Technical Report

TechnoServe

Sofia Knutas, A01831481
Kristina Kristiansen Stapnes, A01830792

Aplicación de métodos multivariados en ciencia de datos (Gpo 601)
Group 6

1. Objective

The aim of this analysis was to evaluate whether the survey dataset was suitable for factor analysis and to extract the underlying latent factor structure using statistical justification.

2. Data Preparation and Screening

- Original dataset loaded and inspected for structure and variable types.
- Non-survey variables were removed (ID fields, timestamps, and outcome metrics).
- Remaining dataset contained only Likert-type survey items.
- Missing values (<5%) were removed to retain data integrity.
- Correlation matrix was generated to confirm sufficient inter-variable relationships.

3. Assessment of Factorability

Two statistical tests were used:

- Kaiser–Meyer–Olkin (KMO) Measure
  - Result: 0.959
  - Interpretation: Sampling adequacy for factor analysis is excellent.
- Average Inter-Item Correlation
  - Result: 0.337
  - Indicates correlations strong enough for latent structure detection but not high enough to imply redundancy.

These indicators confirm that factor analysis is technically appropriate for this dataset.

4. Factor Extraction Procedure

- Data standardized using z-score scaling.
- Initial unrotated factor analysis conducted to obtain eigenvalues.
- Scree plot generated.
- Kaiser criterion (eigenvalues > 1) applied.
- Both methods indicated retention of five factors.

5. Rotation and Loadings

Two rotation methods were applied:

- Varimax Rotation
  - Orthogonal
  - Produces clearer separation between variables
  - Used for primary interpretation
- Promax Rotation
  - Oblique
  - Allows correlation between factors
  - Used for validation comparison

Factor loadings were displayed in heatmaps for visual interpretation.

6. Variance Explained

The five extracted factors account for 50.7% of total variance. This level of variance retention is acceptable for survey-based psychological constructs.

7. Factor Structure Summary

The retained factors represent statistically distinct latent dimensions. Each factor grouped variables with loadings ≥ 0.40, demonstrating clean loading patterns and limited cross-loading.

8. Factor Scores and Correlation Assessment

- Factor scores were computed using the regression method.
- Factor score distributions were plotted.
- Correlation matrices compared factor scores with outcome variables.
- No factor demonstrated strong direct association, indicating independence between latent constructs and external performance indicators.

9. Statistical Visualization Outputs

The following diagnostic visualizations were generated:

- Correlation heatmap

- Scree plot of eigenvalues
- Varimax loading heatmap
- Promax loading heatmap
- Boxplots showing factor score distributions
- Radar chart representing normalized factor score profiles

These visualizations support the robustness of extraction, rotation, and interpretation.

10. Conclusion

- The dataset is statistically well-suited for factor analysis based on sample size, KMO score, and correlation structure.
- Five latent factors were extracted using standard factor retention criteria.
- Rotated factor solutions showed clean loading patterns and acceptable variance explanation.
- The technical process confirms a stable and interpretable factor structure without requiring domain-based interpretation.