SDM4 in R: Multifactor Analysis of Variance (Chapter 27)

Nicholas Horton (nhorton@amherst.edu) and Patrick Frenett

August 12, 2017

Introduction and background

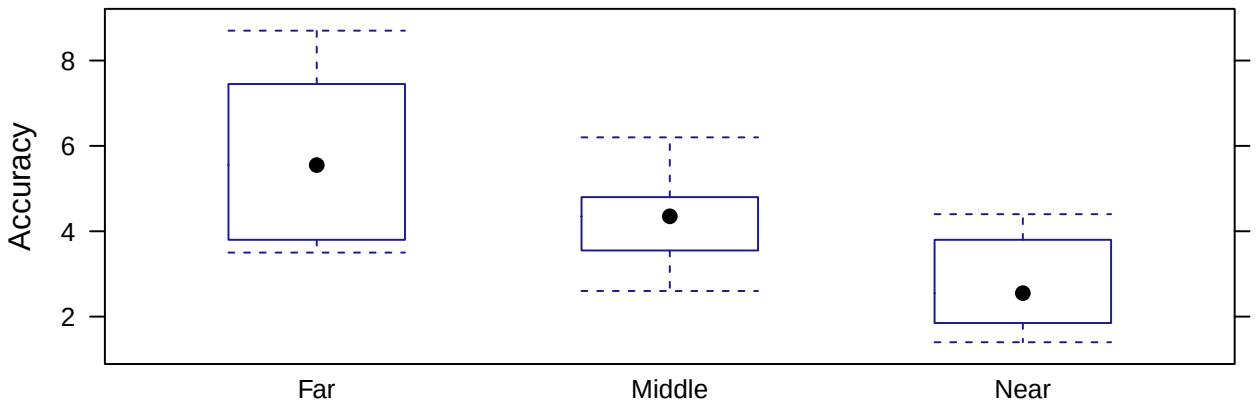
This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Stats: Data and Models* (2014) by De Veaux, Velleman, and Bock. More information about the book can be found at http://wps.aw.com/aw_deveaux_stats_series. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at http://nhorton.people.amherst.edu/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). A paper describing the mosaic approach was published in the *R Journal*: https://journal.r-project.org/archive/2017/RJ-2017-024.

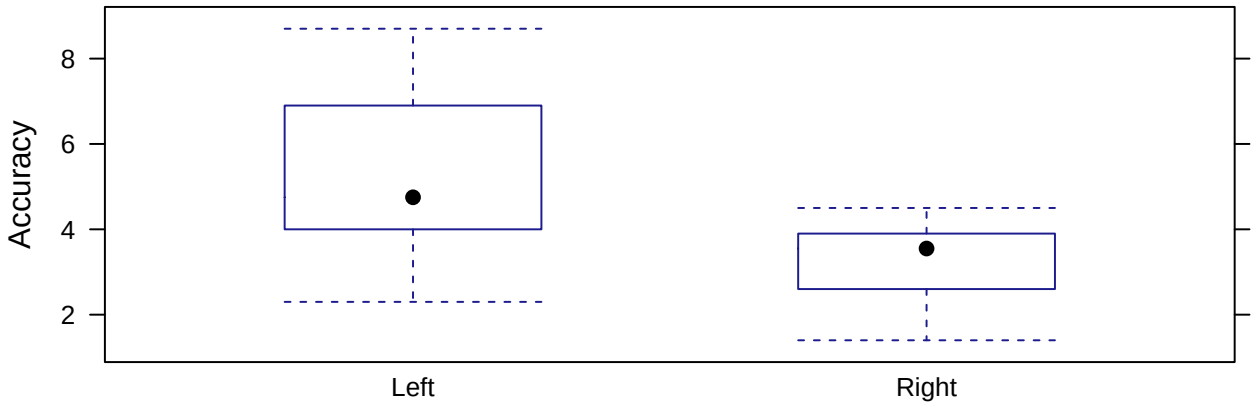
Chapter 27: Multifactor Analysis of Variance

`bwplot` gives the two boxplots from figure 27.1 seen on page 782.

```
Darts <- read.csv("http://nhorton.people.amherst.edu/sdm4/data/Ch29_Darts.csv")
bwplot(Accuracy ~ Distance, data=Darts)
```




```
bwplot(Accuracy ~ Hand, data=Darts)
```



Section 27.1: A Two-Factor ANOVA Model

The `summary` function gives a numerical summary of the linear model seen on page 784.

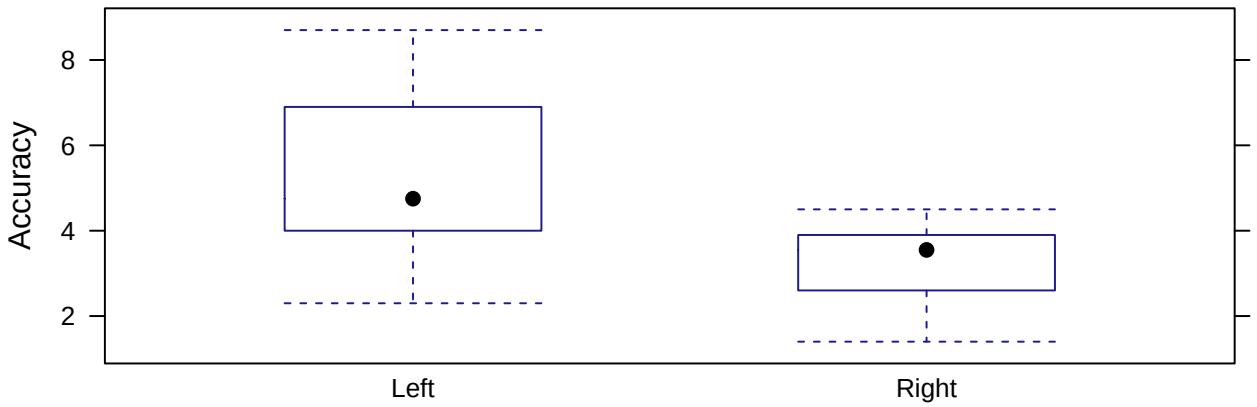
```
lmDarts <- lm(Accuracy ~ Distance + Hand, data=Darts)
summary(aov(lmDarts))
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Distance      2   51.0    25.5    28.6 7.6e-08 ***
## Hand          1   39.7    39.7    44.4 1.6e-07 ***
## Residuals    32   28.6     0.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

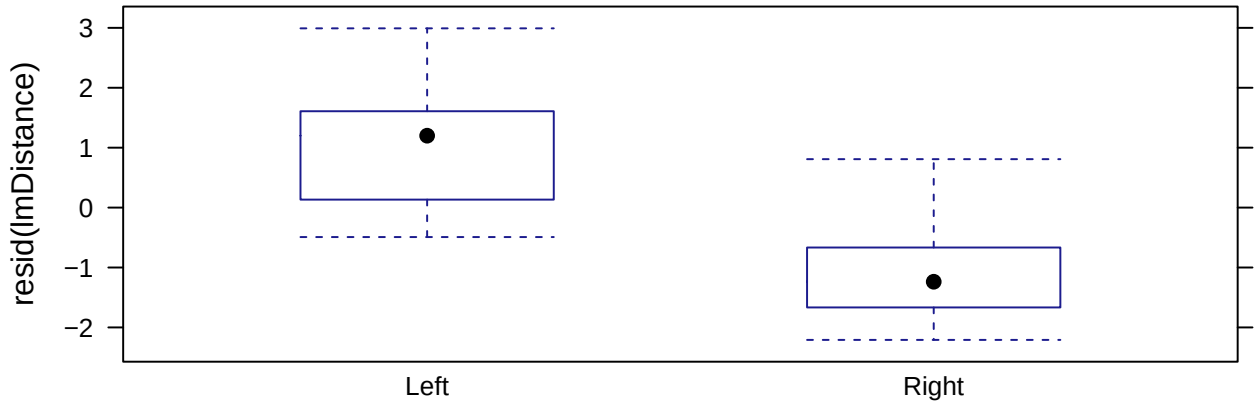
Section 27.2: Assumptions and Conditions

Figures 27.2 and 27.3 showing hand and distance against accuracy (as well as adjusted hand and adjusted distance). Note that the scales on the adjusted plots are centered around 0 rather than the mean as we are working with residuals.

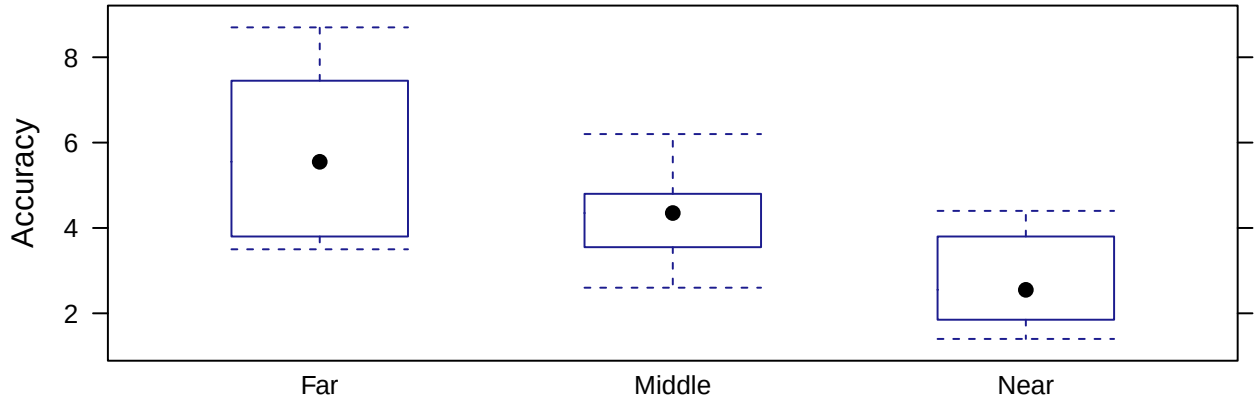
```
lmHand <- lm(Accuracy ~ Hand, data=Darts)
lmDistance <- lm(Accuracy ~ Distance, data=Darts)

bwplot(Accuracy ~ Hand, data=Darts)
```



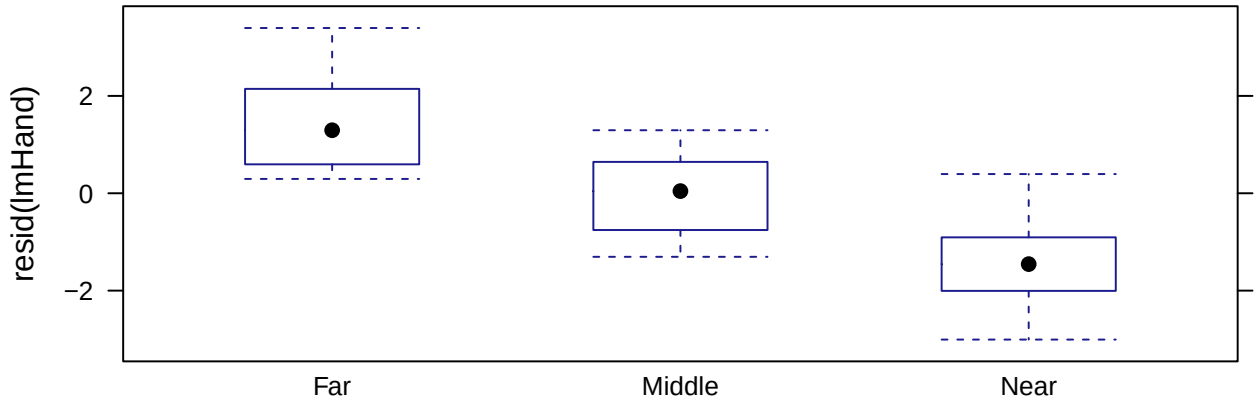
```
bwplot(resid(lmDistance) ~ Hand, data=Darts)
```



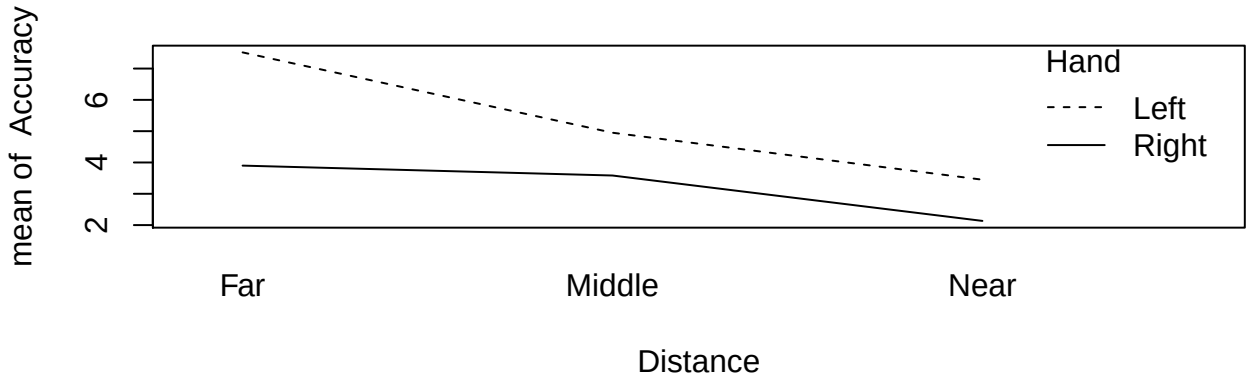

```
bwplot(Accuracy ~ Distance, data=Darts)
```




```
bwplot(resid(lmHand) ~ Distance, data=Darts)
```



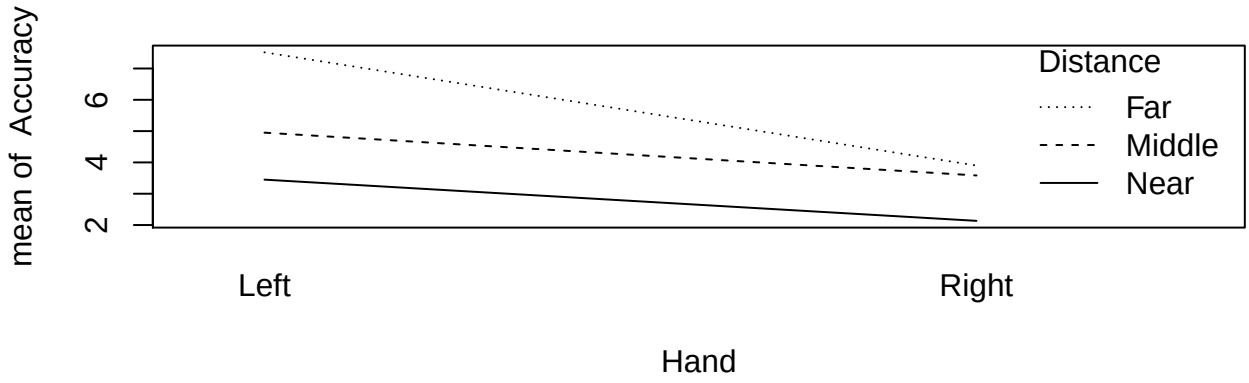

```
with(Darts, interaction.plot(Distance, Hand, Accuracy))
```



Section 27.3: Interactions

Shown below is the interaction plot (figure 27.9) and ANOVA summary on page 797.

```
with(Darts, interaction.plot(Hand, Distance, Accuracy))
```




```
lmDarts <- lm(Accuracy ~ Distance*Hand, data=Darts)
summary(aov(lmDarts))
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## Distance       2   51.0    25.5   41.98 2.0e-09 ***
## Hand           1   39.7    39.7   65.28 5.1e-09 ***
## Distance:Hand  2   10.4     5.2    8.52  0.0012 **
## Residuals     30   18.2     0.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```