Notes on Maxwell & Delaney

PSYCH 710

November 27, 2022

12 higher-order within-subject designs

Previously we 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

First consider a situation in which all combinations of two experimental variables are presented to each subject. This a within-subject factorial design: all of the factors are crossed within every subject. For instance, consider a hypothetical perception experiment that uses a 2×3 within-subject factorial design to examine the effects of stimulus orientation and visual noise (i.e., distractors) 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 once 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

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

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 are defined so they satisfy the sum-to-zero constraint.

The analysis of variance estimates the variation (i.e., sums of squares) associated with each main effect and interaction by calculating the difference between the sum-of-squared residuals for the full model and a model in which the parameters for the effect of interest have been set to zero. Each sum of squares is divided by the relevant number of parameters (i.e., the degrees of freedom) to yield a mean square (Table 1). Inferential statistics are performed by dividing each mean square by an appropriate error term and then using an F test to evaluate the null hypothesis that the effects (i.e., the α's, β's, and $(\alpha\beta)$'s) are zero. For a between-subjects factorial design, the Fs are calculated by dividing all of the mean squares by the mean

square residuals/within. For a within-subjects factorial design, the main effects of A and B and the $A \times B$ interaction are evaluated with different error terms (see Table 1). There is no unbiased F test for any effect that includes *subject*.

Effect	$\mathbf{E}$(Mean Square)	F
S	$\sigma_\epsilon^2 + ab\sigma_\pi^2$	
A	$\sigma_\epsilon^2 + b\sigma_{\alpha\pi}^2 + nb\frac{\sum_j \alpha_j^2}{a-1}$	$\frac{MS_A}{MS_{A \times S}}$
$A \times S$	$\sigma_\epsilon^2 + b\sigma_{\alpha\pi}^2$	
B	$\sigma_\epsilon^2 + a\sigma_{\beta\pi}^2 + na\frac{\sum_k \beta_j^2}{b-1}$	$\frac{MS_B}{MS_{B \times S}}$
$B \times S$	$\sigma_\epsilon^2 + a\sigma_{\beta\pi}^2$	
$A \times B$	$\sigma_\epsilon^2 + \sigma_{\alpha\beta\pi}^2 + n\frac{\sum_j \sum_k (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$	$\frac{MS_{A \times B}}{MS_{A \times B \times S}}$
$A \times B \times S$	$\sigma_\epsilon^2 + \sigma_{\alpha\beta\pi}^2$	

Table 1: Expected Mean Squares and F statistics for a factorial within-subjects design with n subjects (S), a levels of factor A, and b levels of factor B. A and B are assumed to be fixed, and subjects is random.

12.1.1 R example

My example uses the data from Table 12.1 in Maxwell and Delaney (2004), although I modified the numbers to make the F and p values more believable. The data file `rtData-mw-12-1.rda` contains the data in long-format and wide-format data frames.

```
load(url("http://pnb.mcmaster.ca/bennett/psy710/datasets/rtData-mw12-1.rda"))
sapply(rt.wide,class)


##       subj   absent.a0   absent.a4   absent.a8 present.a0 present.a4
##   "factor"   "numeric"   "numeric"   "numeric"  "numeric"  "numeric"
## present.a8
##  "numeric"


summary(rt.long)


##       subj          noise        angle          rt
##   s1     : 6   absent :30   a0:20   Min.   : 284
##   s10    : 6   present:30   a4:20   1st Qu.: 459
##   s2     : 6                a8:20   Median : 571
##   s3     : 6                        Mean   : 588
##   s4     : 6                        3rd Qu.: 686
##   s5     : 6                        Max.   :1000
##   (Other):24
```

First we use `aov`. Note how `angle` and `noise` are listed both inside and outside `Error`.

```
options(width=80,digits=4)
options(contrasts=c("contr.sum","contr.poly"))
# following model is incorrect
rt.aov.00 <- aov(rt ~ 1+Error(subj/(angle*noise)),data=rt.long) # no fixed effects!
# the next 2 models are equivalent:
rt.aov.01 <- aov(rt ~ angle*noise+Error(subj/(angle*noise)),data=rt.long)
# rt.aov.01b <- aov(rt ~ angle*noise+Error(subj/(angle+noise+angle:noise)),data=rt.long)
summary(rt.aov.01)


##
## Error: subj
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  9 344423   38269
##
## Error: subj:angle
##           Df Sum Sq Mean Sq F value Pr(>F)
## angle      2 225422  112711    7.79 0.0036 **
## Residuals 18 260280   14460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subj:noise
##           Df Sum Sq Mean Sq F value Pr(>F)
## noise      1 234625  234625    15.7 0.0033 **
## Residuals  9 134551   14950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subj:angle:noise
##             Df Sum Sq Mean Sq F value Pr(>F)
## angle:noise  2 187983   93992    7.51 0.0043 **
## Residuals   18 225402   12522
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table contains F tests for the main effects and interaction that use the correct error terms. However, the p value for each F test was calculated assuming the variance-covariance matrix for the residuals was spherical. Deviations from sphericity can inflate the Type I error rate for tests of effects with more than 1 degree of freedom, and therefore it is important to check the assumption for the main effect of `angle` and the `angle` $\times$ `noise` interaction.

The following code performs the analysis with `aov_car` in the `afex` package:

```
library(afex)
rt.aov.02 <-aov_car(rt ~ angle*noise+Error(subj/(angle*noise)),data=rt.long)
summary(rt.aov.02)


##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##              Sum Sq num Df Error SS den Df F value   Pr(>F)
## (Intercept) 20765813      1   344423      9  542.62 2.4e-09 ***
## angle         225423      2   260280     18    7.79  0.0036 **
## noise         234625      1   134551      9   15.69  0.0033 **
## angle:noise   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
## angle:noise           0.639   0.167
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
##  for Departure from Sphericity
##
##             GG eps Pr(>F[GG])
## angle         0.764      0.0083 **
## angle:noise   0.735      0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             HF eps Pr(>F[HF])
## angle         0.8880   0.005369
## angle:noise 0.8431    0.007177
```

The first part of the summary contains an ANOVA table that assumes sphericity. 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 **sphericity assumption** implies that all pairwise difference scores among the levels of the within-subject variable have the same variance, and so the assumption *must* be true when the variable has 1 degree of freedom. Hence, we need not check the sphericity assumption for `noise` because it has only two levels and therefore one degree of freedom. The Mauchly tests for sphericity are not significant ($p > 0.15$) in the other two 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 & effect size

Maxwell and Delaney (2004) provide the following 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:

$$\hat{\omega}^2 = \frac{df_{effect}(MS_{effect} - MS_{effect \times S})}{SS_{effect} + SS_{effect \times S} + SS_S + MS_S} \tag{2}$$

Kirk (1995) showed that Cohen's f can be calculated from $\hat{\omega}^2$ with the following equation:

$$\hat{f} = \sqrt{\frac{\hat{\omega}^2_{effect}}{1 - \hat{\omega}^2_{effect}}} \tag{3}$$

The following code calculates association strength and effect size Equations 2 and 3.

```r
omega.sq.angle <- 2*(112711-14460) / (225422 + 260280 + 344423 + 38269)
omega.sq.noise <- 1*(234624-14950) / (234624 + 134551 + 344423 + 38269)
omega.sq.a_x_n <- 2*(93992-12522) / (187983 + 225402 + 344423 + 38269)
c(omega.sq.angle , omega.sq.noise, omega.sq.a_x_n)
```

```
## [1] 0.2263 0.2922 0.2047
```

```r
cohens.f.angle <- sqrt(omega.sq.angle/(1-omega.sq.angle) )
cohens.f.noise <- sqrt(omega.sq.noise/(1-omega.sq.noise) )
cohens.f.noise <- sqrt(omega.sq.a_x_n/(1-omega.sq.a_x_n) )
c(cohens.f.angle , cohens.f.noise, omega.sq.a_x_n)
```

```
## [1] 0.5408 0.5073 0.2047
```

Association strength and effect size also can be calculated using commands in the `effectsize` package.

```r
library(effectsize)
omega_squared(rt.aov.02)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter   | Omega2 (partial) |       95% CI
## -------------------------------------------
## angle       |             0.23 | [0.00, 1.00]
## noise       |             0.29 | [0.00, 1.00]
## angle:noise |             0.20 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

However, note that Cohen's f is calculated by using values of eta-squared, not omega-squared, in Eq 3:

```r
eta_squared(rt.aov.02)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter   | Eta2 (partial) |       95% CI
## -------------------------------------------
## angle       |           0.46 | [0.14, 1.00]
## noise       |           0.64 | [0.23, 1.00]
## angle:noise |           0.45 | [0.13, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```r
cohens_f(rt.aov.02)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter   | Cohen's f (partial) |       95% CI
## -------------------------------------------
## angle       |                0.93 | [0.41, Inf]
## noise       |                1.32 | [0.55, Inf]
## angle:noise |                0.91 | [0.39, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```
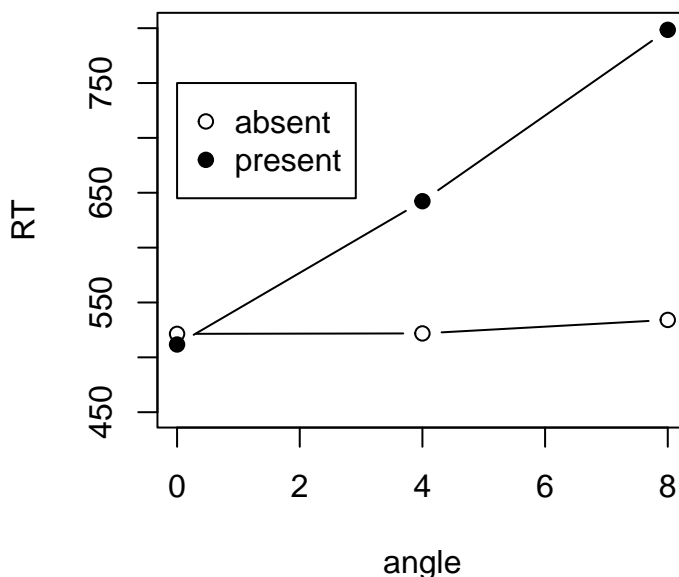
Figure 1: Interaction plot of data in rt.mat.

12.1.3 simple main effects

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

```
( tmp <- with(rt.long,tapply(rt,list(noise,angle),mean)) )


##              a0     a4     a8
## absent   521.4 521.8 534.1
## present 511.7 642.3 798.5


absent.means <- tmp["absent",]
present.means <- tmp["present",]
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. The reason for preferring to use the so-called local error term is that even slight deviations from sphericity can make the F too large for some simple effects and too small for others (Maxwell and Delaney, 2004). Using the local error term means that 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`.

```r
aov.angle.absent <- aov_car(rt~angle+Error(subj/angle),data=subset(rt.long,noise=="absent"))
aov.angle.present <- aov_car(rt~angle+Error(subj/angle),data=subset(rt.long,noise=="present"))
# Hyunh-Feldt correction; pes == partial eta squared:
anova(aov.angle.absent,correction="HF",es="pes")


## Anova Table (Type 3 tests)
##
## Response: rt
##       num Df den Df   MSE    F     pes  Pr(>F)
## angle      2     18 10460 0.05 0.00551   0.95


anova(aov.angle.present,correction="HF",es="pes")


## Anova Table (Type 3 tests)
##
## Response: rt
##       num Df den Df   MSE    F    pes Pr(>F)
## angle      2     18 16522 12.5 0.581  4e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The simple main effect of angle when `noise` is present is significant, $F(2, 18) = 12.5$, $p < .001$, but not when `noise` is absent, $F((2, 18) = 0.05$, $p = 0.95$. Note that the same conclusion would be reached if we used the G-G adjusted p values.

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 $\alpha = .05/2 = .025$. Even with 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` *ignoring* `noise`. Such an analysis is inappropriate in this case because the `angle:noise` interaction is significant, but we continue for illustrative purposes. First I extract the data for the from the wide-format data frame, `rt.wide`, and store the data in a matrix. Then I use the linear trend weights to create linear trend scores for each subject. In what follows, note how my contrast 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 <- as.matrix(rt.wide[,2:7])
rt.mat[1:2,]
```

```
##   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
```

```
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 = 3.5, df = 9, p-value = 0.007
## 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 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 = -5.1, df = 9, p-value = 7e-04
## 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, $t(9) = -5.1$, $p < .001$, so we conclude that the linear trend differs in the noise-present and noise-absent conditions. This result is consistent[1] with our previous ANOVA which found that was a significant `angle` $\times$ `noise`.

---

[1]Although the two analyses yield results that are consistent, each analysis evaluates a different question. The significant `angle` $\times$ `noise` interaction in the overall ANOVA evaluates a *general* hypothesis, namely that the effect of one variable depends

Finally, we can evaluate the linear trend separately at each level of `noise`. The following code applies linear trend weights separately to data in the noise-present and noise-absent conditions in the wide-format data frame, `rt.wide`. The linear trend weights create linear trend scores for each subject, and t tests are used to evaluate the null hypothesis that the linear trend is zero. 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, I reject the null hypothesis of no linear trend in the noise-present condition but not in the noise-absent condition.

```
rt.mat[1,] # inspect column names

##   absent.a0   absent.a4   absent.a8 present.a0 present.a4 present.a8
##         657         446         461        671        474        903


# note the contrast weights in next 2 lines:
rt.absent.lin <- rt.mat %*% c(-1,0,1,0,0,0)
rt.present.lin <- rt.mat %*% c(0,0,0,-1,0,1)
t.test(rt.absent.lin)


##
##      One Sample t-test
##
## data:  rt.absent.lin
## t = 0.32, df = 9, p-value = 0.8
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   -77.93 103.33
## sample estimates:
## mean of x
##      12.7



t.test(rt.present.lin)


##
##      One Sample t-test
##
## data:  rt.present.lin
## t = 4.8, df = 9, p-value = 0.001
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   151.7 421.9
## sample estimates:
## mean of x
##      286.8
```

Previously we used a t test to evaluate the linear trend $\times$ noise interaction, but it is possible to use an F test. The F test is a more general procedure than the t test which could be used in a situation where `noise` had more than two levels, and so I will describe it here. First, I store the linear trend scores for the noise-present and noise-absent conditions in the numeric variable `linTrend`. Then I create the factors `sid` and `nz` and save everything in the data frame `linTrend.df`.

---

on the level of the other variable. The current test evaluates the *specific* hypothesis that the linear trend (across angle) differs across noise levels.

```
N <- length(rt.absent.lin) # number of subjects
linTrend <- c(rt.present.lin,rt.absent.lin)
nz <- as.factor(x=rep(c("present","absent"),each=N))
sid <- factor(x=rep(1:N,times=2),label="s")
linTrend.df <- data.frame(sid,nz,linTrend)
summary(linTrend.df)


##       sid            nz          linTrend
##   s1     :2    absent :10    Min.   :-196
##   s2     :2    present:10    1st Qu.:  19
##   s3     :2                  Median :  104
##   s4     :2                  Mean   :  150
##   s5     :2                  3rd Qu.: 249
##   s6     :2                  Max.   :  575
##   (Other):8
```

Next, I conduct a within-subjects ANOVA on `linTrend`. Note that no Mauchly test was performed: that test is not needed here because the within-subject effect being evaluated has a single degree of freedom.

```
options(contrasts=c("contr.sum","contr.poly"))
linTrend.aov.01 <- aov_car(linTrend~nz+Error(sid/nz),data=linTrend.df)
summary(linTrend.aov.01)


##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##              Sum Sq num Df Error SS den Df F value  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 – are the same at both levels of noise in 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.

Finally, it is worth noting that all of these contrasts can be performed with commands in the `emmeans` package:

```
library(emmeans)
rt.emm <- emmeans(rt.aov.02,specs="angle")
lin.C <- c(-1,0,1) # trend weights

 # linear trend ignoring noise
contrast(rt.emm,method=list(linear=lin.C))


##   contrast estimate SE df t.ratio p.value
##   linear        150 43  9   3.481  0.0069
##
## Results are averaged over the levels of: noise


# linear trend x noise interaction:
rt.emm.2 <- emmeans(rt.aov.02,specs="angle",by="noise")
contrast(rt.emm.2,interaction=list(angle=list(lin.C),noise=list(c(-1,1))),by=NULL)


##   angle_custom noise_custom estimate   SE df t.ratio p.value
##   c(-1, 0, 1)  c(-1, 1)          274 54.3  9   5.051  0.0007


# linear trend separately for each noise
contrast(rt.emm.2,method=list(linear=lin.C))


## noise = absent:
##   contrast estimate   SE df t.ratio p.value
##   linear       12.7 40.1  9   0.317  0.7585
##
## noise = present:
##   contrast estimate   SE df t.ratio p.value
##   linear      286.8 59.7  9   4.801  0.0010
```

12.1.5 alternative analyses with `lmer` & `lme`

In this section I illustrate how to analyze the RT data with `lmer` and `lme`. These routines, which are in the `lme4` and `nlme` packages, respectively, estimate the parameters of linear mixed models using Restricted Maximum Likelihood (i.e., REML, a variant of the maximum likelihood method). Detailed descriptions of the algorithms used by these commands have been published elsewhere (Bates et al., 2015; Pinheiro and Bates, 2000). For our purposes, it is important to note that the methods differ in a fundamental way from those used by ANOVA, and therefore the results obtained by these methods may differ from those obtained by ANOVA. For example, there is debate about the proper way of calculating the degrees of freedom used in F tests of fixed effects (Bates, 2006). Given these issues, why would we consider using these procedures? One reason is that `lmer` and `lme` can be used in situations that are difficult to analyze with ANOVA. For example, mixed model routines work on unbalanced data whereas ANOVA either throws out cases with missing data (e.g., `aov_car`) or fails entirely (e.g., `aov`). In addition, both `lmer` and `lme` provide much greater flexibility for modelling the random components of mixed models.

The following code uses `lmer` to fit a mixed model to the RT data and evaluates the fixed effects with F tests. Note the denominator degrees of freedom, which differ considerably from the values in the ANOVA computed by `aov_car`.

```
# using lmer:
library(lmerTest)
rt.lmer.01 <- lmer(rt~angle*noise+(1|subj),data=rt.long)
```

```
anova(rt.lmer.01) # assumes sphericity


## Type III Analysis of Variance Table with Satterthwaite's method
##             Sum Sq Mean Sq NumDF DenDF F value   Pr(>F)
## angle       225422  112711     2    45    8.18  0.00093 ***
## noise       234575  234575     1    45   17.02  0.00016 ***
## angle:noise 187983   93992     2    45    6.82  0.00259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F tests assume sphericity. We can evaluate the two-way interaction by removing it from the model
and evaluating the change in deviance (i.e., the goodness-of-fit) with a chi-square test. The chi-square test
(which does not assume sphericity) is significant and suggests the interaction should not be removed from
the model. Furthermore, the fact that the interaction is significant indicates that we should not remove the
two main effects (which are marginal to the interaction).

```
rt.lmer.00 <- lmer(rt~angle+noise+(1|subj),data=rt.long)
anova(rt.lmer.00,rt.lmer.01)


## Data: rt.long
## Models:
## rt.lmer.00: rt ~ angle + noise + (1 | subj)
## rt.lmer.01: rt ~ angle * noise + (1 | subj)
##            npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)
## rt.lmer.00    6 771 784   -380      759
## rt.lmer.01    8 762 779   -373      746  13.2  2     0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Alternatively, we could assume that our data are not spherical and evaluate the fixed effects with a conser-
vative F test. The fact that the tests are significant again suggests that the effects should not be removed
from the model.

```
# conservative F tests:
N <- 10 # 10 subjects
F.angle <- 8.18
F.angle.x.noise <- 6.82
# conservative F tests for angle and angle x noise:
1-pf(F.angle,1,N) # p value for angle


## [1] 0.01695


1-pf(F.angle.x.noise,1,N) # p value for angle x noise


## [1] 0.02597
```

Association strength and effect size for the fixed effects can be calculated using commands in **effectsize**.

```
library(effectsize)
omega_squared(rt.lmer.01)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter   | Omega2 (partial) |      95% CI
## ---------------------------------------------
## angle       |             0.23 | [0.06, 1.00]
## noise       |             0.25 | [0.09, 1.00]
## angle:noise |             0.20 | [0.04, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].


cohens_f(rt.lmer.01)


## # Effect Size for ANOVA (Type III)
##
## Parameter   | Cohen's f (partial) |      95% CI
## -----------------------------------------------
## angle       |                0.60 | [0.31, Inf]
## noise       |                0.61 | [0.34, Inf]
## angle:noise |                0.55 | [0.25, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

Finally, the random effect of subject can be evaluated with `ranova` and the variance component can be calculated with `VarCorr`:

```
# random components
ranova(rt.lmer.01) # anova-like table


## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## rt ~ angle + noise + (1 | subj) + angle:noise
##          npar logLik AIC  LRT Df Pr(>Chisq)
## <none>      8   -350 715
## (1 | subj)  7   -352 718 4.81  1     0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# variance components
print(VarCorr(rt.lmer.01),comp=c("Variance","Std.Dev."))


##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept)  4081     63.9
##  Residual             13783    117.4


# association strength
library(performance)
icc(rt.lmer.01)


## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.228
##   Unadjusted ICC: 0.141
```

Here we analyze the RT data using `lme` in the `nlme`. First, we create two mixed models with identical fixed and random effects, but differ in terms of the assumed correlations among the within-subject measures. The first model assumes that the within-subject measures are independent. The second model assumes compound symmetry: residuals in the levels of the within subject factors are assumed to have the same variance and the same covariance. The third model does *not* assume compound symmetry: residual variances, but not covariances, are assumed to be constant across levels of within-subject variables. Another way of stating the difference is that the first model assumes that the correlations between within-subject measures is zero, the second model assumes that the correlations are non-zero but constant, and the third model assumes that the correlations vary.

```
# using lme in nlme
library(nlme)
# assumes independence:
rt.lme.01 <- lme(rt~angle*noise,random=~1|subj,
                data=rt.long)
# assumes sphericity:
rt.lme.02 <- lme(rt~angle*noise,random=~1|subj,
                data=rt.long,
                correlation=corCompSymm(value=0.3,form=~1|subj))
# does not assume sphericity:
rt.lme.03 <- lme(rt~angle*noise,random=~1|subj,
                data=rt.long,
                correlation=corSymm(form=~1|subj))
# no significant difference in fit:
anova(rt.lme.01,rt.lme.02,rt.lme.03)


##           Model df    AIC    BIC  logLik   Test L.Ratio p-value
## rt.lme.01     1  8 715.5 731.4 -349.7
## rt.lme.02     2  9 717.5 735.4 -349.7 1 vs 2    0.00  0.9999
## rt.lme.03     3 23 728.3 774.0 -341.1 2 vs 3   17.23  0.2444
```

Goodness-of-fit does not differ significantly across models, and therefore we can use the simpler models. Note that there is *no* difference between the first and second models, even though the second model uses one more degree of freedom (to estimate the correlation between within-subject measures). Next, we evaluate the fixed effects and calculate association strength and effect size for the first model. Identical results (not shown) are obtained from the second model

```
# fixed effects for model 1
anova(rt.lme.01)


##             numDF denDF F-value p-value
## (Intercept)     1    45   542.6  <.0001
## angle           2    45     8.2  0.0009
## noise           1    45    17.0  0.0002
## angle:noise     2    45     6.8  0.0026


omega_squared(rt.lme.01)


## # Effect Size for ANOVA
##
## Parameter   | Omega2 (partial) |       95% CI
## ---------------------------------------------
## angle       |             0.23 | [0.06, 1.00]
## noise       |             0.25 | [0.09, 1.00]
## angle:noise |             0.20 | [0.04, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
cohens_f(rt.lme.01)


## # Effect Size for ANOVA
##
## Parameter     | Cohen's f (partial) |      95% CI
## ------------------------------------------------
## angle         |                0.60 | [0.31, Inf]
## noise         |                0.62 | [0.34, Inf]
## angle:noise   |                0.55 | [0.25, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

Finally, we estimate the variance components. Here we *do* see a difference between the model that assumed independence and the model that assumed compound symmetry. So allowing the within-subject variables to be correlated did not change our estimates of the fixed effects, but did change the magnitude of the random effect (i.e., `subj`).

```
# variance components (independence)
print(VarCorr(rt.lme.01),comp=c("Variance","Std.Dev."))


## subj = pdLogChol(1)
##             Variance StdDev
## (Intercept)   4081     63.89
## Residual     13783    117.40


( icc.subj <- 4081/(4081+13783) )


## [1] 0.2284


# variance components (compound symmetry)
print(VarCorr(rt.lme.02),comp=c("Variance","Std.Dev."))


## subj = pdLogChol(1)
##             Variance StdDev
## (Intercept)   1142     33.79
## Residual     16722    129.31


( icc.subj <- 1142/(1142+16722) )


## [1] 0.06393
```

12.2   Split plot designs

Some experiments use mixtures of within- and between-subjects factors. Such designs often are called split plot designs. Maxwell and Delaney (2004) illustrate a split-plot ANOVA with data from a hypothetical experiment that measured reaction time in groups of younger and older adults. 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:

```
load(url("http://pnb.mcmaster.ca/bennett/psy710/datasets/rtVisual.rda"))
sapply(rtVisual.wide,class)


##      subj      group         a1         a2         a3
##   "factor"  "factor" "numeric" "numeric" "numeric"


summary(rtVisual)


##        subj        group      angle         rt
##   s1     : 3   young:18    a1:12   Min.   :370
##   s2     : 3   old  :18    a2:12   1st Qu.:448
##   s3     : 3               a3:12   Median :510
##   s4     : 3                       Mean   :517
##   s5     : 3                       3rd Qu.:572
##   s6     : 3                       Max.   :720
##   (Other):18
```

The file contains two data frames: `rtVisual` and `rtVisual.wide`. Both data frames contain the factors `subj` and `group`. The long-format data frame `rtVisual` contains the factor `angle` and the numeric variable `rt`. Note that `group` and `angle` are between- and within-subjects variables, respectively. The wide-format data frame, `rtVisual.wide`, stores the dependent variable in three numeric variable, `a1`, `a2`, and `a3`. In the next block of code, I use the long-format data frame to evaluate the data with a split-plot ANOVA:

```
rtAge.aov.01 <- aov(rt ~ group*angle + Error(subj/angle),data=rtVisual)
summary(rtAge.aov.01)


##
## Error: subj
##            Df Sum Sq Mean Sq F value Pr(>F)
## group       1  16044   16044     1.6   0.23
## Residuals  10 100222   10022
##
## Error: subj:angle
##             Df Sum Sq Mean Sq F value Pr(>F)
## angle        2  28850   14425    2.52  0.105
## group:angle  2  50772   25386    4.44  0.025 *
## Residuals   20 114311    5716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table shows the F and p values for the main effects of `group` and `angle`, and the `group:angle` interaction. Examination of the table shows that the effects of `group` and `angle` are evaluated with different error terms. The `aov` command assumes that the between- and within-subjects variables are *fixed*, and uses the appropriate error term to generate unbiased F tests. However, the p values were calculated assuming that the variance-covariance matrix of the within-subject variables are spherical. The following command adjusts the p values for deviations from sphericity:

```
library(afex)
options(contrasts=c("contr.sum","contr.poly"))
rtAge.aov.02 <- aov_car(rt ~ group*angle + Error(subj/angle),data=rtVisual)
summary(rtAge.aov.02)


##
```

```
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##              Sum Sq num Df Error SS den Df F value  Pr(>F)
## (Intercept) 9610000      1   100222     10  958.87 2.9e-11 ***
## group         16044      1   100222     10    1.60   0.234
## angle         28850      2   114311     20    2.52   0.105
## group:angle   50772      2   114311     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       0.6468    0.13155
## group:angle 0.6468    0.04726
```

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:

```
rt.mat <- rtVisual.wide[,3:5] # extract rt measures
rtVisual.wide$rtAvg <- rowMeans(rt.mat) # calculate average
summary(aov(rtAvg~group,data=rtVisual.wide))


##             Df Sum Sq Mean Sq F value Pr(>F)
## group        1   5348    5348     1.6   0.23
## Residuals   10  33407    3341
```

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 different numbers of data points are analyzed in the two cases, but the ratios 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. As we learned in previous sections, in the case where both variables are within-subject factors each effect is evaluated with an error term that corresponds to the interaction of that effect with subjects: A is evaluated with $A \times S$, B is evaluated with $B \times S$, and $A \times B$ is 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 $\mathrm{MS}_{B \times S}$ 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:

```
rt.young <- aov(rt~angle+Error(subj/angle),data=subset(rtVisual,group=="young"))
rt.old <- aov(rt~angle+Error(subj/angle),data=subset(rtVisual,group=="old"))
```

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

```
summary(rt.young)[[2]]


##            Df Sum Sq Mean Sq F value Pr(>F)
## angle       2  72078   36039    9.71 0.0045 **
## Residuals  10  37122    3712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



summary(rt.old)[[2]]


##            Df Sum Sq Mean Sq F value Pr(>F)
## angle       2   7544    3772    0.49   0.63
## Residuals  10  77189    7719
```

Notice that the sum of squares for the error terms in the two analyses are 37122 and 77189, and that the sum of these values, 114311, 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, $(3712 + 7719)/2 = 5716$, 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 the average of the error terms in a series of within-subjects ANOVAs.

Finally, we can use the `effectsize` package to calculate association strength and effect size for the fixed effects.

```
library(effectsize)
# eta_squared(rtAge.aov.02)
omega_squared(rtAge.aov.02)


## # Effect Size for ANOVA (Type III)
##
## Parameter    | Omega2 (partial) |       95% CI
## -------------------------------------------------
## group        |             0.05 | [0.00, 1.00]
## angle        |             0.07 | [0.00, 1.00]
## group:angle  |             0.14 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
cohens_f(rtAge.aov.02)


## # Effect Size for ANOVA (Type III)
##
## Parameter   | Cohen's f (partial) |      95% CI
## ------------------------------------------------
## group       |                0.40 | [0.00, Inf]
## angle       |                0.50 | [0.00, Inf]
## group:angle |                0.67 | [0.17, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

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 being examined, and therefore is equivalent to a one-way between-subjects ANOVA.

```
levels(rtVisual$angle)


## [1] "a1" "a2" "a3"


summary(aov(rt~1+group,data=subset(rtVisual,angle=="a1")))


##             Df Sum Sq Mean Sq F value Pr(>F)
## group        1  42008   42008    11.3 0.0072 **
## Residuals   10  37217    3722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


summary(aov(rt~1+group,data=subset(rtVisual,angle=="a2")))


##             Df Sum Sq Mean Sq F value Pr(>F)
## group        1  10800   10800    0.84   0.38
## Residuals   10 128100   12810


summary(aov(rt~1+group,data=subset(rtVisual,angle=="a3")))


##             Df Sum Sq Mean Sq F value Pr(>F)
## group        1  14008   14008    2.85   0.12
## Residuals   10  49217    4922
```

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**:

```
anova(aov_car(rt~angle+Error(subj/angle),
              correction="HF",
              es="pes",
              data=subset(rtVisual,group=="young")))


## Anova Table (Type 3 tests)
##
## Response: rt
##         num Df den Df  MSE    F    ges Pr(>F)
## angle   1.27   6.36 5833 9.71 0.615  0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


anova(aov_car(rt~angle+Error(subj/angle),
              correction="HF",
              es="pes",
              data=subset(rtVisual,group=="old")))


## Anova Table (Type 3 tests)
##
## Response: rt
##         num Df den Df   MSE    F    ges Pr(>F)
## angle   1.08   5.39 14318 0.49 0.0426   0.53
```

The simple main effect of angle is significant in the young group, even if we use adjusted p values. Our analyses of the simple main effects 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 software packages, and it 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. The current example has only two levels on the between-subject variable, so a linear contrast is not particularly interesting. Nevertheless, the following code shows how such a comparison is done with **emmeans**:

```
library(emmeans)
rtAge.emm <- emmeans(rtAge.aov.02,specs="group")
contrast(rtAge.emm,method="pairwise") # ignores angle


##  contrast     estimate   SE df t.ratio p.value
##  young - old    -42.2 33.4 10  -1.265  0.2345
##
## Results are averaged over the levels of: angle


# compare age at each angle:
```

```
rtAge.emm.2 <- emmeans(rtAge.aov.02,specs="group",by="angle")
contrast(rtAge.emm.2,method="pairwise")


## angle = a1:
##  contrast     estimate   SE df t.ratio p.value
##  young - old    -118.3 35.2 10  -3.360  0.0072
##
## angle = a2:
##  contrast     estimate   SE df t.ratio p.value
##  young - old      60.0 65.3 10   0.918  0.3801
##
## angle = a3:
##  contrast     estimate   SE df t.ratio p.value
##  young - old     -68.3 40.5 10  -1.687  0.1225
```

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:

```
sapply(rtVisual.wide,class)


##      subj      group        a1        a2        a3      rtAvg
##  "factor"   "factor" "numeric" "numeric" "numeric" "numeric"


rt.mat<-as.matrix(rtVisual.wide[,3:5] )
lin.C <- c(-1,0,1)
rtVisual.wide$linTrend <- rt.mat %*% lin.C
rtVisual.wide[1:3,]


##   subj group  a1  a2  a3 rtAvg linTrend
## 1   s1 young 500 470 510 493.3       10
## 2   s2 young 410 570 430 470.0       20
## 3   s3 young 420 630 400 483.3      -20
```

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

```
t.test(linTrend~group,data=rtVisual.wide)


##
##      Welch Two Sample t-test
##
## data:  linTrend by group
## t = 1.8, df = 10, p-value = 0.1
## alternative hypothesis: true difference in means between group young and group old is not
    equal to 0
## 95 percent confidence interval:
##   -12.63 112.63
```

---

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

```
## sample estimates:
## mean in group young    mean in group old
##                   5                  -45
```

In this case, the test is not significant and therefore I do not reject the null hypothesis that the composite scores do not differ between groups. To test if the contrast is significant ignoring **group**, I do a t test on all of the scores to see if they differ from zero.

```
t.test(rtVisual.wide$linTrend)
```

```
##
##      One Sample t-test
##
## data:  rtVisual.wide$linTrend
## t = -1.3, df = 11, p-value = 0.2
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -53.84   13.84
## sample estimates:
## mean of x
##       -20
```

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 assumes that the **group** $\times$ **contrast** interaction is zero. Suppose we are 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:

```
options(contrasts=c("contr.sum","contr.poly"))
lin.scores.aov <- aov(linTrend~group,data=rtVisual.wide)
summary(lin.scores.aov,intercept=T)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## (Intercept)   1   4800    4800    2.03   0.19
## group         1   7500    7500    3.16   0.11
## Residuals    10  23700    2370
```

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 that the linear trend is zero. 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. The following code illustrates how to repeat these linear contrasts using **emmeans**:

---

[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 $7500 + 23700 = 31200$ and the resulting mean square is $31200/11 = 2836$. Using this mean square value, the F for intercept is $4800/2836 = 1.69$. 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.

```
library(emmeans)
# create aov_car object:
rtAge.aov.10 <- aov_car(rt ~ group*angle + Error(subj/angle),data=rtVisual)
linTrendWeights <- c(-1,0,1)

# linear trend ignoring group
angle.emm <- emmeans(rtAge.aov.10,specs="angle")
contrast(angle.emm,method=list(linTrendWeights))


##   contrast     estimate   SE df t.ratio p.value
##  c(-1, 0, 1)       -20 14.1 10  -1.423  0.1851
##
## Results are averaged over the levels of: group


# linear trend for each group
angle.emm.2 <- emmeans(rtAge.aov.10,specs="angle",by="group")
contrast(angle.emm.2,method=list(linTrendWeights))


## group = young:
##  contrast     estimate   SE df t.ratio p.value
##  c(-1, 0, 1)        5 19.9 10   0.252  0.8065
##
## group = old:
##  contrast     estimate   SE df t.ratio p.value
##  c(-1, 0, 1)      -45 19.9 10  -2.264  0.0470


# linear trend x group interaction
contrast(angle.emm.2,interaction=c("poly","consec"),by=NULL)


##  angle_poly group_consec estimate    SE df t.ratio p.value
##  linear     old - young       -50  28.1 10  -1.779  0.1056
##  quadratic  old - young       307 143.2 10   2.142  0.0578
```

12.2.4 split plot analysis with `lmer`

In this section we analyze the reaction time data first with `lmer` and then with `lme`. First we analyze the fixed effects:

```
# fixed effects:
library(lmerTest)
library(effectsize)
rtAge.lmer.01 <- lmer(rt ~ group*angle + (1|subj),data=rtVisual)
anova(rtAge.lmer.01) # assumes sphericity


## Type III Analysis of Variance Table with Satterthwaite's method
##             Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## group         9153    9153     1    10    1.60  0.234
## angle        28850   14425     2    20    2.52  0.105
## group:angle  50772   25386     2    20    4.44  0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
omega_squared(rtAge.lmer.01)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter    | Omega2 (partial) |        95% CI
## ----------------------------------------------
## group        |             0.05 | [0.00, 1.00]
## angle        |             0.12 | [0.00, 1.00]
## group:angle  |             0.23 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
cohens_f(rtAge.lmer.01)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter    | Cohen's f (partial) |       95% CI
## ------------------------------------------------
## group        |                0.40 | [0.00, Inf]
## angle        |                0.50 | [0.00, Inf]
## group:angle  |                0.67 | [0.17, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

Next we analyze the random effect of subjects:

```
ranova(rtAge.lmer.01)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## rt ~ group + angle + (1 | subj) + group:angle
##            npar logLik AIC  LRT Df Pr(>Chisq)
## <none>        8   -185 386
## (1 | subj)    7   -185 385 1.11  1       0.29
```

```
print(VarCorr(rtAge.lmer.01),comp=c("Variance","Std.Dev."))
```

```
##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept) 1436     37.9
##  Residual             5716     75.6
```

```
library(performance)
icc(rtAge.lmer.01) # intraclass correlation
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.201
##   Unadjusted ICC: 0.145
```

Here we repeat the analysis with `lme` with a model that uses a compound symmetric structure for the within-subject residuals. First we analyze the fixed effects:

```r
library(nlme)
rtAge.lme.01 <- lme(rt ~ group*angle,
                    random=~1|subj,
                    data=rtVisual,
                    correlation=corCompSymm(value=0.3,form=~1|subj))
# fixed effects:
anova(rtAge.lme.01)


##             numDF denDF F-value p-value
## (Intercept)     1    20   958.9  <.0001
## group           1    10     1.6  0.2345
## angle           2    20     2.5  0.1053
## group:angle     2    20     4.4  0.0253


omega_squared(rtAge.lme.01)


## # Effect Size for ANOVA
##
## Parameter   | Omega2 (partial) |       95% CI
## ---------------------------------------------
## group       |             0.05 | [0.00, 1.00]
## angle       |             0.12 | [0.00, 1.00]
## group:angle |             0.23 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].


cohens_f(rtAge.lme.01)


## # Effect Size for ANOVA
##
## Parameter   | Cohen's f (partial) |      95% CI
## -----------------------------------------------
## group       |                0.40 | [0.00, Inf]
## angle       |                0.50 | [0.00, Inf]
## group:angle |                0.67 | [0.17, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

And now we examine the random effect of subjects:

```r
print(VarCorr(rtAge.lme.01),comp=c("Variance","Std.Dev."))


## subj = pdLogChol(1)
##             Variance StdDev
## (Intercept)  554     23.54
## Residual     6597    81.22


554/(554+6597) # small intraclass correlation


## [1] 0.07747
```

The fixed-effects results from `lmer` and `lme` are very similar, but the random results differ slightly. In particular, the variance component for subjects is smaller in the `lme` analysis. This difference is due to the fact that `lmer` assumes that the residuals on the within-subject factor are independent, whereas the model that we used with `lme` allowed them to be correlated. If we forced `lme` to assume independence, then the variance components are the same as those obtained with `lmer`:

```
## subj = pdLogChol(1)
##             Variance StdDev
## (Intercept) 1436       37.89
## Residual     5716       75.60
```

Finally, we examine the residuals of our models to check the normality assumption. The shapiro tests are not significant, and the qq plots looks fine (Figure 2).

```
shapiro.test(residuals(rtAge.lmer.01))
```

```
##
##      Shapiro-Wilk normality test
##
## data:  residuals(rtAge.lmer.01)
## W = 0.99, p-value = 1
```

```
shapiro.test(residuals(rtAge.lme.01))
```

```
##
##      Shapiro-Wilk normality test
##
## data:  residuals(rtAge.lme.01)
## W = 0.99, p-value = 1
```

LMER model

LME model 01

Sample Quantiles

Sample Quantiles

Theoretical Quantiles
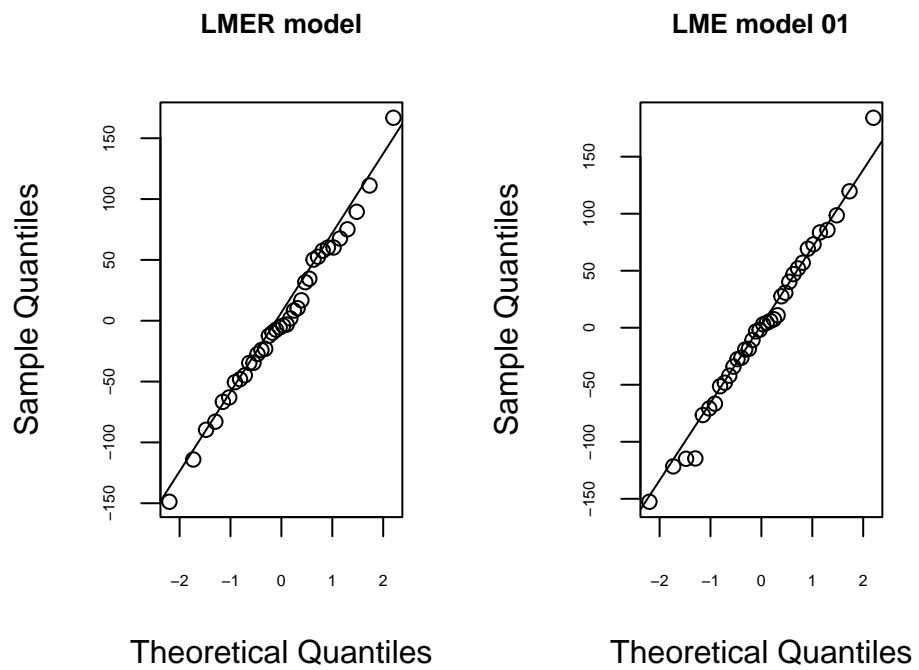
Theoretical Quantiles

Figure 2: Residuals from split-plot mixed model analyses.

References

Bates, D. (2006). lmer, p-values and all that. `stat.ethz.ch/pipermail/r-help/2006-May/094765.html`.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

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.

Maxwell, S. E. and Delaney, H. D. (2004). *Designing experiments and analyzing data: a model comparison perspective.* Lawrence Erlbaum Associates, Mahwah, N.J., 2nd ed edition.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Springer, New York.