Included Analyses

- Exploratory Factor Analysis for 4 variables

Results

Exploratory Factor Analysis

Introduction

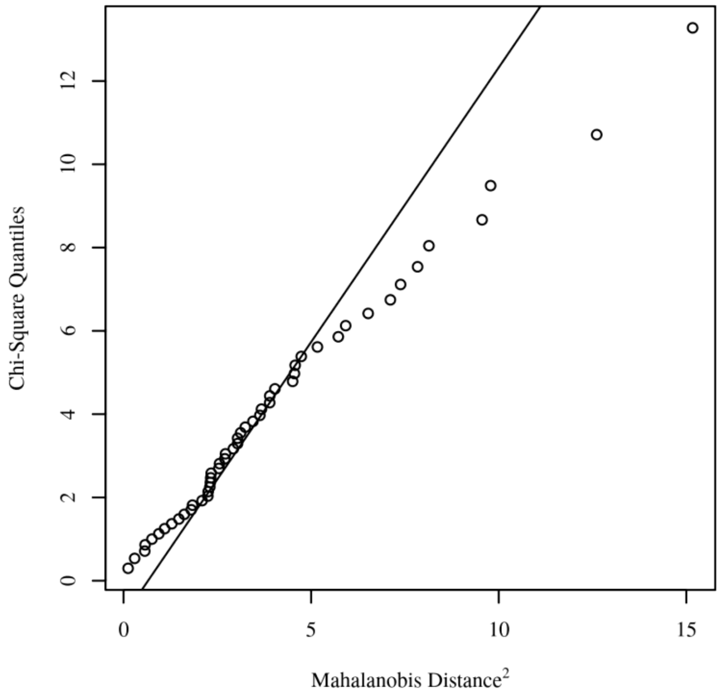
Exploratory factor analysis (EFA) was conducted for 4 variables using the Kaiser criterion for determining the number of factors to retain with varimax rotation.

Assumptions

Multivariate normality. To assess the assumption of multivariate normality, the squared Mahalanobis distances were calculated for the data and plotted against the quantiles of a Chi-square distribution (DeCarlo, 1997; Field, 2017). In the scatterplot, the solid line represents the theoretical quantiles of a normal distribution. Normality can be assumed if the points form a relatively straight line. The scatterplot for normality is presented in Figure 1.

Figure 1

Mahalanobis distance scatterplot testing multivariate normality



Factorability. To assess the factorability of the data, Pearson correlations were calculated to determine the intercorrelations for each variable. According to Tabachnick and Fidell (2019), correlation coefficients should exceed .30 in order to justify comprising the data into factors. All variables had at least one correlation coefficient greater than .30 and appear suitable for factor analysis.

Multicollinearity. Although variables should be intercorrelated with one another, variables that are too highly correlated can cause problems in EFA. To assess multicollinearity, the determinant of the correlation matrix was calculated. A determinant that is ≤ 0.00001 indicates that multicollinearity exists in the data (Field, 2017). The value of the determinant for the correlation matrix was 0.15, indicating that there is no multicollinearity in the data.

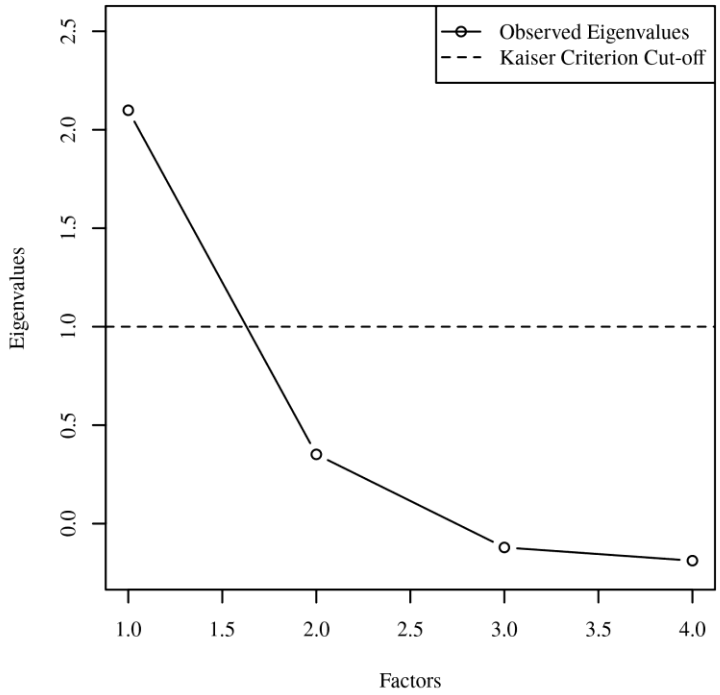
Results

The factor loadings were interpreted by taking the absolute value of each loading and implementing the criterion suggested by Comrey and Lee (2013). Values greater than .71 are considered excellent, values between .63 and .71 are very good, values between .55 and .63 are good, values between .45 and .55 are fair, and values between .32 and .45 are poor. Tabachnick and Fidell (2019) also recommend that .32 should be the minimum threshold used to identify significant factor loadings. These guidelines can help decide which variables to include for a given factor, but the cutoff used to determine which loadings should be included for each factor is a matter of preference to the researcher.

Determination of the number of factors. The Kaiser criterion was chosen for electing how many factors to retain. According to this rule, all factors that have an eigenvalue greater than one are retained for interpretation. The observed eigenvalues were extracted from the correlation matrix with the diagonal of the matrix being replaced by each variable's squared multiple correlations (Ledesma, 2007; Montanelli & Humphereys, 1976) to estimate each variable's communality (DiStefano et al., 2009; Stewart & Ware, 1992). Kaiser's eigenvalue-greater-than-one rule is a simple and common practice used throughout research (Floyd & Widaman, 1995; Ledesma & Valero-Mora, 2007; Yong & Pierce, 2013). Figure 2 shows the scree plot along with the Kaiser criterion for determining the number of significant factors. Looking at Figure 2, there was one factor that had an eigenvalue greater than one. As a result, one factor was used for the EFA.

Figure 2

Scree plot with the Kaiser criterion



Evaluating Sample Size. The sample size for exploratory factor analysis is very important when constructing repeatable and reliable factors. According to Osborne & Costello (2004), the most common guideline for the ratio of sample size to the number of variables (participant to item ratio) included should be at least 10 to 1, but some research indicates a minimum ratio of 5 to 1. The participant to item ratio for this analysis was approximately 12 to 1, where sample size was 50 and the number of variables included was 4. This indicates that the given sample size is sufficient to produce reliable results.

Factor summary. Factor 1 accounted for 54.05% of variance with an eigenvalue of 2.16. The one-factor model accounted for 54.05% of total variance in the data. The factor analysis summary is shown in Table 1. A Chi-square goodness of fit test was conducted to determine if the one-factor model fit the data perfectly based on an alpha value of .05, $\chi^2(2) = 9.15$, $p = .010$. This indicates that the one-factor model did not adequately depict the data.

Table 1

Eigenvalues, Percentages of Variance, and Cumulative Percentages for Factors for the 4 Item Variable Set

Factor	Eigenvalue	% of variance	Cumulative %
1	2.16	54.05	54.05

Note: $\chi^2(2) = 9.15$, $p = .010$.

Factor interpretation. The following variables had excellent loadings for Factor 1: Murder and Assault. The following variables had very good loadings for Factor 1: Rape. Any other loadings were insignificant for Factor 1. The factor analysis loadings are shown in Table 2.

Table 2

Factor Loadings From Exploratory Factor Analysis

	Factor loading	
Variable	1	Communality
Murder	0.82	0.67
Assault	0.98	0.96
UrbanPop		0.07
Rape	0.68	0.47

Note: Factor loadings < .32 are suppressed.

Evaluating the factor structure. According to Costello and Osborne (2005), examining the communality of each variable, checking for crossloadings across multiple factors, and inspecting the number of strong loadings for each factor are good ways to analyze the validity of the factor structure. The following variables had a low communality (< .40): UrbanPop. This indicates that an additional factor may need to be explored and the factor structure does not adequately describe the data (Costello & Osborne, 2005). Crossloadings occur when there are loadings (> .32) for a single variable across multiple factors. There were no variables with crossloadings, which suggests a factor structure that is simple and easy to interpret. Each factor had at least three significant loadings (> .32), which is indicative of a strong and solid factor (Osborne & Costello, 2005). Costello and Osborne (2005) also suggest dropping variables with low communality, crossloadings and any variable that is the only significant loading on a factor which may prevent a weak factor structure and alleviate these problems.

References

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press. https://doi.org/10.4324/9781315827506

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation, 10*(1), 7.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307. https://doi.org/10.1037/1082-989X.2.3.292

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics: North American edition.* Sage Publications

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, *7*(3), 286. https://doi.org/10.1037/1040-3590.7.3.286

Intellectus Statistics [Online computer software]. (2023). Intellectus Statistics. https://statistics.intellectus360.com

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research & evaluation, 12*(2), 1-11.

Montanelli, R. G., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika, 41*(3), 341-348. https://doi.org/10.1007/BF02293559

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. Practical assessment, research & evaluation, 9(11), 8. https://doi.org/10.7275/ktzq-jq66

Stewart, A. L., & Ware, J. E. (Eds.). (1992). *Measuring functioning and well-being: the medical outcomes study approach.* Duke University Press.

Tabachnick, B. G. & Fidell, L. S., (2019). *Using multivariate statistics.* Pearson Education.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology, 9*(2), 79-94. https://doi.org/10.20982/tqmp.09.2.p079

Glossaries

Exploratory Factor Analysis

Exploratory factor analysis (EFA) is a statistical technique to identify underlying relationships between scale variables. It is commonly used to reduce a dataset to a smaller set of summary variables. This is an investigative analysis that allows the researcher to explore theoretical structures (factors) that are represented by a set of variables. There are a several important decisions a researcher needs to make for EFA including: the method used for choosing the number of factors to retain, the rotation method utilized for the factor analysis, and a reasonable cutoff point to determine which variables to include for a given factor.

Chi-Squared Statistic (χ^2): A test statistic based on the χ^2 distribution. Used with the *df* to calculate a *p*-value.

Communality: The percent of explained variance for a variable for all the factors combined. It is used to help determine the reliability of the factor structure.

Crossloading: A variable that has loadings above a given cutoff (> .32) across multiple factors. Crossloadings can make factors difficult to interpret.

Degrees of Freedom (*df*): Refers to the number of values used to compute a statistic; used in conjunction with a test-statistic to calculate the *p*-value.

Determinant: A value calculated from a square (n × n) matrix with useful mathematical properties.

Eigenvalue: The variance that is accounted for by a given factor.

Factor: A set of observed variables that have strong relationships with one another or have a similar pattern.

Factor Loadings: Demonstrates the relationship each variable has to a given factor. Loadings can also be interpreted as a Pearson correlation coefficient with the factor it represents.

Factorability: The assumption that there is at least some level of correlation among the variables so that coherent factors can be identified.

Kaiser Criterion: A method for determining the number of factors to be retained. The number of factors that have an eigenvalue greater than one determines how many factors should be kept for the factor analysis.

Multicollinearity: A state of very high intercorrelations or inter-associations among a set of variables.

Probability Value (*p*): The probability of observing the test statistic under the null hypothesis.

Parallel Analysis: A method for determining the number of factors to be retained. It compares the observed eigenvalues for some given data with the eigenvalues of some randomly generated normal uncorrelated data. The number of factors with a higher observed eigenvalue determines how many factors should be kept for the factor analysis.

Promax Rotation: A rotation method for factor analysis that allows for correlated factors. This rotation method can help prevent crossloadings and is recommended for factor analysis.

Scree Plot: A plot that shows the explained variance (eigenvalue) by each factor. It is commonly used for determining the number of factors to include in factor analysis.

Squared Multiple Correlations: Used in Exploratory Factor Analysis to estimate each variable's communality. Also, referred to as R^2 in multiple linear regression. A value from 0 to 1 that shows the fraction of variance explained.

Varimax Rotation: The most common rotation method for factor analysis that creates uncorrelated factors. This rotation method can help prevent crossloadings, but it can also cause the loss of valuable information if the factors should be correlated.

Raw Output

Exploratory Factor Analysis for 4 Variables with 1 Factors and Varimax Rotation

```
Included Variables:
Murder, Assault, UrbanPop, and Rape

Sample Size (Complete Cases):
N = 50

Multicollinearity:
Determinant of the Correlation Matrix = 0.152

Kaiser Criterion: Observed Eigenvalues > 1
```

Factor	Observed Eigenvalues	% of Variance	Cumulative %
1	2.099	98.007	98.007
2	0.352	16.419	114.426
3	-0.121	0.00000	108.777
4	-0.188	0.00000	100.000

Note. Negative eigenvalues and variance that exceeds 100% can be expected, since squared multiple correlations were used to replace the diagonal of the correlation matrix.

Factor Structure Summary:

Factor	Eigenvalue	% of Variance	Cumulative %
1	2.162	54.046	54.046

Factor Loadings:

	Factor loading	
Variable	1	Communality
Murder	0.818	0.668
Assault	0.979	0.958
UrbanPop	0.262	0.0686
Rape	0.683	0.466

Chi-Square Goodness-of-Fit Test:
$\chi^2(2) = 9.15$, $p = 0.0103$