

Mixed ANOVA: Two-way, Graphing & Follow ups

Contents

1	Mixed Design Factors	2
1.1	Plot of Simulation	2
1.1.1	Spaghetti Plot	2
1.1.2	Means plot	2
1.2	2-Way Mixed ANOVA logic	3
1.2.1	Explained Terms	4
1.2.2	Unexplained Terms	4
1.3	R ANOVA Calculations	4
1.3.1	R Defaults	4
1.3.2	Assumptions	4
1.3.2.1	Sphericity	4
1.3.2.2	Covariance Matrix	5
1.3.2.3	Box's M	5
1.3.2.4	HOV	6
2	Follow Up Testing	7
2.1	Degree of the problem?	7
2.2	Recommendations	7
2.2.1	Example of Follow up (assuming no violations)	8
2.2.1.1	Follow Up Interaction	8
3	References	8

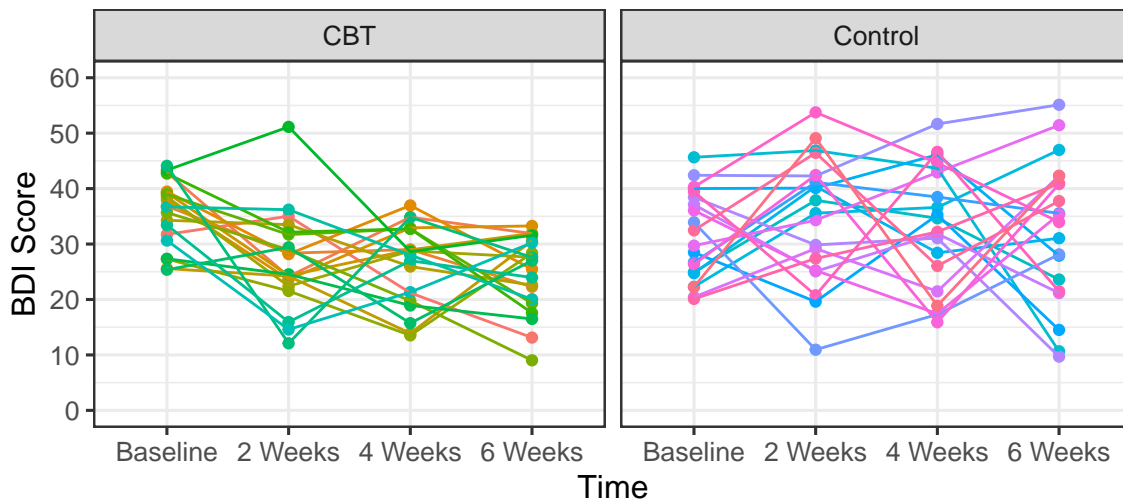
1 Mixed Design Factors

You want to show the effectiveness of CBT therapy against no therapy in reducing depression scores. You give clients (and controls) the Beck depression index (BDI at baseline, and every two weeks afterward for up to 6 Weeks. You assign 30 people to each between-subjects group (control or CBT) with varying levels of depression (from mild to severe; BDI of 20 to 50), and in total, you will collect 4 BDI scores for each person (within subject). You expect BDI scores to drop by the second BDI measurement and continue to decrease over sessions relative to control which should show no change. Keep in mind there is no intervention for the control group, and BDI score has a tendency to be unstable, so you are worried about assumption violations.

1.1 Plot of Simulation

1.1.1 Spaghetti Plot

Each line represents a person's score across conditions



1.1.2 Means plot

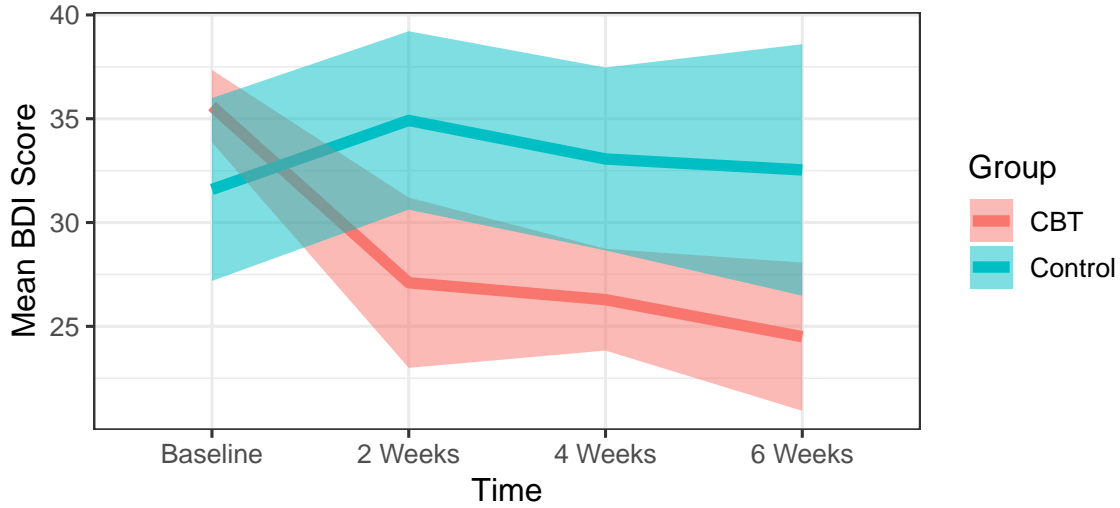
Just like RM ANOVA, we will use Cousineau-Morey Corrected error bars which remove individual differences. The function will correct for the mixed design differences in the correction, just pass which variables are between below. Again, you can select between SE and CI (I plotted CI below).

```
setwd("C:/AlexFiles/SugerSync/UIC/Teaching/Graduate/543-Fall2018/Week 11")
source('HelperFunctions.R')
```

```
MeansforGGPlot <- summarySEwithin(data=DataSim2, "BDI",
                                   withinvars=c("Time"),
                                   betweenvars=c("Group"),
                                   idvar="ID", na.rm=FALSE, conf.interval=.95)
```

```
Means.Within.CI <- ggplot(data = MeansforGGPlot, aes(x = Time, y=BDI, group=Group))+
  scale_y_continuous(breaks = seq(0,40,5))+
  geom_line(aes(colour = Group), size=2)+
  geom_ribbon(aes(ymin=BDI-ci, ymax=BDI+ci, fill = Group), alpha=.5)+
  ylab("Mean BDI Score")+xlab("Time")+
```

```
theme(legend.position = "right")
Means.Within.CI
```



For within error bars (Corrected) you can eye-ball significance (Cumming & Finch, 2005)

95% CIs:

- $p < .05$: when proportion overlap is about .50
- $p < .01$: when proportion overlap is about 0

Notice also you can see problems with the variance

1.2 2-Way Mixed ANOVA logic

Two-way Mixed ANOVA Table With the formulas (conceptually simplified)

Source	SS	DF	MS	F
<i>Between Subject</i>	$CSS_{row\ means}$	$nK - 1$	$\frac{SS_{BS}}{df_{BS}}$	
<i>Group</i>	$CnSS_{Group\ means}$	$K - 1$	$\frac{SS_G}{df_G}$	$\frac{MS_G}{MS_{WG}}$
W_G	$SS_{BS} - SS_G$	$K(n - 1)$	$\frac{SS_{WG}}{df_{WG}}$	
<i>Within Subject</i>	$SS_T - SS_{BS}$	$nK(C - 1)$	$\frac{SS_w}{df_w}$	
<i>RM</i>	$KnSS_{RM\ means}$	$C - 1$	$\frac{SS_{RM}}{df_{RM}}$	$\frac{MS_{RM}}{MS_{SxRM}}$
<i>GxRM</i>	$nSS_{Col\ means} - SS_G - SS_{RM}$	$(K - 1)(C - 1)$	$\frac{SS_{GxRM}}{df_{GxRM}}$	$\frac{MS_{GxRM}}{MS_{SxRM}}$
<i>Residual [Sub x RM]</i>	$SS_{WS} - SS_{RM} - SS_{GxRM}$	$K(n - 1)(C - 1)$	$\frac{SS_{SxRM}}{df_{SxRM}}$	
Total	SS_{scores}	$N - 1$		

Notes:

- $SS_T = SS_{BS} + SS_{WS}$
 - $SS_{BS} = SS_G + SS_{WG}$
 - $SS_{WS} = SS_{RM} + SS_{GxRM} + SS_{SxRM}$
- $df_T = df_{BS} + df_{WS}$
 - $df_{BS} = df_G + df_{WG}$
 - $df_{WS} = df_{RM} + df_{GxRM} + df_{SxRM}$

1.2.1 Explained Terms

- SS_G = caused by group
- SS_{RM} = caused by treatment
- $SS_{G \times RM}$ = caused by group x treatment

1.2.2 Unexplained Terms

- SS_{W_G} = Noise due to group
- $SS_{S \times RM}$ = Noise due to treatment

1.3 R ANOVA Calculations

1.3.1 R Defaults

- `afex` calculates the RM ANOVA for us and it will automatically correct for sphericity violations.
- `Error(Subject ID/RM Factor)`: This tells the `afex` that subjects vary as function RM conditions.
- It reports η_g^2 assuming a manipulated group and treatment.

```
library(afex)
Mixed.1<-aov_car(BDI~ Time*Group + Error(ID/Time),
  data = DataSim2)
Mixed.1

## Anova Table (Type 3 tests)
##
## Response: BDI
##          Effect             df      MSE      F ges p.value
## 1          Group             1, 38 115.22  7.52 ** .06      .009
## 2          Time  2.38, 90.46  95.35   2.52 + .04      .08
## 3 Group:Time  2.38, 90.46  95.35   4.45 * .07      .01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
##
## Sphericity correction method: GG
```

1.3.2 Assumptions

1.3.2.1 Sphericity

Same as RM ANOVA, except we should examine Mauchly's test for Sphericity on each term (that has more than two levels)

```
summary(Mixed.1, return="univariate")

##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##          Sum Sq num Df Error SS den Df    F value    Pr(>F)
## (Intercept) 150794      1  4378.3    38 1308.7751 < 2.2e-16 ***
## Group        866      1  4378.3    38    7.5159  0.009271 **
```

```
## Time          573      3  8625.1    114    2.5239  0.061195 .
## Group:Time    1011      3  8625.1    114    4.4531  0.005363 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##          Test statistic    p-value
## Time          0.58306 0.0013626
## Group:Time     0.58306 0.0013626
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##          GG eps Pr(>F[GG])
## Time          0.79352    0.07603 .
## Group:Time    0.79352    0.01005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          HF eps  Pr(>F[HF])
## Time          0.8501145 0.071636760
## Group:Time    0.8501145 0.008454026
```

Violations of Sphericity are more common in mixed designs, so be cautious

1.3.2.2 Covariance Matrix

We assume Compound Symmetry for each group:

$$\mathbf{Cov} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

However, as in the case of the study that we examined today, we have **two** Covariance Matrices (one for each group: $COV_1 = COV_2$). So we must test to see if they are homogeneous. We can use Box's M.

1.3.2.3 Box's M

Box's M tests to see if the covariance matrices are equal between groups (**HOCV**). This function in R requires we spread DV into columns based on the Within factor (we will use `spread` function in `tidyr` package: `dataset %>% spread(Within factor, DV)`).

```
library(tidyr); library(dplyr)
#Box's M
data_wide <- DataSim2 %>% spread(Time, BDI)
```

Now the data is wide, we will index the columns of the DV (rows 3 to 6) We will send this result to `boxM` function in `heplots` and cut it by group. We can view the covariance matrix for each group and also view the results of the test.

```

library(heplots)
BoxResult<-boxM(data_wide[,3:6],data_wide$Group)
BoxResult$cov
BoxResult

## $CBT
##           Baseline    2 Weeks    4 Weeks    6 Weeks
## Baseline 35.830484  10.082930 39.039303   4.564906
## 2 Weeks  10.082930  74.984873  2.848143 -14.474673
## 4 Weeks  39.039303   2.848143 52.237993   7.832589
## 6 Weeks   4.564906 -14.474673  7.832589  46.560581
##
## $Control
##           Baseline    2 Weeks    4 Weeks    6 Weeks
## Baseline 63.826004   5.222681 43.836602 -41.487667
## 2 Weeks   5.222681 124.377804 28.686090 40.546881
## 4 Weeks  43.836602 28.686090 120.724050 -8.021894
## 6 Weeks -41.487667 40.546881 -8.021894 165.847124
##
##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data:  data_wide[, 3:6]
## Chi-Sq (approx.) = 32.502, df = 10, p-value = 0.0003301

```

Box's M does not work well with small samples and becomes overly sensitive in large samples. We can set alpha here to be $\alpha = .001$. If the result is significant, that means the covariance matrices are not equal which means we have a problem. This increases the rate of which we might commit a Type I error when examining the interaction: the treatment affected the covariance between trials by the group and not the just the means as we expected. So is the interaction due to mean differences or covariance differences?

1.3.2.4 HOV

Box's M is not a great test, so its recommended you also check Levene's test (center the data via mean) or Brown-Forsythe test (center the data via median) in the `car` package. Brown-Forsythe test is more robust to the violation of normality (Note: the output will still say Levene's test when you center via median). Both these tests are sensitive to unequal sample sizes per group. We can test HOV of the groups and `group*treatment`. If the interaction HOV is violated it supports the conclusion of Box's M.

```

library(car)
# Brown-Forsythe test on Group Only
leveneTest(BDI~ Group, data=DataSim2,center=median)

# Brown-Forsythe test on Group X Treatment Only
leveneTest(BDI~ Time*Group, data=DataSim2,center=median)

## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)
## group      1  7.7604 0.005994 **
##           158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)

```

```
## group    7    2.766 0.009832 **
##          152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the case of this study, Box's M and Brown-Forsythe test showed **HOCV** and **HOV** problems. So we are in trouble, what can we do?

Solution?: Cohen (2008) recommends that you do **not** do the mixed ANOVA, meaning you will be unable to test the interaction. You can only do one-way RMs for each group and do ANOVA or independent t-tests on the groups (collapsing over RM term). Is this extremely conservative approach is it justified?

2 Follow Up Testing

We have the same options as we did with RM ANOVA, but we have one extra assumption that will affect both the Modern Approach (using `emmeans` with error term from ANOVA and Satterthwaite's DF) and the Multivariate approach.

2.1 Degree of the problem?

I ran a simulation (see `MonteCarloMixed.R`) on the Type I error rates give the violations. In this simulation ($n = 1000$) of 2 (between) \times 4 (within) design with a sample size of 30. All the cells have mean of 0 and I let the covariance matrices for each group be random but different from each other.

Type I Levels:

- Interaction term in the ANOVA (uncorrected) = **.077**
- **Univariate approach** (uncorrected interaction error term + Satterthwaite's DF)
 - Significant interactions Followed up **within-subject** Pair-wise comparisons (2 families with 6 tests each)
 - * Tukey Corrected & Using the error term from the Interaction = **.175**
 - * Bonferroni Corrected & Using the error term from the ANOVA = **.169**
 - Significant interactions Followed up **Between-subject** Pair-wise comparisons (no correction as 1 per family)
 - * Using the error term from the Interaction = **.289**
- **Multivariate approach** (error term and DF from MANOVA)
 - Significant interactions Followed up **within-subject** Pair-wise comparisons (2 families with 6 tests each)
 - * Tukey Corrected & Using the error term from the Interaction = **.158**
 - * Bonferroni Corrected & Using the error term from the ANOVA = **.152**
 - Significant interactions Followed up **Between-subject** Pair-wise comparisons (no correction as 1 per family)
 - * Using the error term from the Interaction = **.279**

This is extremely high on the follow-ups testing and worse yet when you follow up between-subject way. This is the reason for Cohen suggestions not to test the interaction. In the wild no one does this in practice because usually the whole point of the study.

2.2 Recommendations

- For Box's M, set $\alpha = .001$, double check Brown-Forsythe to be sure there is a problem, and examine Spaghetti plots when you have a small number of subjects.
- When Box's M & B-F are violated:

- Follow it up from the within variable ONLY and the minimum number of tests possible to test your predicted hypothesis.
 - * If you need to contrasts (using the error term of the ANOVA) do only the one you need and keep in mind you might be able to be replicated
 - * If you only need pairwise tests, do NOT use the error term from the ANOVA (do paired sample t-tests and Bonferroni correct the hell out of it)
 - * The multivariate approach will not help you here as you could see
- DO NOT DO ANY EXPLORATORY ANALYSIS; these will be all type I error.
- If you have designed your study well and had no violations, go ahead follow up the way you would follow up an RM ANOVA using emmeans, but try to follow it up only within-subject as you have higher power (and thus lower Type I risk)

2.2.1 Example of Follow up (assuming no violations)

2.2.1.1 Follow Up Interaction

Example using simple effects followed up by polynomial contrast

```
library(emmeans)
Time.by.Group<-emmeans(Mixed.1,~Time|Group)
test(pairs(Time.by.Group), joint=TRUE)
```

```
## Group    df1 df2 F.ratio p.value note
## CBT      3 114  6.459  0.0004  d
## Control  3 114  0.518  0.6706  d
##
## d: df1 reduced due to linear dependence
```

So changes over time in Control group which we can follow up with contrasts: polynomials, consecutive, or any custom contrast you want to test. [Note we can index 1:3 to see only the CBT results]

```
contrast(Time.by.Group, 'poly')[1:3]
```

##	contrast	Group	estimate	SE	df	t.ratio	p.value
##	linear	CBT	-34.135533	8.698216	114	-3.924	0.0001
##	quadratic	CBT	6.720889	3.889960	114	1.728	0.0867
##	cubic	CBT	-8.654832	8.698216	114	-0.995	0.3218

3 References

Cousineau, D., & O'Brien, F. (2014). Error bars in within-subject designs: a comment on Baguley (2012). *Behavior Research Methods*, 46(4), 1149-1151.