Factor Analysis

```
> # Load data
> data("USArrests")
> head(USArrests)
           Murder Assault UrbanPop Rape
Alabama      13.2     236       58 21.2
Alaska       10.0     263       48 44.5
Arizona       8.1     294       80 31.0
Arkansas      8.8     190       50 19.5
California    9.0     276       91 40.6
Colorado      7.9     204       78 38.7
> # -------------------
> # Step 1. Preprocessing
> # -------------------
> # Standardize variables
> usarrests_scaled <- scale(USArrests)
> # -------------------
> # Step 2. Suitability Tests
> # -------------------
> install.packages("psych")
WARNING: Rtools is required to build R packages but is not currently insta
lled. Please download and install the appropriate version of Rtools before
proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Monir/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
also installing the dependencies 'mnormt', 'GPArotation'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/mnormt_2.1.1.
zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/GPArotation_2
025.3-1.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/psych_2.5.6.z
ip'
package 'mnormt' successfully unpacked and MD5 sums checked
package 'GPArotation' successfully unpacked and MD5 sums checked
package 'psych' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Monir\AppData\Local\Temp\RtmpOeBXFd\downloaded_packages
> install.packages("nFactors")
WARNING: Rtools is required to build R packages but is not currently insta
lled. Please download and install the appropriate version of Rtools before
proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Monir/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/nFactors_2.4.
1.2.zip'
Content type 'application/zip' length 222533 bytes (217 KB)
downloaded 217 KB

package 'nFactors' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Monir\AppData\Local\Temp\RtmpOeBXFd\downloaded_packages
> library(psych)

Attaching package: 'psych'

The following objects are masked from 'package:scales':

    alpha, rescale

The following objects are masked from 'package:ggplot2':
```

```
    %+%, alpha

> library(nFactors)

Attaching package: 'nFactors'

The following object is masked from 'package:lattice':

    parallel

> # Correlation matrix
> cor_mat <- cor(usarrests_scaled)
> print(cor_mat)
             Murder   Assault   UrbanPop      Rape
Murder   1.00000000 0.8018733 0.06957262 0.5635788
Assault  0.80187331 1.0000000 0.25887170 0.6652412
UrbanPop 0.06957262 0.2588717 1.00000000 0.4113412
Rape     0.56357883 0.6652412 0.41134124 1.0000000
```

Interpretation:

Murder and Assault are strongly correlated (0.80), indicating they might reflect a common factor. Ur banPop has a weaker relationship with Murder but moderate with Assault and Rape. Rape correlates moderately with Murder (0.56) and Assault (0.67), somewhat with UrbanPop (0.41).

```
> # Kaiser-Meyer-Olkin (KMO) measure
> KMO(cor_mat)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cor_mat)
Overall MSA =  0.65
MSA for each item =
  Murder  Assault UrbanPop     Rape
    0.62     0.64     0.50     0.78
```

Kaiser-Meyer-Olkin (KMO) Test

Overall KMO = 0.65 (mediocre adequacy but acceptable).Individual MSAs: Murder (0.62), Assault (0.64), UrbanPop (0.50), Rape (0.78).Values closer to 1 are better; 0.50 for UrbanPop is low but still acceptable.

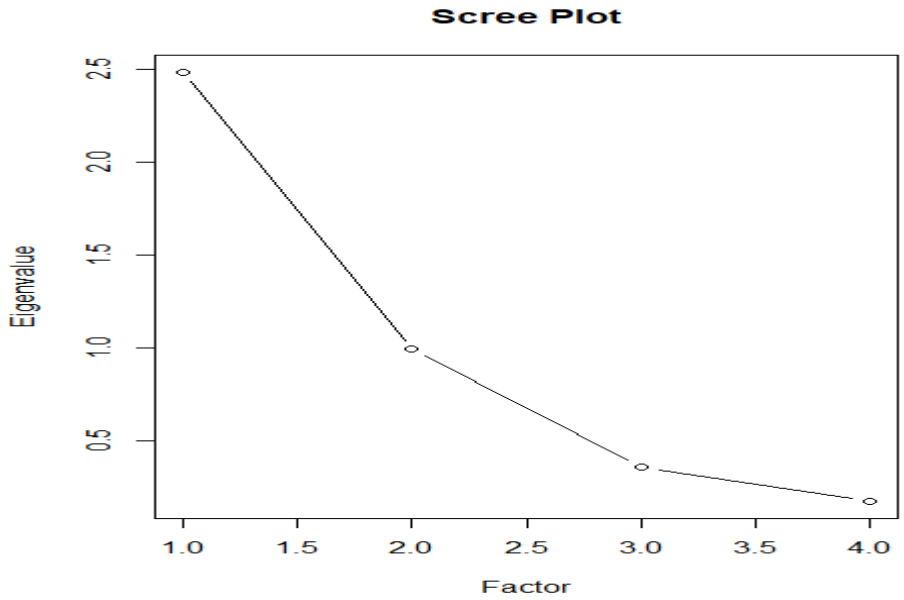
```
> # Bartlett's test of sphericity
> cortest.bartlett(cor_mat, n = nrow(usarrests_scaled))
$chisq
[1] 88.28815

$p.value
[1] 6.868423e-17

$df
[1] 6
```

Bartlett's Test of Sphericity

Chi-square = 88.29, df = 6, p-value ≈ 0 (6.87e-17). Highly significant, rejecting null hypothesis that variables are uncorrelated. Confirms factor analysis is appropriate

```
> # Scree plot (eigenvalues)
```

```
> ev <- eigen(cor_mat)$values
> plot(ev, type="b", main="Scree Plot", xlab="Factor", ylab="Eigenvalue")
```



Interpretation:

Eigenvalues of the correlation matrix are:
1st factor ≈ 2.5, 2nd ≈ 1.0, 3rd ≈ 0.4, 4th ≈ 0.15. According to the Kaiser criterion (eigenvalues > 1), two factors should be retained. Scree plot shows a clear "elbow" after the 2nd factor, confirming the choice of 2 factors.

```
> # -------------------
> # Step 3. Factor Analysis
> # -------------------
> # Suppose we extract 2 factors with varimax rotation
> fa_res <- fa(usarrests_scaled, nfactors=2, rotate="varimax", fm="ml")
> print(fa_res)
Factor Analysis using method =  ml
Call: fa(r = usarrests_scaled, nfactors = 2, rotate = "varimax", fm = "ml"
)
Standardized loadings (pattern matrix) based upon correlation matrix
           ML1  ML2   h2   u2 com
Murder    0.95 0.01 0.89 0.11 1.0
Assault   0.84 0.30 0.80 0.20 1.3
UrbanPop 0.06 0.67 0.46 0.54 1.0
Rape      0.59 0.55 0.65 0.35 2.0
```

Interpretation:

ML1 (Factor 1) loads heavily on violent crime indicators Murder (0.95) and Assault (0.84). **ML2** (Factor 2) loads more on UrbanPop (0.67) and moderately on Rape (0.55). Rape loads moderately on both factors, suggesting some cross-loading. Communalities (h^2) show how much variance of each variable is explained by the factors: Murder and Assault have high

communalities (>0.8), meaning most of their variance is explained. UrbanPop is less well explained (0.46). Uniqueness (u^2) indicates the variance unique to each variable, unexplained by factors. Complexity shows how many factors load significantly on each variable (1 = clean loading, >1 indicates cross-loading).

```
                       ML1  ML2
SS loadings           1.96 0.85
Proportion Var        0.49 0.21
Cumulative Var        0.49 0.70
Proportion Explained  0.70 0.30
Cumulative Proportion 0.70 1.00
```

Interpretation:

Factor 1 explains about 49% of variance, Factor 2 about 21%. Together, they explain 70% of the total variance in the data.

```
Mean item complexity =  1.3
Test of the hypothesis that 2 factors are sufficient.

df null model =  6  with the objective function =  1.89 with Chi Square =
88.29
df of  the model are -1  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  NA

The harmonic n.obs is  50 with the empirical chi square  0  with prob <  N
A
The total n.obs was  50  with Likelihood Chi Square =  0  with prob <  NA

Tucker Lewis Index of factoring reliability =  1.075
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 ML1  ML2
Correlation of (regression) scores with factors   0.96 0.80
Multiple R square of scores with factors          0.92 0.64
Minimum correlation of possible factor scores     0.84 0.28
```

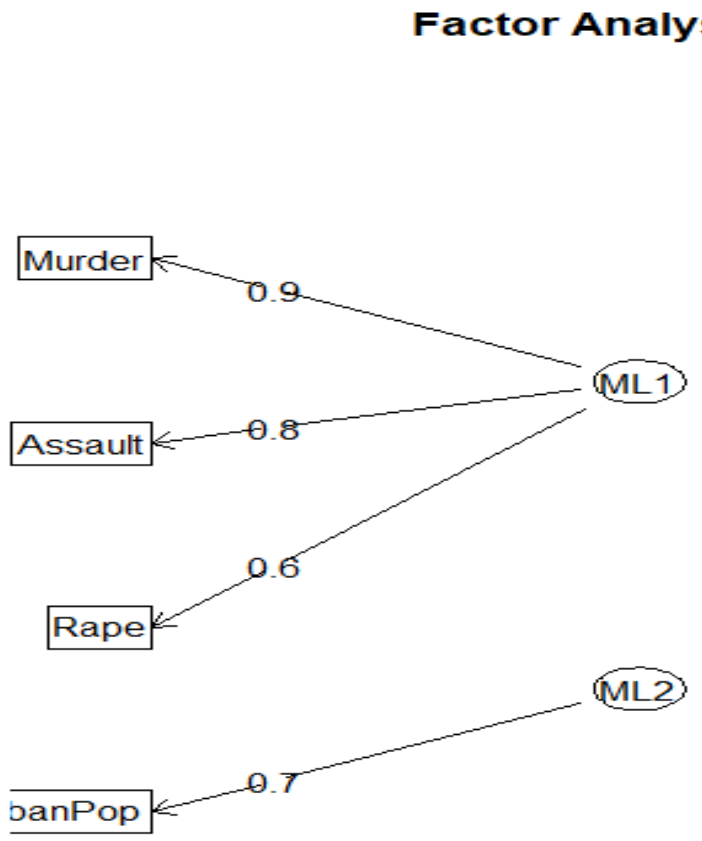
Interpretation:

Root Mean Square Residuals (RMSR) = 0, indicating perfect fit in this case (which is unusual, might be due to small dataset or method). **Tucker Lewis Index (TLI)** = 1.075 (above 1), suggesting excellent model fit (values >0.9 are good).

Factor score adequacy:

- Correlation between estimated factor scores and true factors is high: 0.96 (ML1), 0.80 (ML2).
- R^2 values show good accuracy of factor score prediction.
- Minimum correlation of factor scores: ML1 = 0.84, ML2 = 0.28 (ML2 less reliable).

```
> # -------------------
> # Step 4. Visualization
> # -------------------
> fa.diagram(fa_res)
```



The diagram visually links variables to their factors with arrows weighted by loadings. Murder, Assault, and Rape cluster strongly around Factor 1 (ML1), representing violent crimes. UrbanPop and Rape (less strongly) relate to Factor 2 (ML2), which likely reflects urbanization or social factors. This helps easily interpret and communicate the factor structure.