#### **Data Exercise 7: Factor Analysis**

#### View the Tutorial:

· Factor Analysis: <a href="https://youtu.be/kmyZgwvMx6c">https://youtu.be/kmyZgwvMx6c</a>

#### **Directions**

The following exercises seek to determine what underlying structure exists among the following variables in *profile-a.sav*: highest degree earned (*degree*), hours worked per week (*hrs1*), job satisfaction (*satjob*), years of education (*educ*), hours per day watching TV (*tvhours*), general happiness (*happy*), degree to which life is exciting (*life*), and degree to which the lot of the average person is getting worse (*anomia5*).

- 1) The following output was generated for the initial analysis. Varimax rotation was utilized.
  - a. Assess the eigenvalue criterion. How many components were retained? Is the eigenvalue appropriate, considering the number of factors and the communalities?

Three components were retained: those that are listed as greater than 1.00 in the initial Eigenvalues column of the Table of Total Variances. Mertler and Reinhart (2016) suggest that the eigenvalue criterion is fairly reliable when either the number of variables is less than 30 and communalities are greater than .70, or, when the number of individuals is greater than 250 and their mean communality is greater than or equal to .60. This dataset does not meet the former situation, since the number of variables is less than 30 but not all of the communalities are greater than .70 – in fact, there are only two that are. In that case, eigenvalue criterion is questionable. However, even when removing outliers using Mahalanobis distance, the dataset contains greater than 250 individuals and the mean communality is approximately .61. For that situation, the eigenvalue seems appropriate.

b. Assess the variance explained by the retained components. What is the total variability explained by the model? Is this amount adequate?

For variance, it is advised to retain components that account for at least 70% of total variability. However, the total amount of variability as seen by the initial model is 60.875%, which is lower than the recommended 70% and therefore not adequate. If using this criterion alone, a researcher may consider the inclusion of a fourth component, which would put the total variability explained by the model at 71.952%.

#### c. Assess the scree plot. At which component does the plot begin to level off?

It appears that at around either the third or fourth component the plot begins to level off, suggesting that at least the first two components should be retained, if not the third.

#### d. Assess the residuals. How many residuals exceed the 0.05 criterion?

Ten of the twenty-eight residuals (around 36%) of the residuals exceed the 0.05 criterion.

e. Having applied the four criteria, do you believe the number of components retained in this analysis is appropriate? If not, what is your recommendation?

I would consider adding a fourth component to the model to reevaluate the criteria. Inclusion of the fourth component will allow the model to meet the variance criterion and will likely reduce the number of residuals that exceed the 0.05 criterion.

- 2) Assume that you believe four components should be retained from the analysis in the previous exercise. Conduct a factor analysis with varimax rotation (be sure to retain four components). (\*\*Refer to SPSS Output with the subheading "Factor Analysis With 4 components --)
  - a. Evaluate each of the four criteria. Has the model fit improved? Explain.

With the four-component model:

- 1) Eigenvalue = The communalities have all increased compared to the three-component model provided by Mertler and Reinhart (2016). Still, there are five communalities that remain under the .70 commonly accepted threshold, rendering the eigenvalue criterion questionable. However, when considering the optional case where one has greater than 250 individuals and the mean communality is greater than or equal to .60, the eigenvalue criterion has improved from the three-component model (with a greater mean communality) and is appropriate.
- 2) Variance = The inclusion of four components now indicates that after rotation,

  Component 1 accounts for 23.58%, Component 2 accounts for 20.10%, Component 3

  accounts for 15.38%, and Component 4 for 12.90% of the total variability accounted for by the model. This totals 71.95%, meeting the variance criterion.
- 3) Scree Plot = This is the same, where the plot indicates at least the first two components should be retained (the third one possibly as well).
- 4) Residuals = Indicate that the same amount of residuals are greater than .05 (10/28 or approx 36%). However, the residuals are closer to .05 and thus show improvement.

#### b. Provide two alternatives for improving the model.

- 3.1
- 1) Another component may be added to the model. This again improves all four criteria (\*\*Refer to SPSS output with subheading Factor Analysis With 5 components -- )
- 2) The data could be screened for missing data, outliers, and assumptions of normality and linearity, if not already performed. As Mertler and Reinhart (2016) suggest, both assumptions should be evaluated, and any necessary transformations be made because

"ensuring the quality of data will only improve the quality of the resulting factor or component solution."

#### Communalities

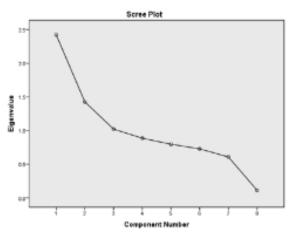
	Initia1	Extraction
degree	1.000	.933
hrs1	1.000	.602
satjob	1.000	.447
educ	1.000	.939
tvhours	1.000	556
happy	1.000	576
life	1.000	.500
anomia5	1.000	.317

Extraction Method: Principal Component Analysis.

Total Variance Explained

		Initial Eigenvalu	ies	Extraction Sums of Squared Loadings		Rotation Sums of Squared Loadings			
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.423	30.293	30.293	2.423	30.293	30.293	1.879	23.488	23.488
2	1.426	17.822	48.115	1.426	17.822	48.115	1.734	21.676	45.165
3	1.021	12.760	60.875	1.021	12.760	60.875	1.257	15.710	60.875
4	.816	11.077	71.952						
5	.796	9.955	81.907						
6	.728	9.094	91.001						
7	.607	7.589	98.590						
8	.113	1.410	100.000						

Extraction Method: Principal Component Analysis.



Reproduced Correlations

		degree	hrs1	satjob	educ	Mhours	happy	life	anomia5
Reproduced Correlation	degree	.933ª	.176	039	.935	239	119	.230	.118
	hrs1	.176	.602ª	239	.194	576	077	.141	049
	satjob	039	239	.447*	062	.214	.469	436	297
	educ	.935	.194	062	.939ª	255	142	.252	.131
	tyhours	239	576	.214	255	.556ª	.066	136	.047
	happy	119	077	.469	142	.066	.576ª	526	412
	life	.230	.141	436	.252	136	526	.500*	.371
	anomia5	.118	049	297	.131	.047	412	.371	.317
Residual <sup>b</sup>	degree		.004	068	050	.032	004	034	037
	hrs1	.004		.104	.011	.361	031	046	.112
	satjob	068	.104		037	105	197	.151	.158
	educ	050	.011	037		.026	002	029	012
	tyhours	.032	.361	105	.026		.014	012	099
	happy	004	031	197	002	.014		.159	.177
	life	034	046	.151	029	012	.159		217
	anomia5	037	.112	.158	012	099	.177	217	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

Residuals are computed between observed and reproduced correlations. There are 12 (42.0%) nonredundant residuals with absolute values greater than 0.05.

#### What to Turn In:

- 1) Submit the SPSS output you created for this exercise as a single PDF document (not multiple). You may want to save the output files from SPSS to your own storage device, in case you need to make changes (PDF documents cannot be edited). In the output window in SPSS, you can use the File>Export feature to create the PDF.
- 2) In addition to the output, you must submit a written report that details what steps were taken for the analysis, providing a rationale where needed, and then you must provide any necessary interpretations of the analysis. The write up should be prepared and saved as a single Word document (.docx) with a 12 point font, in Times New Roman, and it should be double spaced with one inch margins.

## Reference

Mertler, Craig A. and Rachel V. Reinhart. Advanced and Multivariate Statistical Methods, 6th Edition. Routledge, 20161024. VitalBook file.

#### FACTOR

/VARIABLES degree hrs1 satjob educ tvhours happy life anomia5
/MISSING LISTWISE

/ANALYSIS degree hrs1 satjob educ tvhours happy life anomia5
/PRINT INITIAL REPR EXTRACTION ROTATION
/FORMAT SORT
/PLOT EIGEN
/CRITERIA MINEIGEN(.885) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)

/ROTATION VARIMAX

/SAVE REG(ALL)

/METHOD=CORRELATION.

# Factor Analysis -- With 4 components --

#### **Notes**

Output Created		08-DEC-2018 09:08:28
Comments		
Input	Data	C: \Users\linds\AppData\Loca I\Packages\Microsoft. MicrosoftEdge_8wekyb3d 8bbwe\TempState\Downlo ads\profile-a (1).sav
	Active Dataset	DataSet4
	Filter	<none></none>
	Weight	<none></none>
	Split File	<none></none>
	N of Rows in Working Data File	1500
Missing Value Handling	Definition of Missing	MISSING=EXCLUDE: User-defined missing values are treated as missing.
	Cases Used	LISTWISE: Statistics are based on cases with no missing values for any variable used.

## Notes

Syntax		FACTOR /VARIABLES degree hrs1 satjob educ tvhours happy life anomia5 /MISSING LISTWISE /ANALYSIS degree hrs1 satjob educ tvhours happy life anomia5 /PRINT INITIAL REPR EXTRACTION ROTATION /FORMAT SORT /PLOT EIGEN /CRITERIA MINEIGEN(. 885) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /SAVE REG(ALL) /METHOD=CORRELATIO N.
Resources	Processor Time	00:00:00.13
	Elapsed Time	00:00:00.11
	Maximum Memory Required	9904 (9.672K) bytes
Variables Created	FAC1_2	Component score 1
	FAC2_2	Component score 2
	FAC3_2	Component score 3
	FAC4_2	Component score 4

# Communalities (Four-Component Model)

	Initial	Extraction
RS Highest Degree	1.000	.935
Number of Hours Worked Last Week	1.000	.659
Job Satisfaction	1.000	.537
Highest Year of School Completed	1.000	.939
HOURS PER DAY WATCHING TV	1.000	.558
GENERAL HAPPINESS	1.000	.578
Is life exciting or dull	1.000	.610
Lot of average man getting worse	1.000	.940

Extraction Method: Principal Component Analysis.

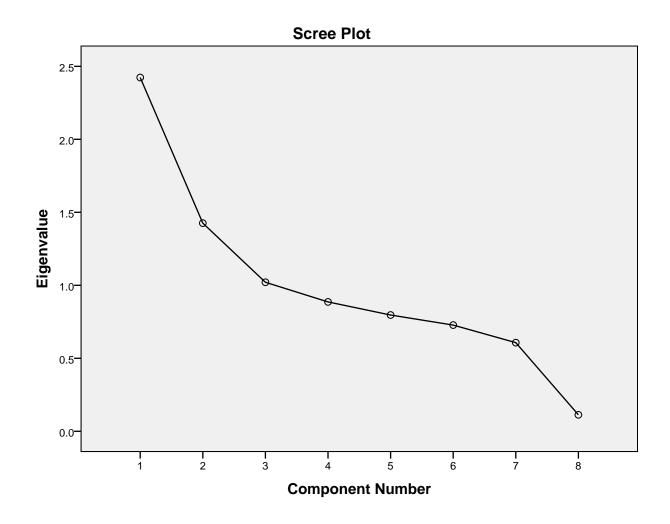
# **Total Variance Explained**

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.423	30.293	30.293	2.423	30.293	30.293
2	1.426	17.822	48.115	1.426	17.822	48.115
3	1.021	12.760	60.875	1.021	12.760	60.875
4	.886	11.077	71.952	.886	11.077	71.952
5	.796	9.955	81.907			
6	.728	9.094	91.001			
7	.607	7.589	98.590			
8	.113	1.410	100.000			

**Total Variance Explained** 

	Rotation Sums of Squared Loadings			
Component	Total	% of Variance	Cumulative %	
1	1.886	23.579	23.579	
2	1.608	20.098	43.677	
3	1.230	15.377	59.054	
4	1.032	12.898	71.952	
5				
6				
7				
8				

Extraction Method: Principal Component Analysis.



# Component Matrix<sup>a</sup>

		Component			
	1	2	3	4	
Highest Year of School Completed	.805	.487	.232	009	
RS Highest Degree	.782	.511	.246	047	
Is life exciting or dull	.548	436	.099	332	
GENERAL HAPPINESS	480	.570	145	.043	
Job Satisfaction	420	.502	.133	.300	
Number of Hours Worked Last Week	.407	.040	660	.239	
HOURS PER DAY WATCHING TV	433	095	.600	046	
Lot of average man getting worse	.324	386	.251	.789	

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

		RS Highest Degree	Number of Hours Worked Last Week	Job Satisfaction
Reproduced Correlation	RS Highest Degree	.935 <sup>a</sup>	.165	053
	Number of Hours Worked Last Week	.165	.659 <sup>a</sup>	167
	Job Satisfaction	053	167	.537 <sup>a</sup>
	Highest Year of School Completed	.936	.191	065
	HOURS PER DAY WATCHING TV	237	586	.200
	GENERAL HAPPINESS	121	067	.481
	Is life exciting or dull	.246	.061	536
	Lot of average man getting worse	.080	.140	060
Residual <sup>b</sup>	RS Highest Degree		.016	054
	Number of Hours Worked Last Week	.016		.032
	Job Satisfaction	054	.032	

		Highest Year of School Completed	HOURS PER DAY WATCHING TV	GENERAL HAPPINESS
Reproduced Correlation	RS Highest Degree	.936	237	121
	Number of Hours Worked Last Week	.191	586	067
	Job Satisfaction	065	.200	.481
	Highest Year of School Completed	.939 <sup>a</sup>	255	142
	HOURS PER DAY WATCHING TV	255	.558 <sup>a</sup>	.064
	GENERAL HAPPINESS	142	.064	.578 <sup>a</sup>
	Is life exciting or dull	.255	121	540
	Lot of average man getting worse	.124	.011	378
Residual <sup>b</sup>	RS Highest Degree	050	.030	002
	Number of Hours Worked Last Week	.013	.372	042
	Job Satisfaction	035	091	209

		Is life exciting or dull	Lot of average man getting worse
Reproduced Correlation	RS Highest Degree	.246	.080
	Number of Hours Worked Last Week	.061	.140
	Job Satisfaction	536	060
	Highest Year of School Completed	.255	.124
	HOURS PER DAY WATCHING TV	121	.011
	GENERAL HAPPINESS	540	378
	Is life exciting or dull	.610 <sup>a</sup>	.109
	Lot of average man getting worse	.109	.940 <sup>a</sup>
Residual <sup>b</sup>	RS Highest Degree	050	-2.048E-5
	Number of Hours Worked Last Week	.033	077
	Job Satisfaction	.250	078

	RS Highest Degree	Number of Hours Worked Last Week	Job Satisfaction
Highest Year of School Completed	050	.013	035
HOURS PER DAY WATCHING TV	.030	.372	091
GENERAL HAPPINESS	002	042	209
Is life exciting or dull	050	.033	.250
Lot of average man getting worse	-2.048E-5	077	078

	Highest Year of School Completed	HOURS PER DAY WATCHING TV	GENERAL HAPPINESS
Highest Year of School Completed	ol	.026	002
HOURS PER DAY WATCHING TV	.026		.016
GENERAL HAPPINE	002	.016	
Is life exciting or dull	032	027	.173
Lot of average man go worse	etting005	063	.143

## **Reproduced Correlations**

	Is life exciting or dull	Lot of average man getting worse
Highest Year of School Completed	032	005
HOURS PER DAY WATCHING TV	027	063
GENERAL HAPPINESS	.173	.143
Is life exciting or dull		.044
Lot of average man getting worse	.044	

Extraction Method: Principal Component Analysis.

- a. Reproduced communalities
- b. Residuals are computed between observed and reproduced correlations. There are 11 (39.0%) nonredundant residuals with absolute values greater than 0.05.

## **Rotated Component Matrix**<sup>a</sup>

	Component				
	1	2	3	4	
RS Highest Degree	.956	087	.118	.017	
Highest Year of School Completed	.951	099	.146	.059	
Is life exciting or dull	.186	758	.018	.006	
Job Satisfaction	.030	.710	176	.033	
GENERAL HAPPINESS	059	.695	.000	302	
Number of Hours Worked Last Week	.067	045	.801	.105	
HOURS PER DAY WATCHING TV	150	.107	721	.062	
Lot of average man getting worse	.052	122	.038	.960	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

#### **Component Transformation Matrix**

Component	1	2	3	4
1	.718	536	.389	.215
2	.595	.720	.085	346
3	.358	061	892	.270
4	051	.437	.213	.872

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

#### FACTOR

/VARIABLES degree hrs1 satjob educ tvhours happy life anomia5 /MISSING LISTWISE

/ANALYSIS degree hrs1 satjob educ tvhours happy life anomia5

/PRINT INITIAL REPR EXTRACTION ROTATION

/FORMAT SORT

/PLOT EIGEN

/CRITERIA MINEIGEN(.795) ITERATE(25)

/EXTRACTION PC

/CRITERIA ITERATE(25)

a. Rotation converged in 5 iterations.

# Factor Analysis -- With 5 components --

## Notes

Output Created		08-DEC-2018 09:29:11
Comments		
Input	Data	C: \Users\linds\AppData\Loca I\Packages\Microsoft. MicrosoftEdge_8wekyb3d 8bbwe\TempState\Downlo ads\profile-a (1).sav
	Active Dataset	DataSet4
	Filter	<none></none>
	Weight	<none></none>
	Split File	<none></none>
	N of Rows in Working Data File	1500
Missing Value Handling	Definition of Missing	MISSING=EXCLUDE: User-defined missing values are treated as missing.
	Cases Used	LISTWISE: Statistics are based on cases with no missing values for any variable used.

## Notes

Syntax		FACTOR /VARIABLES degree hrs1 satjob educ tvhours happy life anomia5 /MISSING LISTWISE /ANALYSIS degree hrs1 satjob educ tvhours happy life anomia5 /PRINT INITIAL REPR EXTRACTION ROTATION /FORMAT SORT /PLOT EIGEN /CRITERIA MINEIGEN(. 795) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /SAVE REG(ALL) /METHOD=CORRELATIO N.
Resources	Processor Time	00:00:00.13
	Elapsed Time	00:00:00.11
	Maximum Memory Required	9904 (9.672K) bytes
Variables Created	FAC1_3	Component score 1
	FAC2_3	Component score 2
	FAC3_3	Component score 3
	FAC4_3	Component score 4
	FAC5_3	Component score 5

## Communalities

	Initial	Extraction
RS Highest Degree	1.000	.939
Number of Hours Worked Last Week	1.000	.929
Job Satisfaction	1.000	.597
Highest Year of School Completed	1.000	.941
HOURS PER DAY WATCHING TV	1.000	.977
GENERAL HAPPINESS	1.000	.578
Is life exciting or dull	1.000	.642
Lot of average man getting worse	1.000	.949

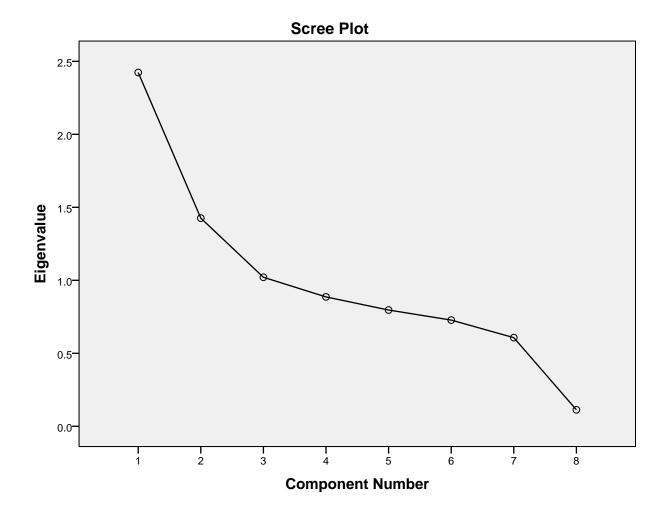
Extraction Method: Principal Component Analysis.

# **Total Variance Explained**

	Initial Eigenvalues			Extraction	on Sums of Square	ed Loadings
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.423	30.293	30.293	2.423	30.293	30.293
2	1.426	17.822	48.115	1.426	17.822	48.115
3	1.021	12.760	60.875	1.021	12.760	60.875
4	.886	11.077	71.952	.886	11.077	71.952
5	.796	9.955	81.907	.796	9.955	81.907
6	.728	9.094	91.001			
7	.607	7.589	98.590			
8	.113	1.410	100.000			

## **Total Variance Explained**

	Rotation Sums of Squared Loadings				
Component	Total	% of Variance	Cumulative %		
1	1.901	23.761	23.761		
2	1.608	20.101	43.863		
3	1.028	12.852	56.714		
4	1.010	12.626	69.340		
5	1.005	12.567	81.907		
6					
7					
8					



# Component Matrix<sup>a</sup>

	Component				
	1	2	3	4	5
Highest Year of School Completed	.805	.487	.232	009	.048
RS Highest Degree	.782	.511	.246	047	.065
Is life exciting or dull	.548	436	.099	332	180
GENERAL HAPPINESS	480	.570	145	.043	013
Job Satisfaction	420	.502	.133	.300	246
Number of Hours Worked Last Week	.407	.040	660	.239	.519
Lot of average man getting worse	.324	386	.251	.789	097
HOURS PER DAY WATCHING TV	433	095	.600	046	.647

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

		RS Highest Degree	Number of Hours Worked Last Week	Job Satisfaction
Reproduced Correlation	RS Highest Degree	.939 <sup>a</sup>	.198	069
	Number of Hours Worked Last Week	.198	.929 <sup>a</sup>	295
	Job Satisfaction	069	295	.597 <sup>a</sup>
	Highest Year of School Completed	.939	.216	077
	HOURS PER DAY WATCHING TV	195	251	.041
	GENERAL HAPPINESS	122	073	.485
	Is life exciting or dull	.234	032	491
	Lot of average man getting worse	.074	.090	036
Residual <sup>b</sup>	RS Highest Degree		018	038
	Number of Hours Worked Last Week	018		.160
	Job Satisfaction	038	.160	

		Highest Year of School Completed	HOURS PER DAY WATCHING TV	GENERAL HAPPINESS
Reproduced Correlation	RS Highest Degree	.939	195	122
	Number of Hours Worked Last Week	.216	251	073
	Job Satisfaction	077	.041	.485
	Highest Year of School Completed	.941 <sup>a</sup>	224	143
	HOURS PER DAY WATCHING TV	224	.977 <sup>a</sup>	.056
	GENERAL HAPPINESS	143	.056	.578 <sup>a</sup>
	Is life exciting or dull	.246	238	538
	Lot of average man getting worse	.119	051	377
Residual <sup>b</sup>	RS Highest Degree	053	012	001
	Number of Hours Worked Last Week	012	.037	035
	Job Satisfaction	023	.068	212

		Is life exciting or dull	Lot of average man getting worse
Reproduced Correlation	RS Highest Degree	.234	.074
	Number of Hours Worked Last Week	032	.090
	Job Satisfaction	491	036
	Highest Year of School Completed	.246	.119
	HOURS PER DAY WATCHING TV	238	051
	GENERAL HAPPINESS	538	377
	Is life exciting or dull	.642 <sup>a</sup>	.126
	Lot of average man getting worse	.126	.949 <sup>a</sup>
Residual <sup>b</sup>	RS Highest Degree	038	.006
	Number of Hours Worked Last Week	.127	027
	Job Satisfaction	.206	102

	RS Highest Degree	Number of Hours Worked Last Week	Job Satisfaction
Highest Year of School Completed	053	012	023
HOURS PER DAY WATCHING TV	012	.037	.068
GENERAL HAPPINESS	001	035	212
Is life exciting or dull	038	.127	.206
Lot of average man getting worse	.006	027	102

	Highest Year of School Completed	HOURS PER DAY WATCHING TV	GENERAL HAPPINESS
Highest Year of School Completed		005	001
HOURS PER DAY WATCHING TV	005		.024
GENERAL HAPPINESS	001	.024	
Is life exciting or dull	024	.089	.171
Lot of average man gett worse	.000	.000	.142

## **Reproduced Correlations**

	Is life exciting or dull	Lot of average man getting worse
Highest Year of School Completed	024	.000
HOURS PER DAY WATCHING TV	.089	.000
GENERAL HAPPINESS	.171	.142
Is life exciting or dull		.027
Lot of average man getting worse	.027	

Extraction Method: Principal Component Analysis.

- a. Reproduced communalities
- b. Residuals are computed between observed and reproduced correlations. There are 10 (35.0%) nonredundant residuals with absolute values greater than 0.05.

# **Rotated Component Matrix**<sup>a</sup>

	Component				
	1	2	3	4	5
RS Highest Degree	.960	087	.014	.068	066
Highest Year of School Completed	.955	098	.058	.082	093
Is life exciting or dull	.173	749	.027	135	182
Job Satisfaction	.009	.720	.065	270	044
GENERAL HAPPINESS	060	.697	297	008	009
Lot of average man getting worse	.048	126	.964	.036	017
Number of Hours Worked Last Week	.124	068	.041	.945	121
HOURS PER DAY WATCHING TV	121	.074	016	119	.971

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

#### **Component Transformation Matrix**

Component	1	2	3	4	5
1	.725	534	.216	.252	280
2	.595	.724	338	.020	082
3	.336	064	.271	684	.585
4	043	.429	.867	.247	040
5	.072	047	120	.638	.755

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

# Index of comments

- 2.1 should be 12
- 3.1 this will not change the underlying problem of communalities/scree plot; might add or remove variables
- 3.2 yes but screening should be a given; you might try another rotation method