Notes on Maxwell & Delaney

PSY710

12 higher-order within-subject designs

Chapter 11 discussed the analysis of data collected in experiments that had a single, within-subject factor. Here we extend those ideas to cover situations in which we have two (or more) within-subjects factors, and one within- and one between-subjects factor. As before, we consider only balanced data. Also, we assume that the subjects variable is random and the experimental factors are fixed. Therefore, we can generalize our results to different subjects, but not to different levels of the independent variables.

12.1 factorial within-subjects designs

We first consider a situation in which all combinations of two experimental variables are presented to each subject. We call this a within-subject factorial design: all of the factors are crossed within every subject. For instance, the data in Table 12.1 in your textbook presents the data from a hypothetical perception experiment that uses a 2x3 within-subject factorial experiment to examine the effects of distractors, or noise, and stimulus orientation on the time required to discriminate two letters. The noise factor has two levels – present vs. absent – and the angle of the stimulus was 0, 4, or 8 deg from upright/vertical. The dependent variable is response time (RT). Data were collected from ten subjects; every subject was tested in all six experimental conditions, and the order of conditions was randomized for each subject.

A full model for this two-way within-subjects design is

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \pi_i + (\alpha\beta)_{jk} + (\alpha\pi)_{ji} + (\beta\pi)_{ki} + (\alpha\beta\pi)_{jki} + \epsilon_{ijk} \quad (1)$$

where Y_{ijk} is the score on the dependent variable from subject i in the jth level of A and the kth level of B; μ is the intercept; α_j is the effect associated with the jth level of A; β_k is the effect associated with the kth level of B; π_i is the effect associated with the ith subject; $(\alpha\beta)_{jk}$ is the effect of the interaction between the jth level of A and the kth level of B; $(\alpha\pi)_{ji}$ is the effect of the interaction between the jth level of A and the ith subject; $(\beta\pi)_{ki}$ is the effect of the interaction between the kth level of B and the ith subject; $(\alpha\beta\pi)_{jki}$ is the effect of the three-way interaction of the jth level of A, the kth level of B, and the ith subject; and ϵ_{ijk} is the error for the observation of the jth level of A, the kth level of B, and the ith subject. α, β, and the $(\alpha\beta)$ are fixed effects, whereas π and all of the interactions with π are random effects. All of the effects satisfy the sum-to-zero constraint.

The magnitude of each effect in Equation 1 can be evaluated by comparing the full model to a restricted model that omits the parameters of the effect begin evaluated. The difference between $SS_{residuals}$ for the full and restricted model represents the sum of squares for the effect being evaluated. The degrees of freedom for the effect is simply the difference between the degrees of freedom for the full and restricted models that are being compared. In other words, the sum of squares and degrees of freedom are calculated just as they are calculated in the analysis of data from a between-subjects factorial design. The only thing that distinguishes the analysis of a within-subjects design from that of a between-subjects design is the choice of the error term for each F test.

In a between-subjects factorial design, the main effects of A and B and the $A \times B$ interaction are evaluated by dividing the appropriate mean square by $MS_{residuals}$, which is sometimes referred to as $MS_{WithinCell}$. Such an approach is inappropriate for within-subjects factorial designs. In fact, there is no $MS_{WithinCell}$

term: each cell in our design consists of a single observation, so it is not possible to use the full model to estimate within-cell error.

Recall that, in the case of a one-way within-subjects design, the effect of the treatment variable was evaluated by comparing $MS_{treatment}$ to a term that represented the combined effects of error and the *treatment* $\times$ *subjects* interaction. The same calculation will be performed here, too. The main effect of A will be evaluated using an error term that represents the $A \times$ *subjects* interaction; the main effect of B will be evaluated using an error term that represents the $B \times$ *subjects* interaction; and the $A \times B$ interaction will be evaluated using an error term that represents the $A \times B \times$ *subjects* interaction. An examination of the expected values of these mean squares, shown in Table 12.3 (page 576) in your textbook, will clarify why these tests are appropriate.

12.1.1 R example

My example uses the data from Table 12.1 in your textbook, although I modified the numbers to make the F and p values more believable. I loaded the data from the CD that came with the textbook, added a variable named `subj` that lists the subject ID number, and then changed then names of the dependent variables to make it easier to identify the conditions. The data file looks like this:

```
> rtData

   subj absent.a0 absent.a4 absent.a8 present.a0 present.a4 present.a8
1    s1       420       420       480        480        600        780
2    s2       420       480       480        360        480        600
3    s3       480       480       540        660        780        780
4    s4       420       540       540        480        780        900
5    s5       540       660       540        480        660        720
6    s6       360       420       360        360        480        540
7    s7       480       480       600        540        720        840
8    s8       480       600       660        540        720        900
9    s9       540       600       540        480        720        780
10  s10       480       420       540        540        660        780
```

Next, I extract the six columns that contain the RTs from the various conditions and store the results in a matrix, `rt.mat`. Then I add some gaussian noise to the numbers:

```
> rt.mat <- as.matrix(rtData[1:10,2:7])
> set.seed(540912);
> rt.nz <- matrix(data=round(rnorm(n=10*6,sd=100)),nrow=10,ncol=6)
> rt.mat <- rt.mat + rt.nz
```

Now I can begin to analyze the data. First, I use `lm` to create a multivariate object. Notice that the formula in `lm` contains multiple dependent variables (i.e., `rt.mat`) and only the intercept.

```
> rt.mlm <- lm(rt.mat~1)
```

I am going to use the `Anova` command in the `car` library to convert the multivariate object to an anova object. To use `Anova` I have to create a data frame that contains the within-subject factors. I created the data frame `rt.idata` in a spread sheet program and then loaded it into R. Notice how the structure of `rt.idata` corresponds to the column names of the dependent variables:

```
> rt.idata

   noise angle
1 absent    a0
```

```
2  absent    a4
3  absent    a8
4 present    a0
5 present    a4
6 present    a8
```

```
> rt.mat

   absent.a0 absent.a4 absent.a8 present.a0 present.a4 present.a8
1        657       446       461        671        474        903
2        450       448       484        284        562        585
3        569       299       548        698        950        727
4        418       519       629        425        828       1000
5        634       661       528        452        683        879
6        392       369       433        481        453        509
7        382       505       581        425        572        847
8        602       762       523        723        695        879
9        532       742       584        364        572        852
10       578       467       570        594        634        804
```

In the Anova command, notice how I specify the within-subjects design with the one-sided formula $\sim$ noise $*$ angle:

```
> library(car)
> rt.aov <- Anova(rt.mlm,idata=rt.idata,idesign=~noise*angle,type="III")
> summary(rt.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                  SS num Df Error SS den Df      F  Pr(>F)
(Intercept) 20765813      1   344423      9 542.62 2.4e-09 ***
noise         234625      1   134551      9  15.69  0.0033 **
angle         225422      2   260280     18   7.79  0.0036 **
noise:angle   187983      2   225402     18   7.51  0.0043 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

            Test statistic p-value
angle                0.690   0.227
noise:angle          0.639   0.167


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

            GG eps Pr(>F[GG])
angle        0.764     0.0083 **
noise:angle  0.735     0.0103 *
---
```

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

             HF eps Pr(>F[HF])
angle         0.888     0.0054 **
noise:angle  0.843     0.0072 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first part of the summary lists the ANOVA table. The `noise:angle` interaction is significant, $F(2, 18) = 7.51$, $p = 0.0043$, as are the main effects of `angle`, $F(2, 18) = 7.79$, $p = 0.0036$, and `noise`, $F(1, 9) = 15.69$, $p = 0.0033$. Note that the denominator degrees of freedom are different for the two main effects because different error terms are used to evaluate the significance of `angle` and `noise`.

The **sphericity assumption** was introduced in Chapter 11. It applies here, too. The results in the ANOVA table assume that sphericity is valid, but of course we need to evaluate it before accepting the p values listed in the table. The second part of the `Anova` output shows the results of the Mauchly test for sphericity. Notice that sphericity tests are done only for the `angle` and `noise:angle` terms, not `noise`. The reason deviations from sphericity are not examined for the main effect of `noise` is that that factor has only two levels and therefore one degree of freedom: sphericity necessarily is valid for F tests that have one degree of freedom in the numerator, and therefore sphericity need not be evaluated for that effect. The Mauchly tests for sphericity are not significant ($p > 0.1$ in both cases), so we could use p values listed in the ANOVA table. However, I still prefer to use the corrected p values listed in the final part of the output. The Greenhouse-Geisser (G-G) estimate of epsilon, $\hat{\epsilon}$, is 0.76 for the main effect of `angle` and 0.73 for the `noise:angle` interaction. The Huynh-Feldt (H-F) estimate of epsilon, $\tilde{\epsilon}$, is 0.89 for the main effect of `angle` and 0.84 for the `noise:angle` interaction. Occasionally, $\tilde{\epsilon} > 1$: In such situations it is standard practice to set $\tilde{\epsilon} = 1$. Either the G-G or the H-F adjusted p values are acceptable, but I prefer to use the H-F adjustment because it is slightly less conservative. So, the `noise:angle` interaction is significant, $F(2, 18) = 7.51$, $\tilde{\epsilon} = 0.84$, $p = 0.0072$, as are the main effects of `angle`, $F(2, 18) = 7.79$, $\tilde{\epsilon} = 0.89$, $p = 0.0054$, and `noise`, $F(1, 9) = 15.69$, $p = 0.0033$.

12.1.2 strength of association

Your textbook gives the equation for omega squared which expresses the variance of an effect relative to the sum of the effect variance, error variance, and the variance due to subjects. Another common measure of the strength of association between the dependent variable and levels of a within-subject factor is partial omega squared, which expresses the variance of the effect relative to the sum of the effect variance and error variance (Keppel and Wickens, 2004; Kirk, 1995). For example, partial omega squared for A is

$$\omega^2_{Y|A\cdot(B,\ A\times B)} = \frac{\sigma^2_\alpha}{\sigma^2_e + \sigma^2_\alpha} \tag{2}$$

and can be calculated from an ANOVA table with the formula:

$$\omega^2_{Y|A\cdot(B,\ A\times B)} = \frac{(a-1)(F_A-1)}{(a-1)(F_A-1)+nab} \tag{3}$$

Partial omega squared for B is calculated by:

$$\omega^2_{Y|B\cdot(A,\ A\times B)} = \frac{(b-1)(F_B-1)}{(b-1)(F_B-1)+nab} \tag{4}$$

Partial omega squared for $A \times B$ is calculated by:

$$\omega^2_{Y|A\times B\cdot(A,\ B)} = \frac{(a-1)(b-1)(F_{AB}-1)}{(a-1)(b-1)(F_{AB}-1)+nab} \tag{5}$$

The key difference between the measure defined in your textbook and the one defined here is how variation due to subjects is handled. Variation due to subjects is included in the denominator of omega squared in your textbook, but it is not included in the denominator of partial omega squared defined here. Hence, the strength of association indexed by Equation 2 will be larger than the value of omega squared defined by Equation 6 in chapter 12 in your textbook. Partial omega squared can be used to calculate Cohen's f (Kirk, 1995). The following equation shows how to calculate f_A:

$$\hat{f}_A = \sqrt{\frac{\omega^2_{Y|A\cdot(B,\ A\times B)}}{1-\omega^2_{Y|A\cdot(B,\ A\times B)}}} \tag{6}$$

It should be obvious how to change Equation 6 to calculate effects sizes for B and $A \times B$. Partial omega squared and f for the data analyzed in section 12.1.1 are listed in Table 1.

Source	Partial ω^2	Cohen's f
noise	0.196	0.494
angle	0.185	0.475
noise:angle	0.178	0.466

Table 1: Strength of association and effect sizes for `rt.mat` data.

12.1.3 simple main effects

The `angle:noise` interaction was significant, so we should examine the simple main effects. First we should plot the data to get an idea of what the interaction might mean. The following commands were used to create Figure 1.

```
> ( absent.means <- colMeans(rt.mat[1:10,1:3]) )

absent.a0 absent.a4 absent.a8
    521.4     521.8     534.1

> ( present.means <- colMeans(rt.mat[1:10,4:6]) )

present.a0 present.a4 present.a8
     511.7      642.3      798.5

> x.angle <- c(0,4,8);
> plot(x=c(0,4,8),absent.means,"b",ylim=c(450,800),ylab="RT",xlab="angle")
> points(x=c(0,4,8),present.means,"b",pch=19)
> legend(x=0,y=750,legend=c("absent","present"),pch=c(21,19))
```

It appears that the difference between the noise absent and present conditions increased with increasing stimulus angle. Alternatively, we can say that the effect of angle appears to be larger in the noise present condition than in the noise absent condition. I will evaluate this idea by conducting tests of the simple main effect of angle at each level of noise.

In a two-way, between-subjects factorial design, simple main effects were evaluated by doing separate one-way ANOVAs, but using $MS_{residuals}$ from the overall analysis as the error term. However, **for within-subjects designs it is better to use separate error terms for each analysis.** Therefore, evaluating simple main effects is essentially identical to conducting a set of one-way, within-subject ANOVAs. The following code illustrates how to conduct simple main effects of `angle` at each level of `noise`. First, we extract the data from the noise present and noise absent conditions:

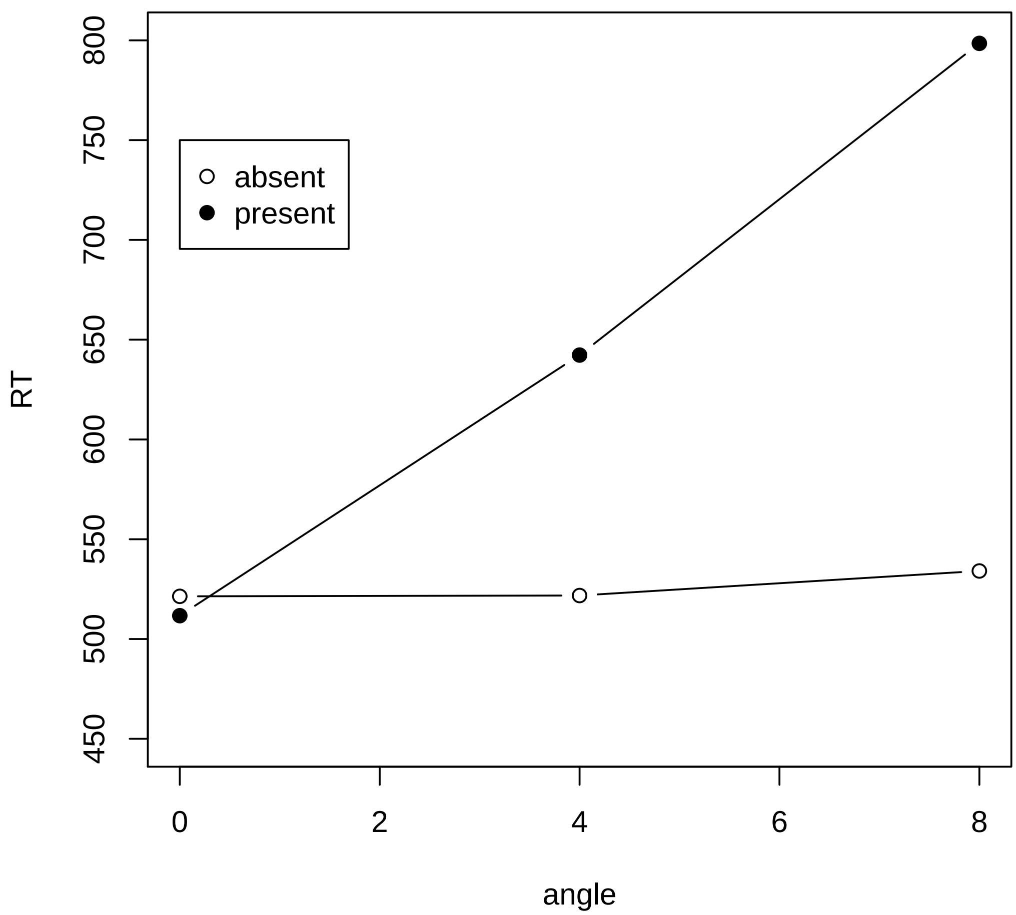

Figure 1: Interaction plot of data in `rt.mat`.

```
> rt.mat

   absent.a0 absent.a4 absent.a8 present.a0 present.a4 present.a8
1        657       446       461        671        474        903
2        450       448       484        284        562        585
3        569       299       548        698        950        727
4        418       519       629        425        828       1000
5        634       661       528        452        683        879
6        392       369       433        481        453        509
7        382       505       581        425        572        847
8        602       762       523        723        695        879
9        532       742       584        364        572        852
10       578       467       570        594        634        804

> rt.absent <- rt.mat[,1:3]
> rt.present <- rt.mat[,4:6]
```

The dependent variables are stored in the matrix `rt.mat`. The syntax `rt.mat[, 1:3]` is a way of specifying "all rows in columns 1 through 3", and is equivalent to `rt[1:10,1:3]`. If we wanted to access data from the data frame `rtData`, we could use the following commands:

```
> rt.absent <- as.matrix(rtData[,2:4])
> rt.present <- as.matrix(rtData[,5:7])
```

However, we added noise to the numbers in `rtData`, and so I will use the numbers in `rt.mat`. Next, we conduct a one-way ANOVA to evaluate the effect of `angle` when the noise is present. First, we use `lm` to create a multivariate lm object:

```
> ang.present.mlm <- lm(rt.present~1)
```

Next, we create a data frame that contains one three-level factor names `angle`

```
> angle <- factor(x=c(1,2,3),label=c("a0","a4","a8"))
> angle.idata <- data.frame(angle)
```

and then use the data frame in conjunction with the `Anova` command to create an aov object. Notice how the following command uses the `idata` and `idesign` parameters to define the within-subjects design for this subset of data:

```
> ang.present.aov <- Anova(ang.present.mlm,idata=angle.idata,idesign=~angle,type="III")
```

Finally, we use the `summary` command to list the ANOVA table and adjustments for sphericity:

```
> summary(ang.present.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                  SS num Df Error SS den Df     F  Pr(>F)
(Intercept) 12707521      1   325698      9 351.1 1.6e-08 ***
angle         412363      2   297401     18  12.5   4e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity
```

```
      Test statistic p-value
angle          0.993   0.972


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle  0.993    0.00041 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      HF eps Pr(>F[HF])
angle   1.27      4e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Mauchly test for deviation from sphericity is not significant ($p = 0.97$), so we can in good conscience use the p values in the ANOVA table: The simple main effect of angle when `noise` is present is significant, $F(2, 18) = 12.5$, $p < .001$. Note that the same conclusion is reached if we use the G-G or H-F adjusted p values.

Here are the commands for evaluating the simple main effect of `angle` when `noise` is absent:

```
> ang.absent.mlm <- lm(rt.absent~1)
> ang.absent.aov <- Anova(ang.absent.mlm,idata=angle.idata,idesign=~angle,type="III")
> summary(ang.absent.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                 SS num Df Error SS den Df      F  Pr(>F)
(Intercept) 8292918      1   153275      9 486.94 3.8e-09 ***
angle          1042      2   188282     18   0.05    0.95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

      Test statistic p-value
angle          0.936   0.769


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle   0.94       0.94

      HF eps Pr(>F[HF])
angle   1.18       0.95
```

The Mauchly test for the deviation from sphericity is not significant, and therefore I will use the unadjusted p value in the ANOVA table: The simple main effect of `angle` when `noise` is absent is not significant, $F((2,18) = 0.05$, $p = 0.95$.

Finally, we could use the Bonferroni adjustment to maintain a familywise Type I error rate of .05 by using a per-comparison Type I error rate of $\alpha = .05/2 = .025$. Using this new criterion, our conclusions remain the same: the simple main effect of `angle` is significant when `noise` is present but not when it is absent.

12.1.4 linear comparisons

The procedures for conducting linear comparisons among treatment levels are similar to those used in to analyze data from one-way within-subject designs. First, we create our set of contrast weights. Next, we use the weights to create composite scores for each subject. Finally, if our null hypothesis is not directional, we use a t test to determine if the composite scores differ from zero.

The following code shows how to test the hypothesis that there is a significant linear trend of RT across `angle`. The `angle:noise` interaction was significant, so I will evaluate the contrast separately at each level of `noise`. In other words, we will analyze each simple main effect separately. I am ignoring the effects of multiple comparisons on Type I error. Obviously, such effects would have to be taken into account when doing a real analysis. Also note that in this situation a test of linear trend is equivalent to comparing the means of the first and third levels of `angle`:

Here is the code to do the comparisons:

```
> lin.C <- c(-1,0,1)
> rt.pres.lin <- rt.present %*% lin.C
> t.test(rt.pres.lin)

        One Sample t-test

data:  rt.pres.lin
t = 4.801, df = 9, p-value = 0.0009719
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 151.7 421.9
sample estimates:
mean of x
    286.8

> rt.absent.lin <- rt.absent %*% lin.C
> t.test(rt.absent.lin)

        One Sample t-test

data:  rt.absent.lin
t = 0.317, df = 9, p-value = 0.7585
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -77.93 103.33
sample estimates:
mean of x
     12.7
```

The t test is significant when `noise` is present, $t(9) = 4.8$, $p < .001$, but not when `noise` is absent, $t(9) = 0.32$, $p = 0.75$. Therefore, we reject the null hypothesis of no linear trend in the noise-present conditions but not in the noise-absent conditions.

Next, we consider a case where the comparison is done on the entire data set, rather than separately on the noise-present and noise-absent data. Such an analysis is inappropriate in this case because the `angle:noise` interaction is significant, but we continue for illustrative purposes. In what follows, note how my contrast now includes six values, not three, and how the values in my contrast correspond to the order of the columns in the data matrix.

```
> rt.mat

   absent.a0 absent.a4 absent.a8 present.a0 present.a4 present.a8
1        657       446       461        671        474        903
2        450       448       484        284        562        585
3        569       299       548        698        950        727
4        418       519       629        425        828       1000
5        634       661       528        452        683        879
6        392       369       433        481        453        509
7        382       505       581        425        572        847
8        602       762       523        723        695        879
9        532       742       584        364        572        852
10       578       467       570        594        634        804

> lin.C <- c(-1,0,1,-1,0,1)
> rt.lin <- rt.mat %*% lin.C
> t.test(rt.lin)

        One Sample t-test
data:  rt.lin
t = 3.481, df = 9, p-value = 0.006925
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 104.9 494.1
sample estimates:
mean of x
    299.5
```

The value of t is significant, so we reject the null hypothesis of no linear trend across `angle` *ignoring the effect of* `noise`.

Now suppose we wanted to know if the linear trend across `angle` *differed* across levels of `noise`. This question is equivalent to asking if the linear trend across `angle` interacts with `noise`. To answer this question, we want to compute the difference between the linear trend in conditions when noise is present and the linear trend when noise was absent. Therefore, the contrast weights needed to address this question are [(-1, 0, 1) - (-1, 0, 1)], or [-1, 0, 1, 1, 0, -1]:

```
> myC <- c(-1,0,1,1,0,-1)
> rt.lin.x.noise <- rt.mat %*% myC
> t.test(rt.lin.x.noise)

        One Sample t-test
data:  rt.lin.x.noise
t = -5.051, df = 9, p-value = 0.000689
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -396.8 -151.4
```

```
sample estimates:
mean of x
   -274.1
```

The contrast is significant, $t(9) = 3.48$, $p = 0.007$, so we conclude that the linear trend differs in the noise-present and noise-absent conditions. This result is consistent[1] with our previous analyses that found that there was a significant interaction between `angle` and `noise`, and that the linear trend across `angle` was significant only in the noise-present conditions.

The previous test of a angle-contrast x noise interaction used a t test, but the same test also can be performed using an F test. The F test is a more general procedure than the t test, and so I will describe it here. First, I re-create separate linear trend scores for the noise-present and noise-absent conditions and then combine the scores into a single two-column matrix:

```
> lin.C <- c(-1,0,1)
> rt.pres.lin <- rt.present %*% lin.C
> rt.absent.lin <- rt.absent %*% lin.C
> lin.scores <- cbind(rt.pres.lin,rt.absent.lin)
> dimnames(lin.scores)[[2]] <- c("nz.pres","nz.absnt")
> lin.scores

   nz.pres nz.absnt
1      232     -196
2      301       34
3       29      -21
4      575      211
5      427     -106
6       28       41
7      422      199
8      156      -79
9      488       52
10     210       -8
```

Next, I conduct a within-subjects ANOVA on `lin.scores`:

```
> lin.scores.mlm <- lm(lin.scores~1)
> nz <- as.factor(c("present","absent"))
> lin.scores.aov <- Anova(lin.scores.mlm,idata=data.frame(nz),idesign=~nz,type="III")
> summary(lin.scores.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
                SS num Df Error SS den Df    F  Pr(>F)
(Intercept) 448501      1   333077      9 12.1 0.00693 **
nz          375654      1   132496      9 25.5 0.00069 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The effect of `nz` is significant, $F(1, 9) = 25.5$, $p = .007$. What does this effect mean? It means that we reject the null hypothesis that the composite scores – i.e., the linear trend – is the same at both levels of noise in

[1] Although the two analyses yield results that are consistent, each analysis evaluates a different question. The significant `angle` × `noise` interaction in the overall ANOVA evaluates a *general* hypothesis, namely that the effect of one variable depends on the level of the other variable. The current test evaluates the *specific* hypothesis that the linear trend (across angle) differs across noise levels.

favor of the alternative hypothesis that the composite scores differed in the noise-present and noise-absent conditions. In other words, there was a significant linear trend x noise interaction. Also note that the F observed here equals the squared t value obtained in the previous t test ($t^2 = (-5.051)^2 = 25.51$): these tests are evaluating the same hypotheses, and therefore the statistics are simple transforms of each other. Also note that the intercept term is significant, $F(1, 9) = 12.1$, $p = 0.0069$. When we use sum-to-zero coding for our effects, the intercept corresponds to the grand mean, and therefore this F test evaluates the null hypothesis that the grand mean of the scores is zero. Normally such a test is not interesting, but in this case the test is informative: it tells us that the average linear trend score differs significantly from zero. Note again that the F value for the intercept is the squared t value obtained in our earlier test of this same hypothesis ($t^2 = 3.481^2 = 12.1$). If we get the same values as our t tests, why would we ever want to use an F test? The F test is useful in situations where the second variable has more than two levels. In the current situation, for example, it is not clear how a t test could be used to evaluate a trend x noise interaction *if noise had more than two levels*. In fact, it is not possible to test for such an interaction using a single t test, but it is possible to perform the F test to see if the trend (or other linear contrast) varies across n levels of the other within-subject variable.

12.2 Split plot designs

Some experiments use mixtures of within- and between-subjects factors. Such designs often are called split plot designs.

Your textbook illustrates the analysis of data from split-plot a experiment using data presented in Tables 12.7 and 12.15. The hypothetical experiment measured RT in a group of young subjects (Table 12.7) and senior subjects (Table 12.15). Each subject participated in three conditions in which the visual stimuli were presented at different angles. In this design, `group` is the between-subject variable (of age) and stimulus `angle` is a within-subjects variable. I will use the same story, but I've created my own fake data:

```
> myData

   group a1 a2 a3
1  young 50 47 51
2  young 41 57 43
3  young 42 63 40
4  young 46 66 47
5  young 45 61 38
6  young 45 57 53
7    old 48 39 38
8    old 55 72 54
9    old 51 44 51
10   old 53 65 53
11   old 68 58 62
12   old 65 37 55
```

Note that the data frame contains the factor `group`, which is a between-subjects variable. Next, I use `lm` to create a multivariate-lm object. Note that the between-subjects formula now includes the between-subjects variable `group`:

```
> myData.mlm <- lm(cbind(a1,a2,a3)~1+group,data=myData)
```

The remaining parts of the analysis are the same as before:

```
> angle <- as.factor(c("a1","a2","a3"))
> myData.idata <- data.frame(angle)
> myData.aov <- Anova(myData.mlm,idata=myData.idata,idesign=~angle,type="III")
> summary(myData.aov,multivariate=F)
```

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
               SS num Df Error SS den Df      F  Pr(>F)
(Intercept) 96100      1     1002     10 958.87 2.9e-11 ***
group         160      1     1002     10   1.60   0.234
angle         289      2     1143     20   2.52   0.105
group:angle   508      2     1143     20   4.44   0.025 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

            Test statistic p-value
angle                0.354 0.00933
group:angle          0.354 0.00933


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

            GG eps Pr(>F[GG])
angle        0.607      0.135
group:angle  0.607      0.051 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            HF eps Pr(>F[HF])
angle        0.647      0.132
group:angle  0.647      0.047 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table shows the uncorrected p values for the main effects of `group` and `angle`, and the `group:angle` interaction. Examination of the table will show that the effects of `group` and `angle` are evaluated with different error terms. The `Anova` command assumes that the between- and within-subjects variables are *fixed*, and uses the appropriate error term to generate unbiased F tests (see Table 12.17, page 596 in your textbook for the expected mean squares for this design). The second part of the output contains the results of the Mauchly tests for sphericity: Note that a test is done for the interaction as well as the within-subjects variable. In general, the sphericity assumption applies to within-subjects factors and all interactions that contain within-subjects factors. The third part of the output shows the G-G and H-F adjusted p values. The Mauchly test is significant, so I will use H-F adjusted p values. There is a significant `group:angle` interaction, $F(2, 20) = 4.44$, $\tilde{\epsilon} = 0.65$, $p = 0.047$, but the main effects of `group` and `angle` are not significant.

Let's think about what the different components of the ANOVA table actually mean. First consider the between-subjects main effect of group. To illustrate what this main effect represents, I am going to do a one-way between-subjects analysis on the *averaged* within-subject scores:

```
> y.mat <- myData[,2:4]
> y.avg <- rowMeans(y.mat)
> summary(aov(y.avg~group,data=myData))

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

```
group        1     53    53.5     1.6   0.23
Residuals   10    334    33.4
```

Notice that the F and p values for the effect of `group` are the same (to within rounding error) as the values listed in the split-plot ANOVA table. In fact, the two analyses of `group` are equivalent: the analysis of the between-subjects factor in a split-plot design is the same as a between-subjects ANOVA on the average score for each subject. The sums of squares and means squares for `group` and residuals listed in the two ANOVA tables differ because the different numbers of data points are analyzed in the two cases, but the ratio of the mean squares are the same. One more thing: if the evaluation of the between-subjects variable is equivalent to a one-way, between-subjects ANOVA, then it should not matter if we have different n in each level of the between-subject variable. And, indeed, it does not matter: having unequal n on the *between-subjects* variable does not cause significant problems with the analysis.

Next, let's consider the error term that is used to evaluate the main effect of the within-subjects factor. In the case where both variables were within-subject factors, each effect was evaluated with an error term that was the interaction of that effect with subjects: A was evaluated with $A \times S$, and B was evaluated with $B \times S$, and $A \times B$ was evaluated with $A \times B \times S$. In the current case, if B is the within-subjects variable, then it is evaluated with $B \times S/A$, which is the mean square interaction of B and subjects *nested within* A. This error term is equivalent to the weighted average of the values of $\text{MS}_{B \times S}$ at each level of the between-subjects variable, A. The following code calculates the ANOVA for the within-subject variable, `angle`, separately for each group:

```
> dat.young <- subset(myData,group=="young")
> dat.old <- subset(myData,group=="old")
> dat.young.mlm <-lm(cbind(a1,a2,a3)~1,data=dat.young)
> dat.old.mlm <-lm(cbind(a1,a2,a3)~1,data=dat.old)
> dat.young.aov <- Anova(dat.young.mlm,idata=myData.idata,idesign=~angle,type="III")
> dat.old.aov <- Anova(dat.old.mlm,idata=myData.idata,idesign=~angle,type="III")
```

Here are the ANOVA tables (I've deleted the other parts of the output):

```
> summary(dat.young.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
               SS num Df Error SS den Df       F  Pr(>F)
(Intercept) 44204      1       80      5 2747.46 4.8e-08 ***
angle         721      2      371     10    9.71  0.0045 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

      Test statistic p-value
angle          0.429   0.184


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle  0.636      0.016 *
---
```

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

      HF eps Pr(>F[HF])
angle  0.756      0.011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(dat.old.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
               SS num Df Error SS den Df      F  Pr(>F)
(Intercept) 52057      1      922      5 282.37 1.4e-05 ***
angle          75      2      772     10   0.49    0.63
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

      Test statistic p-value
angle          0.145   0.021


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle  0.539       0.53

      HF eps Pr(>F[HF])
angle   0.57       0.54
```

Notice that the sum of squares for the error terms in the two analyses are 371 and 772, and that the sum of these values, 1143, equals the sum of squares for `angle's` error term in the original split-plot analysis. Also note that the average of the two mean square values, $(37.1 + 77.2)/2 = 57.15$, is the same as the mean square of the error term in the original analysis. Hence, the error term that is used to evaluate the within-subjects variable in a split-plot analysis can be thought of as an average of the error terms in a series of within-subjects ANOVAs.

12.2.1 simple main effects

Our previous analysis found a significant `group:angle` interaction, so we should examine simple main effects. We start by examining the simple main effect of the between-subject factor, `group` at each level of the within-subject factor, `angle`. Each analysis uses a separate error term calculated on the particular subset of data begin examined, and therefore is equivalent to a one-way between subjects ANOVA.

```
> names(myData)

[1] "group" "a1"    "a2"    "a3"

> summary(aov(a1~1+group,data=myData))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        1    420     420    11.3 0.0072 **
Residuals   10    372      37
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(aov(a2~1+group,data=myData))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        1    108     108    0.84   0.38
Residuals   10   1281     128
```

```
> summary(aov(a3~1+group,data=myData))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        1    140   140.1    2.85   0.12
Residuals   10    492    49.2
```

The analyses indicate that the simple main effect of age `group` is significant only in the first level of `angle`, $F(1, 10) = 11.3$, $p = 0.007$.

We can also look at the simple main effect of the within-subject variable at each level of the between subject variable. Notice that each analysis is simply a one-way within-subjects ANOVA, **and that we use a separate error term for each analysis.** We did these analyses in the previous section, so I will just reprint the summaries here:

```
> summary(dat.young.aov,multivariate=F)
```

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
               SS num Df Error SS den Df       F  Pr(>F)
(Intercept) 44204      1       80      5 2747.46 4.8e-08 ***
angle         721      2      371     10    9.71  0.0045 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

      Test statistic p-value
angle          0.429   0.184


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle  0.636      0.016 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      HF eps Pr(>F[HF])
angle  0.756      0.011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(dat.old.aov,multivariate=F)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
               SS num Df Error SS den Df      F  Pr(>F)
(Intercept) 52057      1      922      5 282.37 1.4e-05 ***
angle          75      2      772     10   0.49    0.63
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Mauchly Tests for Sphericity

      Test statistic p-value
angle          0.145   0.021


Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

      GG eps Pr(>F[GG])
angle  0.539       0.53

      HF eps Pr(>F[HF])
angle   0.57       0.54
```

The simple main effect of angle is significant in the young group, even if we use adjusted p values. Our analysis of the simple main effect of the within-subject variable used individual error terms, but it is possible to use a pooled error term (i.e., the within-subject error term from the overall analysis). This option is the one taken by SPSS, and provides greater power. However, it depends on the assumption that the error variance is the same at each level of the between-subjects variable.

12.2.2 linear comparisons (between-subjects variable)

Previously it was shown that the evaluation of the between-subject effect in a split-plot design is the same as performing a standard between-subject ANOVA on the average of the dependent variables. We can take advantage of this fact when conducting linear comparisons among levels on the between-subjects variable. First, we compute the average score for each subject by averaging across all levels of the within-subject variable. These mean scores can then be analyzed with a standard one-way between-subjects ANOVA. Furthermore, the standard procedures for conducting linear comparisons among levels on between-subjects factors can be used without modification.

12.2.3 linear comparisons (within-subjects variable)

Linear comparisons among levels of the within-subjects variable are done using procedures that are similar to the ones used in standard within-subjects design[2]. First, we construct a set of appropriate contrast weights. Next, we use the weights to create a composite score for each subject. In the following R code, I save the composite scores in the data frame `myData`, but this step is not necessary:

```
> y.mat<-as.matrix( myData[,2:4] )
> lin.C <- c(-1,0,1)
> myData$lin.scores <- y.mat %*% lin.C
> myData
```

[2]The procedure that I am describing is taken directly from Keppel and Wickens (2004, p. 453-456).

```
   group a1 a2 a3 lin.scores
1  young 50 47 51          1
2  young 41 57 43          2
3  young 42 63 40         -2
4  young 46 66 47          1
5  young 45 61 38         -7
6  young 45 57 53          8
7    old 48 39 38        -10
8    old 55 72 54         -1
9    old 51 44 51          0
10   old 53 65 53          0
11   old 68 58 62         -6
12   old 65 37 55        -10
```

To determine if the value of the contrast differs across groups – i.e., to see if there is a contrast x group interaction – I will submit the composite scores to a between-groups t test:

```
> t.test(lin.scores~group,data=myData)

        Welch Two Sample t-test
data:  lin.scores by group
t = 1.779, df = 9.994, p-value = 0.1056
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.263 11.263
sample estimates:
mean in group young   mean in group old
                0.5                -4.5
```

In this case, the test is not significant and therefore I do not reject the null hypothesis that the contrast/composite scores do not differ between groups. To test if the contrast is significant, I do a t test on all of the scores to see if they differ from zero. Note that I have to tell R that the variable `lin.scores` is in `myData`:

```
> t.test(myData$lin.scores)

        One Sample t-test
data:  myData$lin.scores
t = -1.301, df = 11, p-value = 0.2199
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -5.384  1.384
sample estimates:
mean of x
       -2
```

The test is not significant, so I fail to reject the null hypothesis that the contrast scores are zero. It is important to realize that this t test ignores the `group:contrast` interaction: it assumes that the interaction is zero. Suppose we were unwilling to make that assumption: How do we evaluate the contrast while at the same time controlling for a (possibly non-significant) interaction? A simple way of doing this test is to use a one-way between-subjects ANOVA on the composite scores and evaluate the intercept:

```
> lin.scores.aov <- aov(lin.scores~group,data=myData)
> summary(lin.scores.aov,intercept=T)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
(Intercept)  1     48    48.0    2.03   0.19
group        1     75    75.0    3.16   0.11
Residuals   10    237    23.7
```

When we use sum-to-zero coding for our effects, the intercept corresponds to the grand mean of the scores, and the F test evaluates the null hypothesis that the grand mean is zero. Hence, this F test is similar to the t test that we used to determine if the population mean of the composite was zero. There is, however, a subtle difference between the tests: the F test, but not the t test, uses an error term that does *not* include variation associated with the `group:contrast` interaction[3]. In this case, the test is not significant, $F(1, 10) = 2.03$, $p = 0.19$, so we do not reject the null hypothesis. Note, by the way, that the error term in the F test has 10 degrees of freedom, whereas the error term in the t test had 11 degrees of freedom. The difference is due to the inclusion of the `group` factor in the F test but not the t test.

References

Keppel, G. and Wickens, T. (2004). *Design and analysis: A researcher's handbook.* Pearson Education, Inc., 4th edition.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences.* Brooks/Cole, 3rd edition.

[3]One way of proving that this is true is to recalculate F for the intercept using an error term that is based on sum of the variation assigned to the group and residuals terms: the combined sum-of-squares is $75 + 237 = 312$ and the resulting mean square is $312/11 = 28.36$. Using this mean square value, the F for intercept is $48/28.36 = 1.69$. As expected, this new F value equals the squared value of t computed previously ($t^2 = (-1.301)^2 = 1.69$). Hence, this new F test is equivalent to the t test.