Cluster Analysis & Factorial ANOVA

The purpose of the analysis was to examine how social support, loneliness, stress, gender and their combinations (interactions) were related to depression. One challenge when working with multivariate models is that patterns of predictor collinearity can lead to under-representation of some combinations of predictor values. Said differently, if the predictors are inter-correlated, there may be certain combinations of predictor values that are infrequent.

As shown below, there is a substantial collinearity between social support, loneliness, and stress. One way to look for patterns of underrepresentation is to perform a mean-split each of the three quantitative predictors and obtain the sample sizes for the resulting groupings, including gender. Please note, we would likely perform this analysis including social support, loneliness and stress in their original quantitative form along with their interactions. This mean-split was done only to examine whether portions of the resulting multivariate distribution would be "too sparse" to comfortably conform to distributional assumptions of the regression model.

Correlations

		total social support	loneliness	stress	gender
total social support	Pearson Correlation	1	-.628**	-.585**	.109*
	Sig. (2-tailed)		.000	.000	.028
	N		405	405	405
loneliness	Pearson Correlation	-.628**	1	.485**	-.021
	Sig. (2-tailed)	.000		.000	.678
	N	405	405	405	405
stress	Pearson Correlation	-.485**	.285**	1	.018
	Sig. (2-tailed)	.000	.000		.720
	N	405	405	405	405
gender	Pearson Correlation	.109*	-.021	.018	1
	Sig. (2-tailed)	.028	.678	.720	
	N	405	405	405	405

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

There are some substantial collinearities among these predictors, suggesting the some combinations of predictor values will be more poorly represented within the sample.

stress_2 * tss_2 * ruls_2 * gender Crosstabulation

Count

gender	ruls_2			tss_2 Low	tss_2 High	Total
male	Low	stress_2	Low	18	51	69
			High	4	28	32
	High	stress_2	Low	23	14	37
			High	36	6	42
female	Low	stress_2	Low	20	69	89
			High	3	48	51
	High	stress_2	Low	29	14	43
			High	35	7	42

As suggested by the predictor collinearities, notice there are some very small cell sizes, likely leading to sampling instability and "slop" in the position of the model regression lines!

The results of these analyses suggest that not all combinations of predictor values are abundant. Put differently, these results suggest that there may be commonly occurring patterns of predictor values. Recalling an earlier cluster analysis that revealed four distinguishable multivariate profiles(clusters) with respect to social support, loneliness and mental health, we decided to explore whether these clusters might be a useful representation of the patterns of collinearity?

Cluster analyses are usually completed with more than three variables, and it is a good idea whenever possible to include multiple versions of the different kinds of variables represented (i.e., friend, family & significant other social support). However, differentiable profile patterns based on a small set of clustering variables can be obtained and make a point about which of the possible profiles organize the represented population.

Working with the Z-score versions of total social support, loneliness and stress, we obtained the 4-cluster solution for the current sample (including both males and females). The results were very similar those of the earlier cluster analysis conducted with these and other variables, using a female sample!

Report

Mean

Ward Method	Zscore: total social support	Zscore: loneliness	Zscore (stress)
1	1.5171020	-.6149681	-.7154989
2	-.1296110	.8326870	1.3640410
3	-.3514112	.9723187	-.5197010
4	-2.0045627	1.5832024	1.6467230

Report
Mean
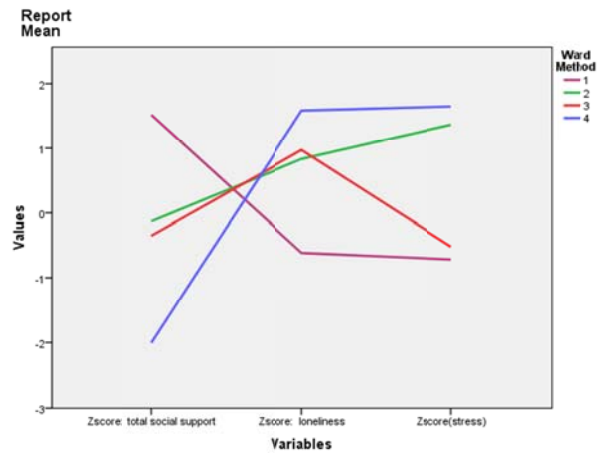
Group 1 was "high support, low lonely & low stress"

Group 2 was "moderate support, lonely & stressed"

Group 3 was "moderate support, lonely & not stressed"

Group 4 was "low support, very lonely & very stressed"

Crosstabs revealed that we had substantial sample sizes for each profile for both males and females.

The intent of this cluster analysis is to represent the predictor collinearity patterns, and the resulting predictor profiles, so that we can capture the "kinds of folks" there are, to see how these "kinds of folks" differ with respect to depression.

Sup_Ruls_Stress * gender Crosstabulation

Count

		gender		
		male	female	Total
Sup_Ruls_Stress	Hi/Lo/Lo	72	93	165
	Med/Hi/Lo	47	66	113
	Med/Hi/Hi	41	42	83
	Lo/Hi/Hi	20	24	44
Total		180	225	405

The results of the factorial ANOVA using cluster membership as one factor and gender as the other is show below.

Descriptive Statistics

Dependent Variable: depression (BDI)

gender	Sup_Ruls_Stress	Mean	Std. Deviation	N
male	Hi/Lo/Lo	3.81	2.958	72
	Med/Hi/Lo	4.25	5.726	47
	Med/Hi/Hi	11.10	6.308	41
	Lo/Hi/Hi	12.30	7.888	20
female	Hi/Lo/Lo	4.33	3.048	93
	Med/Hi/Lo	8.52	5.370	66
	Med/Hi/Hi	9.57	6.801	42
	Lo/Hi/Hi	15.96	12.005	24

Tests of Between-Subjects Effects

Dependent Variable: depression (BDI)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4456.922[a]	7	636.703	19.681	.000
Intercept	25500.102	1	25500.102	788.224	.000
gender	133.286	1	133.286	4.120	.043
CLU4_1	4167.729	3	1389.243	42.942	.000
gender * CLU4_1	396.261	3	132.087	4.080	.007
Error	12543.483	397	32.351		
Total	39805.000	405			
Corrected Total	17300.405	404			

a. R Squared = .258 (Adjusted R Squared = .245)

There are both main and interaction effects involving cluster membership. Notice the differential simple effect pattern of cluster member ship for males and females – suggesting gender differences in what combinations of support, loneliness and stress are related to higher levels of depression!

Males   Hi/Lo/Lo < Med/Hi/Lo < Med/Hi/Hi = Lo/Hi/Hi

Females   Hi/Lo/Lo < Med/Hi/Lo = Med/Hi/Hi < Lo/Hi/Hi