Multivariate Statistics – Assignment 1 (2025-2026)

Ward Vandecruys Paolo Rinaldi Andrea Osmani Duc Nguyen

November 8, 2025

1 Task 1: Principal Component Analysis

1.1 Loading the data and running the PCA

```r
library(rstudioapi)
library(paran)

## Set working directory to the script location
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
load("life.Rdata")

## Standardize
zlife <- scale(life, center = TRUE, scale = TRUE)

## PCA
pca_life <- prcomp(zlife)

# Eigenvalues, proportion and cumulative proportion of explained variance
eigvals <- pca_life$sd^2
round(eigvals, 3)
round(eigvals / ncol(zlife), 3)
round(cumsum(eigvals / ncol(zlife)), 3)
```

Output:

Table 1: Eigenvalues and explained variance for the six principal components.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	2.400	1.589	0.900	0.498	0.384	0.229
Proportion of variance	0.400	0.265	0.150	0.083	0.064	0.038
Cumulative proportion	0.400	0.665	0.815	0.898	0.962	1.000

Interpretation: Each eigenvalue denotes the corresponding variance of the principal component. When we look at table 1 we see that the first and second principle component explain 66.5% of the total variance in our 6 dimensional dataset. This goes up to 81.5% when we include the third principle component, the remaining components give little additional information about the total variance. While at first glance it seems logical to select the first three principal components, we will rely on Horn's procedure to determine which principle components are meaningful enough to include.

1.2 Horn's procedure

```r
set.seed(1425) # Reproducibility
paran(zlife, iterations = 5000, graph = TRUE,
      cfa = FALSE, centile = 0)
```

Output:

Table 2: Results of Horn's parallel analysis (5000 iterations).

Component	Adjusted eigenvalue	Unadjusted eigenvalue	Bias
1	1.914047	2.399644	0.485596
2	1.351613	1.588934	0.237320

Interpretation: After conduction Horn's parallel analysis we can conclude that only the first two components are significant enough to retain in the analysis. This could already be concluded from the table 1 because only the first two components had values larger than one, and were therefore potential candidates under Horn's criterion.

1.3 Loadings and component scores

```
## Loadings (correlations between variables and components)
A <- pca_life$rotation %*% diag(pca_life$sd)
round(A, 2)

## Standardized component scores
Z <- predict(pca_life)
Zs <- Z %*% diag(1 / pca_life$sd)
round(Zs, 2)
```

Output: Matrix A contains the loadings of the six original variables, which are listed in Table 3a. Due to the large number of countries in our dataset we have only listed an excerpt of the standardized component scores on PC1 and PC2 in Table 3b.

(a) Component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6
family	0.73	-0.37	-0.16	-0.48	-0.23	-0.11
friends	-0.33	-0.82	0.12	-0.19	0.41	0.00
leisure_time	-0.48	-0.25	-0.83	0.03	-0.06	0.15
politics	0.15	-0.85	0.17	0.40	-0.26	-0.02
work	0.81	0.05	-0.38	0.26	0.27	-0.23
religion	0.92	-0.03	0.05	0.03	0.11	0.37

(b) Standardized scores on PC1 and PC2 (selected countries).

Country	PC1	PC2
Australia	-1.37	-1.00
Philippines	2.01	-0.11
Qatar	1.03	-2.28
⋮	⋮	⋮
Sweden	-1.44	-1.62
Netherlands	-2.02	-0.01
Morocco	1.17	2.45

Table 3: PCA loadings and country scores.

Interpretation: From these results we can conclude that the first principle component (PC1) describes the effect of traditional versus modern values. Where the loadings of traditional values such as religion (0.92), work (0.81), and family (0.73) show strong positive values in PC1 as opposed to the more modern values such as leisure time (-0.48) and friends (-0.33) have negative loadings. The second principle component is strongly aligned with politics(-0.85) and friends (-0.82), where negative scores show a higher importance towards both factors.

By using these loadings we can now reference the scores in Table 3b to show how countries position themselves on these two dimensions. Countries such as the Philippines, Qatar and Morocco score highly on PC1 showing a strong 'importance' on family, work and religion. Whereas countries such as the Netherlands (-2.02), Sweden (-1.44), and Australia (-1.37) have negative PC1 scores and therefore value leisure time and friends more. In terms of PC2 we note that Qatar (-2.28) has a very low score, meaning a high importance towards politics.

1.4 Biplot

```r
par(cex = 0.8)
biplot(pca_life, pc.biplot = TRUE,
       xlim = c(-2.6, 2.6),
       ylim = c(-2.6, 2.6))
```

Output: The biplot of the first two components is shown in Figure 1.

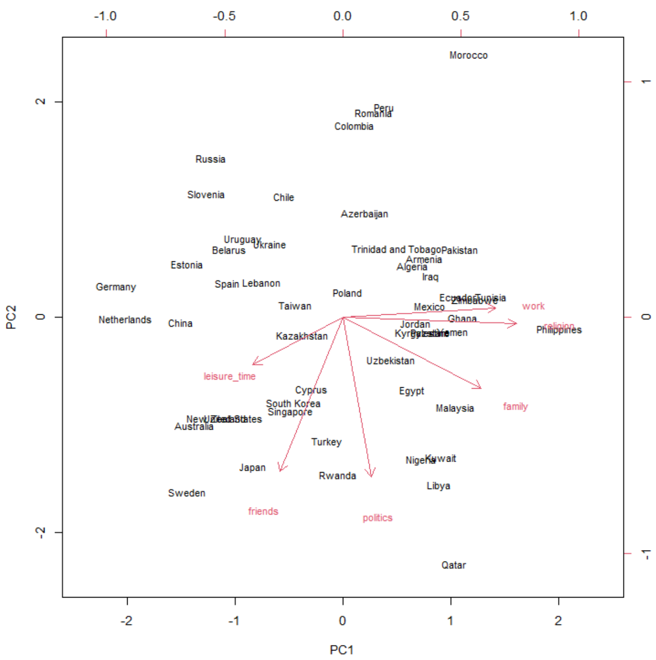

Figure 1: Biplot of the first two principal components.

Interpretation: The biplot in figure 1 visualizes the interpretation in the loading and component scores section. We can indeed see that the Philippines, Qatar, Morocco are positioned to the right side of the biplot showing the emphasis on the listed traditional values. It is interesting to notice that for example countries such as Qatar and Morocco, who scored very similarly on PC1, score strongly opposite of each other on PC2. Note that countries near the origin (such as Poland) display a balanced profile across the 6 life variables in the dataset.

2 Task 2: Factor structure and prediction of financial security

2.1 2a. Exploratory factor analysis (EFA)

```
library(psych)
library(GPArotation)
library(lavaan)
library(rstudioapi)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
load("fsdata.Rdata")

items <- c(
  "FS_pay_bills", "FS_afford_extras", "FS_afford_housing", "FS_save_money",
  "FSF_pay_bills", "FSF_afford_extras", "FSF_afford_housing", "FSF_save_money",
  "SFJ_no_info", "SFJ_no_chance_show", "SFJ_no_training", "SFJ_no_support_findjob",
  "SDJ_help_people", "SDJ_learn_new_things", "SDJ_develop_creativity",
  "SDJ_meet_people", "SDJ_feeling_self_worth",
  "HEALTH_felt_down", "HEALTH_limitation"
)

## standardized / centered versions
fs_complete <- na.omit(fsdata[, c("country", items)])
zfs <- fs_complete
zfs[, items] <- scale(fs_complete[, items], center = TRUE, scale = TRUE)
cfs <- fs_complete
cfs[, items] <- scale(fs_complete[, items], center = TRUE, scale = FALSE)

## 5-factor ML EFA with oblimin rotation
cormat_fs <- cor(zfs[, items])

efa5_obl <- fa(cormat_fs,
               nfactors = 5,
               rotate = "oblimin",
               fm = "ml",
               covar = FALSE,
               n.obs = nrow(zfs))

efa5_obl # pattern, fit indices, factor correlations
round(efa5_obl$communalities, 3)
print(efa5_obl$Structure, digits = 3, cutoff = 0)
round(efa5_obl$residual, 3)

p <- length(items)
resid <- efa5_obl$residual - diag(diag(efa5_obl$residual))
prop_large_resid <- sum(abs(resid) > 0.05) / (p * (p - 1))
prop_large_resid
```

Output:

Table 4: EFA pattern matrix (ML, oblimin), communalities h^2, uniqueness u^2, and complexity.

Item	ML1	ML2	ML3	ML5	ML4	h^2	u^2
FS_pay_bills	-0.03	0.56	-0.02	0.07	-0.20	0.37	0.63
FS_afford_extras	0.00	0.87	0.01	-0.02	0.02	0.75	0.25
FS_afford_housing	0.05	0.57	-0.02	0.00	-0.16	0.40	0.60
FS_save_money	0.02	0.77	0.01	0.00	0.09	0.58	0.42
FSF_pay_bills	0.69	-0.03	-0.02	0.04	-0.15	0.51	0.49
FSF_afford_extras	0.86	0.03	0.01	-0.02	0.07	0.74	0.26
FSF_afford_housing	0.68	0.01	-0.01	0.04	-0.22	0.55	0.45
FSF_save_money	0.85	0.01	0.01	-0.02	0.11	0.72	0.28
SFJ_no_info	-0.01	0.05	0.05	0.48	0.15	0.28	0.72
SFJ_no_chance_show	0.02	-0.03	-0.01	0.68	-0.06	0.45	0.55
SFJ_no_training	-0.01	0.01	0.01	0.64	0.06	0.43	0.57
SFJ_no_support_findjob	-0.03	-0.02	0.00	0.56	0.04	0.33	0.67
SDJ_help_people	-0.03	-0.01	0.61	0.00	0.11	0.40	0.60
SDJ_learn_new_things	0.00	0.00	0.63	0.04	-0.12	0.42	0.58
SDJ_develop_creativity	0.01	-0.03	0.59	0.03	0.06	0.37	0.63
SDJ_meet_people	0.02	0.03	0.65	-0.07	0.04	0.41	0.59
SDJ_feeling_self_worth	0.01	-0.01	0.54	0.06	-0.18	0.34	0.66
HEALTH_felt_down	-0.01	-0.08	0.01	0.07	0.71	0.56	0.44
HEALTH_limitation	0.00	0.02	-0.01	0.03	0.74	0.55	0.45

Table 5: EFA SS loadings and explained variance.

	ML1	ML2	ML3	ML5	ML4
SS loadings	2.43	2.03	1.85	1.47	1.38
Proportion Var	0.13	0.11	0.10	0.08	0.07
Cumulative Var	0.13	0.23	0.33	0.41	0.48
Proportion Explained	0.27	0.22	0.20	0.16	0.15
Cumulative Proportion	0.27	0.49	0.69	0.85	1.00

Table 6: EFA factor correlations (oblimin).

	ML1	ML2	ML3	ML5	ML4
ML1	1.00	0.31	0.04	-0.07	-0.13
ML2	0.31	1.00	0.00	-0.22	-0.19
ML3	0.04	0.00	1.00	0.29	0.04
ML5	-0.07	-0.22	0.29	1.00	0.16
ML4	-0.13	-0.19	0.04	0.16	1.00

Table 7: EFA structure matrix.

Item	ML1	ML2	ML3	ML5	ML4
FS_pay_bills	0.163	0.576	-0.009	-0.091	-0.288
FS_afford_extras	0.264	0.868	0.001	-0.207	-0.145
FS_afford_housing	0.246	0.614	-0.028	-0.156	-0.272
FS_save_money	0.243	0.756	0.012	-0.156	-0.060
FSF_pay_bills	0.700	0.208	0.010	-0.040	-0.236
FSF_afford_extras	0.857	0.282	0.040	-0.078	-0.050
FSF_afford_housing	0.707	0.252	0.017	-0.053	-0.303
FSF_save_money	0.841	0.254	0.040	-0.064	-0.003
SFJ_no_info	-0.048	-0.083	0.191	0.508	0.226
SFJ_no_chance_show	-0.034	-0.158	0.184	0.669	0.055
SFJ_no_training	-0.056	-0.138	0.200	0.651	0.165
SFJ_no_support_findjob	-0.080	-0.160	0.166	0.571	0.141
SDJ_help_people	-0.021	-0.037	0.619	0.201	0.146
SDJ_learn_new_things	0.034	0.016	0.636	0.204	-0.088
SDJ_develop_creativity	0.015	-0.046	0.599	0.220	0.099
SDJ_meet_people	0.053	0.044	0.632	0.116	0.051
SDJ_feeling_self_worth	0.045	0.017	0.550	0.187	-0.149
HEALTH_felt_down	-0.136	-0.233	0.057	0.206	0.740
HEALTH_limitation	-0.092	-0.124	0.030	0.142	0.744

Table 8: Residual correlation matrix from 5-factor EFA (excerpt).

	FS_pay_bills	FS_aff_ext	FS_aff_hsng	FS_save	FSF_pay	FSF_aff_ext	FSF_aff_hsng	FSF_save	SFJ_no_info	SFJ_no_chance
FS_pay_bills	0.631	-0.004	0.076	-0.026	0.082	-0.024	0.020	-0.024	-0.014	-0.004
FS_afford_extras	-0.004	0.246	-0.015	0.008	-0.012	0.019	-0.012	-0.009	0.000	0.000
FS_afford_housing	0.076	-0.015	0.595	-0.008	-0.016	-0.011	0.069	-0.013	0.003	-0.002
FS_save_money	-0.026	0.008	-0.008	0.422	-0.006	-0.021	-0.016	0.034	0.006	0.001
FSF_pay_bills	0.082	-0.012	-0.016	-0.006	0.486	-0.020	0.098	-0.025	-0.006	-0.004
FSF_afford_extras	-0.024	0.019	-0.011	-0.021	-0.020	0.260	-0.022	0.020	0.000	0.002
FSF_afford_housing	0.020	-0.012	0.069	-0.016	0.098	-0.022	0.454	-0.023	-0.007	-0.004
FSF_save_money	-0.024	-0.009	-0.013	0.034	-0.025	0.020	-0.023	0.281	0.005	0.002
SFJ_no_info	-0.014	0.000	0.003	0.006	-0.006	0.000	-0.007	0.005	0.716	0.029
SFJ_no_chance_show	-0.004	0.000	-0.002	0.001	-0.004	0.002	-0.004	0.002	0.029	0.549

	SFJ_no_trn	SFJ_no_sup	SDJ_help	SDJ_learn	SDJ_dev	SDJ_meet	SDJ_self	HEALTH_down	HEALTH_lim
SFJ_no_training	-0.007	0.001	-0.001	0.007	0.009	-0.014	0.013	0.021	0.031
SFJ_no_support_findjob	0.004	0.670	0.012	0.009	-0.003	-0.015	-0.004	0.000	-0.002
SDJ_help_people	0.004	0.012	0.602	-0.045	-0.038	0.101	-0.034	-0.008	-0.006
SDJ_learn_new_things	-0.016	0.009	-0.045	0.581	0.075	-0.055	0.049	0.011	0.011
SDJ_develop_creativity	0.003	-0.003	-0.038	0.075	0.634	-0.039	0.001	-0.002	0.014
SDJ_meet_people	0.017	-0.015	0.101	-0.055	-0.039	0.593	-0.014	-0.015	-0.018
SDJ_feeling_self_worth	-0.012	-0.004	-0.034	0.049	0.001	-0.014	0.664	0.025	0.007
HEALTH_felt_down	-0.013	0.000	-0.008	0.011	-0.002	-0.015	0.025	0.438	0.035
HEALTH_limitation	-0.007	-0.002	-0.006	0.011	0.014	-0.018	0.007	0.035	0.446

Table 9: Global fit indices for the 5-factor EFA model

Statistic	Value	Note
Proportion $\lvert r_{\text{resid}} \rvert > 0.05$	0.041	Off-diagonal residuals
RMSR	0.02	
df-corrected RMSR	0.03	
TLI	0.905	
RMSEA	0.057	90% CI [0.054, 0.059]
Likelihood χ^2	1417.34	$df = 86$, $p < 10^{-200}$
BIC	687.85	
Fit (off-diagonal)	0.99	Model vs. observed R

Table 10: Measures of factor score adequacy.

Measure	ML1	ML2	ML3	ML5	ML4
Correlation of scores with factors	0.94	0.92	0.87	0.85	0.86
Multiple R^2 of scores with factors	0.88	0.85	0.76	0.72	0.75
Minimum correlation of possible scores	0.77	0.70	0.51	0.43	0.49

Interpretation:

From these results, we can interpret that we recover the structure that we intended. Table 4 shows that ML1 captures future financial security (FSF), here all four items have large primary loadings between 0.68 and 0.86 while also having high communalities between 0.51 and 0.72. ML2 does the same for the current financial security (FS) with loadings between 0.56 and 0.87 and communalities between 0.37 and 0.75. The third captures job satisfaction/decent work (SDJ) (0.48–0.68), the fourth captures health problems (0.71–0.74), and lastly, the fifth captures job search difficulties (SFJ) (0.48–0.68). The overall range of the communalities of each factor shows that the item variance is explained adequately. Due to the oblique rotation, we allow for correlation between factors. We can see this in a positive correlation of about 0.31 between current and future financial security. Financial security is also shown to negatively correlate with job and health problems overall. Correspondingly, we do notice that SDJ positively correlates with SFJ and HEALTH. Overall, we can say that the model produces the observed correlation matrix pretty well, because we see that only 4% of the residual correlations exceed the |0.05| threshold. The RMSR (0.02), RMSEA (0.06), and TLI (0.91) all indicate that this is an acceptable fit. What

2.2 2b. Confirmatory factor analysis (CFA)

```
cfa5 <- '
  FS =~ NA*FS_pay_bills + FS_afford_extras + FS_afford_housing + FS_save_money
  FSF =~ NA*FSF_pay_bills + FSF_afford_extras + FSF_afford_housing + FSF_save_money
  SFJ =~ NA*SFJ_no_info + SFJ_no_chance_show + SFJ_no_training + SFJ_no_support_findjob
  SDJ =~ NA*SDJ_help_people + SDJ_learn_new_things + SDJ_develop_creativity +
            SDJ_meet_people + SDJ_feeling_self_worth
  HEALTH =~ NA*HEALTH_felt_down + HEALTH_limitation

  FS ~~ 1*FS
  FSF ~~ 1*FSF
  SFJ ~~ 1*SFJ
  SDJ ~~ 1*SDJ
  HEALTH ~~ 1*HEALTH

  FS ~~ FSF + SFJ + SDJ + HEALTH
  FSF ~~ SFJ + SDJ + HEALTH
  SFJ ~~ SDJ + HEALTH
  SDJ ~~ HEALTH
'

fit_cfa5 <- cfa(cfa5, data = cfs)

summary(fit_cfa5, fit.measures = TRUE, standardized = TRUE)
fitmeasures(fit_cfa5,
           c("chisq","df","pvalue","cfi","tli","rmsea","srmr"))

d <- standardizedSolution(fit_cfa5)

compositereliability <- function(lambda) {
  A <- (sum(lambda))^2
  B <- sum(1 - lambda^2)
  A / (A + B)
}

loadings_std <- subset(d, op == "=~", select = c("lhs","rhs","est.std"))
cov_std <- subset(d, op == "~~" &
                      lhs %in% c("FS","FSF","SFJ","SDJ","HEALTH") &
                      rhs %in% c("FS","FSF","SFJ","SDJ","HEALTH"))

factors <- c("FS","FSF","SFJ","SDJ","HEALTH")
AVE <- CR <- MSV <- numeric(length(factors))

for (i in seq_along(factors)) {
  f <- factors[i]
  lambda_f <- loadings_std$est.std[loadings_std$lhs == f]
  AVE[i] <- mean(lambda_f^2)
  CR[i] <- compositereliability(lambda_f)

  cor_f <- cov_std$est.std[cov_std$lhs == f & cov_std$lhs != cov_std$rhs]
  MSV[i] <- if (length(cor_f) > 0) max(cor_f^2) else 0
}

report <- data.frame(
  Factor = factors,
  AVE = round(AVE, 3),
  MSV = round(MSV, 3),
```

```
    CR = round(CR, 3)
)
report
```

Output:

Table 11: Fit indices for the 5-factor CFA model.

Statistic	Value
χ^2 (df = 142)	2785.51
CFI	0.906
TLI	0.886
RMSEA	0.062 (90% CI [0.060, 0.064])
SRMR	0.044
AIC	203341.49
BIC	203652.64

Table 12: Standardized factor loadings (Std.all) for the CFA model.

Factor	Item	Loading
FS	FS_pay_bills	0.583
FS	FS_afford_extras	0.860
FS	FS_afford_housing	0.626
FS	FS_save_money	0.747
FSF	FSF_pay_bills	0.697
FSF	FSF_afford_extras	0.860
FSF	FSF_afford_housing	0.701
FSF	FSF_save_money	0.834
SFJ	SFJ_no_info	0.526
SFJ	SFJ_no_chance_show	0.637
SFJ	SFJ_no_training	0.665
SFJ	SFJ_no_support_findjob	0.580
SDJ	SDJ_help_people	0.607
SDJ	SDJ_learn_new_things	0.643
SDJ	SDJ_develop_creativity	0.606
SDJ	SDJ_meet_people	0.614
SDJ	SDJ_feeling_self_worth	0.552
HEALTH	HEALTH_felt_down	0.955
HEALTH	HEALTH_limitation	0.613

Table 13: Latent factor correlations (Std.all).

	FS	FSF	SFJ	SDJ	HEALTH
FS	1	0.332	-0.237	-0.004	-0.260
FSF	0.332	1	-0.093	0.043	-0.142
SFJ	-0.237	-0.093	1	0.318	0.260
SDJ	-0.004	0.043	0.318	1	0.049
HEALTH	-0.260	-0.142	0.260	0.049	1

Table 14: AVE, MSV and composite reliability (CFA).

Factor	AVE	MSV	CR
FS	0.507	0.110	0.801
FSF	0.603	0.020	0.858
SFJ	0.365	0.101	0.696
SDJ	0.366	0.002	0.742
HEALTH	0.644	0.000	0.775

Interpretation:

These results show that the CFA model corresponds to the 5-factor structure that was suggested during the EFA in the previous section. Table 11 show the global fit indices which overall indicates that it is acceptable. Both the CFI (0.906) and TLI (0.886) are above 0.9, SRMR (0.044) is below 0.05, only the RMSEA (0.062) barely exceeds the threshold of 0.06. The χ^2 is significant, but due to the large number of datapoints ($N \approx 4800$) it is expected to be so.

Table 12 displays the standardized loadings of each item towards its factor. From these results, we can state that each item is a significant indicator of said factor. Health has the highest AVE (0.644) and loading (0.955 for Felt_Down), though it is a factor based on two items. When we look at the 4-item factors we notice FSF and FS show a stronger convergent validity and loadings compared to SFJ and SDJ. The AVE of every factor also exceeds the MSV with any other factor. So each factor shares more variance with its items than with other factors.

Overall, we can state that the CFA provides evidence that the suggested model fits the data.

2.3 2c. Model modification using correlated errors

```r
mod_cfa5 <- modificationIndices(fit_cfa5)
mod_errors <- subset(mod_cfa5, op == "~~" &
                     lhs %in% items &
                     rhs %in% items &
                     lhs != rhs)

head(mod_errors[order(mod_errors$mi, decreasing = TRUE), ], 20)
```

Output:

Table 15: Largest modification indices for residual covariances.

lhs	rhs	MI	EPC	sepc.all
FSF_afford_extras	FSF_save_money	578.04	0.319	1.059
FSF_pay_bills	FSF_afford_housing	520.57	0.107	0.402
SDJ_help_people	SDJ_meet_people	356.69	0.165	0.383
FS_afford_extras	FS_save_money	248.21	0.274	0.738
FS_afford_housing	FSF_afford_housing	199.37	0.112	0.228
FS_pay_bills	FSF_pay_bills	191.76	0.063	0.221
FS_pay_bills	FS_afford_housing	140.61	0.105	0.201
FS_save_money	FSF_save_money	135.22	0.098	0.221
FSF_afford_extras	FSF_afford_housing	133.18	-0.089	-0.306
FSF_afford_housing	FSF_save_money	110.51	-0.088	-0.251
FSF_pay_bills	FSF_afford_extras	108.23	-0.062	-0.273
FS_afford_extras	FSF_afford_extras	103.18	0.062	0.245
SDJ_learn_new_things	SDJ_feeling_self_worth	98.06	0.052	0.195
SDJ_learn_new_things	SDJ_develop_creativity	95.49	0.067	0.207
FSF_pay_bills	FSF_save_money	94.14	-0.063	-0.229
SDJ_learn_new_things	SDJ_meet_people	87.56	-0.067	-0.201
SDJ_help_people	SDJ_learn_new_things	86.58	-0.065	-0.197
FS_afford_extras	FS_afford_housing	81.25	-0.120	-0.284
FS_pay_bills	FSF_save_money	74.64	-0.057	-0.153
FS_pay_bills	FSF_afford_extras	64.74	-0.046	-0.150

Table 15 shows the 20 largest residual covariances, it immediately becomes apparent that these values are rather high, the highest being linked to the residual covariance between `FSF_afford_extras` and `FSF_save_money` with an MI of 578.04. To increase the value of the CFI and TLI, we will gradually incorporate this in the model until both are above the threshold of 0.9.

```r
cfa5_mod <- '
  FS =~ NA*FS_pay_bills + FS_afford_extras + FS_afford_housing + FS_save_money
  FSF =~ NA*FSF_pay_bills + FSF_afford_extras + FSF_afford_housing + FSF_save_money
```

```
  SFJ =~ NA*SFJ_no_info + SFJ_no_chance_show + SFJ_no_training + SFJ_no_support_findjob
  SDJ =~ NA*SDJ_help_people + SDJ_learn_new_things + SDJ_develop_creativity +
            SDJ_meet_people + SDJ_feeling_self_worth
  HEALTH =~ NA*HEALTH_felt_down + HEALTH_limitation

  FS ~~ 1*FS
  FSF ~~ 1*FSF
  SFJ ~~ 1*SFJ
  SDJ ~~ 1*SDJ
  HEALTH ~~ 1*HEALTH

  FS ~~ FSF + SFJ + SDJ + HEALTH
  FSF ~~ SFJ + SDJ + HEALTH
  SFJ ~~ SDJ + HEALTH
  SDJ ~~ HEALTH

  FSF_afford_extras ~~ FSF_save_money
'

fit_cfa5_mod <- cfa(cfa5_mod, data = cfs)

summary(fit_cfa5_mod, fit.measures = TRUE, standardized = TRUE)
fitmeasures(fit_cfa5_mod,
           c("chisq","df","pvalue","cfi","tli","rmsea","srmr"))
```

Table 16: Fit indices for original vs modified CFA.

Model	χ^2	df	CFI	TLI	RMSEA
Original CFA	2785.51	142	0.906	0.886	0.062
Modified CFA	2215.31	141	0.926	0.910	0.055

Interpretation:

As stated before, the MI output in Table 15 reveals that there are a lot of candidate residual covariances. The largest being the residual convariance between `FSF_afford_extras` and `FSF_save_money` (MI $\approx$ 578, standardized expected correlation $\approx$ 1.06). Meaning that these two items share variance above the FSF factor. Both factors describe financial stability in terms of a financial buffer; it is therefore plausible.

When we incorporated this error we immediately noticed an improvement such that both the TLI (from 0.886 to 0.910) and the CFI (from 0.906 to 0.926) are above the required threshold. The RMSEA (from 0.062 to 0.055) and SRMR (from 0.044 to 0.041) also showed improvements. We can now conclude that all parameters are in the acceptable ranges as opposed to the model suggested in 2b, while the loadings and latent correlations remain the same.

2.4 2d. Multi-group SEM: predictive model for financial security

```
sem_base <- '
  FS =~ FS_pay_bills + FS_afford_extras + FS_afford_housing + FS_save_money
  FSF =~ FSF_pay_bills + FSF_afford_extras + FSF_afford_housing + FSF_save_money
  SFJ =~ SFJ_no_info + SFJ_no_chance_show + SFJ_no_training + SFJ_no_support_findjob
  SDJ =~ SDJ_help_people + SDJ_learn_new_things + SDJ_develop_creativity +
            SDJ_meet_people + SDJ_feeling_self_worth
  HEALTH =~ HEALTH_felt_down + HEALTH_limitation

  FSF ~~ SFJ + SDJ + HEALTH
  SFJ ~~ SDJ + HEALTH
```

```
  SDJ ~~ HEALTH

  FSF_afford_extras ~~ FSF_save_money

  FS ~ FSF + SFJ + SDJ + HEALTH
'

sem_eqReg <- '
  FS =~ FS_pay_bills + FS_afford_extras + FS_afford_housing + FS_save_money
  FSF =~ FSF_pay_bills + FSF_afford_extras + FSF_afford_housing + FSF_save_money
  SFJ =~ SFJ_no_info + SFJ_no_chance_show + SFJ_no_training + SFJ_no_support_findjob
  SDJ =~ SDJ_help_people + SDJ_learn_new_things + SDJ_develop_creativity +
            SDJ_meet_people + SDJ_feeling_self_worth
  HEALTH =~ HEALTH_felt_down + HEALTH_limitation

  FSF ~~ SFJ + SDJ + HEALTH
  SFJ ~~ SDJ + HEALTH
  SDJ ~~ HEALTH

  FSF_afford_extras ~~ FSF_save_money

  FS ~ a1*FSF + a2*SFJ + a3*SDJ + a4*HEALTH
'

fit_conf_free <- sem(sem_base, data = cfs, group = "country")
fit_conf_eqReg <- sem(sem_eqReg, data = cfs, group = "country")
fit_metric_free <- sem(sem_base, data = cfs, group = "country",
                      group.equal = "loadings")
fit_metric_eqReg <- sem(sem_eqReg, data = cfs, group = "country",
                      group.equal = "loadings")

fm_conf_free <- fitmeasures(fit_conf_free,
  c("chisq","df","pvalue","cfi","tli","rmsea","srmr","aic","bic"))
fm_conf_eqReg <- fitmeasures(fit_conf_eqReg,
  c("chisq","df","pvalue","cfi","tli","rmsea","srmr","aic","bic"))
fm_metric_free <- fitmeasures(fit_metric_free,
  c("chisq","df","pvalue","cfi","tli","rmsea","srmr","aic","bic"))
fm_metric_eqReg <- fitmeasures(fit_metric_eqReg,
  c("chisq","df","pvalue","cfi","tli","rmsea","srmr","aic","bic"))

round(rbind(
  config_free_reg = fm_conf_free,
  config_equal_reg = fm_conf_eqReg,
  metric_free_reg = fm_metric_free,
  metric_equal_reg = fm_metric_eqReg
), 3)

anova(fit_conf_free, fit_conf_eqReg)
anova(fit_conf_free, fit_metric_free)
anova(fit_metric_free, fit_metric_eqReg)

std_final <- standardizedSolution(fit_conf_eqReg)
std_final[std_final$op == "~" & std_final$lhs == "FS", ]
```

Output:

Table 17: Multi-group SEM fit indices.

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC
Configural, free regressions	2402.91	282	0.924	0.908	0.056	0.041	201950.2	202831.9
Configural, equal regressions	2408.31	286	0.924	0.910	0.055	0.041	201947.7	202803.3
Metric, free regressions	2484.25	296	0.922	0.910	0.055	0.043	202003.6	202794.4
Metric, equal regressions	2487.09	300	0.922	0.911	0.055	0.043	201998.4	202763.4

Table 18: Chi-square difference tests between SEM models.

Comparison	$\Delta\chi^2$	Δdf	p
Config free vs config equal regressions	5.40	4	0.249
Config free vs metric free	81.34	14	< 0.001
Metric free vs metric equal regressions	2.84	4	0.586

Table 19: Standardized effects on FS in the best model (configural, equal regressions).

Predictor $\rightarrow$ FS	Group 1	Group 2
FSF	0.274	0.263
SFJ	-0.201	-0.153
SDJ	0.072	0.055
HEALTH	-0.163	-0.144

(All coefficients highly significant: $|z| > 3.4$, $p < 0.001$.)

Interpretation:

From Table 17 we can interpret that all four SEM specifications achieve good fit (CFI $\approx$ 0.92–0.93, RMSEA $\approx$ 0.055–0.056, SRMR $\approx$ 0.04). We however want to analyse whether we can use equal regression coefficients across the groups and impose metric invariance without resulting in a significant loss of fit.

The chi-square difference tests in Table 18 show that while using equal regressions in the configural model does not significantly worsen fit ($\Delta\chi^2(4) = 5.40$, $p = 0.25$). The same cannot be said when imposing metric invariance where it does significantly worsen fit ($\Delta\chi^2(14) = 81.34$, $p < 0.001$), therefore a full equality is too restrictive.

When we analyse the information criteria (AIC/BIC) and CFI and TLI we conclude that (`config_equal_reg`) performs best (AIC = 201947.7, BIC = 202803.3, CFI = 0.924, TLI = 0.910).