Measuring Subordination of Women in the Home
A Confirmatory Factor Analysis for Ordinal Variables

Kaylee B. Hodgson

December 13, 2018

Abstract

In this paper, I evaluate whether one factor to measure the subordination of women in the household is sufficient in representing its 11 sub-indicators. I use a path analysis to evaluate the causal theory, and implement a confirmatory factor analysis to examine both a one-factor and two-factor approach. I find that the path analysis does not support the causal theory, and discuss possible reasons for that. I also find that the one-factor model is a good fit for the 11 indicators, and significantly explains the variation in each of the individual variables.

1   Introduction and Background

In 2017, The WomanStats Project released a multivariate country-level scale, named the Patrilineality/Fraternity Syndrome Scale, that indicates overall treatment of women in the household [?]. This scale combines 11 indicators that measure aspects of the subordination of women in the home. These 11 variables measure the laws and customs surrounding certain marital and familial practices that impact women in a society. This scale was built to empirically test the theory that societies that systematically impose gender suppression and inequality in the home will, as a result, perform worse on macro-level indicators of success, including economic performance, political stability, and security.

In this paper, I preliminarily implement a path analysis to evaluate the theory of the Syndrome relationship. I then use confirmatory factor analysis for ordinal data to assess the accuracy of this Patrilineality/Fraternity Syndrome Scale in measuring the proposed 11 indicators of the subordination of women at the household level. Additionally, I simulate data to evaluate how well the model fits in situations with poor data quality or missing data, a common issue especially for data from countries that are less developed and/or have higher levels of corruption.

1.1   The Syndrome Scale

The 11 indicators that make up the Patrilineality/Fraternity Syndrome Scale are all scales developed and coded by The WomanStats Project [?], using both qualitative and quantitative data from a variety of sources, including government documents, journalistic accounts, international government organization (IGO) and non-government organization (NGO) reports, interviews, and scholarly articles. Each of these scales measure data between 2010 and 2015. The variables (including the name of the variable in the database) are: Prevalence and Legality of Polygyny (PW-SCALE-1), Inequitable Family Law and Practice Favoring Males (MUTIVAR-SCALE-3), Bride Price/Dowry/Wedding Costs (MARR-SCALE-3), Women's Property Rights in Law and Practice (LO-SCALE-3), Prevalence and Legality of Cousin Marriage (MARR-SCALE-2), Age of Marriage for Girls in Law and Practice (AOM-SCALE-3), Legal Exoneration for Rapists Offering to Marry Victims (LRW-SCALE-9), Son Preference and Sex Ratios (ISSA-SCALE-1), Prevalence of Patrilocal Marriage (MARR-SCALE-1), Overall Level of Violence against Women (MULTIVAR-SCALE-1), and Societal Sanction for Femicide (MURDER-SCALE-1).
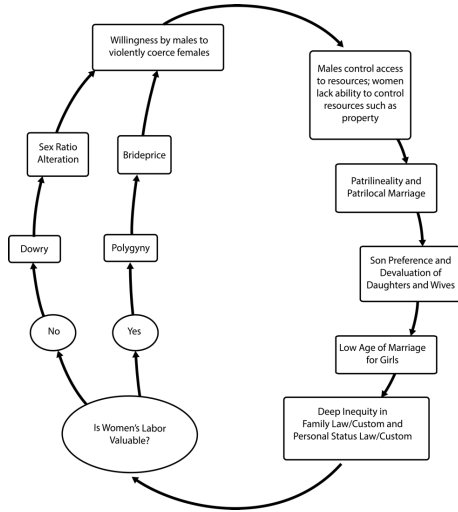
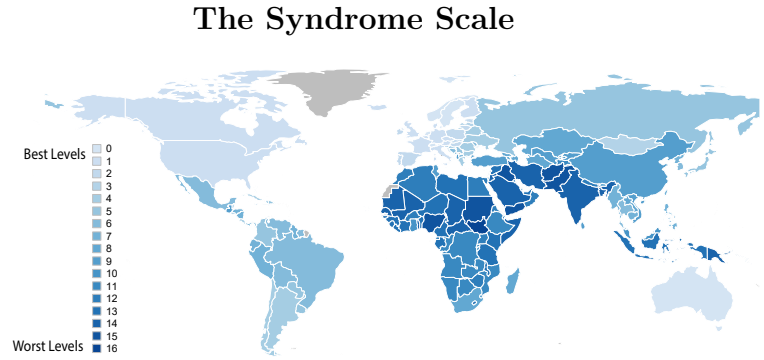Figure 1: Subcomponents and Theory of the Syndrome

Figure 2: Map of countries' levels of systematic subordination of women at the household level

These 11 indicators were combined into the overall scale based on the theory of how and why this systematic subordination of women is embedded into a society. The summary of the theory behind this scale is represented in Figure **??**. All of the scales are coded so that higher values indicate a worse situation for women, so any value added to a country's score is worse. The distribution of the Syndrom scores across countries is shown in Figure **??**.

The scale ranges from 0 to 16, where 16 indicates a country with the worst level of systematic subordination of women in the household.

2 Path Analysis

I first implement a preliminary path analysis, which is essentially an extension of multiple regression analysis, but tests multiple levels of "causal" relationships. This analysis is also, in some ways, a special case of structural equation modeling (SEM), excluding any latent variables. This is done as an initial evaluation of the theory behind the scale, which again is summarized in Figure **??**. Of course, the results of the path analysis should be interpreted cautiously because, despite the apparent directionality of the analysis, it cannot adequately prove causality.

2.1 Results

The path structure output from the analysis can be found in Figure **??**. While there are some strong relationships along the paths, and most are positive as expected, the model fit is poor. The Root Mean Square Error of Approximation (RMSEA) is 0.205 and the Standardized Root Mean Square Residual (SRMR) is 0.3. Typically, in order to conclude that the model is a good fit, these values should be less than 0.08.

The causal relationships defined in the theory for the Syndrome are not a good fit in the path analysis used here. One reason for this could be that the situations described in the theory are sometimes only partially covered by a single variable, and some variables describe multiple phenomena at once. The path created in the analysis may therefore not be an accurate portrayal of the path in the theory.

The path analysis only looks for dependencies among the variables by the relationships specified in the structural equations. Poor fit here does not preclude the possibility that these variables fit well into a single factor. Therefore, despite these results, I proceed to the confirmatory factor analysis to evaluate whether the clustering of all 11 indicators is a good fit.

Path Analysis Results

Rape Exemption • Patrilocal Marriage

0.52 • 0.96 • 0.72

Son Preference • Laws/Customs Favoring Males

0.75 • 0.52

−0.08 • 0.20 • 0.69

0.53

Cousin Marriage • 0.71 • Age of Marriage • 0.98

−0.14 • 0.20

0.53 • 0.84

Private Property • 0.70 • Bride Price / Dowry • 0.51

0.54 • −0.74 • 0.51 • 0.54

0.29

Femicide • 0.96 • Polygyny

Overall Violence

Correlation Plot

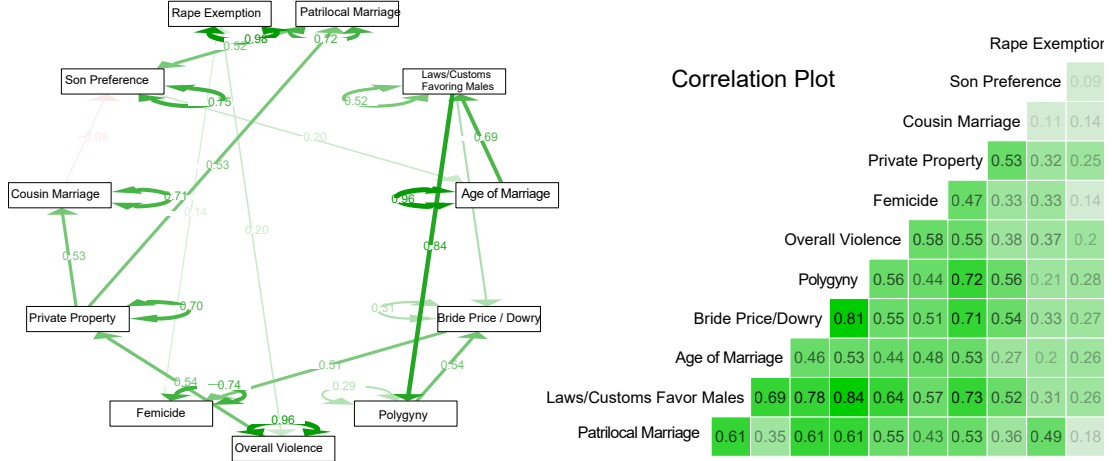|	Rape Exemption	Son Preference	Cousin Marriage	Private Property	Femicide	Overall Violence	Polygyny	Bride Price/Dowry	Age of Marriage	Laws/Customs Favor Males
Son Preference	0.09									
Cousin Marriage	0.11	0.14								
Private Property	0.53	0.32	0.25							
Femicide	0.47	0.33	0.33	0.14						
Overall Violence	0.58	0.55	0.38	0.37	0.2					
Polygyny	0.56	0.44	0.72	0.56	0.21	0.28				
Bride Price/Dowry	0.81	0.55	0.51	0.71	0.54	0.33	0.27			
Age of Marriage	0.46	0.53	0.44	0.48	0.53	0.27	0.2	0.26		
Laws/Customs Favor Males	0.69	0.78	0.84	0.64	0.57	0.73	0.52	0.31	0.26	
Patrilocal Marriage	0.61	0.35	0.61	0.61	0.55	0.43	0.53	0.36	0.49	0.18

Figure 3: Results show that there are some strong relationships in the path analysis. The correlation values are the same as the standardized coefficient estimates if there were only two variables in that portion of the model.

3   Confirmatory Factor Analysis

I implement a confirmatory factor analysis (CFA) to evaluate whether a one-dimensional scale is adequate in representing the 11 components. Each of the 11 variables included in this analysis are ordinal, with anywhere between 2 and 5 levels. While confirmatory factor analysis using maximum likelihood estimators is a useful tool for continuous data that meet multivariate normality assumptions [?], this estimation would likely perform poorly with the ordinal data with so few levels. I therefore explore other options for estimation.

Weighted least square (WLS) has been suggested as a method for estimation when ordinal data is used in CFA, however, because of criticisms of its poor performance when the sample size is small, diagonally weighted least square (DWLS) and unweighted least square (ULS) have become the more popular methods for ordinal data [?]. In 2009, Forero, Maydeu-Olivares, and Gallardo-Pujol used a simulation study to compare DWLS and ULS. They found that ULS outperformed DWLS because ULS gave more accurate estimates, more precise standard errors, and better coverage rates [?]. I therefore use ULS estimation in the CFA model for this analysis.

The general structure for a least squares estimator is:

$$q(\boldsymbol{\theta}; \mathbf{S}) = (\text{vech}\mathbf{S} - \text{vech}\boldsymbol{\Sigma}[\boldsymbol{\theta}])'\hat{\mathbf{V}}^{-1}(\text{vech}\mathbf{S} - \text{vech}\boldsymbol{\Sigma}[\boldsymbol{\theta}]), \qquad (1)$$

where the method finds $\hat{\boldsymbol{\theta}}_{LS}$ that minimizes $q(\boldsymbol{\theta}; \mathbf{S})$. $q(\hat{\boldsymbol{\theta}}_{LS}; \mathbf{S})$ approaches a chi-squared distribution ($\chi^2_{p*-q,\alpha}$, $p* = \frac{1}{2}p(p+1)$). In the case of the unweighted least square model, $\hat{\mathbf{V}}^{-1} = \mathbf{I}$.

In the one-factor model, I fix the Patrilocal Marriage parameter estimate so that the model is estimable. I then evaluate a two-factor model, based on the flow chart in Figure ??. In the chart, there are two separate paths depending on whether women's labor is considered valuable. These paths indicate that if women's labor is not valued, there will be sex ratio alterations in that society, and if it is valuable, polygyny is more likely. I separate the variables measuring these two into separate factors, then allow the other variables to be estimated with both of those factors. Finally, I create a two-factor model that assumes that factor 1 and factor 2 are both affected by a third factor - the second order CFA for the two factors. The models are written out in Table ??.

Table 1: Linear Combination Equations for CFA Models

	One-Factor Model	Two-Factor Models
Polygyny	$y_1 = \lambda_1 f_1 + \epsilon_1$	$y_1 = f_2 + \epsilon_1$
Laws/Customs Favoring Males	$y_2 = \lambda_2 f_1 + \epsilon_2$	$y_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \epsilon_2$
Bride Price/Dowry	$y_3 = \lambda_3 f_1 + \epsilon_3$	$y_3 = \lambda_{31} f_1 + \lambda_{32} f_2 + \epsilon_3$
Property Rights	$y_4 = \lambda_4 f_1 + \epsilon_4$	$y_4 = \lambda_{41} f_1 + \lambda_{42} f_2 + \epsilon_4$
Cousin Marriage	$y_5 = \lambda_5 f_1 + \epsilon_5$	$y_5 = \lambda_{51} f_1 + \lambda_{52} f_2 + \epsilon_5$
Age of Marriage	$y_6 = \lambda_6 f_1 + \epsilon_6$	$y_6 = \lambda_{61} f_1 + \lambda_{62} f_2 + \epsilon_6$
Legal Exoneration for Rapists	$y_7 = \lambda_7 f_1 + \epsilon_7$	$y_7 = \lambda_{71} f_1 + \lambda_{72} f_2 + \epsilon_7$
Son Preference	$y_8 = \lambda_8 f_1 + \epsilon_8$	$y_8 = \lambda_{81} f_1 + \lambda_{82} f_2 + \epsilon_8$
Patrilocal Marriage	$y_9 = f_1 + \epsilon_9$	$y_9 = \lambda_{91} f_1 + \lambda_{92} f_2 + \epsilon_9$
Overall Violence	$y_{10} = \lambda_{10} f_1 + \epsilon_{10}$	$y_{10} = \lambda_{101} f_1 + \lambda_{102} f_2 + \epsilon_{10}$
Societal Santion of Femicide	$y_{11} = \lambda_{11} f_1 + \epsilon_{11}$	$y_{11} = f_1 + \epsilon_{11}$
Only Second-Order CFA		$(f_1 = \beta_1 f_3 + e_1, \ f_2 = \beta_2 f_3 + e_2)$

3.1 Results

The analyses are originally performed in SPSS Amos and SAS Proc Callis, but I turn to R's Lavaan package to estimate the parameters and R^2 values once I choose which CFA model to pursue. I first assess the fit of each of the CFA models, then proceed to discussing the results of the model chosen.

In order to assess the models, I estimate four measures of fit. SAS and SPSS do not output the χ^2 statistic when the unweighted least square estimator is used. Because the χ^2 statistic is not always an accurate measure of model fit, I evaluate other measures, particularly the Adjusted Goodness of Fit (AGFI), the Normed-Fit Index (NFI), the Standardized Root Mean Square Residual (SRMR), and the Average Value Explained (AVE). The AGFI measures the "proportion of variance accounted for by the estimated population covariance" [**?**], but should be evaluated cautiously because it favors a more parsimonious model. The NFI gives the percentage that the model improves the fit from the null model. The SRMR measures the difference between the $\mathbf{S}$ matrix for the data and the $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ estimated from the model. Finally, the AVE is the average of the R^2 values in the factor, which indicate how much variation is explained for each variable by the factor(s).

Table **??** gives both the rules for each fit diagnostic, as well as the estimated values for each model. From these estimates, I find that all three of the models fit well. The two-factor models fit only marginally better than the one-factor model. However, because the theory points heavily towards the one-factor model, and it fits sufficiently for all estimated statistics, I determine that the one-factor model should be used.

Table 2: Model Fit for Different CFA's

CFA	Diagnostics			
	AGFI	NFI	SRMR	AVE
Rule for Good Fit	AGFI $\geq$ 0.90	NFI $\geq$ 0.95	SRMR < 0.08	AVE >0.5
One-Factor Model	0.989	0.990	0.053	0.499
Two-Factor Model	0.994	0.993	0.042	0.537
Two-Factor Second	0.993	0.993	0.044	0.548

The results of the one-factor model are represented in Figure **??** and Table **??**. I find that the RMSEA = 0.000 and CFI = 1.000, indicators that the model fits well. From Figure **??**, I observe that high levels of variation in the individual variables are described by the single factor. This is confirmed by the results in Table **??**. All p-values for the coefficient estimates in the CFA are significant, indicating that all of the variables are significantly affected by the factor. Additionally, many of the R^2 values are decently high, indicating that a large proportion of the variation in the individual variables can be explained by the factor. The exceptions to this are Rape Exemption and Son Preference, with R^2 values around 0.1. However, in general the single

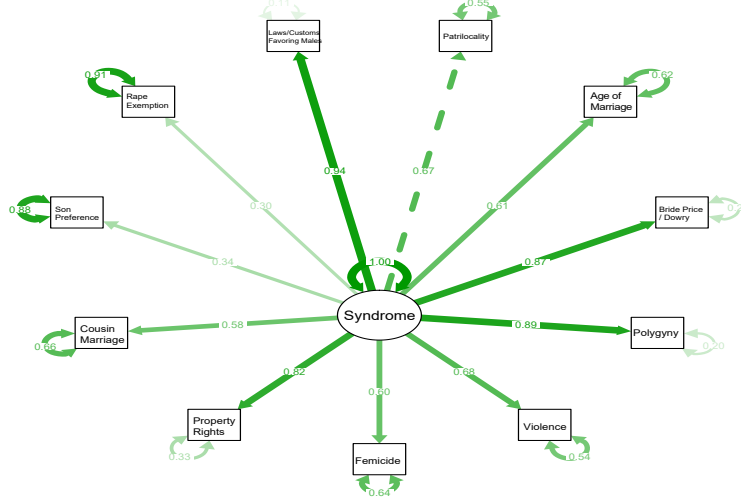**Confirmatory Factor Analysis Visual**

Figure 4: Results show that high amounts of variation in the indicators are described by the single factor. Rape Exemption and Son Preference are the exceptions.

factor appears to describe the variation in the variables well.

Table 3: Estimates from the One-Factor CFA Model

Variable	Estimate	Standard Error	z-value	p-value	R^2
Patrilocality	1.00				0.46
Age of Marriage	1.33	0.09	15.16	0.00	0.38
Bride Price/Dowry	1.519	0.10	15.70	0.00	0.75
Polygyny	2.32	0.14	16.58	0.00	0.80
Violence	0.956	0.072	13.37	0.00	0.46
Femicide	0.621	0.05	10.43	0.00	0.36
Property Rights	1.49	0.10	15.632	0.00	0.67
Cousin Marriage	1.27	0.09	14.95	0.00	0.34
Son Preference	0.45	0.06	8.20	0.00	0.12
Rape Exemption	0.23	0.05	4.59	0.00	0.09
Laws/Customs Favoring Males	1.84	0.11	16.28	0.00	0.89

3.2 Results for Partially-Simulated Data

One common issue with country-level data is that it is more difficult to collect data from less developed and more corrupt or violent countries. This difficulty results in one of two things: 1) data is missing altogether or 2) data is available but poorly measured or estimated. I therefore test the fit of the one-factor model from the CFA proposed and implemented above with the same dataset, but with random changes.

This is done by randomly selecting 1%, 5%, 10%, 25%, 50%, and 90% of the countries, using draws from a Bernoulli distribution, with probability of a success equal to each of these percentages. I then create two data sets for each percentage value. In the first, I delete the entire row of observations if the corresponding Bernoulli draw was 1. Then I create a second data set for each where if a row's corresponding Bernoulli draw is 1, I replace the observed values with randomly drawn values from a Binomial distribution, with the number of trials equal to the number of values in each scale.

I find that for both types of partially-simulated data, the RMSEA, CFI, and SRMR estimates stay essentially the same until the percentage level is high enough that the algorithm cannot estimate anymore.

When the data are randomly missing, the R^2 value estimates are similar and all variable estimates remain significant until 50% of the data is missing. Once that level is missing, the model will no longer estimate. When the data are randomly reassigned values, the R^2 values decrease as I increase the percentage of the data that is randomly reassigned. Again, once 50% of the data is "messy", I run into problems with estimation. I get negative R^2 values, and all of the variable estimates become insignificant in the model.

I find from this simulation study that lower levels of missing data or low-quality data are not too concerning in combining these variables into a single factor. However, once we get high levels of inaccurate estimates, we should be more cautious in the CFA model. Overall, the CFA model is fairly robust to the difficulties faced in measurements for country level data.

4   Discussion and Conclusion

In this study, I find that the combination of the 11 indicators of subordination of women at the household level into one factor is a well fit model, using a confirmatory factor analysis. While the path analysis does not confirm the theory of the causal relationship, I attribute that largely to the variables being structured differently than the phenomena in the theory. Future studies should evaluate the causal theory further by finding variables that can be better separated to fit the theory. I also recommend that future research evaluates how well the actual calculation of the scale represents its individual indicators. Overall, I find sufficient evidence to conclude that the single factor fits the data well and significantly describes variation in the individual variables. These results are robust to lower levels, at least 25% and under, of missing or poor-quality data.

References

[1] Albright, Jeremy J., and Hun Myoung Park. "Confirmatory factor analysis using amos, LISREL, Mplus, SAS/STAT CALIS." (2009).

[2] Forero, Carlos G., Alberto Maydeu-Olivares, and David Gallardo-Pujol. "Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation." Structural Equation Modeling 16, no. 4 (2009): 625-641.

[3] Li, Yingruolan Li. "Confirmatory Factor Analysis with Continuous and Ordinal Data: An Empirical Study of Stress Level." (2014).

[4] Parry, Stephen. "Fit Statistics commonly reported for CFA and SEM." Cornell University Statistical Consulting Unit. https://www.cscu.cornell.edu/news/Handouts/SEM_fit.pdf.

[5] WomanStats Project Database (2018). The WomanStats Project. http://www.womanstats.org.

A   R Code

```
setwd("~/Desktop/Kaylee Work and School/Current Classes/STAT 666/Final Project")
syndat <- read.csv("SyndromeVariables.csv")
head(syndat)

#####################################################################
### Path Analysis

library(lavaan)
library(semPlot)
library(GGally)

model <-'
Patri2016 ~ PRCom16
IFL2016 ~ AOM2015
AOM2015 ~ SonPref2015
BPD2017C ~ IFL2016 + Polygy2016
Polygy2016 ~ IFL2016
PSOW2014 ~ RapeXmp2016
HnrKlg2016 ~ BPD2017C
PRCom16 ~ PSOW2014
CsnMrg2016 ~ PRCom16
SonPref2015 ~ CsnMrg2016 + Patri2016
RapeXmp2016 ~ HnrKlg2016
'
fit <- cfa(model, data=syndat)
summary(fit,fit.measures = TRUE, standardized=T,rsquare=T)

semPaths(fit, 'std', layout = 'circle')

ggcorr(syndat[,2:12], nbreaks = 10, label = T, low = "red3", high = "green3",
        label_round = 2, name = "Correlation Scale", label_alpha = T, hjust = 0.75) +
  ggtitle(label = "Correlation Plot") +
  theme(plot.title = element_text(hjust = 0.6))

#####################################################################
### Confirmatory Factor Analysis
cfamodel <- '
syndrome =~ Patri2016 + IFL2016 + AOM2015 + BPD2017C + Polygy2016 +
PSOW2014 + HnrKlg2016 + PRCom16 + CsnMrg2016 + SonPref2015 + RapeXmp2016'
fitcfa <- cfa(cfamodel,data=syndat,estimator="ULS")
summary(fitcfa,fit.measures=TRUE, standardized=T,rsquare=T)
semPaths(fitcfa, 'std', layout = 'circle')

#####################################################################
### Simulation Study

# y1: 3 levels (0-2)
# y2: 5 levels (0-4)
# y3: 5 levels (0-4)
# y4: 4 levels (0-3)
# y5: 5 levels (0-4)
# y6: 4 levels (1-4)
```

```
# y7: 3 levels (0-2)
# y8: 5 levels (0-4)
# y9: 4 levels (0-3)
# y10: 5 levels (0-4)
# y11: 2 levels (0-1)

# different n sizes for binomial draws
k <- c(3,5,5,4,5,4,3,5,4,5,2)
syndat <- syndat[,2:12]

# observations that are missing or have messy data
set.seed(543675)
ranvals01 <- rbinom(176,1,0.01)
ranvals025 <- rbinom(176,1,0.025)
ranvals05 <- rbinom(176,1,0.05)
ranvals10 <- rbinom(176,1,0.1)
ranvals25 <- rbinom(176,1,0.25)
ranvals50 <- rbinom(176,1,0.5)
ranvals90 <- rbinom(176,1,0.9)

# assign new dataframes
SynVar.M01 <- SynVar.V01 <- syndat
SynVar.M025 <- SynVar.V025 <- syndat
SynVar.M05 <- SynVar.V05 <- syndat
SynVar.M10 <- SynVar.V10 <- syndat
SynVar.M25 <- SynVar.V25 <- syndat
SynVar.M50 <- SynVar.V50 <- syndat
SynVar.M90 <- SynVar.V90 <- syndat

# randomly choose observations that will be missing/random
for(i in 1:176){
  # 1% missing/messy
  if(ranvals01[i]==1){
    SynVar.M01[i,] <- rep(NA,11)
    for(j in 1:11) SynVar.V01[i,j] <- rbinom(1,k[j],0.5)
  }
  # 2.5% missing/messy
  if(ranvals025[i]==1){
    SynVar.M025[i,] <- rep(NA,11)
    for(j in 1:11) SynVar.V025[i,j] <- rbinom(1,k[j],0.5)
  }
  # 5% missing/messy
  if(ranvals05[i]==1){
    SynVar.M05[i,] <- rep(NA,11)
    for(j in 1:11) SynVar.V05[i,j] <- rbinom(1,k[j],0.5)
  }
  # 10% missing/messy
  if(ranvals10[i]==1){
    SynVar.M10[i,] <- rep(NA,11)
    for(j in 1:11) SynVar.V10[i,j] <- rbinom(1,k[j],0.5)
  }
  # 25% missing/messy
  if(ranvals25[i]==1){
    SynVar.M25[i,] <- rep(NA,11)
```

```
      for(j in 1:11) SynVar.V25[i,j] <- rbinom(1,k[j],0.5)
    }
    # 50% missing/messy
    if(ranvals50[i]==1){
      SynVar.M50[i,] <- rep(NA,11)
      for(j in 1:11) SynVar.V50[i,j] <- rbinom(1,k[j],0.5)
    }
    if(ranvals90[i]==1){
      SynVar.M90[i,] <- rep(NA,11)
      for(j in 1:11) SynVar.V90[i,j] <- rbinom(1,k[j],0.5)
  }
}


# write to files for SAS analysis
write.csv(SynVar.M01,"synvarm01.csv")
write.csv(SynVar.V01,"synvarv01.csv")
write.csv(SynVar.M025,"synvarm025.csv")
write.csv(SynVar.V025,"synvarv025.csv")
write.csv(SynVar.M05,"synvarm05.csv")
write.csv(SynVar.V05,"synvarv05.csv")
write.csv(SynVar.M10,"synvarm10.csv")
write.csv(SynVar.V10,"synvarv10.csv")
write.csv(SynVar.M25,"synvarm25.csv")
write.csv(SynVar.V25,"synvarv25.csv")
write.csv(SynVar.M50,"synvarm50.csv")
write.csv(SynVar.V50,"synvarv50.csv")




### Confirmatory Factor Analysis: 1% Random
fitcfa01 <- cfa(cfamodel,data=SynVar.V01,estimator="ULS")
summary(fitcfa01,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.061

### Confirmatory Factor Analysis: 2.5% Random
fitcfa025 <- cfa(cfamodel,data=SynVar.V025,estimator="ULS")
summary(fitcfa025,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.066

### Confirmatory Factor Analysis: 5% Random
fitcfa05 <- cfa(cfamodel,data=SynVar.V05,estimator="ULS")
summary(fitcfa05,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.069

### Confirmatory Factor Analysis: 10% Random
fitcfa10 <- cfa(cfamodel,data=SynVar.V10,estimator="ULS")
summary(fitcfa10,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.057

### Confirmatory Factor Analysis: 25% Random
fitcfa25 <- cfa(cfamodel,data=SynVar.V25,estimator="ULS")
summary(fitcfa25,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.066
```

```
### Confirmatory Factor Analysis: 50% Random
fitcfa50 <- cfa(cfamodel,data=SynVar.V50,estimator="ULS")
summary(fitcfa50,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.057


### Confirmatory Factor Analysis: 90% Random
fitcfa90 <- cfa(cfamodel,data=SynVar.V90,estimator="ULS")
summary(fitcfa90,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.066


#####################

### Confirmatory Factor Analysis: 1% Missing
fitcfa01m <- cfa(cfamodel,data=SynVar.M01,estimator="ULS")
summary(fitcfa01m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.061


### Confirmatory Factor Analysis: 2.5% Missing
fitcfa025m <- cfa(cfamodel,data=SynVar.M025,estimator="ULS")
summary(fitcfa025m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.063


### Confirmatory Factor Analysis: 5% Missing
fitcfa05m <- cfa(cfamodel,data=SynVar.M05,estimator="ULS")
summary(fitcfa05m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.063


### Confirmatory Factor Analysis: 10% Missing
fitcfa10m <- cfa(cfamodel,data=SynVar.M10,estimator="ULS")
summary(fitcfa10m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.057


### Confirmatory Factor Analysis: 25% Missing
fitcfa25m <- cfa(cfamodel,data=SynVar.V25,estimator="ULS")
summary(fitcfa25m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.060


### Confirmatory Factor Analysis: 50% Missing
fitcfa50m <- cfa(cfamodel,data=SynVar.M50,estimator="ULS")
summary(fitcfa50m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.057


### Confirmatory Factor Analysis: 90% Missing
fitcfa90m <- cfa(cfamodel,data=SynVar.V90,estimator="ULS")
summary(fitcfa90m,fit.measures=TRUE, standardized=T,rsquare=T)
# RMSE=0.000,CFI=1.000,SRMR=.066
```

B   SAS Code

```
proc import datafile= 'C:/Users/admin/Desktop/Kaylee Work and School/Current Classes/
  STAT 666/Final Project/SAS Analysis/SyndromeVariables.csv'
        out=work.SyndromeVariables
        dbms=CSV;
run;

title '1-factor CFA';
proc calis method=uls;
  lineqs
    y1 = lam1 f1 + e1,
        y2 = f1 + e2,
        y3 = lam3 f1 + e3,
        y4 = lam4 f1 + e4,
        y5 = lam5 f1 + e5,
        y6 = lam6 f1 + e6,
        y7 = lam7 f1 + e7,
        y8 = lam8 f1 + e8,
        y9 = lam9 f1 + e9,
        y10 = lam10 f1 + e10,
        y11 = lam11 f1 + e11;
  std
    e1-e11= psi1-psi11,
        f1 = phi1;
  bounds
    0 <= phi1,
        0 <= psi1-psi11;
run;

title '2-factor CFA';
proc calis method=uls;
  lineqs
    y1 = lam1 f1 + lam2 f2 + e1,
        y2 = lam3 f1 + lam4 f2 + e2,
        y3 = lam5 f1 + lam6 f2 + e3,
        y4 = lam7 f1 + lam8 f2 + e4,
        y5 = f1 + e5,
        y6 = lam9 f1 + lam10 f2 + e6,
        y7 = f2 + e7,
        y8 = lam11 f1 + lam12 f2 + e8,
        y9 = lam13 f1 + lam14 f2 + e9,
        y10 = lam15 f2 + e10,
        y11 = lam16 f1 + lam17 f2 + e11;
  std
    e1-e11= psi1-psi11,
        f1 f2 = phi1 phi2;
  cov
    f1 f2= phi12;
  bounds
    0 <= phi1-phi2,
        0 <= psi1-psi11;
run;
```

```
title 'Second-order CFA';
proc calis method=uls;
  lineqs
    y1 = lam1 f1 + lam2 f2 + e1,
        y2 = lam3 f1 + lam4 f2 + e2,
        y3 = lam5 f1 + lam6 f2 + e3,
        y4 = lam7 f1 + lam8 f2 + e4,
        y5 = f1 + e5,
        y6 = lam9 f1 + lam10 f2 + e6,
        y7 = f2 + e7,
        y8 = lam11 f1 + lam12 f2 + e8,
        y9 = lam13 f1 + lam14 f2 + e9,
        y10 = lam15 f2 + e10,
        y11 = lam16 f1 + lam17 f2 + e11,
        f1  = beta1 f3 + d1,
    f2  = beta2 f3 + d2;
  std
    e1-e11= psi1-psi11,
    d1-d2=ksi1-ksi2,
    f3 = phi3;
  bounds
        0 <= psi1-psi11,
        0 <= ksi1-ksi2,
    0 <= phi3;
run;
```

C SPSS "Code"