R Answers to Exercises: Module 2

If you have any questions as you go through these, feel free to ask them in the forum.

1. If you haven't done so, read references 1-3 listed on the module 2 page. You can substitute #4 for #1 and #5 for #2, if they're easier to come by or make more sense to you.

2. In your own words, what is Sphericity?
   Sphericity is a quality of a covariance matrix in which the differences in means have equal variance and are independent.

3. What covariance structure does the multivariate approach use?
   Unstructured.

4. Huyhn-Feldt covariance structure for three repeats looks like the matrix on the right:

$$Huynh-Feldt = \begin{bmatrix} \sigma_1^2 & \dfrac{\sigma_1^2+\sigma_2^2}{2}-\lambda & \dfrac{\sigma_1^2+\sigma_3^2}{2}-\lambda \\ \dfrac{\sigma_2^2+\sigma_1^2}{2}-\lambda & \sigma_2^2 & \dfrac{\sigma_2^2+\sigma_3^2}{2}-\lambda \\ \dfrac{\sigma_3^2+\sigma_1^2}{2}-\lambda & \dfrac{\sigma_3^2+\sigma_2^2}{2}-\lambda & \sigma_3^2 \end{bmatrix}$$

   List the parameters that would need to be estimated if there were only two repeats. How many are there? Compare the number of parameters to a compound symmetry and an unstructured structure.

   Two variance parameters—Var1, Var2,plus one lambda. Total of 3.
   Compound symmetry would have two (one variance, one covariance) and
   Unstructured would have 3 (two variances, one covariance).

   Do the same if there were 4 repeats, then 8.
   Four repeats: Four variance parameters—Var1, Var2, Var3, Var4, plus lambda. 5 total. Compound symmetry would have two (one variance, one covariance) and Unstructured would have 10 (four variances, six covariances).
   Eight repeats: Eight variance parameters- Var1 to Var8, plus lambda. 9 total. Compound symmetry would have two (one variance, one covariance) and Unstructured would have 36(eight variances, 28 covariances).

5. Using the Physical Training Data, use GLM Repeated Measures to test if mean LDL levels (which I believe is low-density cholesterol—the bad stuff) change from pre- to post-training equally in the three training regimen groups. Support your answer. What do you conclude about the effects of training regimen on LDL?

Test for Time:

```
Sum of squares and products for the hypothesis:
        time1
time1 32.7861

Multivariate Tests: time
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1 0.5803537 33.19102      1     24 6.1472e-06 ***
Wilks             1 0.4196463 33.19102      1     24 6.1472e-06 ***
Hotelling-Lawley  1 1.3829594 33.19102      1     24 6.1472e-06 ***
Roy               1 1.3829594 33.19102      1     24 6.1472e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time * Group:

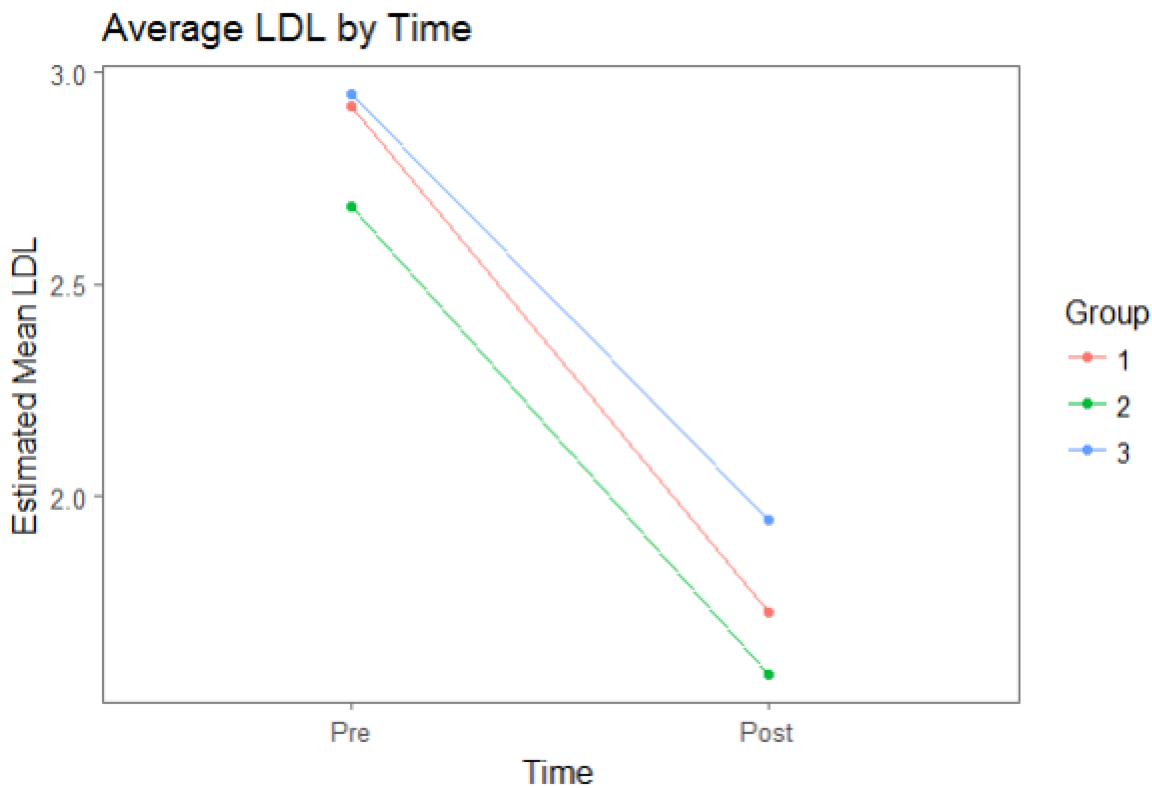
```
Sum of squares and products for the hypothesis:
          time1
time1 0.1672448

Multivariate Tests: as.factor(group):time
                 Df test stat   approx F num Df den Df Pr(>F)
Pillai            2 0.0070052 0.08465515      2     24 0.9191
Wilks             2 0.9929948 0.08465515      2     24 0.9191
Hotelling-Lawley  2 0.0070546 0.08465515      2     24 0.9191
Roy               2 0.0070546 0.08465515      2     24 0.9191
```

Univariate tests:

```
Univariate Type II Repeated-Measures ANOVA Assuming Sphericity

                            SS num Df Error SS den Df        F    Pr(>F)    
(Intercept)            285.518      1   21.570     24 317.6806 2.404e-15 ***
as.factor(group)         0.908      2   21.570     24   0.5051    0.6097    
time                    16.393      1   11.854     24  33.1910 6.147e-06 ***
as.factor(group):time    0.084      2   11.854     24   0.0847    0.9191    
---
```

Means:

```
Response LDL_pre :
                     Estimate Std. Error     t value     Pr(>|t|)
(Intercept)         2.92161611  0.3375976  8.65413741 7.637560e-09
as.factor(group)2 -0.23919200  0.4774351 -0.50099373 6.209396e-01
as.factor(group)3  0.02555578  0.4774351  0.05352723 9.577549e-01

Response LDL_Post :
                    Estimate Std. Error    t value     Pr(>|t|)
(Intercept)        1.7246463  0.2019098  8.5416660 9.708286e-09
as.factor(group)2 -0.1468683  0.2855436 -0.5143464 6.117139e-01
as.factor(group)3  0.2182829  0.2855436  0.7644467 4.520523e-01
```



All three training regimens have a significant effect on LDL—the mean value of LDL decreased in all three groups ($F=33.19$, $p < .001$).

6. Using the County data, test whether the mean number of jobs in Alabama changed across the 5 decades of the study, and whether the change differed for counties classified as rural and non-rural in 1960. Include a mean plot. Describe the findings and support your answer. (Note: It may be easier to read the output if you change the scale of the outcome variable to Thousands of Jobs). Do the univariate and multivariate results differ?

   Yes, the mean number of jobs did change for non-rural counties across every decade, but not at all for rural counties. Although the Time main effect is significant (Wilk's $\Lambda = 11.94$, $p < .001$), the estimated marginal means make it clear that this is only driven by the significant interaction. Within the rural counties, there are no differences at all among years, but within the non-rural counties, all differences are highly significant.

   Test for Year:

```
Sum of squares and products for the hypothesis:
            year1       year2       year3      year4
year1 19692798482 17204065889 11665403632 6091681560
year2 17204065889 15029853851 10191155559 5321828232
year3 11665403632 10191155559  6910223655 3608523403
year4  6091681560  5321828232  3608523403 1884373329

Multivariate Tests: year
                 Df test stat approx F num Df den Df     Pr(>F)    
Pillai            1 0.2845917 6.165948      4     62 0.00030436 ***
Wilks             1 0.7154083 6.165948      4     62 0.00030436 ***
Hotelling-Lawley  1 0.3978031 6.165948      4     62 0.00030436 ***
Roy               1 0.3978031 6.165948      4     62 0.00030436 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tests for Year x Rural1960:

```
Sum of squares and products for the hypothesis:
            year1       year2       year3      year4
year1 23478140400 20542240325 14010922019 7352842221
year2 20542240325 17973469380 12258880916 6433382261
year3 14010922019 12258880916  8361221650 4387915619
year4  7352842221  6433382261  4387915619 2302750039

Multivariate Tests: as.factor(Rural1960):year
                 Df test stat approx F num Df den Df     Pr(>F)    
Pillai            1 0.3219206 7.358678      4     62 6.4607e-05 ***
Wilks             1 0.6780794 7.358678      4     62 6.4607e-05 ***
Hotelling-Lawley  1 0.4747535 7.358678      4     62 6.4607e-05 ***
Roy               1 0.4747535 7.358678      4     62 6.4607e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test of Sphericity:

```
Mauchly Tests for Sphericity

                         Test statistic     p-value
year                         0.00012443 1.3028e-116
as.factor(Rural1960):year    0.00012443 1.3028e-116
```

Univariate tests:

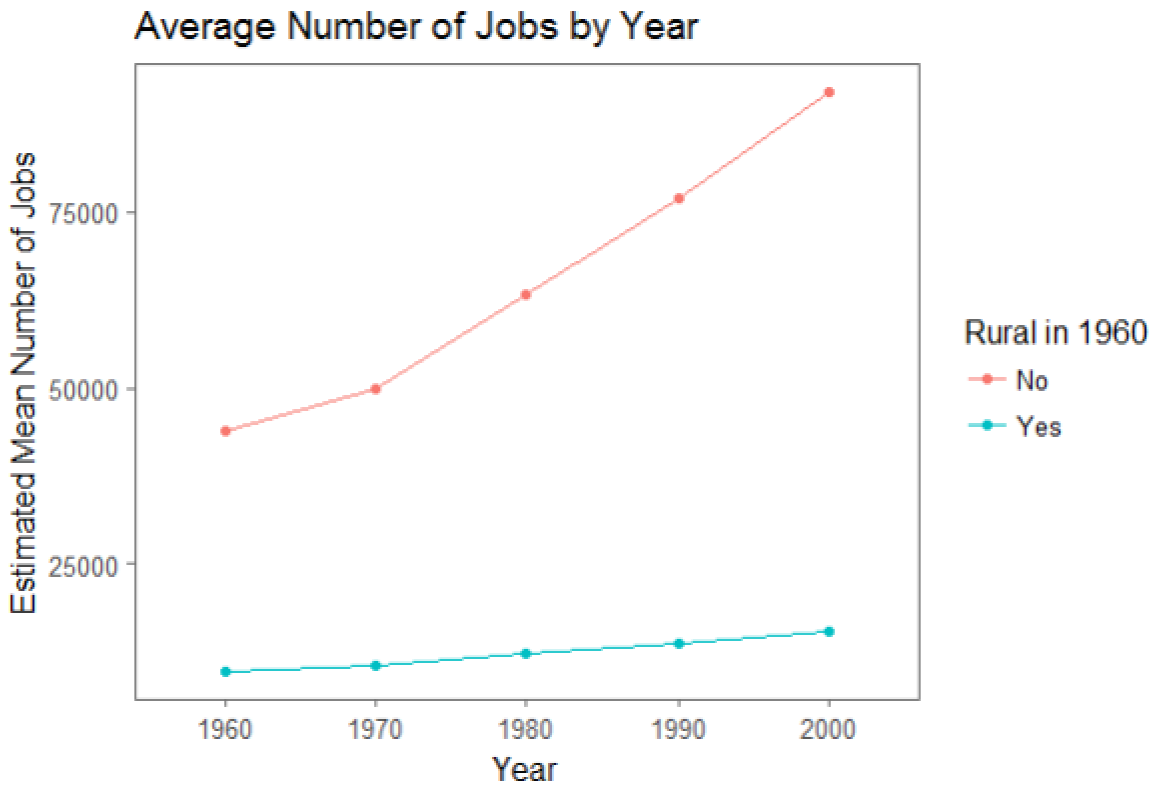
```
Univariate Type II Repeated-Measures ANOVA Assuming Sphericity

                                 SS num Df   Error SS den Df      F    Pr(>F)    
(Intercept)               2.3614e+11      1 6.7082e+11     65 22.881 1.029e-05 ***
as.factor(Rural1960)      1.8312e+11      1 6.7082e+11     65 17.744 7.948e-05 ***
year                      1.3181e+10      4 3.6152e+10    260 23.698 < 2.2e-16 ***
as.factor(Rural1960):year 1.5698e+10      4 3.6152e+10    260 28.224 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

BUT correct for small p-value in Mauchly test:

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

                            GG eps Pr(>F[GG])
year                       0.26701  4.232e-06 ***
as.factor(Rural1960):year 0.26701  7.132e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                              HF eps    Pr(>F[HF])
year                       0.2678258 4.116941e-06
as.factor(Rural1960):year 0.2678258 6.902166e-07
```

Means:

```
Rural1960 = 0:
 year          lsmean        SE df    lower.CL  upper.CL
 Jobs1960 43813.722  8017.357 65 27801.95546  59825.49
 Jobs1970 49842.556  8762.708 65 32342.21942  67342.89
 Jobs1980 63256.389 10613.052 65 42060.65960  84452.12
 Jobs1990 76867.500 12229.185 65 52444.13186 101290.87
 Jobs2000 91843.500 14177.201 65 63529.67496 120157.33

Rural1960 = 1:
 year          lsmean        SE df    lower.CL  upper.CL
 Jobs1960  9777.878  4859.252 65    73.28139  19482.47
 Jobs1970 10525.755  5311.003 65   -81.05035  21132.56
 Jobs1980 12191.367  6432.481 65  -655.18458  25037.92
 Jobs1990 13826.245  7412.005 65  -976.55158  28629.04
 Jobs2000 15576.306  8592.681 65 -1584.46332  32737.08
```



7. Using the Teacher data, test whether children's summer expectancies and gender predict teacher's ratings of rapport with each student (STRS) over time. Treat student as the subject.

According to both the univariate and multivariate tests, the only significant effect in the model is gender. There are no effects of children's summer expectancies on teacher's ratings of relationship quality with students. Likewise, there is no effect of time. At all three time points, girls are rated higher than boys.

Test of Time:

```
Sum of squares and products for the hypothesis:
        time1     time2
time1  138.72 -141.4400
time2 -141.44  144.2133

Multivariate Tests: time
                 Df test stat approx F num Df den Df Pr(>F)
Pillai            1 0.0577120 2.174257      2     71 0.1212
Wilks             1 0.9422880 2.174257      2     71 0.1212
Hotelling-Lawley  1 0.0612467 2.174257      2     71 0.1212
Roy               1 0.0612467 2.174257      2     71 0.1212
```

Time x Gender:

```
Sum of squares and products for the hypothesis:
          time1    time2
time1  3.430103 18.50503
time2 18.505030 99.83259

Multivariate Tests: as.factor(Gender):time
                 Df test stat  approx F num Df den Df  Pr(>F)
Pillai            1 0.0117326 0.4214509      2     71 0.65772
Wilks             1 0.9882674 0.4214509      2     71 0.65772
Hotelling-Lawley  1 0.0118719 0.4214509      2     71 0.65772
Roy               1 0.0118719 0.4214509      2     71 0.65772
```

Time x tOTchExp:

```
Sum of squares and products for the hypothesis:
         time1    time2
time1 557.7308 518.2919
time2 518.2919 481.6419

Multivariate Tests: tOTchExp:time
                 Df test stat approx F num Df den Df   Pr(>F)
Pillai            1 0.0664371 2.526363      2     71 0.087116 .
Wilks             1 0.9335629 2.526363      2     71 0.087116 .
Hotelling-Lawley  1 0.0711652 2.526363      2     71 0.087116 .
Roy               1 0.0711652 2.526363      2     71 0.087116 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Univariate analyses:

```
Univariate Type II Repeated-Measures ANOVA Assuming Sphericity

                          SS num Df Error SS den Df         F    Pr(>F)
(Intercept)          3005831      1  29686.9     72 7290.0732 < 2.2e-16 ***
tOTchExp                  57      1  29686.9     72    0.1376 0.7117699
as.factor(Gender)       6493      1  29686.9     72   15.7469 0.0001693 ***
time                     283      2   9506.5    144    2.1427 0.1210598
tOTchExp:time            347      2   9506.5    144    2.6310 0.0754642 .
as.factor(Gender):time    57      2   9506.5    144    0.4280 0.6526661
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

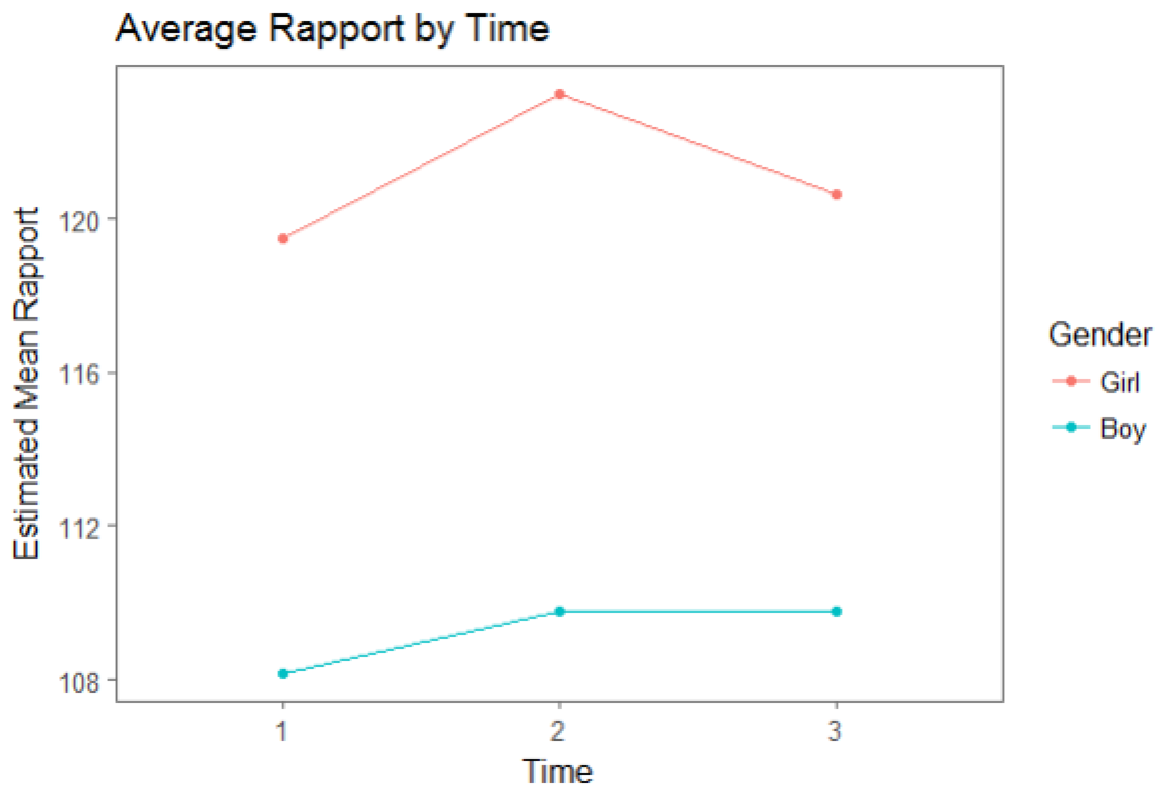
Remember that the effect of time is NOT correct for a type III effect even when type = 3 is used in the R syntax! This IS a correct type II test. This examines the effect of time BEFORE the interactions with t0TchExp and Gender are included in the model.

Mauchly test of sphericity:

```
Mauchly Tests for Sphericity

                        Test statistic p-value
time                           0.99915 0.97036
t0TchExp:time                  0.99915 0.97036
as.factor(Gender):time         0.99915 0.97036
```

Means:

```
Gender = 1:
 time     lsmean       SE df lower.CL upper.CL
 STRS.1 119.5048 2.207068 72 115.1051 123.9045
 STRS.2 123.2210 2.243612 72 118.7484 127.6935
 STRS.3 120.6442 2.252955 72 116.1530 125.1354

Gender = 2:
 time     lsmean       SE df lower.CL upper.CL
 STRS.1 108.1659 2.374839 72 103.4318 112.9001
 STRS.2 109.8046 2.414161 72 104.9920 114.6171
 STRS.3 109.7780 2.424214 72 104.9454 114.6106

Confidence level used: 0.95
```


Average Rapport by Time
Estimated Mean Rapport
120
116
112
108
1
2
3
Time
Gender
Girl
Boy

8. Using the Swallowing data, test whether there are differences in mean Release Slope for the following four tasks: ESS (Effortful Saliva Swallow), NESS (Noneffortful Saliva Swallow), Water (DSW), and NectarThick Apple Juice (ANEC). You won't be able to have more than one swallow for each task per person, so the best we can do in this analysis is take the mean for the 5 trials on each task, and use it as the outcome variable. Make sure you restrict the data to the posterior bulb.

According to both the univariate and multivariate tests, there is a significant effect of Task. A look at the mean comparisons indicates that this effect is only due to a higher mean for ESS compared to NESS. No other pairwise comparisons are significant.

Test of Task:

```
Sum of squares and products for the hypothesis:
         task1     task2    task3
task1 471842.5 171877.53 440310.5
task2 171877.5  62609.63 160391.4
task3 440310.5 160391.39 410885.6

Multivariate Tests: task
                 Df test stat approx F num Df den Df   Pr(>F)
Pillai            1 0.5263229 5.185333      3     14 0.012852 *
Wilks             1 0.4736771 5.185333      3     14 0.012852 *
Hotelling-Lawley  1 1.1111427 5.185333      3     14 0.012852 *
Roy               1 1.1111427 5.185333      3     14 0.012852 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Univariate test of Task:

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                 SS num Df Error SS den Df       F    Pr(>F)
(Intercept) 6534824      1  1102135     16 94.8679 3.958e-08 ***
task         322714      3  1175055     48  4.3942  0.008231 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test of sphericity:

```
Mauchly Tests for Sphericity

     Test statistic p-value
task        0.76281 0.55221
```

Means:

```
task                lsmean       SE df lower.CL upper.CL
 ReleaseSlope.ANEC 380.9120 44.66976 16 286.2163 475.6076
 ReleaseSlope.DSW  274.9994 42.14332 16 185.6595 364.3392
 ReleaseSlope.ESS  369.7785 52.05643 16 259.4238 480.1332
 ReleaseSlope.NESS 214.3123 43.48220 16 122.1342 306.4905
```