

Data Analytics

Bilal Hassan

22/12/2022

First we need to load the libraries.

```
library(lavaan)
```

```
## Warning: package 'lavaan' was built under R version 4.1.2
```

```
## This is lavaan 0.6-9  
## lavaan is FREE software! Please report any bugs.
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(knitr)  
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4  
## v tibble  3.1.4      v stringr 1.4.0  
## v tidyr   1.1.3      v forcats 0.5.1  
## v readr   2.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()    masks stats::lag()
```

```
dat <- PoliticalDemocracy  
dat %>% head() %>% kable()
```

y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3
2.50	0.000000	3.333333	0.000000	1.250000	0.000000	3.726360	3.333333	4.442651	3.637586	2.557615
1.25	0.000000	3.333333	0.000000	6.250000	1.100000	6.666666	0.736999	5.384495	5.062595	3.568079
7.50	8.800000	9.999998	9.199991	8.750000	8.094061	9.999998	8.211809	5.961005	6.255750	5.224433
8.90	8.800000	9.999998	9.199991	8.907948	8.127979	9.999998	4.615086	6.285998	7.567863	6.267495
10.00	3.333333	9.999998	6.666666	7.500000	3.333333	9.999998	6.666666	5.863631	6.818924	4.573679
7.50	3.333333	6.666666	6.666666	6.250000	1.100000	6.666666	0.368500	5.533389	5.135798	3.892270

Describe each variable

?PoliticalDemocracy

starting httpd help server ... done

```
# A data frame of 75 observations of 11 variables.
#
# y1
# Expert ratings of the freedom of the press in 1960
#
# y2
# The freedom of political opposition in 1960
#
# y3
# The fairness of elections in 1960
#
# y4
# The effectiveness of the elected legislature in 1960
#
# y5
# Expert ratings of the freedom of the press in 1965
#
# y6
# The freedom of political opposition in 1965
#
# y7
# The fairness of elections in 1965
#
# y8
# The effectiveness of the elected legislature in 1965
#
# x1
# The gross national product (GNP) per capita in 1960
#
# x2
# The inanimate energy consumption per capita in 1960
#
# x3
# The percentage of the labor force in industry in 1960
```

Correlations

```
dat %>% cor() %>% round(2) %>% kable()
```

	y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3
y1	1.00	0.60	0.68	0.69	0.74	0.65	0.67	0.67	0.38	0.32	0.25
y2	0.60	1.00	0.45	0.72	0.54	0.71	0.58	0.61	0.21	0.25	0.21
y3	0.68	0.45	1.00	0.61	0.58	0.43	0.65	0.53	0.33	0.31	0.23
y4	0.69	0.72	0.61	1.00	0.65	0.66	0.68	0.74	0.47	0.44	0.39
y5	0.74	0.54	0.58	0.65	1.00	0.56	0.68	0.63	0.56	0.52	0.43
y6	0.65	0.71	0.43	0.66	0.56	1.00	0.61	0.75	0.34	0.35	0.33
y7	0.67	0.58	0.65	0.68	0.68	0.61	1.00	0.71	0.39	0.40	0.35
y8	0.67	0.61	0.53	0.74	0.63	0.75	0.71	1.00	0.46	0.46	0.37
x1	0.38	0.21	0.33	0.47	0.56	0.34	0.39	0.46	1.00	0.89	0.80
x2	0.32	0.25	0.31	0.44	0.52	0.35	0.40	0.46	0.89	1.00	0.85
x3	0.25	0.21	0.23	0.39	0.43	0.33	0.35	0.37	0.80	0.85	1.00

Define a FA model

```
cfamodel <- 'economic factor=~ x1 + x2 + x3;  
democracy factor for 1960=~ y1 + y2 + y3 + y4;  
democracy factor for 1965=~ y5 + y6 + y7 + y8'
```

Factor Analysis

```
mycfa <- cfa(model=cfamodel, data = dat)  
mycfa %>% summary(fit.measures=T)
```

```

## lavaan 0.6-9 ended normally after 47 iterations
##
## Estimator ML
## Optimization method NLMINB
## Number of model parameters 25
##
## Number of observations 75
##
## Model Test User Model:
##
## Test statistic 72.462
## Degrees of freedom 41
## P-value (Chi-square) 0.002
##
## Model Test Baseline Model:
##
## Test statistic 730.654
## Degrees of freedom 55
## P-value 0.000
##
## User Model versus Baseline Model:
##
## Comparative Fit Index (CFI) 0.953
## Tucker-Lewis Index (TLI) 0.938
##
## Loglikelihood and Information Criteria:
##
## Loglikelihood user model (H0) -1564.959
## Loglikelihood unrestricted model (H1) -1528.728
##
## Akaike (AIC) 3179.918
## Bayesian (BIC) 3237.855
## Sample-size adjusted Bayesian (BIC) 3159.062
##
## Root Mean Square Error of Approximation:
##
## RMSEA 0.101
## 90 Percent confidence interval - lower 0.061
## 90 Percent confidence interval - upper 0.139
## P-value RMSEA <= 0.05 0.021
##
## Standardized Root Mean Square Residual:
##
## SRMR 0.055
##
## Parameter Estimates:
##
## Standard errors Standard
## Information Expected
## Information saturated (h1) model Structured
##
## Latent Variables:
## Estimate Std.Err z-value P(>|z|)
## economicfactor =~
## x1 1.000
## x2 2.182 0.139 15.714 0.000
## x3 1.819 0.152 11.956 0.000
## democracyfactorfor1960 =~

```

```
##      y1                1.000
##      y2                1.354    0.175    7.755    0.000
##      y3                1.044    0.150    6.961    0.000
##      y4                1.300    0.138    9.412    0.000
##      democracyfactorfor1965 =~
##      y5                1.000
##      y6                1.258    0.164    7.651    0.000
##      y7                1.282    0.158    8.137    0.000
##      y8                1.310    0.154    8.529    0.000
##
## Covariances:
##                                Estimate Std.Err  z-value  P(>|z|)
##      economicfactor ~~
##      dmcrcyfctr1960      0.660    0.206    3.202    0.001
##      dmcrcyfctr1965      0.774    0.208    3.715    0.000
##      democracyfactorfor1960 ~~
##      dmcrcyfctr1965      4.487    0.911    4.924    0.000
##
## Variances:
##                                Estimate Std.Err  z-value  P(>|z|)
##      .x1                0.082    0.020    4.180    0.000
##      .x2                0.118    0.070    1.689    0.091
##      .x3                0.467    0.090    5.174    0.000
##      .y1                1.942    0.395    4.910    0.000
##      .y2                6.490    1.185    5.479    0.000
##      .y3                5.340    0.943    5.662    0.000
##      .y4                2.887    0.610    4.731    0.000
##      .y5                2.390    0.447    5.351    0.000
##      .y6                4.343    0.796    5.456    0.000
##      .y7                3.510    0.668    5.252    0.000
##      .y8                2.940    0.586    5.019    0.000
##      economicfactor      0.448    0.087    5.169    0.000
##      dmcrcyfctr1960      4.845    1.088    4.453    0.000
##      dmcrcyfctr1965      4.345    1.051    4.134    0.000
```

Goodness of fit

```
mycfa %>% fitmeasures(c('chisq', 'pvalue', 'cfi', 'tli', 'rmsea', 'srmr'))
```

```
##  chisq pvalue   cfi   tli  rmsea  srmr
## 72.462  0.002  0.953  0.938  0.101  0.055
```

cfi, tli, rmsea, srmr: good.

Factor loadings

```
mycfa %>% coef() %>% kable()
```

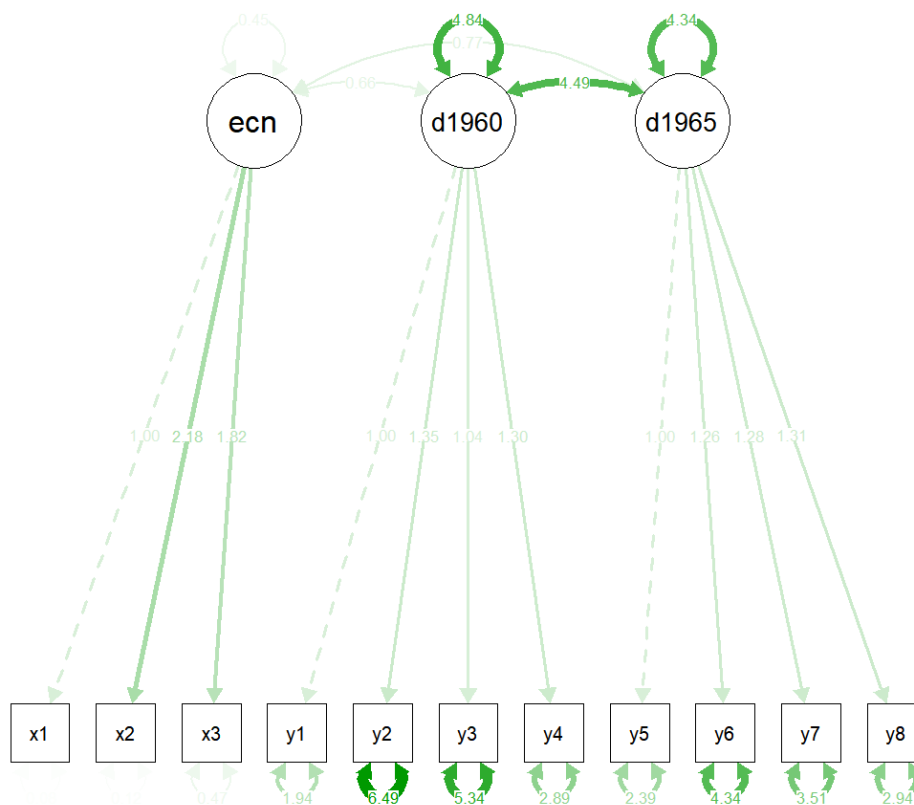
	x
economicfactor=~x2	2.18174687
economicfactor=~x3	1.81876676
democracyfactorfor1960=~y2	1.35400730

democracyfactorfor1960=~y3	1.04400960
democracyfactorfor1960=~y4	1.29953852
democracyfactorfor1965=~y6	1.25847783
democracyfactorfor1965=~y7	1.28248550
democracyfactorfor1965=~y8	1.30977070
x1~~x1	0.08182862
x2~~x2	0.11843712
x3~~x3	0.46720892
y1~~y1	1.94188425
y2~~y2	6.48962793
y3~~y3	5.33992561
y4~~y4	2.88715966
y5~~y5	2.39011846
y6~~y6	4.34288264
y7~~y7	3.50957687
y8~~y8	2.94032667
economicfactor~~economicfactor	0.44815787
democracyfactorfor1960~~democracyfactorfor1960	4.84496934
democracyfactorfor1965~~democracyfactorfor1965	4.34456222
economicfactor~~democracyfactorfor1960	0.66046897
economicfactor~~democracyfactorfor1965	0.77403436
democracyfactorfor1960~~democracyfactorfor1965	4.48732293

The first factor correlates most highly with x2, second with y2, and the third with y8.

Plot

```
library(semPlot)
semPaths(mycfa, 'par')
```



Predict the latent variables

```
mycfa %>% lavPredict() %>% summary() %>% kable()
```

economicfactor	democracyfactorfor1960	democracyfactorfor1965
Min. :-1.31339	Min. :-3.7447	Min. :-3.7825
1st Qu.: -0.54193	1st Qu.: -1.6449	1st Qu.: -1.6593
Median : 0.09651	Median : -0.5155	Median : -0.5598
Mean : 0.00000	Mean : 0.0000	Mean : 0.0000
3rd Qu.: 0.42491	3rd Qu.: 1.8412	3rd Qu.: 1.7942
Max. : 1.31410	Max. : 4.1268	Max. : 4.0814

```
latent <- mycfa %>% lavPredict()
```

The minimum value for the economic latent variable is -1.31. This mean subject is 1.31 standard deviation below the mean