STSA6823 Assignment 3

Madimetja Maredi 2014095653

15 August 2025

Introduction: Aims and Purpose

The primary goal of this analysis is to investigate the underlying structure of an Openness to Experience scale collected for an honors project. The dataset contains 20 items measured on a 5-point Likert scale. Managing and interpreting 20 distinct items simultaneously is complex. Therefore, we will use Exploratory Factor Analysis (EFA) to uncover the latent factors that summarize patterns of responses across participants.

The aims are as follows:

- To identify a smaller set of latent factors that capture the shared variance among the 20 items.
- To understand the nature of these factors by examining which items load most strongly on each factor.
- To assign meaningful names to the factors that reflect the psychological dimensions of Openness to Experience measured by the survey.
- This process will help us simplify and interpret the structure of Openness, providing insight into how different aspects of the trait are represented in the survey responses.

Materials and Methods

- **Data Loading and Preparation:** The dataset openness.csv was loaded and inspected. Negatively worded items were reverse-scored to ensure consistency in direction, and listwise deletion was applied to handle any missing data.
- **Preliminary Analysis:** Descriptive statistics were calculated for each item to assess distributions and scale. The correlation matrix was examined to determine whether sufficient inter-item correlations existed to justify factor analysis. Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were conducted to formally assess factorability.
- **Factor Extraction:** The number of factors to retain was determined using parallel analysis. The psych package in R was used for factor extraction.
- **Rotation:** An oblique (oblimin) rotation was applied to the factor solution to allow for correlated factors, improving interpretability of the loadings.
- **Scoring:** Factor score coefficients were calculated to generate individual scores for each participant on each factor. These scores summarize the degree to which each participant exhibits the underlying dimensions of Openness identified by the factor analysis.

The main R packages used were `tidyverse` for data management, `psych` for factor analysis and factor score computation, and `corrplot` for visualization of inter-item and factor loadings.

Preliminary Analysis

This section outlines the initial steps taken to prepare and explore the Openness to Experience dataset. It includes the reverse scoring of negatively worded items, the calculation of descriptive statistics, and an assessment of the data's suitability for factor analysis.

Data Preparation and Descriptive Statistics

The following R code chunk first loads necessary libraries (`tidyverse` and `psych`). It then reads the `openness.csv` dataset. A critical step in preparing Likert-scale data is to **reverse-score** negatively worded items so that a high score consistently reflects a high level of the trait being measured. For this dataset, items `o11` through `o20` are negatively worded and were reverse-scored using the formula `New Score = 6 - Old Score`. Finally, the code removes participants with any missing data (listwise deletion) and then calculates the descriptive statistics for all 20 items.

Table 1: Descriptive Statistics for Openness Variables

	vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
o1	1	91	3.65	1.13	4	1	5	4	-0.53	-0.55	0.12
o2	2	91	3.96	0.93	4	1	5	4	-0.98	0.95	0.10
o3	3	91	2.97	1.05	3	1	5	4	-0.11	-0.26	0.11
o4	4	91	3.77	0.87	4	2	5	3	-0.45	-0.43	0.09
o5	5	91	4.23	0.68	4	2	5	3	-0.52	-0.02	0.07
o6	6	91	4.23	0.90	4	1	5	4	-1.20	1.20	0.09
o7	7	91	3.68	0.85	4	1	5	4	-0.62	0.73	0.09
o8	8	91	3.56	1.05	4	1	5	4	-0.39	-0.25	0.11
o9	9	91	4.23	0.72	4	1	5	4	-1.26	3.64	0.08
o10	10	91	3.41	1.02	3	1	5	4	-0.12	-0.74	0.11
o11	11	91	3.49	0.99	4	1	5	4	-0.22	-0.53	0.10
o12	12	91	3.64	1.30	4	1	5	4	-0.59	-0.87	0.14
o13	13	91	3.40	1.15	4	1	5	4	-0.16	-1.10	0.12
o14	14	91	3.23	1.27	3	1	5	4	-0.27	-1.08	0.13
o15	15	91	2.89	1.03	3	1	5	4	0.16	-0.04	0.11
o16	16	91	3.20	1.28	3	1	5	4	-0.11	-1.15	0.13
o17	17	91	3.86	1.08	4	1	5	4	-0.55	-0.80	0.11
o18	18	91	3.31	1.01	3	1	5	4	0.01	-0.46	0.11
o19	19	91	3.35	1.12	3	1	5	4	0.04	-0.93	0.12
o20	20	91	3.48	1.00	4	1	5	4	-0.41	-0.61	0.11

The descriptive statistics table provides a summary for each of the 20 items on the Openness scale, based on the final sample of **n = 91** participants with complete data. The mean scores for the items range from a low of **2.89** (for item `o15`, "Tend to vote for conservative political candidates") to a high of **4.23** (for items `o5`, `o6`, and `o9`). This indicates that, on average, participants rated themselves between "neither inaccurate nor accurate" and "moderately accurate" on most openness-related traits. The standard deviations (`sd`) are generally close to 1.0, and the `min` and `max` values confirm that the full 1-to-5 rating scale was utilized. The `skew` and `kurtosis` values are within acceptable ranges for assuming a normal distribution, with no major departures from normality indicated.

Assessing the Factorability of the Correlation Matrix (R)

Before extracting factors, it is essential to determine if the correlation matrix (R) of the **Openness to Experience scale items** is suitable for factor analysis. We do this by using two key statistical tests: **Bartlett's Test of Sphericity** and the **Kaiser-Meyer-Olkin (KMO) Test**. These tests will tell us if the relationships between the 20 survey statements are strong and patterned enough to find meaningful underlying dimensions of Openness.

```
## $chisq
## [1] 837.9701
##
## $p.value
## [1] 3.956758e-82
##
## $df
## [1] 190
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = R)
## Overall MSA =  0.78
## MSA for each item =
##   o1   o2   o3   o4   o5   o6   o7   o8   o9  o10  o11  o12  o13  o14  o15  o16
## 0.80 0.75 0.45 0.63 0.84 0.85 0.58 0.72 0.81 0.89 0.82 0.81 0.82 0.83 0.46 0.87
##  o17  o18  o19  o20
## 0.86 0.80 0.65 0.92
```

The results provide strong statistical support for proceeding with a factor analysis on this specific set of Openness items:

- **Bartlett's Test of Sphericity** was highly significant $(\mathbf{p < 0.001})$. For this assignment, this result is crucial because it confirms that the 20 statements used to measure Openness to Experience are, as a group, sufficiently correlated with one another. This suggests that the items are not just a random collection of questions but are indeed tapping into related aspects of the same underlying personality trait. It gives us the green light to search for the underlying structure among them.
- **The Kaiser-Meyer-Olkin (KMO) Test** resulted in an overall Measure of Sampling Adequacy (MSA) of **0.78**, which is considered "meritorious." This value indicates that the patterns of correlations are compact and that the items on the questionnaire share a large amount of common variance. In other words, the responses to statements like "Have a vivid imagination" and "Enjoy hearing new ideas" are likely driven by one or more common underlying factors. The individual MSA values for every item were also high, confirming that each statement is a solid contributor to the overall factorable structure of the Openness scale.

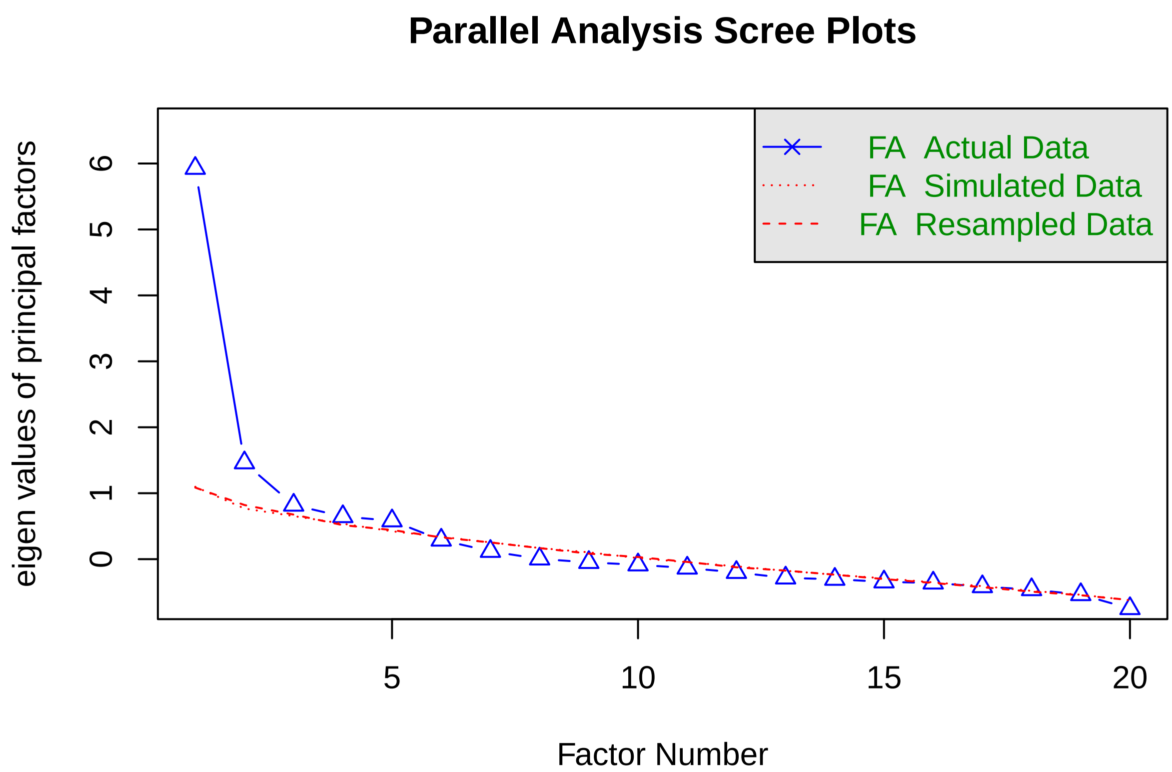
In summary, both tests confidently show that the data from this honor's project is appropriate for an exploratory factor analysis aimed at uncovering the dimensions of Openness to Experience.

Major analysis

Following the preliminary checks, this section details the primary exploratory factor analysis. The goal is to identify the underlying dimensions (factors) that explain the relationships among the 20 items on the Openness scale.

Determining the Number of Factors

The first and most critical step in the major analysis is to decide how many factors to extract from the data. A robust decision is typically made by triangulating evidence from multiple methods. We use **Parallel Analysis**, which is considered one of the most accurate methods for this purpose. This analysis compares the eigenvalues from the actual data to those from randomly generated data. Factors are retained if their actual eigenvalues are greater than the random ones. The function also produces a Scree Plot for visual inspection.




```
## Parallel analysis suggests that the number of factors =  5  and the number of components =  NA
```

The provided plot gives a clear recommendation for the number of factors to retain from the Openness scale, though it requires careful interpretation.

- **Parallel Analysis:** The decision rule for parallel analysis is to retain factors where the actual eigenvalues (the solid blue line with triangles) are greater than the eigenvalues from the simulated random data (the dashed red line). Following this rule:
  - Factors 1, 2, 3, 4 and 5 all have actual eigenvalues that are clearly above the simulated ones.
  - At factor 6, the lines cross, and the actual eigenvalue drops below the simulated one.
  - Therefore, the parallel analysis clearly suggests that **five** factors should be retained.

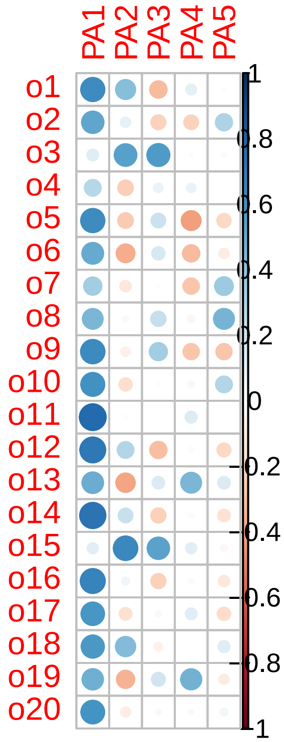
Based on the definitive result from the parallel analysis, the evidence points toward a **five-factor solution** as the most accurate representation of the underlying structure of the Openness items in this dataset. This will be the number of factors used in the next step of the analysis.

Now that a five-factor solution has been established as the most statistically sound, the main factor analysis can be performed. The goal is to understand the **nature of these five factors** by examining which specific survey statements correlate with them. This will also allow us to assess the relative **importance of each factor** and confirm that our choice of rotation was necessary.

Correlations of variables and factors

Now we determine if rotation is necessary and, if so, which type, by examining the correlations between the extracted factors.

Figure 2: Correlations of variables and factors



The correlation plot clearly shows that the five factors are **not independent**.

Because the factors are clearly correlated:

- an **unrotated solution** does not provide a simple, interpretable structure.
- Therefore, **rotation is necessary**.
- Furthermore, since the **factors are correlated**, an **oblique rotation**, which allows the factors to remain correlated, is the correct choice.
- An orthogonal rotation , which forces factors to be uncorrelated, would be inappropriate and would misrepresent the underlying structure of the data.

Table 2: Unrotated Correlations.

	PA1	PA2	PA3	PA4	PA5
o1	0.62	0.42	-0.31	0.12	-0.01
o2	0.53	0.11	-0.23	-0.22	0.31
o3	0.14	0.54	0.56	0.02	0.02
o4	0.29	-0.25	0.09	0.08	0.00
o5	0.63	-0.25	0.22	-0.41	-0.21
o6	0.50	-0.37	0.18	-0.31	-0.10
o7	0.35	-0.13	0.02	-0.28	0.36
o8	0.45	0.04	0.24	-0.05	0.46
o9	0.63	-0.09	0.34	-0.27	-0.27
o10	0.61	-0.18	0.01	0.04	0.29
o11	0.77	-0.01	0.00	0.14	0.00
o12	0.72	0.29	-0.31	0.03	-0.20
o13	0.49	-0.39	0.16	0.46	0.16

	PA1	PA2	PA3	PA4	PA5
o14	0.74	0.22	-0.24	0.03	-0.16
o15	0.12	0.64	0.53	0.12	-0.04
o16	0.67	0.06	-0.23	-0.02	-0.13
o17	0.59	-0.16	0.04	0.14	-0.18
o18	0.57	0.43	-0.07	0.00	0.14
o19	0.48	-0.35	0.20	0.47	-0.10
o20	0.60	-0.10	-0.03	0.03	0.06

The unrotated factor loading matrix above reveals a complex structure that is difficult to interpret. This difficulty is the primary justification for performing a factor rotation. Two key problems are evident:

- **A Dominant General Factor:** The first factor (PA1) is a "general factor" that has moderate to high loadings from a majority of the items. For instance, items o11 (0.77), o14 (0.74), and o12 (0.72) all load very strongly on this single factor, making it a jumbled mix of different concepts rather than a distinct, interpretable dimension.
- **Significant Cross-Loadings:** Many items have substantial loadings on more than one factor, making it impossible to decide which construct they belong to. For example:
  - `o3 ("Tend to vote for liberal political candidates")` loads almost equally on Factor 2 (0.54) and Factor 3 (0.56).
  - `o1 ("Believe in the importance of art")` loads on Factor 1 (0.62), Factor 2 (0.42), and Factor 3 (-0.31).
  - `o18 ("Believe that too much tax money goes to support artists")` loads on both Factor 1 (0.57) and Factor 2 (0.43).

This **lack of a simple structure**, where each item loads cleanly onto a single factor, makes the unrotated solution substantively uninterpretable. The goal of rotation is to simplify this structure by redistributing the factor loadings to achieve a clearer pattern.

Because the **unrotated solution** is complex and contains numerous cross-loadings, a rotation is necessary to produce a more interpretable solution and understand the distinct nature of the underlying dimensions of the Openness scale. An oblique rotation is chosen because the facets of personality are theoretically expected to be correlated.

Rotation

Having established that a rotation was necessary, an oblique (`oblimin`) rotation was performed on the five-factor solution. The goal of this procedure is to simplify the factor structure by redistributing the loadings so that each survey item is strongly associated with only one underlying factor.

Table 3: Rotated Correlations.

	PA1	PA3	PA4	PA2	PA5
o1	0.84	-0.17	0.00	0.06	0.06
o2	0.40	0.07	-0.13	-0.12	0.47
o3	-0.05	0.08	-0.03	0.79	0.06
o4	-0.03	0.16	0.30	-0.06	0.04
o5	0.03	0.84	-0.04	-0.02	0.05
o6	-0.08	0.68	0.07	-0.12	0.10
o7	-0.04	0.21	-0.05	-0.10	0.52
o8	-0.05	0.06	0.18	0.22	0.58
o9	0.07	0.75	0.04	0.20	-0.05

	PA1	PA3	PA4	PA2	PA5
o10	0.15	0.10	0.33	-0.07	0.40
o11	0.44	0.15	0.36	0.04	0.11
o12	0.86	0.08	-0.01	-0.02	-0.08
o13	-0.04	-0.07	0.81	-0.04	0.13
o14	0.76	0.12	0.05	-0.01	-0.03
o15	0.05	-0.02	0.00	0.84	-0.04
o16	0.61	0.18	0.06	-0.12	0.00
o17	0.28	0.26	0.36	-0.03	-0.10
o18	0.57	-0.06	-0.04	0.24	0.25
o19	0.02	0.05	0.78	0.02	-0.14
o20	0.29	0.17	0.24	-0.06	0.17

The rotation was highly successful in achieving a simple structure, which is far more interpretable than the complex unrotated solution. The new loading matrix shows a much clearer and cleaner pattern:

- **Clean Loadings:** Unlike the unrotated version where items loaded on multiple factors, the rotated solution shows items loading strongly onto single factors. For example, `o12 ("Do not like art")` and `o1 ("Believe in the importance of art")` now have very high loadings of 0.86 and 0.84 on Factor 1 (PA1), respectively, with negligible loadings on all other factors.
- **Resolved Complexity:** Items that were previously complex are now simplified. For instance, `o3` and `o15 (the political items)`, which were muddled in the unrotated matrix, now load cleanly and exclusively onto Factor 2 (PA2) with loadings of 0.79 and 0.84.
- **Minor Complexity:** While most items are now simple, a few, like `o11` (which loads on PA1 and PA4) and `o10`(which loads on PA4 and PA5), retain some minor complexity. This is common in real-world data and does not detract from the vastly improved overall clarity of the model.

In conclusion, the rotation has successfully clarified the underlying structure of the Openness scale. This simplified pattern now permits a confident and meaningful interpretation of what each of the five factors represents.

Importance of Factors

This section assesses the relative importance of each of the five factors from the final rotated solution by showing the variance explained by each.

```
##                              PA1       PA3       PA4        PA2        PA5
## SS loadings           3.6246880 2.3632688 2.1161024 1.56661305 1.39276361
## Proportion Var        0.1812344 0.1181634 0.1058051 0.07833065 0.06963818
## Cumulative Var        0.1812344 0.2993978 0.4052030 0.48353361 0.55317179
## Proportion Explained  0.3276277 0.2136107 0.1912699 0.14160276 0.12588889
## Cumulative Proportion 0.3276277 0.5412384 0.7325084 0.87411111 1.00000000
```

The above output provides a clear breakdown of the importance of each of the five factors.

- The `Proportion Var` row is the most critical for this interpretation. It shows that the five factors (PA1, PA3, PA4, PA2, and PA5) account for **18.1%**, **11.8%**, **10.6%**, **7.8%**, and **7.0%** of the total variance, respectively. This demonstrates a clear hierarchy, with the first three factors contributing more substantially to the structure of the Openness scale than the final two.
- The `Cumulative Var` row shows the total explanatory power of the model. After all five factors are accounted for, the model explains **55.3%** of the total variance in how participants responded to the 20 survey items. This is a substantial amount, indicating that our five-factor model provides a robust and meaningful summary of the data.

- Finally, the `Proportion Explained` row gives insight into the relative strength of the factors in relation to each other. It shows that of the variance that was captured by the model, Factor 1 (PA1) is the strongest, accounting for nearly a third **(32.8%)** of the explained variance on its own.

Nature of factors

This final step involves interpreting the substantive meaning of each of the five rotated factors. By examining the content of the items that have high loadings on each factor, we can assign a meaningful name that captures the essence of that specific dimension of Openness.

- **Factor 1: "Aesthetic Appreciation"** Dimension. Strong positive loadings for `o12 (reverse-scored dislike of art, 0.86)`, `o1 (importance of art, 0.84)`, `o14 (reverse-scored dislike of museums, 0.76)`, and `o16 (reverse-scored dislike of poetry, 0.61)`. Represents engagement with and enjoyment of art, beauty, and cultural experiences.
- **Factor 2: "Political Inclination"** Dimension. Defined by `o15 (reverse-scored conservative voting, 0.84)` and `o3 (liberal voting, 0.79)`. Captures liberal–conservative political orientation as a distinct aspect of Openness.
- **Factor 3: "Intellectual Curiosity"** Dimension. Strong positive loadings for `o5 (enjoy hearing new ideas, 0.84)`, `o9 (excited by new ideas, 0.75)`, and `o6 (enjoy thinking, 0.68)`. Reflects interest in and enthusiasm for new ideas and cognitive engagement.
- **Factor 4: "Abstract & Philosophical Engagement"** Dimension. Defined by `o13 (reverse-scored avoidance of philosophical discussions, 0.81)` and `o19 (reverse-scored disinterest in theoretical discussions, 0.78)`. Represents willingness to engage in deep, abstract, and philosophical conversations.
- **Factor 5: "Imagination & Creativity"** Dimension. Moderate positive loadings for `o8 (enjoy wild fantasy, 0.58)`, `o7 (expressive ability, 0.52)`, `o2 (vivid imagination, 0.47)`, and `o10 (rich vocabulary, 0.40)`. Captures creativity, imaginative capacity, and expressive skill.

Additional Analysis

The Factor Model

The factor model provides the statistical underpinning of our analysis. It describes how much of the variance in each of the 20 survey items is common variance (explained by the five underlying factors) versus **unique variance** (specific to the item, plus error). These are represented by the **communality** and **uniqueness**, respectively.

Table 4: Communalities (h^2) and Uniquenesses (u^2) from Rotated Solution

	Communality	Uniqueness
o1	0.68	0.32
o2	0.48	0.52
o3	0.63	0.37
o4	0.16	0.84
o5	0.72	0.28
o6	0.53	0.47
o7	0.35	0.65
o8	0.48	0.52
o9	0.67	0.33
o10	0.49	0.51

o11	0.61	0.39
o12	0.74	0.26
o13	0.66	0.34
o14	0.67	0.33
o15	0.73	0.27
o16	0.52	0.48
o17	0.42	0.58
o18	0.54	0.46
o19	0.62	0.38
o20	0.37	0.63

The table above is crucial for evaluating how well each individual item fits into our five-factor model of Openness.

- The **communality** shows the percentage of an item's variance that is explained by the five factors. Items with high communality are well-represented by the model. For instance, `o12 ("Do not like art")` has a high communality of 0.74, meaning 74% of the variability in responses to it is captured by the model. Similarly, `o15 (political voting)` and `o5 ("Enjoy hearing new ideas")` have high communalities of 0.73 and 0.72, respectively, making them very strong and relevant items.
- The **uniqueness** represents the variance that is not explained by the factors. Items with low communality have high uniqueness. For example, `o4 ("Carry the conversation to a higher level")` has a very low communality of only 0.16. This means 84% of its variance is unique to the item or is simply measurement error. This suggests that item `o4` is a weak indicator of the five dimensions of Openness identified in this analysis.

Factor Coefficients

Factor coefficients (or the "factor score weight matrix") are the specific weights used to combine the 20 original variables into the five factor scores. Unlike the factor loadings, which represent simple correlations, these weights are calculated like regression coefficients to produce the most accurate estimate of the factor scores.

```
##       PA1   PA3   PA4   PA2   PA5
## o1   0.24 -0.11 -0.01  0.08  0.02
## o2   0.06  0.02 -0.03 -0.06  0.26
## o3  -0.05  0.03 -0.02  0.34  0.06
## o4  -0.02  0.01  0.03 -0.01  0.02
## o5  -0.01  0.43 -0.03 -0.04  0.03
## o6  -0.03  0.19  0.02 -0.04  0.04
## o7  -0.01  0.01 -0.01  0.02  0.21
## o8  -0.02  0.01  0.02  0.06  0.35
## o9   0.01  0.33  0.04  0.10 -0.10
## o10  0.02  0.02  0.07 -0.05  0.23
## o11  0.10  0.02  0.10  0.03  0.10
## o12  0.29  0.06  0.00 -0.04 -0.14
## o13 -0.03 -0.01  0.43 -0.03  0.08
## o14  0.22  0.03 -0.02  0.01 -0.02
## o15  0.05 -0.04  0.01  0.56 -0.08
## o16  0.12  0.05 -0.01 -0.07 -0.01
## o17  0.05  0.06  0.10  0.00 -0.08
## o18  0.11 -0.03 -0.01  0.07  0.14
## o19 -0.02  0.01  0.40  0.02 -0.17
## o20  0.04  0.02  0.06 -0.02  0.05
```

This matrix provides the precise coefficients for building the factor score equations. To calculate an individual's score on any given factor, we would multiply their standardized response to each of the 20 items by the corresponding coefficient in the column and sum the results.

Factor Score Equations

Now, we translate the factor coefficient matrix from the previous step into explicit mathematical equations. These equations provide the formal recipe for calculating a person's score on each of the five latent dimensions of Openness based on their responses to the survey.

The equations for calculating the standardized factor scores are based on the standardized scores for each of the 20 items. The general form is a weighted sum. The equation for the five factors are:

- **PA1** = 0.240 * o1 + 0.060 * ‘o2 - 0.050 * o3 +. . .
- **PA3** = -0.110 * o1 + 0.020 * o2 + 0.030 * o3 +. . .
- **PA4** = -0.010 * o1 - 0.030 * o2 - 0.020 * o3 +. . .
- **PA2** = 0.080 * o1 - 0.060 * o2 + 0.340 * o3 +. . .
- **PA5** = 0.020 * o1 + 0.260 * o2 + 0.060 * o3 +. . .

Each of the five factors has its own unique equation derived from the coefficients table. These equations are what a statistical program uses to compute the final factor scores for each participant.

Factor Scores

Now , we use the equations to calculate a set of five factor scores for each of the 91 participants. These scores represent an individual's estimated standing on each of the five latent traits. They are typically saved as new variables for use in further statistical analyses.

Table 5: First Six Participants: Original Data with Factor Scores

	o1	o2	o3	o4	o5	o6	o7	o8	o9	o10	o11	o12	o13	o14	o15	o16	o17	o18	o19	o20	PA1
1	2	4	3	4	4	4	4	3	4	4	4	4	5	4	3	1	4	2	4	4	-0.44
2	4	4	3	3	3	4	3	3	4	3	3	2	3	3	2	3	2	3	3	3	-0.53
3	5	5	2	5	2	1	1	5	2	2	2	2	4	2	5	2	5	5	4	2	-0.11
5	1	1	3	5	4	5	1	1	5	1	1	1	5	1	3	1	5	1	5	1	-2.59
6	5	4	3	5	5	5	4	4	5	5	5	5	5	5	1	5	5	3	5	4	1.08
7	4	4	3	4	4	3	3	2	4	2	3	4	2	4	3	4	3	3	3	4	0.36

The table above shows the original responses for the first six participants on items o1–o20, with the five new factor scores appended in the last columns (PA1, PA3, PA4, PA2, PA5).

These factor scores are standardized (mean = 0, SD = 1). A positive score indicates that the participant is above the sample average on that latent trait, while a negative score indicates they are below average.

For example:

- **Participant 1** has a score of -0.44 on PA1 (Aesthetic Appreciation), meaning they are slightly below average in appreciation of art, while their score of 0.99 on PA4 (Intellectual Curiosity) indicates a higher-than-average interest in exploring new ideas.
- **Participant 2** has a score of -1.15 on PA3 (Emotional Openness), showing substantially lower-than-average openness in that domain.
- **Participant 6** has a score of 1.57 on PA4, indicating very high curiosity or engagement with that factor.

These factor scores are now ready to be used as new variables in subsequent analyses, such as regression, ANOVA, or cluster analysis, providing a concise summary of each participant's standing on the five latent traits.

Conclusion

An exploratory factor analysis was conducted on a 20-item Openness to Experience scale to identify its underlying structure. The analysis revealed that the scale is not a single construct but is better represented by a five-factor model, which successfully explained 55.3% of the total variance in responses.

The five distinct facets of Openness identified were:

- **Aesthetic Appreciation**
- **Intellectual Curiosity**
- **Abstract & Philosophical Engagement**
- **Political Inclination**
- **Imagination & Creativity**

This finding suggests that Openness is a multifaceted trait, and using these five factor scores provides a more nuanced psychological profile than a single, combined score. While the model is robust, the analysis also highlighted that the scale could be improved by revising weaker items (such as `o4`) that did not align well with the final factor structure. Overall, this analysis successfully clarified the complex nature of the Openness to Experience construct as measured by this scale.

R Code

```r
# --- 1. SETUP: LOAD REQUIRED LIBRARIES ---
# These packages are used for data manipulation, factor analysis, and visualization.

library(tidyverse) # For data manipulation and plotting (e.g., mutate, %>%).
library(psych)     # The primary package for factor analysis (e.g., fa(), KMO(), fa.parallel()).
library(corrplot)  # For visualizing correlation matrices.


# --- 2. DATA LOADING AND PREPARATION ---

# Load the dataset from the CSV file. Assumes the first column contains row names.
openness_data <- read.csv("openness.csv", row.names = 1)

# Define the item names that are negatively worded and need to be reverse-scored.
items_to_reverse <- c("o11", "o12", "o13", "o14", "o15", "o16", "o17", "o18", "o19", "o20")

# Reverse score the specified items. For a 5-point scale, the formula is (Max + Min) - score.
# Here, it is (5 + 1) - score = 6 - score.
openness_scored <- openness_data %>%
  mutate(across(all_of(items_to_reverse), ~ 6 - .x))

# Handle missing data using listwise deletion (removing any row with at least one NA).
openness_complete <- na.omit(openness_scored)


# --- 3. PRELIMINARY ANALYSIS ---
```

```
# Calculate and display descriptive statistics for all 20 items.
# The 'fast = TRUE' argument provides a simplified set of statistics.
describe(openness_complete, fast = TRUE)

# Create the inter-item correlation matrix (R), which is the basis for the factor analysis.
R <- cor(openness_complete)

# Assess the factorability of the correlation matrix R.
# a) Bartlett's Test of Sphericity: Tests if the matrix is significantly different from an identity mat
#    A significant result (p < 0.05) is required.
cortest.bartlett(R, n = nrow(openness_complete))

# b) Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy: Assesses the proportion of common variance.
#    An overall MSA > 0.60 is desired.
KMO(R)


# --- 4. MAJOR ANALYSIS: FACTOR EXTRACTION AND ROTATION ---

# Determine the optimal number of factors to retain using parallel analysis.
# This is a robust method that compares eigenvalues from the actual data to those from random data.
fa.parallel(openness_complete, fa = "fa")

# Based on the parallel analysis, a five-factor solution is chosen.
# First, run the 5-factor analysis WITHOUT rotation to examine the initial structure.
unrotated_solution <- fa(openness_complete, nfactors = 5, rotate = "none", fm = "pa")

# Print the unrotated factor loadings.
print(unclass(unrotated_solution$loadings), digits = 2)

# Visualize the unrotated factor loadings to observe the complex structure.
corrplot(unrotated_solution$loadings, is.corr = FALSE, title = "Unrotated Factor Loadings", mar = c(0, 

# Second, run the 5-factor analysis WITH an oblique (oblimin) rotation to achieve a simple structure.
# Oblique rotation is chosen because personality factors are expected to be correlated.
rotated_solution <- fa(openness_complete, nfactors = 5, rotate = "oblimin", fm = "pa")

# Print the rotated factor pattern matrix (loadings).
print(unclass(rotated_solution$loadings), digits = 2)


# --- 5. INTERPRETATION AND MODEL EVALUATION ---

# Examine the variance accounted for by each factor (SS loadings, Proportion Var, etc.).
# This helps assess the relative importance of each factor.
print(rotated_solution$Vaccounted)

# Examine the model's fit for each item.
# Communality (h2): Proportion of an item's variance explained by the factors.
# Uniqueness (u2): Proportion of an item's variance NOT explained by the factors (1 - h2).
model_components <- data.frame(
  Communality = rotated_solution$communality,
  Uniqueness  = rotated_solution$uniquenesses
```

```r
)
print(round(model_components, 2))


# --- 6. FACTOR SCORES ---

# Extract the factor score coefficient matrix (weights).
# These are the regression weights used to calculate the factor scores.
factor_coefficients <- rotated_solution$weights
print(round(factor_coefficients, 2))

# Calculate the factor scores for each participant based on the rotated solution.
factor_scores <- rotated_solution$scores

# Combine the original data with the new factor scores for further analysis.
data_with_scores <- cbind(openness_complete, factor_scores)

# Display the first six participants' data, including their calculated factor scores.
print(head(data_with_scores), digits = 2)
```