DA410_Project8_MattGraham

Confirmatory Factor Analysis (CFA)

EFA is used to explore possible underlying factor structure of a set of observed variables without imposing a preconceived structure of outcome.

CFA is a stat technique to verify the factor strucutre of a set of observed variables. Allows peeps to test hypotheses of relations between observed variables and underlying structure. This is provided by experience, theory, or research. Mostly prior=known and test to see if statistically significant.

Similar to EFA - CFA uses the common factor model. It uses covariance between observed variables as reflection of the influence of one or more factors, and if a variance is not explained. Different from network as it allows covariance between items to have a cost. It is believed covariance of items has a latent factor that explains it.

If not common factor, it may be network model.

What is it and when do we apply CFA? - Serves to estimate structure of instrument to determine how well measured variables measures number of constructs - Must state which instrument to test - generally used when previous study shows previous measurement of instrument - Use case: Explaining how well structure of US measurements fit Brazilian - Use case: Explore dimensionality and test different models w/ CFA

Fit test: How well our model we are estimating with testing. E.g. A well fitting model reduces discrepancy between population covariance matrix and S matrix, or sample covariance matrix.

Can have great fit, but it does not apply to the population.

```r
library(nnspat)  # used for dist2full()
library("dplyr")  # used to select numeric datatypes
library("ggplot2")
library(reshape)  # used for melting matricies
library(klaR)
library(ggvis)
library(class)
library(gmodels)
library(MASS)
library(readxl)
library(psych)
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.2.2
```

```r
library(lavaan)
```

```
## Warning: package 'lavaan' was built under R version 4.2.2
```

```r
library(semPlot)
```

```
## Warning: package 'semPlot' was built under R version 4.2.2
```

```
library(semTable)
```

```
## Warning: package 'semTable' was built under R version 4.2.2
```

```
library(kutils)
```

```
## Warning: package 'kutils' was built under R version 4.2.2
```

Get data

```
probe <- read.table("C:/mattgraham93.github.io/school/22_3_DA410/data/T3_6_PROBE.DAT", header
=FALSE)[-1]
colnames(probe) <- c('r_time1', 'r_time2', 'r_time3', 'r_time4', 'r_time5')
probe
```

r_time1 <int>	r_time2 <int>	r_time3 <int>	r_time4 <int>	r_time5 <int>
51	36	50	35	42
27	20	26	17	27
37	22	41	37	30
42	36	32	34	27
27	18	33	14	29
43	32	43	35	40
41	22	36	25	38
38	21	31	20	16
36	23	27	25	28
26	31	31	32	36

1-10 of 11 rows Previous **1** 2 Next

Correlation matrix

```
probe.cor <- cor(probe)
probe.cor
```

```
##           r_time1   r_time2   r_time3   r_time4   r_time5
## r_time1 1.0000000 0.6143902 0.7571850 0.5750730 0.4130573
## r_time2 0.6143902 1.0000000 0.5473897 0.7497770 0.5476595
## r_time3 0.7571850 0.5473897 1.0000000 0.6052716 0.6918927
## r_time4 0.5750730 0.7497770 0.6052716 1.0000000 0.5238876
## r_time5 0.4130573 0.5476595 0.6918927 0.5238876 1.0000000
```

Covariance matrix

```
probe.cov <- var(probe)
probe.cov
```

```
##          r_time1  r_time2  r_time3  r_time4  r_time5
## r_time1 65.09091 33.64545 47.59091 36.77273 25.42727
## r_time2 33.64545 46.07273 28.94545 40.33636 28.36364
## r_time3 47.59091 28.94545 60.69091 37.37273 41.12727
## r_time4 36.77273 40.33636 37.37273 62.81818 31.68182
## r_time5 25.42727 28.36364 41.12727 31.68182 58.21818
```

Hypothesis

Ho: The model is a perfect fit

Ha: The model is not a perfect fit

Testing

```
probe.fit1 <- factanal(covmat=probe.cor, factors=1, rotation="none", n.obs = 55)
probe.fit1
```

```
##
## Call:
## factanal(factors = 1, covmat = probe.cor, n.obs = 55, rotation = "none")
##
## Uniquenesses:
## r_time1 r_time2 r_time3 r_time4 r_time5
##   0.373   0.409   0.276   0.395   0.515
##
## Loadings:
##         Factor1
## r_time1 0.792
## r_time2 0.769
## r_time3 0.851
## r_time4 0.778
## r_time5 0.696
##
##                Factor1
## SS loadings      3.032
## Proportion Var   0.606
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 33.45 on 5 degrees of freedom.
## The p-value is 3.06e-06
```

```
probe.fit2 <- factanal(covmat=probe.cov, factors=1, rotation="none", n.obs = 55)
probe.fit2
```

```
##
## Call:
## factanal(factors = 1, covmat = probe.cov, n.obs = 55, rotation = "none")
##
## Uniquenesses:
## r_time1 r_time2 r_time3 r_time4 r_time5
##   0.373   0.409   0.276   0.395   0.515
##
## Loadings:
##         Factor1
## r_time1 0.792
## r_time2 0.769
## r_time3 0.851
## r_time4 0.778
## r_time5 0.696
##
##                Factor1
## SS loadings      3.032
## Proportion Var   0.606
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 33.45 on 5 degrees of freedom.
## The p-value is 3.06e-06
```

Assess the goodness of fit

```
model <- "f1=~r_time1+r_time2+r_time3+r_time4+r_time5"

fit1 <- cfa(model=model, data=probe, likelihood="wishart")
summary(fit1, fit.measures=TRUE, standardized=TRUE)
```

```
## lavaan 0.6-12 ended normally after 46 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        10
##
##   Number of observations                            11
##
## Model Test User Model:
##
##   Test statistic                                 6.581
##   Degrees of freedom                                 5
##   P-value (Chi-square)                           0.254
##
## Model Test Baseline Model:
##
##   Test statistic                                31.962
##   Degrees of freedom                                10
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.928
##   Tucker-Lewis Index (TLI)                       0.856
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)               -175.822
##   Loglikelihood unrestricted model (H1)       -172.202
##
##   Akaike (AIC)                                 371.643
##   Bayesian (BIC)                               375.622
##   Sample-size adjusted Bayesian (BIC)          345.512
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.178
##   90 Percent confidence interval - lower         0.000
##   90 Percent confidence interval - upper         0.500
##   P-value RMSEA <= 0.05                          0.267
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.071
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
```

```
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   f1 =~
##     r_time1           1.000                               6.389    0.792
##     r_time2           0.816    0.320    2.548    0.011    5.217    0.769
##     r_time3           1.037    0.362    2.862    0.004    6.629    0.851
##     r_time4           0.965    0.373    2.585    0.010    6.165    0.778
##     r_time5           0.832    0.367    2.267    0.023    5.314    0.696
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .r_time1          24.271   13.777    1.762    0.078   24.271    0.373
##    .r_time2          18.861   10.300    1.831    0.067   18.861    0.409
##    .r_time3          16.752   11.200    1.496    0.135   16.752    0.276
##    .r_time4          24.812   13.743    1.805    0.071   24.812    0.395
##    .r_time5          29.977   15.181    1.975    0.048   29.977    0.515
##     f1               40.820   28.311    1.442    0.149    1.000    1.000
```

Discussion

<<insertHTML:[probe.fitted.html]

```
htmltools::includeHTML("probe.fitted.html")
```

	Model			
	Estimate	Std. Err.	z	p
	Factor Loadings			
f1				
r.time1	1.00+			
r.time2	0.82	0.32	2.55	.011
r.time3	1.04	0.36	2.86	.004
r.time4	0.96	0.37	2.59	.010
r.time5	0.83	0.37	2.27	.023
	Residual Variances			
r.time1	24.27	13.78	1.76	.078
r.time2	18.86	10.30	1.83	.067
r.time3	16.75	11.20	1.50	.135
r.time4	24.81	13.74	1.81	.071
r.time5	29.98	15.18	1.97	.048
	Latent Variances			
f1	40.82	28.31	1.44	.149
	Fit Indices			
χ^2	6.58(5)			.254
CFI	0.93			
TLI	0.86			
RMSEA	0.18			

+Fixed parameter

Summary

When looking at our output, we can conclude that our chi-squared is valuable and can state our model is generally adequate when looking at differencs between expected and observed covariances. With TLI and CFA falling under 0.95, we cannot state the model is adequate for 95% in relation, which is compounded through our RMSEA value, concluding our fit index model is inadequate. Finally, our SRMR is adequate, which makes sense as this is most-related to chi-squared.