# Multivariate Analysis

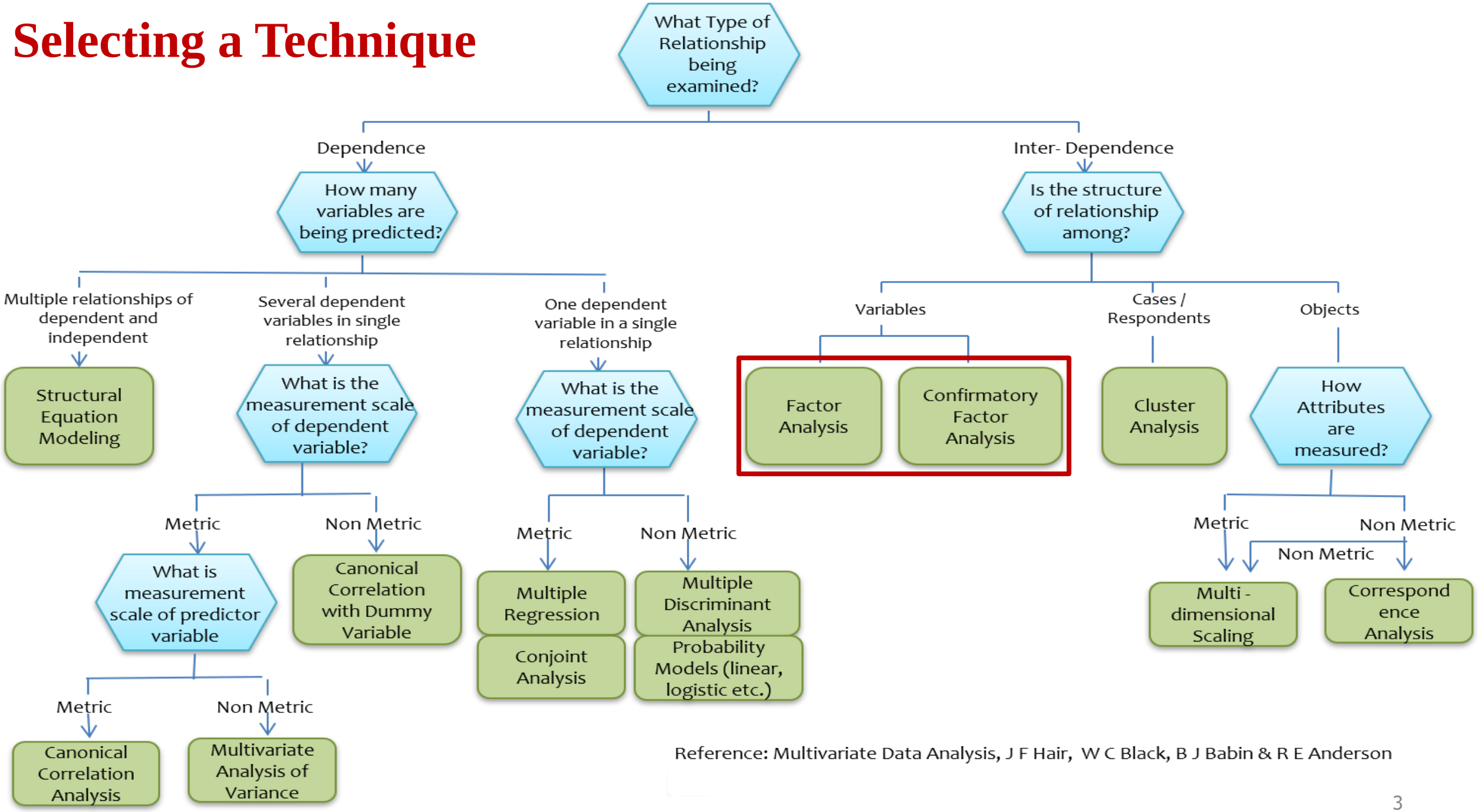
## Exploratory Factor Analysis Using R

**Dr. Hakeem–Ur–Rehman**
**IQTM–PU**

# Outline

- Selecting a Multivariate Technique
- Factor Analysis & It’s Types
- Exploratory Factor Analysis (EFA): Definitions
- EFA Using R: The R anxiety questionnaire (RAQ)
  - Initial preparation and analysis
  - Factor extraction Method
    - PCA
  - Factor Rotation & Factor Loadings
    - Orthogonal rotation
    - *Oblique rotation*
  - Factor Score
  - How to report factor analysis?

# Selecting a Technique



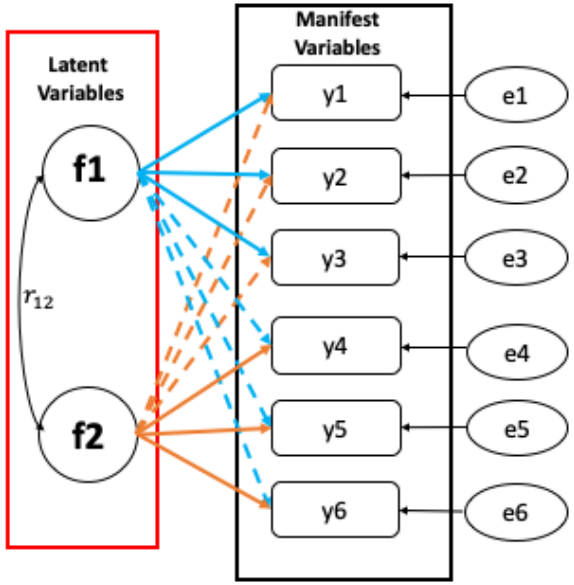

Reference: Multivariate Data Analysis, J F Hair, W C Black, B J Babin & R E Anderson

# Factor Analysis & It's Types

**Factor Analysis**

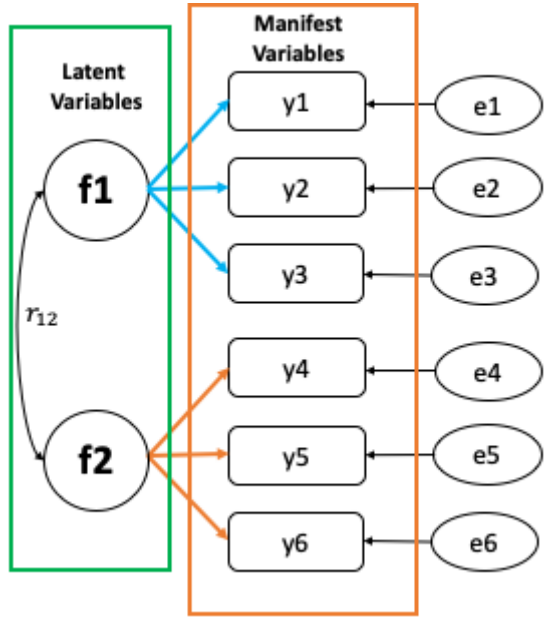
A method for *modeling observed variables, and their covariance structure,* in terms of a smaller number of *underlying unobservable (latent) "factors."*

## Exploratory




- EFA is used to *discover the factor structure of a construct* and examine its reliability.
- It is data driven.

## Confirmatory



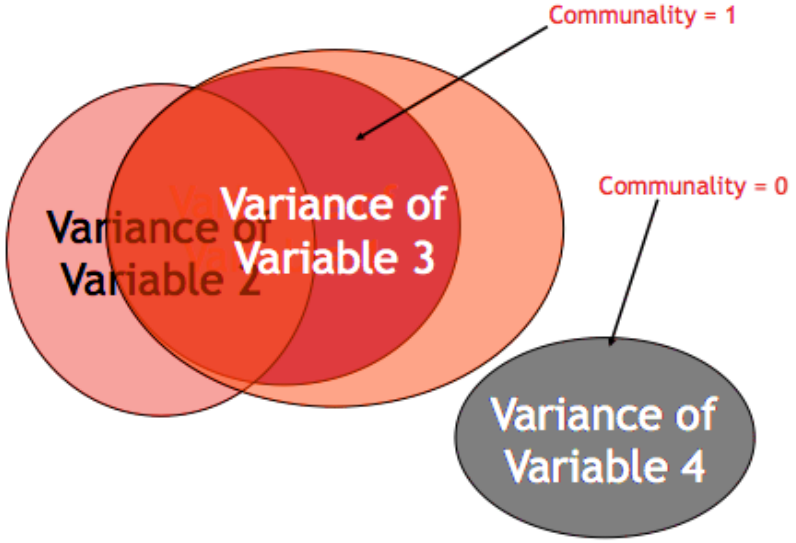

- CFA is used to *confirm the fit of the hypothesized factor structure to observed (sample) data*.
- It is theory driven.

# Exploratory Factor Analysis (EFA): Definitions

- ***Factor:*** A factor is an underlying dimension that account for several observed variables.

- ***Factor-loadings:*** Values that explain *how closely the variables are related to each one of the factors discovered*. They are also known as *factor-variable correlations*.
  - It is the absolute size (rather than the signs, plus or minus) of the loadings.

- ***Variance Types***
  - ***Common variance*** = overlapping variance between items (systematic variance)
  - ***Unique variance*** = variance only related to that item (error variance)
  - ***Communality*** **($h^2$)** – the common variance for the item (how much of each variable is accounted for by the underlying factor )
    - We can think it as $R^2$ for that item (The estimate of variance in each variable that is explained by the factors)
    - e.g. For variable A, communality of 0.65 indicates that 65% of the variance in variable A is explained in the terms of factors




***$h^2$* of the $i^{th}$ Variable = (*i*$^{th}$ factor loading of factor *A*)$^2$ + (*i*$^{th}$ factor loading of factor *B*)$^2$ + …**

- ***Eigen value (or latent root):*** *The sum of squared values of factor loadings relating to a factor* (referred to as Eigen Value)
  - Eigen value indicates the relative importance of each factor in accounting for the particular set of variables being analyzed.

# EFA Using R: The R anxiety questionnaire (RAQ)

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

| | | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| 1 | Statistics make me cry | O | O | O | O | O |
| 2 | My friends will think I'm stupid for not being able to cope with R | O | O | O | O | O |
| 3 | Standard deviations excite me | O | O | O | O | O |
| 4 | I dream that Pearson is attacking me with correlation coefficients | O | O | O | O | O |
| 5 | I don't understand statistics | O | O | O | O | O |
| 6 | I have little experience of computers | O | O | O | O | O |
| 7 | All computers hate me | O | O | O | O | O |
| 8 | I have never been good at mathematics | O | O | O | O | O |
| 9 | My friends are better at statistics than me | O | O | O | O | O |
| 10 | Computers are useful only for playing games | O | O | O | O | O |
| 11 | I did badly at mathematics at school | O | O | O | O | O |
| 12 | People try to tell you that R makes statistics easier to understand but it doesn't | O | O | O | O | O |
| 13 | I worry that I will cause irreparable damage because of my incompetence with computers | O | O | O | O | O |
| 14 | Computers have minds of their own and deliberately go wrong whenever I use them | O | O | O | O | O |
| 15 | Computers are out to get me | O | O | O | O | O |
| 16 | I weep openly at the mention of central tendency | O | O | O | O | O |
| 17 | I slip into a coma whenever I see an equation | O | O | O | O | O |
| 18 | R always crashes when I try to use it | O | O | O | O | O |
| 19 | Everybody looks at me when I use R | O | O | O | O | O |
| 20 | I can't sleep for thoughts of eigenvectors | O | O | O | O | O |
| 21 | I wake up under my duvet thinking that I am trapped under a normal distribution | O | O | O | O | O |
| 22 | My friends are better at R than I am | O | O | O | O | O |
| 23 | If I am good at statistics people will think I am a nerd | O | O | O | O | O |

**Data file → raq.dat**

```
> head(raqData)
  Q01 Q02 Q03 Q04 Q05 Q06 Q07 Q08 Q09 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23
1   4   5   2   4   4   4   3   5   5   4   5   4   4   4   4   3   5   4   3   4   4   4   1
2   5   5   2   3   4   4   4   4   1   4   4   3   5   3   2   3   4   4   3   2   2   2   4
3   4   3   4   4   2   5   4   4   4   4   3   3   4   2   4   3   4   3   5   2   3   4   4
4   3   5   5   2   3   3   2   4   4   2   4   4   4   3   3   3   4   2   4   2   2   2   3
5   4   5   3   4   4   3   3   4   2   4   4   3   3   4   4   4   4   3   3   2   4   2   2
6   4   5   3   4   2   2   2   4   2   3   4   2   3   3   1   4   3   1   5   1   3   5   2
```

```
> tail(raqData)
     Q01 Q02 Q03 Q04 Q05 Q06 Q07 Q08 Q09 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23
2566   4   4   3   3   2   3   2   2   4   4   4   4   4   5   4   3   3   4   3   2   2   3   3
2567   3   5   4   4   4   5   4   3   4   4   3   2   2   3   3   3   3   4   4   3   3   4   2
2568   3   5   4   3   2   2   2   3   4   5   4   1   2   2   2   1   1   3   5   1   2   3   5
2569   3   5   4   3   3   4   4   4   5   2   4   3   3   4   4   3   4   4   4   2   3   4   2
2570   3   5   3   4   2   4   3   2   5   3   3   4   4   3   3   3   3   4   4   2   2   4   3
2571   3   5   2   4   3   4   3   4   4   5   3   4   4   3   3   3   3   4   4   4   3   2   2
```

# EFA Using R: Initial preparation and analysis

1. **Correlation matrix:**
   - Check the correlations between variables
     - There are essentially two potential problems:
       1. Correlations that are not high enough
       2. Correlations that are too high.

**Bartlett's Test of Sphericity:**

- Compares an observed correlation matrix to the identity matrix
  - $H_0$: R Matrix = Identity Matrix (i.e., There is No Correlation Between Variables)
  - $H_1$: R Matrix ≠ Identity Matrix (i.e., There is a Correlation Between Variables)

```
$chisq
[1] 19334.49

$p.value
[1] 0

$df
[1] 253
```

- P-Value < 0.01; so, Bartlett's test is highly significant (i.e., *R*-matrix is not an identity matrix); $\chi^2(253) = 19334.49$ , $p < .01$, and therefore factor analysis is appropriate.

**The Determinant of the *R*-matrix should be greater than 0.00001**

```
> # Determinant of the R-matrix
> det(raqMatrix)
[1] 0.0005271037
```

# EFA Using R: Initial preparation and analysis…

**3. Sample Size:**

- KMO Test (Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy):

KMO → The ratio of the squared correlation between variables to the squared partial correlation between variables.

- The KMO statistic varies between 0 and 1.
- KMO = 0 indicates diffusion in the pattern of correlations → factor analysis is likely to be inappropriate
- KMO close to 1 indicates that patterns of correlations are relatively compact → factor analysis should yield distinct and reliable factors.

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = raqData)
Overall MSA =  0.93
MSA for each item =
 Q01  Q02  Q03  Q04  Q05  Q06  Q07  Q08  Q09  Q10  Q11  Q12  Q13  Q14  Q15  Q16  Q17  Q18
0.93 0.87 0.95 0.96 0.96 0.89 0.94 0.87 0.83 0.95 0.91 0.95 0.95 0.97 0.94 0.93 0.93 0.95
 Q19  Q20  Q21  Q22  Q23
0.94 0.89 0.93 0.88 0.77
```

| KMO Value | Level of Acceptance |
|---|---|
| Above 0.90 | Superb |
| 0.80 to 0.90 | Great |
| 0.70 to 0.80 | Good |
| 0.50 to 0.70 | Mediocre |
| Below 0.50 | Unacceptable |

Source: Field [56].

- Because Overall KMO = 0.93 & values of KMO for all the variables greater than 0.5 so, sample size and the data are adequate for factor analysis
- If any variables with KMO values below .5 then you should consider excluding them from the analysis.

# EFA Using R: Factor extraction

- **Methods of Factor Extraction**
  - PCA Method
  - Maximum Likelihood
  - Generalized Least Square
  - Alpha Factoring

# EFA Using R: Factor extraction

**Factor Loading Matrix**

```
Principal Components Analysis
Call: principal(r = raqData, nfactors = 23, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10  PC11  PC12  PC13  PC14  PC15  PC16  PC17  PC18  PC19  PC20  PC21  PC22  PC23 h2       u2 com
Q01  0.59  0.18 -0.22  0.12 -0.40 -0.11 -0.22 -0.08  0.01 -0.10  0.11 -0.12  0.30 -0.25  0.18  0.12 -0.05 -0.17  0.16 -0.01 -0.21  0.05  0.01  1 -1.1e-15 2.0
Q02 -0.30  0.55  0.15  0.01 -0.03 -0.38  0.19 -0.39  0.01 -0.12  0.30  0.27 -0.02  0.01 -0.24 -0.05 -0.08  0.00  0.01 -0.02 -0.02  0.03  0.02  1 -3.8e-15 4.1
Q03 -0.63  0.29  0.21 -0.07  0.02  0.00  0.01 -0.05  0.20  0.10  0.15  0.03  0.10  0.13  0.40 -0.06  0.43  0.08  0.09  0.05  0.01  0.00  0.05  1  6.7e-16 4.4
Q04  0.63  0.14 -0.15  0.15 -0.20 -0.12 -0.06  0.11 -0.11 -0.01 -0.03  0.34 -0.32 -0.17  0.12  0.31  0.19  0.05 -0.21  0.04  0.09 -0.02  0.02  1 -1.1e-15 4.9
Q05  0.56  0.10 -0.07  0.14 -0.42 -0.17 -0.06  0.11  0.24  0.09 -0.30  0.16  0.12  0.48 -0.07 -0.08 -0.04  0.01 -0.04  0.00 -0.02  0.02  0.01  1 -6.7e-16 5.2
Q06  0.56  0.10  0.57 -0.05  0.17  0.01  0.00  0.05  0.00  0.00 -0.13  0.20  0.24 -0.03  0.08  0.20 -0.14  0.05  0.09 -0.07  0.04 -0.32 -0.11  1 -4.4e-16 4.4
Q07  0.69  0.04  0.25  0.10  0.17 -0.08  0.05  0.03 -0.08  0.13 -0.27  0.20  0.04 -0.22  0.00 -0.23  0.03 -0.15  0.20  0.16  0.14  0.24  0.09  1 -4.4e-16 4.1
Q08  0.55  0.40 -0.32 -0.42  0.15  0.10 -0.07 -0.04  0.01 -0.05 -0.09  0.03 -0.01  0.04 -0.04  0.03  0.10  0.07  0.12 -0.15  0.06  0.16 -0.36  1 -1.3e-15 5.7
Q09 -0.28  0.63 -0.01  0.10  0.17 -0.27 -0.01 -0.03  0.16  0.32 -0.22 -0.37 -0.17 -0.07  0.12  0.11 -0.19 -0.02 -0.08 -0.03  0.04 -0.01  0.03  1 -1.3e-15 5.0
Q10  0.44  0.03  0.36 -0.10 -0.34  0.22  0.44 -0.03  0.37 -0.22 -0.11 -0.21 -0.17 -0.15 -0.07  0.03  0.07 -0.01  0.00  0.04 -0.03  0.02 -0.04  1  2.2e-16 7.7
Q11  0.65  0.25 -0.21 -0.40  0.13  0.18 -0.01  0.03  0.10 -0.14  0.00  0.03  0.02  0.03 -0.02  0.07 -0.05  0.07  0.07 -0.18  0.06  0.00  0.41  1 -1.3e-15 4.1
Q12  0.67 -0.05  0.05  0.25  0.04 -0.08 -0.14  0.08  0.01 -0.11  0.19 -0.07 -0.45  0.17  0.09 -0.10 -0.08  0.04  0.36  0.00 -0.04 -0.10 -0.02  1 -1.6e-15 3.8
Q13  0.67  0.08  0.28 -0.01  0.13  0.03 -0.21  0.05  0.08 -0.22  0.24 -0.08  0.01  0.12  0.14 -0.11 -0.06 -0.32 -0.30 -0.06  0.16  0.08 -0.05  1 -1.6e-15 4.2
Q14  0.66  0.02  0.20  0.14  0.08 -0.03 -0.10 -0.06 -0.14  0.16  0.08 -0.29  0.07  0.14 -0.37  0.25  0.34 -0.09  0.06  0.02  0.03 -0.01  0.05  1 -1.1e-15 4.3
Q15  0.59  0.01  0.12 -0.11 -0.07  0.29  0.32 -0.12 -0.27  0.41  0.15  0.09 -0.09  0.16  0.16  0.06 -0.12 -0.10 -0.04 -0.07 -0.19  0.10  0.00  1 -6.7e-16 5.6
Q16  0.68  0.01 -0.14  0.08 -0.32  0.00  0.12 -0.14 -0.19  0.15  0.16 -0.19  0.12 -0.08  0.06 -0.22 -0.03  0.22 -0.02 -0.04  0.35 -0.12 -0.01  1 -1.8e-15 4.0
Q17  0.64  0.33 -0.21 -0.34  0.10  0.05 -0.02  0.03 -0.04  0.02  0.01 -0.03 -0.01 -0.01 -0.05 -0.18  0.04 -0.04 -0.10  0.42 -0.15 -0.23 -0.01  1 -1.6e-15 4.3
Q18  0.70  0.03  0.30  0.13  0.15 -0.09 -0.10  0.06 -0.06 -0.12  0.05 -0.11  0.09  0.00  0.03 -0.01 -0.06  0.45 -0.15  0.08 -0.18  0.23  0.01  1 -1.1e-15 3.4
Q19 -0.43  0.39  0.10 -0.01 -0.15  0.07  0.05  0.68  0.02  0.16  0.29  0.04  0.06 -0.09 -0.16 -0.03 -0.06  0.01  0.05 -0.02  0.02  0.04 -0.02  1 -8.9e-16 3.5
Q20  0.44 -0.21 -0.40  0.30  0.33 -0.01  0.34  0.03  0.33  0.02  0.21  0.04  0.17  0.07  0.05  0.22 -0.09  0.00  0.04  0.18  0.10  0.06 -0.04  1  4.4e-16 3.7
Q21  0.66 -0.06 -0.19  0.28  0.24 -0.15  0.18  0.10  0.12  0.08 -0.02  0.04  0.03 -0.15 -0.04 -0.27  0.20 -0.03 -0.11 -0.31 -0.20 -0.13 -0.01  1 -2.2e-16 4.6
Q22 -0.30  0.47 -0.12  0.38  0.07  0.12  0.31  0.12 -0.41 -0.39 -0.19 -0.10  0.08  0.15  0.09  0.01  0.04 -0.06  0.02  0.00  0.01 -0.01  0.01  1  0.0e+00 7.2
Q23 -0.14  0.37 -0.02  0.51  0.02  0.62 -0.28 -0.22  0.18  0.08  0.00  0.13 -0.01 -0.07 -0.12 -0.06 -0.03  0.05 -0.03  0.01 -0.01 -0.02  0.00  1  1.1e-16 4.2

                       PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10 PC11 PC12 PC13 PC14 PC15 PC16 PC17 PC18 PC19 PC20 PC21 PC22 PC23
SS loadings           7.29 1.74 1.32 1.23 0.99 0.90 0.81 0.78 0.75 0.72 0.68 0.67 0.61 0.58 0.55 0.52 0.51 0.46 0.42 0.41 0.38 0.36 0.33
Proportion Var        0.32 0.08 0.06 0.05 0.04 0.04 0.04 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01
Cumulative Var        0.32 0.39 0.45 0.50 0.55 0.59 0.62 0.65 0.69 0.72 0.75 0.78 0.80 0.83 0.85 0.88 0.90 0.92 0.94 0.95 0.97 0.99 1.00
Proportion Explained  0.32 0.08 0.06 0.05 0.04 0.04 0.04 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01
Cumulative Proportion 0.32 0.39 0.45 0.50 0.55 0.59 0.62 0.65 0.69 0.72 0.75 0.78 0.80 0.83 0.85 0.88 0.90 0.92 0.94 0.95 0.97 0.99 1.00

Mean item complexity =  5
Test of the hypothesis that 23 components are sufficient.

The root mean square of the residuals (RMSR) is  0
 with the empirical chi square  0  with prob <  NA

Fit based upon off diagonal values = 1
```

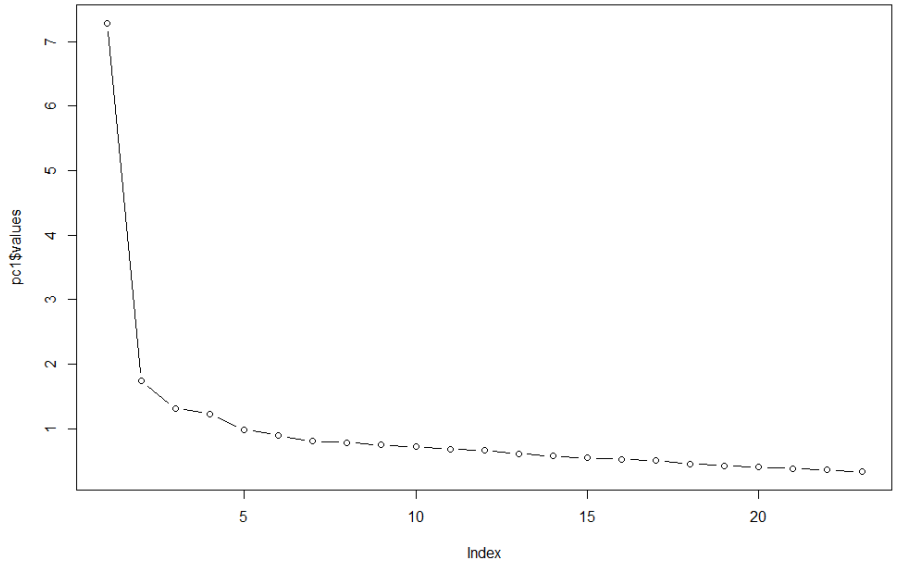
- **PCA Method:**
  - By extracting the factors, inspect two columns, labelled
    - h2 (communalities) → *All equal to 1 (Explained all of the variance in every variable)*
      - Because Factor extracted 23 = number of variables
      - When we extract fewer factors (or components) we'll have lower communalities.
    - u2 (amount of unique variance for each variable) → (1 − communalities) → all of the uniqueness's are 0

# EFA Using R: Factor Extraction

```
                        PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10 PC11 PC12 PC13 PC14 PC15 PC16 PC17 PC18 PC19 PC20 PC21 PC22 PC23
SS loadings            7.29 1.74 1.32 1.23 0.99 0.90 0.81 0.78 0.75 0.72 0.68 0.67 0.61 0.58 0.55 0.52 0.51 0.46 0.42 0.41 0.38 0.36 0.33
Proportion Var         0.32 0.08 0.06 0.05 0.04 0.04 0.04 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01
Cumulative Var         0.32 0.39 0.45 0.50 0.55 0.59 0.62 0.65 0.69 0.72 0.75 0.78 0.80 0.83 0.85 0.88 0.90 0.92 0.94 0.95 0.97 0.99 1.00
Proportion Explained   0.32 0.08 0.06 0.05 0.04 0.04 0.04 0.03 0.03 0.03 0.03 0.03 0.03 0.03 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01
Cumulative Proportion  0.32 0.39 0.45 0.50 0.55 0.59 0.62 0.65 0.69 0.72 0.75 0.78 0.80 0.83 0.85 0.88 0.90 0.92 0.94 0.95 0.97 0.99 1.00

Mean item complexity =  5
Test of the hypothesis that 23 components are sufficient.

The root mean square of the residuals (RMSR) is  0
 with the empirical chi square  0  with prob <  NA

Fit based upon off diagonal values = 1
```

- **Eigen Values:**
  - The eigenvalues associated with each factor represent the *variance explained by that particular linear component*.
  - **R** calls these SS loadings (sums of squared loadings)
    - **Factor 1** → explains 7.29 units of variance out of a possible 23 (the number of factors) so as a proportion this is 7.29/23 = 0.32; so, factor 1 explains 32% of the total variance.

- **Factor Extraction Criteria:**
  - According to Kaiser's criterion (eigenvalues > 1) → We can pick four components (or factors)
  - By Jolliffe's criterion (retain factors with eigenvalues > 0.7) → we can pick10 factors,
  - We should also consider the ***scree plot***.




**The evidence from the scree plot and from the eigenvalues suggests a four-component solution may be the best.**

# EFA Using R: Factor Extraction Tips

- To decide **how many factors to extract**:
  - look at the eigenvalues and the scree plot.

- If number of variables < 30 then using eigenvalues greater than 1 is OK (Kaiser's criterion) as long as your communalities are all over .7

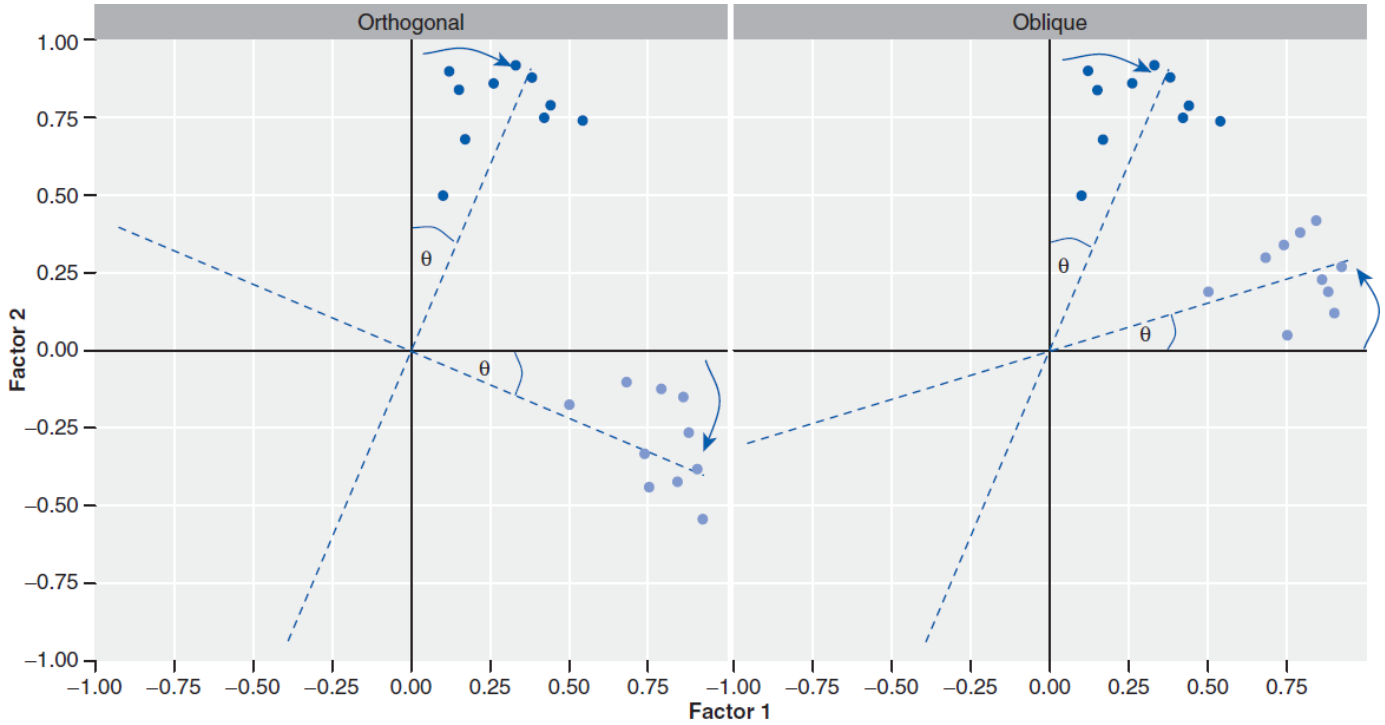
- Likewise, if your sample size exceeds 250 and the average of the communalities is .6 or greater then this is also fine.

- Alternatively, with 200 or more participants the scree plot can be used.

# EFA Using R: Factor Rotation

**After the factors extraction:**

- **Factor Loading:** Calculate to what degree variables load on these factors (i.e., calculate the loading of the variable on each factor).
  - Generally, most variables have *high loadings on the most important factor* and *small loadings on all other factors*.




- *Orthogonal rotation ensures that the factors remain independent or uncorrelated (perpendicular).*
- *Oblique rotation allow the factors to correlate (hence, do not perpendicular).*

***Rotation maximizes the loading of each variable on one of the extracted factors while minimizing the loading on all other factors.***

# EFA Using R: Factor Rotation → Orthogonal rotation (varimax)

- Having theoretical grounds to think that the factors are independent (unrelated) then
  - Choose one of the orthogonal rotations
    - **Varimax** (Recommended), quartimax, BentlerT & geominT
- pc3 <- principal(raqData, nfactors = 4, rotate = "varimax")
- print.psych(pc3, cut = 0.3, sort = TRUE)
  - *i. Remove loadings that are below the cut point*
  - *ii. Reorders the items to try to put them into their factors, using the sort option*

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

| | | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| 1 | Statistics make me cry | ○ | ○ | ○ | ○ | ○ |
| 2 | My friends will think I'm stupid for not being able to cope with R | ○ | ○ | ○ | ○ | ○ |
| 3 | Standard deviations excite me | ○ | ○ | ○ | ○ | ○ |
| 4 | I dream that Pearson is attacking me with correlation coefficients | ○ | ○ | ○ | ○ | ○ |
| 5 | I don't understand statistics | ○ | ○ | ○ | ○ | ○ |
| 6 | I have little experience of computers | ○ | ○ | ○ | ○ | ○ |
| 7 | All computers hate me | ○ | ○ | ○ | ○ | ○ |
| 8 | I have never been good at mathematics | ○ | ○ | ○ | ○ | ○ |
| 9 | My friends are better at statistics than me | ○ | ○ | ○ | ○ | ○ |
| 10 | Computers are useful only for playing games | ○ | ○ | ○ | ○ | ○ |
| 11 | I did badly at mathematics at school | ○ | ○ | ○ | ○ | ○ |
| 12 | People try to tell you that R makes statistics easier to understand but it doesn't | ○ | ○ | ○ | ○ | ○ |
| 13 | I worry that I will cause irreparable damage because of my incompetence with computers | ○ | ○ | ○ | ○ | ○ |
| 14 | Computers have minds of their own and deliberately go wrong whenever I use them | ○ | ○ | ○ | ○ | ○ |
| 15 | Computers are out to get me | ○ | ○ | ○ | ○ | ○ |
| 16 | I weep openly at the mention of central tendency | ○ | ○ | ○ | ○ | ○ |
| 17 | I slip into a coma whenever I see an equation | ○ | ○ | ○ | ○ | ○ |
| 18 | R always crashes when I try to use it | ○ | ○ | ○ | ○ | ○ |
| 19 | Everybody looks at me when I use R | ○ | ○ | ○ | ○ | ○ |
| 20 | I can't sleep for thoughts of eigenvectors | ○ | ○ | ○ | ○ | ○ |
| 21 | I wake up under my duvet thinking that I am trapped under a normal distribution | ○ | ○ | ○ | ○ | ○ |
| 22 | My friends are better at R than I am | ○ | ○ | ○ | ○ | ○ |
| 23 | If I am good at statistics people will think I am a nerd | ○ | ○ | ○ | ○ | ○ |

- All these questions related to using computers or **R**.
  - we can label this factor ***fear of computers***.

```
Principal Components Analysis
Call: principal(r = raqData, nfactors = 4, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
     item   RC3   RC1   RC4   RC2   h2   u2 com
Q06     6  0.80                    0.65 0.35 1.0
Q18    18  0.68  0.33              0.60 0.40 1.5
Q13    13  0.65                    0.54 0.46 1.6
Q07     7  0.64  0.33              0.55 0.45 1.7
Q14    14  0.58  0.36              0.49 0.51 1.8
Q10    10  0.55                    0.33 0.67 1.2
Q15    15  0.46                    0.38 0.62 2.6
Q20    20        0.68              0.48 0.52 1.1
Q21    21        0.66              0.55 0.45 1.5
Q03     3       -0.57        0.37 0.53 0.47 2.3
Q12    12  0.47  0.52              0.51 0.49 2.1
Q04     4  0.32  0.52  0.31        0.47 0.53 2.4
Q16    16  0.33  0.51  0.31        0.49 0.51 2.6
Q01     1        0.50  0.36        0.43 0.57 2.4
Q05     5  0.32  0.43              0.34 0.66 2.5
Q08     8              0.83        0.74 0.26 1.1
Q17    17              0.75        0.68 0.32 1.5
Q11    11              0.75        0.69 0.31 1.5
Q09     9                    0.65 0.48 0.52 1.3
Q22    22                    0.65 0.46 0.54 1.2
Q23    23                    0.59 0.41 0.59 1.4
Q02     2       -0.34        0.54 0.41 0.59 1.7
Q19    19       -0.37        0.43 0.34 0.66 2.2

                       RC3  RC1  RC4  RC2
SS loadings           3.73 3.34 2.55 1.95
Proportion Var        0.16 0.15 0.11 0.08
Cumulative Var        0.16 0.31 0.42 0.50
Proportion Explained  0.32 0.29 0.22 0.17
Cumulative Proportion 0.32 0.61 0.83 1.00

Mean item complexity =  1.8
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is  0.06
 with the empirical chi square  4006.15  with prob <  0

Fit based upon off diagonal values = 0.96
```

# EFA Using R: Factor Rotation → Orthogonal rotation (varimax)

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

| | | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| 1 | Statistics make me cry | ○ | ○ | ○ | ○ | ○ |
| 2 | My friends will think I'm stupid for not being able to cope with R | ○ | ○ | ○ | ○ | ○ |
| 3 | Standard deviations excite me | ○ | ○ | ○ | ○ | ○ |
| 4 | I dream that Pearson is attacking me with correlation coefficients | ○ | ○ | ○ | ○ | ○ |
| 5 | I don't understand statistics | ○ | ○ | ○ | ○ | ○ |
| 6 | I have little experience of computers | ○ | ○ | ○ | ○ | ○ |
| 7 | All computers hate me | ○ | ○ | ○ | ○ | ○ |
| 8 | I have never been good at mathematics | ○ | ○ | ○ | ○ | ○ |
| 9 | My friends are better at statistics than me | ○ | ○ | ○ | ○ | ○ |
| 10 | Computers are useful only for playing games | ○ | ○ | ○ | ○ | ○ |
| 11 | I did badly at mathematics at school | ○ | ○ | ○ | ○ | ○ |
| 12 | People try to tell you that R makes statistics easier to understand but it doesn't | ○ | ○ | ○ | ○ | ○ |
| 13 | I worry that I will cause irreparable damage because of my incompetence with computers | ○ | ○ | ○ | ○ | ○ |
| 14 | Computers have minds of their own and deliberately go wrong whenever I use them | ○ | ○ | ○ | ○ | ○ |
| 15 | Computers are out to get me | ○ | ○ | ○ | ○ | ○ |
| 16 | I weep openly at the mention of central tendency | ○ | ○ | ○ | ○ | ○ |
| 17 | I slip into a coma whenever I see an equation | ○ | ○ | ○ | ○ | ○ |
| 18 | R always crashes when I try to use it | ○ | ○ | ○ | ○ | ○ |
| 19 | Everybody looks at me when I use R | ○ | ○ | ○ | ○ | ○ |
| 20 | I can't sleep for thoughts of eigenvectors | ○ | ○ | ○ | ○ | ○ |
| 21 | I wake up under my duvet thinking that I am trapped under a normal distribution | ○ | ○ | ○ | ○ | ○ |
| 22 | My friends are better at R than I am | ○ | ○ | ○ | ○ | ○ |
| 23 | If I am good at statistics people will think I am a nerd | ○ | ○ | ○ | ○ | ○ |

```
Principal Components Analysis
Call: principal(r = raqData, nfactors = 4, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
     item   RC3   RC1   RC4   RC2   h2   u2 com
Q06     6  0.80                    0.65 0.35 1.0
Q18    18  0.68  0.33              0.60 0.40 1.5
Q13    13  0.65                    0.54 0.46 1.6
Q07     7  0.64  0.33              0.55 0.45 1.7
Q14    14  0.58  0.36              0.49 0.51 1.8
Q10    10  0.55                    0.33 0.67 1.2
Q15    15  0.46                    0.38 0.62 2.6
Q20    20        0.68              0.48 0.52 1.1
Q21    21        0.66              0.55 0.45 1.5
Q03     3       -0.57        0.37 0.53 0.47 2.3
Q12    12  0.47  0.52              0.51 0.49 2.1
Q04     4  0.32  0.52  0.31        0.47 0.53 2.4
Q16    16  0.33  0.51  0.31        0.49 0.51 2.6
Q01     1        0.50  0.36        0.43 0.57 2.4
Q05     5  0.32  0.43              0.34 0.66 2.5
Q08     8              0.83        0.74 0.26 1.1
Q17    17              0.75        0.68 0.32 1.5
Q11    11              0.75        0.69 0.31 1.5
                                0.65 0.48 0.52 1.3
                                0.65 0.46 0.54 1.2
                                0.59 0.41 0.59 1.4
                                0.54 0.41 0.59 1.7
                                0.43 0.34 0.66 2.2

                       RC3  RC1  RC4  RC2
SS loadings           3.73 3.34 2.55 1.95
Proportion Var        0.16 0.15 0.11 0.08
Cumulative Var        0.16 0.31 0.42 0.50
Proportion Explained  0.32 0.29 0.22 0.17
Cumulative Proportion 0.32 0.61 0.83 1.00

Mean item complexity =  1.8
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is  0.06
 with the empirical chi square  4006.15  with prob <  0

Fit based upon off diagonal values = 0.96
```

- All these questions related to different aspects of statistics.
  - we can label this factor ***fear of Statistics***.

# EFA Using R: Factor Rotation → Orthogonal rotation (varimax)

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

| | | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| 1 | Statistics make me cry | ○ | ○ | ○ | ○ | ○ |
| 2 | My friends will think I'm stupid for not being able to cope with R | ○ | ○ | ○ | ○ | ○ |
| 3 | Standard deviations excite me | ○ | ○ | ○ | ○ | ○ |
| 4 | I dream that Pearson is attacking me with correlation coefficients | ○ | ○ | ○ | ○ | ○ |
| 5 | I don't understand statistics | ○ | ○ | ○ | ○ | ○ |
| 6 | I have little experience of computers | ○ | ○ | ○ | ○ | ○ |
| 7 | All computers hate me | ○ | ○ | ○ | ○ | ○ |
| 8 | I have never been good at mathematics | ○ | ○ | ○ | ○ | ○ |
| 9 | My friends are better at statistics than me | ○ | ○ | ○ | ○ | ○ |
| 10 | Computers are useful only for playing games | ○ | ○ | ○ | ○ | ○ |
| 11 | I did badly at mathematics at school | ○ | ○ | ○ | ○ | ○ |
| 12 | People try to tell you that R makes statistics easier to understand but it doesn't | ○ | ○ | ○ | ○ | ○ |
| 13 | I worry that I will cause irreparable damage because of my incompetence with computers | ○ | ○ | ○ | ○ | ○ |
| 14 | Computers have minds of their own and deliberately go wrong whenever I use them | ○ | ○ | ○ | ○ | ○ |
| 15 | Computers are out to get me | ○ | ○ | ○ | ○ | ○ |
| 16 | I weep openly at the mention of central tendency | ○ | ○ | ○ | ○ | ○ |
| 17 | I slip into a coma whenever I see an equation | ○ | ○ | ○ | ○ | ○ |
| 18 | R always crashes when I try to use it | ○ | ○ | ○ | ○ | ○ |
| 19 | Everybody looks at me when I use R | ○ | ○ | ○ | ○ | ○ |
| 20 | I can't sleep for thoughts of eigenvectors | ○ | ○ | ○ | ○ | ○ |
| 21 | I wake up under my duvet thinking that I am trapped under a normal distribution | ○ | ○ | ○ | ○ | ○ |
| 22 | My friends are better at R than I am | ○ | ○ | ○ | ○ | ○ |
| 23 | If I am good at statistics people will think I am a nerd | ○ | ○ | ○ | ○ | ○ |

```
Principal Components Analysis
Call: principal(r = raqData, nfactors = 4, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
     item   RC3   RC1   RC4   RC2   h2   u2 com
Q06     6  0.80                    0.65 0.35 1.0
Q18    18  0.68  0.33              0.60 0.40 1.5
Q13    13  0.65                    0.54 0.46 1.6
Q07     7  0.64  0.33              0.55 0.45 1.7
Q14    14  0.58  0.36              0.49 0.51 1.8
Q10    10  0.55                    0.33 0.67 1.2
Q15    15  0.46                    0.38 0.62 2.6
Q20    20        0.68              0.48 0.52 1.1
Q21    21        0.66              0.55 0.45 1.5
Q03     3       -0.57        0.37 0.53 0.47 2.3
Q12    12  0.47  0.52              0.51 0.49 2.1
Q04     4  0.32  0.52  0.31        0.47 0.53 2.4
Q16    16  0.33  0.51  0.31        0.49 0.51 2.6
Q01     1        0.50  0.36        0.43 0.57 2.4
Q05     5  0.32  0.43              0.34 0.66 2.5
Q08     8              0.83        0.74 0.26 1.1
Q17    17              0.75        0.68 0.32 1.5
Q11    11              0.75        0.69 0.31 1.5
Q09     9                    0.65 0.48 0.52 1.3
Q22    22                    0.65 0.46 0.54 1.2
Q23    23                    0.59 0.41 0.59 1.4
Q02     2       -0.34        0.54 0.41 0.59 1.7
Q19    19       -0.37        0.43 0.34 0.66 2.2

                       RC1  RC4  RC2
                      .34 2.55 1.95
Proportion Var        0.10 0.15 0.11 0.08
Cumulative Var        0.16 0.31 0.42 0.50
Proportion Explained  0.32 0.29 0.22 0.17
Cumulative Proportion 0.32 0.61 0.83 1.00

Mean item complexity =  1.8
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is  0.06
 with the empirical chi square  4006.15  with prob <  0

Fit based upon off diagonal values = 0.96
```

- All these questions related to mathematics.
  - Label this factor ***fear of Mathematics***.

# EFA Using R: Factor Rotation → Orthogonal rotation (varimax)

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

| | | SD | D | N | A | SA |
|---|---|---|---|---|---|---|
| 1 | Statistics make me cry | ○ | ○ | ○ | ○ | ○ |
| 2 | My friends will think I'm stupid for not being able to cope with R | ○ | ○ | ○ | ○ | ○ |
| 3 | Standard deviations excite me | ○ | ○ | ○ | ○ | ○ |
| 4 | I dream that Pearson is attacking me with correlation coefficients | ○ | ○ | ○ | ○ | ○ |
| 5 | I don't understand statistics | ○ | ○ | ○ | ○ | ○ |
| 6 | I have little experience of computers | ○ | ○ | ○ | ○ | ○ |
| 7 | All computers hate me | ○ | ○ | ○ | ○ | ○ |
| 8 | I have never been good at mathematics | ○ | ○ | ○ | ○ | ○ |
| 9 | My friends are better at statistics than me | ○ | ○ | ○ | ○ | ○ |
| 10 | Computers are useful only for playing games | ○ | ○ | ○ | ○ | ○ |
| 11 | I did badly at mathematics at school | ○ | ○ | ○ | ○ | ○ |
| 12 | People try to tell you that R makes statistics easier to understand but it doesn't | ○ | ○ | ○ | ○ | ○ |
| 13 | I worry that I will cause irreparable damage because of my incompetence with computers | ○ | ○ | ○ | ○ | ○ |
| 14 | Computers have minds of their own and deliberately go wrong whenever I use them | ○ | ○ | ○ | ○ | ○ |
| 15 | Computers are out to get me | ○ | ○ | ○ | ○ | ○ |
| 16 | I weep openly at the mention of central tendency | ○ | ○ | ○ | ○ | ○ |
| 17 | I slip into a coma whenever I see an equation | ○ | ○ | ○ | ○ | ○ |
| 18 | R always crashes when I try to use it | ○ | ○ | ○ | ○ | ○ |
| 19 | Everybody looks at me when I use R | ○ | ○ | ○ | ○ | ○ |
| 20 | I can't sleep for thoughts of eigenvectors | ○ | ○ | ○ | ○ | ○ |
| 21 | I wake up under my duvet thinking that I am trapped under a normal distribution | ○ | ○ | ○ | ○ | ○ |
| 22 | My friends are better at R than I am | ○ | ○ | ○ | ○ | ○ |
| 23 | If I am good at statistics people will think I am a nerd | ○ | ○ | ○ | ○ | ○ |

```
Principal Components Analysis
Call: principal(r = raqData, nfactors = 4, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
     item   RC3   RC1   RC4   RC2   h2   u2 com
Q06     6  0.80                    0.65 0.35 1.0
Q18    18  0.68  0.33              0.60 0.40 1.5
Q13    13  0.65                    0.54 0.46 1.6
Q07     7  0.64  0.33              0.55 0.45 1.7
Q14    14  0.58  0.36              0.49 0.51 1.8
Q10    10  0.55                    0.33 0.67 1.2
Q15    15  0.46                    0.38 0.62 2.6
Q20    20        0.68              0.48 0.52 1.1
Q21    21        0.66              0.55 0.45 1.5
Q03     3       -0.57        0.37  0.53 0.47 2.3
Q12    12  0.47  0.52              0.51 0.49 2.1
Q04     4  0.32  0.52  0.31        0.47 0.53 2.4
Q16    16  0.33  0.51  0.31        0.49 0.51 2.6
Q01     1        0.50  0.36        0.43 0.57 2.4
Q05     5  0.32  0.43              0.34 0.66 2.5
                                   0.74 0.26 1.1
                                   0.68 0.32 1.5
                                   0.69 0.31 1.5
                                   0.65 0.48 0.52 1.3
                                   0.65 0.46 0.54 1.2
Q23    23                          0.59 0.41 0.59 1.4
Q02     2       -0.34               0.54 0.41 0.59 1.7
Q19    19       -0.37               0.43 0.34 0.66 2.2

                       RC3  RC1  RC4  RC2
SS loadings           3.73 3.34 2.55 1.95
Proportion Var        0.16 0.15 0.11 0.08
Cumulative Var        0.16 0.31 0.42 0.50
Proportion Explained  0.32 0.29 0.22 0.17
Cumulative Proportion 0.32 0.61 0.83 1.00

Mean item complexity =  1.8
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is  0.06
 with the empirical chi square  4006.15  with prob <  0

Fit based upon off diagonal values = 0.96
```

- All these items contain some component of social evaluation from friends.
  - Label this factor ***peer evaluation***.

# EFA Using R: Factor Rotation → oblique rotation (oblimin)

- Having theoretical grounds to think that the factors are related then
  - Choose one of the orthogonal rotations
    - **oblimin**, promax, simplimax, BentlerQ and geominQ

```
> # 2) Oblique rotation (oblimin)
> pc4 <-  principal(raqData, nfactors = 4, rotate = "oblimin")
> print.psych(pc4, cut = 0.3, sort = TRUE)
Principal Components Analysis
Call: principal(r = raqData, nfactors = 4, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
    item   TC1   TC4   TC3   TC2   h2   u2 com
Q06    6  0.87                    0.65 0.35 1.1
Q18   18  0.70                    0.60 0.40 1.1
Q07    7  0.64                    0.55 0.45 1.2
Q13   13  0.64                    0.54 0.46 1.1
Q10   10  0.57                    0.33 0.67 1.2
Q14   14  0.57                    0.49 0.51 1.3
Q12   12  0.45        0.43        0.51 0.49 2.0
Q15   15  0.40                    0.38 0.62 1.9
Q08    8         0.90              0.74 0.26 1.0
Q11   11         0.78              0.69 0.31 1.0
Q17   17         0.78              0.68 0.32 1.0
Q20   20               0.71        0.48 0.52 1.1
Q21   21               0.60        0.55 0.45 1.3
Q03    3              -0.51        0.53 0.47 1.8
Q04    4               0.41        0.47 0.53 2.6
Q16   16               0.41        0.49 0.51 2.4
Q01    1         0.33  0.40        0.43 0.57 2.4
Q05    5               0.34        0.34 0.66 2.7
Q22   22                     0.65  0.46 0.54 1.2
Q09    9                     0.63  0.48 0.52 1.4
Q23   23                     0.61  0.41 0.59 1.6
Q02    2              -0.36  0.51  0.41 0.59 1.9
Q19   19              -0.35  0.38  0.34 0.66 2.1

                       TC1  TC4  TC3  TC2
SS loadings           3.90 2.88 2.94 1.85
Proportion Var        0.17 0.13 0.13 0.08
Cumulative Var        0.17 0.29 0.42 0.50
Proportion Explained  0.34 0.25 0.25 0.16
Cumulative Proportion 0.34 0.59 0.84 1.00
```

**In our Study:**
All of our factors related to fear; It's likely that these will be correlated: people with fear of one thing might have fear of other things.

The **same four factors** seem to have emerged although they are in a different order.

- Factor 1 seems to represent **fear of computers**,
- factor 2 represents fear of **peer evaluation**,
- factor 3 represents **fear of statistics** and
- factor 4 represents **fear of mathematics**.

```
 With component correlations of
      TC1   TC4   TC3   TC2
TC1  1.00  0.44  0.36 -0.18
TC4  0.44  1.00  0.31 -0.10
TC3  0.36  0.31  1.00 -0.17
TC2 -0.18 -0.10 -0.17  1.00
```

- *These correlations exist tell us that the constructs measured can be interrelated*

# EFA Using R: Factor Score

```
# Factor Score
'------------'
pc5 <- principal(raqData, nfactors = 4, rotate = "oblimin", scores = TRUE)
pc5$scores

raqData <- cbind(raqData, pc5$scores)
raqData
```

- Factor scores can be used in this way to assess the relative fear of one person compared to another.
- After saving Factor score they can be further used as
    - Independent variable
    - Dependent variable
- Used in ANOVA / MANOVA / Regression Analysis
- Used in Cluster Analysis

# How to report factor analysis?

Summary of exploratory factor analysis results for the **R** anxiety questionnaire ($N$ = 2571)

| | Varimax rotated factor loadings | | | |
|---|---|---|---|---|
| *Item* | *Fear of computers* | *Fear of statistics* | *Peer evaluation* | *Fear of maths* |
| I have little experience of computers | **.80** | −.01 | −.07 | .10 |
| R always crashes when I try to use it | **.68** | .33 | −.08 | .13 |
| I worry that I will cause irreparable damage because of my incompetence with computers | **.65** | .23 | −.10 | .23 |
| All computers hate me | **.64** | .33 | −.08 | .16 |
| Computers have minds of their own and deliberately go wrong whenever I use them | **.58** | .36 | −.07 | .14 |
| Computers are useful only for playing games | **.55** | .00 | −.12 | .13 |
| Computers are out to get me | **.46** | .22 | −.19 | .29 |
| I can't sleep for thoughts of eigen vectors | −.04 | **.68** | −.14 | .08 |
| I wake up under my duvet thinking that I am trapped under a normal distribution | .29 | **.66** | −.07 | .16 |
| Standard deviations excite me | −.20 | **−.57** | .37 | −.18 |
| People try to tell you that R makes statistics easier to understand but it doesn't | **.47** | **.52** | −.08 | .10 |
| I dream that Pearson is attacking me with correlation coefficients | .32 | **.52** | .04 | .31 |
| I weep openly at the mention of central tendency | .33 | **.51** | −.12 | .31 |
| Statistics makes me cry | .24 | **.50** | .06 | .36 |
| I don't understand statistics | .32 | **.43** | .02 | .24 |
| I have never been good at mathematics | .13 | .17 | .01 | **.83** |
| I slip into a coma whenever I see an equation | .27 | .22 | −.04 | **.75** |
| I did badly at mathematics at school | .26 | .21 | −.14 | **.75** |
| My friends are better at statistics than me | −.09 | −.20 | **.65** | .12 |
| My friends are better at R than I am | −.19 | .03 | **.65** | −.10 |
| If I'm good at statistics my friends will think I'm a nerd | −.02 | .17 | **.59** | −.20 |
| My friends will think I'm stupid for not being able to cope with R | −.01 | −.34 | **.54** | .07 |
| Everybody looks at me when I use R | −.15 | −.37 | **.43** | −.03 |
| Eigenvalues | 3.73 | 3.34 | 1.95 | 2.55 |
| % of variance | 16.22 | 14.52 | 8.48 | 11.10 |
| $\alpha$ | .82 | .82 | .57 | .82 |

*Note*: Factor loadings over .40 appear in bold.

# QUESTIONS