Principal Component Analysis

Karen Grace-Martin

How this Webinar Will Work:

- ~40 minutes of teaching
- Questions answered at the very end
- The slides handout is available for download on the GoToWebinar panel
- Recording will be sent via email within 48 hours

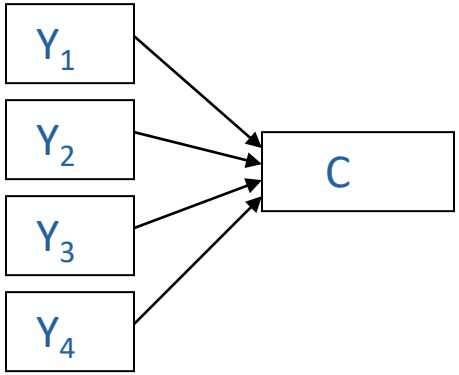
What You Will Learn:

1. What Principal Component Analysis Does
2. The Steps to Conduct a PCA

    1. Extract the Components
    2. Determine the Number of Components to Retain
    3. Rotate
    4. Interpret the Rotated Solution
    5. Create Component Scores

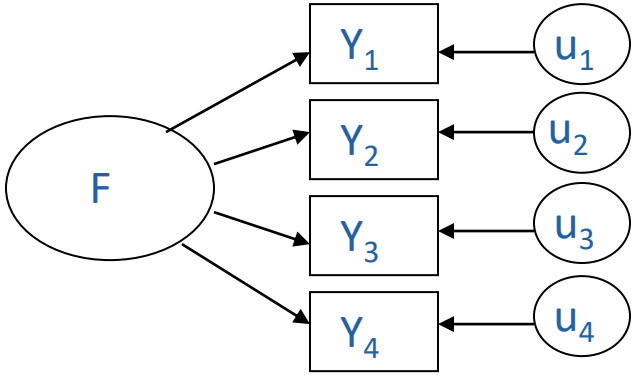
What is PCA?

Principal Component Analysis



$$C = w_1(Y_1) + w_2(Y_2) + w_3(Y_3) + w_4(Y_4)$$

Factor Analysis



$$Y_1 = b_1*F + u_1$$
$$Y_2 = b_2*F + u_2$$
$$Y_3 = b_3*F + u_3$$
$$Y_4 = b_4*F + u_4$$

A Motivating Example

weight	body weight in kg
brainwt	brain weight in g
slosleep	slow wave ("nondreaming") sleep (hrs/day)
parsleep	paradoxical ("dreaming") sleep (hrs/day)
totsleep	total sleep (hrs/day) (sum of slow wave and paradoxical sleep)
lifespan	maximum life span (years)
gestatn	gestation time (days)
predatn	predation index (1-5)
expos	sleep exposure index (1-5)
danger	overall danger index (1-5)

Animal Sleep Example

Correlation Matrix

		weight	brainwt	slosleep	parsleep	totsleep	lifespan	gestatn	predatn	expos	danger
Correlation	weight	1.000	.956	-.394	-.075	-.343	.470	.714	.096	.406	.259
	brainwt	.956	1.000	-.387	-.074	-.337	.629	.734	-.015	.323	.151
	slosleep	-.394	-.387	1.000	.518	.968	-.372	-.606	-.353	-.580	-.535
	parsleep	-.075	-.074	.518	1.000	.717	-.268	-.409	-.398	-.504	-.572
	totsleep	-.343	-.337	.968	.717	1.000	-.382	-.614	-.405	-.621	-.604
	lifespan	.470	.629	-.372	-.268	-.382	1.000	.646	-.170	.316	.015
	gestatn	.714	.734	-.606	-.409	-.614	.646	1.000	.091	.573	.306
	predatn	.096	-.015	-.353	-.398	-.405	-.170	.091	1.000	.626	.927
	expos	.406	.323	-.580	-.504	-.621	.316	.573	.626	1.000	.790
	danger	.259	.151	-.535	-.572	-.604	.015	.306	.927	.790	1.000

The Steps

1. Extract the Components
2. Determine the Number of Components to Retain
3. Rotate
4. Interpret the Rotated Solution
5. Create Component Scores

Step 1: Extract the Components

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	5.104	51.039	51.039
2	2.381	23.815	74.854
3	1.086	10.857	85.711
4	.565	5.646	91.357
5	.384	3.844	95.201
6	.287	2.868	98.069
7	.151	1.505	99.574
8	.025	.253	99.828
9	.017	.172	100.000
10	-1.665E-16	-1.665E-15	100.000

Extraction Method: Principal Component Analysis.

Step 2: Determine the Number of Components to Retain

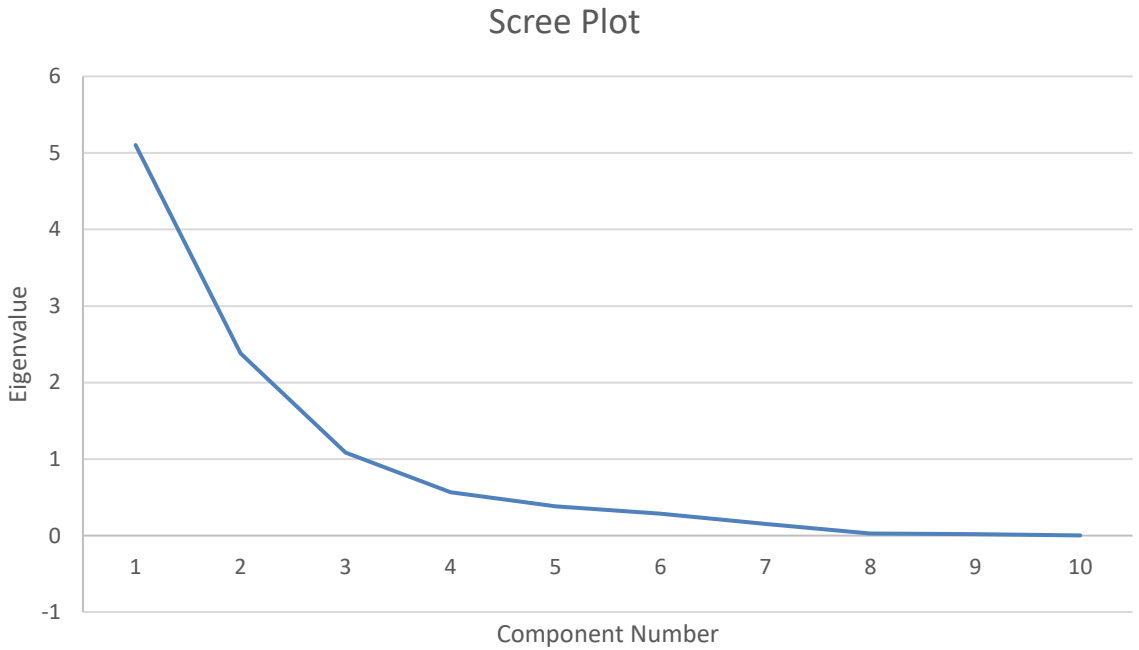
1. Kaiser Criterion: Eigenvalues > 1
2. Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	5.104	51.039	51.039
2	2.381	23.815	74.854
3	1.086	10.857	85.711
4	.565	5.646	91.357
5	.384	3.844	95.201
6	.287	2.868	98.069
7	.151	1.505	99.574
8	.025	.253	99.828
9	.017	.172	100.000
10	-1.665E-16	-1.665E-15	100.000

Extraction Method: Principal Component Analysis.

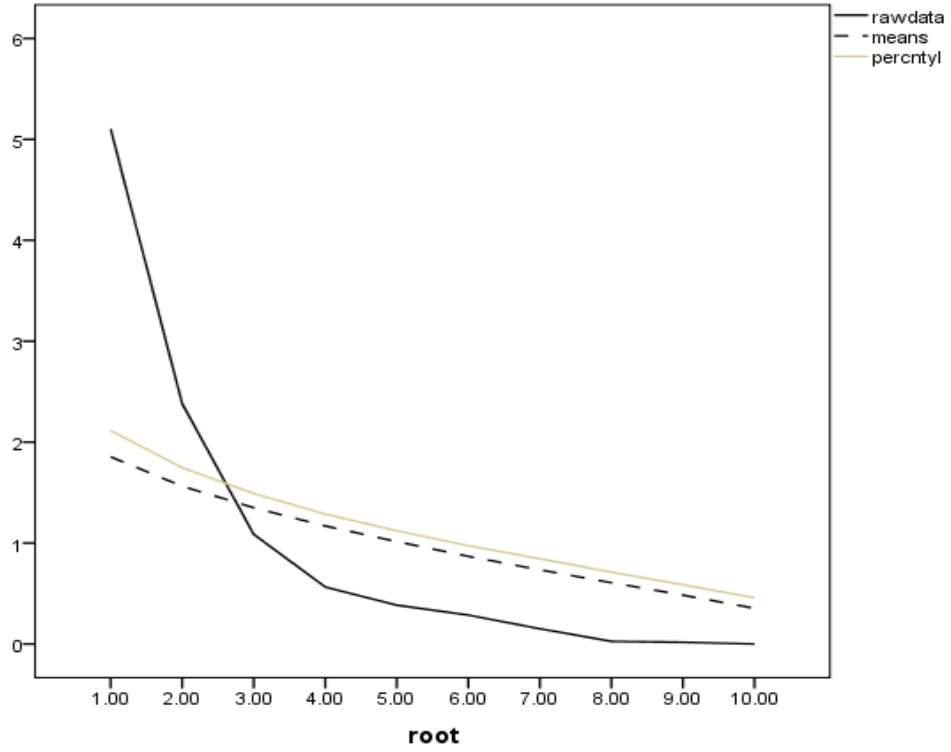
Step 2: Determine the Number of Components to Retain

3. Scree Plot



Step 2: Determine the Number of Components to Retain

4. Parallel Analysis



Step 2: Determine the Number of Components to Retain

5. Interpretability of Rotated Components

Component Matrix[a]

	Component	
	1	2
weight	.653	.566
brainwt	.630	.680
slosleep	-.832	.079
parsleep	-.648	.357
totsleep	-.869	.169
lifespan	.542	.583
gestatn	.809	.423
predatn	.517	-.703
expos	.817	-.264
danger	.728	-.595

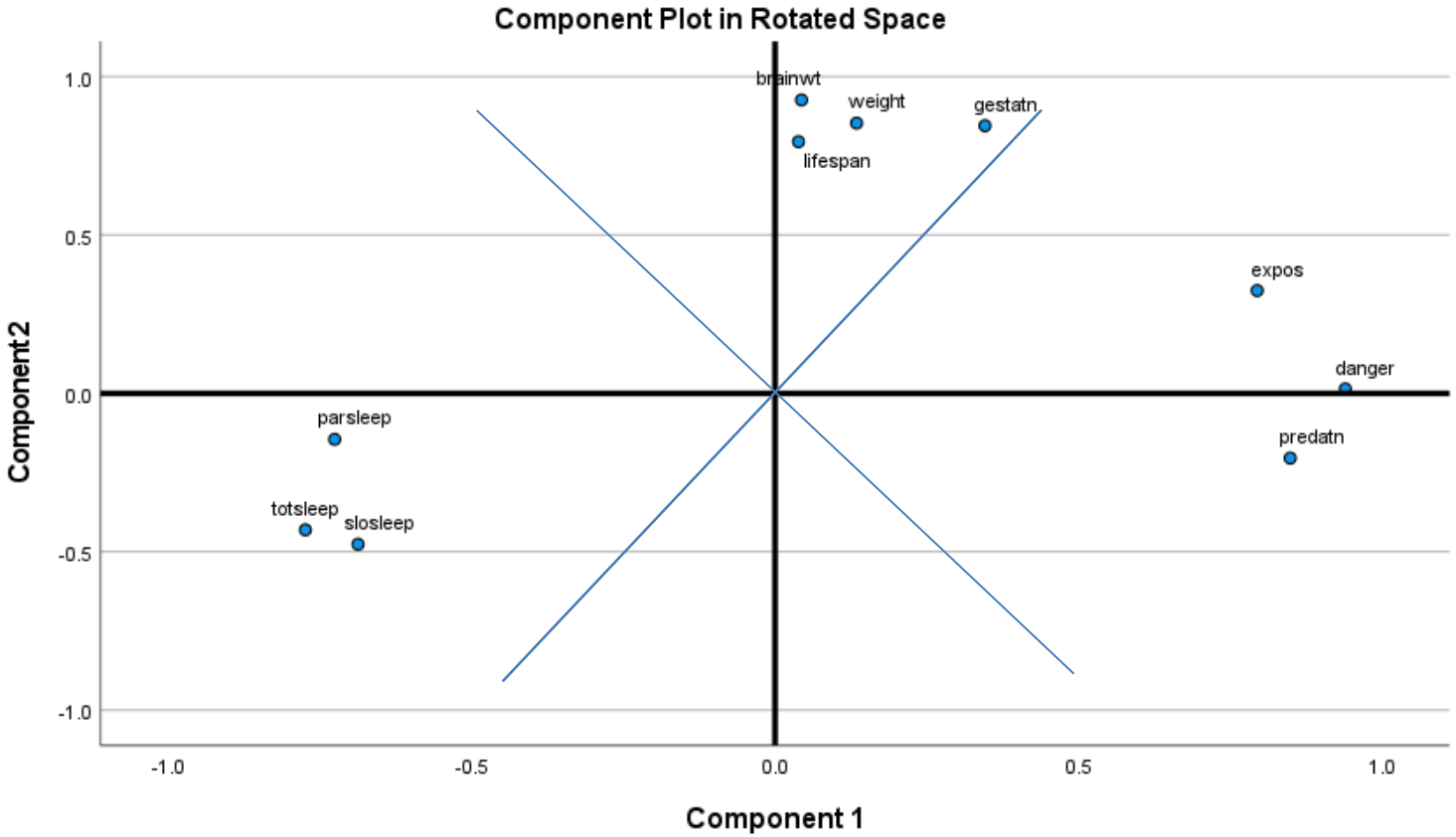
Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component	
	1	2
weight	.134	.853
brainwt	.044	.926
slosleep	-.687	-.476
parsleep	-.726	-.145
totsleep	-.774	-.431
lifespan	.039	.795
gestatn	.346	.845
predatn	.849	-.204
expos	.794	.325
danger	.940	.014

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

Step 3: Rotate



Step 4: Interpret Rotated Component Loadings

Component Matrix[a]

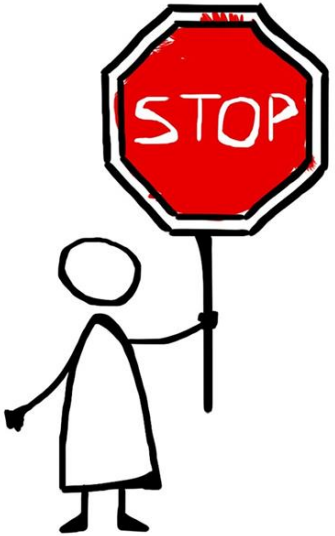
	Component	
	1	2
weight	.653	.566
brainwt	.630	.680
slosleep	-.832	.079
parsleep	-.648	.357
totsleep	-.869	.169
lifespan	.542	.583
gestatn	.809	.423
predatn	.517	-.703
expos	.817	-.264
danger	.728	-.595

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component	
	1	2
weight	.134	.853
brainwt	.044	.926
slosleep	-.687	-.476
parsleep	-.726	-.145
totsleep	-.774	-.431
lifespan	.039	.795
gestatn	.346	.845
predatn	.849	-.204
expos	.794	.325
danger	.940	.014

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



Step 5: Create Component Scores

Component Scores are weighted averages for each individual

Rotated Component Matrix[a]

	Component 1	Component 2
weight	.134	.853
brainwt	.044	.926
slosleep	-.687	-.476
parsleep	-.726	-.145
totsleep	-.774	-.431
lifespan	.039	.795
gestatn	.346	.845
predatn	.849	-.204
expos	.794	.325
danger	.940	.014

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

Component Score Coefficient Matrix

	Component 1	Component 2
weight	-.055	.264
brainwt	-.090	.298
slosleep	-.146	-.080
parsleep	-.194	.033
totsleep	-.176	-.055
lifespan	-.077	.255
gestatn	.007	.238
predatn	.268	-.160
expos	.194	.018
danger	.270	-.099

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

References and Further Reading

- Hatcher, L. (1994). A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling. SAS Publishing. Chapter 1: Principal Component Analysis.

- Wang, Yun. (2002). Principal Components: Not Just Another Factor Analysis.
  https://cscu.cornell.edu/wp-content/uploads/49_pca.pdf

Resources Page: https://www.theanalysisfactor.com/resources/by-topic/pca-fa/

Questions?