Strathmore
UNIVERSITY

Institute of Mathematical Sciences

End of Semester Exams: Masters of Science (Statistical Sciences)

STA 8404: LONGITUDINAL DATA ANALYSIS

Sept-Dec 2018 Group

DATE: 19th December 2018 **TIME: 2 Hours**

INSTRUCTIONS

1. **Answer question one and any other two**
2. **All answers should be computed in R and scripts send to codhiambo@strathmore.edu before leaving exam room**

Question 1

a. Install the packages HSAUR, HSAUR2, lattice, scatterplot3d and ggplot2 and load the library of the packages. Attach BtheB data and generate two box plots with treatment as usual and Beat the blues. **(5 marks)**

b. Interpret you results in (b). **(4 marks)**

c. Install.packages("geepack") and load library("geepack"). Load the dataset ohio. Fit a logistic GEE model while treating time (age) as continuous. Use binomial link function and

   i. corstr = "exchangeable" **(3 marks)**

   ii. corstr = "unstructured" **(3 marks)**

d. Interpret the results in (c) **(3 marks)**

e. Test the effect of smoke for ohio using anova() i.e.

   i. by fitting fit1 <- geeglm(resp~factor(age)+smoke with correlation structure "exchangeable"

   ii. by fitting fit2 <- geeglm(resp~factor(age) with correlation structure "exchangeable"

   (6 marks)

f. By loading library("doBy"), generate individual Wald test and confidence interval for each parameter in e(i) **(6 marks)**

Question 2

Using library(HSAUR) and library(lattice) attach data epilepsy

a) Draw a dotplot(I(seizure.rate / base) ~ period | subject, data = epilepsy, subset = treatment == "Progabide") and interpret the output **(6 marks)**
b) Instal and load library("lme4") and by using BtheB data fit Fit_lmer1 = bdi ~ bdi.pre + time + treatment + drug + length + (1 | subject) and Fit_lmer2 =bdi ~ bdi.pre + time + treatment + drug + length + (time | subject) **(7 marks)**
c) Compare Fit_lmer1 and Fit_lmer2 using anova and discuss your results **(7 marks)**

Question 3

Using library(MASS) generate the following two longitudinal datasets

a) Data1 <- mvrnorm(n=250, mu=c(30, 20, 28), Sigma=matrix(c(25.0, 17.5, 12.3, 17.5, 25.0, 17.5, 12.3, 17.5, 25.0), nrow=3, byrow=TRUE)) **(3 marks)**
b) Data2 <- mvrnorm(n=250, mu=c(30, 20, 22), Sigma=matrix(c(25.0, 17.5, 12.3, 17.5, 25.0, 17.5, 12.3, 17.5, 25.0), nrow=3, byrow=TRUE)) **(3 marks)**
c) Merge the two data sets i.e. Data.T contains bothe Data1 and Data2 **(3 marks)**
d) names(Data.T) = c('measure.1', 'measure.2', 'measure.3') **(3 marks)**
e) Data.T <- data.frame(subject.id=factor(1:500), tx=rep(c('A', 'B'), each=250), Data.T)
f) Rewrite the dataset Data.T to long format **(8 marks)**

Question 4

Clearly interpret the following codes and results

```
model.f <- lme(alcuse ~ coa+cpeer*age_14 , data=alcohol1, random= ~ age_14 | id,
method="ML")
summary(model.f)
```

Linear mixed-effects model fit by maximum likelihood

Data: alcohol1

AIC BIC logLik

606.7033 638.2513 -294.3516

Random effects:

Formula: ~age_14 | id

Structure: General positive-definite, Log-Cholesky parametrization

StdDev Corr

(Intercept) 0.4908389 (Intr)

age_14 0.3730381 -0.034

Residual 0.5807690

*Fixed effects: alcuse ~ coa + cpeer * age_14*

Value Std.Error DF t-value p-value

(Intercept) 0.3938745 0.10460706 162 3.765276 0.0002

coa 0.5711970 0.14773612 79 3.866333 0.0002

cpeer 0.6951827 0.11240368 79 6.184696 0.0000

age_14 0.2705847 0.06189978 162 4.371336 0.0000

cpeer:age_14 -0.1513771 0.08538523 162 -1.772873 0.0781

Correlation:

(Intr) coa cpeer age_14

coa -0.637

cpeer 0.094 -0.146

age_14 -0.336 0.000 0.000

cpeer:age_14 0.000 0.000 -0.431 0.001

Standardized Within-Group Residuals:

Min Q1 Med Q3 Max

-2.59554441 -0.40414064 -0.08351877 0.45549526 2.29975184

Number of Observations: 246

Number of Groups: 82

(10 marks)

model.g <- lme(alcuse ~ ccoa+cpeer*age_14 , data=alcohol1, random= ~ age_14 | id, method="ML")

summary(model.g)

Linear mixed-effects model fit by maximum likelihood

Data: alcohol1

AIC BIC logLik

606.7033 638.2513 -294.3516

Random effects:

Formula: ~age_14 | id

Structure: General positive-definite, Log-Cholesky parametrization

StdDev Corr

(Intercept) 0.4908386 (Intr)

age_14 0.3730384 -0.034

Residual 0.5807690

*Fixed effects: alcuse ~ ccoa + cpeer * age_14*

Value Std.Error DF t-value p-value

(Intercept) 0.6514843 0.08060971 162 8.081959 0.0000

ccoa 0.5711970 0.14773606 79 3.866334 0.0002

cpeer 0.6951827 0.11240365 79 6.184698 0.0000

age_14 0.2705847 0.06189981 162 4.371334 0.0000

cpeer:age_14 -0.1513771 0.08538526 162 -1.772872 0.0781

Correlation:

(Intr) ccoa cpeer age_14

ccoa 0.000

cpeer 0.001 -0.146

age_14 -0.436 0.000 0.000

cpeer:age_14 0.000 0.000 -0.431 0.001

Standardized Within-Group Residuals:

Min Q1 Med Q3 Max

-2.59554430 -0.40414090 -0.08351925 0.45549506 2.29975158

Number of Observations: 246

Number of Groups: 82

(10 marks)

Question 5

Using Data_treeGrowth provided.

a) Generate and interpret boxplots for logsize against chamber and zone **(7 marks)**
b) Perform lmer fit without covariates but include i.e. logsize~ 1+ (1|tree) and another fit with covariates i.e. logsize~ days+chamber+year+(days|tree). Use anova to determine the best fit and interpret your results

(13 marks)