Complex data lab 2

Anna Zaleska

Task 1

1. what is L for this H_0 ? (p.3)
L = $(0, 0, 1, 0)$
2. What is L for H_0 : the mean blood lead level is the same at week 4 and week 6? (p.3)
L = $(0, 0, 1, -1)$
3. what is r for the H_0 in (1) above? (p.9)
r = 3. Output for the test:

```
lead <- read.table(file = "../data/lead.txt", header = FALSE)
names(lead) <- c("id", paste("y", 1:4, sep=""))
lead.uni <- data.frame(id=rep(lead$id, each=4),
  y=as.numeric(t(as.matrix(lead[,2:5]))),
  time=rep(c(0,1,4,6)),
  time.cat=rep(1:4))

lead.cat.ml <- gls(y~factor(time),
correlation=corSymm(form= ~1 | id),
weights=varIdent(form= ~1 | factor(time)),
method = "ML",
data=lead.uni)

lead.cat.no.ml <- gls(y~1,
correlation=corSymm(form= ~1 | id),
weights=varIdent(form= ~1 | factor(time)),
method = "ML",
data=lead.uni)

anova(lead.cat.ml, lead.cat.no.ml)
```

```
##                Model df      AIC      BIC    logLik   Test  L.Ratio
## lead.cat.ml        1 14 1314.459 1360.635 -643.2294
## lead.cat.no.ml     2 11 1381.866 1418.148 -679.9331 1 vs 2 73.40745
##                p-value
## lead.cat.ml
## lead.cat.no.ml  <.0001
```

4. However it should not be used to perform LRTs for nested models for mean. Why? (p.10)
REML can't be used to compare models with different fixed effects. REML's estimates of random effects are derived at the subsequence which is orthogonal to the one spanned by fixed effects. If two models have different projection matrices they are not nested (and not comprable anymore) and thus the likelyhood ratio test should not be performed.

5. What type of test was used? What is our conclusion for testing parallelism? (p.15)
We used Wald test for model parameters. From p-value for factor(diet):factor(week) which is above signifficance level we can conclude that there is no signifficant group by time effect - we cannot reject the null hipotesis.

6. What are our conclusions for testing main effects? What other type test could be used?

(p.16)
For all parameters p-value is under the signifficance level of 0.05 so we conclude that both time and diet are signifficant. We could use likelihood ratio test or score test.

Task 2

(a) The most general model

$X_{1ij} = 1$ for all measurements
$X_{2ij} = 1$ if the jth measurement was taken at $time = 2$ weeks, 0 otherwise
$X_{3ij} = 1$ if the jth measurement was taken at $time = 3$ weeks, 0 otherwise
$X_{4ij} = 1$ if the jth measurement was taken at $time = 4$ weeks, 0 otherwise
$X_{5ij} = 1$ if the ith cow ate barley and lupins, 0 otherwise
$X_{6ij} = 1$ if the ith cow ate only barley, 0 otherwise
$X_{7ij} = 1$ if the ith cow ate barley and lupins and its jth measurement is at $time = 2$, 0 otherwise
$X_{8ij} = 1$ if the ith cow ate barley and lupins and its jth measurement is at $time = 3$, 0 otherwise
$X_{9ij} = 1$ if the ith cow ate barley and lupins and its jth measurement is at $time = 4$, 0 otherwise
$X_{10ij} = 1$ if the ith cow ate only barley and its jth measurement is at $time = 2$, 0 otherwise
$X_{11ij} = 1$ if the ith cow ate only barley and its jth measurement is at $time = 3$, 0 otherwise
$X_{12ij} = 1$ if the ith cow ate only barley and its jth measurement is at $time = 4$, 0 otherwise
$i = 1, \ldots, 52$ cows
$j = 1, \ldots, 4$ protein measurements

Model:

$$Y_{ij} = \epsilon_{ij} + \beta_1 + \sum_{k=2}^{12} \beta_k X_{kij}$$

(b) Test for parallelism
We will fit the most general model under REML to test the significance of the diet x week interaction.

```
moo <- read.table(file = "../data/mooAll.txt", header = TRUE)
colnames(moo) <- c("protein", "week", "cow", "diet")

moo.gls.interaction <- gls(protein~factor(diet)*factor(week),
correlation=corSymm(form= ~1 | cow),
weights=varIdent(form= ~1 | factor(week)),
data=moo)
summary(moo.gls.interaction)
```

```
## Generalized least squares fit by REML
##   Model: protein ~ factor(diet) * factor(week)
##   Data: moo
##        AIC      BIC    logLik
##   108.8282 190.5304 -32.41412
##
## Correlation Structure: General
##  Formula: ~1 | cow
##  Parameter estimate(s):
##  Correlation:
##   1     2     3
## 2 0.440
## 3 0.474 0.485
## 4 0.321 0.515 0.600
## Variance function:
```

```
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | factor(week)
##  Parameter estimates:
##         1         2         3         4
## 1.0000000 0.6789112 0.6399626 0.6305705
##
## Coefficients:
##                                            Value  Std.Error  t-value
## (Intercept)                             3.885532 0.08052206 48.25425
## factor(diet)barley+lupins              -0.024420 0.11178982 -0.21845
## factor(diet)lupins                     -0.127383 0.11178982 -1.13949
## factor(week)2                          -0.246580 0.07526628 -3.27611
## factor(week)3                          -0.387148 0.07211824 -5.36824
## factor(week)4                          -0.510722 0.07988426 -6.39327
## factor(diet)barley+lupins:factor(week)2 -0.074531 0.10420488 -0.71523
## factor(diet)lupins:factor(week)2       -0.083790 0.10420488 -0.80409
## factor(diet)barley+lupins:factor(week)3 -0.128408 0.10012548 -1.28247
## factor(diet)lupins:factor(week)3        0.001963 0.10012548  0.01960
## factor(diet)barley+lupins:factor(week)4 -0.072611 0.11113362 -0.65337
## factor(diet)lupins:factor(week)4        0.046648 0.11113362  0.41975
##                                         p-value
## (Intercept)                              0.0000
## factor(diet)barley+lupins                0.8272
## factor(diet)lupins                       0.2554
## factor(week)2                            0.0012
## factor(week)3                            0.0000
## factor(week)4                            0.0000
## factor(diet)barley+lupins:factor(week)2  0.4750
## factor(diet)lupins:factor(week)2         0.4220
## factor(diet)barley+lupins:factor(week)3  0.2007
## factor(diet)lupins:factor(week)3         0.9844
## factor(diet)barley+lupins:factor(week)4  0.5140
## factor(diet)lupins:factor(week)4         0.6750
##
##  Correlation:
##                                         (Intr) fct()+ fctr() fct()2 fct()3
## factor(diet)barley+lupins              -0.720
## factor(diet)lupins                     -0.720  0.519
## factor(week)2                          -0.749  0.539  0.539
## factor(week)3                          -0.777  0.560  0.560  0.724
## factor(week)4                          -0.800  0.576  0.576  0.770  0.822
## factor(diet)barley+lupins:factor(week)2  0.541 -0.752 -0.390 -0.722 -0.523
## factor(diet)lupins:factor(week)2         0.541 -0.390 -0.752 -0.722 -0.523
## factor(diet)barley+lupins:factor(week)3  0.560 -0.777 -0.403 -0.522 -0.720
## factor(diet)lupins:factor(week)3         0.560 -0.403 -0.777 -0.522 -0.720
## factor(diet)barley+lupins:factor(week)4  0.575 -0.800 -0.414 -0.554 -0.591
## factor(diet)lupins:factor(week)4         0.575 -0.414 -0.800 -0.554 -0.591
##                                         fct()4 f()+:()2 f():()2 f()+:()3
## factor(diet)barley+lupins
## factor(diet)lupins
## factor(week)2
## factor(week)3
## factor(week)4
## factor(diet)barley+lupins:factor(week)2 -0.556
```

```
## factor(diet)lupins:factor(week)2            -0.556  0.522
## factor(diet)barley+lupins:factor(week)3 -0.592  0.728     0.377
## factor(diet)lupins:factor(week)3            -0.592  0.377     0.728   0.519
## factor(diet)barley+lupins:factor(week)4 -0.719  0.774     0.400   0.823
## factor(diet)lupins:factor(week)4            -0.719  0.400     0.774   0.426
##                                           f():()3 f()+:()4
## factor(diet)barley+lupins
## factor(diet)lupins
## factor(week)2
## factor(week)3
## factor(week)4
## factor(diet)barley+lupins:factor(week)2
## factor(diet)lupins:factor(week)2
## factor(diet)barley+lupins:factor(week)3
## factor(diet)lupins:factor(week)3
## factor(diet)barley+lupins:factor(week)4  0.426
## factor(diet)lupins:factor(week)4            0.823    0.517
##
## Standardized residuals:
##        Min          Q1         Med          Q3         Max
## -2.65092978 -0.66554849  0.04504866  0.69691506  2.77838379
##
## Residual standard error: 0.4029334
## Degrees of freedom: 315 total; 303 residual
```

Now we will test for parallelism.

$$H_0 : \beta_7 = \ldots = \beta_{12} = 0$$

$$H_1 : \text{at least one is non-zero}$$

```
anova(moo.gls.interaction)
```

```
## Denom. DF: 303
##                          numDF   F-value p-value
## (Intercept)                  1 19681.141  <.0001
## factor(diet)                 2     2.736  0.0664
## factor(week)                 3    49.492  <.0001
## factor(diet):factor(week)    6     1.152  0.3322
```

The output shows that p-value for $factor(week) : factor(diet)$ in the multivariate Wald test is 0.3322 which is above standard significance level of 0.05. We do not reject H_0 and conclude that there is no significant group by time effect.

(c) Test the main effects
    We will fit the model without considering iteractions:

$$Y_{ij} = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij}$$

and test

$$H_0^1 : \beta_5 = \beta_6 = 0,$$

$$H_0^2 : \beta_2 = \beta_3 = \beta_4 = 0.$$

```
moo.gls.fixed <- gls(protein~factor(diet) + factor(week),
correlation=corSymm(form= ~1 | cow),
weights=varIdent(form= ~1 | factor(week)),
data=moo)
summary(moo.gls.fixed)
```

4

```
## Generalized least squares fit by REML
##   Model: protein ~ factor(diet) + factor(week)
##   Data: moo
##        AIC      BIC    logLik
##    82.4687 142.2022 -25.23435
##
## Correlation Structure: General
##  Formula: ~1 | cow
##  Parameter estimate(s):
##  Correlation:
##   1     2     3
## 2 0.448
## 3 0.467 0.479
## 4 0.324 0.505 0.605
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | factor(week)
##  Parameter estimates:
##        1         2         3         4
## 1.0000000 0.6865158 0.6423451 0.6332031
##
## Coefficients:
##                              Value  Std.Error    t-value p-value
## (Intercept)               3.918225 0.05769067   67.91782  0.0000
## factor(diet)barley+lupins -0.112933 0.06020377   -1.87585  0.0616
## factor(diet)lupins        -0.133910 0.06020377   -2.22427  0.0269
## factor(week)2             -0.301282 0.04183371   -7.20190  0.0000
## factor(week)3             -0.430633 0.04068026  -10.58580  0.0000
## factor(week)4             -0.519752 0.04487244  -11.58287  0.0000
##
##  Correlation:
##                           (Intr) fct()+ fctr() fct()2 fct()3
## factor(diet)barley+lupins -0.543
## factor(diet)lupins        -0.543  0.520
## factor(week)2             -0.583 -0.001 -0.001
## factor(week)3             -0.609  0.002  0.002  0.722
## factor(week)4             -0.626  0.002  0.002  0.765  0.825
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.72722805 -0.70856383  0.08694052  0.70834687  2.89455153
##
## Residual standard error: 0.4012554
## Degrees of freedom: 315 total; 309 residual
```

```
anova(moo.gls.fixed)
```

```
## Denom. DF: 309
##              numDF   F-value p-value
## (Intercept)      1 19672.521  <.0001
## factor(diet)     2     2.726  0.0671
## factor(week)     3    49.096  <.0001
```

p-value for $factor(diet)$ is 0.0671, so we do not reject H_0^1 and do not confirm the hypothesis that diet is a significant factor here. We do reject H_0^2 due to low value of $factor(time)$ p-value.