Name ________________________

Psychology 5068
Hierarchical Linear Models
Homework 7
Due March 26, 2018

For this assignment, you will extend our analyses of the daily diary data (daily_diary.csv). Time series analyses often use lagged variables. A lagged variable is collected at a different point in time compared to other measures with which it is paired. For example, in the daily diary data, self-esteem was collected each day and one question we might have is whether self-esteem on a particular day is related to self-esteem on the previous day. This would address a question about the persistence of self-esteem, at least in the short run.

To examine that question, we need a lagged self-esteem variable in the data file. The following is a small piece of the daily diary data file, for Participant 5, showing the self-esteem measures that this participant provided for 10 consecutive days. In the last column is the lagged self-esteem variable. It simply carries to a particular day the self-esteem measure from the previous day (there is no "previous day" for the first row, so data are missing [NA] for the lagged variable). For this particular participant, we could calculate the correlation between the last two columns and would find a value of .67. In other words, for this participant, there is some persistence in self-esteem. If self-esteem is high on one day, it tends to be high on the following day too.

	id	day	rse	rse_lag_1
49	5	6	5.25	NA
50	5	7	5.50	5.25
51	5	8	6.75	5.50
52	5	9	7.00	6.75
53	5	10	7.00	7.00
54	5	11	6.50	7.00
55	5	12	6.50	6.50
56	5	13	6.50	6.50
57	5	14	6.75	6.50
58	5	15	7.00	6.75

1. Begin by creating lagged variables for self-esteem (rse_lag_1), the reported average positive events (posmn_lag_1), and the reported average negative events (negmn_lag_1). You may find functions that can help you do this or you may decide to do this with your own code. Either way, you will need to do this carefully because the reports by participants do not always reflect consecutive days. That is, a person might provide data on the following days: 1, 2, 3, 6, 8, 9, 10, 12, 13, and 14. Lagged variables would only be available for days 2, 3, 9, 10, 13, and 14. Those are the only days when there was also a measure available on the immediately preceding day. So, we will lose some data in the process of creating lagged variables but that will insure that the meaning of the lagged variable is consistent.

2. First, let's see how persistent self-esteem is for all of the participants. Fit the following model:

Level 1

$$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$

   (a) What is the value of the fixed effect for rse_lag_1? Is it significant?
   (b) Provide a caterpillar plot showing the individual estimated coefficients for rse_lag_1. The individual coefficients can be obtained using the coef( ) function for the fit object. Each individual coefficient has a standard error that can be obtained with the se.coef( ) function, a helper function from the *arm* package. Is persistence in self-esteem a common occurrence? What makes you say so?

3. Now let's see how persistent positive events are for all of the participants. Fit the following model:

Level 1

$$posmn_{ti} = \pi_{0i} + \pi_{1i} posmn\_lag\_1_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$

   (a) Is there significant persistence for positive events? How would you interpret this effect? Speculate about why it might occur.
   (b) Provide a caterpillar plot of the individual estimated coefficients for posmn_lag_1. Is this form of persistence common?

4. Finally, let's see how persistent negative events are for all of the participants. Fit the following model:

Level 1

$$negmn_{ti} = \pi_{0i} + \pi_{1i} negmn\_lag\_1_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$

   (a) Is there significant persistence for negative events?
   (b) Provide a caterpillar plot of the individual estimated coefficients for negmn_lag_1 and comment on how common this form of persistence is among participants.

5. Now fit a model with self-esteem as the outcome and with lagged self-esteem, current positive events, and current negative events as predictors:

Level 1

$$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn_{ti} + \pi_{3i} negmn_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$
$$\pi_{2i} = \beta_{20} + r_{2i}$$
$$\pi_{3i} = \beta_{30} + r_{3i}$$

(a) This model provides a different way of examining the impact of current events on self-esteem. By partialling the previous day's self-esteem, how would the influence of current events be interpreted in this model?
(b) Fit the model and comment on the magnitude and interpretation of the coefficients for positive and negative events.

6. Fit a full lagged model:

Level 1

$$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn_{ti} + \pi_{3i} negmn_{ti} + \pi_{4i} posmn\_lag\_1_{ti} + \pi_{5i} negmn\_lag\_1_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$
$$\pi_{2i} = \beta_{20} + r_{2i}$$
$$\pi_{3i} = \beta_{30} + r_{3i}$$
$$\pi_{4i} = \beta_{40} + r_{4i}$$
$$\pi_{5i} = \beta_{50} + r_{5i}$$

(a) With current positive events controlled, is there any residual effect of the previous day's positive events on self-esteem?
(b) With current negative events controlled, is there any residual effect of the previous day's negative events on self-esteem?

7. Lastly, take the model from the previous problem and add the standardized CES-D depression measure (you will need to create this) as a predictor in each Level 2 equation:

Level 1

$$rse_{ti} = \pi_{0i} + \pi_{1i} rse\_lag\_1_{ti} + \pi_{2i} posmn_{ti} + \pi_{3i} negmn_{ti} + \pi_{4i} posmn\_lag\_1_{ti} + \pi_{5i} negmn\_lag\_1_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + \beta_{01} cesd\_z_i + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11} cesd\_z_i + r_{1i}$$
$$\pi_{2i} = \beta_{20} + \beta_{21} cesd\_z_i + r_{2i}$$
$$\pi_{3i} = \beta_{30} + \beta_{31} cesd\_z_i + r_{3i}$$
$$\pi_{4i} = \beta_{40} + \beta_{41} cesd\_z_i + r_{4i}$$
$$\pi_{5i} = \beta_{50} + \beta_{51} cesd\_z_i + r_{5i}$$

Comment on any moderating effects of depression score.

Some of the models you will fit are pretty complex. If you run into problems getting them to converge, try the following:

(a) Increase the number of iterations used by the estimation algorithm
(b) Try a different algorithm
(c) Simply the model by removing random effects (i.e., setting Level 2 variances to 0).