**Homework 5 Due on Monday by 1:10 pm, 4/08, 2019**

The data are from a longitudinal study to identify salient parent and adolescent psychosocial factors related to emotional distress in adolescents, assessed using the Global Severity Index (GSI). A total of 409 adolescents whose parents had HIV were included, and their GSI measurements were obtained over 6 years (mean = 63.5 months). (Reference of the study**:** B. Bursch et. al.(2008). Psychosocial Predictors of Somatic Symptoms in Adolescents of Parents with HIV: a Six-Year Longitudinal Study. *AIDS Care*, 20, 667-676.)

The data “**hw5p1\_data**” is uploaded on blackboard. The data has been cleaned so all the missing values have been removed. The following variables are included in the data.

* ID: adolescent’s ID number
* Parent\_ID: parent’s id number
* GSI: Global severity index (outcome, continuous)
* True\_month: actual time since baseline (months)
* Gender: adolescent’s gender
* Hispanic: whether adolescent is Hispanic (not Hispanic vs. hispanic)
* Treatment: whether family was in the treatment =1 or control =0 group
* Season: 3 categories separated into months (spring=3-6, summer=7-10 ,winter=11-2)
* Parent\_base\_age: parent’s baseline age
* Parent\_died: whether parent died during the study(Yes/No)
* Parent\_gender: adolescent’s parent’s gender (male/female)
* Parent\_drug\_status: Parental hard drug use during the last 3 months:
* nonuser: parent never used hard drugs during the study
* nonusing-user: parent used hard drugs during the study, but not in the last 3 months
* using-user: parent used hard drugs during the last 3 months
* Parent\_alcohol: whether parent drank alcohol in the last 3 month (no=0, yes=1)
* Parent\_marijuana: whether parent smoked marijuana in the last 3 months (no=0,yes=1)
* Parent\_diagnosis: 3 level parent self-reported baseline illness level (asymptomatic, symptomatic, diagnosed\_with\_AIDS)

Perform a logarithm transformation with a base of 2 on the outcome variable GSI as follows:

**data** one; set hw5p1\_data; logGSI=log2(GSI + **1**/**53**); **run**;

The following variables are **time-dependent**: GSI, True\_month, Season, Parent\_died, Parent\_drug\_status, Parent\_alcohol, and Parent\_marijuana.

**Questions:**

1. Use the model building strategies to build a model for these data. Due to the presence of multiple time-dependent variables, assess the random slopes for significant variables only. If the test for random slope is not significant then it is necessary to keep the slope fixed. When comparing two models, if the likelihood ratio test is not significant then choose the simpler model.
2. From the final model, use the type 3 analysis to find significant effects and interpret them.

Data overview:

Level 1 variable: True\_month, Season, Parent\_died, Parent\_drug\_status, Parent\_alcohol, and Parent\_marijuana.

Level 2 variable: Gender, Hispanic.

Level 3 variable: Parent\_base\_age, Parent\_gender and Parent\_diagnosis Treatment.

**Model 1: intercept only model:**

/\* 1. Intercept only model\*/

**PROC** **MIXED** DATA = HW5.dat NOCLPRINT METHOD=ML COVTEST NOITPRINT;

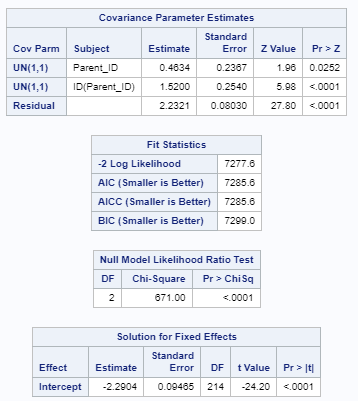
CLASS ID Parent\_ID;

MODEL logGSI = / solution ddfm=betwithin;

random intercept / subject=Parent\_ID type=UN;

random intercept / subject=ID(Parent\_ID) type=UN;

**RUN**;



**Model 2: Adding level 1 Time-dependent Variable**

/\*Step 2: Adding Time-dependent Variable\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

class ID Parent\_ID season Parent\_died Parent\_drug\_status;

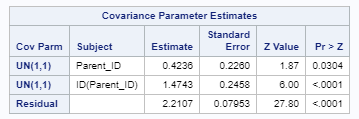
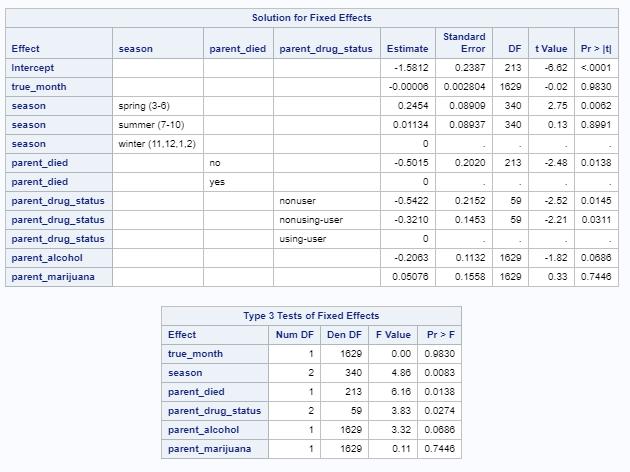
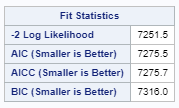
model logGSI = True\_month Season Parent\_died

Parent\_drug\_status Parent\_alcohol Parent\_marijuana/ solution ddfm=bw ;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

**run**;

Since deviance test gives a p-value less than 0.05, this model is significantly better than the intercept only model.

**Model 3: Testing Random Slope for Signifciant Variables in Step 2**

Since the G matrix is not positive significant if I add season as the random slope, we did not add it in the random slope part.

/\*Step 3: Testing Random Slope for Signifciant Variables in Step 2 \*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

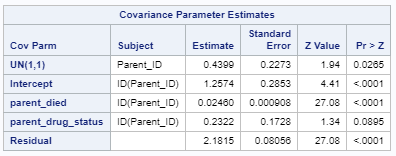
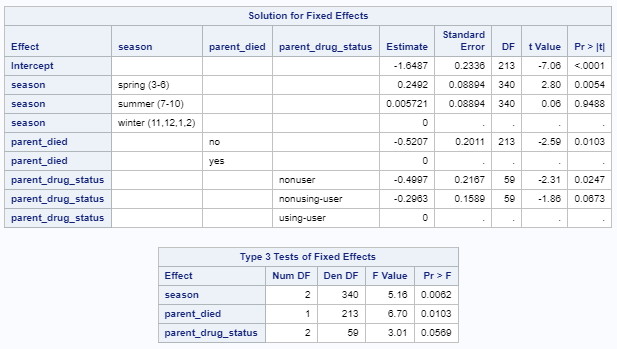
class ID Parent\_ID season Parent\_died Parent\_drug\_status;

model logGSI = Season Parent\_died Parent\_drug\_status / solution ddfm=bw ;

random intercept / subject=Parent\_ID type=un;

random intercept Parent\_died Parent\_drug\_status / subject=ID(Parent\_ID);

**run**;

Since the -2loglikelihood was increased, it suggests that the model fit did not increase after adding random slopes. Model 2 was better than model 3 and we did not even need to use deviance test.

**Model 4: adding level 2 variables**

/\*Step 4: Adding Level-2 Variables\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

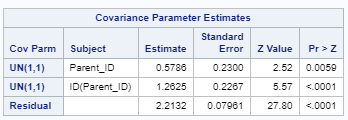
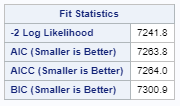
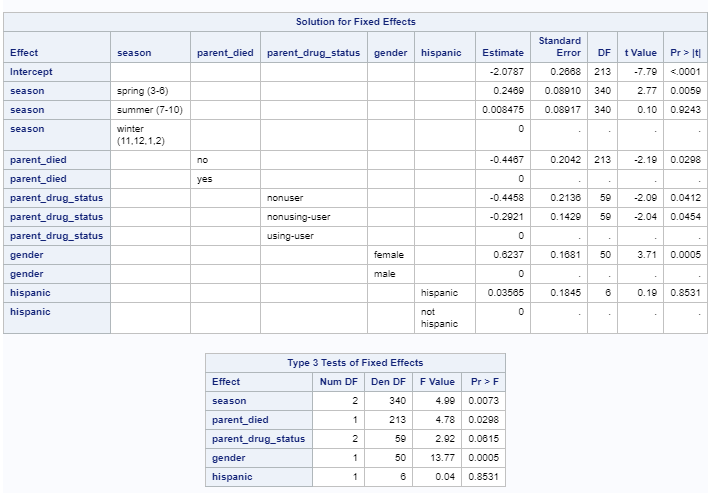
class ID Parent\_ID season Parent\_died Parent\_drug\_status Gender Hispanic;

model logGSI = Season Parent\_died Parent\_drug\_status Gender Hispanic/ solution ddfm=bw ;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

**run**;

The deviance test produces a p-value less than 0.05, so this model is significantly better than model 2.

**Model 5: Adding Level-3 Variables**

/\*Step 5: Adding Level-3 Variables\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

class ID Parent\_ID season Parent\_died Parent\_drug\_status Gender

Parent\_gender Parent\_diagnosis;

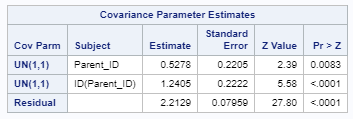
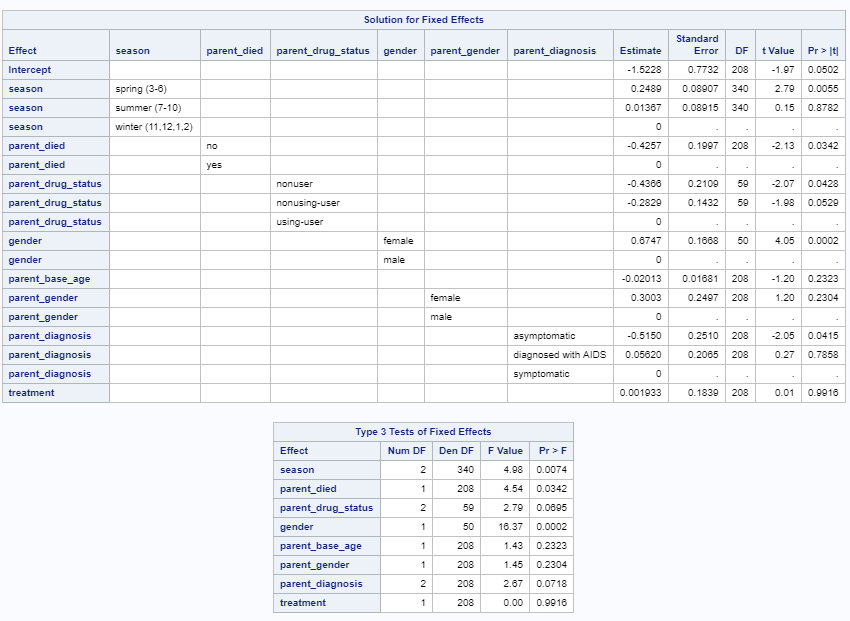
model logGSI = Season Parent\_died Parent\_drug\_status Gender

Parent\_base\_age Parent\_gender Parent\_diagnosis Treatment/ solution ddfm=bw ;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

**run**;



Since all level 3 variables are not significant, and the deviance test suggests that model 5 is no better than model 4. Model 4 is a better model than model 5.

**Model 6: test interactions between levels**

1) interaction between Season and Gender

/\*Step 6: Testing the Interaction Between Season and Gender\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

class ID Parent\_ID season Parent\_died Parent\_drug\_status Gender

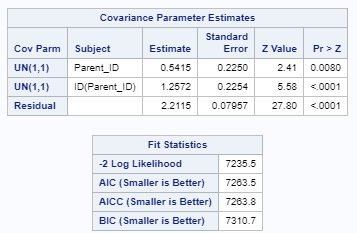
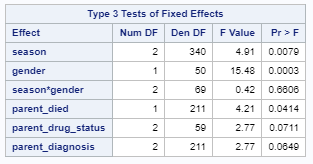
Parent\_gender Parent\_diagnosis;

model logGSI = Season|Gender Parent\_died Parent\_drug\_status Parent\_diagnosis/ solution ddfm=bw;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

**run**;

The type 3 effect for the interaction was not significant, neither does the p-value from the deviance test. Therefore, this model is no better than model 4.

2) Interaction between Parent\_died and Gender

/\*Step 6: Testing the Interaction Between Parent\_died and Gender\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

class ID Parent\_ID season Parent\_died Parent\_drug\_status Gender

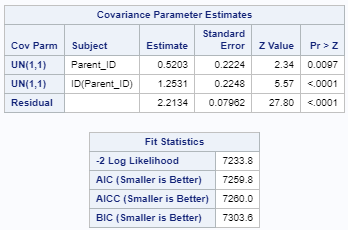
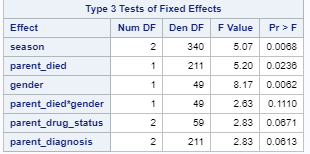
Parent\_gender Parent\_diagnosis;

model logGSI = Season Parent\_died|Gender Parent\_drug\_status Parent\_diagnosis/ solution ddfm=bw;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

**run**;

The type 3 effect for the interaction was not significant, neither does the p-value from the deviance test. Therefore, this model is no better than model 4.

3) Interaction Between Parent\_drug\_status and Gender

/\*Step 6: Testing the Interaction Between Parent\_drug\_status and Gender\*/

**proc** **mixed** data=HW5.dat noclprint method=ml covtest noitprint;

class ID Parent\_ID season Parent\_died Parent\_drug\_status Gender

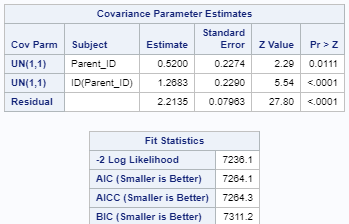
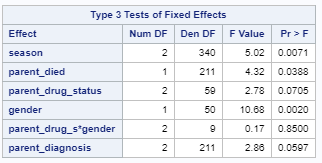
Parent\_gender Parent\_diagnosis;

model logGSI = Season Parent\_died Parent\_drug\_status|Gender Parent\_diagnosis/ solution ddfm=bw;

random intercept / subject=Parent\_ID type=un;

random intercept / subject=ID(Parent\_ID) type=un;

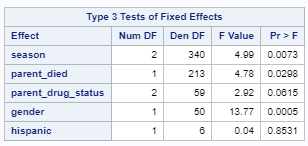
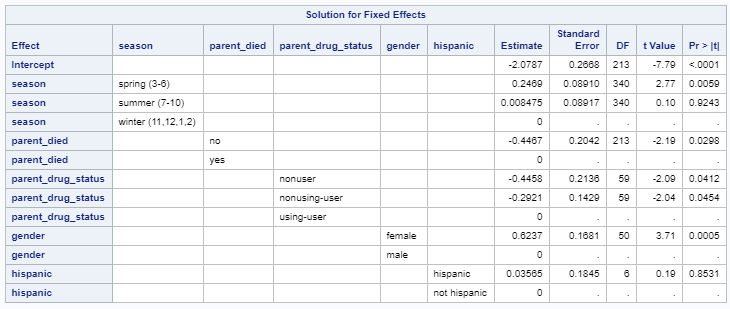
**run**;

The type 3 effect for the interaction was not significant, neither does the p-value from the deviance test. Therefore, this model is no better than model 4.

In conclusion, the best fit model is Model 4.

**Type 3 analysis:**

Type 3 analysis suggests that the significant variables are season, patient\_died, and gender.

1. Compared with winter, spring is positive associated with the log of GSI score.
2. Children with their parents live have lower GSI score.
3. Female children have higher GSI score, compared with male children.