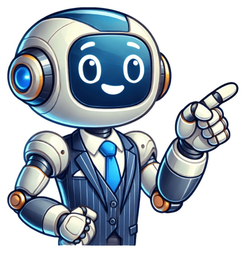
Click to prove you're human



Share copy and redistribute the material in any medium or format for any purpose, even commercially. Adapt remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms. Attribution You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. ShareAlike If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. No additional restrictions You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits. You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation . No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material. Evaluate if data are appropriate for Factor analysisThe goal of Factor Analysis (and Principal Components Analysis) is toreduce the dimensionality of the data with minimal loss of informationby identifying and using the structure in the correlation matrix of thevariables included in the analysis. The researcher will often try tolink the original variables (or items) to an underlying factorand provide a descriptive label for each. First, go to the Data > Manage tab, selectexamples from the Load data of typedropdown, and press the Load button. Then select thetoothpaste dataset. The dataset contains information from60 consumers who were asked to respond to six questions to determinetheir attitudes towards toothpaste. The scores shown for variables v1-v6indicate the level of agreement with statements on a 7-point scale where1 = strongly disagree and 7 = strongly agree.The first step in factor analysis is to determine if the data has therequired characteristics. Data with limited or no correlation betweenthe variables are not appropriate for factor analysis. We will use threecriteria to test if the data are suitable for factor analysis: Bartlett,KMO, and Collinearity for each variableThe KMO and Bartlett test evaluate all available data together. A KMOvalue over 0.5 and a significance level for the Bartletts test below0.05 suggest there is substantial correlation in the data. Variablecollinearity indicates how strongly a single variable is correlated withother variables. Values above 0.4 are considered appropriate. KMOmeasures can also be calculated for each variable. Values above 0.5 areacceptable.As can be seen in the output from Multivariate > Factor >Pre-factor below, Bartletts test statistic is large andsignificant (p.value close to 0) as desired. The Kaiser-Meyer-Olkin(KMO) measure is larger than .5 and thus acceptable. The variablecollinearity values are above 0.4 and the KMO values are above 0.5 soall variables can be used in the analysis.To replicate the results shown in the screenshot make sure you havethe toothpaste data loaded. Then select variablesv1 through v6 and click theEstimate button or press CTRL-enter(CMD-enter on mac) to generate results.The next step is to determine the number of factors needed to capturethe structure underlying the data. Factors that do not capture even asmuch variance as could be expected by chance are generally omitted fromfurther consideration. These factors have eigenvalues < 1 in theoutput.A further criteria that is often used to determine the number offactors is the scree-plot. This is a plot of the eigenvalues against thenumber of factors, in order of extraction. Often a break orelbow is visible in the plot. Factors up to and including thiselbow are selected for further analysis if they all have eigenvaluesabove 1. A set of factors that explain more than 70% of the variance inthe original data is generally considered acceptable. The eigenvaluesfor all factors are shown above. Only two factors have eigenvalues above1.At first glance the scree-plot of the Eigenvalues shown below seemsto suggest that 3 factors should be extracted (i.e., look for theelbow). The bar plot confirms this insight, i.e., the change inEigenvalues between factors 1 and 2 is small but the drop-off from 2 to3 is much larger. However, because the value for the third factor isless than one we will extract only 2 factors.The increase in cumulative % explained variance is relatively smallgoing from 2 to 3 factors (i.e., from 82% to 90%). This is confirmed bythe fact that the eigenvalue for factor 3 is smaller than 1 (0.44).Again, we choose 2 factors. The first 2 factors capture 82% of thevariance in the original data which is excellent. The pre-factor analysis diagnostics are calculated using PrincipalComponents Analysis (PCA). The correlation matrix used as input for PCAcan be calculated for variables of type numeric,integer, date, and factor. Whenvariables of type factor are included theAdjust for categorical variables box should be checked.When correlations are estimated with adjustment, variables that are oftype factor will be treated as (ordinal) categoricalvariables and all other variables will be treated as continuous. Add code toReport> Rmd to (re)create the analysis by clicking the icon on the bottomleft of your screen or by pressing ALT-enter on yourkeyboard.If a plot was created it can be customized using ggplot2commands or with gridExtra. See example below andData> Visualize for details.plot(result, plots = "scree", custom = TRUE) + labs(caption = "Data used from ...") For an overview of related R-functions used by Radiant to conductfactor analysis seeMultivariate> Factor.The key functions used in the pre_factor tool arecor from the stats package, eigenfrom base, and cortest.bartlett andKMO from the psych package. Vincent Nijs (2024) Statistical measure to determine how suited data is for factor analysis.The KaiserMeyerOlkin (KMO) test is a statistical measure to determine how suited data is for factor analysis. The test measures sampling adequacy for each variable in the model and the complete model. The statistic is a measure of the proportion of variance among variables that might be common variance. The higher the proportion, the higher the KMO-value, the more suited the data is to factor analysis.[1]Henry Kaiser introduced a Measure of Sampling Adequacy (MSA) of factor analytic data matrices in 1970.[2] Kaiser and Rice then modified it in 1974.[3]The measure of sampling adequacy is calculated for each indicator as M S A j = k j r j k 2 k j r j k 2 + k j p j k 2 {\displaystyle MSA_{j}={\frac {\displaystyle \sum _{keq j}r_{jk}^{2}}{\displaystyle \sum _{keq j}r_{jk}^{2}+\sum _{keq j}p_{jk}^{2}}}} and indicates to what extent an indicator is suitable for a factor analysis.The KaiserMeyerOlkin criterion is calculated and returns values between 0 and 1. K M O = j k r j k 2 j k r j k 2 + j k p j k 2 {\displaystyle KMO={\frac {\displaystyle {\underset {jeq k}{\sum \sum }}r_{jk}^{2}}{\displaystyle {\underset {jeq k}{\sum \sum }}r_{jk}^{2}+{\underset {jeq k}{\sum \sum }}p_{jk}^{2}}}} Here r j k {\displaystyle r_{jk}} is the correlation between the variable in question and another, and p j k {\displaystyle p_{jk}} is the partial correlation.This is a function of the squared elements of the `image' matrix compared to the squares of the original correlations. The overall MSA as well as estimates for each item are found. The index is known as the KaiserMeyerOlkin (KMO) index.[4]In flamboyant fashion, Kaiser proposed that a KMO>0.9 was marvelous, in the 0.80s, meritorious, in the 0.70s, middling, in the 0.60s, mediocre, in the 0.50s, miserable, and less than 0.5 would be unacceptable. [3] In general, KMO values between 0.8 and 1 indicate the sampling is adequate. KMO values less than 0.6 indicate the sampling is not adequate and that remedial action should be taken. In contrast, others set this cutoff value at 0.5.[5] A KMO value close to zero means that there are large partial correlations compared to the sum of correlations. In other words, there are widespread correlations which would be a large problem for factor analysis.[1]An alternative measure of whether a matrix is factorable is the Bartlett test, which tests the degree that the matrix deviates from an identity matrix.[1]If the following is run in R with the library(psych)library(psych)set.seed(5L)five.samples .9 were marvelous, in the .80s,meritorious, in the .70s, middling, in the .60s, mediocre, in the .50s,miserable, and less than .5, unacceptable. Hair et al. (2006) suggestaccepting a value > 0.5. Values between 0.5 and 0.7 are mediocre, and valuesbetween 0.7 and 0.8 are good.Variables with individual KMO values below 0.5 could be considered forexclusion them from the analysis (note that you would need to re-compute theKMO indices as they are dependent on the whole dataset). This function is a wrapper around the KMO and the cortest.bartlett()functions in the psych package (Revelle, 2016).Revelle, W. (2016). How To: Use the psych package for Factor Analysisand data reduction.Bartlett, M. S. (1951). The effect of standardization on a Chi-squareapproximation in factor analysis. Biometrika, 38(3/4), 337-344.Kaiser, H. F. (1970). A second generation little jiffy.Psychometrika, 35(4), 401-415.Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. Educationaland psychological measurement, 34(1), 111-117.Kaiser, H. F. (1974). An index of factorial simplicity.Psychometrika, 39(1), 31-36. check_clusterstructure(). library(performance)check_factorstructure(mtcars)#> # Is the data suitable for Factor Analysis?#> #> #> - Sphericity: Bartlett's test of sphericity suggests that there is sufficient significant correlation in the data for factor analysis (Chisq(55) = 408.01, p < .001).#> - KMO: The Kaiser, Meyer, Olkin (KMO) overall measure of sampling adequacy suggests that data seems appropriate for factor analysis (KMO = 0.83). The individual KMO scores are: mpg (0.93), cyl (0.90), disp (0.76), hp (0.84), drat (0.95), wt (0.74), qsec (0.74), vs (0.91), am (0.88), gear (0.85), carb (0.62).# One can also pass a correlation matrixr # Is the data suitable for Factor Analysis?#> #> #> - Sphericity: Bartlett's test of sphericity suggests that there is sufficient significant correlation in the data for factor analysis (Chisq(55) = 408.01, p < .001).#> - KMO: The Kaiser, Meyer, Olkin (KMO) overall measure of sampling adequacy suggests that data seems appropriate for factor analysis (KMO = 0.83). The individual KMO scores are: mpg (0.93), cyl (0.90), disp (0.76), hp (0.84), drat (0.95), wt (0.74), qsec (0.74), vs (0.91), am (0.88), gear (0.85), carb (0.62). We have already discussed factor analysis in the previous article, and how it should be conducted using SPSS. In this article, we will be discussing how the output of Factor analysis can be interpreted. The first output from the analysis is a table of descriptive statistics for all the variables under investigation. Typically, the mean, standard deviation, and the number of respondents (N) who participated in the survey are given. The mean value describes the characteristics of the most common response among the stated dataset. Therefore there is no minimum value required. Looking at the mean values in Table 1 below, one can conclude that the respectability of product is the most important variable that influences customers to buy the product. The lowest value of 2.42 for cost of the product indicates that the respondents approximately strongly disagree on the cost of product role. All the variables roles in consumers decisions to buy a product can be interpreted in a similar way. Table 1: Descriptive statistics The next output from the analysis is the correlation coefficient. A correlation matrix is simply a rectangular array of numbers that gives the correlation coefficients between a single variable and every other variable in the investigation. The correlation coefficient between a variable and itself is always 1, hence the principal diagonal of the correlation matrix contains 1s (See Red Line in Table 2 below). The correlation coefficients above and below the principal diagonal are the same. The determinant of the correlation matrix is shown at the foot of the table below. With respect to the correlation matrix if any pair of variables has a value less than 0.5, consider dropping one of them from the analysis. For this factor, analysis needs to be reperformed with the exclusion of pair of variables with less than 0.5 value. The off-diagonal elements (The values on the left and right sides of the diagonal in the table below) should all be very small (close to zero) in a good model. Table 2: Correlation matrix The KMO measures the sampling adequacy (which determines if the responses given with the sample are adequate or not) which should be close to 0.5 for satisfactory factor analysis to proceed. Kaiser (1974) recommends 0.5 (value for KMO) as a minimum (barely accepted), values between 0.7-0.8 are acceptable, and values above 0.9 are superb. Looking at the table below, the KMO measure is 0.417, which is close to 0.5 and therefore can be barely accepted (Table 3). There is no significant answer to the question How many cases respondents do I need to factor analysis?, and methodologies differ. A common rule is to suggest that a researcher has at least 10-15 participants per variable. Fiedel (2005) says that in general over 300 Respondents for sampling analysis is probably adequate. There is universal agreement that factor analysis is inappropriate when the sample size is below 50. Bartletts test is another indication of the strength of the relationship among variables. This tests the null hypothesis that the correlation matrix is an identity matrix. An identity matrix is a matrix in which all of the diagonal elements are 1 (See Table 1) and all off-diagonal elements (term explained above) are close to 0. You want to reject this null hypothesis. From the same table, we can see that Bartletts Test Of Sphericity is significant (0.12). That is, the significance is less than 0.05. In fact, it is actually 0.012, i.e. the significance level is small enough to reject the null hypothesis. This means that the correlation matrix is not an identity matrix. Table 3: KMO and Barletts test The next item from the output is a table of commonalities which shows how much of the variance (i.e. the communality value which should be more than 0.5 to be considered for further analysis. Else these variables are to be removed from further steps of factor analysis) in the variables has been accounted for by the extracted factors. For instance over 90% of the variance in Quality of product is accounted for, while 73.5% of the variance in Availability of product is accounted for (Table 4). Table 4: Communalities Eigenvalue actually reflects the number of extracted factors whose sum should be equal to the number of items that are subjected to factor analysis. The next item shows all the factors extractable from the analysis along with their eigenvalues. The Eigenvalue table has been divided into three sub-sections: Initial Eigen ValuesExtracted Sums of Squared LoadingsRotation of Sums of Squared Loadings. For analysis and interpretation purposes we are concerned only with Initial Eigenvalues and Extracted Sums of Squared Loadings. The requirement for identifying the number of components or factors stated by selected variables is the presence of eigenvalues of more than 1. Table 5 herein shows that for 1st component the value is 3.709 > 1, 2nd component is 1.478 > 1, 3rd component is 1.361 > 1, and 4th component is 0.600 < 1. Thus, the stated set of 8 variables with 12 observations represents three components. Further, the extracted sum of squared holding % of variance depicts that the first factor accounts for 46.367% of the variance features from the stated observations, the second 18.471% and the third 17.013% (Table 5). Thus, 3 components are effective enough in representing all the characteristics or components highlighted by the stated 8 variables. Component: As can be seen in the Communalities table 3 above, there 8 components shown in column 1 under table 3.Initial Eigenvalues Total: Total variance.Initial Eigenvalues % of the variance: The percent of variance attributable to each factor.Initial Eigenvalues Cumulative %: Cumulative variance of the factor when added to the previous factors.Extraction sums of Squared Loadings Total: Total variance after extraction.Extraction Sums of Squared Loadings % of the variance:The percent of variance attributable to each factor after extraction. This value is of significance to us and therefore we determine in this step that they are three factors which contribute to why would someone buy a particular product.Extraction Sums of Squared Cumulative %: Cumulative variance of the factor when added to the previous factors after extraction.Rotation of Sums of Squared Loadings Totals: Total variance after rotation.Rotation of Sums of Squared Loadings % of the variance: The percent of variance attributable to each factor after rotation.Rotation of Sums of Squared Loadings Cumulative %: Cumulative variance of the factor when added to the previous factors. Table 5: Total variance Offer ID is invalid The scree plot is a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 onwards has an eigenvalue of less than 1, so only three factors have been retained. Figure 1: Scree plot Table 6 below shows the loadings (extracted values of each item under 3 variables) of the eight variables on the three factors extracted. The higher the absolute value of the loading, the more the factor contributes to the variable. We have extracted three variables wherein the 8 items are divided into 3 variables according to the most important items which are similar responses in component 1 and simultaneously in components 2 and 3. The gap (empty spaces) on the table represents loadings that are less than 0.5, this makes reading the table easier. We suppressed all loadings less than 0.5. As the requirement of having precise computation of each factor component, Table 6 depicts that there is the presence of cross loading i.e. one factor measuring more than one component. As this cross-loading is very high in Table 6 i.e. cost of the product, the popularity of the product, the prestige of the product, and the quality of the product cross-loading, thus, for deriving more adequate results, these cross-loadings need to be eliminated. For this, the solution is to redistribute the factor loading by having rotation, and the hence rotated component matrix is examined for the identification of components. Table 6: Component matrix The idea of rotation is to reduce the number of factors on which the variables under investigation have high loadings. Rotation does not actually change anything but makes the interpretation of the analysis easier. Looking at the table below, we can see that availability of a product, and the cost of the product is substantially loaded on Factor (Component) 3. In contrast, experience with the product, the popularity of the product, and the quantity of the product are substantially loaded on Factor 2. Sometimes the loading of variables is there on two components or more. Therefore there is a requirement of checking the factor loading value. If the value is lower than the required value of 0.5 or the set limit (which could be 0.6 too as per the researchers need of including the desired factor loading) for one of the components, then that variable could be considered for further analysis. But as the presence of more than 0.5 (or 0.6) loading in more than one component represents that this variable represents two components, thus, it is not effective in measuring a specific category. Hence, need to be excluded. As in Table 7 experience with the product, and quality of the product measures more than one component, thus, they cant be considered for further analysis. Hence, further processing i.e. impact analysis or any other statistical analysis includes all variables except experience with the product, and quality of the product (Table 7). Table 7: Rotated component matrix

What is kmo and bartlett's test. What is the purpose of kmo and bartlett's test. Kmo and bartlett's test values. Kmo and bartlett's test interpretation.