

Research Methods

Factor Analysis with R

Andy Gillett

1

Factor Analysis

Factor analysis is a statistical technique for analysing the correlations between a number of variables in order to reduce them to a smaller number of underlying dimensions, called factors, and to determine the correlation of each of the original variables with each factor.

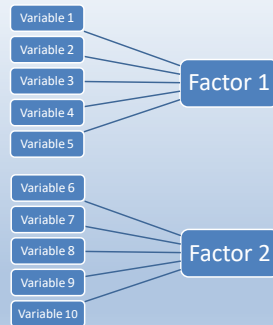
2

Factor Analysis

- Variable 1
- Variable 2
- Variable 3
- Variable 4
- Variable 5
- Variable 6
- Variable 7
- Variable 8
- Variable 9
- Variable 10

3

Factor Analysis



The amount of the total variance explained by that factor is called the **eigenvalue** of a factor.

The extent to which a particular variable contributes to a given factor is the **factor loading** of a variable.

4

Factor Analysis

Job Satisfaction

interest	Job Interest
supervis	Relationship with Supervisor
facilit	Facilities in the Workplace
peerrel	Relationship with Peers/Colleagues
persdev	Opportunities for Personal Development & Promotion
workhours	Hours of Work
holidays	Holidays

5

Factor Analysis

	A	B	C	D	E	F	G	H	I	J	K
1	interest	supervis	facilit	peerrel	persdev	workhour	holidays				
2	5	5	2	3	8	6	7				
3	4.5	2	3	4	3	2	4				
4	7	4	4	8	7	5	5				
5	6.25	3	3	9	8	7	3				
6	5.25	1	5	7	9	6	4				
7	4	4	4	3	3	4	1				
8	3.5	2	2	4	5	7	3				
9	6.75	7	3	5	8	6	4				
10	7.75	4	6	7	8	4	8				
11	5.75	1	3	5	5	1	9				
12	4	2	4	3	3	5	7				
13	4.75	4	2	5	5	7	3				
14	6	3	4	7	9	4	4				
15	6.5	5	1	8	8	1	5				
16	3.75	3	1	3	2	4	7				
17	5.75	4	2	8	9	6	5				
18	8.5	5	5	8	7	5	3				
19	6.75	3	3	7	6	9	5				
20	7.75	4	3	6	8	5	7				
21	4.5	2	5	4	4	4	8				
22	4	3	4	3	4	6	3				
23	7	5	3	5	7	7	8				
24	7.75	3	5	7	8	5	8				
25	8	3	3	6	6	3	7				
26	4	1	3	4	5	4	6				
27	4.5	2	4	5	5	6	4				
28	8	4	1	3	3	7	3				
29	7.5	1	2	5	6	8	5				
30	6.75	5	3	6	4	8	7				

6

[illegible]

7

The screenshot shows the RStudio interface. The script editor contains the following R code:

```
RStudio (64-bit)
File Edit View Misc Packages Windows Help

> x = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
> x = factor(x)
> prcomp(x)
> prcomp(x)$x1
[1] 0.3717537 0.6431910 0.2171469 0.4252194 0.5403609 0.7825595
[7] 0.1532814 0.3143279 0.4838384 0.6431910
```

8

[illegible]

9

Factor Analysis

The screenshot shows an R console window with the following text:

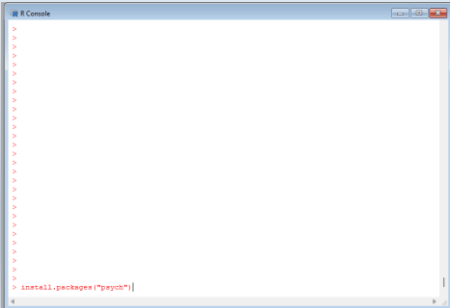
```
>
>
> jobstat <- read.csv("job-satisfaction.csv", header=TRUE, colClasses=c("numeric", "mul
> jobstat
1      interest supervisor facilities personal services workhours holidays
2      9.500      5      2      3      8      6      7
3      4.500      2      3      4      3      2      4
4      7.500      4      4      8      7      5      5
5      4.250      3      3      9      8      7      3
6      3.250      1      5      7      9      6      4
7      3.000      4      4      3      3      4      1
8      3.500      2      5      4      1      7      1
9      6.750      7      3      5      8      6      4
10     7.750      4      6      7      5      4      8
11     8.750      1      3      5      5      1      9
12     4.000      2      4      3      3      5      7
13     6.750      4      3      5      5      7      3
14     6.000      3      4      7      9      4      4
15     6.500      5      1      8      9      3      5
16     5.750      3      1      3      2      4      7
17     5.750      4      2      8      9      4      5
18     8.500      5      5      7      5      5      3
19     4.750      3      3      7      6      9      5
20     7.750      4      3      4      2      5      7
21     4.500      2      5      4      4      4      8
```

10

[illegible]

11

Factor Analysis



The screenshot shows an R Console window with the title "R Console". The command prompt is at the bottom, showing the command `> install.packages("psych")` entered. The console output area is empty, and the command line is currently blank.

12

12

[illegible]

13

Factor Analysis

```
> describe(jobsat)
      vars  n mean   sd median trimmed  mad min max range   skew kurtosis   se
interest  1 44 5.92 1.52   5.88   5.90 2.04 3.5 8.5    5 0.08   -1.41 0.23
supervis  2 44 3.34 1.41   3.00   3.36 1.48 1.0 7.0    6 0.08   -0.58 0.21
facilit  3 44 3.16 1.24   3.00   3.17 1.48 1.0 6.0    5 0.14   -0.66 0.19
peerrel  4 44 5.20 1.86   5.00   5.19 2.22 1.0 9.0    8 -0.02   -0.85 0.28
persdev  5 44 5.84 1.99   6.00   5.83 2.97 2.0 9.0    7 -0.01   -1.24 0.30
workhours 6 44 5.34 1.92   5.00   5.33 2.22 1.0 9.0    8 -0.01   -0.90 0.29
holidays 7 44 5.07 1.93   5.00   5.06 2.97 1.0 9.0    8 0.02   -0.80 0.29
>
```

14

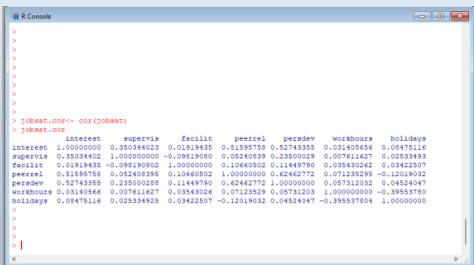
14

[illegible]

Factor Analysis

16

Factor Analysis



```
R Console
```

```
>
>
>
>
> jolibat.coe<- cor(jolibat)
> jolibat.coe
               interest      supervise      facilit      peerrel      peerdev      workhours      holidays
interest  1.000000000  0.305846023  0.01819495  0.51595750  0.52783305  0.03140546  0.08475116
supervise  0.30584602  1.000000000  -0.09819002  0.08240359  0.23800029  0.00761167  0.02533983
facilit    0.01819495 -0.09819002  1.00000000  0.10460502  0.11449790  0.03945202  0.03422507
peerrel    0.51595750  0.08240359  0.10460502  1.00000000  0.6246772  0.07121329  0.12119032
peerdev    0.52783305  0.23800029  0.11449790  0.6246772  1.00000000  0.00781203  0.04524047
workhours  0.03140546  0.00761167  0.03945202  0.07121329  0.00781203  1.00000000  -0.39583780
holidays   0.08475116  0.02533983  0.03422507  0.12119032  0.04524047  -0.39583780  1.00000000
>
>
>
>
```

Factor Analysis

```
> jobstat.cor<- cor(jobstat)
> jobstat.cor
```

	interest	supervis	facilit	peerrel	perader	workhours	holidays
interest	1.000000000	0.350344023	-0.0919435	0.51595758	0.52743355	0.03140566	0.08475116
supervis	0.35034402	1.000000000	-0.09819080	0.05240839	0.23500029	0.007611627	0.02533493
facilit	0.01918435	-0.098190802	1.000000000	0.10660502	0.11449790	0.035430262	0.03422507
peerrel	0.51595758	0.052408395	0.10660502	1.000000000	0.62462772	0.071235295	-0.12018032
perader	0.52743355	0.235000029	0.11449790	0.62462772	1.000000000	0.057312032	0.04524047
workhours	0.03140566	0.007611627	0.03543026	0.07123529	0.05731203	1.000000000	-0.39553780
holidays	0.08475116	0.025334925	0.03422507	-0.12019032	0.04524047	-0.395537804	1.000000000

18

Factor Analysis

```
RStudio Console
```

```
> 
> 
> 
> 
> 
> 
> 
> 
> RMD(jobsat)
Falconer-Meyer-Olkin factor adequacy
Call: RMD(x = jobsat)
Overall NGA = 0.6
NGA for each item =
      interest    supervis     facilit   peerrel     perdev workhours   holidays
       0.69        0.49         0.57       0.59        0.67        0.49        0.41
> 
>
```

19

Factor Analysis

```
R Console
```

```
>
>
>
>
>
>
>
>
>
>
> source("test.jobnet")
R was not square, finding R from data
finding
[1] 56.27619

Sp.value
[1] 9.006324e-05

Sdf
[1] 21

>
>
```

20

Factor Analysis



The screenshot shows an R Console window. The title bar reads "R Console". The window contains a series of red prompt characters (>) on the left margin. The first line of code entered is `det(jobast.cox)`, which has been executed, resulting in the output `[1] 0.2559909` displayed on the line immediately following the prompt.

21

Factor Analysis

[illegible]

22

Factor Analysis

[illegible]

23

Factor Analysis

```
> jobsat.eigen$values
[1] 2.2399384 1.4173697 1.1273358 0.8457473 0.6128181 0.4503572 0.3064335
```

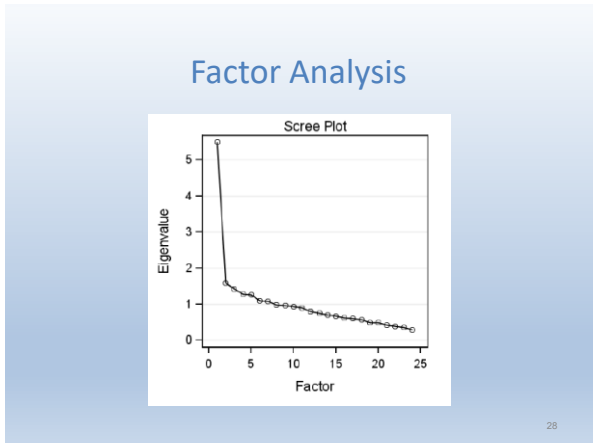
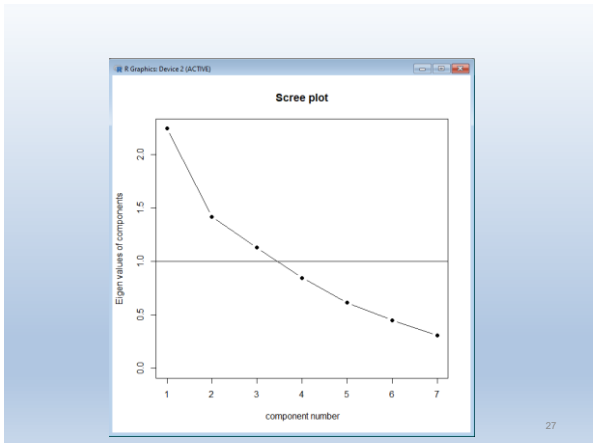
Tabachnick and Fidell (2013) recommend that researchers adopt an exploratory approach, experimenting with different numbers of factors until a satisfactory solution is found.

24

Factor Analysis

```
R Console  
>  
>  
# load eigen -<= eigen (jobstat.cov)  
> library(eigen)  
eigen(1, decomposition="eigen")  
Principle components  
[1] 2.3999304 4.4173697 1.1273358 0.8467473 0.4129181 0.4503572 0.3064335  
  
Eigenvalues  
      [1]          [2]          [3]          [4]          [5]          [6]  
[1,] = -0.54604620  -0.11850580  -0.12633116  -0.02366216  -0.14656492  -0.78300592  
[2,] = -0.27766193  -0.14672983  -0.09496920  -0.01411116  -0.24890008  -0.19424321  
[3,] = -0.08068661  -0.04763072  -0.73371359  -0.44500037  -0.14766129  -0.08137623  
[4,] = -0.83771890  -0.11423006  -0.21846331  -0.37038460  -0.54727000  -0.02707142  
[5,] = -0.84705488  -0.00312514  -0.10358464  -0.10060394  -0.04674613  -0.45203665  
[6,] = -0.07667554  -0.47368450  -0.09624333  -0.32876149  -0.47677910  -0.08425493  
[7,] = 0.01112475  -0.703941720  -0.19334026  -0.00094660  -0.63681545  -0.10851456  
[8,] = -0.252912390  
  
# load eigenValues  
[1] 2.3999304 4.4173697 1.1273358 0.8467473 0.4129181 0.4503572 0.3064335  
  
# score[jobstat,FAC1$E]  
>
```

The screenshot shows a terminal window titled "R Console". It contains several lines of R code and their corresponding outputs. The code starts with loading the "eigen" package and performing an eigen decomposition on a matrix. This is followed by printing the principal components and eigenvalues. The output displays two matrices: one for principal components (labeled "Principle components") and one for eigenvalues (labeled "Eigenvalues"). Both matrices have 8 columns and 1 row. The values are displayed in scientific notation. The final line of code is a comment "# score[jobstat,FAC1\$E]" followed by a prompt character ">".



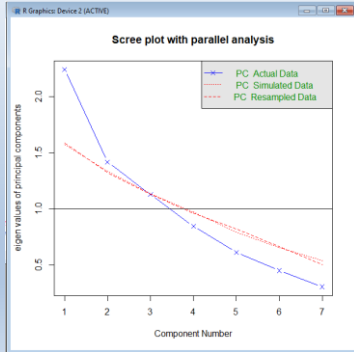
28

Factor Analysis

[illegible]

30

Factor Analysis



31

Factor Analysis

```
> jobstat.analysis <- principal(jobstat, nfactors=3, rotate="none")
> jobstat.analysis
Principal Components Analysis
Call: principal(x = jobstat, nfactors = 3, rotate = "none")
Standardized Loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	h2	u2	com
interest	0.82	0.14	-0.13	0.71	0.29	1.1
supervis	0.42	0.17	-0.64	0.61	0.39	1.9
facilit	0.13	-0.06	0.78	0.63	0.37	1.1
peerrel	0.80	-0.13	0.23	0.72	0.28	1.2
persdev	0.85	0.04	0.12	0.74	0.26	1.0
workhours	0.12	-0.80	-0.10	0.67	0.33	1.1
holidays	-0.02	0.84	0.14	0.72	0.28	1.1

```
SS loadings: 2.24 1.42 1.13
Proportion Var: 0.32 0.20 0.16
Cumulative Var: 0.32 0.52 0.68
Proportion Explained: 0.47 0.30 0.24
Cumulative Proportion: 0.47 0.76 1.00

Mean item complexity = 1.2
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 29.54 with prob < 1.7e-06

Fit based upon off diagonal values = 0.75>
```

32

Factor Analysis

```
> jobstat.analysis
Principal Components Analysis
Call: principal(x = jobstat, nfactors = 3, rotate = "none")
Standardized Loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	h2	u2	com
interest	0.82	0.14	-0.13	0.71	0.29	1.1
supervis	0.42	0.17	-0.64	0.61	0.39	1.9
facilit	0.13	-0.06	0.78	0.63	0.37	1.1
peerrel	0.80	-0.13	0.23	0.72	0.28	1.2
persdev	0.85	0.04	0.12	0.74	0.26	1.0
workhours	0.12	-0.80	-0.10	0.67	0.33	1.1
holidays	-0.02	0.84	0.14	0.72	0.28	1.1

```
SS loadings: 2.24 1.42 1.13
Proportion Var: 0.32 0.20 0.16
Cumulative Var: 0.32 0.52 0.68
Proportion Explained: 0.47 0.30 0.24
Cumulative Proportion: 0.47 0.76 1.00

Mean item complexity = 1.2
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 29.54 with prob < 1.7e-06

Fit based upon off diagonal values = 0.75>
```

33

Factor Analysis

	PC1	PC2	PC3	h2	u2	com
interest	0.82	0.14	-0.13	0.71	0.29	1.1
supervis	0.42	0.17	-0.64	0.61	0.39	1.9
facilit	0.13	-0.06	0.78	0.63	0.37	1.1
peerrel	0.80	-0.13	0.23	0.72	0.28	1.2
persdev	0.85	0.04	0.12	0.74	0.26	1.0
workhours	0.12	-0.80	-0.10	0.67	0.33	1.1
holidays	-0.02	0.84	0.14	0.72	0.28	1.1

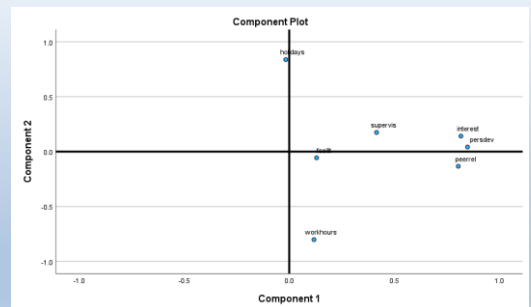
34

Factor Analysis

	PC1	PC2	PC3
SS loadings	2.24	1.42	1.13
Proportion Var	0.32	0.20	0.16
Cumulative Var	0.32	0.52	0.68
Proportion Explained	0.47	0.30	0.24
Cumulative Proportion	0.47	0.76	1.00

35

Factor Analysis

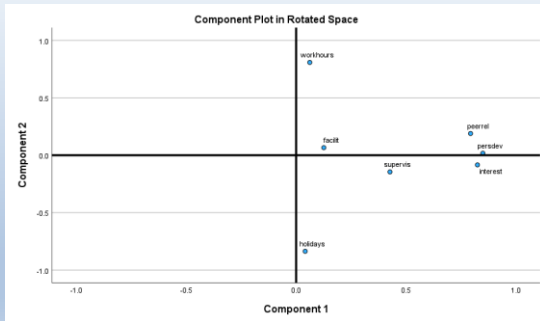


36

35

36

Factor Analysis



37

Factor Analysis

```
R Console
> jobstat.analysis <- principal(jobstat, nfactors=3, rotate="varimax")
> jobstat.analysis
Principal Components Analysis
Call: principal(r = jobstat, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
      RC1    RC2    RC3    h2    u2    com
interest 0.80 0.07 -0.25 0.71 0.29 1.2
supervis 0.34 0.05 -0.70 0.61 0.39 1.5
facilit 0.22 0.05 0.76 0.63 0.37 1.2
peerrel 0.82 -0.14 0.16 0.72 0.28 1.1
persdev 0.86 0.01 0.01 0.74 0.26 1.0
workhours 0.07 -0.81 0.01 0.67 0.33 1.0
holidays 0.03 0.85 0.01 0.72 0.28 1.0

SS loadings      2.22 1.41 1.15
Proportion Var   0.32 0.20 0.16
Cumulative Var   0.32 0.52 0.68
Proportion Explained 0.46 0.30 0.24
Cumulative Proportion 0.46 0.76 1.00

Mean item complexity = 1.1
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 29.54 with prob < 1.7e-06
Fit based upon off diagonal values = 0.75
```

38

Factor Analysis

```
> jobstat.analysis
Principal Components Analysis
Call: principal(r = jobstat, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
      RC1    RC2    RC3    h2    u2    com
interest 0.80 0.07 -0.25 0.71 0.29 1.2
supervis 0.34 0.05 -0.70 0.61 0.39 1.5
facilit 0.22 0.05 0.76 0.63 0.37 1.2
peerrel 0.82 -0.14 0.16 0.72 0.28 1.1
persdev 0.86 0.01 0.01 0.74 0.26 1.0
workhours 0.07 -0.81 0.01 0.67 0.33 1.0
holidays 0.03 0.85 0.01 0.72 0.28 1.0

SS loadings      2.22 1.41 1.15
Proportion Var   0.32 0.20 0.16
Cumulative Var   0.32 0.52 0.68
Proportion Explained 0.46 0.30 0.24
Cumulative Proportion 0.46 0.76 1.00

Mean item complexity = 1.1
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 29.54 with prob < 1.7e-06
Fit based upon off diagonal values = 0.75
```

39

Factor Analysis

	RC1	RC2	RC3	h2	u2	com
interest	0.80	0.07	-0.25	0.71	0.29	1.2
supervis	0.34	0.05	-0.70	0.61	0.39	1.5
facilit	0.22	0.05	0.76	0.63	0.37	1.2
peerrel	0.82	-0.14	0.16	0.72	0.28	1.1
persdev	0.86	0.01	0.01	0.74	0.26	1.0
workhours	0.07	-0.81	0.01	0.67	0.33	1.0
holidays	0.03	0.85	0.01	0.72	0.28	1.0

40

Factor Analysis

	RC1	RC2	RC3	h2	u2	com
interest	0.80	0.07	-0.25	0.71	0.29	1.2
supervis	0.34	0.05	-0.70	0.61	0.39	1.5
facilit	0.22	0.05	0.76	0.63	0.37	1.2
peerrel	0.82	-0.14	0.16	0.72	0.28	1.1
persdev	0.86	0.01	0.01	0.74	0.26	1.0
workhours	0.07	-0.81	0.01	0.67	0.33	1.0
holidays	0.03	0.85	0.01	0.72	0.28	1.0

41

Factor Analysis

	RC1	RC2	RC3
SS loadings	2.22	1.41	1.15
Proportion Var	0.32	0.20	0.16
Cumulative Var	0.32	0.52	0.68
Proportion Explained	0.46	0.30	0.24
Cumulative Proportion	0.46	0.76	1.00

42

37

38

39

40

41

42

Factor Analysis

```
R Console
Standardized loadings (pattern matrix) based upon correlation matrix
    RC1  RC2  RC3  h2  u2  com
interest 0.80 0.07 -0.25 0.71 0.29 1.2
supervis 0.34 0.05 -0.70 0.43 0.59 1.5
facilit  0.22 0.05 0.76 0.43 0.57 1.2
peerrel  0.82 -0.14 0.16 0.72 0.28 1.1
persdev  0.86 0.01 0.01 0.74 0.26 1.0
workhours 0.07 -0.81 0.01 0.47 0.53 1.0
holidays 0.03 0.85 0.01 0.72 0.28 1.0

SS loadings:    RC1  RC2  RC3
Proportion Var  0.32 0.20 0.16
Cumulative Var  0.32 0.52 0.68
Proportion Explained 0.46 0.30 0.24
Cumulative Proportion 0.46 0.76 1.00

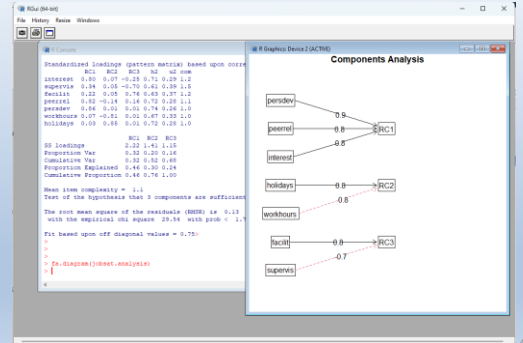
Mean item complexity = 1.1
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.13
with the empirical chi square 29.54 with prob < 1.7e-06

Fit based upon off diagonal values = 0.75>
>
>
>
> fa.diagram(jobtest.analysis)
>
<
```

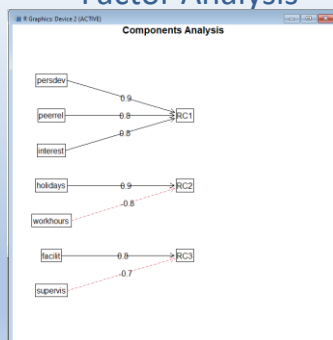
43

Factor Analysis



44

Factor Analysis



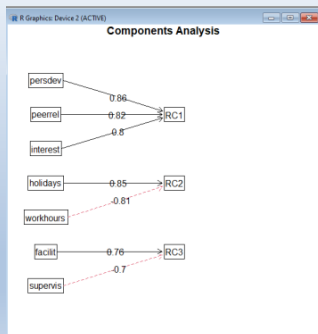
45

Factor Analysis

```
R Console
>
>
>
>
> fa.diagram(jobtest.analysis, digits=2)
>
<
```

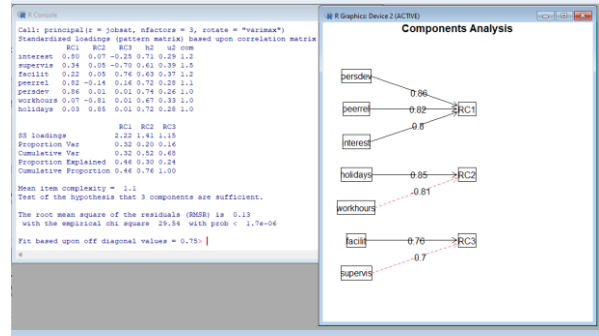
46

Factor Analysis



47

Factor Analysis



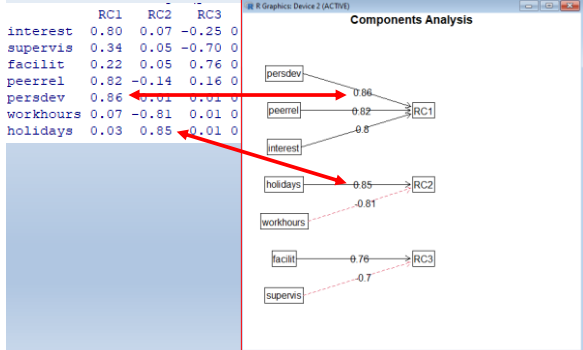
48

43

45

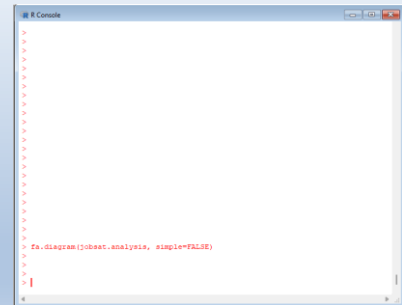
47

Factor Analysis

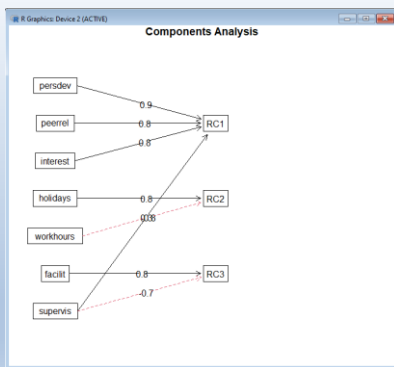


49

Factor Analysis

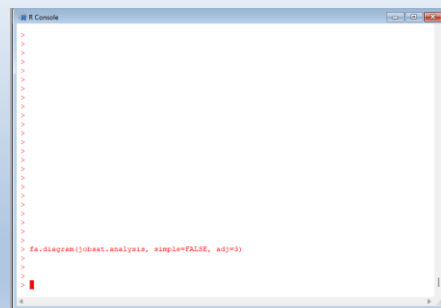


50



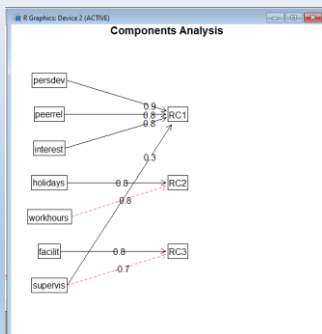
51

Factor Analysis



52

Factor Analysis



53

Factor Analysis

```
fa.diagram(jobsat.analysis, simple=FALSE)
```

```
fa.diagram(jobsat.analysis, digits=3)
fa.diagram(jobsat.analysis, cut=.7)
```

```
fa.diagram(jobsat.analysis, simple=FALSE, l.cex=3)
fa.diagram(jobsat.analysis, simple=FALSE, adj=4)
```

54

Factor Analysis

[illegible]

55

Factor Analysis

[illegible]

56

Factor Analysis

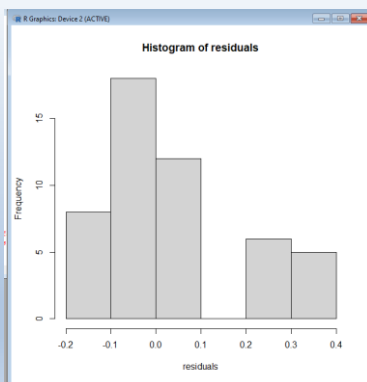
[illegible]

57

Factor Analysis

[illegible]

58



59

Factor Analysis

[illegible]

60

Factor Analysis

```
R Console
>
> jobset.scores.round <- round(jobset.scores,2)
> jobset.scores.round
      RC1  RC2  RC3
[1,] -0.19 0.47 -1.35
[2,] -1.31 0.40 0.43
[3,] 1.19 0.07 0.36
[4,] 1.01 -1.29 0.39
[5,] 0.83 -0.86 2.30
[6,] -1.37 -0.82 0.04
[7,] -1.23 -1.27 0.02
[8,] 0.86 -0.40 -1.70
[9,] 1.62 1.40 1.35
[10,] -0.44 2.47 0.96
[11,] -1.44 0.71 0.99
[12,] -0.57 -1.23 -0.86
[13,] 1.04 0.06 0.84
[14,] 1.00 0.53 -1.61
[15,] -1.86 0.86 -1.07
[16,] 1.00 -0.36 -0.55
[17,] 1.70 -0.43 0.20
[18,] 0.87 -1.18 0.12
[19,] 1.97 0.77 -0.40
[20,] -0.81 1.35 1.62
```

61

Factor Analysis

```
R Console
> jobsetplus <- cbind(jobset,jobset.scores.round)
> jobsetplus
      interest supervisor facilities peerrel peerdev workhours holidays  RC1  RC2  RC3
1  5.00  5  2  3  5  6  7  -0.19 0.47 -1.35
2  7.00  4  4  8  7  5  5  -1.31 0.40 0.43
3  7.00  4  4  8  7  5  5  1.19 0.07 0.36
4  6.25  3  3  9  8  7  3  1.01 -1.29 0.39
5  5.25  1  5  7  9  6  4  0.83 -0.86 2.30
6  6.00  4  4  3  3  4  1  -1.37 -0.82 0.04
7  3.50  2  2  4  5  7  3  -1.23 -1.27 0.02
8  6.75  7  3  5  5  6  4  0.86 -0.40 -1.70
9  7.75  4  6  7  8  4  8  1.62 1.40 1.35
10  5.75  1  3  5  5  1  9  -0.44 2.47 0.96
11  6.00  2  4  3  3  5  7  -1.44 0.71 0.99
12  6.75  4  2  5  5  7  3  -0.57 -1.23 -0.86
13  6.00  3  4  7  9  6  4  1.04 0.06 0.84
14  6.50  5  1  8  8  3  5  1.00 0.53 -1.61
15  3.75  3  1  3  2  4  7  -1.86 0.86 -1.07
16  5.75  4  2  8  9  6  5  1.00 -0.36 -0.55
17  8.50  5  5  8  7  5  3  1.70 -0.43 0.20
18  6.75  3  3  7  6  9  5  0.87 -1.18 0.12
19  7.75  4  3  6  8  5  7  1.97 0.77 -0.40
20  4.50  2  5  4  4  4  8  -0.81 1.35 1.62
```

62

Factor Analysis

```
R Console
>
> colnames(jobsetplus)[c(8,9,10)] <- c("Component 1", "Component 2", "Component 3")
> jobsetplus
      interest supervisor facilities peerrel peerdev workhours holidays Component 1 Component 2 Component 3
1  5.00  5  2  3  5  6  7  -0.19 0.47 -1.35
2  7.00  4  4  8  7  5  5  -1.31 0.40 0.43
3  7.00  4  4  8  7  5  5  1.19 0.07 0.36
4  6.25  3  3  9  8  7  3  1.01 -1.29 0.39
5  5.25  1  5  7  9  6  4  0.83 -0.86 2.30
6  6.00  4  4  3  3  4  1  -1.37 -0.82 0.04
7  3.50  2  2  4  5  7  3  -1.23 -1.27 0.02
8  6.75  7  3  5  5  6  4  0.86 -0.40 -1.70
9  7.75  4  6  7  8  4  8  1.62 1.40 1.35
10  5.75  1  3  5  5  1  9  -0.44 2.47 0.96
11  6.00  2  4  3  3  5  7  -1.44 0.71 0.99
12  6.75  4  2  5  5  7  3  -0.57 -1.23 -0.86
13  6.00  3  4  7  9  6  4  1.04 0.06 0.84
14  6.50  5  1  8  8  3  5  1.00 0.53 -1.61
15  3.75  3  1  3  2  4  7  -1.86 0.86 -1.07
16  5.75  4  2  8  9  6  5  1.00 -0.36 -0.55
17  8.50  5  5  8  7  5  3  1.70 -0.43 0.20
18  6.75  3  3  7  6  9  5  0.87 -1.18 0.12
19  7.75  4  3  6  8  5  7  1.97 0.77 -0.40
20  4.50  2  5  4  4  4  8  -0.81 1.35 1.62
```

63

Factor Analysis

Factor analysis, using principal components extraction and varimax rotation, was performed on the data. Three factors were extracted, Factor 1 explaining 32% of the variance in the original measures; Factor 2 explaining 20% and Factor 3 explaining an additional 16% cent of the variance. The job interest, relationship with colleagues and opportunities for personal development variables were found to have high factor loadings (.80, .82, .86) on Factor 1, hours of work and holidays had high factor loadings (.81, .85) on Factor 2 and facilities and relationship with supervisor had quite high loadings on Factor 3 (.70 - .76).

Based on Colman & Pulford (2008).

64

Factor Analysis

Factor Analysis with R - 2

65

Factor Analysis

Student Satisfaction

id	Identity
ethnicgp	Ethnic Group
age	Age
gender	Gender
years	Number of Years in Institution
ss1-ss12	Twelve Measures of Satisfaction of Course
attendance	Number of Hours Attending
skill	Initial Skill Level
participation	Class Participation
quality	Quality of Work
absence	Number of Absences

66

Factor Analysis

67

Factor Analysis

student-satisfaction.csv

68

Factor Analysis

69

Factor Analysis

70

Factor Analysis

71

Factor Analysis

72

Factor Analysis

```
R Console
> student <- studentall[, c(6:17)]
> student
  ss1 ss2 ss3 ss4 ss5 ss6 ss7 ss8 ss9 ss10 ss11 ss12
1 0 3 4 4 2 4 2 2 2 2 3 2
2 0 0 2 3 2 2 1 2 3 4 4 4
3 4 4 4 1 2 1 2 2 2 1 2 3
4 2 3 2 4 1 2 2 2 3 2 2 3
5 3 4 4 1 2 3 3 4 5 5 4
6 2 5 3 3 2 1 1 2 4 4 4 4
7 3 3 2 3 2 2 3 2 3 5 4 2
8 2 2 4 2 3 4 3 2 3 3 2
9 3 1 2 4 2 3 4 1 2 2 3 2
10 3 2 3 3 2 1 3 2 3 4 3 5
11 2 4 1 4 2 1 1 2 5 4 3 4
12 1 5 2 4 1 1 1 3 4 3 3
13 3 3 4 2 2 2 3 3 5 5 5 4
14 3 4 2 4 2 2 3 1 4 3 4 4
15 3 2 1 3 2 4 4 2 3 3
16 3 4 2 2 2 3 1 2 4 2 2 2
17 3 3 2 4 1 2 2 4 5 4 4
18 2 4 2 4 2 3 2 5 5 1
19 4 2 3 2 3 4 3 3 2 3 2
20 2 3 3 3 2 1 2 5 4 4 5
21 4 4 3 3 3 2 3 3 2 4 4
```

73

Factor Analysis

```
R Console
>
>
> library("psych")
> describe(student)
  variable  N  mean  sd median trimmed  mad  min  max  range  skew  kurtosis  ss
1  ss1      21  1.70  1.48  3    2.60  1.48  0  5    5 -0.16    -0.24  0.13
2  ss2      21  2.75  3.24  1.19  3    3.30  1.48  0  5    5 -0.34    -0.44  0.14
3  ss3      21  2.66  1.08  3    2.64  1.48  1  5    4  0.22    -0.79  0.13
4  ss4      21  3.23  1.11  3    3.27  1.48  1  5    4 -0.32    -0.84  0.13
5  ss5      21  2.41  0.89  2    2.36  1.48  1  5    4  0.49    0.52  0.11
6  ss6      21  2.40  1.05  2    2.32  1.48  1  5    4  0.46    -0.02  0.12
7  ss7      21  2.34  0.87  2    2.34  1.48  1  4    3 -0.55    -0.03  0.10
8  ss8      21  2.23  0.80  2    2.20  0.00  1  4    3  0.41    -0.22  0.10
9  ss9      21  3.37  1.17  3    3.38  1.48  1  5    4  0.01    -1.18  0.14
10 ss10     21  3.34  1.25  3    3.41  1.48  1  5    4 -0.22    -0.89  0.15
11 ss11     21  3.47  1.13  4    3.50  1.48  1  5    4 -0.20    -1.00  0.13
12 ss12     21  3.24  1.21  3    3.23  1.48  1  5    4  0.02    -1.27  0.14
```

74

Factor Analysis

```
R Console
>
>
> student_cor <- cor(student)
> student_cor
  ss1 ss2 ss3 ss4 ss5 ss6 ss7 ss8 ss9 ss10 ss11 ss12
ss1 1.000000 -0.2669509 0.3682194 -0.4172057 0.5630020 0.3667360 0.5058963 0.6968714 -0.3648597 -0.3921494 -0.3610664 -0.2996336
ss2 -0.2669509 1.000000 -0.2707700 0.4511641 -0.4572441 -0.5877323 -0.3948807 -0.3923232 0.3684047 0.4276402 -0.3887790 -0.4099930
ss3 0.3682194 -0.2707700 1.000000 -0.5424897 0.4974079 0.2467762 0.3303330 0.2960004 -0.3004938 -0.2240232 -0.2839463 -0.2471494
ss4 -0.4172057 0.4511641 -0.5424897 1.000000 -0.4007146 -0.2984249 -0.2947812 -0.3960132 0.4444487 -0.3985953 -0.2831134
ss5 0.5630020 -0.4572441 0.4974079 0.4007146 1.000000 0.4804895 0.4138104 0.4783309 -0.3578154 -0.3240579 -0.3223258
ss6 0.3667360 -0.3677323 0.2467762 -0.2984249 0.4804895 1.000000 0.3858004 0.3861466 -0.4392191 -0.3793360 -0.3534445 -0.5227912
ss7 0.5058963 -0.3948807 0.3303330 -0.2947812 0.4138104 0.3858004 1.000000 0.4502455 -0.2846777 -0.3319751 -0.3328264 -0.3360070
ss8 0.6968714 -0.3648597 0.2960004 -0.3960132 0.4783309 0.3861466 0.4502455 1.000000 -0.3547463 -0.2818753 -0.2977049 -0.2375259
ss9 -0.3985953 0.4444487 -0.3985953 0.4007146 -0.3240579 -0.3793360 -0.3319751 -0.3547463 1.000000 0.4950614 -0.7898814 -0.7249312
ss10 -0.3921494 0.4276402 -0.2240232 0.4144587 -0.3240579 -0.3793360 -0.3319751 -0.2818753 0.4950614 1.000000 0.6251351 0.5098674
ss11 -0.3610664 0.3887790 -0.2831134 0.3985953 -0.3223258 -0.3534445 -0.3328264 -0.2977049 0.7898814 0.6251351 1.000000 0.4387049
ss12 -0.2996336 0.4099930 -0.2471494 0.2831134 -0.3223258 -0.5227912 -0.3360070 -0.2375259 0.7249312 0.5098674 0.4387049 1.000000
```

75

Factor Analysis

```
R Console
>
>
> student_eigen <- eigen(student_cor)
> student_eigen$values
[1] 5.4386836 1.6560017 0.9329142 0.8094490 0.6725782 0.6567179 0.4513982
[8] 0.4156085 0.3588073 0.2842936 0.1781016 0.1454461
```

76

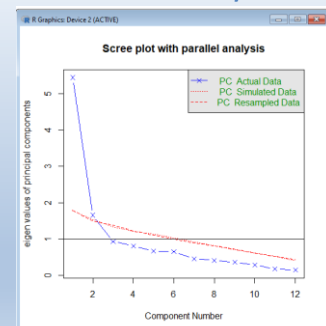
Factor Analysis

```
student.eigen$values
[1] 5.4386836 1.6560017 0.9329142 0.8094490 0.6725782 0.6567179 0.4513982
[8] 0.4156085 0.3588073 0.2842936 0.1781016 0.1454461
```

Tabachnick and Fidell (2013) recommend that researchers adopt an exploratory approach, experimenting with different numbers of factors until a satisfactory solution is found.

77

Factor Analysis



78

Factor Analysis

```

R Console
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 
```

79

79

Factor Analysis

```
R Console
```

```
>  
>  
>  
>  
R> cor.test(bartlett[student])  
R was not square, finding R from data  
Schlupf  
[1] 417.6944  
  
Op.value  
[1] 1.80092e-52  
  
Std  
[1] .66  
  
>  
>  
>
```

80

80

Factor Analysis

[illegible]

81

81

Factor Analysis

```

RStudio Console

> student.analysis <- principal(student, factor2=3, rotate="varimax")
> student.analysis
Principal Component Analysis

Call: principal(x = student, factor2 = 3, rotate = "varimax")

Standardized principal component analysis material based upon correlation matrix

    PC1    PC2    PC3    PC4
scl  -0.24    0.55    0.62    0.10
sei  -0.01   -0.45    0.46    0.20
sco  -0.16    0.42    0.41    0.39
s11  -0.32    0.42    0.49    0.51
s12  -0.22    0.40    0.42    0.52
s13  -0.50    0.48    0.45    0.58
s14  -0.38    0.70    0.49    0.51
s15  -0.16    0.78    0.43    0.57
s16  -0.89   -0.23    0.48    0.14
s17  -0.75   -0.24    0.43    0.17
s18  -0.05   -0.16    0.75    0.29
s19  -0.04   -0.12    0.71    0.30
s20  -0.04   -0.12    0.71    0.30

SS loadings          PC1    PC2    PC3
Fingerprint Var      3.94  3.64
Cumulative Var      0.30  0.29
Fingerprint Explained 0.30  0.50
Cumulative Proportion 0.30  1.00

Mean item complexity = 1.3
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMR) is 0.08
with the empirical chi square 59.55 with pval < 0.007

Fit based upon giv diagonal values = 0.94

```

82

82

Factor Analysis

	RC1	RC2	h2	u2	com
ss1	-0.24	0.70	0.55	0.45	1.2
ss2	0.51	-0.45	0.46	0.54	2.0
ss3	-0.16	0.62	0.41	0.59	1.1
ss4	0.32	-0.62	0.49	0.51	1.5
ss5	-0.22	0.80	0.68	0.32	1.2
ss6	-0.50	0.45	0.45	0.55	2.0
ss7	-0.08	0.70	0.49	0.51	1.0
ss8	-0.16	0.78	0.63	0.37	1.1
ss9	0.89	-0.23	0.84	0.16	1.1
ss10	0.75	-0.24	0.63	0.37	1.2
ss11	0.85	-0.16	0.75	0.25	1.1
ss12	0.84	-0.12	0.71	0.29	1.0

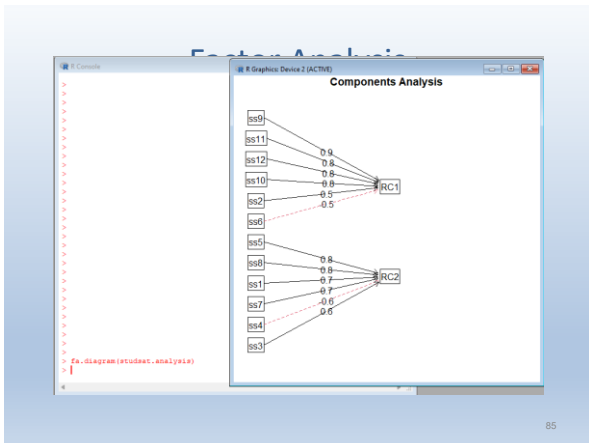
83

Factor Analysis

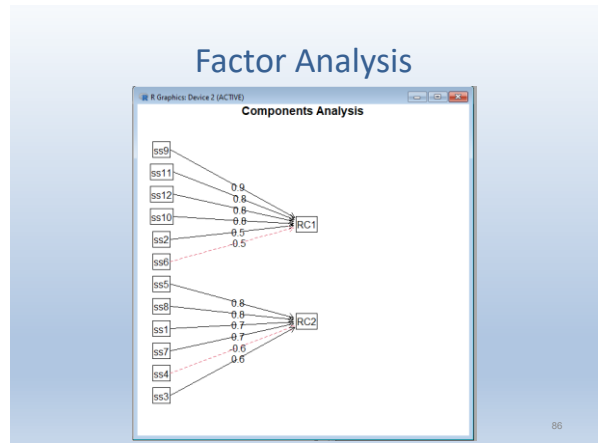
	RC1	RC2
SS loadings	3.56	3.54
Proportion Var	0.30	0.29
Cumulative Var	0.30	0.59
Proportion Explained	0.50	0.50
Cumulative Proportion	0.50	1.00

84

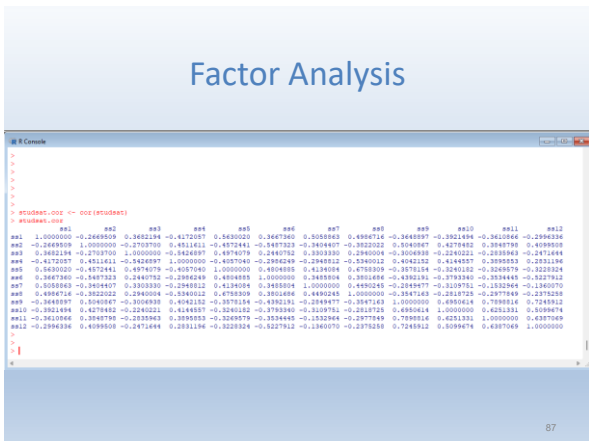
84



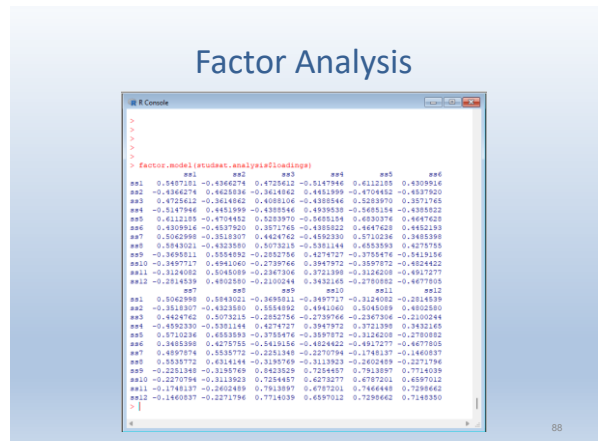
85



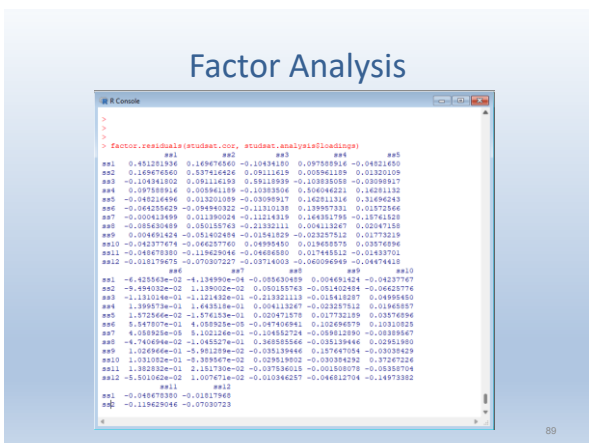
86



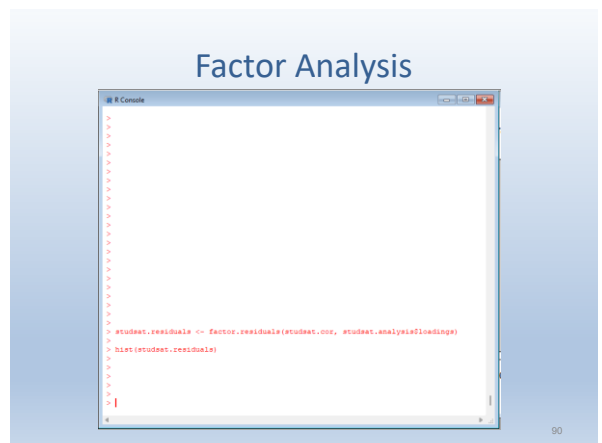
87



88

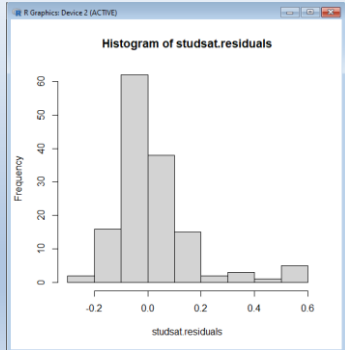


89



90

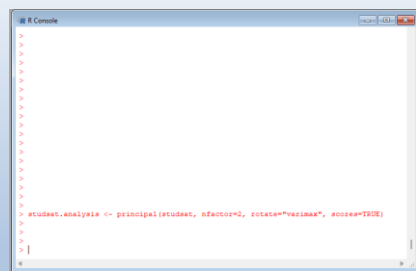
Factor Analysis



91

91

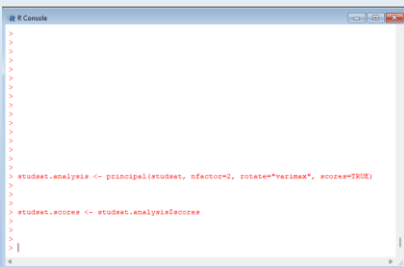
Factor Analysis



92

92

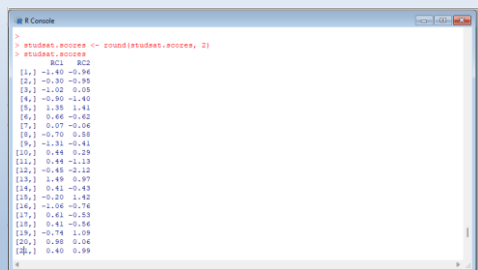
Factor Analysis



93

93

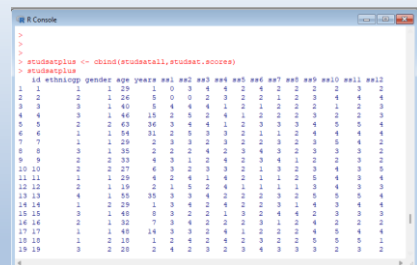
Factor Analysis



94

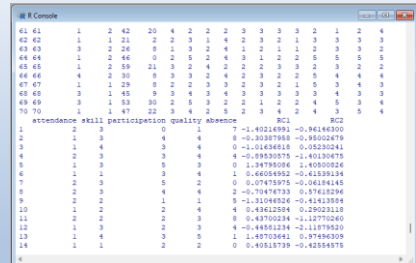
94

Factor Analysis



95

Factor Analysis



96

96

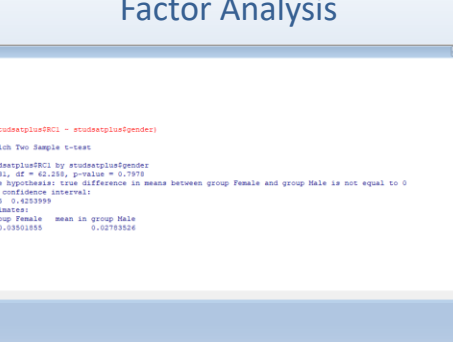
Factor Analysis

```
R Console  
>  
>  
>  
>  
>  
>  
>  
>  
>  
>  
> t.test(studstatpluseBC1 ~ studstatplusegender)  
  
Welch Two Sample t-test  
  
data: studstatpluseBC1 by studstatplusegender  
t = 0.29731, df = 62.286, p-value = 0.7978  
alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0  
95 percent confidence interval:  
-0.4219969 0.5111075  
sample estimates:  
mean in group 1 mean in group 2 [REDACTED]  
0.02789356 -0.03601888  
  
>  
<
```

97

97

Factor Analysis



```
R Console
>
>
>
>
>
> t.test(studentplusTBC1 ~ studentplusTgender)

Welch Two Sample t-test

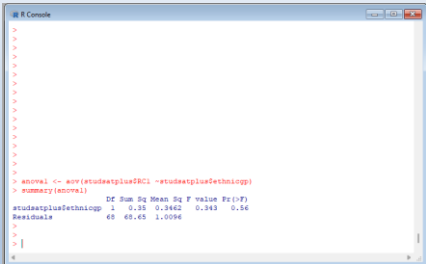
data: studentplusTBC1 by studentplusTgender
t = -0.20793, df = 62.286, p-value = 0.7976
alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
95 percent confidence interval:
 -0.5110778  0.4213989
sample estimates:
mean in group Female      mean in group Male
-0.03501855              0.02783526

>
>
>
> |
> |
4
```

98

98

Factor Analysis



```
R Console
```

```
> anova1 <- aov(studatplusf2C1 ~studatplusfethn1opp)
> summary(anova1)
```

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
studatplusfethn1opp	1	0.35	0.3462	0.343	0.56
Residuals	48	65.45	1.094		

```
> |
```

99

99

Factor Analysis

The 12 items of the Job Satisfaction Scale were subjected to factor analysis with PCA and varimax rotation using R version 4.3.2. Prior to performing the factor analysis, the suitability of factor for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of several coefficients of .3 and above. The Kaiser-Meyer-Olkin value was .80, exceeding the recommended value of .6 (Kaiser 1960) and Bartlett's Test of Sphericity (Bartlett 1954) reached statistical significance, supporting the factorability of the correlation matrix.

Factor analysis revealed the presence of two components with eigenvalues exceeding 1, explaining 30% & 29% of the variance respectively. An inspection of the screeplot revealed a clear after the second component. Using Cattell's (1966) scree test, it was decided to retain two components for further investigation.

The two solution explained a total of 59% of the variance, with Component 1 contributing 30% and Component 2 adding the other 29%. To aid in the interpretation of these three components, varimax rotation was performed. The rotated solution revealed the presence of simple structure, with all components showing a number of strong loadings and most variables loading substantially on only one component.

A test was conducted to see whether the factor model is sufficient to explain the observed data. The null hypothesis is that a 2-factor model is sufficient. For this model, the chi-square statistic is 58.55. The p-value for the chi-square test is 0.057 which is larger than .05. Therefore, we fail to reject the null hypothesis that the factor model have a good fit to the data.

100

References & Further Reading

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society* 16 (Series B), 296-8.

Cattell, R. B. (1966). The screw test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

Colman, A. M. & Pulford, B. D. (2008). *A crash course in SPSS for Windows* (4th ed.). Oxford: Blackwell

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

De Vries, A. & Meyers, J. (2015). *R for dummies* (2nd ed.). London: Wiley.

Field, A., Miles, J. & Field, Z. (2013). *Discovering statistics using R*. London: Sage.

Hatch, E. & Lazaratou, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.

Hayton, J. C., Allen, D. G. & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods* 7, 191-204.

Kabacoff, R. I. (2015). *R in action* (2nd ed.). Shelter Island, NY: Manning.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). London: Routledge.

Loewen, S. & Goumal, T. (2015). Exploratory factor analysis and principal component analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 182-212). London: Routledge.

Loewen, S., Laviolette, E., Spino, L., Papi, M., Schmidtke, J., Sterling, S., & Wolff, D. (2014). Statistical literacy among applied linguists and second language acquisition researchers. *TESOL Quarterly*, 48, 360-388.

Lowie, W. & Seton, B. (2013). *Essential statistics for applied linguistics*. London: Palgrave Macmillan.

Popham, W. J. & Sirotnik, K. A. (1973). *Educational statistics: Use and interpretation* (2nd ed.). New York: Harper & Row

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. (5th ed.). Boston: Allyn & Bacon.

101

101