SDM4 in R: Paired Samples and Blocks (Chapter 23)

Nicholas Horton (nhorton@amherst.edu)

April 12, 2016

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Stats: Data and Models* (2014) 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 http://www.amherst.edu/~nhorton/sdm4.

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).

Chapter 23: Paired samples and blocks

Section 23.1: Paired data

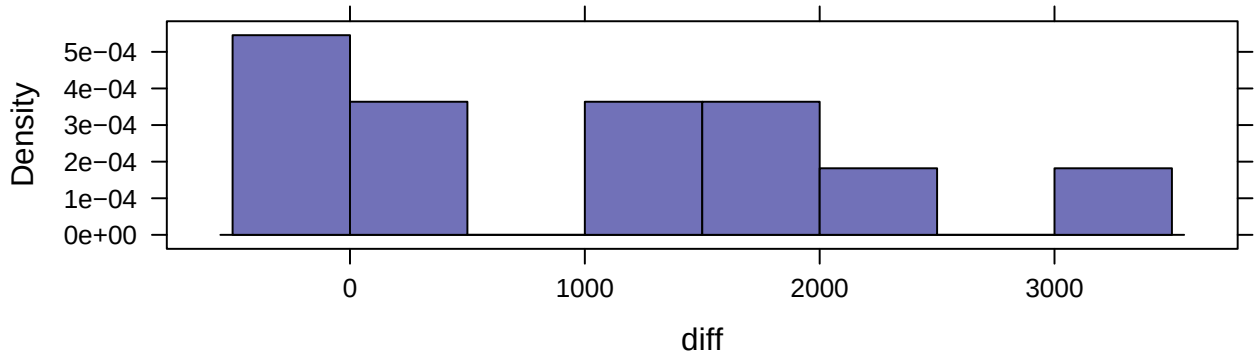
The example on page 631 compares the mileage of 11 field workers using either a 5 day or 4 day schedule.

```
fiveday <- c(2798, 7724, 7505, 838, 4592, 8107, 1228, 8718, 1097, 8089, 3807)
fourday <- c(2914, 6112, 6177, 1102, 3281, 4997, 1695, 6606, 1063, 6392, 3362)
ds <- data.frame(fiveday, fourday)
ds <- mutate(ds, diff = fiveday - fourday); ds
```

```
##    fiveday fourday diff
## 1     2798    2914 -116
## 2     7724    6112 1612
## 3     7505    6177 1328
## 4      838    1102 -264
## 5     4592    3281 1311
## 6     8107    4997 3110
## 7     1228    1695 -467
## 8     8718    6606 2112
## 9     1097    1063   34
## 10    8089    6392 1697
## 11    3807    3362  445
```

Section 23.2: Assumptions and conditions

```
histogram(~ diff, width=500, center=500/2, data=ds)   # page 634
```




```
t.test(~ diff, data=ds)
```

```
##
##  One Sample t-test
##
## data:  ds$diff
## t = 2.86, df = 10, p-value = 0.017
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   216.43 1747.57
## sample estimates:
## mean of x
##       982
```

Section 23.3: Confidence intervals for matched pairs

The same result is seen as on page 640 for the confidence interval for the population difference in mileage using the (results not shown).

```
t.test(~ diff, data=ds)$conf.int
```

Section 23.4: Blocking

The sign test on page 642 can be calculated using the `binom.test()` function.

```
binom.test(119, 151)
```

```
##
##
##
## data:  119 out of 151
## number of successes = 119, number of trials = 151, p-value =
## 5.6e-13
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.71421 0.85030
## sample estimates:
## probability of success
##                0.78808
```