

IPS9 in R: Two-way analysis of variance (Chapter 13)

Bonnie Lin and Nicholas Horton (nhorton@amherst.edu)

July 20, 2018

Introduction and background

These documents are intended to help describe how to undertake analyses introduced as examples in the Ninth Edition of *Introduction to the Practice of Statistics* (2017) by Moore, McCabe, and Craig.

More information about the book can be found [here](#). The data used in these documents can be found under Data Sets in the Student Site. 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/ips9/>.

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>.

Chapter 13: Two-way analysis of variance

This file replicates the analyses from Chapter 13: Two-way analysis of variance.

First, load the packages that will be needed for this document:

```
library(mosaic)
library(readr)
```

Section 13.1: The two-way ANOVA model

```
HRTRATE <- read_csv("https://nhorton.people.amherst.edu/ips9/data/chapter13/EG13-08HRTRATE.csv")
head(HRTRATE)
```

```
## # A tibble: 6 x 3
##   Sex      Control Runners
##   <chr>    <int>    <int>
## 1 Female      159      119
## 2 Female      183       84
## 3 Female      140       89
## 4 Female      140      119
## 5 Female      125      127
## 6 Female      155      111
```

By default, the `read_csv()` function will output the types of columns, as we see above. To improve readability for future coding, we will suppress the “Parsed with column specification” message by adding `message = FALSE` at the top of the code chunks.

We need to transform the data from wide to tall format using the `gather()` function.

```
HRTRATE_tidy <- HRTRATE %>%
  tidyr::gather(key = Group, value = Heart_Rate, Control, Runners)
head(HRTRATE_tidy)
```

```
## # A tibble: 6 x 3
##   Sex    Group Heart_Rate
##   <chr> <chr>    <int>
## 1 Female Control      159
## 2 Female Control      183
## 3 Female Control      140
## 4 Female Control      140
## 5 Female Control      125
## 6 Female Control      155
```

```
## Figure 13.4, age 710
favstats(Heart_Rate ~ Sex + Group, data = HRTRATE_tidy)
```

```
##           Sex.Group min  Q1 median  Q3 max    mean      sd  n missing
## 1 Female.Control 105 137    147 160 196 148.000 16.27095 200      0
## 2 Male.Control   77 119    130 142 172 130.000 17.10035 200      0
## 3 Female.Runners  78 106    116 126 164 115.985 15.97154 200      0
## 4 Male.Runners   69  96    103 112 146 103.975 12.49942 200      0
```

```
## Figure 13.5, age 711
lm_HRTRATE <- lm(Heart_Rate ~ Group * Sex, data = HRTRATE_tidy)
msummary(lm_HRTRATE)
```

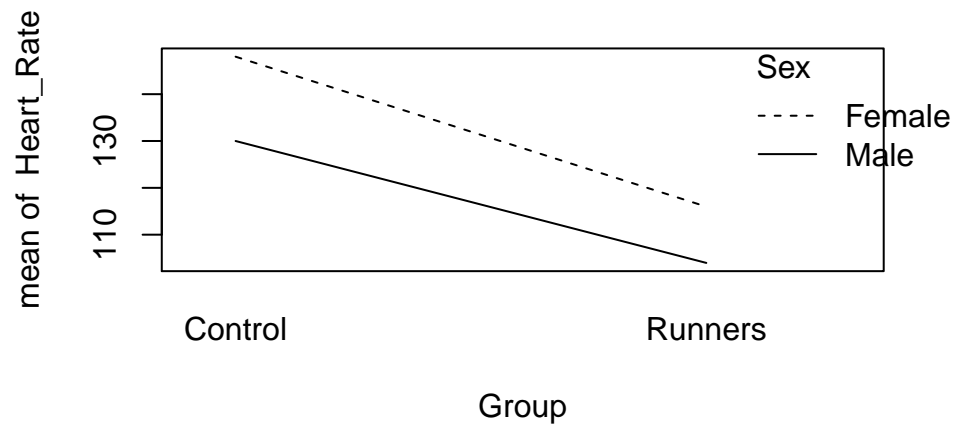
```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    148.000      1.100 134.511 < 2e-16 ***
## GroupRunners   -32.015      1.556 -20.575 < 2e-16 ***
## SexMale        -18.000      1.556 -11.568 < 2e-16 ***
## GroupRunners:SexMale  5.990      2.201   2.722 0.00663 **
##
## Residual standard error: 15.56 on 796 degrees of freedom
## Multiple R-squared:  0.5276, Adjusted R-squared:  0.5258
## F-statistic: 296.3 on 3 and 796 DF, p-value: < 2.2e-16
```

```
anova(lm_HRTRATE)
```

```
## Analysis of Variance Table
##
## Response: Heart_Rate
##           Df Sum Sq Mean Sq F value Pr(>F)
## Group      1 168432 168432 695.6470 < 2e-16 ***
## Sex        1  45030  45030 185.9799 < 2e-16 ***
## Group:Sex   1   1794   1794   7.4095 0.00663 **
## Residuals 796 192730    242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that there is a significant interaction ($p=0.007$): the sex difference between the heart rates differs by groups.

```
### Figure 13.6, page 712
with(HRTRATE_tidy, interaction.plot(Group, Sex, Heart_Rate))
```



Section 13.2: Inference for two-way ANOVA