

# Human-Computer Interaction Subjective Measures + Scale Construction

Professor Bilge Mutlu

# Questions

To ask questions during class:

- » Go to [slido.com](https://slido.com) and use code #**2938904** or [direct link](#) or scan QR code
- » Anonymous
- » I will monitor during class



# Today's Agenda

- » Topic overview
  - » Overview of subjective measures
  - » How to design good questions (quick review)
  - » Scale construction
- » Tutorial on factor analysis

**Recap:** What are different kinds of measurements we can take?

1. **Objective:** Measurement from participants against an objective standard, e.g., performance in a test
2. **Behavioral:** Measurement of the actions and behaviors of participants, E.g., how much eye-contact participants maintain with a robot
3. **Subjective:** Measurement of self-report data on subjective evaluations, e.g., preferences, personality ➤ our focus today
4. **Physiological:** Measurements taken directly from participants' bodies, e.g., body temperature, GSR, EEG, EMG, fMRI

*What are subjective measures?*

**Definition:** Measurements of the subjective perceptions, e.g., thoughts, feelings, preferences, and individual traits using self-reported data collection instruments, often *questions*.

Types of subjective measurement instruments:

1. **Standardized responses:** Questionnaires
2. **Open-ended responses:** Interviews

These instruments can be administered by the researcher or by the respondent.

*What is a survey then?*

**Definition:** A survey is a quantitative empirical research method that uses questionnaires or interviews to collect, analyze, and interpret data from populations of interest.

*What does a subjective measure look like?*

Subjective measures follow the archetypal formats below:

1. Open-ended question
2. Closed-ended question

# Open-ended question

**Definition:** Questions designed to prompt rich, unstructured responses for participants that will generate data that can be qualitatively or quantitatively analyzed.

*What barriers did you face, in attempting to use the Banjee software to complete your tasks?*<sup>1</sup>

[Open-ended answer]

---

<sup>1</sup>Lazar et al., 2017, Chapter 5 - Surveys

# Closed-ended question

**Definition:** Questions where standardized response instruments are used to standardize responses so that statistical methods can be used.

*What is your impression of using the website for www.veggieworld.com?*<sup>1</sup>

Please circle one number:

(Frustrating) 1 – 2 – 3 – 4 – 5 – 6 – 7 (Satisfying)

---

<sup>1</sup>Lazar et al., 2017, Chapter 5 - Surveys

*What are different standardized response instruments?*

1. Likert scales
2. Rating scales
3. Semantic differential scales

## *What is a Likert scale?*

**Definition:** A Likert scale includes a number of rank-ordered items that respondents use to express their level of agreement with a statement or a question.

- » Strongly disagree
- » Disagree
- » Neither agree nor disagree
- » Agree
- » Strongly agree

*What is a rating scale?*

**Definition:** A rating scale is a numerical range with which participants can express their level of agreement with a statement or a question.

1 2 3 4 5 6 7

1 2 3 4 5

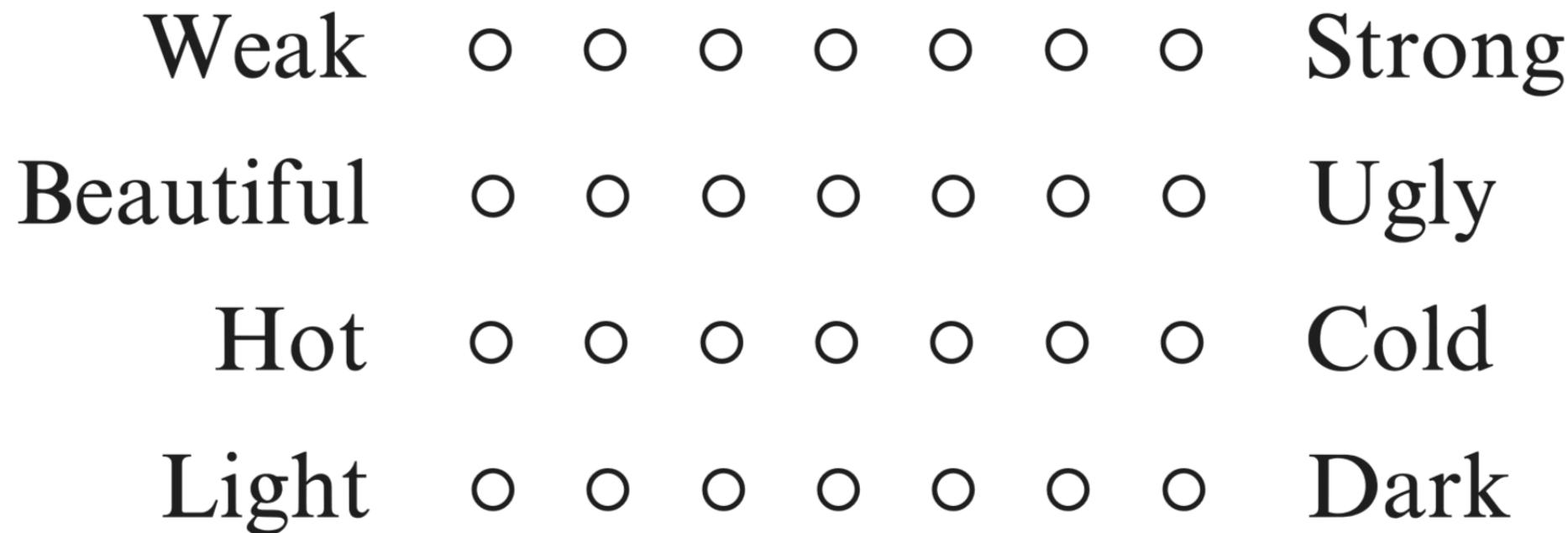
-1 0 1

An anchored rating scale provides anchor terms that ground the ends of the scale in descriptive terms:

(Frustrating) 1 – 2 – 3 – 4 – 5 – 6 – 7 (Satisfying)

## *What is a semantic differential scale?*

**Definition:** The semantic differential technique involves presenting pairs of bipolar, or opposite, adjectives at either end of a series of scales.<sup>1</sup>



---

<sup>1</sup>Lazar et al., 2017, Chapter 5 – Surveys

*What are scales (or metrics)?*

**Definition:** A scale (or metric) is an instrument made up of individual *items* that measures self-reported data on a *construct*.

*What is a construct?*

**Definition:** A concept or attribute of interest that we wish to measure and that has been conceptualized to aid in the measurement.

*What is an item?*

**Definition:** A question or a survey prompt that participants respond to.

*Wait, I am confused. What are scales again?*

Scale as a **response** instrument:

A response instrument is a mechanism to capture the extent to which participants agree with statements made in items.

Scale as a **measurement** instrument:

A scale (or metric) is an instrument made up of individual *items* that measures self-reported data on a *construct*.

**Recap:** What are different types of variables?

**Nominal:** names of groups or categories, e.g., experts vs. novices, juniors vs. seniors

**Ordinal:** rank-ordering of measurements, e.g., very satisfied, satisfied, neutral, unsatisfied, very unsatisfied

**Interval:** measurements along a scale with no real zero, e.g., happiness in a scale of 1 to 7 ➔

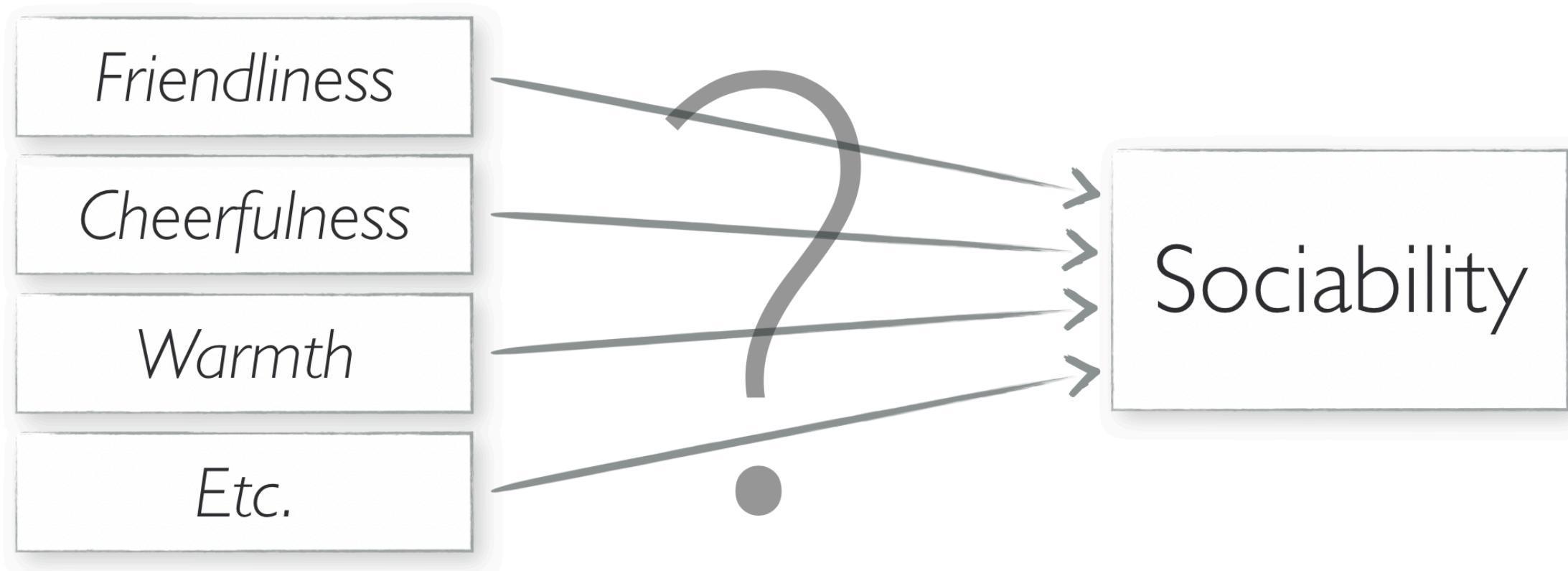
**variable type used for scales**

**Ratio:** absolute measurements along a scale with a real zero, e.g., a person's weight

*Why are scales important?*

**An example:** how can we measure *sociability*?

Too vague and multifaceted to be measured directly. Might be made up of sub-constructs, e.g., *friendliness*, *cheerfulness*, *warmth*, etc.



*How do we make up scales?*

A proposed algorithm:

1. Consider the construct you wish to measure, e.g., sociability, trust, usability
2. Using mind-mapping or lists, write down potential components of the construct, e.g.,  
sociability → friendliness; trust → credibility; usability → ease of use (e.g., use a thesaurus or  
WordNet)
3. The connections will serve as hypotheses that will be tested statistically through a process  
called *factor analysis*
4. Prune connections that may not be substantiated in data later

Alternatively, use a validated scale, where one exist. Examples:

- » Usability scales
- » International Personality Item Pool scales

# What does a validated scale look like? SUS, UEQ

	Strongly disagree					Strongly agree	
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
	1	2	3	4	5		
1. I think that I would like to use this system frequently						✓	4
2. I found the system unnecessarily complex						✓	1
3. I thought the system was easy to use			✓				1
4. I think that I would need the support of a technical person to be able to use this system	✓						4
5. I found the various functions in this system were well integrated			✓				1
6. I thought there was too much inconsistency in this system				✓			2
7. I would imagine that most people would learn to use this system very quickly			✓				1
8. I found the system very cumbersome to use					✓		1
9. I felt very confident using the system						✓	4
10. I needed to learn a lot of things before I could get going with this system			✓				3

Total= 22

SUS Score=  $22 * 2.5 = 55$

	1	2	3	4	5	6	7	
annoying	○	○	○	○	○	○	○	enjoyable
not understandable	○	○	○	○	○	○	○	understandable
creative	○	○	○	○	○	○	○	dull
easy to learn	○	○	○	○	○	○	○	difficult to learn
valuable	○	○	○	○	○	○	○	inferior
boring	○	○	○	○	○	○	○	exciting
not interesting	○	○	○	○	○	○	○	interesting
unpredictable	○	○	○	○	○	○	○	predictable
fast	○	○	○	○	○	○	○	slow
inventive	○	○	○	○	○	○	○	conventional
obstructive	○	○	○	○	○	○	○	supportive
good	○	○	○	○	○	○	○	bad
complicated	○	○	○	○	○	○	○	easy
unlikable	○	○	○	○	○	○	○	pleasing
usual	○	○	○	○	○	○	○	leading edge
unpleasant	○	○	○	○	○	○	○	pleasant
secure	○	○	○	○	○	○	○	not secure
motivating	○	○	○	○	○	○	○	demotivating
meet expectations	○	○	○	○	○	○	○	does not meet expectations
inefficient	○	○	○	○	○	○	○	efficient
clear	○	○	○	○	○	○	○	confusing
impractical	○	○	○	○	○	○	○	practical
organized	○	○	○	○	○	○	○	cluttered
attractive	○	○	○	○	○	○	○	unattractive
friendly	○	○	○	○	○	○	○	unfriendly
conservative	○	○	○	○	○	○	○	innovative

*What do we do when validated scales and metrics don't exist?*

We develop our own by designing **good** questions!

*How do we design good questions?*

## ***Principle 1: Avoid “leading” or “loaded” questions***

### **Example**

*Don’t you agree that social workers should earn more money than they currently earn?*

- » Yes, they should earn more
- » No, they should not earn more
- » Don’t know/no opinion

## ***Principle 2: Avoid double negatives***

### **Example**

*Do you agree or disagree with the following statement?*

Teachers should not be required to supervise their students during recess.

**Principle 3:** Always aim at capturing firsthand experiences and beware of asking about information that is acquired only secondhand

## Tip

People are very good at describing criminal activity directed at them but terrible at describing how much crime happens in their neighborhood

## ***Principle 4: Beware of asking hypothetical questions***

### **Tip**

People are not good at predicting what they will do as they have limited direct experience with future situations

## ***Principle 5: Beware of asking about causality***

### **Tip**

Events mostly have more than one reason

People are not good at describing why they do the things they do

### **Example**

*Were you limited in your daily activities because of your back problem?*

*What is the main reason why you did not vote?*

*Were you homeless because of high cost of housing?*

## ***Principle 6: Beware of asking about solutions to complex problems***

### **Tip**

People in general do not have informed opinions about complex issues

## ***Principle 7: Avoid asking more than one question at a time***

### **Tip**

The answers to two questions can be dramatically different

### **Example**

*Would you like to be rich and famous?*

*Are you physically able to do things like swim and run without difficulty?*

## ***Principle 8: Avoid asking questions that impose unwarranted assumptions***

### **Tip**

Double-barreled or one-and-a-half-barreled questions

### **Example**

*Should the organization reduce paperwork required of employees by hiring more administrators?  
With the economy the way it is, do you think investing in the stock market is a good idea?*

## ***Principle 9: Beware of questions that include hidden contingencies***

### **Tip**

Questions must apply to the majority of your sample

### **Example**

To measure social activity:

*How often do you interact with people on Facebook?*

**Principle 10:** *The words in questions should be chosen so that all respondents understand their meaning and have the same sense of what the meaning is*

~

**Principle 11:** When words or terms that have meanings that are likely no to be shared, definitions should be provided to all respondents

## Example

*In the past 12 months, how many times have you seen or talked with a medical doctor about your health? Include visits to psychiatrists, ophthalmologists, and any other professional with a medical degree.*

**Principle 12:** If definitions are provided, they should be given before the question itself is asked

## Example

*How many days in the past week have you done any exercise? When you consider exercise be sure to include walking, work around the house, or work on a job, if you think they constituted exercise.*

## Better:

*The next question is going to ask you about how often you've engaged in exercise. We want to you to include walking, anything you may do around the house, or work around the house, or work on a job, if you think they constituted exercise. Using this definition, in the last week, on how many days did you do any exercise?*

**Principle 13:** *The time period referred to by a question should be unambiguous and questions about feelings or behaviors must refer to a period of time*

## **Tip**

Questions about feelings or behaviors must refer to a period of time.

## **Example**

*Are you able to run half a mile without stopping?*

*How many drinks do you usually have on days when you drink any alcoholic beverages at all?*

**Principle 14:** If what is to be covered is too complex to be included in a single question, ask multiple questions

~

***Principle 15: Use multiple questions to measure the same thing***

~

**Principle 16:** A question should end with the question itself. If there are response alternatives, they should constitute the final part of the question

## Example

*Would you say that you are very likely, fairly likely, or not likely to move out of this house in the next year?*

## Better:

*In the coming year, how likely are you to move to another home? Would you say very likely, fairly likely, or not very likely?*

**Principle 17:** Clearly communicate to all respondents the kind of answer that constitutes an adequate answer to a question

## Example

*When did you move to this community?*

*Possible answers:*

- » When I was sixteen.
- » Right after I was married.
- » In 1953.

**Better:**

*In what year did you move to this community?*

**Principle 18:** Specify the number of responses to be given to questions for which more than one answer is possible

## Example

*What was it about the brand you bought that made you buy it rather than some other brand? List all that apply.*

**Principle 19:** Design survey instruments to make the tasks of reading questions, following instructions, and recording answers as easy as possible for interviewers and respondent

~

**Principle 20:** Measurements will be better to the extent that people answering questions are oriented to the task in a consistent way

## Tip

Train your respondents!

*So, I have good questions. How do I turn them into a scale or metric?*

## *What is factor analysis?*

**Definition:** A statistical test to explore relationships among items that make up a provisional scale used for scale construction and data reduction.

## *What does factor analysis do?*

- » Removes redundancy/duplication from a set of correlated variables
- » Represents correlated variables with a smaller set of derived variables, called **factors**, that are relatively independent of one another
- » Represents the statistical relationships between items and factors as **loadings**; greater loading indicates stronger relationship

*Are there different types of factor analyses?*

**Exploratory factor analysis** aims to discover relationships among items and constructs. E.g., *what are components of sociability?*

**Confirmatory factor analysis** aims to confirm whether proposed relationships are substantiated in data. E.g., *how well do my questions measure sociability?*

**Exploratory-confirmatory factor analysis** aims to first discover and then confirm relationships in order to develop usable scales. E.g., *what is a good scale of sociability?*

## *How do we do exploratory factor analysis?*

- » Collect and explore data — choose relevant variables
- » Extract initial factors via principal components analysis (PCA)
- » Choose number of factors to retain
- » Choose estimation method, estimate model
- » Rotate and interpret
- » Decide if changes need to be made and estimate, rotate, and implement again
- » Construct scales and use in further analysis

## *How do we do confirmatory factor analysis?*

- » Define the factor model
- » Collect measurements
- » Obtain the correlation matrix
- » Fit the model to the data
- » Evaluate model adequacy
- » Compare with other models

## *What is the exploratory-confirmatory factor analysis?*

- » Perform an exploratory factor analysis and decide on the number of factors,  $m$ .
- » Fit an  $m$ -factor model, and rotate to simple structure using, e.g., varimax.
- » For each column of the factor pattern, find the largest loading, then constrain all the other loadings in that row to be zero, and fit the resulting model as a confirmatory factor model.
- » Examine the factor pattern and test all factor loadings. Delete non-significant loadings from the model.

## *Where do we get started?*

Imagine that we are interested in measuring factors that might affect people's decisions about buying a car.

We design a questionnaire with a number of items that we think will be relevant: price, safety, exterior appearance, space/comfort, technology, after sales service, resale value, fuel type, fuel efficiency, color, maintenance, test drive, product reviews, testimonials.

# How important is the following factors in your decision to purchase?

<b>Price</b>	Not important	1	2	3	4	5	Important
<b>Safety</b>	Not important	1	2	3	4	5	Important
<b>Exterior appearance</b>	Not important	1	2	3	4	5	Important
<b>Space/comfort</b>	Not important	1	2	3	4	5	Important
<b>Technology</b>	Not important	1	2	3	4	5	Important
<b>After sales service</b>	Not important	1	2	3	4	5	Important
<b>Resale value</b>	Not important	1	2	3	4	5	Important
<b>Fuel type</b>	Not important	1	2	3	4	5	Important
<b>Fuel efficiency</b>	Not important	1	2	3	4	5	Important
<b>Color</b>	Not important	1	2	3	4	5	Important
<b>Maintenance</b>	Not important	1	2	3	4	5	Important
<b>Test drive</b>	Not important	1	2	3	4	5	Important
<b>Product reviews</b>	Not important	1	2	3	4	5	Important
<b>Testimonials</b>	Not important	1	2	3	4	5	Important

Given  $m$  factors and  $n$  observed variables:

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \dots + \lambda_{1m}F_m + e_1$$

$$X_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \dots + \lambda_{2m}F_m + e_2$$

...

$$X_n = \lambda_{n1}F_1 + \lambda_{n2}F_2 + \dots + \lambda_{nm}F_m + e_n$$

In matrix notation:

$$X_{n \times 1} = \Lambda_{n \times m} F_{m \times 1} + e_{n \times 1}$$

$$\begin{matrix} X_1 \\ \vdots \\ X_n \end{matrix} = \begin{matrix} \lambda_{11} & \dots & \dots & \lambda_{1m} \\ \vdots & & & \vdots \\ \lambda_{n1} & \dots & \dots & \lambda_{nm} \end{matrix} \begin{matrix} F_1 \\ \vdots \\ F_m \end{matrix} + \begin{matrix} e_1 \\ \vdots \\ e_m \end{matrix}$$

*How do we interpret the factor matrix?*

$$\begin{matrix} X_1 \\ \vdots \\ X_n \end{matrix} = \begin{matrix} \lambda_{11} & \dots & \dots & \lambda_{1m} \\ \vdots & & & \vdots \\ \lambda_{n1} & \dots & \dots & \lambda_{nm} \end{matrix} \begin{matrix} F_1 \\ \vdots \\ F_m \end{matrix} + \begin{matrix} e_1 \\ \vdots \\ e_m \end{matrix}$$

## *What is factor rotation?*

**Definition:** Factor rotation is a statistical technique that allows us to make more clear-cut decisions by spreading variability more evenly among factors by redefining factors to force loadings to be very high (-1 or 1) or very low (0).

There are different methods of factor rotation. We will use *varimax*, which maximizes squared loading variance across variables (sum over factors).

# *Let's try it out!*<sup>2</sup><sup>3</sup>

---

<sup>2</sup>We'll use R.

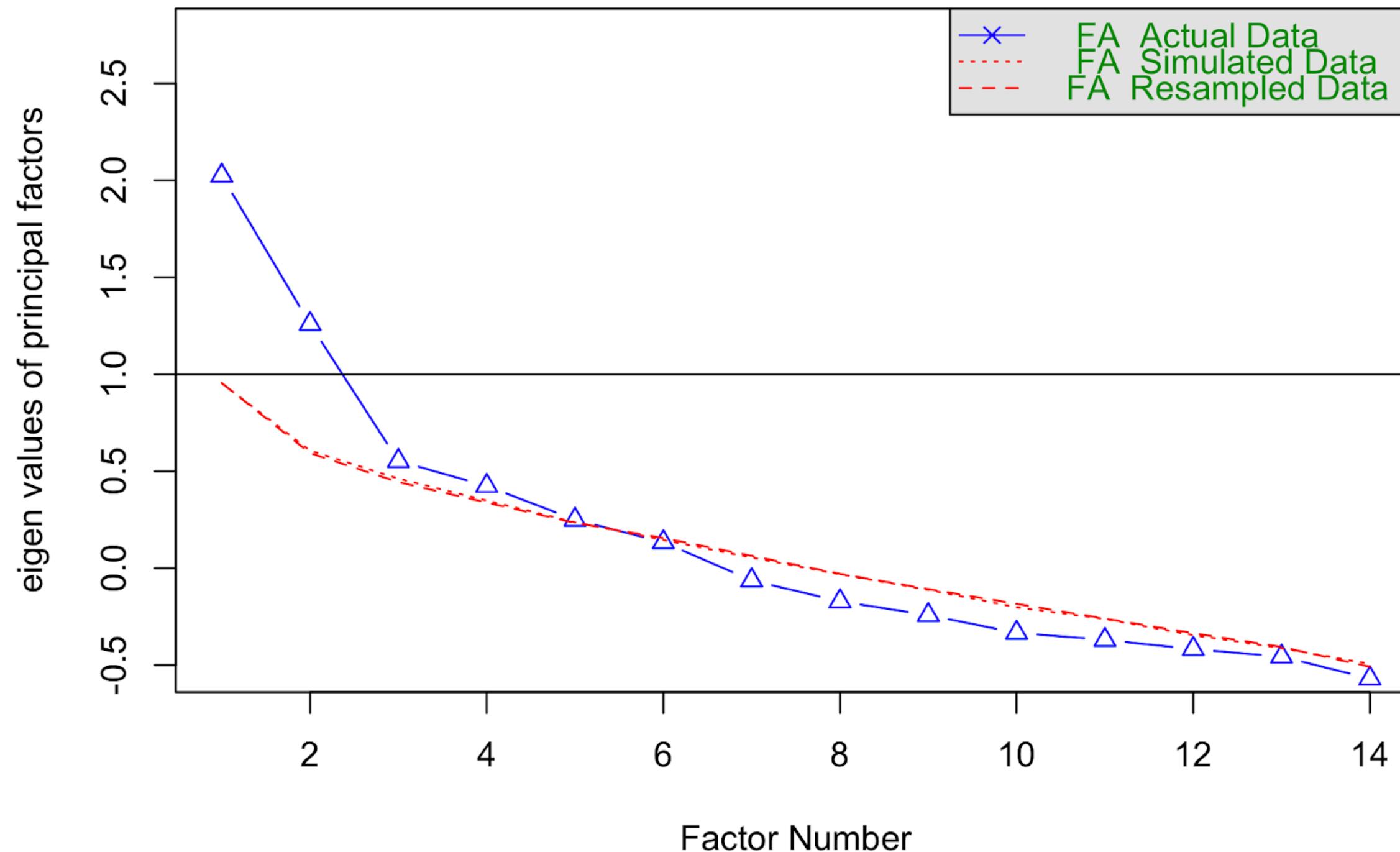
<sup>3</sup>We will use same data from PromptCloud.

*Step 1. Determine the number of factors using PCA*

```
install.packages("psych")
library(psych)
library(readr)
EFA <- read_csv("EFA.csv")
pa = fa.parallel(EFA, fm = 'minres', fa = 'fa')
```

This will produce what is called a Scree plot that will plot eigenvalues on the Y axis and number of factors on the X axis.

## Parallel Analysis Scree Plots



*How do we determine the number of factors?*

There are two methods:

1. **Kaiser Criterion:**<sup>4</sup> take eigenvalues that are larger than 1.
2. **Scree test:**<sup>5</sup> find point of inflection and consider the factors up to the leveling off.

---

<sup>4</sup>Kaiser, 1960, The application of electronic computers to factor analysis<sup>1</sup>

<sup>5</sup>Cattell, 1966, The Scree test for the number of factors

## *Step 2: Factor rotation*

Calculate loadings for each variable on each factor:

$$\text{corr}(F_i, X_j) = \lambda_{ji}$$

Apply factor rotation to spread the variability evenly among variables:

```
fit = fa(data, nfactors = 3, rotate = "varimax", fm="minres")
```

Visualize the factor matrix:

```
print(fit$loadings, cutoff = 0.3)
```

This will print out the following factor matrix:

Loadings:

	MR1	MR2	MR3
Price	0.444		
Safety		0.311	
Exterior_Looks			
Space_comfort		0.832	
Technology		0.342	
After_Sales_Service		0.460	
Resale_Value	0.599		
Fuel_Type		0.573	
Fuel_Efficiency	0.655		
Color	0.464		
Maintenance	0.668		
Test_drive			0.328
Product_reviews	0.424		
Testimonials			0.742

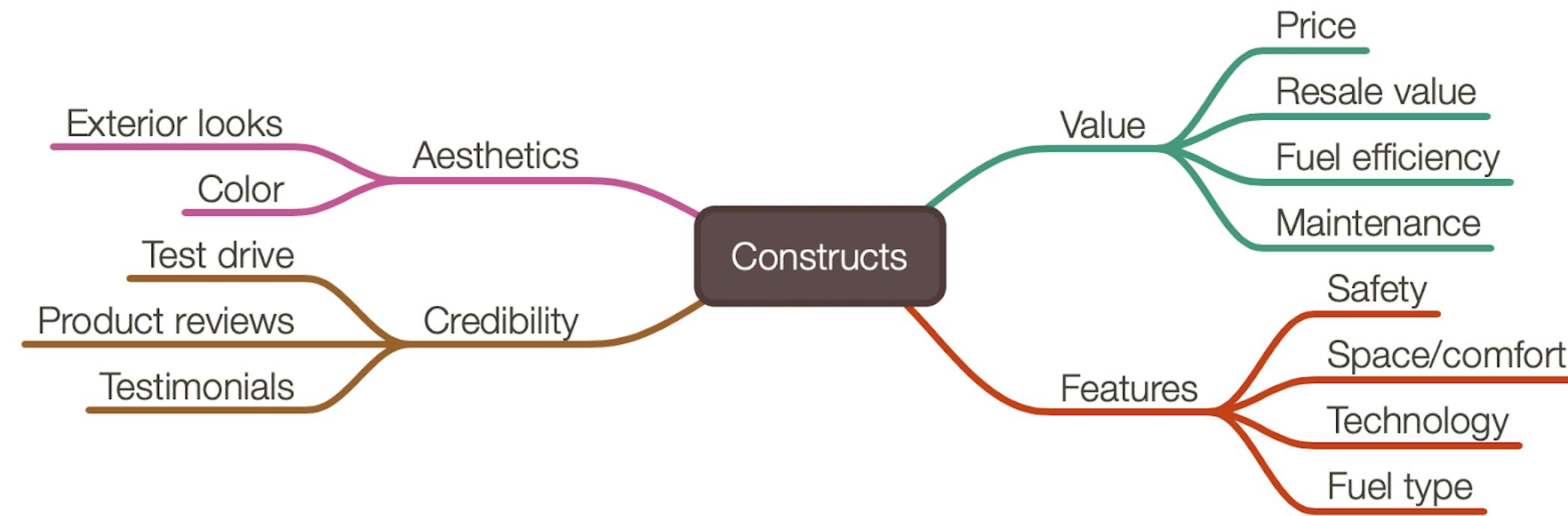
We can iteratively interpret and recalculate:

Loadings:

	MR1	MR2	MR3	MR4
Price	0.535			
Safety		0.356		
Exterior_Looks			-0.544	
Space_comfort		0.753		
Technology		0.349		
After_Sales_Service		0.528		
Resale_Value	0.724			
Fuel_Type		0.557		
Fuel_Efficiency	0.492			
Color			0.706	
Maintenance	0.603			
Test_drive			0.407	
Product_reviews	0.336		0.429	
Testimonials			0.677	

### Step 3: Scale construction

We inspect all factors and items that load to them:



To create a scale, we combine the items that load to that scale:

```
scale_value = cbind(EFA$Price,EFA$Resale_Value,EFA$Fuel_Efficiency,EFA$Maintenance)
```

## *Step 4: Test scale reliability*

**Recap:** Most commonly used measure of scale reliability is Cronbach's  $\alpha$ .

### **Cronbach's alpha** — Internal consistency

- »  $\alpha \geq .9$  — Excellent
- »  $.9 > \alpha \geq .8$  — Good
- »  $.8 > \alpha \geq .7$  — Acceptable
- »  $.7 > \alpha \geq .6$  — Questionable
- »  $.6 > \alpha \geq .5$  — Poor
- »  $.5 > \alpha$  — Unacceptable

To calculate Cronbach's  $\alpha$ :

```
alpha(scale_value, na.rm = TRUE)
```

This will produce:

**Reliability analysis**

Call: alpha(x = scale\_value, na.rm = TRUE)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.65	0.67	0.61	0.34	2	0.055	3.9	0.61	0.31

lower	alpha	upper	95% confidence boundaries
0.54	0.65	0.76	

**Reliability if an item is dropped:**

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r
scale_value	0.60	0.62	0.53	0.35	1.6	0.067	0.00556	
Resale_Value	0.54	0.55	0.45	0.29	1.2	0.081	0.00015	
Fuel_Efficiency	0.60	0.64	0.55	0.37	1.8	0.061	0.00467	
Maintenance	0.55	0.59	0.50	0.33	1.5	0.072	0.00233	

**med.r**

scale_value	0.33
Resale_Value	0.30
Fuel_Efficiency	0.38
Maintenance	0.33

To use the scale, we will average items out:

```
average <- rowMeans(scale_value)
scale_numerical <- data.frame(Average = average)
print(scale_numerical)
```

*Let's try it out!* → **Tutorial**

```

# Load libraries & data
install.packages("psych")
library(psych)
library(readr)
EFA <- read_csv("EFA.csv")

# Calculate factor matrix using Parallel Analysis
pa = fa.parallel(EFA, fm = 'minres', fa = 'fa')

# Factor rotation
fit = fa(data,nfactors = 3,rotate = "varimax",fm="minres")
print(fit$loadings,cutoff = 0.3)

# Calculate scale reliability
scale_value = cbind(EFA$Price,EFA$Resale_Value,EFA$Fuel_Efficiency,EFA$Maintenance)
alpha(scale_value, na.rm = TRUE)

# Use the scale
average <- rowMeans(scale_value)
scale_numerical <- data.frame(Average = average)
print(scale_numerical)

```