Application of Longitudinal Data for Multilevel Models Approach on Diabetes Mellitus Disease

1. Wudneh Ketema Moges , Debre Berhan University, Department of Statistics , Ethiopia
   Email: wudnehketema@gmail.com
2. Tena Manaye Endalamaw, Debre Berhan University, Department of Statistics , Ethiopia

Email: tenamanaye@gmail.com

Abstract:

As multilevel models, hierarchical models and individual growth models increase in popularity, the need for credible and flexible software that can be used to fit them to data increases. To determine the prevalence of diabetic mellitus patients and identify associated risk factors using multilevel longitudinal model and understand multilevel model changes for level-1 and level-2 models. Multilevel models were developed for analyzing hierarchically structured data. Some examples of hierarchically structured data are students nested within schools and employees nested within companies. The large proportion of the variability in the by follow up time explained by diabetes mellitus emphasizes the importance of accounting for the hierarchical structure of the data. The fixed effects $\hat{\gamma}_{00}$and $\hat{\gamma}_{10}$ estimates the starting point(y-intercept) and slope of the population average change trajectory for time points. The parameters are significant (t-value of 206.52 and 0.75 respectively) which indicates that they should both be included in the model.

For a two-level longitudinal MLM, the software requires input of values regarding the duration of the study, frequency of observations, the level 1 variance , the between –person variability in the parameter of interest and an estimates of the effect size.

Keywords: Diabetes, Multilevel, Longitudinal ,ANOVA , MACOVA.

1. INTRODUCTION

Diabetes has become one of the most common chronic non-communicable diseases both in developed and developing countries over the past several decades. According to the International Diabetes Federation (IDF), there were 382 million people living with diabetes worldwide in 2013, with that number projected to reach 592 million by 2035 [1]. Due to the large economic burden of diabetes and its adverse impact on health, diabetes has become one of the foremost public health challenges of the 21st century [2].

Diabetes screening identifies individuals with undiagnosed, asymptomatic type 2 diabetes or prediabetes that are eligible for evidence –based interventions to prevent or delay type 2 diabetes and its complications. National screening guidelines, such as the American Diabetes Association (ADA) [3] and U.S. Preventive Services Task Force (USPSTF) [4], help clinical practice. Despite screening recommendations,>7 million U.S. adults with type 2 diabetes and 74 million U.S. adults with prediabetes remain undiagnosed [5]. Nationally representative data indicates

significant gaps between screening eligibility and screening completion, with only half of individuals ADA and USPSTF guidelines reporting a completed screening test [4, 5].

In the presence of follow up measurements per subject and the interest of modeling changes of response variables over time, longitudinal modeling techniques are required to appropriately account for the dependence that exists among repeated measurements per subject (Liang & Zeger, 1986). Using cross-sectional models, which are appropriate only for data with a single measurement per subject, while dealing with longitudinal data is inappropriate and will result in unreliable conclusions. Due to the complexity introduced to the study because of the correlated nature of the longitudinal data, more advanced models are required to account for the dependence between multiple outcome values observed for each subject at different time points.

As multilevel models, hierarchical models and individual growth models increase in popularity, the need for credible and flexible software that can be used to fit them to data increases.

The analysis of longitudinal data has the primary purpose of examining the covariate effects of each level in general on the responses and effects in response changes over time. One of the important benefits of longitudinal data analysis is that it allows the separation of cross-sectional and repeated measurement effects. In addition, in the analysis of longitudinal data can also be examined the diversity between units both within the level of the response and in changes between times. The variation not obtained from the observed variation results in a dependency between responses as well after the covariates are controlled. This violates assumptions in ordinary linear regression models and must be overcome to avoid fallacy in inference.

2. Objective

To determine the prevalence of diabetic mellitus patients and identify associated risk factors using multilevel longitudinal model and understand multilevel model changes for level-1 and level-2 models.

3. Methodology

Data : All Diabetes mellitus patients who were both Type I and Type II, and placed under insulin and metformin follow up the case unit of 1 st September, 2012- 30th August, 2015 G. C, in Debre Berhan referral hospital for a period of three years. Longitudinal data were collected from 248 diabetic were collected in Debre Berhan referral hospital.

Patients were measured four times:

- ✓ When the program was initiated at baseline (time T_0)
- ✓ After six 6 months (time T_1)
- ✓ After 12 months (time T_2)
- ✓ After 18 months (time T_3)

Multilevel models are used to analyze data that are clustered in some way. In this work, multilevel models are used to analyze longitudinal data from a case study.

Multilevel Models

Analyzing data sets that contain variables measured at different levels of hierarchy is known as multilevel modeling. In a multilevel data set subjects in the same level or cluster may be more similar to one another than subjects in other levels or clusters. Multilevel models have extensive use in social sciences. They are more generally used when group level effects need to be analyzed.

Simple Random intercept multilevel models

A random intercept model allows the intercept to vary across different clusters. When we consider a simple linear regression model, one intercept is common between all observations in a data set.

Random slope coefficient multilevel model

Random slope coefficient models are obtained by expanding the simple intercept-only model. This is done by adding independent predictor variables at the individual level (level-1) to simple intercept-only models.

Multilevel models for Longitudinal Data

Multilevel models were developed for analyzing hierarchically structured data. Some examples of hierarchically structured data are students nested within schools and employees nested within companies. Therefore, a hierarchy consists of lower-level observations (individual-level data) nested within higher levels (group-level data).Analysis of models that contain variables measured at different levels of the hierarchy are known as multilevel models [1]. In the case of longitudinal data, multilevel models are useful in the analysis of within person and between person changes by distinguishing two things: how individuals change over time and, how these changes vary across individuals [2].

In the study, we will analyze the diabetic data (from the longitudinal data section) from a multilevel point of view. Here, we will explore whether individual change in glucose level
And time varying to them.

Multilevel Model for Longitudinal Data

Studying diabetic patients over time

Level 1: Within-Person (WP) Variation: that means "INTRA-individual Differences" – Time-Varying. Only longitudinal studies can provide this extra information

Level 2: Between-Person (BP) Variation: "INTER-individual Differences" Time-Invariant. All longitudinal studies begin as cross-sectional studies

Longitudinal studies allow examination of both types of relationships simultaneously (and their interactions).

Multilevel models were developed for analyzing hierarchically structured data. A hierarchy consists of lower –level observations (individual –level data) nested within higher levels (grouped –level data). Analyses of models that contain variables measured at different levels of

the hierarchy are known as multilevel models. In the case longitudinal data, multilevel models are useful in the analysis of within person and between person changes by distinguishing two things: how individual change over time and how these changes vary across individuals. Therefore, we will analyze the diabetes mellitus data (from the longitudinal data perception) from a multilevel point of view. Here, we will explore whether individual trajectories of change in fasting glucose level

4. Analysis of the Result

There were 248 individual with Diabetes mellitus patients under the study are who had for at least one time point. Among these 251 individual, 211 had complete data for all four time points (i.e. baseline, at month 6, month 12 and month 18 of diabetes patients).
Under the descriptive statistics for the given samples are presented in Table 1. We conducted on all socio demographic variables. The functional status of diabetes mellitus patients about 236(95.16%) who had working diabetes mellitus patients whereas 12(4.86%) who had ambulatory diabetes mellitus patients; this indicates that majority of diabetes mellitus patients categorized under working status. Under the clinical diagnosis DM patients accounted under this study about 159(64.01%) were Type I whereas 89(35.99%) were Type II DM patients.

Table 1: Socio Demographic Characteristics of Diabetes Mellitus Patients Debre Berhan Referral Hospital

Variables		Frequency	Percent
Sex	Male	137	55.24%
	Female	111	44.76%
Functional status	Working	236	95.16%
	Ambulatory	12	4.86%
Marital status	Married	194	78.23%
	Single	54	21.77%
Address	Urban	159	64.11%
	Rural	89	35.89%
Educational status	Illiterate	87	35.08%
	Primary	67	27.02%
	Secondary	44	17.74%
	Tertiary	50	20.16%
Occupation	Full -time	137	55.24%
	No-working	111	54.76%
Clinical diagnosis	Type I	159	64.01%
	Type II	89	35.99%

From Table 2, the mean of fasting blood sugar level of patients increases with an increasing rate until at time two, then decreases time three. The largest value of standard deviation with a

baseline time point with the value was 94.80 compare from the other time point, so the number of measurements for fasting blood sugar level count showed increasing and decreasing observations between follow up times for the response indicating that the data have both intermittent and dropout missing observations and also, missing value was increasing over time.

As we observed the 95% CL for given time point, the narrowest confidence interval accounted under month 18 (time point 3) and at bassline time point whereas the widest 95% CL were found at month 12 (time point 2) and at month 6 (time point 1). Based on the result we concluded that the smallest fasting glucose level under diabetes mellitus patients for within variation at month 18 (time point 3) and at bassline time point.

Table 2: Descriptive Statistics of Continuous Response by Time

Time	Baseline	1	2	3
N	248(100%)	232(93.5%)	239(96.41%)	236(95.2%)
Mean of Fasting	179.53	189.31	190.38	184.87
St DV	94.800	92.950	96.855	89.998
Max.	600	574	600	587
Min.	22	22	29	31
95% CL	(167.68,191.34)	(177.28,201.33)	(178.04,202.73)	(173.33,196.41)
Missing	0	16	9	12

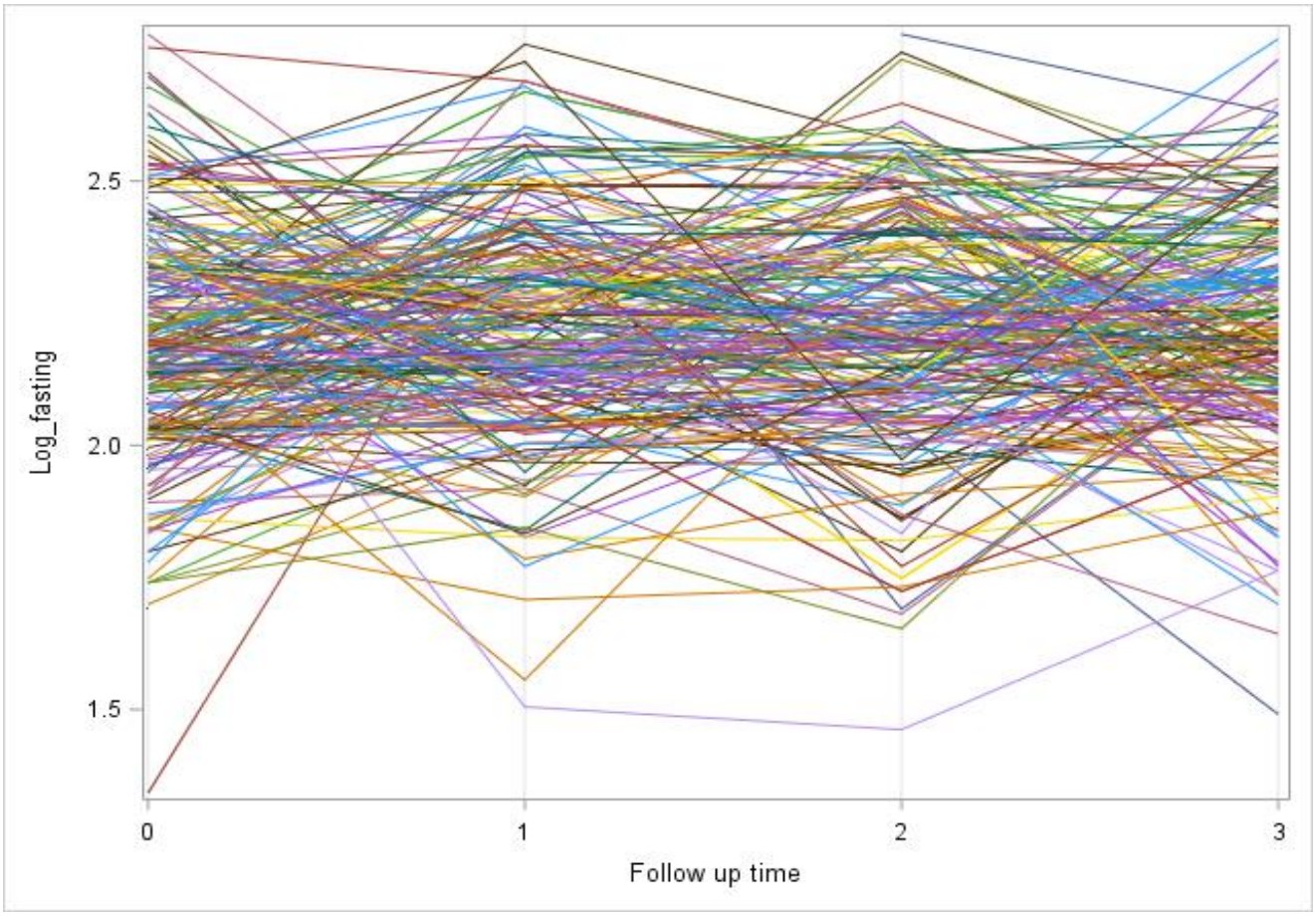

Figure 1: Predicted Individual Growth Curves Random Intercept Model

From the above Figure 1, we depicted that the shape of the response –profile curve is approximately the same for all subjects for each follow up time .However; the curves are shifted up or down to better match the subject individual profile.

The Null model or Unconditional Means Model

The results of this model are helpful in analyzing the partitions in outcome variation i.e it distinguishes between the individual variation and within group variation.

Level 1 Model: Log Fasting = β_0+ intercept

Level 2 Model: $\boldsymbol{\beta_0=\gamma_{00}+u_0}$

SAS Result

```
Proc mixed data =dm covtset;
Model lo_fasting=/ solution;
Raadom intercept /subject =class type=un;
Run;
```

The "covtest" option is needed to report the standard errors of the variance component estimates. Also, you need to specify the unstructured covariance matrix type.

Table 3: Null model

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	2.2173	0.006955	3	318.82	<.0001

Based on the results obtained using SAS, we notice that the estimated grand mean, γ_{00} is 2.2173. This is overall mean fasting glucose level across all patients. The population mean is different from significantly (t-value of 318.82) which indicates that it should be included in the model. Similarly, the "within-person" variance ($\hat{\delta}^2$) 0.04619 and the "between -person" variance ($\delta^2{}_0$) is 0.3505.

Table 4: Covariance parameter Estimates

Covariance Parameter Estimates					
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**
Intercept	**Time**	0.3505	0.001525	3.84	<.0001
Residual		0.04619	0.002115	21.84	<.0001

This PROC MIXED syntax generated the Covariance Parameter Estimates table shown above. Using the estimates presented in the table, we can compute the Intraclass Correlation Coefficient (ICC) that indicates how much of the total variation of patient is accounted for by time.

The general equation for calculating the ICC is provided below. In this equation,$\boldsymbol{\sigma^2}$**time** refers to the covariance estimates for the intercept and $\boldsymbol{\sigma^2}$**error** refers to the covariance estimate for the residual.

$$\text{ICC}= \frac{\sigma^2\, Time}{\sigma^2 time + \sigma^2}$$

Thus, based on the output above we calculated the ICC as

$$\text{ICC}= \frac{0.3505}{0.3505+0.04619} =0.8835$$

This indicates that 88.35% of the variability in diabetes mellitus patients accounted by for the follows up time in our study, leaving 11.65% of the variability in diabetes mellitus to be accounted by follow up time .These results answer our first research question and provide support for using a two-level model. The large proportion of the variability in the by follow up time explained by diabetes mellitus emphasizes the importance of accounting for the hierarchical structure of the data.

Generally, this suggests that about 88.35 percent of the variation in fasting glucose of diabetes mellitus patients can be attributed to difference among individuals.

Table 5: Unconditional Growth Model

Covariance Parameter Estimates					
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**
Time	**ID**	0.001856	0.000464	4.00	<.0001
Residual		0.03980	0.002113	18.83	<.0001

Fit Statistics

-2 Res Log Likelihood	235.0
AIC (Smaller is Better)	231.0
AICC (Smaller is Better)	231.0
BIC (Smaller is Better)	223.9

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	2.2106	0.01070	707	206.52	<.0001
Time	0.004820	0.006387	246	0.75	0.4512

Based on the results obtained, we observing the outputs results, the fixed effect $\hat{\gamma}_{00}$,is 2.2159.This is the overall mean log fasting sugar level across all individual with slope for time zero or baseline fasting . Similarly, $\hat{\gamma}_{10}$ is 0.00482, this is rate at which Y_{ij} changes for individual i when the predictor variable time is included as a level-l predictor within group. The fixed effects $\hat{\gamma}_{00}$and $\hat{\gamma}_{10}$ estimates the starting point(y-intercept) and slope of the population average change trajectory for time points. The parameters are significant (t-value of 206.52 and 0.75 respectively) which indicates that they should both be included in the model.

Similarly, the "within –person" variance ($\hat{\delta}^2{}_{\varepsilon}$) is 0.03980 this number summarizes the average scatter of an individual's observed outcome values around his or her own true change trajectory of Diabetes Mellitus patients[10]. $\hat{\delta}^2{}_{\varepsilon}$ is still significantly smaller for al Diabetes Mellitus in unconditional growth model than what we obtained for Null model or Unconditional Means Model . Therefore, when analysis is performed on all Diabetes Mellitus, we can say that a lot of the within-person variance in Y_{ij} is explained by the duration of times.

Likewise, the level -2 variance components $\hat{\delta}^2{}_0$ *and* $\hat{\delta}^2{}_1$ give the amount of unpredicted variation in the individual growth parameters. $\hat{\delta}^2{}_0$ is the unpredicted variability in the true initial status i.e when time equals zero or baseline. It represents the scatter of the π_{0i} around $\hat{\gamma}_{00}[10]$. the value of $\hat{\delta}^2{}_0$ 4.8868(2.2106^2) .This value is much smaller for just within a group of patients. Similarly, $\hat{\delta}^2{}_1$ gives the unpredictable variability in true rates of change.

Table 6: Conditional Growth Model (Full model)

Solution for Fixed Effects						
Effect		**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept		1.7677	0.03094	0	57.13	.
Age		-0.00003	0.000205	939	-0.16	0.8731
Sex	**Male**	-0.00816	0.006012	939	-1.36	0.01749
	Female	0	.	.	.	.
Weight		0.000326	0.000418	939	0.78	0.4353
Baseline_F		-9.78E-6	0.000026	939	-0.38	0.7016
Fasting_B		0.002156	0.000025	939	84.88	<.0001
F_status	**Working**	0.005912	0.01131	939	0.52	0.6014
	Ambulatory	0	.	.	.	.
M_status	**Married**	-0.00522	0.007577	939	-0.69	0.4913
	Single	0	.	.	.	.
Address	**Urban**	-0.00475	0.006356	939	-0.75	0.4547
	Rural	0	.	.	.	.
Edu_L	**Illiterate**	0.01651	0.008370	939	1.97	0.0489
	Primary	0.02068	0.007530	939	2.75	0.0061
	Secondary	0.01315	0.007838	939	1.68	0.0938
	Tertiary	0	.	.	.	.
Occupation	**Full time**	0.002835	0.005718	939	0.50	0.6201
	Part time	0	.	.	.	.
BMI		0.000762	0.001321	939	0.58	0.5640

Solution for Fixed Effects						
Effect		**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Clinical_D	**Type I**	0.009031	0.008032	939	1.12	0.2611
	Type II	0	.	.	.	.

From the above Table, we emphasized that the covariates for each individual values of the log fasting blood sugar level count at baseline rather than the other covariates of individual measurements. The intercept represents the estimated mean for the last level of log fasting blood sugar level counts. The variables sex of male is about *exp*(-0.00816) = 0.99 was higher than that of female(the reference group) p = 0.01749 , this indicates that sex is a significant effect for diabetes mellitus patient with the progression of time .

The Educational level of Diabetes mellitus patients was a significant effect through the follow up time, so this showed that when an individual have had a diabetes patients a physician should aware the nature of the disease .

For a two-level longitudinal MLM, the software requires input of values regarding the duration of the study, frequency of observations, the level 1 variance , the between –person variability in the parameter of interest and an estimates of the effect size . It is also important that researchers build in estimates of expected attrition rates in their calculations. In sum, MLM can address all the research questions that repeated measures ANOVA/MANOVA tests address without being constrained by the rigid assumptions of the latter [12]. Further, MLM can be used to pursue research questions that cannot be answered with repeated measures ANOVA/MANOVA.

Conclusion

This paper briefly describes the characteristics and the application areas of multilevel modeling with the use of longitudinal data.

Additionally, we investigated the value of ρ , the intra-class correlation coefficient, in each model. By adding level-1 predictors the ICC increased .However, when we added level-2 predictors the ICC dramatically decreased to an even lower value than the unconditional model. This is due to a decrease in the unexplained Level -2 variation, the random intercept term u_{0j} when a predictor was added at the class level. Multilevel model can be used when all individuals are assessed on the same number of occasions which are equally spaced over time. However, MLM can also be used when the spacing of measurement points is not identical across individuals (e.g. , the time interval between diabetes mellitus patient screenings might vary across participants), and also when the number of measurement waves is not the same across individuals. The latter is a particularly important feature, given the attrition of participants recorded in longitudinal studies.

Reference

1. International Diabetes Federation IDF Diabetes Atlas, 6th edition, 2013.[(accessed on 4 March 2014)].
2. Zimmet P.Z., Magliano D.J., Herman W.H., Shaw J.E. Diabetes: A 21st century challenge .Lancet Diabetes Endocrinol.2014; 2:56-64.
3. American Diabetes Association. Classification and diagnosis of diabetes :Standards of Medical Care in Diabetes-2019.Diabetes Care;42(Suppl.1): S13-S28
4. Siu AL; U S Preventive Service Task Force. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015: 163:861-868.
5. National Center for Chronic Disease Preventive and Health Promotion, Division of Diabetes Translation. National Diabetes Statistics Report, 2017. Estimates of Diabetes and its Burden in the United States.
6. Bullard KM, Ali MK, Imperatore G,et al. Receipt of glucose testing and performance of two US diabetes screening guidelines,2007-2012.
7. Kiefer MM,Silverman JB,Young BA, Nelson KM. National patterns in diabetes screening data from the National Health and Nutrition Examination Survey(NHANES) 2005-2012.J Gen Inern Med 2015; 30:612-618 .
8. Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13-22.
9. Introducing Multilevel Modeling, 1st edition, by I. Kreft, J. D. Leeuw, Published By Sage Publications 1999.
10. Applied Longitudinal Data Analysis, 1st edition, by J. D. Singer, J. B. Willet, and Published by Oxford University Press 2003.
11. Introducing Multilevel Modeling, 1st edition, I Kreft,J.D.Leeuw, Published by Sage Publications 1999.
12. Kwok ,O.M., Underhill,A.T.,Berry,J.W.,Luo, W.,Elliott,T.R.,and Yoon,M.(2008) Analyzing longitudinal data with individuals living with multilevel models, an example with individuals living with lower extremity intra-articular fractures.
13. Rong Su, Le Cai, Wenlong Cui, Jianhui He, Dingyun You and Allison Golden (2016). Multilevel Analysis of Socioeconomic Determinants on Diabetes Prevalence, Awareness, Treatment and Self-Management in Ethnic Minorities of Yunnan Province, China. Int. J. Environ. Res. Public Health, 13(8), 751.
14. Bertho Tantular, I Gede Nyoman Mindra Jaya and Zulhanif (2018). Multilevel models approach on longitudinal studies for disease data. Journal of Physics: Conference Series, 1025.
15. Wudneh Ketema Moges, A. R. Muralidharan, and Haymanot Zeleke Tadesse (2017). Gaussian Longitudinal Analysis of Progression of Diabetes Mellitus Patients Using Fasting Blood Sugar Level: A Case of Debre Berhan Referral Hospital, Ethiopia. American Journal of Theoretical and Applied Statistics, 7(1): 21-28.
16. Bethany A. Bell, Mihaela Ene, Whitney Smiley, and Jason A. Schoeneberger (2013). A Multilevel Model Primer Using SAS® PROC MIXED. Statistics and Data Analysis, 433, 1-19.?
17. The Department of Statistics and Data Sciences (2015). Multilevel Modeling Tutorial: Using SAS, Stata, HLM, R, SPSS, and Mplus. Statistics + Data Sciences.