Module 1 Lab Submission

Ben Tankus

In this made-up example, we have three different nutritional supplements (A, B, and C are the group labels), and we are interested in how these supplements affect sleep duration. We randomly assign 15, 20, and 30 people to the supplements, respectively (so 15 to group A, 20 to group B, and 30 to group C). We record the number of hours that each person sleeps during one night after a week of taking their assigned supplement.

In the code below, we simulate data for this scenario. There are many ways to do this, and the following code is written for readability rather than coding aesthetics/speed or runtime efficiency. Note that in our simulation, the *average* number of hours of sleep per night is the same for all three groups (that is, the null hypothesis tested by one-way ANOVA is true). Also note that the *standard deviations* (and therefore *variances*) of the three populations differ!

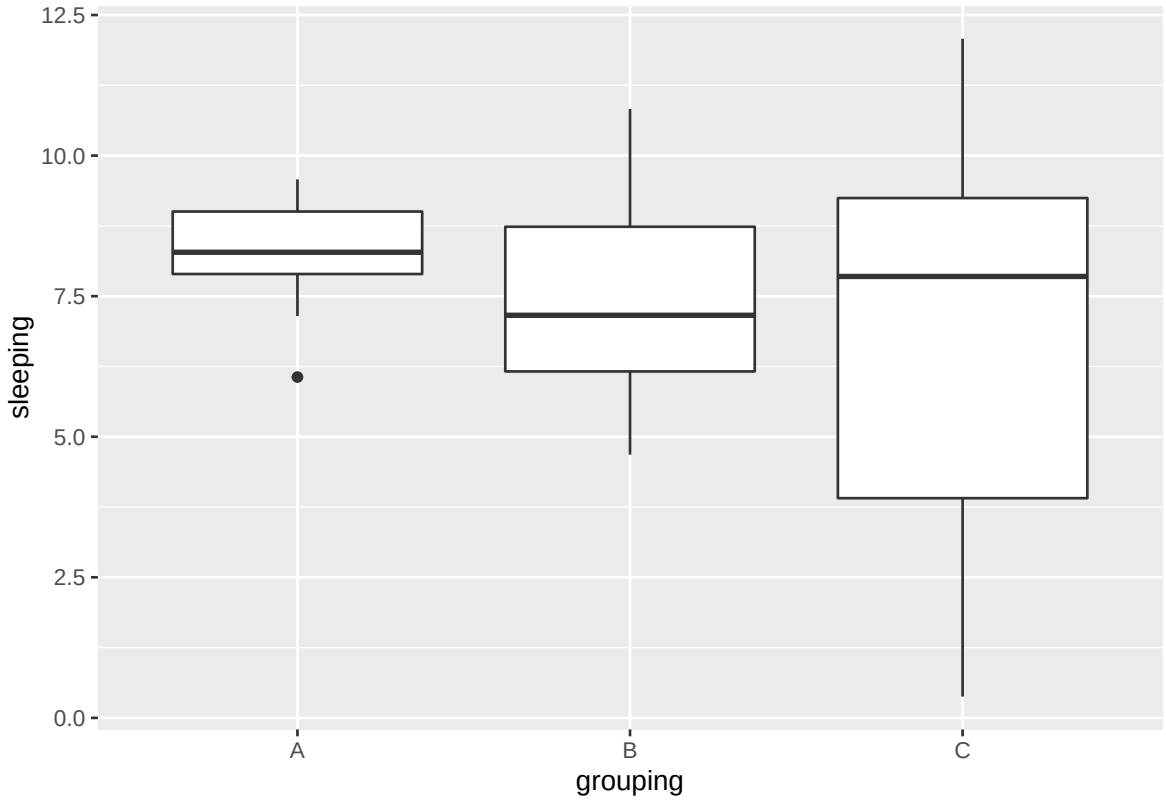
```
set.seed(2881)

group_sizes <- c(15, 20, 30)
group_means <- c(8, 8, 8)
group_sds <- c(1, 2, 3)

group1 <- rnorm(group_sizes[1], group_means[1], group_sds[1])
group2 <- rnorm(group_sizes[2], group_means[2], group_sds[2])
group3 <- rnorm(group_sizes[3], group_means[3], group_sds[3])

sleepData <- data.frame(sleeping=c(group1, group2, group3),
                        grouping=rep(c('A', 'B', 'C'), group_sizes))

head(sleepData)
```

```
##   sleeping grouping
## 1 8.281356        A
## 2 7.993048        A
## 3 8.672573        A
## 4 7.805599        A
## 5 8.903858        A
## 6 8.251239        A
```

Write code in the chunks below to answer the following questions:

- **Make boxplots of the sleep durations by group**

```
qplot(x = grouping, y = sleeping, data = sleepData, geom = "boxplot")
```



- **What is the sample mean for each of the three groups A, B, and C?**

```
print("Sample Means:")
```

```
## [1] "Sample Means:"
```

```
tapply( sleepData$sleeping, sleepData$grouping, mean )
```

```
##        A        B        C
## 8.294607 7.370158 6.839735
```

- **Perform a one-way ANOVA to assess whether the population means differ across treatments A, B, and C, assuming all of the population variances are equal.**

```
oneway.test(sleeping~grouping, sleepData, var.equal = T)
```

```
##
##  One-way analysis of means
##
## data:  sleeping and grouping
## F = 1.9301, num df = 2, denom df = 62, p-value = 0.1538
```

We can say with a decent level of certainty that our groups have the same means. Fail to reject null hypothesis of X_bar1 = X_Bar2 = X_bar3

- **Perform a one-way ANOVA to assess whether the population means differ across treatments A, B, and C, *without* assuming all of the population variances are equal.**

```
oneway.test(sleeping~grouping, sleepData, var.equal = F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  sleeping and grouping
## F = 4.0138, num df = 2.000, denom df = 40.329, p-value = 0.02572
```

We can say with a decent level of certainty that our groups have DIFFERENT means. Reject null hypothesis of X_bar1 = X_Bar2 = X_bar3

- **Perform a Kruskal-Wallace Test to compare the sleep durations for treatments A, B, and C.**

```
kruskal.test(sleeping~grouping, sleepData)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sleeping by grouping
## Kruskal-Wallis chi-squared = 1.6593, df = 2, p-value = 0.4362
```

The kruskal-wallace test does not provide sufficient evidence to reject the null hypothesis that mean ranks are different between A, B, and C.