Lab03 - F tests for ANOVA

Goals

The goal in this lab is to practice F tests for ANOVA

Loading packages

Here are some packages with functionality you may need for this lab. Run this code chunk now.

```
library(readr)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
library(gridExtra)
library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 3.5.2

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.5.2

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggformula

## Warning: package 'ggformula' was built under R version 3.5.2

## Loading required package: ggstance

##
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh

##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData
```

```
## Loading required package: Matrix

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
##
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':
##
##     mean

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following object is masked from 'package:ggplot2':
##
##     stat

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median,
##     prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```r
library(dplyr)

options("pillar.sigfig" = 10) # print 10 significant digits in summarize output
```

A study was conducted to examine the effectiveness of four different hand-washing methods for eliminating bacteria

```r
soap <- read_csv("http://www.evanlray.com/data/sdm4/Bacterial_Soap.csv")
```

```
## Parsed with column specification:
## cols(
##   `Bacterial Counts` = col_double(),
##   Method = col_character()
## )
```

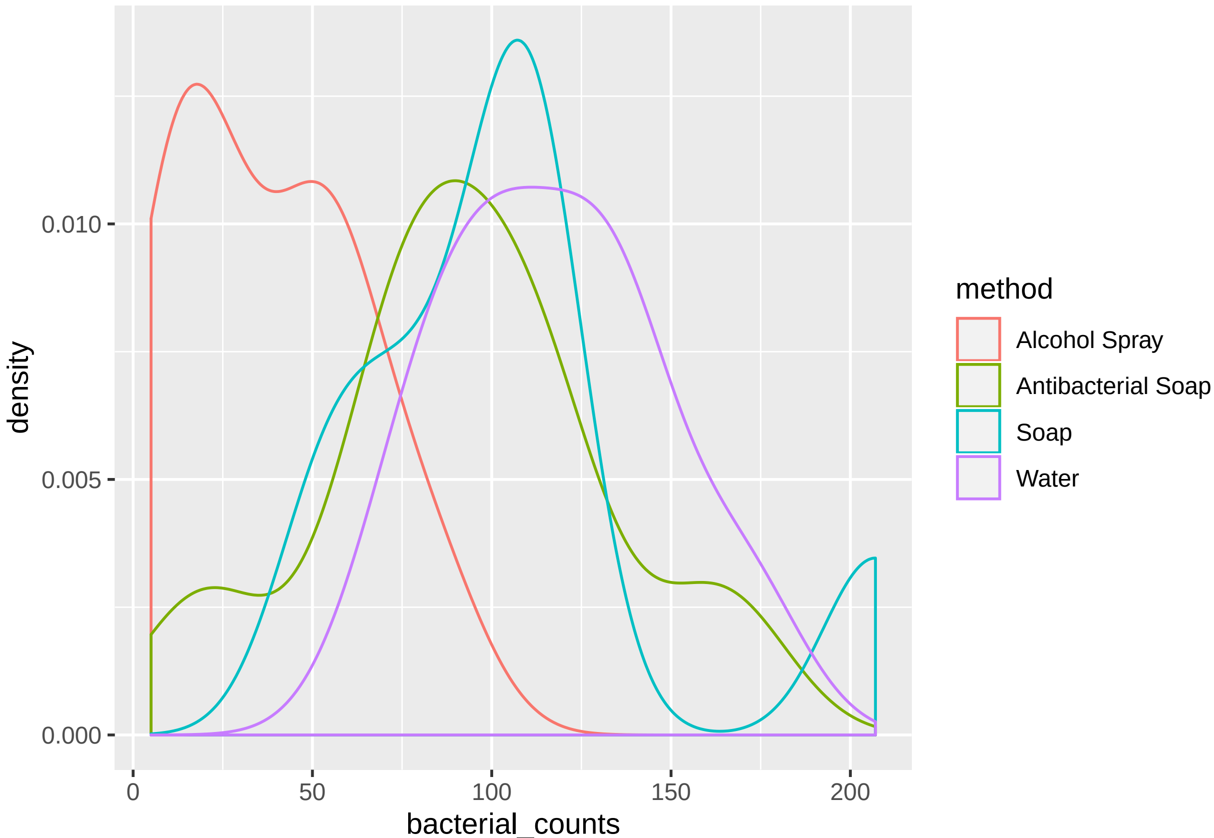
```r
names(soap) <- c("bacterial_counts", "method")

soap %>%
  count(method)
```

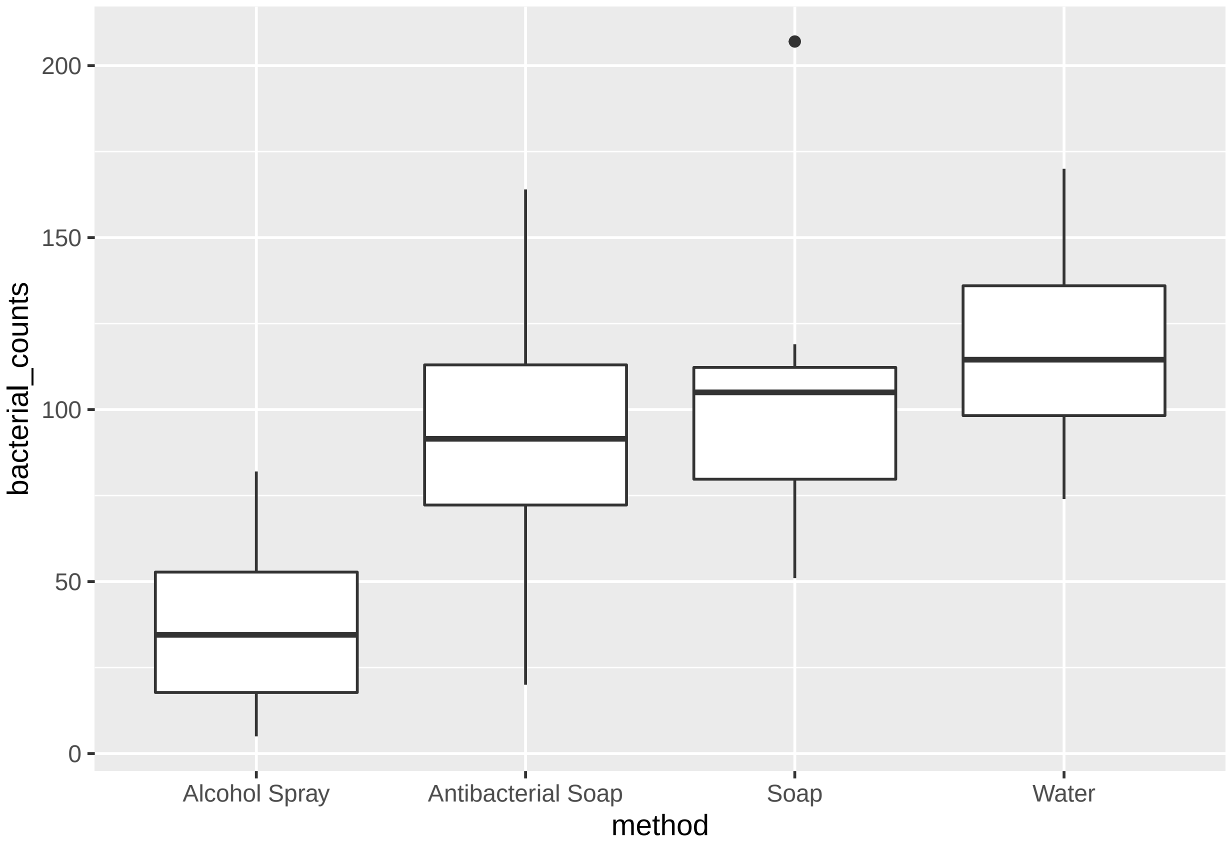
```
## # A tibble: 4 x 2
##   method                 n
##   <chr>              <int>
## 1 Alcohol Spray          8
## 2 Antibacterial Soap     8
## 3 Soap                   8
## 4 Water                  8
```

1. Make an appropriate plot of the data.

```r
ggplot(data = soap, mapping = aes(x = bacterial_counts, color = method)) +
  geom_density()
```




```r
ggplot(data = soap, mapping = aes(x = method, y = bacterial_counts)) +
  geom_boxplot()
```



2. Conduct a test of the claim that all four methods are equally effective.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ H_A : at least one of μ_1, μ_2, μ_3, and μ_4 is not equal to the others

```r
model_fit <- lm(bacterial_counts ~ method, data = soap)
summary(model_fit)
```

```
##
## Call:
## lm(formula = bacterial_counts ~ method, data = soap)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -72.50 -20.88  -1.00  18.12 101.00
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 37.50      13.28   2.825 0.008629 **
## methodAntibacterial Soap    55.00      18.78   2.929 0.006686 **
## methodSoap                  68.50      18.78   3.648 0.001070 **
## methodWater                 79.50      18.78   4.234 0.000224 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.55 on 28 degrees of freedom
## Multiple R-squared:  0.4308, Adjusted R-squared:  0.3698
## F-statistic: 7.064 on 3 and 28 DF,  p-value: 0.001111
```

The p-value for the test is 0.001111. There is very strong evidence against the null hypothesis that all four methods are equally effective.

3. We will do this part on Monday: Conduct a test of the claim that the "Antibacterial Soap", "Soap", and "Water" methods are equally effective.

$H_0 : \mu_2 = \mu_3 = \mu_4$ H_A : at least one of μ_2, μ_3, and μ_4 is not equal to the others

```
soap <- soap %>%
  mutate(
    grouped_method = ifelse(method %in% c("Antibacterial Soap", "Soap", "Water"),
      "grouped", "Alcohol Spray")
  )

reduced_model_fit <- lm(bacterial_counts ~ grouped_method, data = soap)
anova(reduced_model_fit, model_fit)
```

```
## Analysis of Variance Table
##
## Model 1: bacterial_counts ~ grouped_method
## Model 2: bacterial_counts ~ method
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     30 41893
## 2     28 39484  2    2409.3 0.8543 0.4364
```

The p-value for the test is 0.436. The data do not offer any evidence against the null hypothesis that the "Antibacterial Soap", "Soap", and "Water" methods are equally effective.

4. If you have extra time and want a refresher, calculate the mean and standard deviation of observations in each group

```
soap %>%
  group_by(method) %>%
  summarize(
    mean_counts = mean(bacterial_counts),
    sd_counts = sd(bacterial_counts)
  )
```

```
## # A tibble: 4 x 3
##   method             mean_counts   sd_counts
##   <chr>                    <dbl>       <dbl>
## 1 Alcohol Spray             37.5 26.55990534
## 2 Antibacterial Soap        92.5 41.96256835
## 3 Soap                     106   46.95894864
## 4 Water                    117   31.13105936
```