Solution in R: Exercises Day 5

PSY8003

Matthias Mittner

spring 2022

```
library(tidyverse)
```

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

v ggplot2 3.3.5     v purrr   0.3.4
v tibble  3.1.6     v dplyr   1.0.7
v tidyr   1.1.4     v stringr 1.4.0
v readr   2.1.1     v forcats 0.5.1

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

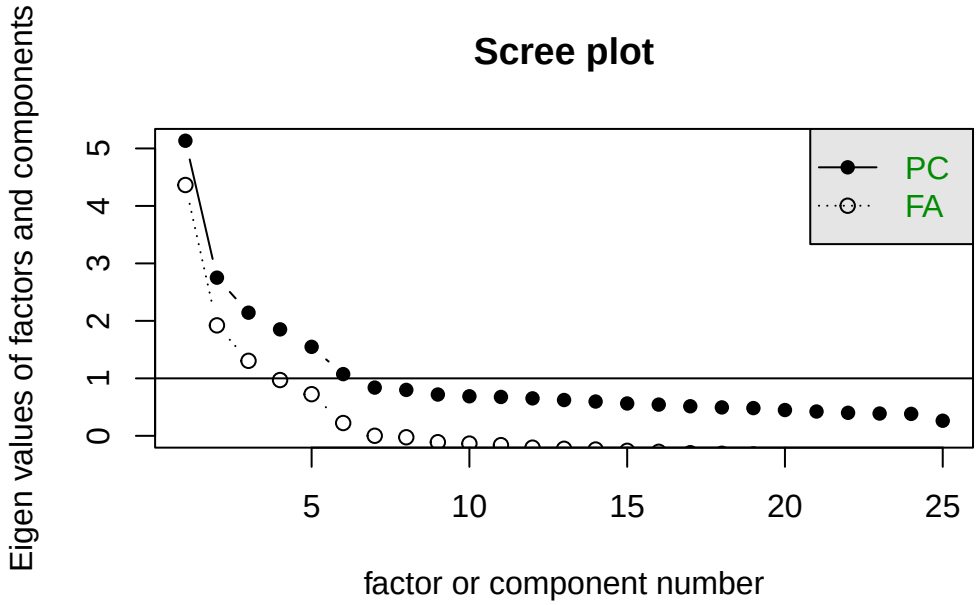
Exercise 1: Factor analysis

- Scree-plot suggests bw 4 and 5 factors
- Parallel analysis suggests somewhat more, up to 7 or 8
- we go with 5 because of theory (Big-Five)
- the loadings are as expected but there are quite some cross-loadings as well
- the CFA fit is not great and there are quite a few cross-loadings; not sure why that is the case, this is supposed to be a well-validated questionnaire

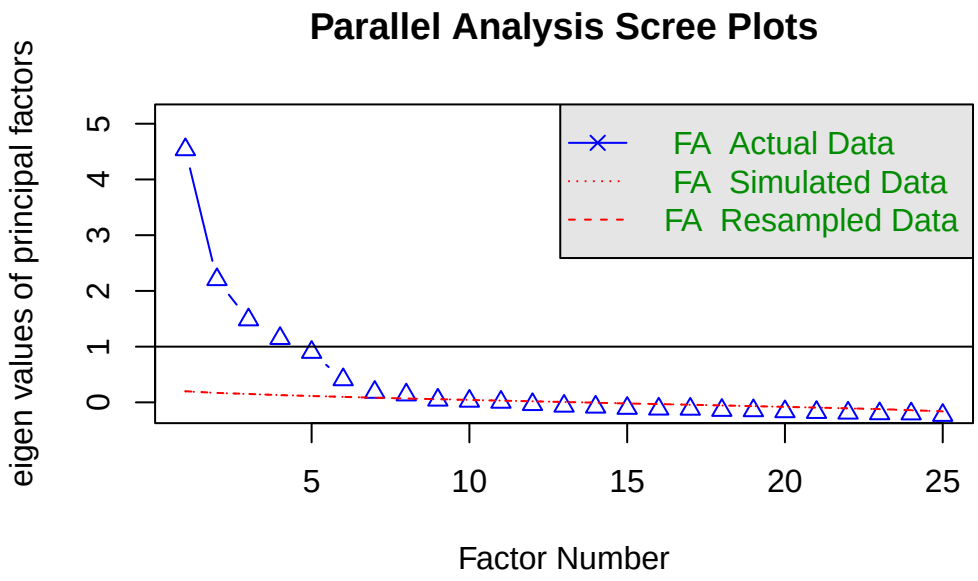
```
library(psych)
```

```
Attaching package: 'psych'

The following objects are masked from 'package:ggplot2':

    %+%, alpha
```

```r
bfi <- haven::read_dta("../data/psychbfi.dta")

scree(bfi)
```




```r
paranalysis <- fa.parallel(bfi, n.iter = 50, fm="pa", fa="fa", SMC="TRUE")
```




```
Parallel analysis suggests that the number of factors =  8  and the number of components =  1
```

```
fmod <- fa(bfi, nfactors = 5, fm="pa",  rotate = "varimax")
summary(fmod)
```

```
Factor analysis with Call: fa(r = bfi, nfactors = 5, rotate = "varimax", fm = "pa")

Test of the hypothesis that 5 factors are sufficient.
The degrees of freedom for the model is 185  and the objective function was  0.64
The number of observations was  2436  with Chi Square =  1538.69  with prob <  8e-212

The root mean square of the residuals (RMSA) is  0.03
The df corrected root mean square of the residuals is  0.04

Tucker Lewis Index of factoring reliability =  0.877
RMSEA index =  0.055  and the 10 % confidence intervals are  0.052 0.057
BIC =  96.03
```

```
print(fmod$loadings, digits=4, cutoff=0.4)
```

```
Loadings:
   PA2     PA1     PA3     PA5     PA4
A1                         -0.4283
A2                          0.6270
A3                          0.6505
A4                          0.4356
A5                          0.5370
C1                  0.5459
C2                  0.6486
C3                  0.5571
C4                 -0.6338
C5                 -0.5625
E1         -0.5749
E2         -0.6786
E3          0.5369
E4          0.6469
E5          0.5041
N1  0.7864
N2  0.7542
N3  0.7318
N4  0.5907
N5  0.5379
O1                                  0.5049
O2                                 -0.4690
O3                                  0.5957
O4
O5                                 -0.5339

                 PA2    PA1    PA3    PA5    PA4
SS loadings    2.7095 2.4734 2.0411 1.8441 1.5220
Proportion Var 0.1084 0.0989 0.0816 0.0738 0.0609
Cumulative Var 0.1084 0.2073 0.2890 0.3627 0.4236
```

```r
cbind(h2=fmod$communality, u2=fmod$uniquenesses)
```

```
          h2        u2
A1 0.2040368 0.7959632
A2 0.4628830 0.5371170
A3 0.5394745 0.4605255
A4 0.3019232 0.6980768
```

```
A5 0.4700367 0.5299633
C1 0.3484215 0.6515785
C2 0.4537165 0.5462835
C3 0.3243261 0.6756739
C4 0.4767162 0.5232838
C5 0.4354071 0.5645929
E1 0.3478146 0.6521854
E2 0.5453487 0.4546513
E3 0.4411086 0.5588914
E4 0.5413144 0.4586856
E5 0.4071865 0.5928135
N1 0.6806448 0.3193552
N2 0.6082398 0.3917602
N3 0.5445960 0.4554040
N4 0.5058311 0.4941689
N5 0.3493755 0.6506245
O1 0.3173782 0.6826218
O2 0.2675234 0.7324766
O3 0.4743090 0.5256910
O4 0.2460417 0.7539583
O5 0.2964037 0.7035963
```

```
library(lavaan)
```

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


Attaching package: 'lavaan'

The following object is masked from 'package:psych':

    cor2cov
```

```
model="
Agreeableness =~ A1+A2+A3+A4+A5
Openness =~ O1+O2+O3+O4+O5
Conscientiousness =~ C1+C2+C3+C4+C5
Extraversion =~ E1+E2+E3+E4+E5
Neuroticism =~ N1+N2+N3+N4+N5
```

```
"
cfamod <- cfa(model, bfi)
summary(cfamod, fit.measures=T, estimates=T, standardized=T)
```

```
lavaan 0.6-9 ended normally after 55 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        60

  Number of observations                          2436

Model Test User Model:

  Test statistic                              4165.467
  Degrees of freedom                               265
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                             18222.116
  Degrees of freedom                               300
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.782
  Tucker-Lewis Index (TLI)                       0.754

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -99840.238
  Loglikelihood unrestricted model (H1)     -97757.504

  Akaike (AIC)                              199800.476
  Bayesian (BIC)                            200148.363
  Sample-size adjusted Bayesian (BIC)       199957.729

Root Mean Square Error of Approximation:

  RMSEA                                          0.078
```

```
  90 Percent confidence interval - lower         0.076
  90 Percent confidence interval - upper         0.080
  P-value RMSEA <= 0.05                          0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.075

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
  Agreeableness =~
    A1                  1.000                               0.484    0.344
    A2                 -1.579    0.108  -14.650    0.000   -0.764   -0.648
    A3                 -2.030    0.134  -15.093    0.000   -0.983   -0.749
    A4                 -1.564    0.115  -13.616    0.000   -0.757   -0.510
    A5                 -1.804    0.121  -14.852    0.000   -0.873   -0.687
  Openness =~
    O1                  1.000                               0.635    0.564
    O2                 -1.020    0.068  -14.962    0.000   -0.648   -0.418
    O3                  1.373    0.072   18.942    0.000    0.872    0.724
    O4                  0.437    0.048    9.160    0.000    0.277    0.233
    O5                 -0.960    0.060  -16.056    0.000   -0.610   -0.461
  Conscientiousness =~
    C1                  1.000                               0.680    0.551
    C2                  1.148    0.057   20.152    0.000    0.781    0.592
    C3                  1.036    0.054   19.172    0.000    0.705    0.546
    C4                 -1.421    0.065  -21.924    0.000   -0.967   -0.702
    C5                 -1.489    0.072  -20.694    0.000   -1.012   -0.620
  Extraversion =~
    E1                  1.000                               0.920    0.564
    E2                  1.226    0.051   23.899    0.000    1.128    0.699
    E3                 -0.921    0.041  -22.431    0.000   -0.847   -0.627
    E4                 -1.121    0.047  -23.977    0.000   -1.031   -0.703
    E5                 -0.808    0.039  -20.648    0.000   -0.743   -0.553
  Neuroticism =~
    N1                  1.000                               1.300    0.825
    N2                  0.947    0.024   39.899    0.000    1.230    0.803
```

N3	0.884	0.025	35.919	0.000	1.149	0.721
N4	0.692	0.025	27.753	0.000	0.899	0.573
N5	0.628	0.026	24.027	0.000	0.816	0.503

Covariances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Agreeableness ~~						
Openness	-0.093	0.011	-8.446	0.000	-0.303	-0.303
Conscientisnss	-0.110	0.012	-9.254	0.000	-0.334	-0.334
Extraversion	0.304	0.025	12.293	0.000	0.683	0.683
Neuroticism	0.141	0.018	7.712	0.000	0.223	0.223
Openness ~~						
Conscientisnss	0.130	0.014	9.190	0.000	0.301	0.301
Extraversion	-0.265	0.021	-12.347	0.000	-0.453	-0.453
Neuroticism	-0.093	0.022	-4.138	0.000	-0.112	-0.112
Conscientiousness ~~						
Extraversion	-0.224	0.020	-11.121	0.000	-0.357	-0.357
Neuroticism	-0.250	0.025	-10.117	0.000	-0.283	-0.283
Extraversion ~~						
Neuroticism	0.292	0.032	9.131	0.000	0.244	0.244

Variances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.A1	1.745	0.052	33.725	0.000	1.745	0.882
.A2	0.807	0.028	28.396	0.000	0.807	0.580
.A3	0.754	0.032	23.339	0.000	0.754	0.438
.A4	1.632	0.051	31.796	0.000	1.632	0.740
.A5	0.852	0.032	26.800	0.000	0.852	0.528
.O1	0.865	0.032	27.216	0.000	0.865	0.682
.O2	1.990	0.063	31.618	0.000	1.990	0.826
.O3	0.691	0.039	17.717	0.000	0.691	0.476
.O4	1.346	0.040	34.036	0.000	1.346	0.946
.O5	1.380	0.045	30.662	0.000	1.380	0.788
.C1	1.063	0.035	30.073	0.000	1.063	0.697
.C2	1.130	0.039	28.890	0.000	1.130	0.650
.C3	1.170	0.039	30.194	0.000	1.170	0.702
.C4	0.960	0.040	24.016	0.000	0.960	0.507
.C5	1.640	0.059	27.907	0.000	1.640	0.615
.E1	1.814	0.058	31.047	0.000	1.814	0.682
.E2	1.332	0.049	26.928	0.000	1.332	0.512
.E3	1.108	0.038	29.522	0.000	1.108	0.607
.E4	1.088	0.041	26.732	0.000	1.088	0.506
.E5	1.251	0.040	31.258	0.000	1.251	0.694

.N1	0.793	0.037	21.575	0.000	0.793	0.320
.N2	0.836	0.036	23.458	0.000	0.836	0.356
.N3	1.222	0.043	28.271	0.000	1.222	0.481
.N4	1.654	0.052	31.977	0.000	1.654	0.672
.N5	1.969	0.060	32.889	0.000	1.969	0.747
Agreeableness	0.234	0.030	7.839	0.000	1.000	1.000
Openness	0.404	0.033	12.156	0.000	1.000	1.000
Conscientisnss	0.463	0.036	12.810	0.000	1.000	1.000
Extraversion	0.846	0.062	13.693	0.000	1.000	1.000
Neuroticism	1.689	0.073	23.034	0.000	1.000	1.000

Exercise 2: Multi-level regression

- there is a clear positive effect between homework score and math score across all students
- however, in the MLM, this effect is gone. Looking at the individual slopes, there is huge variability between schools. In some schools, students profit a lot from homework, in others they don't.
- at the individual level, socio-economic status (`ses`) is a good predictor for math achievement
- at the school level, mean-SES is predictive of a school's students math achievements

```
library(lmerTest)
```

```
Loading required package: lme4
```

```
Loading required package: Matrix
```

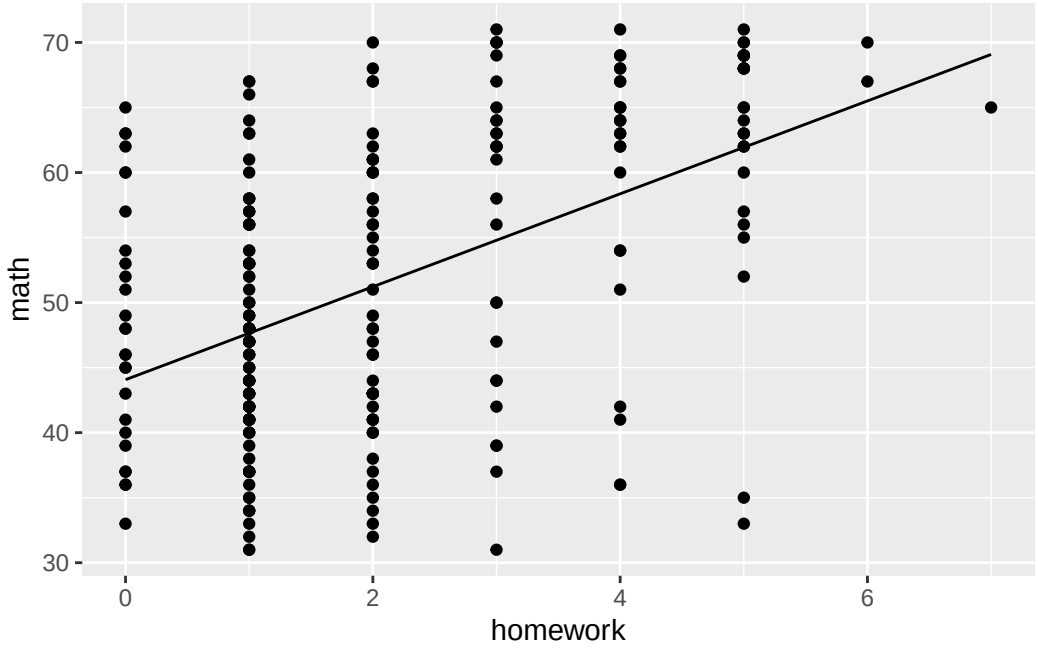
```

Attaching package: 'Matrix'
```

```
The following objects are masked from 'package:tidyr':

    expand, pack, unpack
```

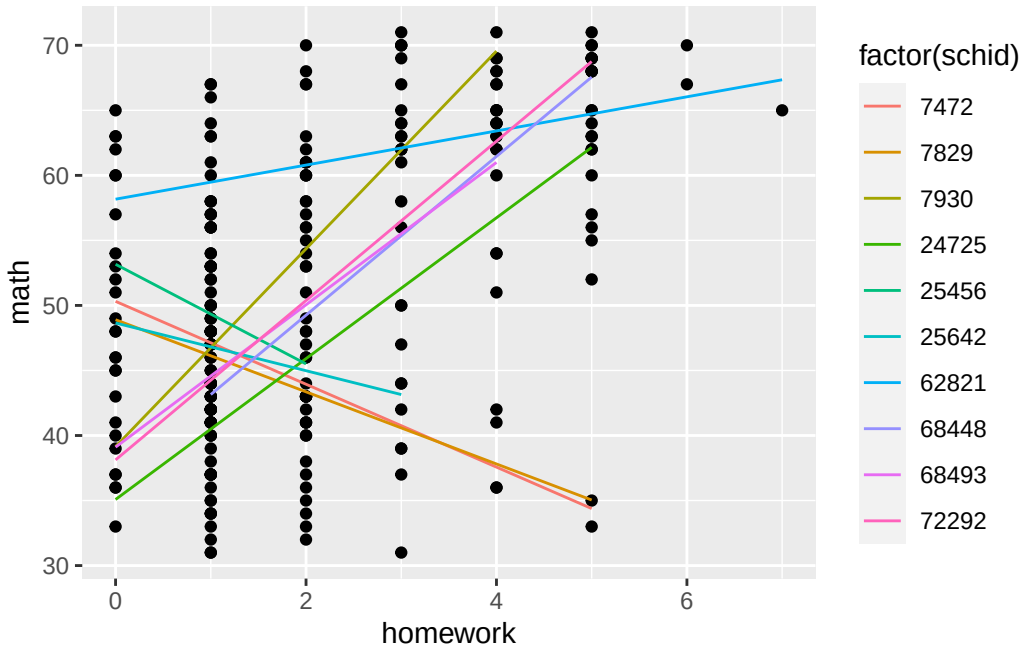
```

Attaching package: 'lmerTest'
```

```
The following object is masked from 'package:lme4':

    lmer

The following object is masked from 'package:stats':

    step
```

```
imm10 <- haven::read_dta("../data/imm10.dta")

mod.lm <- lm(math ~ homework, data=imm10)
ggplot(imm10, aes(x=homework, y=math))+
  geom_point()+geom_line(aes(y=predict(mod.lm)))
```




```
mod.lmer <- lmer(math ~ homework + (1+homework|schid), data=imm10)
ggplot(imm10, aes(x=homework, y=math))+
  geom_point()+geom_line(aes(y=predict(mod.lmer), color=factor(schid)))
```




```
mod <- lmer(math ~ homework + ses + meanses + (1+homework|schid), REML=FALSE, data = imm10
summary(mod)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: math ~ homework + ses + meanses + (1 + homework | schid)
   Data: imm10

     AIC      BIC   logLik deviance df.resid
  1752.0   1780.5   -868.0   1736.0      252

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.65746 -0.67497  0.03454  0.63597  2.65142

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 schid    (Intercept) 49.19    7.013
          homework    17.07    4.132    -0.99
 Residual             41.32    6.428
Number of obs: 260, groups:  schid, 10
```

```
Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  48.0222     2.3521   9.9748  20.416 1.82e-09 ***
homework      1.8024     1.3598   9.6955   1.325 0.215413
ses           2.3765     0.6359 239.2005   3.737 0.000233 ***
meanses       6.2303     1.0196  11.6260   6.110 6.02e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) homwrk ses
homework -0.973
ses       0.042 -0.037
meanses   0.066 -0.009 -0.611
```