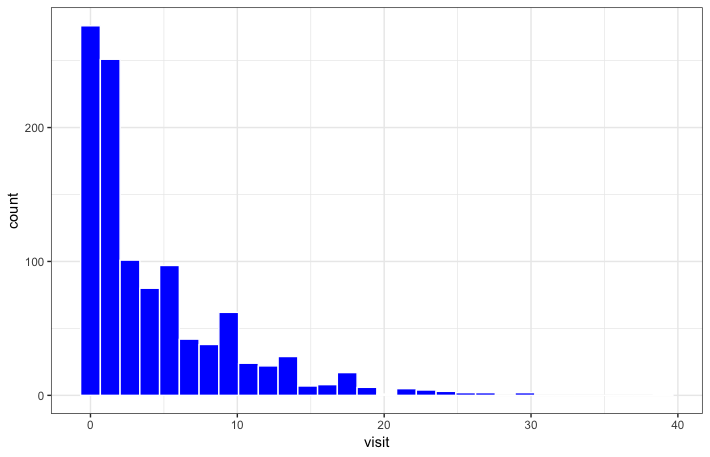
Predicting the number of visits to physician using a hierarchical Poisson model: A case in the General Hospital of the city of St. John’s, Canada

BST 5220 Multilevel and Longitudinal Analysis Final Project

Cheng-You Tsai, Miao Cai

Instructor: Dr. Hong Xian

May 6, 2019

Figure 1. Count data of visit

Yes, the plot appears to be right-skewed.

**1.Random-intercept only model**

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class family\_ID ID;**

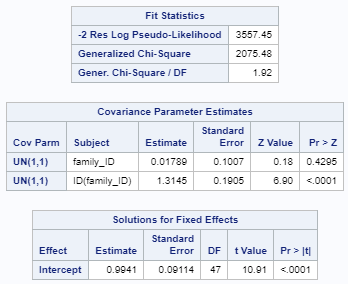
**model visit = / solution dist = poisson link = log;**

**random int / subject=family\_ID type=un;**

**random int / subject=ID(family\_ID) type=un;**

**COVTEST / WALD;**

**run;**



Since the intercept at the family level is not significant (estimate: 0.01789, P-value = 0.4295), we excluded the random intercept at this level.

**2. Exclude insignificant random intercepts at family level**

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

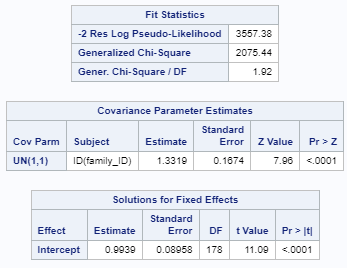
**class ID family\_ID;**

**model visit = / solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=un;**

**COVTEST / WALD;**

**run;**



The level 2 (individual ID) random intercepts are significant (P-value = 0.029). Therefore, we keep the individual ID in the model. However, the generalized Chi-square/df = 1.92, suggesting overdispersion or underfit of the model.

**3. Adjust for overdispersion**

We added a scale parameter to the model – the R-side variance through the “**random \_residual\_ /subject=ID(family\_ID)**” statement

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID;**

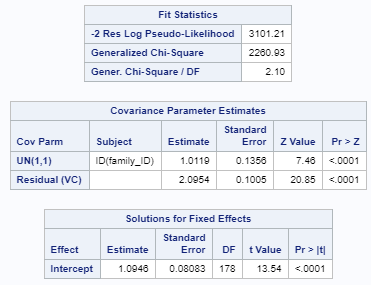
**model visit = / solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=un;**

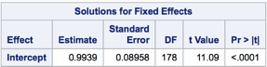
**random \_residual\_ /subject=ID(family\_ID);**

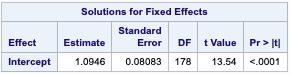
**COVTEST / WALD;**

**run;**



* R-side variance = 2.10 : the overdispersion parameter
* The level-2 variance = 1.01
* Both level-1 and level-2 variance are significantly different from 0

 Model parameter without the scale parameter

Model parameter with the scale parameter

* At the third step, we can see that the model parameter results with and without the scale parameter are very similar

According to - 2 Res Log Pseudo-Likelihood, after adjusting for overdispersion, Step 3 model is significantly better than step 2.

**4. Adding level-1 variable - Year**

procglimmix **data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID;**

**model visit = year / solution dist = poisson link = log;**

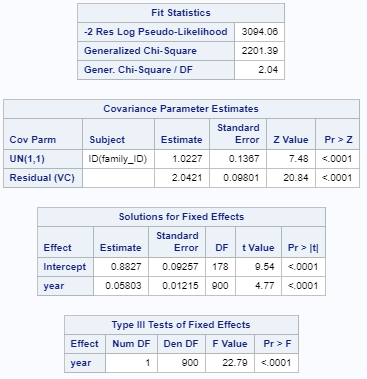
**random int year / subject=ID(family\_ID) type=UN;**

**random \_residual\_ /subject=ID(family\_ID);**

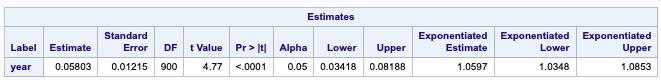
**estimate 'year' year** 1 **/exp cl;**

**COVTEST / WALD;**

run**;**



* Level-1 variance reduced from 2.10 to 2.04
* Level-2 variance slightly increased (from 1.01 to 1.02)



* For every one additional year the number of visits to physician is individual increased by 5.97 %

**5. Add year random slope effect**

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID;**

**model visit = year/ solution dist = poisson link = log;**

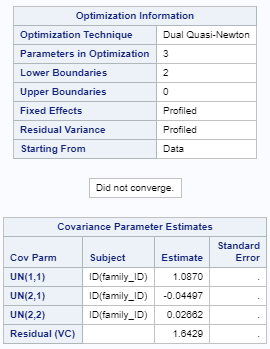
**random int year/ subject=ID(family\_ID) type=un;**

**random \_residual\_ /subject=ID(family\_ID);**

**Estimate 'year' year 1/expcl;**

**COVTEST / WALD;**

**run;**



After adding time random slopes, the algorithm cannot converge, so we exclude the year random effects and only use year as fixed effects.

**6. Add level 2 variables**

**proc glimmix data=F.FinalProjectData\_bst522 noitprint noclprint;**

**class ID family\_ID gender (ref=first) education (ref=first) ;**

**model visit = year gender chronic education age\_base / solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=UN;**

**random \_residual\_ /subject=ID(family\_ID);**

**estimate 'year' year 1 /exp cl;**

**estimate 'F vs M' gender 1 -1 /exp cl;**

**estimate 'chronic' chronic 1 /exp cl;**

**estimate 'high school vs less than high school' education 1 0 0 -1 /exp cl;**

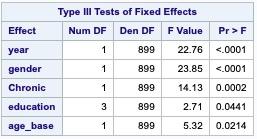
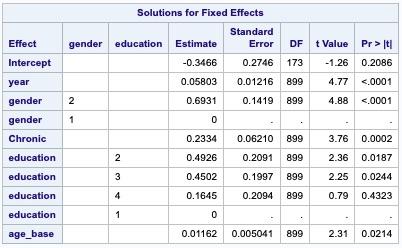
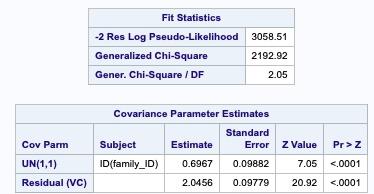
**estimate 'university graduate vs less than high school' education 0 1 0 -1 /exp cl;**

**estimate 'post graduate vs less than high school' education 0 0 1 -1 /exp cl;**

**estimate 'age\_base' age\_base 1 /exp cl;**

**COVTEST / WALD;**

**run;**



* Level-2 variance reduced from 1.02 to 0.70
* Level-1 variance (2.04) remained the same

All the type 3 tests of fixed effects are significant for the level 2 variables.



Controlling/adjusted for all other factors,

* The number of visits to physician for every additional year is multiplied by exp(0.06) = 1.06
* The number of visits to physician for female is two time that for male.
* For every unit change in chronic status the number of visits to a physician by each individual is multiplied by exp(0.23)=1.26
* The number of visits to a physician by each individual for people having an education of high school is exp(0.49)=1.64 times that for less than high school.
* The number of visits to a physician by each individual for people having an education of university graduate is exp(0.45)=1.57 times that for less than high school.
* For every unit change in age\_basethe number of visits to a physician by each individual is multiplied by exp(0.01)=1.01

**7. Testing interaction between level 1 and level 2 variables**

1. The interaction between year and gender

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID gender education(ref='1');**

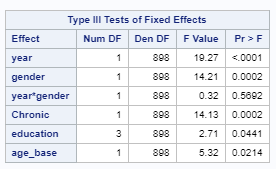
**model visit = year gender chronic education age\_base/ solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=un;**

**random \_residual\_ /subject=ID(family\_ID);**

**COVTEST / WALD;**

**run;**



The type 3 effects for the interaction between year and gender was not significant. Therefore, this model is not better than model 6.

1. The interaction between year and chronic disease

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID gender education(ref='1');**

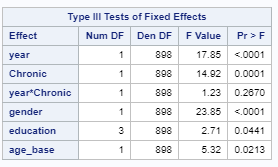
**model visit = year|chronic gender education age\_base/ solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=un;**

**random \_residual\_ /subject=ID(family\_ID);**

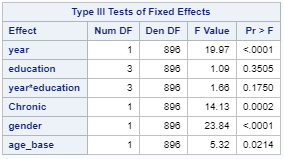
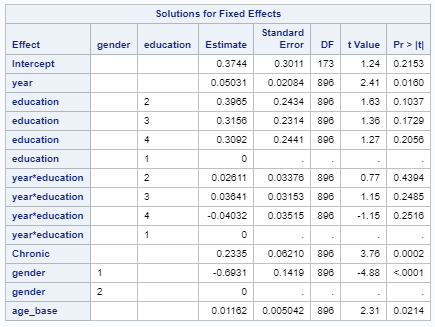
**COVTEST / WALD;**

**run;**



The type 3 effects for the interaction between year and chronic disease was not significant. Therefore, this model is not better than model 6.

1. The interaction between year and education



The type 3 effects for the interaction between year and education was not significant. Therefore, this model is not better than model 6.

1. The interaction between year and age\_base

**proc glimmix data=F.finalprojectdata\_bst522 noitprint noclprint;**

**class ID family\_ID gender education(ref='1');**

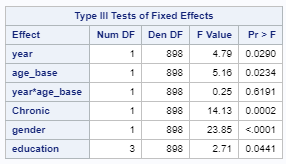
**model visit = year|age\_base chronic gender education / solution dist = poisson link = log;**

**random int / subject=ID(family\_ID) type=un;**

**random \_residual\_ /subject=ID(family\_ID);**

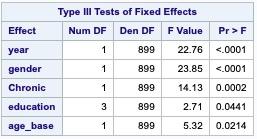
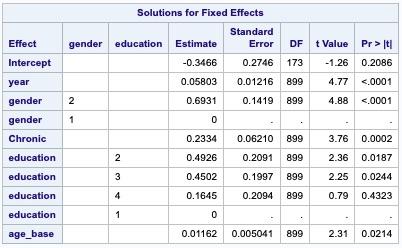
**COVTEST / WALD;**

**run;**



The type 3 effects for the interaction between year and age\_base was not significant. Therefore, this model is not better than model 6.

**Therefore, the final model is at step 6.**



Type 3 analysis suggests that the significant variables are year, gender, chronic disease, education, and age\_base.



Controlling/adjusted for all other factors,

* The number of visits to physician for every additional year is multiplied by exp(0.06) = 1.06
* The number of visits to physician for female is two time that for male.
* For every unit change in chronic status the number of visits to a physician by each individual is multiplied by exp(0.23)=1.26
* The number of visits to a physician by each individual for people having an education of high school is exp(0.49)=1.64 times that for less than high school.
* The number of visits to a physician by each individual for people having an education of university graduate is exp(0.45)=1.57 times that for less than high school.
* For every unit change in age\_basethe number of visits to a physician by each individual is multiplied by exp(0.01)=1.01