# Stats 112 Homework 3

Owen Lin

# Problem 1

1a:

```
birthwt = read.table("birthwt-data.txt", header=FALSE)
names(birthwt) = c("MID", "Order", "Wt", "Age", "CID")
mod = lme(Wt ~ 1 + Age, random = ~ 1| MID, data = birthwt, method = "ML")
summary(mod)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: birthwt
##        AIC      BIC    logLik
##   67079.34 67104.89 -33535.67
##
## Random effects:
##  Formula: ~1 | MID
##         (Intercept) Residual
## StdDev:    354.6307 434.1899
##
## Fixed effects:  Wt ~ 1 + Age
##                 Value Std.Error   DF  t-value p-value
## (Intercept) 2785.1495  44.98839 3511 61.90819       0
## Age           17.1431   1.98003 3511  8.65803       0
##  Correlation:
##     (Intr)
## Age -0.953
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.04734611 -0.46642860  0.05826707  0.56462185  3.13023577
##
## Number of Observations: 4390
## Number of Groups: 878
```

1b:

$y_{ij} = b_{0i} + 2785.1495 + 17.1431 * Age_{ij} + \epsilon_{ij}$

The estimate of $\sigma_\epsilon$ is 434.1899.

1c:

The p-value is smaller than 0.05, so Age is a significant covariate of Wt and it should not be dropped from the model.

```
mod0 = lme(Wt ~ 1, random = ~ 1 | MID, data = birthwt, method = "ML")
anova(mod0, mod)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod0     1  3 67150.64 67169.80 -33572.32
## mod      2  4 67079.34 67104.89 -33535.67 1 vs 2 73.30363  <.0001
```

1d:

```
mod1 = lme(Wt ~ 1 + Age, random = ~ 1 + Age | MID, data = birthwt, method = "ML")
summary(mod1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: birthwt
##        AIC      BIC    logLik
##   67031.89 67070.21 -33509.94
##
## Random effects:
##  Formula: ~1 + Age | MID
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 437.9291 (Intr)
## Age          24.9588 -0.761
## Residual    426.4561
##
## Fixed effects:  Wt ~ 1 + Age
##                 Value Std.Error   DF  t-value p-value
## (Intercept) 2805.7881  46.68000 3511 60.10686       0
## Age           15.9444   2.23187 3511  7.14394       0
##  Correlation:
##     (Intr)
## Age -0.959
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.96974922 -0.45714553  0.05445116  0.56136668  3.32690915
##
## Number of Observations: 4390
## Number of Groups: 878
```

1e:

$$y_80 = -126.0165 + 2805.7881 + 15.9444 * Age_{80,j} + 21.6985 * Age_{80,j}$$

1f:

The random slope is needed on top of a random intercept as the p-value is less than 0.0001.

```
mod1.reml = lme(Wt ~ 1 + Age, random = ~1+Age | MID, data=birthwt, method="REML")
mod.reml = lme(Wt ~ 1 + Age, random = ~1 | MID, data=birthwt, method="REML")
anova(mod.reml, mod1.reml)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod.reml      1  4 67069.07 67094.62 -33530.53
## mod1.reml     2  6 67021.44 67059.76 -33504.72 1 vs 2 51.63041  <.0001
```

# Problem 2

Within a high school, we would expect most student to have a comparable socioeconomical background, and sex is something that was initially determined at birth. Thus, it is reasonable that the variance of Math Achievement score is constant across different students at the same high school. Also, if students are picked randomly, there is no reason to say student1 and student2 has a different correlation than student 1 and student 4. Then the corrlelation between any pair of score would be the same. That is the precise definition of a compound symmetry structure.

# Problem 3

3a:

```
NCGS = read.table("cholesterol-data.txt", na.strings=".")
names(NCGS) = c("Trt", "ID", "M0", "M6", "M12", "M20", "M24")
NCGS$Trt = factor(NCGS$Trt, levels=c(2,1), labels=c("Placebo","HighDose"))
NCGS.long = reshape(NCGS, varying=list(3:7), idvar="ID", timevar="Month", times=c(0,6,12,20,24),
v.names="Chol", direction="long")
mod.lme  = lme(Chol~Trt+I(Month)+I(Month^2), random=~I(Month)+I(Month^2)|ID, data=NCGS.long, na.
action=na.omit, method="ML") # 1+, the intercept is default
summary(mod.lme)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: NCGS.long
##        AIC      BIC    logLik
##   4364.344 4409.472 -2171.172
##
## Random effects:
##  Formula: ~I(Month) + I(Month^2) | ID
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev      Corr
## (Intercept) 40.87206089 (Intr) I(Mnt)
## I(Month)     1.10825273 -0.679
## I(Month^2)   0.04961792  0.425 -0.770
## Residual    22.33614017
##
## Fixed effects:  Chol ~ Trt + I(Month) + I(Month^2)
##                 Value Std.Error  DF  t-value p-value
## (Intercept) 230.04326  6.544943 342 35.14824  0.0000
## TrtHighDose   1.03269  7.794134 101  0.13250  0.8949
## I(Month)      2.29456  0.467767 342  4.90534  0.0000
## I(Month^2)   -0.05230  0.019363 342 -2.70116  0.0073
##  Correlation:
##             (Intr) TrtHgD I(Mnt)
## TrtHighDose -0.717
## I(Month)    -0.325  0.001
## I(Month^2)   0.237  0.005 -0.945
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.52995542 -0.59020108 -0.02986747  0.50313451  2.37731630
##
## Number of Observations: 447
## Number of Groups: 103
```

3b:

Treatment main effect is not significant with a high p-value of 0.8949.

3c:

```
getVarCov(mod.lme, type = "random.effects")
```

```
## Random effects variance covariance matrix
##             (Intercept)   I(Month) I(Month^2)
## (Intercept)  1670.50000 -30.772000  0.8615100
## I(Month)      -30.77200   1.228200 -0.0423430
## I(Month^2)      0.86151  -0.042343  0.0024619
##   Standard Deviations: 40.872 1.1083 0.049618
```

3d:

The estimated variance of the random intercepts is $1670.5^2 = 2790570$.

The estimated variances of the random slopes are $1.2282^2$ and $0.00246^2$.

The estimated correlation between the random intercepts and slopes are $\frac{-30.77200}{1670.5*1.228}$ and $\frac{0.86151}{1670.5*0.00246}$

3e:
The random effects for the subject with id = 96 are [-48.069, 0.625, -0.0229]

```
random.effects(mod.lme)[96,]
```

```
##    (Intercept)  I(Month)  I(Month^2)
## 96   -48.06953 0.6253606 -0.02291275
```

3f:
The estimated response for Month 6 for the subject with ID = 71 is 215.6939.

```
fitted(mod.lme)[names(fitted(mod.lme))=="71"]
```

```
##       71       71       71       71       71
## 201.1366 215.6939 224.7885 228.4168 226.5892
```

3g:
At a 5% significance level, we should not have random effects on the time covariates with a p-value of 0.2121 from the likelihood ratio test.

```
mod.lme.reml = lme(Chol~Trt+I(Month)+I(Month^2), random=~I(Month)+I(Month^2)|ID, data=NCGS.long,
na.action=na.omit, method="REML")
mod.lme.notime = lme(Chol~Trt+I(Month)+I(Month^2), random=~1|ID, data=NCGS.long, na.action=na.om
it, method="REML")
anova(mod.lme.notime, mod.lme.reml)
```

```
##                Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod.lme.notime     1  6 4358.967 4383.528 -2173.483
## mod.lme.reml       2 11 4361.850 4406.879 -2169.925 1 vs 2 7.116648  0.2121
```