: true difference in means between group F and group M is not equal to 0
```

```
## 95 percent confidence interval:
##  -24.531 -20.124
## sample estimates:
## mean in group F mean in group M
##          63.499          85.826
```