ST 790, Homework 2 Solutions
Spring 2015

1. (a) Let $U_{gh} = \sum_{i=1}^{N} I(Y_{i1} = g, Y_{i2} = h)$. The joint density of the U_{gh}, $g = 1, \dots, G$, $h = 1, \dots, H$, is the multinomial density

$$\frac{N!}{\prod_{g=1}^{G}\prod_{h=1}^{H} U_{gh}!}\prod_{g=1}^{G}\prod_{h=1}^{H}\theta_{gh}^{U_{gh}}.$$

Taking logarithms and ignoring the constant term, we thus want to maximize

$$\sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh}\log(\theta_{gh}) \quad \text{subject to} \quad \sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} = 1.$$

Most of you imposed the constraint by using a Lagrange multiplier approach. Letting

$$L(\theta) = \sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh}\log(\theta_{gh}) - \lambda\left(\sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} - 1\right),$$

taking derivatives of $L(\theta)$ with respect to each θ_{gh} and λ and setting equal to zero yields

$$\theta_{gh} = \frac{U_{gh}}{\lambda}, \quad g = 1, \dots, G, h = 1, \dots, H,$$

and

$$\sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} = 1.$$

Using the fact that

$$\sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh} = N,$$

we get

$$\lambda = N,$$

from whence it follows that

$$\widehat{\theta}_{gh} = \frac{U_{gh}}{N}, \quad g = 1, \dots, G, h = 1, \dots, H.$$

(b) Clearly, from the full data loglikelihood, $U_{gh} = \sum_{i=1}^{N} I(Y_{i1} = g, Y_{i2} = h)$ is a sufficient statistic for θ_{gh}, $g = 1, \dots, G$, $h = 1, \dots, H$. Accordingly, from pages 68-69 of the notes, the E-step involves finding for each $g = 1, \dots, G$, $h = 1, \dots, H$

$$E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\}.$$

Here, R takes on values $r = (1,1)^T$, $(1,0)^T$, or $(0,1)^T$. Trivially, when $R = (1,1)^T$, $Z_{(R)} = (Y_1, Y_2)$, and

$$E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} = I(Y_1 = g, Y_2 = h).$$

When $R = (1,0)^T$, $Z_{(R)} = Y_1$, and

$$\begin{aligned}
E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} &= E\{I(Y_1 = g)I(Y_2 = h)|Y_1 = g)\} \\
&= I(Y_1 = g)E\{I(Y_2 = h)|Y_1 = g)\} = I(Y_1 = g)\text{pr}(Y_2 = h|Y_1 = g) \\
&= I(Y_1 = g)\frac{\theta_{gh}}{\theta_{g\cdot}},
\end{aligned}$$

using $\theta_{gh} = \text{pr}(Y_1 = g, Y_2 = h)$ and $\text{pr}(Y_1 = g) = \theta_{g\cdot} = \sum_{h=1}^{H}\theta_{gh}$. Analogously, when $R = (0,1)^T$,

$$E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} = I(Y_2 = h)E\{I(Y_1 = g)|Y_2 = h\} = I(Y_2 = h)\frac{\theta_{gh}}{\theta_{\cdot h}},$$

where $\theta_{\cdot h} = \sum_{g=1}^{G}\theta_{gh}$. Thus, given the tth iterate $\theta_{gh}^{(t)}$ for $g = 1, \dots, G,$, $h = 1, \dots, H$, and defining $\theta_{g\cdot}^{(t)}$ and $\theta_{\cdot h}^{(t)}$ in the obvious way, the E-step involves calculating for each $i = 1, \dots, N$

$$\theta_{gh,i}^{(t+1)} = R_{i1}R_{i2}I(Y_{i1} = g, Y_{i2} = h) + R_{i1}(1 - R_{i2})I(Y_{i1} = g)\frac{\theta_{gh}^{(t)}}{\theta_{g\cdot}^{(t)}} + (1 - R_{i1})R_{i2}I(Y_{i2} = h)\frac{\theta_{gh}^{(t)}}{\theta_{\cdot h}^{(t)}}.$$

The M-step is then

$$\theta_{gh}^{(t+1)} = N^{-1}\sum_{i=1}^{N}\theta_{gh,i}^{(t+1)}, \quad g = 1, \dots, G, h = 1, \dots, H.$$

2. All of you summarized the missingness patterns for (a). For (b) – (d), you needed to massage and reconfigure the data set to use SAS `proc mixed` or R `gls` to fit the model as parameterized in (4), namely

$$Y_{ij} = \mu_{0j} + \beta_j A_i + \epsilon_{ij}, \quad \epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{i5})^T \sim \mathcal{N}(0, \Sigma).$$

Here, the model is parameterized in terms of the means for treatment 0 (placebo) at each of the 5 time points (μ_{0j}, $j = 1, \dots, 5$) and the differences in means between treatments 1 and 0 (active and placebo) at each of the time points (β_j, $j = 1, \dots, 5$). To fit this mean model with the common unstructured covariance matrix using SAS `proc mixed`, the basic specification is

```
proc mixed method=ml;
    class week pid;
    model visual = week week*treat / noint solution;
    repeated / subject=pid type=un;  *  or repeated week / subject=pid type=un;
```

Using R `gls`, it is of the form

```
#  Need to give the corSymm structure an integer time sequence so it
#  can identify which observations correspond to which positions
time <- rep(seq(1,5,1),nrow(armd.data))
#  need week to be a classification variable so we get a separate
#  mean for each week
week <- factor(weeks)
mle.result <- gls(visual ~ -1 + week + week:treat,
           correlation=corSymm(form = ~ time | factor(pid)),
           weights = varIdent(form = ~ 1 |week),method="ML",na.action=na.omit)
```

Recall that `corSymm` specifies the unstructured *correlation* structure. `gls` assumes by default that all outcomes have the *same variance*, so to generalize this to have different variances at each week/time, we also need to use the `weights = varIdent` option as shown above.

Here are results for each case:

(b) Available case analysis: R `gls` yields

```
Coefficients:
                Value Std.Error  t-value p-value
week0        55.33613  1.367406 40.46795  0.0000
week4        54.05487  1.460954 36.99971  0.0000
week12       52.98447  1.588584 33.35327  0.0000
week24       49.31615  1.721048 28.65473  0.0000
week52       44.02513  1.766815 24.91780  0.0000
week0:treat  -0.75762  1.925797 -0.39341  0.6941
week4:treat  -2.96186  2.063044 -1.43567  0.1514
week12:treat -4.26556  2.253360 -1.89298  0.0586
week24:treat -3.82741  2.453133 -1.56021  0.1190
week52:treat -5.62378  2.545241 -2.20953  0.0273

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 220.50 201.61 188.79 182.45 143.79
[2,] 201.61 250.77 228.67 219.40 175.34
[3,] 188.79 228.67 295.71 262.48 224.76
[4,] 182.45 219.40 262.48 342.44 290.86
[5,] 143.79 175.34 224.76 290.86 350.16
  Standard Deviations: 14.849 15.836 17.196 18.505 18.713
```

SAS `proc mixed` yields

```
                         Solution for Fixed Effects

                                          Standard
   Effect        week     Estimate           Error      DF     t Value     Pr > |t|

   week            0       55.3361          1.3612     240       40.65       <.0001
   week            4       54.0549          1.4543     240       37.17       <.0001
   week           12       52.9845          1.5814     240       33.50       <.0001
   week           24       49.3162          1.7133     240       28.78       <.0001
   week           52       44.0251          1.7588     240       25.03       <.0001
   treat*week      0       -0.7576          1.9171     240       -0.40       0.6931
   treat*week      4       -2.9619          2.0537     240       -1.44       0.1506
   treat*week     12       -4.2656          2.2432     240       -1.90       0.0584
   treat*week     24       -3.8274          2.4421     240       -1.57       0.1184
   treat*week     52       -5.6238          2.5338     240       -2.22       0.0274

                         Estimated R Matrix for pid 2
```

```
  Row       Col1       Col2       Col3       Col4       Col5

    1     220.50     201.61     188.80     182.46     143.80
    2     201.61     250.77     228.68     219.41     175.34
    3     188.80     228.68     295.72     262.49     224.77
    4     182.46     219.41     262.49     342.45     290.87
    5     143.80     175.34     224.77     290.87     350.17
```

(c) Complete case analysis: R `gls` yields

```
Coefficients:
                Value Std.Error  t-value p-value
week0        55.39216  1.473343 37.59624  0.0000
week4        54.47059  1.551355 35.11163  0.0000
week12       53.07843  1.673451 31.71795  0.0000
week24       49.79412  1.808359 27.53553  0.0000
week52       44.43137  1.835730 24.20365  0.0000
week0:treat  -0.54332  2.178380 -0.24941  0.8031
week4:treat  -2.86594  2.293722 -1.24947  0.2118
week12:treat -2.89238  2.474245 -1.16900  0.2427
week24:treat -3.27086  2.673710 -1.22334  0.2215
week52:treat -4.71044  2.714179 -1.73549  0.0830

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 219.06 205.22 194.81 191.84 151.26
[2,] 205.22 242.87 221.54 217.76 174.92
[3,] 194.81 221.54 282.61 254.05 218.11
[4,] 191.84 217.76 254.05 330.01 279.89
[5,] 151.26 174.92 218.11 279.89 340.07
  Standard Deviations: 14.801 15.584 16.811 18.166 18.441
```

SAS `proc mixed` yields

```
                       Solution for Fixed Effects

                                       Standard
   Effect        week    Estimate         Error       DF     t Value     Pr > |t|

   week            0      55.3922        1.4655      188       37.80       <.0001
   week            4      54.4706        1.5431      188       35.30       <.0001
   week           12      53.0784        1.6645      188       31.89       <.0001
   week           24      49.7941        1.7987      188       27.68       <.0001
   week           52      44.4314        1.8259      188       24.33       <.0001
   treat*week      0      -0.5433        2.1668      188       -0.25       0.8023
   treat*week      4      -2.8659        2.2815      188       -1.26       0.2106
```

```
   treat*week    12       -2.8924      2.4611      188       -1.18       0.2414
   treat*week    24       -3.2709      2.6595      188       -1.23       0.2203
   treat*week    52       -4.7104      2.6997      188       -1.74       0.0827
```

```
                         Estimated R Matrix for pid 4

         Row        Col1        Col2        Col3        Col4        Col5

           1      219.06      205.22      194.81      191.84      151.27
           2      205.22      242.87      221.54      217.76      174.92
           3      194.81      221.54      282.61      254.05      218.12
           4      191.84      217.76      254.05      330.01      279.89
           5      151.27      174.92      218.12      279.89      340.08
```

(d) LOCF analysis: R `gls` yields

```
Coefficients:
                Value Std.Error  t-value p-value
week0        55.33613  1.366928 40.48211  0.0000
week4        54.05882  1.457555 37.08869  0.0000
week12       53.01681  1.579395 33.56780  0.0000
week24       49.46218  1.700184 29.09226  0.0000
week52       44.70588  1.729726 25.84564  0.0000
week0:treat  -0.75762  1.925124 -0.39354  0.6940
week4:treat  -2.78610  2.052759 -1.35724  0.1750
week12:treat -3.91763  2.224353 -1.76125  0.0785
week24:treat -2.87541  2.394467 -1.20086  0.2300
week52:treat -3.69762  2.436073 -1.51786  0.1293

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 220.50 202.55 190.01 184.35 152.43
[2,] 202.55 250.70 230.54 224.94 186.46
[3,] 190.01 230.54 294.37 267.19 232.11
[4,] 184.35 224.94 267.19 341.12 296.17
[5,] 152.43 186.46 232.11 296.17 353.08
  Standard Deviations: 14.849 15.834 17.157 18.469 18.79
```

SAS `proc mixed` yields

```
                           Solution for Fixed Effects

                                             Standard
   Effect        week     Estimate             Error      DF     t Value     Pr > |t|

   week            0       55.3361            1.3612     240       40.65       <.0001
   week            4       54.0588            1.4515     240       37.24       <.0001
   week           12       53.0168            1.5728     240       33.71       <.0001
```

```
week             24          49.4622        1.6931      240        29.21        <.0001
week             52          44.7059        1.7225      240        25.95        <.0001
treat*week        0          -0.7576        1.9171      240        -0.40        0.6931
treat*week        4          -2.7861        2.0442      240        -1.36        0.1742
treat*week       12          -3.9176        2.2151      240        -1.77        0.0782
treat*week       24          -2.8754        2.3845      240        -1.21        0.2291
treat*week       52          -3.6976        2.4259      240        -1.52        0.1288

                        Estimated R Matrix for pid 2

       Row        Col1        Col2        Col3        Col4        Col5

         1      220.50      202.55      190.01      184.36      152.43
         2      202.55      250.71      230.54      224.94      186.47
         3      190.01      230.54      294.38      267.20      232.11
         4      184.36      224.94      267.20      341.13      296.18
         5      152.43      186.47      232.11      296.18      353.08
```

(e) The estimates of β_5 from the three analyses are

	Est	SE	P-value
Available	-5.624	2.534	0.03
Complete Case	-4.710	2.700	0.08
LOCF	-3.698	2.426	0.13

The available case analysis based on all observed data, which is the appropriate observed data likelihood analysis under the assumptions of MAR and multivariate normality, shows evidence supporting a mean difference at 52 weeks (level 0.05 test) and yields a point estimate of the difference of -5.6. The inappropriate complete case and LOCF analyses do not suggest (at level 0.05) evidence supporting a difference, and the point estimate of the difference is smaller in each case.

(f) The code to fit the model (5) for each treatment $a = 0, 1$ using SAS `proc mixed` is

```
proc mixed method=ml; by treat;
    class week pid;
    model visual = week /noint solution;
    repeated / subject=pid type=un r=2 rcorr=2;
run;
```

This yields

```
----------------------------------- treat=0 -----------------------------------

                        Solution for Fixed Effects

                                     Standard
    Effect     week     Estimate        Error       DF     t Value     Pr > |t|

    week          0      55.3361       1.3694      119       40.41       <.0001
```

week	4	54.0529	1.4463	119	37.37	<.0001
week	12	52.9808	1.5695	119	33.76	<.0001
week	24	49.3409	1.7409	119	28.34	<.0001
week	52	44.0346	1.7566	119	25.07	<.0001

Estimated R Matrix for pid 4

Row	Col1	Col2	Col3	Col4	Col5
1	223.15	206.17	188.80	188.82	148.25
2	206.17	248.10	223.43	219.37	176.35
3	188.80	223.43	291.25	251.31	217.76
4	188.82	219.37	251.31	353.02	296.47
5	148.25	176.35	217.76	296.47	349.42

----------------------------------- treat=1 -----------------------------------

Solution for Fixed Effects

Effect	week	Estimate	Standard Error	DF	t Value	Pr > |t|
week	0	54.5785	1.3419	121	40.67	<.0001
week	4	51.0966	1.4587	121	35.03	<.0001
week	12	48.7181	1.6020	121	30.41	<.0001
week	24	45.5213	1.6941	121	26.87	<.0001
week	52	38.4559	1.8278	121	21.04	<.0001d

Estimated R Matrix for pid 2

Row	Col1	Col2	Col3	Col4	Col5
1	217.90	196.94	189.02	174.60	139.30
2	196.94	253.38	233.98	217.41	173.77
3	189.02	233.98	300.15	272.20	230.93
4	174.60	217.41	272.20	326.03	282.20
5	139.30	173.77	230.93	282.20	351.13

Using R `gls`, code is, for treatment 0 for example,

```
time0 <- time[treat==0]
visual0 <- visual[treat==0]
week0 <- factor(weeks[treat==0])
mle.treat0 <- gls(visual0 ~  -1 + week0,
           correlation=corSymm(form = ~ time0 | factor(pid)),
           weights = varIdent(form = ~ 1 |week0),method="ML",na.action=na.omit)
```

This yields

```
Coefficients:
           Value Std.Error  t-value p-value
week00  55.33613  1.375417 40.23226       0
week04  54.05293  1.452678 37.20918       0
week012 52.98078  1.576440 33.60786       0
week024 49.34091  1.748577 28.21775       0
week052 44.03456  1.764363 24.95777       0

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 223.15 206.17 188.80 188.81 148.25
[2,] 206.17 248.10 223.43 219.36 176.35
[3,] 188.80 223.43 291.25 251.30 217.76
[4,] 188.81 219.36 251.30 353.02 296.47
[5,] 148.25 176.35 217.76 296.47 349.42
  Standard Deviations: 14.938 15.751 17.066 18.789 18.693

Coefficients:
           Value Std.Error  t-value p-value
week10  54.57851  1.348227 40.48171       0
week14  51.09662  1.465549 34.86517       0
week112 48.71815  1.609522 30.26870       0
week124 45.52128  1.702002 26.74573       0
week152 38.45591  1.836415 20.94076       0

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 217.90 196.93 189.01 174.60 139.30
[2,] 196.93 253.38 233.97 217.40 173.76
[3,] 189.01 233.97 300.14 272.19 230.92
[4,] 174.60 217.40 272.19 326.02 282.19
[5,] 139.30 173.76 230.92 282.19 351.13
  Standard Deviations: 14.761 15.918 17.324 18.056 18.738
```

(g) The code to implement the EM algorithm to fit (5) for each treatment using SAS `proc mi` is

```
proc mi data=armd seed=370252 simple nimpute=0; by treat;
    em initial=ac maxiter=200 converge=1e-4 itprint;
    var visual0 visual4 visual12 visual24 visual52;
run;
```

and yields

```
----------------------------------- treat=0 ------------------------------------

                        EM (MLE) Parameter Estimates
```

```
 _TYPE_    _NAME_           visual0        visual4       visual12       visual24

 MEAN                     55.336134      54.052925      52.980780      49.340990
 COV       visual0       223.147518     206.169154     188.798637     188.815488
 COV       visual4       206.169154     248.105146     223.429001     219.361610
 COV       visual12      188.798637     223.429001     291.253074     251.299053
 COV       visual24      188.815488     219.361610     251.299053     353.015463
 COV       visual52      148.249535     176.348266     217.761896     296.465010

                                        visual52

                                       44.034562
                                      148.249535
                                      176.348266
                                      217.761896
                                      296.465010
                                      349.424750

----------------------------------- treat=1 ------------------------------------

                            EM (MLE) Parameter Estimates

 _TYPE_    _NAME_           visual0        visual4       visual12       visual24

 MEAN                     54.578512      51.096624      48.718150      45.521305
 COV       visual0       217.896728     196.936457     189.016355     174.599657
 COV       visual4       196.936457     253.382251     233.979421     217.406702
 COV       visual12      189.016355     233.979421     300.146006     272.195593
 COV       visual24      174.599657     217.406702     272.195593     326.030258
 COV       visual52      139.305009     173.764660     230.928394     282.195016

                                        visual52

                                       38.455976
                                      139.305009
                                      173.764660
                                      230.928394
                                      282.195016
                                      351.124415
```

For R `gls` the code is

```
prelim.em0 <- prelim.norm(visual0)
theta.init0 <- em.norm(prelim.em0,showits=FALSE)
theta.init0 <- getparam.norm(prelim.em0,theta.init0,corr=TRUE)
theta.init0 <- makeparam.norm(prelim.em0,theta.init0)
trt0.em <- em.norm(prelim.em0,start=theta.init0,showits=TRUE,maxits=200,criterion=1e-5)
theta.em.trt0 <- getparam.norm(prelim.em0,trt0.em,corr=TRUE)
```

```
theta.em.trt0$mu       #  means
[1] 55.33613 54.05293 52.98078 49.34093 44.03457
#  covariance matrix
theta.em.trt0.cov <-diag(theta.em.trt0$sdv)%*%theta.em.trt0$r%*%diag(theta.em.trt0$sdv)
theta.em.trt0.cov
         [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 223.1475 206.1692 188.7986 188.8160 148.2489
[2,] 206.1692 248.1052 223.4291 219.3650 176.3470
[3,] 188.7986 223.4291 291.2531 251.3030 217.7606
[4,] 188.8160 219.3650 251.3030 353.0215 296.4667
[5,] 148.2489 176.3470 217.7606 296.4667 349.4235

prelim.em1 <- prelim.norm(visual1)
theta.init1 <- em.norm(prelim.em1,showits=FALSE)
theta.init1 <- getparam.norm(prelim.em1,theta.init1,corr=TRUE)
theta.init1 <- makeparam.norm(prelim.em1,theta.init1)
trt1.em <- em.norm(prelim.em1,start=theta.init1,showits=TRUE,maxits=200,criterion=1e-5)
theta.em.trt1 <- getparam.norm(prelim.em1,trt1.em,corr=TRUE)
theta.em.trt1$mu       #  means
[1] 54.57851 51.09662 48.71815 45.52130 38.45595
#  covariance matrix
theta.em.trt1.cov <- diag(theta.em.trt1$sdv)%*%theta.em.trt1$r%*%diag(theta.em.trt1$sdv)
 theta.em.trt1.cov
         [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 217.8967 196.9365 189.0164 174.5998 139.3036
[2,] 196.9365 253.3823 233.9794 217.4068 173.7648
[3,] 189.0164 233.9794 300.1460 272.1957 230.9290
[4,] 174.5998 217.4068 272.1957 326.0305 282.1967
[5,] 139.3036 173.7648 230.9290 282.1967 351.1272
```

(h) The estimates from the available observed data likelihood analyses for each treatment group obtained from `gls` or `proc mixed` are virtually identical. This is entirely expected. Both analyses are doing the same thing – maximizing the observed data likelihood. `gls` and `proc mixed` are carrying out the maximization using standard optimization techniques. The EM algorithm is just another numerical technique to maximize the same likelihood.