MWSUG 2024 - Paper 49

Evaluating Questionnaire Data Reliability with SAS

Yida Bao, University of Wisconsin Stout
Zheng Zhang, Murray State University
Jiafeng Ye, Shanghai Jiaotong University
Liangliang Xu, Auburn University
Tao Jiang, University of Florida
Yongli Zhao, Auburn University
Philippe Gaillard, Augusta University

Abstract:

In the field of questionnaire data analysis, evaluating the internal consistency of scales is crucial for ensuring the reliability of measurement instruments. This paper explores the use of SAS for analyzing questionnaire data, focusing on methods that assess internal consistency and reliability. It presents the application of several statistical techniques, including Split-Half Reliability, Cronbach's Alpha, and McDonald's Omega, using both simulated and real datasets. The primary objective is not merely to evaluate these reliability measures but to illustrate how to utilize SAS effectively for conducting reliability analysis in questionnaire research. By demonstrating the implementation of these methods in SAS, this paper aims to provide researchers with a practical guide for assessing questionnaire quality, where the focus on methodology supersedes the importance of results.

Key words: Split-Half Reliabilit ● McDonald's Omega ● questionnaire data ● Cronbach's Alpha ● internal consistency

The Importance of Questionnaire Data Analysis:

Questionnaire data analysis is a cornerstone in numerous fields such as psychology, education, market research, and social sciences. The primary objective of employing questionnaires is to collect data that can provide insights into various aspects of human behavior, attitudes, and experiences. The reliability and validity of the conclusions drawn from such data are heavily dependent on the quality of the questionnaire and the robustness of the data analysis methods used. Analyzing questionnaire data allows researchers to identify patterns, understand underlying constructs, and make informed decisions based on empirical evidence. Without rigorous analysis, the data collected might lead to misleading conclusions, ultimately affecting the outcomes of the research or interventions based on these findings.

One of the critical aspects of questionnaire data analysis is the evaluation of internal consistency, which refers to the extent to which all items in a questionnaire measure the same construct. Ensuring internal consistency is essential because it validates that the items are cohesively reflecting the intended concept, providing a reliable measure. Various statistical methods, such as Cronbach's Alpha, Split-Half Reliability, McDonald's Omega, , are employed to assess this consistency. Each method has its unique advantages and applications, allowing researchers to choose the most appropriate one based on their specific research design and data characteristics. By rigorously evaluating the internal consistency of questionnaires, researchers can enhance the reliability and credibility of their findings, ensuring that the insights derived are both accurate and meaningful. This process not only strengthens the foundation of the research but also ensures that subsequent applications of the findings are based on sound and dependable data.

Questionnaire Data Analysis:

Structure of Simulated Questionnaire Data:

The structure of questionnaire data can vary depending on the type of questionnaire and the nature of the items (e.g., binary, Likert scale). Here's a basic outline:

- Sample ID: A unique identifier for each respondent.
- Items: Columns representing individual questionnaire items. These items can have various types of responses, such as binary (yes/no), Likert scale (e.g., 1 to 5), or open-ended responses.

For example, a dataset might look like this for a binary questionnaire with five items:

Sample	Item 1	Item 2	Item 3	Item 4	Item 5
1	1	0	1	1	0
2	1	1	0	1	1
3	0	0	1	0	1
...	...	...	...	...	...

Generating Simulated Questionnaire Data in SAS:

To generate a simulated dataset suitable for evaluating reliability methods such as Split-Half Reliability, Cronbach's Alpha, Guttman's Lambda, and McDonald's Omega, we can create a dataset with 100 variables (items). These variables will follow a Likert scale (e.g., 1 to 5), which is common in questionnaire data. This section will describe the process of generating this data in SAS, focusing on automating the generation of a large number of variables.

Below is an example of how to generate a simulated dataset with 100 variables, each following a Likert scale from 1 to 5.

```
/* Set the random seed for reproducibility */
%let seed = 12345;

/* Define the number of respondents and items */
```

```
%let num_respondents = 1000;
%let num_items = 100;

/* Generate simulated data */
data simulated_data;
    call streaminit(&seed);
    do id = 1 to &num_respondents;
        array items[&num_items];
        do i = 1 to &num_items;
            items[i] = ceil(rand("Uniform") * 5); /* Randomly generate
values between 1 and 5 */
        end;
        output;
    end;
    drop i;
run;

/* Display a sample of the generated data */
proc print data=simulated_data(obs=10);
run; proc corr data=simulated_data alpha;
    var items1-items100;
run;
```

This SAS code generates a dataset with 1,000 respondents and 100 Likert-scale items, suitable for evaluating reliability measures like Split-Half Reliability, Cronbach's Alpha, Guttman's Lambda, and McDonald's Omega.

Obs	id	items1	items2	items3	items4	items5	items6	items7	items8	items9	items10	items11	items12	items13	items14	items15	items16
1	1	3	5	3	5	5	2	4	2	3	5	3	5	2	5	1	2
2	2	5	4	3	5	5	2	1	4	1	5	4	2	3	2	3	3
3	3	2	2	5	4	1	3	1	1	1	4	5	3	2	1	5	1
4	4	4	2	4	3	5	3	5	3	4	5	2	4	1	1	5	1
5	5	2	5	1	1	1	2	4	5	1	5	3	4	4	4	3	3
6	6	2	1	4	1	1	1	4	2	2	2	3	3	2	4	5	4
7	7	1	2	3	3	1	3	3	3	3	5	2	5	5	1	3	3
8	8	2	1	5	2	2	4	1	3	4	2	3	3	1	4	5	2
9	9	4	5	1	5	5	4	2	3	4	2	4	2	1	2	1	4
10	10	2	2	2	1	4	4	5	4	2	2	3	4	1	5	3	3

This code snippet consists of a nested loop structure to create the dataset. The outer loop iterates over each respondent (id), creating 1,000 unique respondents. Inside this loop, an array named items with 100 elements is declared to represent the 100 questionnaire items. The inner loop iterates over each of these items, assigning a random Likert-scale value (1 to 5) to each item using ceil(rand("Uniform") * 5). This generates a uniform distribution of responses, simulating realistic questionnaire data. The output statement saves the data for each respondent after all items have been assigned values. Finally, the drop i; statement removes the loop variable i from the dataset to keep it clean. This approach efficiently creates a comprehensive dataset ideal for evaluating reliability measures like Split-Half Reliability, Cronbach's Alpha, Guttman's Lambda, and McDonald's Omega.

This simulated dataset is ideal for comprehensive reliability analysis, enabling detailed comparison and validation of different methods.

- **Split-Half Reliability:**
  With 100 items, multiple splits can be tested, ensuring robust internal consistency estimates.
- **Cronbach's Alpha:**
  The large number of items and responses ensures stable and reliable alpha coefficients.
- **McDonald's Omega:**
  Facilitates factor analysis and omega calculations, providing deeper insights into the questionnaire's structure.

This simulated dataset is ideal for comprehensive reliability analysis, enabling detailed comparison and validation of different methods.

Data Description in SAS:

Accurate and detailed description of questionnaire data is a fundamental part of any research study involving surveys. It sets the stage for understanding the dataset, providing crucial insights into its structure, content, and the demographic characteristics of the respondents. By outlining the specifics of how the data were collected, organized, and processed, researchers ensure that the dataset is transparent and comprehensible, facilitating reliable and valid analysis in subsequent sections of the study.

Correlation Matrix Heatmap

A Correlation Matrix Heatmap is a powerful tool for visualizing relationships between items in questionnaire data. It graphically represents the correlation coefficients between items, making it easy to identify patterns at a glance. In questionnaire data, each item typically represents a question or statement rated by respondents. High positive correlations between items indicate that respondents who rate one item highly also tend to rate the other item highly, suggesting they measure similar constructs. Negative correlations, on the other hand, suggest that respondents who rate one item highly tend to rate the other item lowly, indicating potential differences in the underlying constructs.

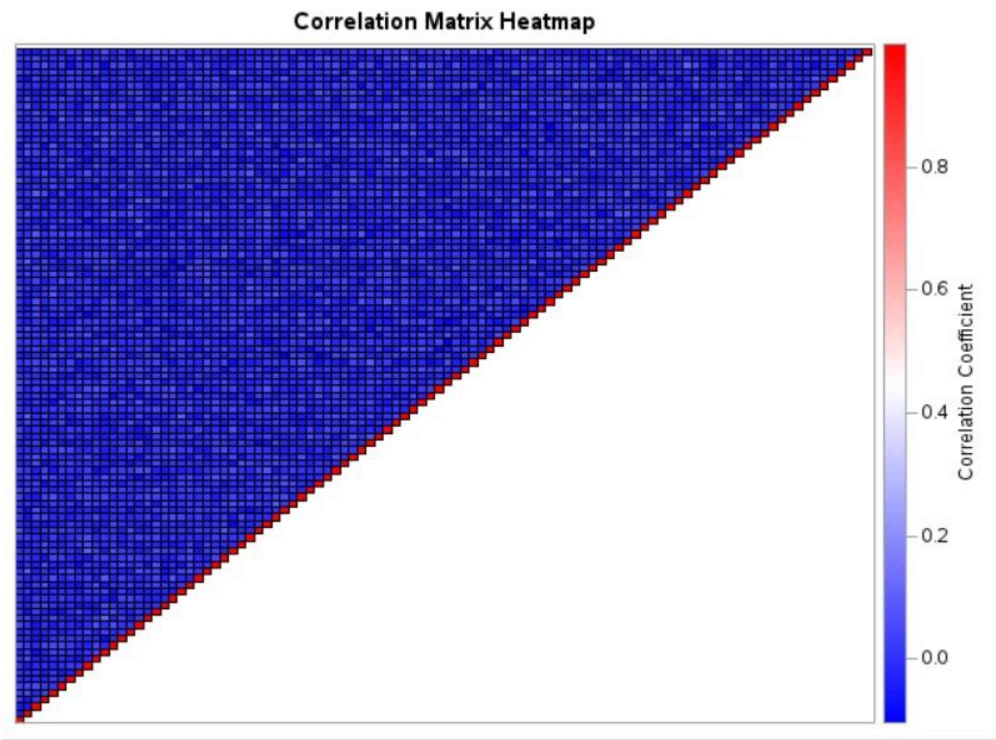
```
/* Calculate the correlation matrix */
 proc corr data=simulated_data alpha;
    var items1-items100;
run;

/* Prepare data for heatmap */
data corr_matrix_long;
    set corr_matrix;
```

```
    array items[*] items1-items100;
    do i = 1 to dim(items);
        do j = i to dim(items);
            corr_value = items[j];
            if _TYPE_ = 'CORR' and _NAME_ = vname(items[i]) then do;
                item_x = vname(items[i]);
                item_y = vname(items[j]);
                output;
            end;
        end;
    end;
    keep item_x item_y corr_value;
run;

/* Plot the Correlation Matrix Heatmap */

proc sgplot data=corr_matrix_long noautolegend;
    title "Correlation Matrix Heatmap";
    heatmapparm x=item_x y=item_y colorresponse=corr_value /
colormodel=(blue white red) outline;
    gradlegend / title="Correlation Coefficient" position=right;
    xaxis display=none;
    yaxis display=none;
    refline 0 / axis=x lineattrs=(pattern=solid color=black);
    refline 0 / axis=y lineattrs=(pattern=solid color=black);
run;
```

First, the PROC CORR procedure calculates the correlation matrix for the items in the dataset. The alpha option is included to compute Cronbach's alpha for reliability analysis, although it is not used in the subsequent heatmap.

Next, the corr_matrix_long dataset is prepared for the heatmap. The code reshapes the correlation matrix from a wide format to a long format, suitable for plotting. The array statement creates an array for the items, and nested loops iterate through each pair of items. For each pair, the correlation value is extracted and stored along with the item names in a new dataset.

Finally, the PROC SGPLOT procedure is used to create the Correlation Matrix Heatmap. The heatmapparm statement plots the heatmap, with the colormodel option specifying a blue-white-red color scheme to represent the strength and direction of correlations. The gradlegend statement customizes the legend. The xaxis and yaxis statements hide the axis labels for a cleaner look, and reference lines are added at the origin.



Since the dataset is simulated with random values between 1 and 5, the heatmap predominantly shows blue shades, indicating correlations close to zero or slightly negative. In this heatmap, blue shades represent negative or near-zero correlations, showing that most item pairs do not have a strong relationship. White areas indicate correlations near zero, suggesting no significant association between items. Red shades would indicate positive correlations but are rare in this random dataset. Overall, the blue-dominated heatmap is expected, reflecting the lack of inherent relationships in the simulated data.

Analyzing Simulated Data for Reliability in SAS:

Crobach's Alpha

Cronbach's Alpha is one of the most widely used methods for assessing the internal consistency of a questionnaire. It measures how closely related a set of items are as a group, providing a coefficient that ranges from 0 to 1. A higher Cronbach's Alpha indicates better internal consistency. The formula for Cronbach's Alpha is:

$$\alpha = \frac{N}{N-1}(1 - \frac{\sum_{i=1}^{N} \sigma_i^2}{\sigma_{i\,total}^2})$$

Where N is the number of items, $\sigma_i^2$is the variance of each item, and$\sigma_{itotal}^2$is the total variance of all items. However, it assumes that all items contribute equally to the construct being measured, which might not always be the case.

Cronbach's Alpha calculation using PROC CORR has been provided. Here is the output for reference:

```
proc corr data=simulated_data alpha;
    var items1-items100;
run;
```

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	-.019474
Standardized	-.018792

When using SAS to calculate Cronbach's alpha, a negative result usually indicates issues with the internal consistency of the scale items. This can occur due to several reasons: items may have negative or low correlations with each other, indicating they may not be measuring the same underlying construct; some items might need reverse scoring but were not properly reversed; the characteristics of simulation data might not accurately reflect real measurement properties, leading to low or negative correlations; a very small sample size can cause instability in correlation estimates; and potential calculation errors should also be considered. Reviewing the items, scoring procedures, and data used in the analysis is essential to address these issues.

Raw Cronbach's alpha and standardized Cronbach's alpha differ in their calculation methods. Raw alpha uses the original item scores and is suitable when item scores are on the same scale and have consistent units. In contrast, standardized alpha is based on standardized item scores, calculated by dividing each item's score by its standard deviation, making it more appropriate for comparing items with different scales or units.

$$\alpha = \frac{N_{\bar{r}}}{1 + (N + 1)\bar{r}}$$

Where

- N is the number of items.
- $\bar{r}$ is the average of all pairwise correlations between the items.

If there is a significant difference between raw and standardized alpha, it may indicate discrepancies in the scales or units of item scores. Standardized alpha can provide a more accurate reflection of internal consistency in such cases.

Also, The "Cronbach Coefficient Alpha with Deleted Variable" table provides insights into the reliability of a questionnaire by showing the Cronbach's Alpha coefficient for the dataset if each specific item were removed. Items that, when deleted, result in a higher alpha value may be negatively impacting the internal consistency and could be considered for removal or revision. Conversely, items whose removal decreases the alpha value are positively contributing to the overall reliability and should be retained. This analysis helps in refining the questionnaire to ensure it effectively measures the intended construct with high internal consistency.

Cronbach Coefficient Alpha with Deleted Variable				
	Raw Variables		Standardized Variables	
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
items1	0.021835	-.024293	0.021777	-.023545
items2	-.008636	-.017894	-.009011	-.017107
items3	-.002392	-.019180	-.002425	-.018481
items4	-.005478	-.018538	-.005117	-.017919
items5	0.024342	-.024894	0.024333	-.024081
items6	0.029787	-.025923	0.030505	-.025377
items7	-.016947	-.016141	-.016913	-.015462
items8	-.066171	-.005790	-.066267	-.005243
items9	-.009939	-.017603	-.009533	-.016999
items10	0.025318	-.024957	0.025927	-.024415
items11	0.005560	-.020868	0.004912	-.020014
items12	0.057779	-.031821	0.058410	-.031257
items13	-.026343	-.014123	-.026207	-.013530
items14	-.029460	-.013476	-.029742	-.012796
items15	0.015087	-.022849	0.015619	-.022254
items16	0.001669	-.020015	0.001615	-.019325
items17	-.011234	-.017350	-.011247	-.016642
items18	-.020478	-.015327	-.020323	-.014753
items19	-.045178	-.010265	-.045214	-.009590
items20	-.018780	-.015684	-.018851	-.015059
items21	0.029121	-.025757	0.029953	-.025261
items22	-.002577	-.019135	-.003032	-.018354

Split-Half Reliability

Split-Half Reliability involves splitting the questionnaire into two halves and correlating the scores from each half. This method provides a measure of consistency but requires correction using the Spearman-Brown formula:

$$r_{SB} = \frac{2r}{1 + r}$$

Where r is the correlation between the two halves. Split-half reliability can vary depending on how the questionnaire is split, which might affect the consistency of the results. This method is straightforward but may not always provide the most stable reliability estimate.

To complete the Split-Half Reliability calculation, we need to compute the Spearman-Brown prophecy formula, which corrects the split-half correlation for the fact that it is based on a split test.

```
/* Calculate the correlation between the two halves */
proc corr data=split_half noprint outp=split_half_corr;
    var half1 half2;
run;

/* Extract the correlation coefficient */
data split_half_corr;
    set split_half_corr;
    if _TYPE_ = 'CORR' and _NAME_ = 'half1';
    corr_half = half2; /* The correlation between half1 and half2 */
    keep corr_half;
run;

/* Calculate Spearman-Brown corrected split-half reliability */
data split_half_final;
    set split_half_corr;
    spearman_brown = 2 * corr_half / (1 + corr_half);
run;

/* Print the final results */
proc print data=split_half_final;
    var corr_half spearman_brown;
run;
```

Obs	corr_half	spearman_brown
1	-0.015329	-0.031135

The corr_half value of -0.0153 indicates a very weak and slightly negative correlation between the two halves of the questionnaire, suggesting that the items in the two halves do not consistently measure the same construct. The spearman brown value of -0.0311, which is also negative, indicates poor overall reliability.

However, since these results are based on simulated data, such outcomes can make sense. The simulated data might not perfectly reflect real-world measurement properties, leading to low or negative correlations. These results highlight the importance of carefully designing and validating simulated datasets to ensure they accurately represent the constructs being measured.

McDonald's Omega

McDonald's Omega is a more sophisticated method that addresses some of the limitations of Cronbach's Alpha. It provides a more accurate estimate of internal consistency by accounting for the different contributions of individual items to the overall construct. The formula for Omega is:

$$\omega = \frac{\sum_{i=1}^{k} \lambda_i{}^2}{\sum_{i=1}^{k} \lambda_i{}^2 + \sum_{i=1}^{k} \delta_i{}^2}$$

whereλ_i are the factor loadings and δ_i are the error variances. Omega is particularly useful in factor analysis, where it can provide a clearer picture of the underlying structure of the questionnaire.

For a more accurate calculation of McDonald's Omega, we need to use both the factor loadings and the residual variances.

```
/* Perform factor analysis to get factor loadings */
proc factor data=simulated_data nfact=1 outstat=fact_loadings;
    var items1-items100;
run;
```

We performed a factor analysis using Principal Axis Factoring to extract a single common factor from the dataset containing the items items1 through items100. This step was essential to determine the factor loadings (λ_i) for each item, which reflect the degree to which each item correlates with the underlying latent factor.

```
/* Extract factor loadings and calculate Omega */
data omega_calc;
    set fact_loadings(where=(_TYPE_='PATTERN'));
    array loadings[100] _numeric_;
    sum_squared_loadings = 0;
    sum_loadings_squared = 0;
    sum_residual_variances = 0;

    /* Loop through each loading to calculate required sums */
    do i = 1 to dim(loadings);
        sum_squared_loadings + loadings[i];
        sum_loadings_squared + loadings[i]**2;
        sum_residual_variances + (1 - loadings[i]**2);
    end;
```

After obtaining the factor loadings, we extracted them from the fact_loadings dataset and computed the necessary sums for the Omega calculation.Specifically, we calculated:

- The sum of squared factor loadings: $\sum \lambda_i{}^2$.
- The sum of residual variances: $\sum \delta_i{}^2$, where $\delta_i{}^2 = 1 - \lambda_i{}^2$

```
    /* Calculate Omega numerator and denominator */
    omega_num = sum_loadings_squared;
    omega_denom = omega_num + sum_residual_variances;

    /* Calculate Omega */
    omega = omega_num / omega_denom;

    /* Keep only relevant variables */
    keep omega;
run;
```

With the sums computed, we calculated the Omega coefficient using the formula:

$$\omega = \frac{\sum_{i=1}^{k} \lambda_i{}^2}{\sum_{i=1}^{k} \lambda_i{}^2 + \sum_{i=1}^{k} \delta_i{}^2}$$

```
/* Print the calculated Omega value */
proc print data=omega_calc;
    var omega;
run;
```

Obs	omega
1	0.017200

An Omega value of 0.0172 indicates that the items on the questionnaire are not reliably measuring the same underlying construct. This means that the variance in the composite score is mostly due to measurement error or noise rather than true score variance.

Since the data is simulated, such a low Omega value can make sense depending on how the data was generated. Simulated data might not always reflect the complexities and inherent structures of real-world data.

Compare Three Method:

Assessing the reliability of psychological tests often involves methods like Split-Half Reliability, Cronbach's Alpha, and McDonald's Omega, each differing in approach and applicability.

Split-Half Reliability divides a test into two halves and calculates the correlation between them. It's straightforward and requires minimal statistical effort. However, its reliability estimate can vary depending on how the test is split, and it doesn't consider all possible split combinations. This method may not be suitable for tests with heterogeneous items, as it can yield inconsistent results.

Cronbach's Alpha computes the average correlation among all items, providing an overall estimate of internal consistency. It's widely used due to its ease of calculation with standard statistical software. The main limitation is its assumption of tau-equivalence—that all items contribute equally to the construct. If items differ significantly or the scale is multidimensional, Cronbach's Alpha may overestimate or underestimate the true reliability.

McDonald's Omega addresses some limitations of Cronbach's Alpha by considering the factor loadings of items from a factor analysis. It provides a more accurate reliability estimate for scales with hierarchical or multidimensional structures. Unlike Cronbach's Alpha, it doesn't assume equal item contributions. The downside is its computational complexity and the requirement for a factor analysis, making it less accessible to those unfamiliar with advanced statistical techniques.

In essence, Split-Half Reliability is simple but can be inconsistent; Cronbach's Alpha is convenient but assumes equal item impact; McDonald's Omega is precise for complex scales but demands more sophisticated analysis.

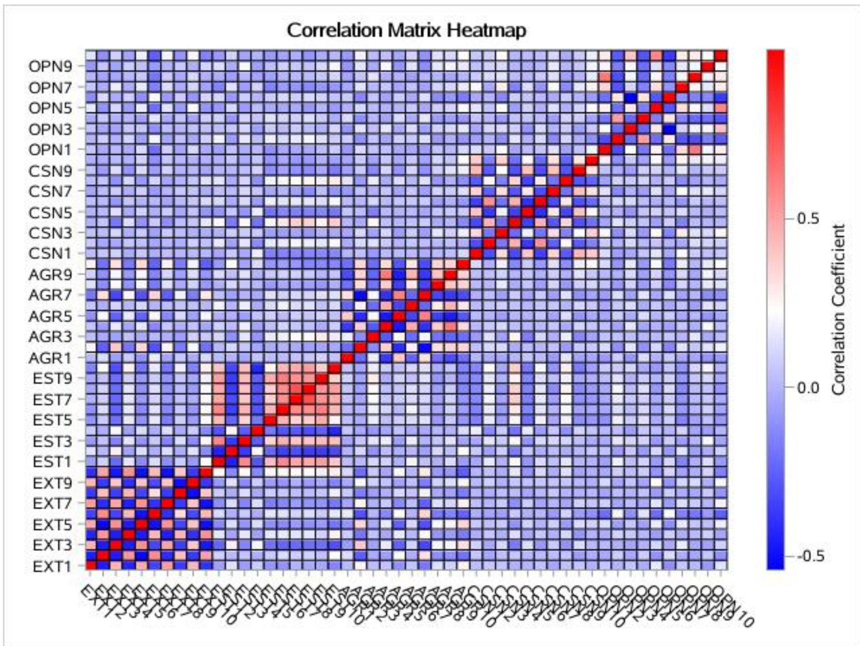
Method	Key Feature	Advantages	Limitation
Split-Half Reliability	Correlation between two test halves	Simple computation	Results vary by split; not comprehensive
Cronbach's Alpha	Average inter-item correlation	widely used	Assumes equal item impact;
McDonald's Omega	Considers item factor loadings	Accurate for complex scales	Requires factor analysis; more complex

Test Method on Real Data

We decide to use a real data to test our methods. the dataset in question is the Big Five Personality Test dataset, which contains over one million responses to a 50-item questionnaire designed to assess the five major dimensions of personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Collected online by Open Psychometrics, this extensive dataset provides a rich source for analyzing the reliability of psychological measures. Using the Big Five Personality Test dataset, which includes over a million questionnaire responses, we randomly selected 1,000 samples to evaluate Split-Half Reliability, Cronbach's Alpha, and McDonald's Omega.

These three methods differ notably in how they assess internal consistency, especially when applied to a multidimensional instrument like the Big Five.

Here the heatmap:



The heatmap represents the correlations between 50 different question from the Big Five Personality Test, grouped by trait. The variation in color intensity shows that some items within the same trait have stronger correlations (red areas), indicating better internal consistency, while others show weaker or even negative correlations (blue areas), suggesting less consistency or possible multidimensionality.

Despite using 50 items to cover five broad traits, the reliability values remain relatively low across all methods. The **Split-Half Reliability** (0.15395) and **McDonald's Omega** (0.15278) indicate inconsistencies in internal structure, likely due to differences in how the items relate to their respective traits. **Cronbach's Alpha** (0.23) also shows moderate consistency but is influenced by the assumption of equal item contribution, which doesn't hold here, as shown by the heatmap's varying correlation patterns.

Method	Value
Split-Half Reliability	0.15395
Cronbach's Alpha	0.23
McDonald's Omega	0.15278

Conclusion:

In this study, we demonstrated how to use SAS for comprehensive reliability analysis of questionnaire data. By applying methods such as Split-Half Reliability, Cronbach's Alpha, and McDonald's Omega, we showcased how each approach can be utilized to assess internal consistency. Although our results from simulated and real-world data indicated some limitations and varying reliability levels, the emphasis of this paper was on the application of these techniques rather than the reliability outcomes themselves. The diverse methodologies presented offer researchers the tools necessary for rigorous reliability evaluation, enhancing the robustness of questionnaire-based studies. This methodological approach is particularly useful for those who aim to ensure the accuracy and credibility of the insights derived from their survey data, ultimately contributing to more reliable research findings across multiple disciplines.

REFERENCE:

- Acharya, B. (2010). Questionnaire design. Central Department of population studies.
- Arsham, H. (2005). Questionnaire design and surveys sampling. University of Baltimore.
- Boparai, J. K., Singh, S., & Kathuria, P. (2018). How to design and validate a questionnaire: a guide. Current Clinical Pharmacology, 13(4), 214-222.
- El-Jardali, F., Dumit, N., Jamal, D., & Mouro, G. (2008). Migration of Lebanese nurses: a questionnaire survey and secondary data analysis. International Journal of Nursing Studies, 45(11), 1490-1500.
- Falissard, B. (2012). Analysis of questionnaire data with R.
- Zhang, Z., Das, A., Rahgouy, M., Bao, Y., & Baskiyar, S. (2023). Multi-label classification of CS papers using natural language processing models. 2023 International Conference on Machine Learning and Applications (ICMLA), 1907–1912.
- Gorter, J. P., Fox, J. P., & Twisk, J. W. R. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Medical Research Methodology, 15(1), 50.
- Ayine, P., Selvaraju, V., Venkatapoorna, C. M. K., Bao, Y., Gaillard, P., & Geetha, T. (2021). Eating behaviors in relation to child weight status and maternal education.

Children, 8(32).

- Harris, L. R., & Brown, G. T. L. (2019). Mixing interview and questionnaire methods: Practical problems in aligning data. Practical Assessment, Research & Evaluation, 15(1), 1.
- Ye, J., Poudel, A., Liu, J., Vinel, A., Silva, D., Shao, S., & Shamsaei, N. (2023). Machine learning augmented X-ray computed tomography features for volumetric defect classification in laser beam powder bed fusion. The International Journal of Advanced Ma
- Marshall, G. (2005). The purpose, design and administration of a questionnaire for data collection. Radiography, 11(2), 131-136.
- Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. Journal of Clinical Nursing, 16(2), 234-243.
- Sijtsma, K., & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. British Journal of Mathematical and Statistical Psychology, 70(1), 137-158.
- Ye, J., Yasin, M. S., Muhammad, M., Liu, J., Vinel, A., Slvia, D., ... & Shao, S. (2021). Bayesian process optimization for additively manufactured nitinol.
- Zhang, Z., Xu, L., Bao, Y., & Baskiyar, S. (2023). Towards the diagnosis of heart disease using an ensemble learning approach. 2023 International Conference on Machine Learning and Applications (ICMLA), 1901–1906.

Contact Information:

Your comments and questions are valued and encouraged. Contact the author at:

YIDA BAO, Assistant Professor, SAS certified advanced programmer

Mathematics, Statistics & Computer Science Department, University of Wisconsin-Stout

E-mail: yzb0010@auburn.edu

Philippe Gaillard, Associate Professor

Administration Department of Biostatistics, Data Science, & Epidemiology, Augusta University

E-mail: pgaillard@augusta.edu