**HW2 Due 2/25 Monday**

The ECLS-K is a national survey of children enrolled in kindergarten in 1998-1999, representative of all kindergarteners attending school that year. The survey is administered by the National Center for Education Statistics, U.S. Department of Education (NCES), and uses a multistage probability sample design (counties, schools, then randomly selected children). The baseline (fall and spring of kindergarten) waves of the ECLS-K, conducted in 1998-1999, included about 20,000 children, with sizable over-samples of at-risk children, including poor and minority children. Children and families were followed in the fall and spring of first grade (1999-2000), spring of third grade (2002), and spring of fifth grade (2004). The entire longitudinal data can be found from this link: <http://nces.ed.gov/ecls/dataproducts.asp>

**The data for this homework is a subset of the above data, which is the baseline data. The name of the data set is “DataForHW2.sas7bdat”. Using Proc freq to find the code for the categorical variables in the data set.**

In this homework, we will evaluate the relationship between child’s **age- and sex-specific BMI percentile** (bmipct) and variables at both school and child levels at the baseline (spring of 1999), by fitting a multilevel model to nest children (level 1) within schools (level 2). The multilevel approach allows us to model variations in BMIPCT within and between schools. The analysis is based on part of the study in the reference paper, “The role of local food availability in explaining obesity risk among young”, which can be found in the Blackboard.

Child, school, and neighborhood measures are described as follows.

**Child-level outcome variable**:

* **Outcome**: **the age- and sex-specific BMI percentile** (bmipct) for each child was calculated using the SAS program (SAS Institute, Inc., Cary, North Carolina) developed by the Centers for Disease Control and Prevention (CDC) based on the updated 2000 Growth Charts.Height and weight were measured by ECLS-K trained assessors using a Shorr Board (Shorr Productions, Olney, MD) and a digital bathroom scale (Seca Model 840, Hanover, MD), which reduced concerns regarding the reliability of parental reports and BMI was calculated from measured height and weight. Children with unreasonable or implausible values were also dropped. These included cases where: BMI values were less than 10 or greater than 50 in the kindergarten. These were likely data entry, coding, or reporting errors. Those with a BMI percentile of 95 or higher were classified as obese.

**Child-level outcome independent variables**:

The ECLS-K attempted to collect parental information at each wave (generally from the child’s mother). Using data from the parent interviews, we consider measures of the child’s racial/ ethnic background, gender, age, household income, educational attainment of parents, parental health (self-reported), and family structure (e.g., single parent). These socio-demographic factors are important correlates of childhood obesity and may also be correlated with residential location and other neighborhood characteristics of interest. More proximate behavioral factors potentially related to BMI include parental reports of how many days per week the child engaged in 20 min or more of vigorous activity or exercise (where the child’s heart rate is consistently elevated) outside the school context, and parental reports of hours the child spends watching television and videos. These two measures serve as indicators of how sedentary children’s lives are outside the school environment.

* Gender: gender of the focal child: 1=male, 2=female
* Childgender: gender of the focal child: “1=female”, “2=male”
* ChildAge: Age of child in months (Continuous)
* Childbmi: child composite bmi
* Meducation: A composite variable of the mother’s highest attained education levels.
* ChildRace: Race and ethnicity of the focal child
* FamilyStructure: Classification of the focal child’s parents who reside in the household
* HouseIncome: Household income (Continuous)
* Phealth: Child’s parent health status
* ExerciseFreeTime : Child’s physically active free time(Continuous)
* TV: Number of hours child spent on watching TV (Continuous).

**School-level variables**

Several school-based factors (public or private, urbanicity, and percent of students who are eligible for school meal programs e a proxy for school-level disadvantage) may be indirectly related to obesity risk and tap into school-level dynamics.

* Region (Census region): Indicates the geographic region in which the child lives
* Urban: School urbanicity designation by the Census Bureau
* Schooltype: School type from the school administrator questionnaire
* PctMinority: School Percentage of minority students in school(Continuous)
* PCTFreeLunch: School Percentage of students eligible for free lunch in school(Continuous)

**Questions:**

Run proc mixed using data set DataForHW2 on the outcome variable BMIPCT along with the above independent variables. Use model building strategies to build the final model. Interpret significant effects. Find the variances at different levels.