

Lab04 - 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)
library(gridExtra)
library(mosaic)

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

```
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

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

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

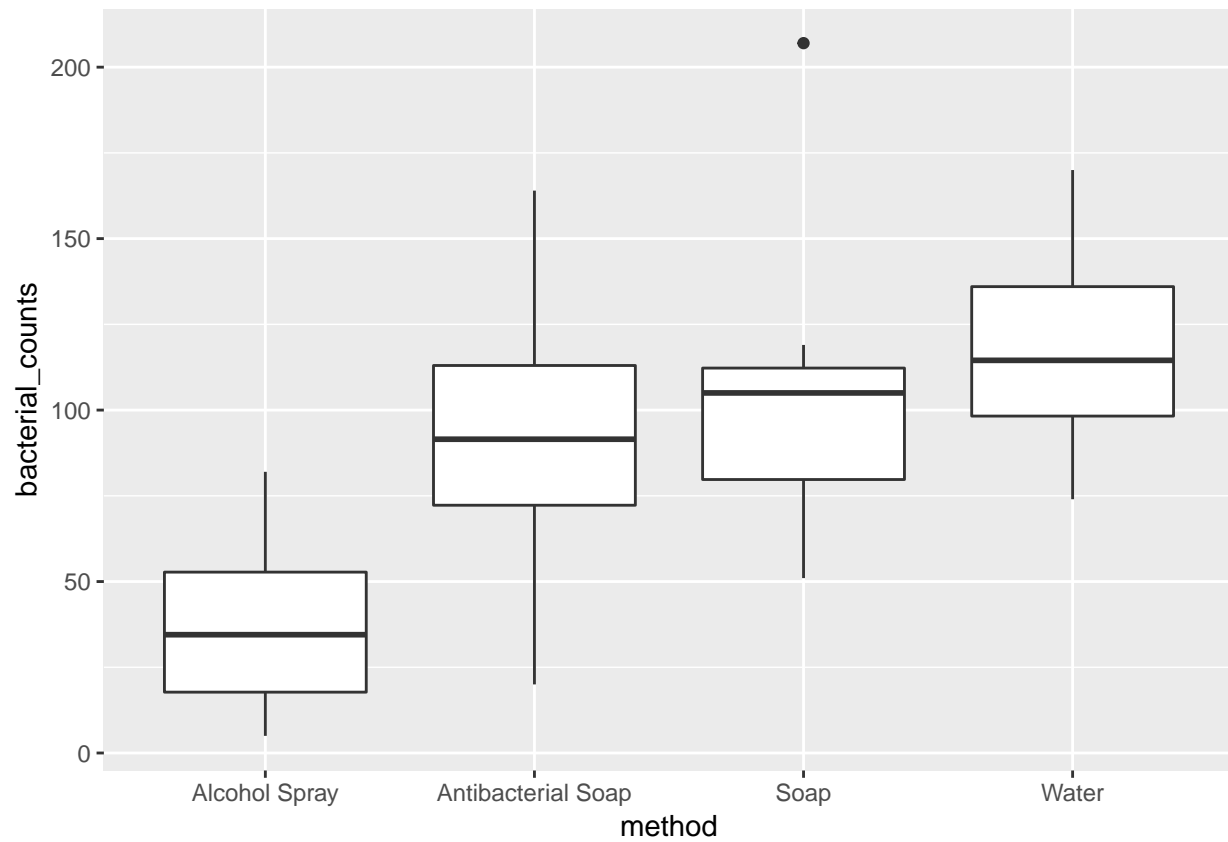
```
names(soap) <- c("bacterial_counts", "method")
```

```
soap %>%
  count(method)
```

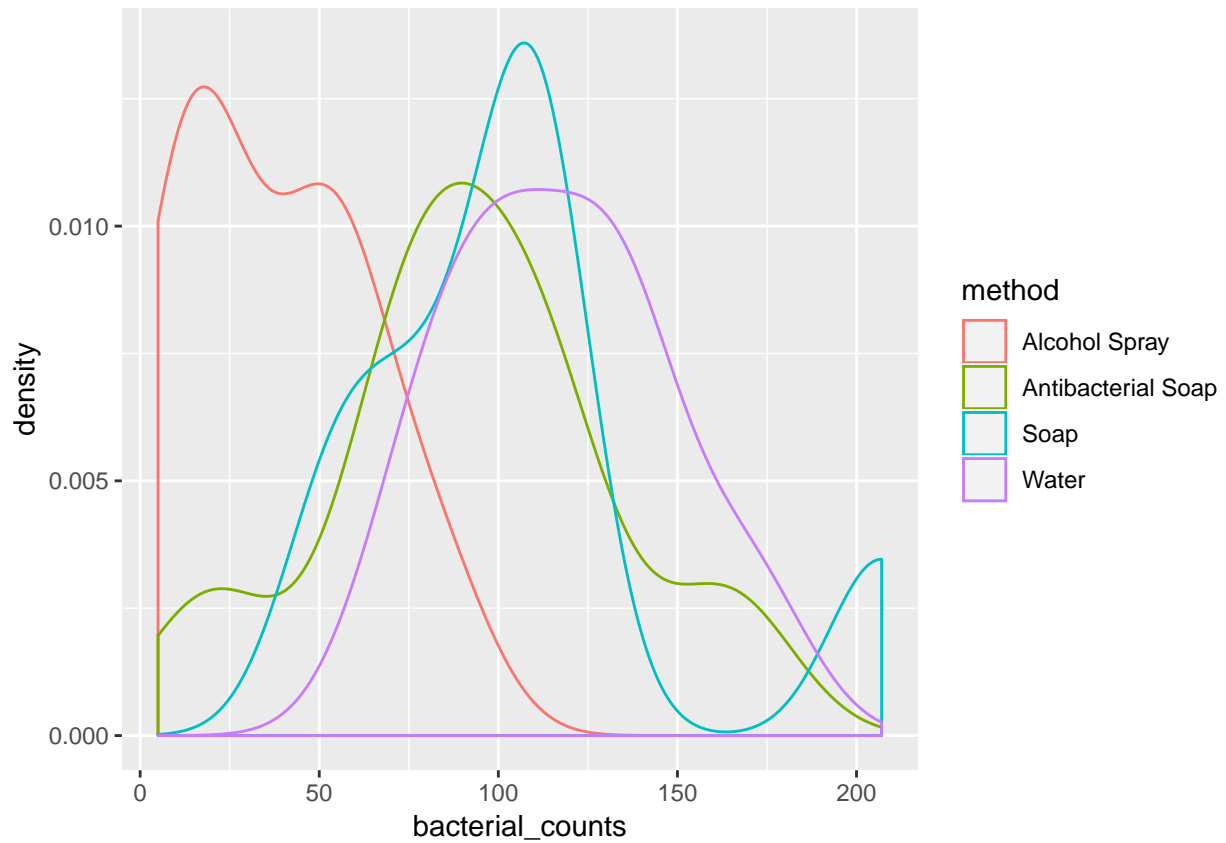
```
## # A tibble: 4 x 2
##   method      n
##   <chr>    <int>
## 1 Alcohol Spray      8
## 2 Antibacterial Soap  8
## 3 Soap              8
## 4 Water              8
```

1. Make an appropriate plot of the data.

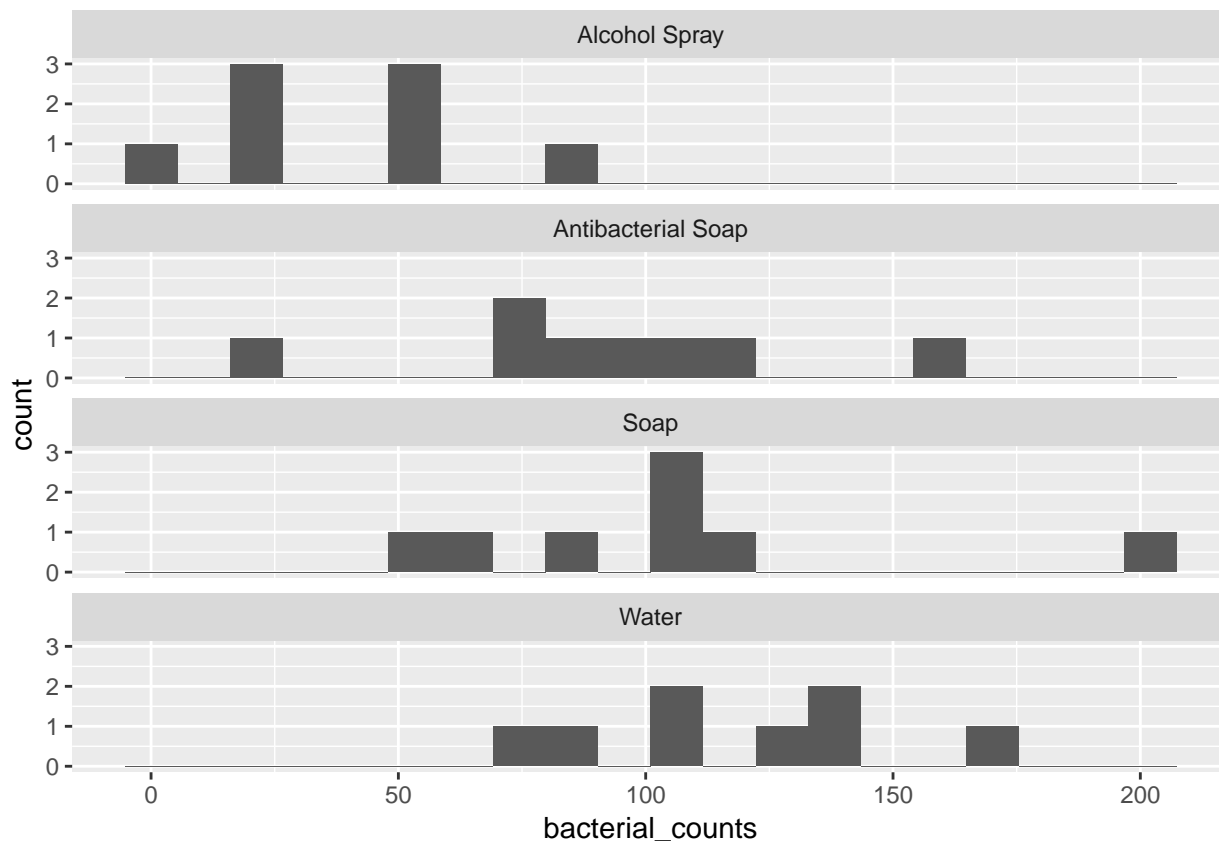
```
ggplot(data = soap, mapping = aes(x = method, y = bacterial_counts)) +  
  geom_boxplot()
```



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



```
ggplot(data = soap, mapping = aes(x = bacterial_counts)) +  
  geom_histogram(bins = 20) +  
  facet_wrap(~ method, ncol = 1)
```



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

Define the parameters

μ_1 = mean bacterial count when using alcohol spray

μ_2 = mean bacterial count when using antibacterial soap

μ_3 = mean bacterial count when using soap

μ_4 = mean bacterial count when using water

Our hypotheses are:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ The average bacterial count is the same for all four handwashing methods.

H_A : Not all of μ_1 , μ_2 , μ_3 , and μ_4 are equal to each other. The average bacterial count is different for at least one of the handwashing methods.

```
fit_full <- lm(bacterial_counts ~ method, data = soap)
anova(fit_full)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: bacterial_counts
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## method      3  29882  9960.7    7.0636 0.001111 **
```

```
## Residuals  28  39484  1410.1
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value for this test is 0.001111. The data provide very strong evidence that the mean bacterial counts are not equal for the four handwashing methods.

3. Conduct a test of the claim that the “Antibacterial Soap”, “Soap”, and “Water” methods are equally effective.

```
soap <- soap %>%
  mutate(
    not_alcohol_spray = ifelse(method %in% c("Antibacterial Soap", "Soap", "Water"), "Not Alcohol", "Alcohol")
  )
fit_reduced <- lm(bacterial_counts ~ not_alcohol_spray, data = soap)
anova(fit_reduced, fit_full)
```

```
## Analysis of Variance Table
##
## Model 1: bacterial_counts ~ not_alcohol_spray
## 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
```

In terms of the parameters defined in the answer to part 2, we are testing the following hypotheses:

$H_0 : \mu_2 = \mu_3 = \mu_4$ The mean bacterial count is the same for each of the antibacterial soap, soap, and water methods.

H_A : At least one of μ_2 , μ_3 , and μ_4 is different from the other two. The mean bacterial count is not the same for all three of these methods.

The p-value for the test is 0.436. The data do not provide strong evidence that there is a difference in mean bacterial counts.

4. (Optional) 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_bacteria = mean(bacterial_counts),
    sd_bacteria = sd(bacterial_counts)
  )
```

```
## # A tibble: 4 x 3
##   method          mean_bacteria sd_bacteria
##   <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
```