Comments for MEDB 5502, Week 05

Topics to be covered

- What you will learn
  - Principal components analysis
  - Applications of principal components
  - Factor analysis
  - Criticisms of principal components analysis and factor analysis

What is a linear combination?

- Multiply each variable by a constant and add
- Examples
  - $X_1 + X_2 + X_3 + X_4$
  - $\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + \frac{1}{4}X_4$
  - $\frac{1}{2} \times X_1 + 0X_2 + 0X_3 + \frac{1}{2}X_4$
  - $\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 - 1 \times X_4$
  - $\frac{4}{10}X_1 + \frac{4}{10}X_2 + \frac{1}{10}X_3 + \frac{1}{10}X_4$
  - $9X_1 + 3X_2 + 5X_3 - X_4$

Speaker notes

The simplest linear combination is just a sum. Multiply each variable by one and add the values together.

Almost as simple is an average. With four variables, multiply each variable by one fourth and add the values together.

A linear combination could effectively exclude one or more of the variables just by multiplying them by zero. The third linear combination is an average of just the first and last variables.

You can put negative numbers in the linear combination. The fourth linear combination computes the average of the first three variables and subtracts the last variable.

The fifth linear combination is a weighted average. The first two variables are weighted four times as much as the last two variables.

A linear combination does not necessarily have to involve averages. Taking 9 times the first variable plus 3 times the second variable plus 5 times the third variable minus the fourth variable.

Not a linear combination?

- $X_1 \times X_2 + X_3 + X_4$
- $X_1^4 + X_2 + X_3 + X_4$
- $log(X_1) + X_2 + X_3 + X_4$

Speaker notes

Multiply two or more of the variables together and it is no longer a linear combination. Raise one of the variables to a power and it is no longer a linear combination. Taking a function of one of the variables and it is no longer a linear combination.

Practical examples

- Apgar score is a linear combination
  - Rate appearance, pulse, grimace response, activity, and respiration
  - Add the values together
- Body mass index is not a linear combination
  - Weight divided by height squared

Speaker notes

Apgar score is a linear combination. Rate five features of a newborn baby on a scale of 0, 1, or 2. Then add the five values together to produce a linear combination that equals zero for a baby in extreme distress all the way to a happy baby with an apgar score of 10. Actually, the baby has to have a loud cry to get a 10, so perhaps this is more of a healthy baby than a happy baby.

Body mass index is not a linear combination. It involved division, which is just as much of a no-no as multiplication. It also involve a power of two which would also make it not a linear combination.

Philosophy behind principal components, 1 of 4

- Reduce complexity by modeling inter-relationships
- Inter-relationships represented by linear combinations
  - There is no dependent or outcome variable in principal components analysis

Philosophy behind principal components, 2 of 4

- First principal component
  - Linear combination that accounts most variation
  - This linear combination is the first eigenvector
  - The amount of variation accounted for is the first eigenvalue

Philosophy behind principal components, 3 of 4

- Need to resolve an ambiguity
  - $3X_1 + 5X_2 - 4X_3 + 7X_4 - 1X_5$ versus $6X_1 + 10X_2 - 8X_3 + 14X_4 - 2X_5$
  - Solution: require sum of squared coefficients to equal 1
  - Note: $3^2 + 5^2 + (-4)^2 + 7^2 + (-1)^2 = 100$
  - Use $\frac{3}{10}X_1 + \frac{5}{10}X_2 - \frac{4}{10}X_3 + \frac{7}{10}X_4 - \frac{1}{10}X_5$

Philosophy behind principal components, 4 of 4

- Second principal component
  - Linear combination that accounts second most variation
  - Must be uncorrelated with first principal component
  - This linear combination is the second eigenvector
  - The amount of variation accounted for is the second eigenvalue
- Third principal component defined similarly

Covariance matrix or correlation matrix

- Correlation matrix equivalent to standardizing
  - Absolute requirement if differing units
- Covariance matrix de-emphasizes low variance variables

How many components?

- Percentage of variation accounted for
  - Scree plot
- Eigenvalues > 1
- Researcher preference/convenience

Communality

- Amount of shared variation
  - Always between 0 and 1
  - Similar interpretation to R-squared
  - “One of these things is not like the others”

Factor score matrix

- Linear combination coefficients
- Needed if you score by hand
- No obvious interpretation
- First component is often only positive values

Component matrix

- Interpret as correlation matrix
  - Rows are individual variables
  - Columns are principal components

Health-related quality of life, 1 of 6

data_dictionary: Health-related Quality of Life of Persons Living with Tuberculosis.sav

source: |
  Tornu, Eric ; Quarcoopome, Louisa (2022), "Health-related Quality of Life of Persons Living with Tuberculosis", Mendeley Data, V2, doi: 10.17632/jg4xp7883w.2

description: |
  From the original source: This data is on a study which accessed the health-related quality of life of persons living with Tuberculosis.

Speaker notes

Here are a few pieces of the data dictionary for a file I will be using. I won’t show the entire data dictionary, as it is quite long.

Health-related quality of life, 2 of 6

```
Howwouldyourateyourqualityoflife_1:
  values:
    '1': Poor
    '2': Neither poor nor good
    '3': Good
    '4': Very Good

HowSatisareyouwithyourhealth_2:
  values:
    '1': Very dissatisfied
    '2': Dissatisfied
    '3': Neither satisfied nor satisfied
    '4': Satisfied
    '5': Very Satisfied

Physicalpainprevents_3:
  values:
    '1': An extreme amount
    '2': Very much
    '3': A moderate amount
    '4': A little
    '5': Not at all
```

Health-related quality of life, 3 of 6

```
Howmuchmedicaltreatment_4:
  values:
    '1': An extreme amount
    '2': Very much
    '3': A moderate amount
    '4': A little
    '5': Not at all

Howmuchdoyouenjoylife_5:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Towhatextentdoyoufeelyourlifetobemeaningful_6:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely
```

Health-related quality of life, 4 of 6

```
Howwellareyouabletoconcentrate_7:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Howsafedoyoufeelinyourdailylife_8:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Howhealthyisyourphysicalenvironment_9:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely
```

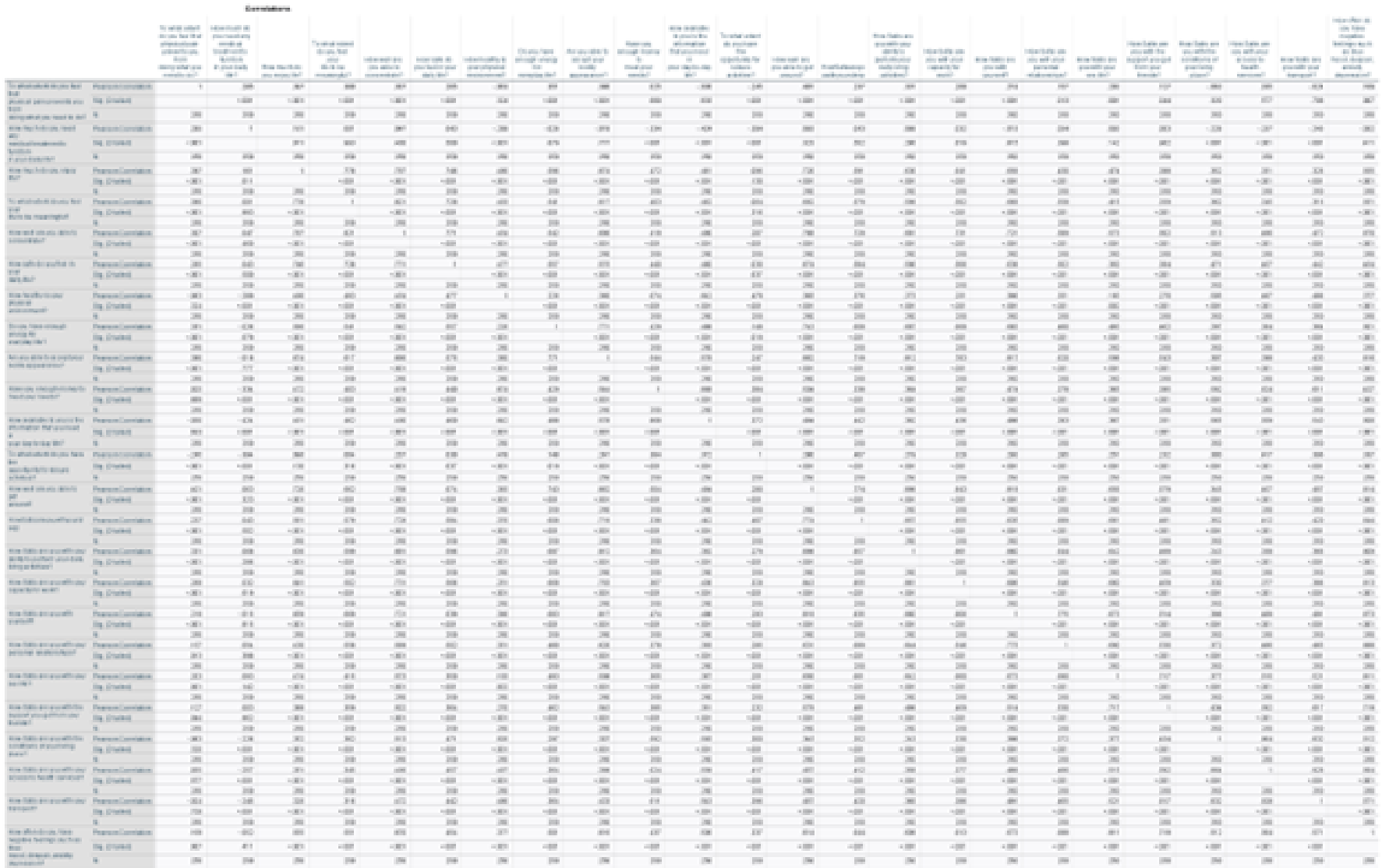
Health-related quality of life, 5 of 6

```
Doyouhaveenoughenergyforeverydaylife_10:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Areyouabletoacceptyourbodilyappearance_11:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Haveyouenoughmoneytomeetyourneeds_12:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely
```

Health-related quality of life, 6 of 6

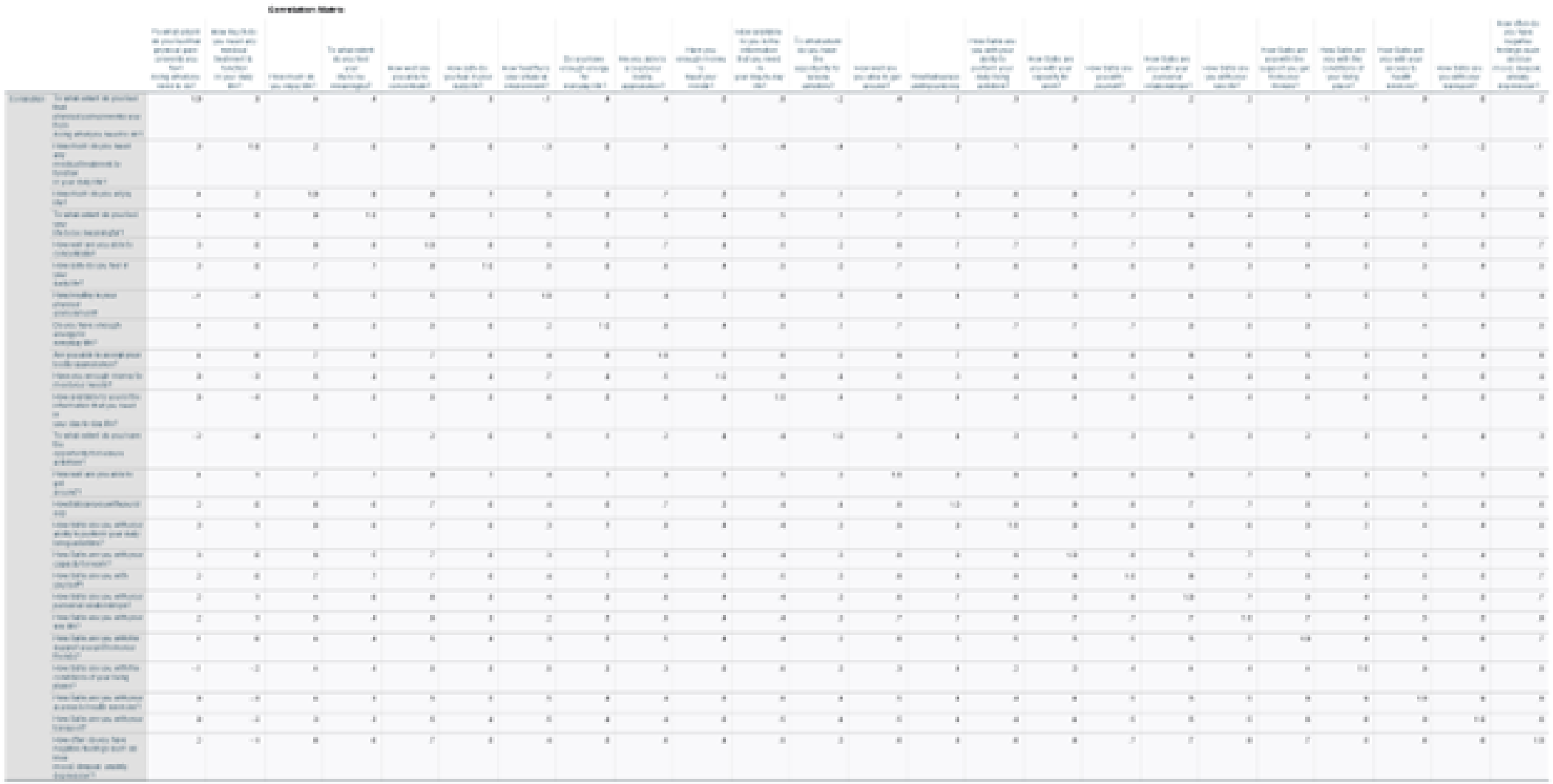
```
AvailableInformation_13:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Opportunityforleisureactivities_14:
  values:
    '1': Not at all
    '2': A little
    '3': A moderate amount
    '4': Very much
    '5': Extremely

Howwellareyouabletogetaround_15:
  values:
    '1': Very poor
    '2': Poor
    '3': Neither poor nor good
    '4': Good
    '5': Very Good
```

Correlation matrix, 1 of 3



Speaker notes

Correlation matrix, 2 of 3



Correlation matrix, 3 of 3

Correlation Matrix

		To what extent do you feel that physical pain prevents you from doing what you need to do?	How much do you need any medical treatment to function in your daily life?	How much do you enjoy life?	To what extent do you feel your life to be meaningful?
Correlation	To what extent do you feel that physical pain prevents you from doing what you need to do?	1.0	.3	.4	.4
	How much do you need any medical treatment to function in your daily life?	.3	1.0	.2	.0
	How much do you enjoy life?	.4	.2	1.0	.8
	To what extent do you feel your life to be meaningful?	.4	.0	.8	1.0

Communalities

Communalities

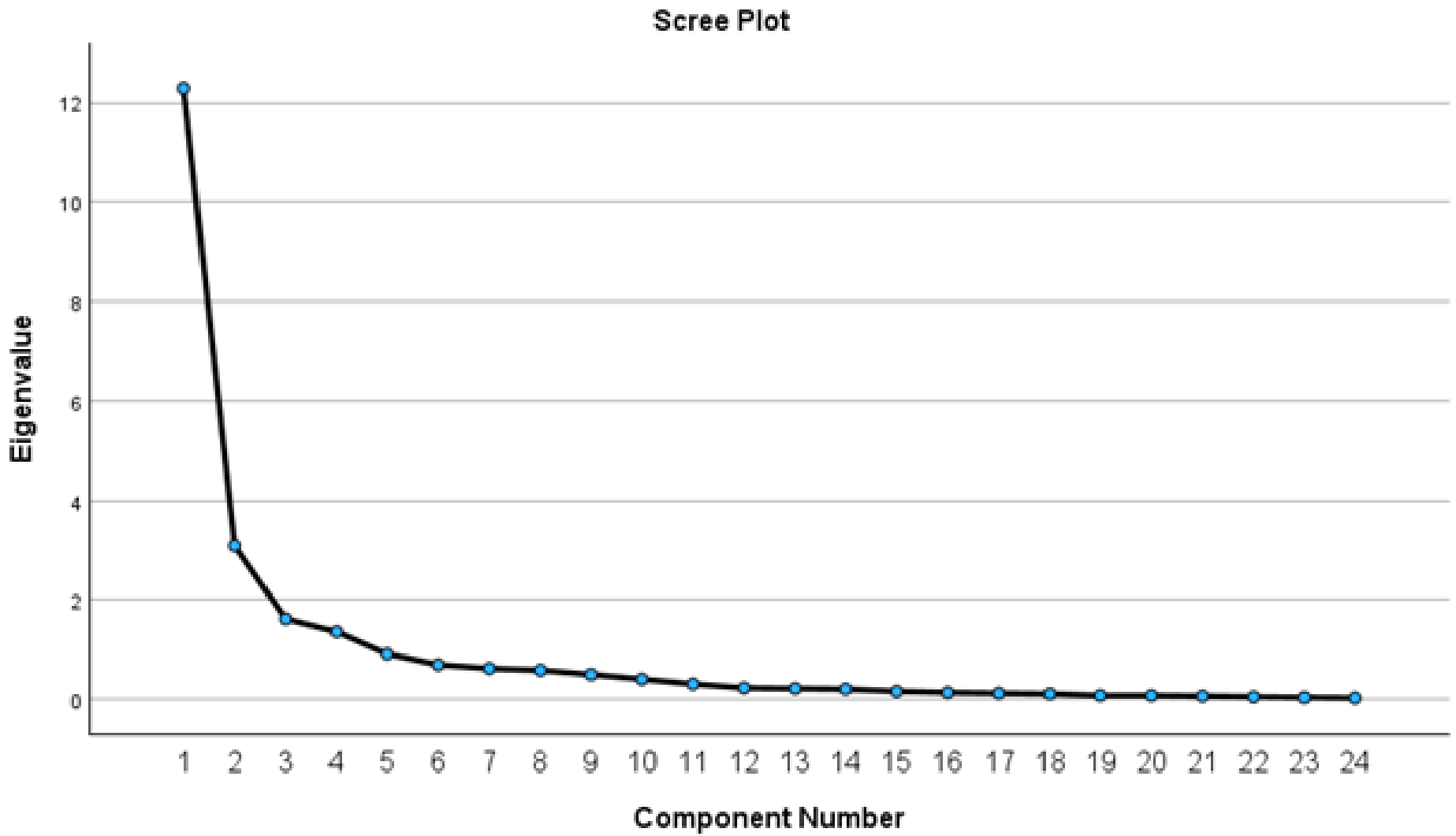
	Initial	Extraction
To what extent do you feel that physical pain prevents you from doing what you need to do?	1.000	.496
How much do you need any medical treatment to function in your daily life?	1.000	.607
How much do you enjoy life?	1.000	.814
To what extent do you feel your life to be meaningful?	1.000	.774
How well are you able to concentrate?	1.000	.797
How safe do you feel in your daily life?	1.000	.764
How healthy is your	1.000	.699

Eigenvalues

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.285	51.185	51.185	12.285	51.185	51.185
2	3.101	12.923	64.108	3.101	12.923	64.108
3	1.623	6.762	70.870	1.623	6.762	70.870
4	1.366	5.691	76.561	1.366	5.691	76.561
5	.918	3.824	80.385			
6	.701	2.919	83.304			
7	.627	2.612	85.915			
8	.592	2.465	88.380			
9	.505	2.103	90.483			
10	.410	1.710	92.193			

Scree plot


Scree Plot
Eigenvalue
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Component Number

Component matrix

Component Matrix[a]

	Component			
	1	2	3	4
To what extent do you feel that physical pain prevents you from doing what you need to do?	.282	.557	.199	.257
How much do you need any medical treatment to function in your daily life?	-.073	.642	-.130	.415
How much do you enjoy life?	.770	.257	.376	.116
To what extent do you feel your life to be meaningful?	.745	.206	.402	.120
How well are you able to concentrate?	.852	.155	.174	.132
How safe do you feel in	.743	.125	.395	.203

Live demo, Principal components analysis

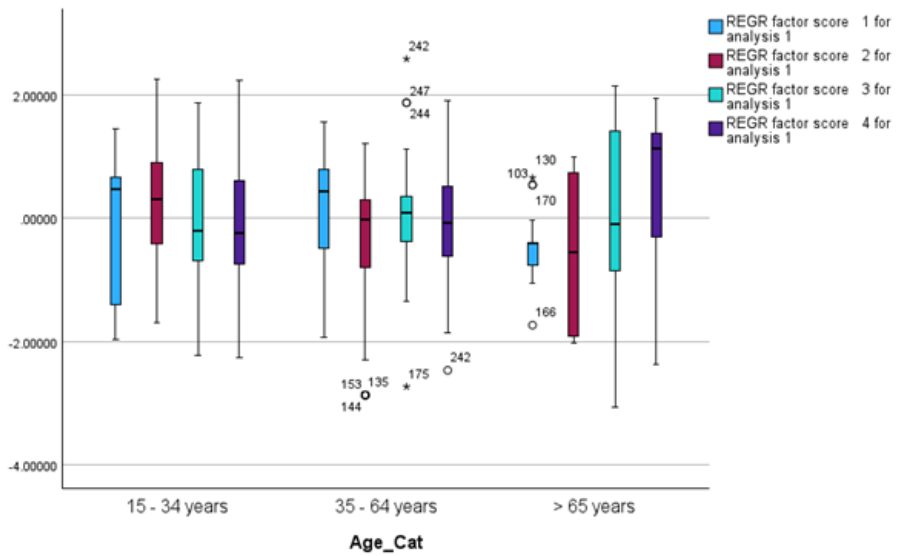
Break #1

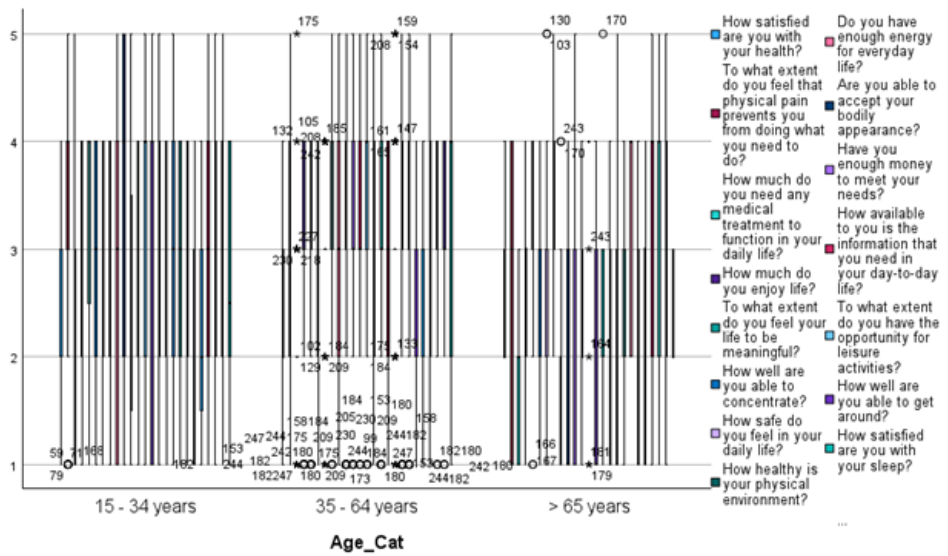
- What you have learned
  - Principal components analysis
- What’s coming next
  - Applications of principal components

Applications

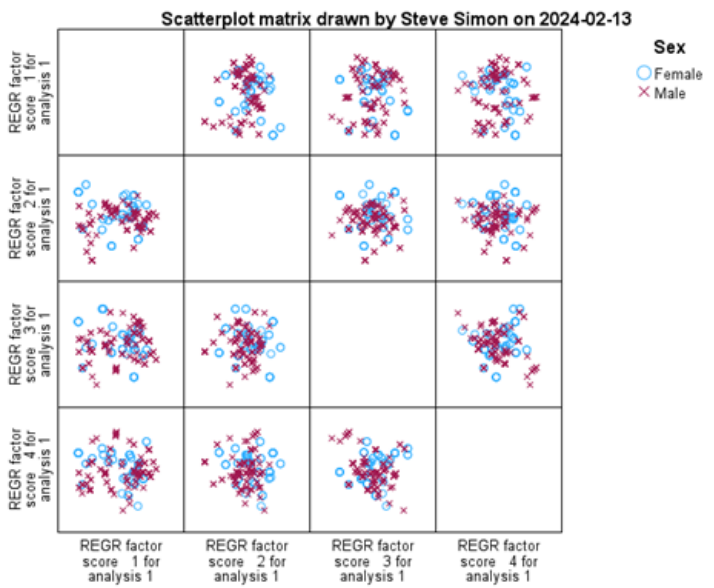
- Visualization
  - Reduce high dimensional visualization
  - Fewer graphs
- Regression analysis
  - Fewer independent variables (rule of 15)
  - Removes collinearity

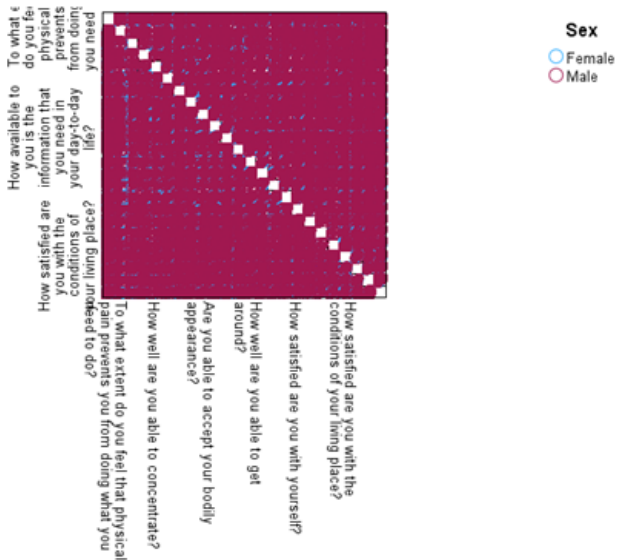
Boxplots of first four principal components





Scatterplot of first four principal components





R-squared using four principal components

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.785[a]	.617	.610	.528

a. Predictors: (Constant), REGR factor score 4 for analysis 1, REGR factor score 3 for analysis 1, REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

R-squared using all 24 variables

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.841[a]	.707	.676	.482

a. Predictors: (Constant), How often do you have negative feelings such as blue mood, despair, anxiety, depression?, How much do you need any medical treatment to function in your daily life?, To what extent do you feel that physical pain prevents you from doing what you need to do?, How healthy is your physical environment?, Do you have enough energy for everyday life?, To what extent do you have the opportunity for leisure activities?, How satisfied are you with the conditions of your living place?, How satisfied are you with your personal relationships?, How satisfied are you with the support you get from your friends?, How safe do you feel in your daily life?, Have you enough money to meet your needs?, How satisfied are you with your capacity for work?, To what extent do you feel your life to be meaningful?, How satisfied are you with your sex life?, Are you able to accept your bodily appearance?, How satisfied are you with your transport?, How much do you enjoy life?, How satisfied are you with your ability to perform your daily living activities?, How available to you is the information that you need in your day-to-day life?, How satisfied are you with your sleep?, How well are you able to concentrate?, How satisfied are you with yourself?, How satisfied are you with your access to health services?, How well are you able to get around?

Live demo, Applications of principal components

Break #2

- What you have learned
  - Applications of principal components
- What’s coming next
  - Factor analysis

Philosophy behind factor analysis

- Variance equals information
- Covariance (correlation) equals shared information
- Modeling shared information creates latent variables

Speaker notes

Most of the time, statisticians talk about variation as if it were a bad thing. It’s noise. It is what makes our confidence intervals so wide. It’s what forces us to collect such large sample sizes.

But, in a different context, variation is a good thing. It provides valuable information.

Consider studies of student success done at Harvard and at Johnson County Community College. There is relatively little variation at Harvard in things like high school GPA and SAT scores because everyone is clustered near the upper limits of these scales. At Johnson County Community College–a very good school, don’t get me wrong, there is a bit more variation. So you can see enough of a spread to better understand factors for success after graduation.

Some clinical trials deliberately focus on the most seriously ill patients. Mildly ill patients don’t have a lot of variation because there is only a small amount of improvement needed to get them back to full health. Serioulsly ill patients have more variation, giving you more opportunity to identify successful treatment options. Now, there are exceptions. If the patients are so seriously ill that no one recovers, you fall back into the same problem of too little variation.

If variance equals information, then covariance (or equivalently correlation) equals shared information.

Factor rotation

- Recombine factors
- Strive for simple interpretation
  - Components close to -1, 0, or 1
  - Each variable has one and only one non-zero components
  - Not always achievable

Rotated factor pattern, 1 of 3

Rotated Factor Matrix[a]

	Factor			
	1	2	3	4
To what extent do you feel that physical pain prevents you from doing what you need to do?	.278	.351	-.066	-.185
How much do you need any medical treatment to function in your daily life?	.083	.160	-.133	-.510
How much do you enjoy life?	.464	.731	.108	.156
To what extent do you feel your life to be meaningful?	.391	.789	.131	.122
How well are you able to concentrate?	.522	.696	.307	.074
How safe do you feel in	.387	.693	.266	.131

Rotated factor pattern, 2 of 3

Rotated Factor Matrix[a]

	Factor			
	1	2	3	4
How satisfied are you with your ability to perform your daily living activities?	.911	.276	.096	.047
How satisfied are you with your capacity for work?	.843	.274	.139	.077
How well are you able to get around?	.819	.396	.170	.159
How satisfied are you with your sleep?	.813	.293	.196	.058
How satisfied are you with yourself?	.810	.338	.230	.153
Are you able to accept your	.778	.340	.090	.287

Rotated factor pattern, 3 of 3

Rotated Factor Matrix[a]

	Factor			
	1	2	3	4
How satisfied are you with your ability to perform your daily living activities?	.911			
How satisfied are you with your capacity for work?	.843			
How well are you able to get around?	.819			
How satisfied are you with your sleep?	.813			
How satisfied are you with yourself?	.810			
Are you able to accept your	.778			

Live demo, Factor analysis

Break #3

- What you have learned
  - Factor analysis
- What's coming next
  - Criticisms of principal components analysis and factor analysis

Criticisms of principal components analysis

- Advantages
  - Makes collection of many variables manageable
  - Eliminates collinearity issues
  - Focus only on important sources of variation
- Disadvantages
  - Components often uninterpretable
  - False sense of parsimony

Speaker notes

Criticisms of factor analysis

- Advantages
  - Explore underlying structure
  - Create or validate subscales
- Disadvantages
  - Difficulty in choosing number of factors
  - Reification

Speaker notes

Summary

- What you have learned
  - Principal components analysis
  - Applications of principal components
  - Factor analysis
  - Criticisms of principal components analysis and factor analysis

Additional topics??

Speaker notes

- Learning objectives
  - To be determined