

How to Check for Correlation Between Variables Using Bartlett's Test of Sphericity

Authored by
stats writer

December 30, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Check for Correlation Between Variables Using Bartlett's Test of Sphericity*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=110024>

Introduction to Bartlett's Test of Sphericity

Bartlett's Test of Sphericity is a crucial statistical procedure in multivariate analysis, specifically designed to evaluate the presence of significant interrelationships among a set of variables. Developed by Maurice Stevenson Bartlett, this test provides a formal statistical confirmation that the data structure is appropriate for subsequent dimensionality reduction techniques. It serves as an essential gatekeeper before researchers apply methods intended to summarize variance, such as Factor Analysis or Principal Component Analysis (PCA).

The test operates by assessing the fundamental assumption that variables in a system are not statistically independent. It quantifies the degree to which the observed correlation matrix deviates from a theoretical structure where all correlations are zero. This comparison is vital because if variables are truly uncorrelated (orthogonal), attempts to combine them into fewer latent factors would be statistically unjustified and yield meaningless component structures.

Mathematically, the test calculates a statistic based on the determinant of the correlation matrix. This statistic asymptotically follows a chi-square distribution, allowing for the calculation of a p-value. If the determinant of the observed matrix is substantially smaller than what would be expected under the null hypothesis, it indicates strong correlation, justifying the use of dimension reduction.

The Core Purpose: Assessing Data Suitability

The primary function of Bartlett's Test of Sphericity is to verify that the dataset contains sufficient redundancy or shared variance among the variables. Data reduction techniques, such as Principal Component Analysis, rely entirely on the idea that the information contained within numerous measured variables can be summarized efficiently by a smaller number of underlying constructs or components. If the variables are independent, this reduction is impossible without severe loss of information.

Essentially, the test checks to see if the variables are correlated enough to support the extraction of factors. If the variables are orthogonal, meaning they stand alone and do not share variance, then attempting to compress them into linear combinations will not effectively summarize the data. The resulting factors would simply reflect random noise or negligible structures, failing the goal of parsimonious data representation.

Therefore, performing this test is a mandatory preliminary step in many advanced statistical pipelines. It ensures that the researcher is working with a data structure where variables exhibit meaningful interdependence. Only when the test yields a statistically significant result can the analyst proceed confidently, knowing that the data is structured to benefit from techniques that seek to identify latent variables or principal components.

Note: Bartlett's Test of Sphericity focuses exclusively on evaluating the pattern of correlation coefficients. It should not be confused with other tests that share the name "Bartlett's Test," particularly those used to test for homogeneity of variances (e.g., in ANOVA), which address a completely different statistical assumption.

Null and Alternative Hypotheses

To properly frame the results of Bartlett's Test of Sphericity, we must define the statistical hypotheses that govern the evaluation of the correlation structure. The core of any statistical test lies in distinguishing between a baseline assumption and the potential existence of a true effect.

The **Null Hypothesis** (H_0) asserts the condition of sphericity: that the population correlation matrix is an identity matrix. Under this assumption, all variables are perfectly orthogonal; that is, the off-diagonal correlation coefficients are zero. If we fail to reject this null hypothesis, we conclude that the variables are independent, making data compression inappropriate.

The **Alternative Hypothesis** (H_a) proposes that the population correlation matrix is not an identity matrix. This is the condition researchers hope to confirm. Acceptance of H_a indicates that there are statistically significant correlations among at least some of the variables, implying that the data contains the necessary redundancy to be summarized by common factors.

The test procedure is designed to determine if the observed data provides enough evidence to reject the parsimonious, yet restrictive, scenario defined by the null hypothesis. A significant test result (rejection of H_0) is the prerequisite evidence needed to proceed with the primary objective of data reduction.

Understanding the Correlation Matrix

A correlation matrix is fundamental to multivariate statistics, providing a symmetric square table that displays the pairwise correlation coefficients between all variables in a dataset. If a dataset has p variables, the matrix will be p times p . Understanding its structure is key to understanding what Bartlett's test is evaluating.

By convention, the elements along the main diagonal (where a variable correlates with itself) are always 1.0. The elements in the off-diagonal positions represent the strength and direction of the linear relationship between two distinct variables. These values range from -1.0 to +1.0. Coefficients close to zero signify weak or non-existent linear correlation, while coefficients closer to the extremes (1 or -1) indicate stronger relationships and greater shared variance.

Consider the correlation matrix below, which might represent interrelationships between various performance metrics for professional basketball teams. The non-zero values off the diagonal

indicate some degree of shared information.

NBA Correlation Matrix				
	Age	Wins	True Shooting %	Turnover %
Age	1.00	0.48	0.56	-0.29
Wins		1.00	0.77	-0.15
True Shooting %			1.00	0.10
Turnover %				1.00

The goal of Bartlett's test is to quantify whether this collective set of off-diagonal values is significantly different from zero. If these coefficients were all zero, the variables would be orthogonal, and the matrix would collapse into the structure of the identity matrix.

The Role of the Identity Matrix

The identity matrix (I) serves as the theoretical benchmark for the null hypothesis in Bartlett's test. It is a square matrix defined by having ones along the main diagonal and zeros everywhere else.

Interpreted as a correlation matrix, the identity matrix signifies a perfect scenario of zero correlation among all pairs of distinct variables. The ones on the diagonal reflect that each variable is perfectly correlated with itself, but the zeros in the off-diagonal positions mean that no variable shares variance with any other variable.

Identity Matrix				
	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	1	0	0	0
Variable 2	0	1	0	0
Variable 3	0	0	1	0
Variable 4	0	0	0	1

If the data's observed correlation matrix closely approximates this identity matrix, it implies that the variables are highly independent. In such a case, a data reduction technique like Principal Component Analysis would fail to compress the data in any meaningful way, as there is no latent structure to uncover. Thus, the reason we conduct Bartlett's Test of Sphericity is specifically to ensure that the correlation matrix of our variables diverges significantly from this orthogonal ideal,

thereby confirming that data reduction methods are suitable.

Interpreting the Test Statistic and P-Value

The computational heart of Bartlett's test involves converting the structural information of the correlation matrix into a single test statistic, which is then assessed using the chi-square distribution. The transformation utilizes the determinant of the matrix, rewarding smaller determinants (which occur when high correlations exist) with larger test statistics.

The resulting chi-square statistic is compared to a critical value associated with the degrees of freedom (calculated as $p(p-1)/2$, where p is the number of variables) and the pre-selected significance level (α , commonly 0.05). If the calculated statistic is larger than the critical value, it suggests that the observed pattern of correlations is unlikely to have occurred purely by chance if the variables were truly uncorrelated.

The p-value provides the definitive measure for decision-making. It represents the probability of observing a correlation matrix as extreme as, or more extreme than, the one observed, assuming the null hypothesis of sphericity is true. If the p-value falls below the chosen significance level (e.g., $p < 0.05$), we reject H_0 . This rejection is the desired outcome, confirming that sufficient correlation exists for data reduction.

If the p-value is large ($p > \alpha$), we fail to reject the null hypothesis. This implies that the correlations are not collectively significant enough to warrant further multivariate analysis. In this case, proceeding with Factor Analysis or PCA is generally advised against, as the derived factors would be unstable and lack substantive meaning.

Practical Application in R using `cortest.bartlett()`

To conduct Bartlett's Test of Sphericity in the R environment, the researcher typically utilizes the **psych** package, which provides the convenient **cortest.bartlett()** function. This function requires the input of the correlation matrix (R) and the sample size (n) from which the matrix was derived.

The use of the sample size (n) is critical because the chi-square distribution is only an asymptotic approximation of the true distribution of the test statistic. The larger the sample size, the more accurate this approximation becomes, which affects the reliability of the calculated p-value.

The general syntax for implementing this statistical check is concise, requiring that the preliminary step of calculating the correlation matrix has already been completed:

```
cortest.bartlett(R, n)
```

R: The input correlation matrix derived from the dataset.

n: The total sample size (number of rows) used to generate the correlation matrix.

The following R demonstration illustrates the full process, from generating sample data to calculating the matrix and performing the test. Note how the resulting p-value dictates the conclusion regarding data suitability.

Set seed for reproducibility of random data generation

set.seed(0)

Create a sample data frame with 50 observations for three weakly correlated variables (A, B, C)

```
data <- data.frame(A = rnorm(50, 1, 4), B = rnorm(50, 3, 6), C = rnorm(50, 5, 8))
```

Inspect the structure of the data by viewing the first few rows

```
head(data)
```

```
# A B C
```

```
#1 6.0518171 4.5968242 11.25487348
```

```
#2 -0.3049334 0.7397837 -1.21421297
```

```
#3 6.3191971 17.6481878 0.07208074
```

```
#4 6.0897173 -1.7720347 5.37264242
```

```
#5 2.6585657 2.6707352 -4.04308622
```

```
#6 -5.1598002 4.5008479 9.61375026
```

Calculate the correlation matrix (R) from the sample data

```
cor_matrix <- cor(data)
```

Display the calculated correlation matrix

```
cor_matrix
```

```
# A B C
```

```
#A 1.0000000 0.1600155667 0.2825308511
```

```
#B 0.1600156 1.0000000000 0.0005358384
```

```
#C 0.2825309 0.0005358384 1.0000000000
```

Load the required psych library

```
library(psych)
```

Execute Bartlett's Test of Sphericity using the calculated matrix and sample size (n)

```
cortest.bartlett(cor_matrix, n = nrow(data))
```

```
#$chisq
```

```
# 5.252329
```

```
#  
#$p.value  
# 0.1542258  
#  
#$df  
# 3
```

Conclusion: Interpreting the R Results

The output from the `cortest.bartlett()` function provides three critical values: the Chi-Square test statistic, the p-value, and the degrees of freedom. In this specific scenario, the Chi-Square statistic is **5.252329**, and the corresponding p-value is **0.1542258**.

When comparing the p-value (0.1542258) to a conventional significance level ($\alpha = 0.05$), we find that the p-value is substantially higher. Based on this outcome, we **fail to reject the null hypothesis** of sphericity.

The interpretation is clear: the observed correlations among variables A, B, and C are statistically weak and do not collectively differ enough from zero to justify the assumption of underlying factors. The variables are largely orthogonal. Consequently, this dataset is likely not suitable for advanced data reduction methods like Principal Component Analysis or Factor Analysis. Attempting to force a data compression model onto such a dataset would result in factors that explain minimal variance and offer little practical insight.

The conclusion reinforces the importance of using Bartlett's Test of Sphericity as a mandatory diagnostic. Only when the test yields a statistically significant rejection of the null hypothesis can researchers be confident that the data possesses a robust structure of interdependencies suitable for effective dimensional simplification.