|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Workflow:**  **Principal Components Analysis (PCA)** | | | | | | |
|  |  | | | |  |  |
|  | **Structure-up your dataframe(s) to have** | | | | | |
|  | ● | | Reverse-score any items are negatively worded (i.e., the item is scaled opposite the other items) | | | |
|  | ● | | Create a df with *only* the items (scaled in the proper direction) | | | |
|  |  | | | |  |  |
|  | **Check the assumptions associated with PCA** | | | | | |
|  | ● | | Kaiser-Meyer-Olkin (KMO) for sampling adequacy (KMO should be > .50) | | | |
|  | ● | | Barlett’s to rule out the possibility of an *identity matrix* (Barlett’s *p-value* should be < .05) | | | |
|  | ● | | Determinant of the correlation matrix to rule out *multicollinearity*  or *singularity* (the determinant should be > 0.00001) | | | |
|  |  | | | |  |  |
|  | **Determine the Number of Components to Extract** | | | | | |
|  | ● | | Run a PCA with the number of components equal to the number of variables/items in the data | | | |
|  | ● | | Examine **eigenvalues** (“SS loadings” in the psych package), consider the eigenvalues-greater-than-1 criteria for determining the number of components to extract | | | |
|  | ● | | Examine the **scree plot**, consider extracting the number of components above the plateau | | | |
|  | ● | | Consider the proportion of variance accounted for. How much additional variance does each additional component add? | | | |
|  | ● | | Consider any a priori theory regarding the number of factors | | | |
|  |  | | | |  |  |
|  | **Specify and Evaluate the Solution with the** | | | | | |
|  | ● | | Rerun the PA with the determined number of components | | | |
|  | ● | | Examine **communalities** individually and as an average.Criteria for acceptability range from being greater than .40 to being greater than .70 | | | |
|  | ● | | Examine **residuals**. Determine the number and percentage that are greater than .05. A common practice is to stay below 50% larger than .05. | | | |
|  |  | | | |  |  |
|  | **Component Rotation: Orthogonal or Oblique?** Expect to run both (with varying component structures), report only one. | | | | | |
|  | **Orthogonal** – to be used when we think the components are independent/unrelated; this approach minimizes cross-loadings | | | | | |
|  |  | ● | | Examine the components loadings. Are they above .3? Do any cross-load on other components (the highest loading should be on their theorized subscale)? | | |
|  |  | ● | | Are the communalities greater than at least .4? (ideally higher). Consider removing items when the communalities are .2 or lower. | | |
|  | **Oblique** – to be used when we expect the components to be correlated; this approach may be *truer* to our overarching hypothese (i.e., subscales are likely related to each other), but add complexity because cross-loadings are more likely | | | | | |
|  |  | ● | | Examine the components loadings. Are they above .3? Do any cross-load on other components (the highest loading should be on their theorized subscale)? | | |
|  |  | ● | | Are the communalities greater than at least .4? (ideally higher). Consider removing items when the communalities are .2 or lower. | | |
|  |  | ● | | The **pattern matrix** reports component loadings and is most comparable to the orthogonal output. | | |
|  |  | ● | | The **structure matrix** is a product of the pattern matrix and matrix containing the correlation coefficients between the components/scales). This can serve as a useful check to the pattern matrix. | | |
|  |  | |  | | | |
|  | **Engage in the *iterative* nature of this process** | | | | | |
|  | ● | | Try different models and evaluate their results before determining which model (number of components, orthogonal or oblique rotation) best fits the data | | | |
|  |  | |  | | |  |

Diagram

Description automatically generatedA picture containing text, antenna

Description automatically generated

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Workflow:**  **Principal Axis Factoring (PAF)** | | | | | | |
|  |  | | | |  |  |
|  | **Structure-up your dataframe(s) to have** | | | | | |
|  | ● | | Reverse-score any items are negatively worded (i.e., the item is scaled opposite the other items) | | | |
|  | ● | | Create a df with *only* the items (scaled in the proper direction) | | | |
|  |  | | | |  |  |
|  | **Check the assumptions associated with PAF** | | | | | |
|  | ● | | Kaiser-Meyer-Olkin (KMO) for sampling adequacy (KMO should be > .50) | | | |
|  | ● | | Barlett’s to rule out the possibility of an *identity matrix* (Barlett’s *p-value* should be < .05) | | | |
|  | ● | | Determinant of the correlation matrix to rule out *multicollinearity*  or *singularity* (the determinant should be > 0.00001) | | | |
|  |  | | | |  |  |
|  | **Determine the Number of Factors to Extract** | | | | | |
|  | ● | | Run a PCA with the number of factors equal to the number of variables/items in the data | | | |
|  | ● | | Examine **eigenvalues** (“SS loadings” in the psych package), consider the eigenvalues-greater-than-1 criteria for determining the number of factors to extract | | | |
|  | ● | | Examine the **scree plot**, consider extracting the number of factors above the plateau | | | |
|  | ● | | Consider the proportion of variance accounted for. How much additional variance does each additional factor add? | | | |
|  | ● | | Consider any a priori theory regarding the number of factors | | | |
|  |  | | | |  |  |
|  | **Specify and Evaluate the Solution with the** | | | | | |
|  | ● | | Rerun the PA with the determined number of factors | | | |
|  | ● | | Examine **communalities** individually and as an average.Criteria for acceptability range from being greater than .40 to being greater than .70 | | | |
|  | ● | | Examine **residuals**. Determine the number and percentage that are greater than .05. A common practice is to stay below 50% larger than .05. | | | |
|  |  | | | |  |  |
|  | **Factor Rotation: Orthogonal or Oblique?** Expect to run both (with varying factor structures), report only one. | | | | | |
|  | **Orthogonal** – to be used when we think the factors are independent/unrelated; this approach minimizes cross-loadings | | | | | |
|  |  | ● | | Examine the factors loadings. Are they above .3? Do any cross-load on other factors (the highest loading should be on their theorized subscale)? | | |
|  |  | ● | | Are the communalities greater than at least .4? (ideally higher). Consider removing items when the communalities are .2 or lower. | | |
|  | **Oblique** – to be used when we expect the factors to be correlated; this approach may be *truer* to our overarching hypothese (i.e., subscales are likely related to each other), but add complexity because cross-loadings are more likely | | | | | |
|  |  | ● | | Examine the factors loadings. Are they above .3? Do any cross-load on other factors (the highest loading should be on their theorized subscale)? | | |
|  |  | ● | | Are the communalities greater than at least .4? (ideally higher). Consider removing items when the communalities are .2 or lower. | | |
|  |  | ● | | The **pattern matrix** reports factor loadings and is most comparable to the orthogonal output. | | |
|  |  | ● | | The **structure matrix** is a product of the pattern matrix and matrix containing the correlation coefficients between the factors/scales). This can serve as a useful check to the pattern matrix. | | |
|  |  | |  | | | |
|  | **Engage in the *iterative* nature of this process** | | | | | |
|  | ● | | Try different models and evaluate their results before determining which model (number of factors, orthogonal or oblique rotation) best fits the data | | | |
|  |  | |  | | |  |