Multivariate Statistics

Assignment 1

Tom Leppens r0716859

Wenhan Cu r0973145

Zeynep Deniz Guvenol r0768138

Thien Nyugen

**CHANGE THE VARIABLES SO THAT THEY MEAN THE SAME**

# Question A:

We load the data, rename the variables to following the factors, compute the covariance matrix, and we fit a CFA model with three correlated factors (one for each attitude concept), and assuming each item only has a loading on the concept it aims to measure. We print fit measures, the standardized solution and we compute, for each latent variable, the composite reliability, the average variance extracted and the maximum shared variance with other latent variables.

load("ess.Rdata")

ess <- ess[-1]

names(ess)[1:13]<-c("sotru1","sotru2","sotru3","truin1","truin2","truin3","truin4","webe1","webe2","webe3","webe4","webe5","webe6")

covmat<-cov(ess)

##specify model with 3 correlated factors

cfa1<-'sotru =~NA\*+sotru1+sotru2+sotru3

truin =~NA\*truin1+truin2+truin3+truin4

webe =~NA\*webe1+webe2+webe3+webe4+webe5+webe6

sotru ~~1\*sotru

truin ~~1\*truin

webe ~~1\*webe'

#fit model on covariance matrix

fitcfa1<-cfa(cfa1,sample.cov=covmat,sample.nobs=4046)

> #standardized solution

> d<-standardizedSolution(fitcfa1)

> d

lhs op rhs est.std se z pvalue ci.lower ci.upper

1 sotru =~ sotru1 0.684 0.013 52.036 0 0.658 0.709

2 sotru =~ sotru2 0.648 0.013 48.322 0 0.622 0.674

3 sotru =~ sotru3 0.626 0.014 46.031 0 0.600 0.653

4 truin =~ truin1 0.789 0.008 93.956 0 0.773 0.805

5 truin =~ truin2 0.718 0.010 74.774 0 0.699 0.737

6 truin =~ truin3 0.581 0.012 48.194 0 0.557 0.605

7 truin =~ truin4 0.802 0.008 97.758 0 0.786 0.818

8 webe =~ webe1 0.661 0.011 60.710 0 0.640 0.683

9 webe =~ webe2 0.670 0.011 62.343 0 0.649 0.691

10 webe =~ webe3 0.589 0.012 48.379 0 0.565 0.612

11 webe =~ webe4 -0.718 0.010 72.725 0 0.699 0.738

12 webe =~ webe5 -0.677 0.011 63.729 0 0.656 0.698

13 webe =~ webe6 -0.595 0.012 49.291 0 0.571 0.618

14 sotru ~~ sotru 1.000 0.000 NA NA 1.000 1.000

15 truin ~~ truin 1.000 0.000 NA NA 1.000 1.000

16 webe ~~ webe 1.000 0.000 NA NA 1.000 1.000

17 sotru1 ~~ sotru1 0.533 0.018 29.674 0 0.498 0.568

18 sotru2 ~~ sotru2 0.580 0.017 33.355 0 0.546 0.614

19 sotru3 ~~ sotru3 0.608 0.017 35.629 0 0.574 0.641

20 truin1 ~~ truin1 0.377 0.013 28.488 0 0.352 0.403

21 truin2 ~~ truin2 0.485 0.014 35.192 0 0.458 0.512

22 truin3 ~~ truin3 0.662 0.014 47.280 0 0.635 0.690

23 truin4 ~~ truin4 0.357 0.013 27.162 0 0.331 0.383

24 webe1 ~~ webe1 0.562 0.014 39.020 0 0.534 0.591

25 webe2 ~~ webe2 0.551 0.014 38.292 0 0.523 0.579

26 webe3 ~~ webe3 0.654 0.014 45.641 0 0.626 0.682

27 webe4 ~~ webe4 0.484 0.014 34.112 0 0.456 0.512

28 webe5 ~~ webe5 0.542 0.014 37.693 0 0.514 0.570

29 webe6 ~~ webe6 0.646 0.014 45.059 0 0.618 0.675

30 sotru ~~ truin 0.555 0.016 34.183 0 0.524 0.587

31 sotru ~~ webe 0.287 0.020 14.604 0 0.248 0.326

32 truin ~~ webe 0.185 0.018 10.022 0 0.149 0.221

#print fit measures

> fitmeasures(fitcfa1,c("chisq","df","pvalue","cfi","tli","rmsea","srmr"))

chisq df pvalue cfi tli rmsea srmr

1526.049 62.000 0.000 0.912 0.889 0.076 0.040

> factorscore<-c("sotru","truin","webe")

> #composite reliability

> reliability<-round(c(compositerel(d[1:3,4]),compositerel(d[4:6,4]),compositerel(d[7:9,4])),3)

> #average variance extracted

> average\_var\_extracted<-round(c(mean(d[1:3,4]^2),mean(d[4:6,4]^2),mean(d[7:9,4]^2)),3)

> #maximum shared variance

> max\_shared\_var<-round(c(max(d[c(22,23),4]^2),max(d[c(22,24),4]^2),max(d[c(23,24),4]^2)),3)

> data.frame(factorscore,reliability,average\_var\_extracted,max\_shared\_var)

factorscore reliability average\_var\_extracted max\_shared\_var

1 sotru 0.690 0.427 0.439

2 truin 0.741 0.492 0.439

3 webe 0.756 0.510 0.316

The **fit measures** indicate that the model is rejected by an absolute goodness of fit test, i.e. the fit of the model is significantly lower than for a perfectly fitting model (chisquare= 1526.049, df=62, p<.001). Furthermore, descriptive fit measures also indicate that the model cannot reproduce the observed covariance matrix well: CFI (.912) and TLI (.889) both are lower than 0.95 and hence do not meet the cutoff of good fit. RMSEA (.076) and SRMR(0.04) indicate a good fit as they are below 0.08. Given these results, it can be argued that further modifications to the model are needed.

As can be seen in the standardized solution, all variables have significant standardized loadings, but they are positive apart from only for the variables webe4, webe5, and webe6. Note that there are only 4 variables having loadings which exceed 0.7. Hence, the square of these loadings i.e. the individual reliabilities are larger than 0.5 only for these 4 variables. This indicates that the other variables do not have sufficient reliability and therefore **convergent validity** is not satisfied for these other variables in the model. When we checked the correlations between the factors (.555, .287, .18) we see that they are poorly correlated. Furthermore, **divergent validity** is satisfied for all latent variables . Divergent validity is also confirmed using the criterion of Fornell and Lanker as we see that for, each latent variable, the average variance extracted in the observed indicator variables is larger than the maximum variance that is shared with other latent variables except social trust factor since the scores are almost the same.

Finally, we see that composite reliability of the factor scores is not goodbut acceptable as for all latent variables **composite reliabilities** are around 0.7.

## Question B

To improve our model, we can use the ‘modificationIndices()’ function to get an idea of which error terms correlation we can add to improve our model.

> modificationindices(fitcfa1)

lhs op rhs mi epc sepc.lv sepc.all sepc.nox   
sotru =~ truin1 15.852 -0.181 -0.181 -0.077 -0.07734   
sotru =~ truin2 11.370 0.157 0.157 0.067 0.06735   
sotru =~ truin3 35.878 0.288 0.288 0.129 0.12936   
sotru =~ truin4 8.738 -0.125 -0.125 -0.057 -0.05737   
sotru =~ webe1 1.248 0.013 0.013 0.018 0.01838   
sotru =~ webe2 0.967 -0.011 -0.011 -0.016 -0.01639   
sotru =~ webe3 5.258 0.030 0.030 0.039 0.03940   
sotru =~ webe4 2.192 -0.019 -0.019 -0.023 -0.02341   
sotru =~ webe5 0.033 -0.002 -0.002 -0.003 -0.00342   
sotru =~ webe6 0.065 -0.004 -0.004 -0.004 -0.00443   
truin =~ sotru1 20.844 0.242 0.242 0.112 0.11244   
truin =~ sotru2 19.103 -0.206 -0.206 -0.104 -0.10445   
truin =~ sotru3 0.078 -0.014 -0.014 -0.007 -0.00746   
truin =~ webe1 1.275 0.012 0.012 0.016 0.01647   
truin =~ webe2 0.666 0.008 0.008 0.012 0.01248   
truin =~ webe3 1.761 0.015 0.015 0.020 0.02049   
truin =~ webe4 8.275 -0.032 -0.032 -0.040 -0.04050   
truin =~ webe5 0.185 -0.005 -0.005 -0.006 -0.00651   
truin =~ webe6 0.461 0.009 0.009 0.010 0.01052   
webe =~ sotru1 3.265 -0.069 -0.069 -0.032 -0.03253   
webe =~ sotru2 12.584 0.123 0.123 0.062 0.06254   
webe =~ sotru3 2.954 -0.063 -0.063 -0.030 -0.03055   
webe =~ truin1 0.131 -0.011 -0.011 -0.005 -0.00556   
webe =~ truin2 7.008 0.086 0.086 0.037 0.03757   
webe =~ truin3 11.255 0.115 0.115 0.051 0.05158   
webe =~ truin4 16.240 -0.117 -0.117 -0.054 -0.05459   
sotru1 ~~ sotru2 1.114 -0.088 -0.088 -0.037 -0.03760   
sotru1 ~~ sotru3 6.018 -0.207 -0.207 -0.080 -0.08061   
sotru1 ~~ truin1 0.838 0.045 0.045 0.020 0.02062   
sotru1 ~~ truin2 2.162 0.076 0.076 0.030 0.03063   
sotru1 ~~ truin3 0.404 -0.035 -0.035 -0.012 -0.01264   
sotru1 ~~ truin4 2.323 0.069 0.069 0.033 0.03365   
sotru1 ~~ webe1 0.167 0.007 0.007 0.008 0.00866   
sotru1 ~~ webe2 0.559 0.012 0.012 0.015 0.01567   
sotru1 ~~ webe3 9.170 0.056 0.056 0.058 0.05868   
sotru1 ~~ webe4 6.366 -0.044 -0.044 -0.051 -0.05169   
sotru1 ~~ webe5 7.718 -0.051 -0.051 -0.055 -0.05570   
sotru1 ~~ webe6 0.008 0.002 0.002 0.002 0.00271   
sotru2 ~~ sotru3 11.982 0.252 0.252 0.102 0.10272   
sotru2 ~~ truin1 3.196 -0.081 -0.081 -0.037 -0.03773   
sotru2 ~~ truin2 1.005 0.048 0.048 0.020 0.02074   
sotru2 ~~ truin3 3.177 0.090 0.090 0.033 0.03375   
sotru2 ~~ truin4 12.337 -0.147 -0.147 -0.075 -0.07576   
sotru2 ~~ webe1 0.000 0.000 0.000 0.000 0.00077   
sotru2 ~~ webe2 3.723 -0.028 -0.028 -0.037 -0.03778   
sotru2 ~~ webe3 13.839 -0.064 -0.064 -0.069 -0.06979   
sotru2 ~~ webe4 8.465 0.048 0.048 0.057 0.05780   
sotru2 ~~ webe5 29.089 0.093 0.093 0.103 0.10381   
sotru2 ~~ webe6 0.001 -0.001 -0.001 -0.001 -0.00182   
sotru3 ~~ truin1 15.508 -0.190 -0.190 -0.081 -0.08183   
sotru3 ~~ truin2 0.487 0.036 0.036 0.013 0.01384   
sotru3 ~~ truin3 30.187 0.297 0.297 0.100 0.10085   
sotru3 ~~ truin4 0.077 -0.012 -0.012 -0.006 -0.00686   
sotru3 ~~ webe1 0.075 0.004 0.004 0.005 0.00587   
sotru3 ~~ webe2 1.452 -0.019 -0.019 -0.023 -0.02388   
sotru3 ~~ webe3 8.985 0.055 0.055 0.055 0.05589   
sotru3 ~~ webe4 0.012 0.002 0.002 0.002 0.00290   
sotru3 ~~ webe5 5.836 -0.044 -0.044 -0.046 -0.04691   
sotru3 ~~ webe6 1.311 -0.024 -0.024 -0.021 -0.02192   
truin1 ~~ truin2 52.242 -0.492 -0.492 -0.210 -0.21093   
truin1 ~~ truin3 211.717 -0.854 -0.854 -0.326 -0.32694   
truin1 ~~ truin4 559.300 1.707 1.707 0.914 0.91495   
truin1 ~~ webe1 0.319 0.008 0.008 0.011 0.01196   
truin1 ~~ webe2 2.238 0.021 0.021 0.030 0.03097   
truin1 ~~ webe3 2.379 -0.026 -0.026 -0.030 -0.03098   
truin1 ~~ webe4 2.427 -0.025 -0.025 -0.032 -0.03299   
truin1 ~~ webe5 1.698 0.022 0.022 0.026 0.026100   
truin1 ~~ webe6 0.003 0.001 0.001 0.001 0.001101   
truin2 ~~ truin3 478.146 1.275 1.275 0.430 0.430102   
truin2 ~~ truin4 168.787 -0.834 -0.834 -0.395 -0.395103   
truin2 ~~ webe1 0.000 0.000 0.000 0.000 0.000104   
truin2 ~~ webe2 0.671 0.013 0.013 0.015 0.015105   
truin2 ~~ webe3 1.245 0.020 0.020 0.020 0.020106   
truin2 ~~ webe4 0.004 0.001 0.001 0.001 0.001107   
truin2 ~~ webe5 0.006 0.001 0.001 0.001 0.001108   
truin2 ~~ webe6 0.196 0.009 0.009 0.008 0.008109   
truin3 ~~ truin4 73.909 -0.471 -0.471 -0.199 -0.199110   
truin3 ~~ webe1 2.761 -0.028 -0.028 -0.029 -0.029111   
truin3 ~~ webe2 0.001 0.000 0.000 0.000 0.000112   
truin3 ~~ webe3 1.206 0.021 0.021 0.019 0.019113   
truin3 ~~ webe4 3.041 0.032 0.032 0.032 0.032114   
truin3 ~~ webe5 0.204 0.009 0.009 0.008 0.008115   
truin3 ~~ webe6 0.967 0.021 0.021 0.017 0.017116   
truin4 ~~ webe1 1.231 0.015 0.015 0.022 0.022117   
truin4 ~~ webe2 0.493 -0.009 -0.009 -0.014 -0.014118   
truin4 ~~ webe3 0.022 0.002 0.002 0.003 0.003119   
truin4 ~~ webe4 3.942 -0.030 -0.030 -0.041 -0.041120   
truin4 ~~ webe5 3.947 -0.031 -0.031 -0.040 -0.040121   
truin4 ~~ webe6 0.012 -0.002 -0.002 -0.002 -0.002122   
webe1 ~~ webe2 153.517 0.068 0.068 0.255 0.255123   
webe1 ~~ webe3 50.754 0.045 0.045 0.137 0.137124   
webe1 ~~ webe4 62.404 -0.051 -0.051 -0.173 -0.173125   
webe1 ~~ webe5 68.018 -0.054 -0.054 -0.171 -0.171126   
webe1 ~~ webe6 5.876 -0.017 -0.017 -0.047 -0.047127   
webe2 ~~ webe3 128.133 0.068 0.068 0.219 0.219128   
webe2 ~~ webe4 57.439 -0.047 -0.047 -0.168 -0.168129   
webe2 ~~ webe5 86.209 -0.058 -0.058 -0.194 -0.194130   
webe2 ~~ webe6 23.523 -0.033 -0.033 -0.094 -0.094131   
webe3 ~~ webe4 114.875 -0.075 -0.075 -0.219 -0.219132   
webe3 ~~ webe5 150.356 -0.088 -0.088 -0.239 -0.239133   
webe3 ~~ webe6 33.002 0.045 0.045 0.105 0.105134   
webe4 ~~ webe5 612.318 0.182 0.182 0.553 0.553135   
webe4 ~~ webe6 0.199 0.004 0.004 0.009 0.009136   
webe5 ~~ webe6 3.493 0.015 0.015 0.037 0.037

Based on this output we included every error term correlation that lowers the chi square by at least 100. This yields the following model:

cfa2<- 'sotru =~NA\*sotru1+sotru2+sotru3  
 truin =~NA\*truin1+truin2+truin3+truin4  
 webe =~NA\*webe1+webe2+webe3+webe4+webe5+webe6  
 sotru ~~1\*sotru  
 truin ~~1\*truin  
 webe ~~1\*webe  
 truin1 ~~ truin3  
 truin1 ~~ truin4  
 truin2 ~ ~ truin3  
 truin2 ~~ truin4  
 webe1 ~~ webe2  
 webe2 ~~ webe3  
 webe3 ~~ webe4  
 webe3 ~~ webe5  
 webe4 ~~ webe5'

Looking at the fit measurements we see indeed an improvement. CFI and TLI are above 0.95 and RMSEA and SRMR are below 0.08. The chi-square test is still significantly different from the perfectly fitted model (chi-square= 169.929, df=53, p<.001). This is probably due to the large number of observations in the dataset. We can therefore conclude that these additions add enough value to be included in the final model. This is also confirmed with the LR ratio test (LR= 1356.1, df=9, p<.001).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Chisq | df | p-value | CFI | TLI | RMSEA | SRMR |
| cfa1 | 1526.049 | 62 | 0.000 | 0.912 | 0.889 | 0.076 | 0.040 |
| cfa2 | 169.929 | 53 | 0.000 | 0.993 | 0.990 | 0.023 | 0.017 |

Chi-Squared Difference Test  
  
 Df AIC BIC Chisq Chisq diff RMSEA Df diff Pr(>Chisq)   
fitcfa2 53 164265 164504 169.93   
fitcfa1 62 165603 165786 1526.05 1356.1 0.19234 9 < 2.2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Comparing the standardized estimates of cfa2 and cfa1. Overall, we can see some differences between the two models but what stands out the most is that only three estimates are above 0.7. This is also reflected in the difference in composite reliability as this is decreased for Trust institution and wellbeing.

|  |  |  |
| --- | --- | --- |
|  | **CFA2 (est.std)** | **CFA1 (est.std)** |
| sotru =~ sotru1 | 0.681 | 0.648 |
| sotru =~ sotru2 | 0.648 | 0.626 |
| sotru =~ sotru3 | 0.629 | 0.626 |
| truin =~ truin1 | 0.713 | 0.789 |
| truin =~ truin2 | 0.750 | 0.718 |
| truin =~ truin3 | 0.628 | 0.581 |
| truin =~ truin4 | 0.689 | 0.802 |
| webe =~ webe1 | 0.623 | 0.661 |
| webe =~ webe2 | 0.613 | 0.670 |
| webe =~ webe3 | 0.691 | 0.589 |
| webe =~ webe4 | -0.703 | -0.718 |
| webe =~ webe5 | -0.657 | -0.677 |
| webe =~ webe6 | -0.602 | -0.595 |
|  |  |  |
| truin1 ~~ truin3 | -0.121 |  |
| truin1 ~~ truin4 | 0.386 |  |
| truin2 ~~ truin3 | 0.200 |  |
| truin2 ~~ truin4 | 0.030 |  |
|  |  |  |
| webe1 ~~ webe2 | 0.232 |  |
| webe2 ~~ webe3 | 0.096 |  |
| webe3 ~~ webe4 | 0.266 |  |
| webe3 ~~ webe5 | 0.270 |  |
| webe4 ~~ webe5 | 0.307 |  |

|  |  |  |
| --- | --- | --- |
| **Factorscore** | **Reliability cfa1** | **Reliability cfa2** |
| sotru | 0.690 | 0.690 |
| truin | 0.741 | 0.740 |
| webe | 0.756 | 0.678 |

Looking at the correlations between the error terms they are all significant (p<0.001). Overall, the correlations between the different variables are positively correlated except for truin1 (Trust in the country's parliament) and truin3 (Trust in the police) are negatively correlated. This value ranges from 0.030 to 0.386. For well-being, all terms have a positive correlation ranging from 0.096 to 0.307.

# Question C

For this question we use the same model from the previous question (cfa2) and transform it to a structural equation model. Sem2 is used to fit a model where the coefficients are constraint to be equal among countries. Sem1 was used to let the coefficient of the regressions range freely.

sem1<-'

# measurements model

sotru =~NA\*sotru1+sotru2+sotru3  
truin =~NA\*truin1+truin2+truin3+truin4  
webe =~NA\*webe1+webe2+webe3+webe4+webe5+webe6  
sotru ~~1\*sotru  
truin ~~1\*truin  
webe ~~1\*webe  
truin1 ~~ truin3  
truin1 ~~ truin4  
truin2 ~~ truin3  
truin2 ~~ truin4  
webe1 ~~ webe2  
webe2 ~~ webe3  
webe3 ~~ webe4  
webe3 ~~ webe5  
webe4 ~~ webe5

#structural model  
webe ~ sotru + truin'

sem2<-'

sotru =~NA\*sotru1+sotru2+sotru3  
truin =~NA\*truin1+truin2+truin3+truin4  
webe =~NA\*webe1+webe2+webe3+webe4+webe5+webe6  
sotru ~~1\*sotru  
truin ~~1\*truin  
webe ~~1\*webe  
truin1 ~~ a\*truin3  
truin1 ~~ b\*truin4  
truin2 ~~ c\*truin3  
truin2 ~~ d\*truin4  
webe1 ~~ e\*webe2  
webe2 ~~ f\*webe3  
webe3 ~~ g\*webe4  
webe3 ~~ h\*webe5  
webe4 ~~ i\*webe5  
webe ~ j\*sotru + k\*truin'

Looking at the fitmeasures we can conclude that the best model to fit the data is config1. This model has the lowest chi-square, this is still different from the perfectly fitted model, both CFI and TLI are the highest and RMSEA and SRMR are the lowest. Looking at the AIC and BIC again config1 yields the lowest value, indicating this the best fitting model.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Chisq | df | p-value | CFI | TLI | RMSEA | SRMR | AIC | BIC |
| config1 | 218.858 | 106 | 0.000 | 0.993 | 0.990 | 0.023 | 0.020 | 163211 | 163854 |
| config2 | 259.497 | 117 | 0.000 | 0.991 | 0.989 | 0.025 | 0.022 | 163229 | 163803 |
| metric1 | 261.703 | 119 | 0.000 | 0.991 | 0.989 | 0.024 | 0.027 | 163227 | 163789 |
| metric2 | 311.430 | 130 | 0.000 | 0.989 | 0.987 | 0.026 | 0.031 | 163255 | 163747 |

Using The LR test we can conclude that all models are significantly different (p<0.001) except for config2 and metric1 (LR: 2.2057, df=2, p=0.03319). We can thus confidently say that config1 is the best model. Looking at the standardized solution of config1:

lhs op rhs group est.std se z pvalue ci.lower ci.upper  
1 sotru =~ sotru1 1 0.611 0.021 29.231 0.000 0.570 0.652  
2 sotru =~ sotru2 1 0.686 0.020 33.767 0.000 0.646 0.726  
3 sotru =~ sotru3 1 0.570 0.021 26.608 0.000 0.528 0.612  
4 truin =~ truin1 1 0.688 0.024 28.458 0.000 0.640 0.735  
5 truin =~ truin2 1 0.737 0.024 30.183 0.000 0.689 0.785  
6 truin =~ truin3 1 0.603 0.028 21.815 0.000 0.549 0.657  
7 truin =~ truin4 1 0.666 0.028 24.182 0.000 0.612 0.720  
8 webe =~ webe1 1 0.639 0.018 34.716 0.000 0.603 0.675  
9 webe =~ webe2 1 0.656 0.020 33.221 0.000 0.617 0.694  
10 webe =~ webe3 1 0.692 0.021 32.594 0.000 0.650 0.734  
11 webe =~ webe4 1 -0.656 0.020 -32.239 0.000 -0.696 -0.616  
12 webe =~ webe5 1 -0.616 0.021 -29.024 0.000 -0.658 -0.575  
13 webe =~ webe6 1 -0.608 0.018 -33.498 0.000 -0.644 -0.573

Overall, all the correlations with the latent variables are significant however all but one (truin =~ truin2) are below 0.7. Also note that webe4, webe5, webe6 are negatively correlated with well-being.

Looking at the coefficients of social trust and trust institution we see that in the first group, social trust has a significant effect on well-being while this is not the case for trust institution. In group 2, however, both predictors are significant in explaining well-being but here social trust has a bigger impact than trust institution.

lhs op rhs group est.std se z pvalue ci.lower ci.upper

webe ~ sotru 1 0.263 0.043 6.191 0.000 0.180 0.347  
webe ~ truin 1 0.040 0.042 0.955 0.340 -0.042 0.121

webe ~ sotru 2 0.218 0.036 6.120 0.000 0.148 0.288  
webe ~ truin 2 0.070 0.035 1.996 0.046 0.001 0.139

# Question D

|  |
| --- |
| #standardize variables  sess<-ess  sess[,2:14]<-scale(ess[,2:14],center=TRUE,scale=TRUE)  head(sess)  #conduct canonical correlation analysis  cancor.sess <- cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)  ~ppltrst+pplfair+pplhlp+trstprl+trstlgl+trstplc+trstplt,  data=sess)  #print summary result  summary(cancor.sess)  #print canonical loadings  cancor.sess$structure$X.xscores  cancor.sess$structure$Y.yscores    #print redundancies (this tells us how much variance in Y is explained by X)  redu <- redundancy(cancor.sess)  round(redu$Xcan.redun, 3)  round(redu$Ycan.redun, 3) |

We load the data with the original variable names, standardize the variables, use the candisc() procedure to conduct canonical correlation analysis and print a summary of the results and compute redundancies.

The canonical correlation analysis extracts 6 pairs of canonical variances. Hypotheses tests indicate that only the first two correlations are significant i.e., H0: corr(u3,t3)=0 cannot be rejected at the 5% level (p= 0.2378).

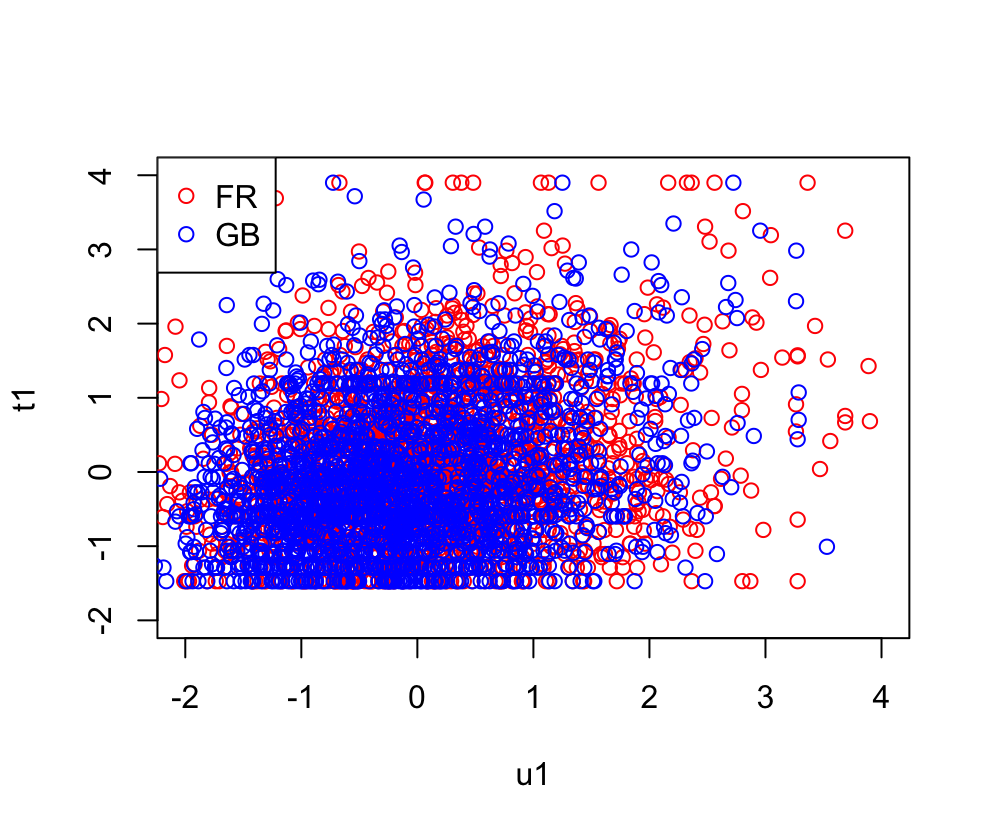
The first canonical correlation equals 0.24. This means that the canonical variate u1 accounts for 5.89% of the variance in the canonical variate t1. The second canonical correlation equals 0.11. This means that the canonical variate u2 accounts for 1.21% of the variance in the canonical variate t2.

U1 accounts for 3% variance in Y and u2 accounts for 0.2% variance in Y. Since only the first two correlations are significant, we can say that X variables account for 3.2% of variance in the Y variables. The u2 barely contributed. Only a small portion of variance in Y is explained by X.

To interpret the first pairs of canonical variates, we print the canonical loadings (corr(u,X), corr(t, Y)). In addition, we make a scatter plot of the first pair of canonical variates and indicate a different color for observations of each country.

|  |
| --- |
| > summary(cancor.sess)  CanR CanRSQ Eigen percent cum scree  1 0.242629 5.887e-02 6.255e-02 77.37503 77.38 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  2 0.110279 1.216e-02 1.231e-02 15.22875 92.60 \*\*\*\*\*\*  3 0.063206 3.995e-03 4.011e-03 4.96159 97.57 \*\*  4 0.041142 1.693e-03 1.696e-03 2.09741 99.66 \*  5 0.016167 2.614e-04 2.614e-04 0.32339 99.99  6 0.003343 1.118e-05 1.118e-05 0.01383 100.00  Test of H0: The canonical correlations in the  current row and all that follow are zero    CanR LR test stat approx F numDF denDF Pr(> F)  1 0.242629 0.92415 7.6396 42 18920 < 2.2e-16 \*\*\*  2 0.110279 0.98196 2.4539 30 16138 1.618e-05 \*\*\*  3 0.063206 0.99405 1.2056 20 13384 0.2378  4 0.041142 0.99804 0.6617 12 10678 0.7897  5 0.016167 0.99973 0.1834 6 8074 0.9815  6 0.003343 0.99999 NaN 2 NaN NaN |

|  |
| --- |
| > #print redundancies (this tells us how much variance in Y is explained by X)  > redu <- redundancy(cancor.ess)  > round(redu$Ycan.redun, 3)  Ycan1 Ycan2 Ycan3 Ycan4 Ycan5 Ycan6  0.030 0.002 0.000 0.000 0.000 0.000 |
| > #print canonical loadings  > cancor.sess$structure$X.xscores  Xcan1 Xcan2 Xcan3 Xcan4 Xcan5 Xcan6  ppltrst -0.6763363 4.724725e-01 0.13525466 0.05282290 -0.01131331 -0.5374706  pplfair -0.8416969 -3.845485e-01 -0.24453178 0.24728799 -0.07490891 -0.1232267  pplhlp -0.6092904 4.750760e-01 -0.38969676 0.24141612 0.20860056 0.2809934  trstprl -0.5441966 -8.041144e-05 0.71491848 0.03091828 0.08653009 0.2225929  trstlgl -0.5990731 1.842299e-01 0.38310315 -0.18269912 -0.08660209 0.4894876  trstplc -0.5543285 1.441351e-01 0.01696339 -0.64550485 -0.40052327 0.2577618  trstplt -0.4406747 2.170732e-01 0.49023591 0.29713029 -0.55893911 0.2729865  > cancor.sess$structure$Y.yscores  Ycan1 Ycan2 Ycan3 Ycan4 Ycan5 Ycan6  fltdpr -0.7396063 0.12576041 0.18270422 0.56931392 -0.043348111 0.27890610  fltsd -0.6871178 0.12922116 0.43606685 -0.29073751 0.322580432 0.36390976  fltanx -0.6772528 0.67065702 -0.07524753 -0.08650432 0.099240187 -0.26182813  wrhpp -0.7485314 -0.18422145 -0.51344794 -0.17741963 -0.005463235 0.33260617  enjlf -0.7927583 -0.49962698 -0.07182186 -0.04339612 0.138156486 -0.30948079  fltpcfl -0.6438757 0.01868394 0.17137384 -0.27206474 -0.693025695 -0.03744553 |



X variables have relatively high negative correlations with u1. A higher score on u1 indicates low social trust and low trust in institutions.

The canonical loadings indicate that Y variables have a relatively high negative correlation with t1. A high score on t1 means that people in a country with low social trust and trust in institutions experience lower welling-being. The Great Britain has better trust in social and institution than France.

# Question E

# Use the split-half approach to assess the validity of the solution. Assign even- numbered observations to the calibration set and assign odd-numbered observations to the validation set when conducting this analysis. Discuss what you can conclude about the validity of the solution.

|  |
| --- |
| #split data and standardize data  train<-sess[seq(2,4046,by=2),]  valid<-sess[seq(1,4046,by=2),]  train[,2:14]<-scale(train[,2:14],center=TRUE,scale=TRUE)  valid[,2:14]<-scale(valid[,2:14],center=TRUE,scale=TRUE)    #conduct canonical correlation analysis on train data  cancor.train <- cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)  ~ppltrst+pplfair+pplhlp+trstprl+trstlgl+trstplc+trstplt,  data=train)  #print summary result  summary(cancor.train)  #print canonical loadings of train  cancor.train$structure$X.xscores  cancor.train$structure$Y.yscores    #conduct canonical correlation analysis on calibration data  cancor.valid <- cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)  ~ppltrst+pplfair+pplhlp+trstprl+trstlgl+trstplc+trstplt,  data=valid)    #print summary result  summary(cancor.valid)    #print canonical loadings of train  cancor.valid$structure$X.xscores  cancor.valid$structure$Y.yscores    #to obtain U and T from train data  trainU <- cancor.train$scores$X  trainT <- cancor.train$scores$Y  # to compute U\* and T\*  # U\* = Xb\*  # T\*= Ya\*  U\_star <-as.matrix(train[,2:8])%\*%cancor.valid$coef$X  T\_star <-as.matrix(train[,9:14])%\*%cancor.valid$coef$Y |

The absolute value of the diagonal elements of R(T,T\*) and R(U,U\*) represent the reliabilities of the canonical variates for Y and X variables. The reliabilities of t1, t2 equal .989, .817. And the reliabilities of u1, u2 equal .982, .514. In other words the first pairs of canonical variates have excellent reliability, but the reliability of u2 is unacceptable. The off-diagonal correlations are low.

A comparison of R(U\*,T\*) and R(U,T) shows that R(u1, t1) 0.253 is only a little higher than R(u1\*, t1\*) 0.246. In other words overestimation of the first canonical correlation due to the maximization involved is not an issue. Yet, the overestimation in the second canonical correlation is also rather large(.129 versus .044).

The off-diagonal elements of R(T\*,T\*) and R(U\*,U\*) are close to 0, which indicates that canonical variates of Y variables and of X variables computed on calibration data but based on the coefficients from validation data have as expected correlations that are close to 0(canonical variates are independent).

|  |
| --- |
| > # we need to compute R(T,T\*), R(U,U\*) for t1,t2,u1,u2  > round(cor(trainT,T\_star)[1:2,1:2],3)  Ycan1 Ycan2  Ycan1 0.989 -0.111  Ycan2 0.101 0.817  > round(cor(trainU,U\_star)[1:2,1:2],3)  Xcan1 Xcan2  Xcan1 0.982 -0.042  Xcan2 0.029 0.514  > # we need to compute R(U\*,T\*), R(U,T)  > round(cor(U\_star,T\_star)[1:2,1:2],3)  Ycan1 Ycan2  Xcan1 0.246 -0.028  Xcan2 -0.002 0.044  > round(cor(trainU,trainT)[1:2,1:2],3)  Ycan1 Ycan2  Xcan1 0.253 0.000  Xcan2 0.000 0.129  > # we need to compute R(U\*,U\*), R(T\*,T\*)  > round(cor(U\_star,U\_star)[1:2,1:2],3)  Xcan1 Xcan2  Xcan1 1.000 -0.007  Xcan2 -0.007 1.000  > round(cor(T\_star,T\_star)[1:2,1:2],3)  Ycan1 Ycan2  Ycan1 1.000 -0.019  Ycan2 -0.019 1.000 |

Question F

In summary, t1, t2 and u1 are both important and reliable.

This confirmatory factor analysis model has three correlated factors and assumes each item only has a loading on the concept it aims to measure. To get an idea of the fit of the model we used fit measures, the standardized solution, and computed, for each latent variable, the composite reliability (CR), the average variance extracted (AVE), and the maximum shared variance (MSV) with other latent variables.

To test the validity and reliability we used the following cutoff values:

* CFI > 0.95
* TLI > 0.95
* RMSEA < 0.08
* SRMR < 0.08
* Convergent Validity: AVE > 0.5
* Discriminant Validity: MSV < AVE
* Reliability: CR > 0.7

Fitting the model yields the following results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Chisq | df | p-value | CFI | TLI | RMSEA | SRMR |
| 689.143 | 32 | 0.000 | 0.944 | 0.921 | 0.071 | 0.034 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Factor score** | **AVE** | **MSV** | **CR** |
| **social trust** | 0.4264263 | 0.4732259 | 0.6901061 |
| **trust institutions** | 0.5294827 | 0.6433114 | 0.8160278 |
| **well being** | 0.4873276 | 0.5577240 | 0.7396447 |

Measures we can state the following: The chi-square test tells us that the model is significantly different from a perfectly fitted model (chi-square=689.143, df=32, p-value<0.000). However, as this dataset has a large sample size (n=4046) the goodness of fit test can be significantly different even if the difference between the two models is very small. Nonetheless, the other measures indicate that the model cfa1 can be improved. Both CFI (0.944) and TLI (0.921) are lower than 0.95 and the RMSEA (0.071) and SRMR (0.034) are higher than 0.08. Furthermore, the validity measurements indicate that there could be improvements: the MSV is not always lower than the AVE, the CR for social trust is not bigger than 0.7, and the AVE is not always above 0.5.

**Question B**: Use modification indices to see how you can obtain a model that meets the criteria of good fit in (in terms of TLI, CFI, RMSEA, SRMR) by including a few well-chosen correlated error terms for pairs of items. Try to justify the correlated error terms from a substantive point of view

To improve our model, we can use the ‘modificationIndices()’ function to get an idea of which error terms correlation we can add to improve our model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **lhs** | **op** | **rhs** | **mi** | **epc** | **sepc.lv** | **sepc.all** | **sepc.nox** |
| social\_trust | =~ | trstprl | 16.004 | -0.182 | -0.182 | -0.078 | -0.078 |
| social\_trust | =~ | trstlgl | 11.091 | 0.155 | 0.155 | 0.066 | 0.066 |
| social\_trust | =~ | trstplc | 34.495 | 0.283 | 0.283 | 0.126 | 0.126 |
| social\_trust | =~ | trstplt | 7.946 | -0.119 | -0.119 | -0.055 | -0.055 |
| social\_trust | =~ | fltdpr | 1.150 | 0.013 | 0.013 | 0.019 | 0.019 |
| social\_trust | =~ | fltsd | 7.992 | -0.034 | -0.034 | -0.050 | -0.050 |
| social\_trust | =~ | fltanx | 4.218 | 0.027 | 0.027 | 0.036 | 0.036 |
| trust\_institutions | =~ | ppltrst | 18.110 | 0.228 | 0.228 | 0.106 | 0.106 |
| trust\_institutions | =~ | pplfair | 15.552 | -0.185 | -0.185 | -0.094 | -0.094 |
| trust\_institutions | =~ | pplhlp | 0.168 | -0.020 | -0.020 | -0.010 | -0.010 |
| trust\_institutions | =~ | fltdpr | 0.188 | 0.005 | 0.005 | 0.007 | 0.007 |
| trust\_institutions | =~ | fltsd | 0.723 | -0.009 | -0.009 | -0.013 | -0.013 |
| trust\_institutions | =~ | fltanx | 0.251 | 0.006 | 0.006 | 0.008 | 0.008 |
| well\_being | =~ | ppltrst | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| well\_being | =~ | pplfair | 0.590 | 0.028 | 0.028 | 0.014 | 0.014 |
| well\_being | =~ | pplhlp | 0.594 | -0.029 | -0.029 | -0.014 | -0.014 |
| well\_being | =~ | trstprl | 0.159 | -0.013 | -0.013 | -0.006 | -0.006 |
| well\_being | =~ | trstlgl | 5.796 | 0.083 | 0.083 | 0.035 | 0.035 |
| well\_being | =~ | trstplc | 5.440 | 0.084 | 0.084 | 0.037 | 0.037 |
| well\_being | =~ | trstplt | 9.847 | -0.095 | -0.095 | -0.044 | -0.044 |
| ppltrst | ~~ | pplfair | 0.488 | -0.059 | -0.059 | -0.025 | -0.025 |
| ppltrst | ~~ | pplhlp | 11.715 | -0.296 | -0.296 | -0.116 | -0.116 |
| ppltrst | ~~ | trstprl | 0.702 | 0.041 | 0.041 | 0.018 | 0.018 |
| ppltrst | ~~ | trstlgl | 2.060 | 0.074 | 0.074 | 0.029 | 0.029 |
| ppltrst | ~~ | trstplc | 0.378 | -0.033 | -0.033 | -0.012 | -0.012 |
| ppltrst | ~~ | trstplt | 1.744 | 0.059 | 0.059 | 0.029 | 0.029 |
| ppltrst | ~~ | fltdpr | 0.912 | -0.016 | -0.016 | -0.020 | -0.020 |
| ppltrst | ~~ | fltsd | 1.018 | -0.016 | -0.016 | -0.023 | -0.023 |
| ppltrst | ~~ | fltanx | 4.422 | 0.039 | 0.039 | 0.042 | 0.042 |
| pplfair | ~~ | pplhlp | 16.453 | 0.297 | 0.297 | 0.119 | 0.119 |
| pplfair | ~~ | trstprl | 2.782 | -0.076 | -0.076 | -0.035 | -0.035 |
| pplfair | ~~ | trstlgl | 1.289 | 0.055 | 0.055 | 0.022 | 0.022 |
| pplfair | ~~ | trstplc | 3.774 | 0.099 | 0.099 | 0.036 | 0.036 |
| pplfair | ~~ | trstplt | 12.011 | -0.145 | -0.145 | -0.073 | -0.073 |
| pplfair | ~~ | fltdpr | 6.819 | 0.041 | 0.041 | 0.053 | 0.053 |
| pplfair | ~~ | fltsd | 0.043 | 0.003 | 0.003 | 0.004 | 0.004 |
| pplfair | ~~ | fltanx | 4.821 | -0.038 | -0.038 | -0.043 | -0.043 |
| pplhlp | ~~ | trstprl | 15.802 | -0.192 | -0.192 | -0.082 | -0.082 |
| pplhlp | ~~ | trstlgl | 0.500 | 0.036 | 0.036 | 0.014 | 0.014 |
| pplhlp | ~~ | trstplc | 30.995 | 0.301 | 0.301 | 0.101 | 0.101 |
| pplhlp | ~~ | trstplt | 0.165 | -0.018 | -0.018 | -0.009 | -0.009 |
| pplhlp | ~~ | fltdpr | 0.132 | -0.006 | -0.006 | -0.007 | -0.007 |
| pplhlp | ~~ | fltsd | 6.440 | -0.040 | -0.040 | -0.054 | -0.054 |
| pplhlp | ~~ | fltanx | 7.170 | 0.049 | 0.049 | 0.052 | 0.052 |
| trstprl | ~~ | trstlgl | 51.686 | -0.489 | -0.489 | -0.209 | -0.209 |
| trstprl | ~~ | trstplc | 210.055 | -0.850 | -0.850 | -0.324 | -0.324 |
| Trstprl | ~~ | trstplt | 554.999 | 1.700 | 1.700 | 0.911 | 0.911 |
| Trstprl | ~~ | fltdpr | 0.156 | 0.006 | 0.006 | 0.008 | 0.008 |
| Trstprl | ~~ | fltsd | 2.284 | 0.022 | 0.022 | 0.034 | 0.034 |
| Trstprl | ~~ | fltanx | 3.768 | -0.033 | -0.033 | -0.039 | -0.039 |
| trstlgl | ~~ | trstplc | 479.201 | 1.277 | 1.277 | 0.430 | 0.430 |
| trstlgl | ~~ | trstplt | 169.431 | -0.835 | -0.835 | -0.395 | -0.395 |
| trstlgl | ~~ | fltdpr | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| trstlgl | ~~ | fltsd | 0.645 | 0.012 | 0.012 | 0.017 | 0.017 |
| trstlgl | ~~ | fltanx | 1.307 | 0.021 | 0.021 | 0.022 | 0.022 |
| Trstplc | ~~ | trstplt | 73.607 | -0.470 | -0.470 | -0.198 | -0.198 |
| Trstplc | ~~ | fltdpr | 1.134 | -0.018 | -0.018 | -0.020 | -0.020 |
| Trstplc | ~~ | fltsd | 0.586 | 0.013 | 0.013 | 0.015 | 0.015 |
| Trstplc | ~~ | fltanx | 3.155 | 0.034 | 0.034 | 0.032 | 0.032 |
| Trstplt | ~~ | fltdpr | 0.030 | 0.002 | 0.002 | 0.004 | 0.004 |
| Trstplt | ~~ | fltsd | 4.301 | -0.028 | -0.028 | -0.047 | -0.047 |
| Trstplt | ~~ | fltanx | 0.458 | -0.011 | -0.011 | -0.014 | -0.014 |
| fltdpr | ~~ | fltsd | 3.441 | 0.041 | 0.041 | 0.182 | 0.182 |
| Fltdpr | ~~ | fltanx | 6.709 | -0.049 | -0.049 | -0.167 | -0.167 |
| fltsd | ~~ | fltanx | 1.012 | 0.020 | 0.020 | 0.077 | 0.077 |

Based on the output we included the following every proposal that lowers the chi square by at least 50 which yields the following model:

cfa2 <-' social\_trust =~ NA\*ppltrst + pplfair + pplhlp

trust\_institutions =~ NA\*trstprl + trstlgl + trstplc + trstplt

well\_being =~ NA\*fltdpr + fltsd + fltanx

social\_trust ~~ 1\*social\_trust

trust\_institutions ~~ 1\*trust\_institutions

well\_being ~~ 1\*well\_being

trstprl ~~ trstplc

trstlgl ~~ trstplc

trstlgl ~~ trstplt

trstplc ~~ trstplt

trstprl ~~ trstplt

trstprl ~~ trstlgl

'

Looking at the fit measures we see that the goodness of fit test is still significantly different (chi-square=85.457, df=27, p-value<0.000) but CFI and TLI are both above 0.95 and RMSEA and SRMR have decreased. However, they are still not below the cutoff of 0.08. Overall we can conclude that the model is improved and fits the data better.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Chisq | df | p-value | CFI | TLI | RMSEA | SRMR |
| 85.457 | 26 | 0.000 | 0.995 | 0.991 | 0.024 | 0.014 |

**c.** Fit a multi-group structural equation model (with country as the group variable) on the matrix of centered variables to investigate how the latent variables “social trust” and “trust in public institutions” affect the latent variable “well-being”. Estimate four versions of the multi-group structural equation model:

1) a configural measurement invariance model with country-specific regression coefficients in the regression equation of the structural model

2) a configural measurement invariance model with regression coefficients that are constrained to be equal across countries

3) a metric measurement invariance model with country-specific regression coefficients in the regression equation of the structural model

4) a metric measurement invariance model with regression coefficients that are constrained to be equal across countries

Compare the fit measures of the four estimated models and/or use model comparison tests to select the best model. Next discuss the results of this final model (e.g., model fit, estimated intercepts, (standardized) regression coefficients, etc.).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Chisq** | **Df** | **CFI** | **TLI** | **RMSEA** | **SRMR** | **AIC** |
| **config\_1** | 709.055 | 64 | 0.945 | 0.923 | 0.071 | 0.034 | 139051 |
| **config\_2** | 709.377 | 66 | 0.945 | 0.925 | 0.069 | 0.034 | 139047 |
| **metric\_1** | 775.804 | 74 | 0.94 | 0.927 | 0.068 | 0.043 | 139098 |
| **metric\_2** | 778.457 | 76 | 0.94 | 0.929 | 0.068 | 0.043 | 139096 |

> sem1<-'sotru =~NA\*+sotru1+sotru2+sotru3

+ truin =~NA\*truin1+truin2+truin3+truin4

+ webe =~NA\*webe1+webe2+webe3+webe4+webe5+webe6

+ sotru ~~1\*sotru

+ truin ~~1\*truin

+ webe ~~1\*webe

+ truin1 ~~ truin3

+ truin1 ~~ truin4

+ truin2 ~~ truin3

+ truin2 ~~ truin4

+ webe1 ~~ webe2

+ webe2 ~~ webe3

+ webe3 ~~ webe4

+ webe3 ~~ webe5

**d.**

> zess<- ess

> zess[,2:14]<-scale(ess[,2:14],center=TRUE,scale=FALSE)

>

> cancor.out<-cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)

+ ~ppltrst+ pplfair+ pplhlp+ trstprl+ trstlgl+ trstplc+ trstplt, data=zess)

Warning message:

In model.matrix.default(mt, mf, contrasts) :

non-list contrasts argument ignored

> summary(cancor.out)

Canonical correlation analysis of:

7 X variables: ppltrst, pplfair, pplhlp, trstprl, trstlgl, trstplc, trstplt

with 6 Y variables: fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl

CanR CanRSQ Eigen percent cum scree

1 0.242629 5.887e-02 6.255e-02 77.37503 77.38 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

2 0.110279 1.216e-02 1.231e-02 15.22875 92.60 \*\*\*\*\*\*

3 0.063206 3.995e-03 4.011e-03 4.96159 97.57 \*\*

4 0.041142 1.693e-03 1.696e-03 2.09741 99.66 \*

5 0.016167 2.614e-04 2.614e-04 0.32339 99.99

6 0.003343 1.118e-05 1.118e-05 0.01383 100.00

Test of H0: The canonical correlations in the

current row and all that follow are zero

CanR LR test stat approx F numDF denDF Pr(> F)

1 0.242629 0.92415 7.6396 42 18920 < 2.2e-16 \*\*\*

2 0.110279 0.98196 2.4539 30 16138 1.618e-05 \*\*\*

3 0.063206 0.99405 1.2056 20 13384 0.2378

4 0.041142 0.99804 0.6617 12 10678 0.7897

5 0.016167 0.99973 0.1834 6 8074 0.9815

6 0.003343 0.99999 NaN 2 NaN NaN

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Raw canonical coefficients

X variables:

Xcan1 Xcan2 Xcan3 Xcan4 Xcan5 Xcan6

ppltrst -0.108405 0.2853267 0.108076 -0.0495087 0.013165 -0.4336276

pplfair -0.282079 -0.4593998 -0.159970 0.1408960 -0.075323 -0.0624360

pplhlp -0.075600 0.2800853 -0.260620 0.1410840 0.205822 0.2527408

trstprl -0.097725 -0.1535614 0.320595 -0.0890089 0.367309 -0.0048426

trstlgl -0.080180 0.0573613 0.099537 -0.0092008 0.111209 0.2620818

trstplc -0.081954 -0.0060607 -0.126314 -0.4440381 -0.209336 -0.0123684

trstplt 0.067069 0.1206186 0.053729 0.3492442 -0.542852 0.0720115

Y variables:

Ycan1 Ycan2 Ycan3 Ycan4 Ycan5 Ycan6

fltdpr -0.38390 0.072259 0.26753 1.64697 -0.246829 0.463516

fltsd -0.20901 -0.016788 1.14942 -0.96373 0.813397 0.793988

fltanx -0.32340 1.188032 -0.46127 -0.10370 0.304109 -0.832190

wrhpp -0.22731 -0.035385 -1.35617 -0.33622 -0.046116 0.980694

enjlf -0.46398 -0.941008 0.22904 0.12495 0.474741 -1.137365

fltpcfl -0.18295 -0.034972 0.40065 -0.42358 -1.232723 -0.058797

>

> #redundancies

> redu<-redundancy(cancor.out)

> round(redu$Ycan,3)

Ycan1 Ycan2 Ycan3 Ycan4 Ycan5 Ycan6

0.030 0.002 0.000 0.000 0.000 0.000

>

> #computation redundancies from output

> R2tu<-cancor.out$cancor^2

> VAFYbyt<-apply(cancor.out$structure$Y.yscores^2,2,sum)/4

> redund<-R2tu\*VAFYbyt

> round(cbind(R2tu,VAFYbyt,redund,total=cumsum(redund)),3)

R2tu VAFYbyt redund total

Ycan1 0.059 0.770 0.045 0.045

Ycan2 0.012 0.192 0.002 0.048

Ycan3 0.004 0.132 0.001 0.048

Ycan4 0.002 0.131 0.000 0.048

Ycan5 0.000 0.154 0.000 0.048

Ycan6 0.000 0.122 0.000 0.048

**e.**

> samplesize<-dim(ess)[1]

> train<-ess[seq(2,samplesize,by=2),2:14]

> valid<-ess[seq(1,samplesize,by=2),2:14]

> train<-as.data.frame(scale(train,center=TRUE,scale=TRUE))

> valid<-as.data.frame(scale(valid,center=TRUE,scale=TRUE))

>

> #conduct CCA on training data

> cancor.train<-cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)

+ ~ppltrst+ pplfair+ pplhlp+ trstprl+ trstlgl+ trstplc+ trstplt, data=train)

Warning message:

In model.matrix.default(mt, mf, contrasts) :

non-list contrasts argument ignored

>

> #conduct CCA on validation data

> cancor.valid<-cancor(cbind(fltdpr, fltsd, fltanx, wrhpp, enjlf, fltpcfl)

+ ~ppltrst+ pplfair+ pplhlp+ trstprl+ trstlgl+ trstplc+ trstplt, data=valid)

Warning message:

In model.matrix.default(mt, mf, contrasts) :

non-list contrasts argument ignored

>

> # canonical variates calibration set

> train.X1<-cancor.train$score$X

> train.Y1<-cancor.train$score$Y

>

> # compute canonical variates using data of calibration set and coefficients estimated on validation set

> train.X2<-as.matrix(train[,1:7])%\*%cancor.valid$coef$X

> train.Y2<-as.matrix(train[,8:13])%\*%cancor.valid$coef$Y

>

>

> #R(T,T\*) and R(U,U\*)

> round(cor(train.Y1,train.Y2)[1:3,1:3],3)

Ycan1 Ycan2 Ycan3

Ycan1 0.989 -0.111 -0.029

Ycan2 0.101 0.817 0.372

Ycan3 -0.053 -0.207 0.288

> round(cor(train.X1,train.X2)[1:3,1:3],3)

Xcan1 Xcan2 Xcan3

Xcan1 0.982 -0.042 -0.090

Xcan2 0.029 0.514 0.370

Xcan3 0.082 -0.095 0.405

>

> #R(U,T) and R(U\*,T\*)

> round(cor(train.X1,train.Y1)[1:3,1:3],3)

Ycan1 Ycan2 Ycan3

Xcan1 0.253 0.000 0.000

Xcan2 0.000 0.129 0.000

Xcan3 0.000 0.000 0.067

> round(cor(train.X2,train.Y2)[1:3,1:3],3)

Ycan1 Ycan2 Ycan3

Xcan1 0.246 -0.028 -0.001

Xcan2 -0.002 0.044 0.031

Xcan3 -0.019 0.039 0.043

>

> #R(T\*,T\*) and R(U\*,U\*)

> round(cor(train.Y2,train.Y2)[1:3,1:3],3)

Ycan1 Ycan2 Ycan3

Ycan1 1.000 -0.019 0.001

Ycan2 -0.019 1.000 0.015

Ycan3 0.001 0.015 1.000

> round(cor(train.X2,train.X2)[1:3,1:3],3)

Xcan1 Xcan2 Xcan3

Xcan1 1.000 -0.007 0.015

Xcan2 -0.007 1.000 -0.002

Xcan3 0.015 -0.002 1.000

**f.**

> as.matrix(round(cancor.out$structure$X.xscores[,1],3))

[,1]

ppltrst -0.676

pplfair -0.842

pplhlp -0.609

trstprl -0.544

trstlgl -0.599

trstplc -0.554

trstplt -0.441

> as.matrix(round(cancor.out$structure$Y.yscores[,1],3))

[,1]

fltdpr -0.740

fltsd -0.687

fltanx -0.677

wrhpp -0.749

enjlf -0.793

fltpcfl -0.644