Homework 3
Due Tues, June 10, 10 AM (pdf file please)

1. This problem uses the aging and control data set from HW 2, agingcontrol2.dat or agingcontrol2.sav, http://web.pdx.edu/~newsomj/data.htm. For any of the SEM models below, **use Mplus or lavaan**.

a. Obtain descriptive information from SPSS or R about the skewness and kurtosis of all six variables (`sleep, effort, getgo, enjoy, hopeful, energy`). Report your findings and evaluate the degree to which the data are likely to violate distributional assumptions. You can reference the criteria provided by West, Finch, and Curran (1995) for these univariate statistics, or, optionally, you can compute the normalized Mardia's coefficient using DeCarlo's macro with SPSS or the MVN package in R with `mvn(data = dataframename, mvnTest = "mardia")` and interpret in terms of the Bentler & Wu (2002) standard.

b. Using maximum likelihood with scaled chi-square and robust standard errors (`estimator=MLM` in Mplus or `estimator="mlm"` in `lavaan`), test the same two-factor model (`sleep`, `effort`, `getgo` on one factor and `enjoy`, `hopeful`, and `energy` on a second factor) as in HW 2 Problem 2. Report your findings, including information about model fit and standardized loadings. Compare your output to that obtained with the standard ML solution, and describe any differences in the fit statistics, the loadings, and the standard errors.

c. Show how the scaling correction factor was obtained, and explain what it tells you about nonnormality and the chi-square test.

2. Use the same data set that you used for Problem 7 in HW 2 (your own data set). Retest whichever model fit better from 7a and 7b using maximum likelihood with scaled chi-square and robust standard errors (`estimator=MLM` in Mplus or `estimator="mlm"` in lavaan). Report your findings, including information about model fit and standardized loadings. Then, compare your output to that obtained with the standard ML solution (Your comparison should use the same number of cases, so if you used ML with missing data last time, use MLR with missing cases this time). Be sure to describe any differences in the fit statistics, the loadings, and the standard errors.

3. For the following question, I have created another version of the aging and control data set (`agingcontrol3`) that includes a grouping variable (`vol`) to indicate whether the participant did any volunteer work or not (0 = no, 1 = yes). The data page has a new raw data file, new Mplus and `lavaan` input files, and SPSS version of the file (http://web.pdx.edu/~newsomj/data.htm). This version of the data set has missing data (all missing values are coded -99). Include your output and show your work for any hand computations.

The new version of the data set contains the following variables:

`vol` whether the participant did any volunteer work
Four negative affect items from the CES-D (8-point frequency scale; higher values indicate more negative affect):
- `mind` "I could not keep my mind on things"
- `sad` "I felt sad"
- `lonely` "I felt lonely"
- `blues` "I could not shake the blues"

Four social support items (4-point rating scale of agreement, for "strongly disagree" to "strongly agree"):
- `turn` "Have someone can turn to for support"
- `talk` "Have someone can really talk to"
- `help` "Have someone who would help me out with things"
- `sick` "Have someone who would help if I were sick"

`negaff` average of negative affect items
`support` average of support items
`strenex` frequency of strenuous exercise (same as HW 2)

For Problems 3a and 3b, **allow missing data** estimation and **assume the variables are not multivariate normal**.

a. Test a one-factor confirmatory factor model of the four negative affect variables (`mind`, `sad`, `lonely`, `blues`). Report and interpret your results, including fit indices (chi-square, CFI, SRMR), standardized loadings, and loading significance.

b. Test a one-factor confirmatory factor model of the four support items (`turn`, `talk`, `help`, `sick`), using the WLSMV estimator. Report and interpret your results, including fit indices (chi-square, CFI, SRMR), standardized loadings, and loading significance.

For the multigroup models below (Problems 3c and 3d), use the effects coding identification rather than referent item identification, no means structure, and use standard FIML estimation (`estimator=ML`) with no missing data (using `listwise=on` in Mplus or `missing = 'listwise'` in `lavaan`) to simplify chi-square difference testing for now)[1].

c. Run two multigroup structural equation models in Mplus or `lavaan` to compare the one-factor negative affect model (`mind`, `sad`, `lonely`, `blues`) across non-volunteer and volunteer groups to see if there is at least weak measurement invariance (to simplify things, we are not comparing means or intercepts across groups in this exercise). In the first model, allow all parameters to differ across groups (same form model). In a second model, constrain only the loadings to be equal across volunteer groups, leaving factor variances and measurement variances to be freely estimated in each group (partial or weak invariance). Report and interpret your findings, including a description of the standardized loadings, the chi-square difference test, and Cohen's w and ΔNCI magnitude of effect measures for the chi-square difference.

d. Run two multigroup structural equation models in Mplus or `lavaan` to examine moderated mediation. Use the composite measures for support, strenuous exercise, and negative affect to compare a mediational model (`support` → `strenex` → `negaff`) across the two volunteer groups by constraining the path from support to strenuous exercise to be equal across the groups in one model and allowing all paths to be freely estimated in the other model. Report and interpret your results, including the chi-square difference in fit of the two models, the within-group fit indices and direct effects (unstandardized and standardized paths) for the free model. To reduce the amount of work, I am not requiring indirect effects tests or simple slopes tests for this problem.

4. The data for the following problems come from the Early Childhood Longitudinal Study (ECLS) following kids from kindergarten to fifth grade.[2] The data sets (.sav and .dat) as well as Mplus and `lavaan` input files are available at http://web.pdx.edu/~newsomj/data.htm. Variables include adult's (i.e., parent's) food insecurity score at Time 2 and Time 4 (`afs2`, `afs4`), child's food insecurity score at Time 2 and Time 4 (`cfs2`, `cfs4`), math ability scores for grades K through 3 (`math1`, `math2`, `math3`, `math4`), whether or not the family received food stamps in one of the years (`foodstmp`). There are missing data present (all coded -99) and you should assume multivariate nonnormality for these variables, so use the appropriate estimation approach for nonnormal missing data.

a. Use a cross-lagged panel model in Mplus or `lavaan` to investigate the causal directionality of the relationship between adult's and child's food insecurity (whether the parent's food insecurity leads to higher child's food insecurity or child's food insecurity leads to higher parent's food insecurity), using `afs2`, `afs4`, `cfs2`, and `cfs4`. Report and interpret your results.

b. Using the same data set and Mplus or `lavaan`, test a latent growth curve model of math scores over the four grades, `math1`, `math2`, `math3`, and `math4`. Use the 0, 1, 2 coding for the growth factor. Show how the degrees of freedom are derived for this model. Describe and interpret your results from the model. Be sure your write-up contains information about all of the following questions: Is there an overall increase or decrease in math scores

[1] Analyses using the Satorra-Bentler correction for nonnormality did not lead to any differences in the substantive conclusions, although in actual practice it would likely be preferable to use this method. We will use regular ML here to make life easier for you.
[2] This data set is a random subset of cases from the study. For more details on the study, see https://nces.ed.gov/ecls/

over time? Do the intercepts or slopes vary significantly across respondents? Do the intercepts and slopes *covary* across respondents? Draw a quick sketch by hand that depicts the pattern of the slopes in this case.

c. Use the variable for whether or not the family has received food stamps (`foodstmp`) to predict the intercepts and slopes of math scores. Describe and interpret your results. No simple slopes test is required for this problem.

5. Draw the following hypothetical model, which is described in "all-y" format. (* indicates parameter is free to be estimated, 1 or 0 indicates a fixed value, and all others are assumed to be zero). Label the parameters with the correct Greek notation. Hand drawn is acceptable, but this is a good opportunity to try drawing the model with PowerPoint or another graphic design software program.

Λ

	ETA 1	ETA 2	ETA 3	ETA 4
y1	*			
y2	*			
y3	*			
y4		*		
y5		*		
y6		*		
y7		*		
y8		*		
y9			1	
y10				*
y11				*
y12				*

Ψ

	ETA 1	ETA 2	ETA 3	ETA 4
ETA 1	1			
ETA 2	*	1		
ETA 3			*	
ETA 4				*

B

	ETA 1	ETA 2	ETA 3	ETA 4
ETA 1				
ETA 2				
ETA 3	*	*		
ETA 4			*	

Θε

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12
y1	*											
y2		*										
y3			*									
y4				*								
y5					*							
y6					*	*						
y7					*		*					
y8								*				
y9									0			
y10										*		
y11											*	
y12												*