

Analysis of Variance (Chapter 24)

Patrick Frenett, Vickie Ip, Sarah McDonald, and Nicholas Horton (nhorton@amherst.edu)

June 29, 2018

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Intro Stats* (2013) by De Veaux, Velleman, and Bock. More information about the book can be found at http://wps.aw.com/aw_deveaux_stats_series. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at <https://nhorton.people.amherst.edu/is4>.

This work leverages initiatives undertaken by Project MOSAIC (<http://www.mosaic-web.org>), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the `mosaic` package vignettes (<http://cran.r-project.org/web/packages/mosaic>). A paper describing the `mosaic` approach was published in the *R Journal*: <https://journal.r-project.org/archive/2017/RJ-2017-024>.

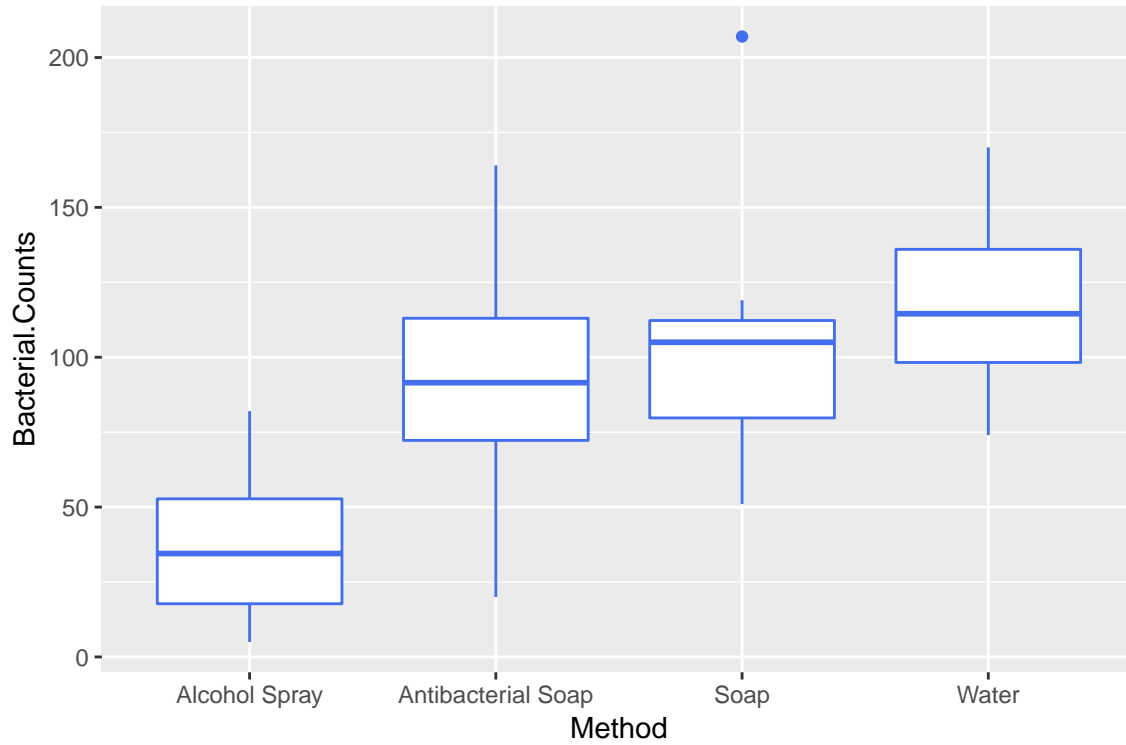
Note that some of the figures in this document may differ slightly from those in the IS4 book due to small differences in datasets. However in all cases the analysis and techniques in R are accurate.

Chapter 24: Analysis of variance

Section 24.1: Testing whether the means of several groups are zero

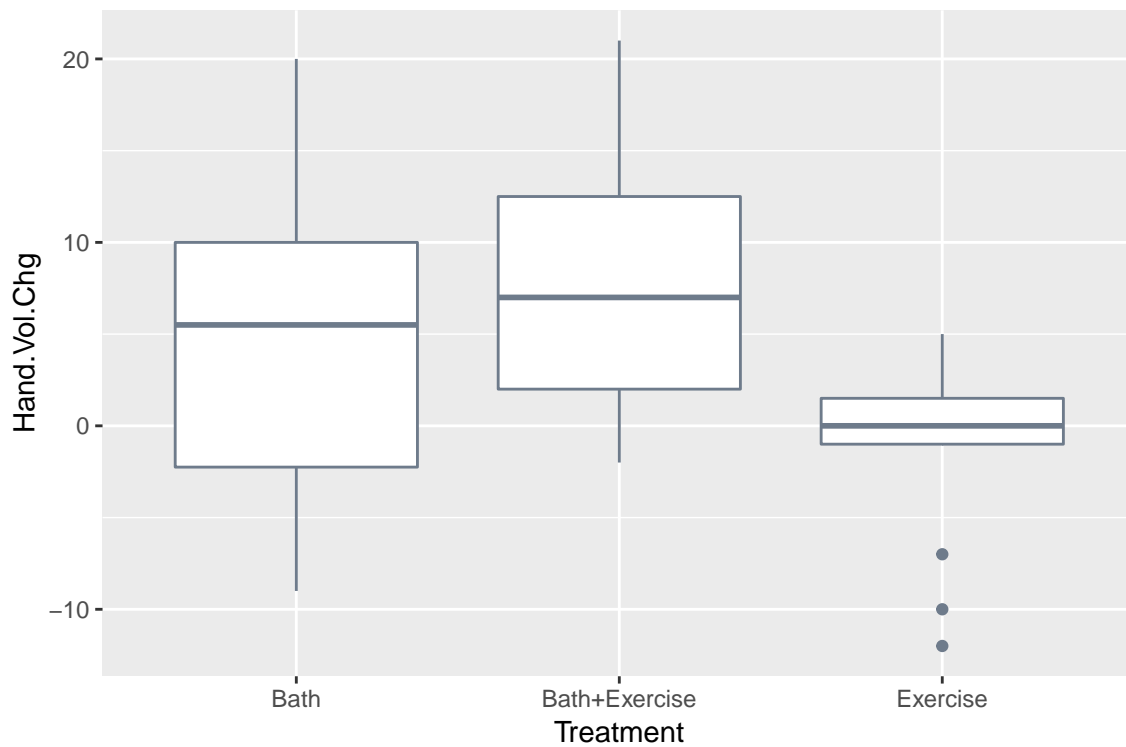
The graph in Figure 24.1 (page 701) can be generated using the `gf_boxplot()` function.

```
Soap <- read.csv("https://nhorton.people.amherst.edu/sdm4/data/Bacterial_Soap.csv")
gf_boxplot(Bacterial.Counts ~ Method, data = Soap, col = "royalblue2")
```



The example on page 704 considers the outcomes in hand volumes for three treatments post surgery.

```
Contrast <- read.csv("https://nhorton.people.amherst.edu/sdm4/data/Contrast_baths.csv")
gf_boxplot(Hand.Vol.Chg ~ Treatment, data = Contrast, col = "lightsteelblue4")
```



The summary statistics at the bottom of page 705 can be calculated using `favstats()` (or `df_stats()`).

```
favstats(Bacterial.Counts ~ Method, data = Soap)
```

```
##           Method min   Q1 median   Q3 max  mean    sd n missing
## 1   Alcohol Spray  5 17.75  34.5  52.75  82  37.5 26.55991 8      0
## 2 Antibacterial Soap 20 72.25  91.5 113.00 164  92.5 41.96257 8      0
## 3           Soap  51 79.75 105.0 112.25 207 106.0 46.95895 8      0
## 4           Water  74 98.25 114.5 136.00 170 117.0 31.13106 8      0
```

Section 24.2: The ANOVA table

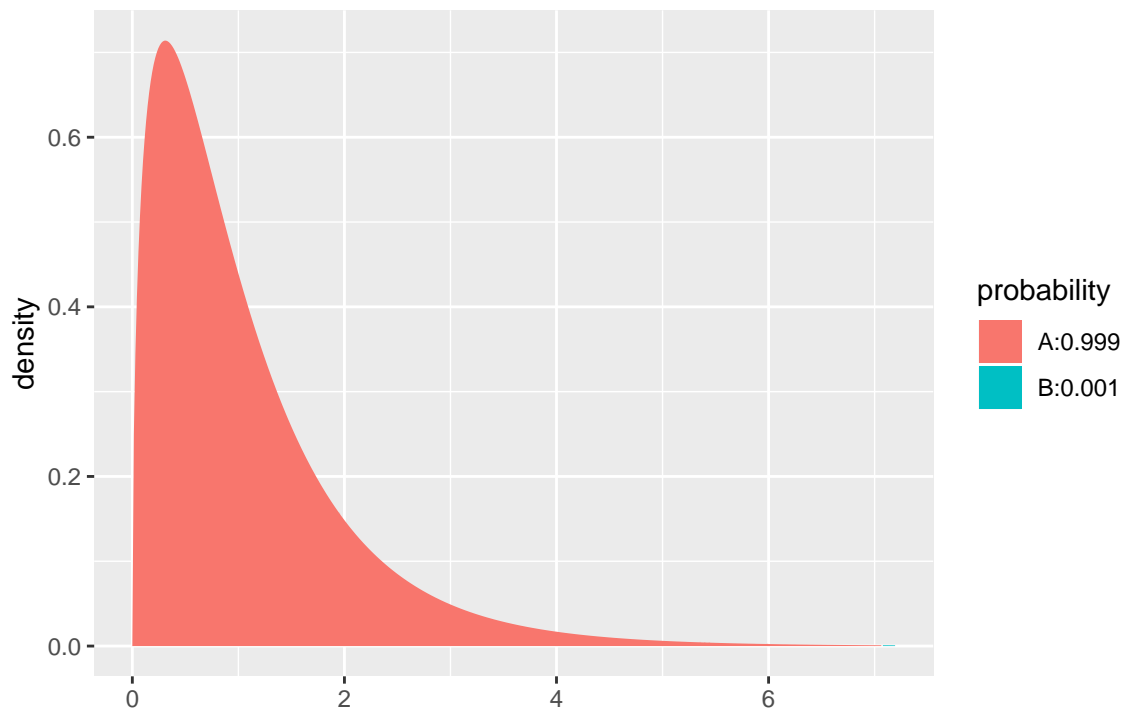
The `aov()` function can be used to fit an analysis of variance model.

```
aovmod <- aov(Bacterial.Counts ~ Method, data = Soap)
summary(aovmod)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Method      3  29882    9961  7.064 0.00111 **
## Residuals  28  39484    1410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This model has 3 degrees of freedom for the model (numerator) and 28 degrees of freedom for the error (denominator). The `xpf()` function can replicate the calculation of the exact p-value (and generate Figure 24.4, page 708).

```
xpf(7.0636, df1 = 3, df2 = 28)
```



```
## [1] 0.9988886
```

The treatment means can be generated using `model.tables()` (see page 711).

```
model.tables(aovmod)
```

```
## Tables of effects
##
## Method
## Method
##      Alcohol Spray Antibacterial Soap          Soap
##      -50.75          4.25          17.75
##      Water
##      28.75
```

The residual standard deviation can be calculated (page 713).

```
n <- 32
k <- 4
sp <- sqrt(sum(resid(aovmod)^2/(n - k)))
sp
```

```
## [1] 37.55187
```

```
sqrt(1410)
```

```
## [1] 37.54997
```

We can also see how the results are equivalent when fitting a regression model with indicators.

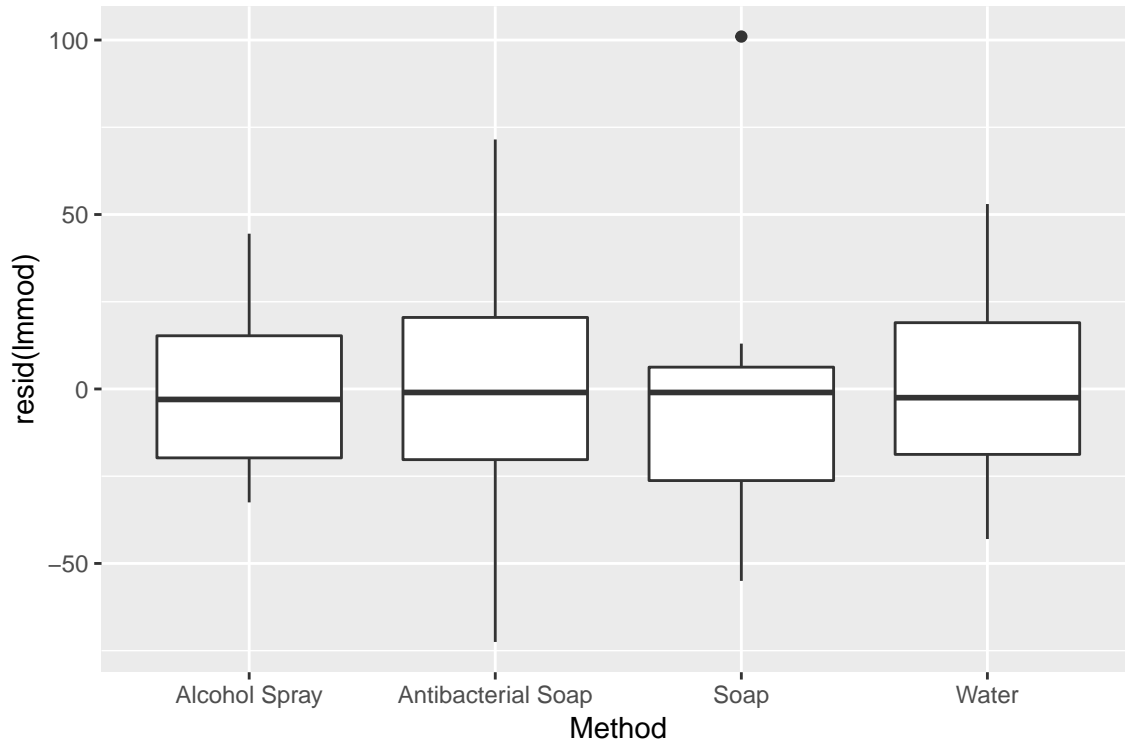
```
lmmod <- lm(Bacterial.Counts ~ Method, data = Soap)
msummary(lmmod)
```

```
##              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 ***
##
## 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
```

Section 24.3: Assumptions and Conditions

A box plot of the residuals shown by figure 24.5 on page 715 can be generated with the `gf_boxplot()` function. Figure 24.6 is made using the `gf_qq()` function.

```
gf_boxplot(resid(lmmod) ~ Method, data = Soap)
```



```
gf_qq(~ resid(lmmod), data = Soap, col = "lightcoral") %>%
  gf_labs(x = "qnorm", y = "resid(lmmod)")
```

