Complex Data lab 1

Anna Zaleska

Task 1

Univariate format data:

id	y	time	time.cat
1	26.5	0	1
1	14.8	1	2
1	19.5	4	3
1	21.0	6	4
2	25.8	0	1
2	23.0	1	2

Summary of unstructured gls model:

```
## Generalized least squares fit by REML
##   Model: y ~ factor(time.cat)
##   Data: lead.uni
##        AIC      BIC    logLik
##   1308.337 1354.231 -640.1687
##
## Correlation Structure: General
##  Formula: ~1 | id
##  Parameter estimate(s):
##  Correlation:
##   1     2     3
## 2 0.401
## 3 0.384 0.731
## 4 0.495 0.507 0.455
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | factor(time.cat)
##  Parameter estimates:
##        1        2        3        4
## 1.000000 1.528090 1.563894 1.841550
##
## Coefficients:
##                     Value Std.Error   t-value p-value
## (Intercept)        26.540 0.7100685  37.37668       0
## factor(time.cat)2 -13.018 1.0309693 -12.62695       0
## factor(time.cat)3 -11.026 1.0638672 -10.36408       0
## factor(time.cat)4  -5.778 1.1378398  -5.07804       0
##
##  Correlation:
##                   (Intr) fc(.)2 fc(.)3
## factor(time.cat)2 -0.266
## factor(time.cat)3 -0.267  0.704
## factor(time.cat)4 -0.055  0.387  0.332
##
```

```
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.8020152 -0.8258034 -0.1227678  0.5184782  4.6654262
##
## Residual standard error: 5.020942
## Degrees of freedom: 200 total; 196 residual
```

Variance covariance matrix:

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 25.210 15.466 15.138 22.985
## [2,] 15.466 58.867 44.029 35.966
## [3,] 15.138 44.029 61.657 33.022
## [4,] 22.985 35.966 33.022 85.494
##   Standard Deviations: 5.0209 7.6725 7.8522 9.2463
```

Task 2

Summary of model when "ML" method is used:

```
## Generalized least squares fit by maximum likelihood
##   Model: y ~ factor(time.cat)
##   Data: lead.uni
##        AIC      BIC    logLik
##   1314.459 1360.635 -643.2294
##
## Correlation Structure: General
##  Formula: ~1 | id
##  Parameter estimate(s):
##  Correlation:
##   1     2     3
## 2 0.401
## 3 0.384 0.731
## 4 0.495 0.507 0.455
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | factor(time.cat)
##  Parameter estimates:
##        1        2        3        4
## 1.000000 1.528096 1.563892 1.841571
##
## Coefficients:
##                    Value Std.Error   t-value p-value
## (Intercept)        26.540 0.7100676  37.37673       0
## factor(time.cat)2 -13.018 1.0309793 -12.62683       0
## factor(time.cat)3 -11.026 1.0638743 -10.36401       0
## factor(time.cat)4  -5.778 1.1378407  -5.07804       0
##
##  Correlation:
##                   (Intr) fc(.)2 fc(.)3
## factor(time.cat)2 -0.266
## factor(time.cat)3 -0.267  0.704
## factor(time.cat)4 -0.055  0.387  0.332
##
```

```
## Standardized residuals:
##        Min         Q1        Med         Q3        Max 
## -1.8202919 -0.8341887 -0.1240145  0.5237409  4.7127447 
## 
## Residual standard error: 4.970473 
## Degrees of freedom: 200 total; 196 residual
```

Variance covariance matrix when "ML" method is used:

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 24.706 15.156 14.835 22.526
## [2,] 15.156 57.689 43.148 35.247
## [3,] 14.835 43.148 60.424 32.362
## [4,] 22.526 35.247 32.362 83.786
##   Standard Deviations: 4.9705 7.5954 7.7733 9.1535
```

We can observe that coefficients, variance multipliers and correlations derived from both models are the same. The difference in variance covariance matrices arises from different residual standard errors - σ_{11}.

Task 3

Based on the observations of estimated variance covariance matrices we can assume that variances at each coordinate are different. We do not observe any particular structure except their growth in time. Also correlations between variables vary. We should not use any simple AR correlation class due to differences in time gaps between measured observations.

Task 4

Let us see how the variance covariance matrix looks like if we assume equal variances (using varIdent()) weights.

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 57.154 24.226 22.812 24.093
## [2,] 24.226 57.154 41.108 26.098
## [3,] 22.812 41.108 57.154 23.499
## [4,] 24.093 26.098 23.499 57.154
##   Standard Deviations: 7.56 7.56 7.56 7.56
```

Let us now compare this model to unstructured REML model by likelihood ratio test.

```
##                  Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.id.var     1 11 1324.325 1360.384 -651.1625                 
## lead.cat            2 14 1308.337 1354.231 -640.1687 1 vs 2 21.98756
##                  p-value
## lead.cat.id.var         
## lead.cat          1e-04
```

We can see that p-value in likelihood ratio test is below 0.05. Thus we reject the hipotesis stating that there is no significant difference in both models' fit. Unstructed model explaines more variation of data.

Let us now test the model in which variances are powers of the time coeffitient (using varPower(form = ~time+1)).

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
```

```
## [1,] 32.233 14.660 18.917 25.833
## [2,] 14.660 44.602 38.422 28.417
## [3,] 18.917 38.422 68.520 32.385
## [4,] 25.833 28.417 32.385 80.222
##   Standard Deviations: 5.6774 6.6785 8.2777 8.9567
```

```
##                   Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.pow.var     1 12 1310.289 1349.627 -643.1447
## lead.cat             2 14 1308.337 1354.231 -640.1687 1 vs 2 5.951938
##                   p-value
## lead.cat.pow.var
## lead.cat            0.051
```

This time we performed better. We receive p-value just above the significance level so we have no basis to claim that models are signifficantly different.

What is interesting, if we fit model depending on time.cat coeffitient (using varPower(form = ~time.cat)) we get the following results.

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 28.110 14.526 16.913 24.442
## [2,] 14.526 48.578 40.488 31.074
## [3,] 16.913 40.488 66.897 33.274
## [4,] 24.442 31.074 33.274 83.947
##   Standard Deviations: 5.3019 6.9698 8.179 9.1622
```

```
##                    Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.pow.var2     1 12 1306.798 1346.135 -641.3989
## lead.cat              2 14 1308.337 1354.231 -640.1687 1 vs 2 2.460452
##                    p-value
## lead.cat.pow.var2
## lead.cat            0.2922
```

We get even bigger p-value from likelihood ratio test. But the assumpion that variations depend on the observation number (rather than observation time) does not seem reasonable.

Testing exponential dependance on time (varExp(form = ~time +1)), we reject the hipotesis thet there is no difference between models.

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 36.921 14.871 19.439 28.974
## [2,] 14.871 42.230 35.287 27.682
## [3,] 19.439 35.287 63.194 31.456
## [4,] 28.974 27.682 31.456 82.677
##   Standard Deviations: 6.0762 6.4985 7.9495 9.0927
```

```
##                   Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.exp.var     1 12 1312.620 1351.957 -644.3099
## lead.cat             2 14 1308.337 1354.231 -640.1687 1 vs 2 8.282431
##                   p-value
## lead.cat.exp.var
## lead.cat           0.0159
```

We also tried to put speciffic structure to correlations but the assumption that there is no structure seems mostly reasonable.

Test for compound symmetry (corCompSymm(form= ~1 | id)):

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 26.804 19.321 19.961 23.915
## [2,] 19.321 56.362 28.945 34.678
## [3,] 19.961 28.945 60.157 35.827
## [4,] 23.915 34.678 35.827 86.350
##   Standard Deviations: 5.1773 7.5074 7.7561 9.2925
```

```
##                   Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.cor.comp     1  9 1311.470 1340.973 -646.7351
## lead.cat              2 14 1308.337 1354.231 -640.1687 1 vs 2 13.13289
##                   p-value
## lead.cat.cor.comp
## lead.cat           0.0222
```

Test for exponential spatial correlation. (corExp(form= ~1 | id)):

```
## Marginal variance covariance matrix
##         [,1]   [,2]   [,3]    [,4]
## [1,] 26.9970 20.370 10.790  7.0357
## [2,] 20.3700 55.899 29.611 19.3080
## [3,] 10.7900 29.611 57.048 37.1980
## [4,]  7.0357 19.308 37.198 88.2170
##   Standard Deviations: 5.1959 7.4766 7.553 9.3924
```

```
##                  Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.cor.exp     1  9 1318.916 1348.419 -650.4579
## lead.cat             2 14 1308.337 1354.231 -640.1687 1 vs 2 20.57846
##                  p-value
## lead.cat.cor.exp
## lead.cat           0.001
```

Test for rational quadratics spatial correlation. (corRatio(form= ~1 | id)):

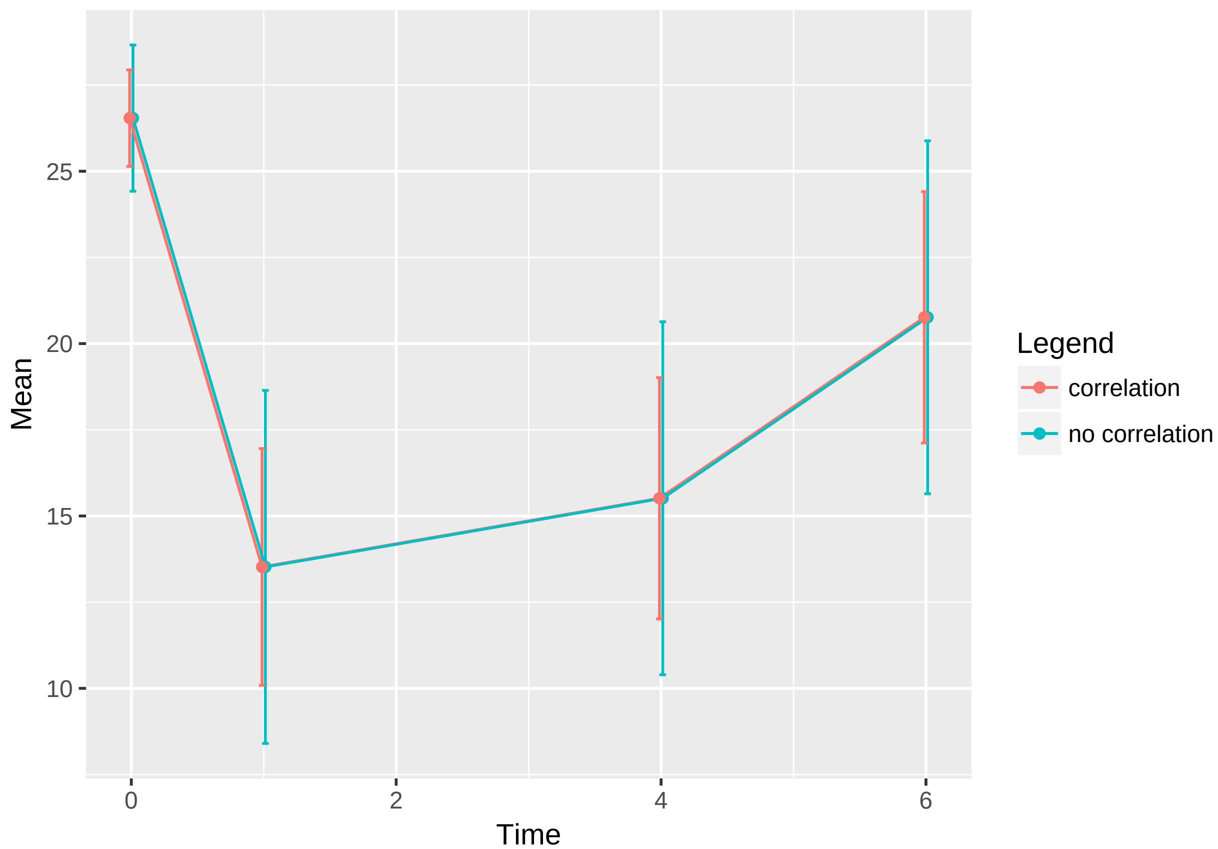
```
## Marginal variance covariance matrix
##         [,1]   [,2]    [,3]    [,4]
## [1,] 26.3770 18.461  7.3358  4.5998
## [2,] 18.4610 54.131 26.6320 13.2380
## [3,]  7.3358 26.632 54.8940 33.7850
## [4,]  4.5998 13.238 33.7850 87.1110
##   Standard Deviations: 5.1359 7.3574 7.4091 9.3333
```

```
##                  Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.cor.rat     1  9 1320.457 1349.960 -651.2286
## lead.cat             2 14 1308.337 1354.231 -640.1687 1 vs 2 22.11977
##                  p-value
## lead.cat.cor.rat
## lead.cat           5e-04
```

Task 5

The plot below presents mean values of observations in consecutive points of time together with their confidence intervals calculated with and without the respect of correlations. To extract confidence intervals we used methods: intervals from gls and confint from lm.

We can observe that confidence intervals taking correlations into account are narrower than the standard ones.



Task 6

To estimate the mean difference between 2nd and 3rd time points we will use anova function to test a contrast L=c(0,1,-1,0).

```
## Denom. DF: 196
##  F-test for linear combination(s)
## factor(time.cat)2 factor(time.cat)3
##                 1                -1
##   numDF  F-value p-value
## 1     1 6.111043  0.0143
```

p-value is below significance level, so we cannot conlude that the difference between time points is significant.