

RESEARCH PAPER

Statistical Verification of Facet Intercorrelations Using Parametric Bootstrapping in Differential Personality Diagnostics

Tobey Gross

Chairman, International Council of Academics for Progressive Education (ICAPE), Germany

Corresponding author: chair@icape-edu.com

Received: 17 Sept., 2024

Revised: 30 Nov., 2024

Accepted: 05 Dec., 2024

ABSTRACT

Building on our previous work that examined intra-individual intercorrelations in differential psychological diagnostics, this paper focuses on the statistical verification of normative facet intercorrelations using parametric bootstrapping^[2]. Traditional methods rely on Pearson correlations and z-scores to compare subject profiles to a normative matrix^{[6][1]}. However, these approaches assume a normal distribution of correlations—a premise that may not hold under all conditions^[4]. Here, we describe how parametric bootstrapping, with direct normal approximation and Fisher-z transformation^{[2][3]}, can empirically assess the variability and distributional characteristics of correlation coefficients. The insights gained help establish an objective diagnostic threshold for identifying significant deviations in individual profiles. This methodological enhancement is critical to refining personality diagnostics, merging rigorous statistical validation with clinical applicability.

Keywords: Differential Psychology, Diagnostics, Parametric Bootstrapping, Facet Intercorrelations, Pearson Correlation, Fisher-z

Previously, we reviewed conventional techniques in differential psychological diagnostics, underscoring the limitations inherent in comparing individual personality profiles solely via traditional z-score computations and Pearson correlation coefficients^[6]. While such methods offer a useful starting point, they assume that the intercorrelations observed in a normative population accurately capture the dynamics at the individual level^[9]. This assumption, however, can obscure subtle yet diagnostically significant deviations within a subject's profile.

The present work extends that analysis by embedding a statistical verification procedure—parametric bootstrapping^[2]—into the diagnostic framework. We focus on assessing whether the normative facet intercorrelations are sufficiently normally distributed to justify the application of a $\pm 2\sigma$ threshold^[1].

How to cite this article: Gross, T. (2024). Statistical Verification of Facet Intercorrelations Using Parametric Bootstrapping in Differential Personality Diagnostics. *IJISC.*, 11(02): 83-93.

Source of Support: None; **Conflict of Interest:** None



This step is essential because deviations from normality may lead to misinterpretations when applying conventional diagnostic thresholds^[4].

Specifically, we outline two variants of bootstrapping: one based on the direct normal approximation and the other utilizing the Fisher- z transformation^[3]. These approaches allow us to generate empirical confidence intervals for correlation estimates^[7] and objectively identify when a subject's intercorrelation matrix diverges from the norm. By validating these assumptions statistically, we integrate rigorous methodological improvements with clinical insights, ultimately advancing the precision of personality diagnostics.

In the sections that follow, we detail the theoretical rationale, implementation steps, and diagnostic implications of our parametric bootstrapping approach, thereby offering a refined toolset for practitioners in both research and applied settings.

BACKGROUND

In the present analysis, the intercorrelations among the 30 facets f in a normative population N are examined. The facet values are based on the means of the individual item components that were rated on a Likert scale. The intercorrelations in the normative matrix R^N are calculated as Pearson correlations between each pair of f_n, f_m ^[6]:

$$r(f_n, f_m) = \frac{\text{cov}(f_n, f_m)}{\sigma(f_n) \cdot \sigma(f_m)} \quad \dots(1)$$

$$r(f_1^X, f_2^Y) = \frac{\frac{1}{N} \sum_{i=1}^N (f_{1,i}^X - \bar{f}_1^X)(f_{2,i}^Y - \bar{f}_2^Y)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (f_{1,i}^X - \bar{f}_1^X)^2} \cdot \sqrt{\frac{1}{N} \sum_{i=1}^N (f_{2,i}^Y - \bar{f}_2^Y)^2}} \quad \dots(2)$$

A central goal of this analysis is to determine the diagnostic relevance of the deviations of individual (subject) correlations from the normative correlations^[9]. For this purpose, the subject matrix R^P has been constructed by computing the correlations for a single sample with a modified covariance formula. The difference matrix

$$D = R^P - R^N \quad \dots(3)$$

shows the deviations of r^P from r^N and serves as the basis for analyzing the individual differences.

To determine the diagnostic relevance, it is necessary to establish an *objective* threshold above which deviations are considered significant. A commonly used statistical measure is the threshold of $\pm 2\sigma^{[1]}$, with values outside this range deemed unusual. The application of this threshold assumes, however, that the correlations in the normative matrix R^N approximately follow a normal distribution^[4]:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95 \quad \dots(4)$$

Since correlations do not necessarily follow a normal distribution—especially when derived from a limited sample or when the underlying variables are not normally distributed—it is necessary to verify this assumption empirically.

Important methodological note: If *raw data* (item-level) are available, one could perform a classical *case resampling* (non-parametric bootstrapping)^{[2][8]}, i.e., drawing subjects with replacement and recalculating a full 30×30 correlation matrix each time. However, if *no raw data* are available, this is not possible. In that case, only a *parametric* approach remains feasible, which simulates the *distribution* of the correlations approximately using standard error and an appropriate transformation (e.g., Fisher-z).

In the following, we will use such *parametric bootstrapping* to test whether the correlations in the normative matrix R^N are sufficiently normally distributed (or at least symmetric enough) to justify using the $\pm 2\sigma$ threshold. Should the distributions deviate too strongly from a normal form, a percentile approach (e.g., the 2.5%- and 97.5% percentiles) or a more robust method (e.g., rank tests) would be appropriate^[5].

PARAMETRIC BOOTSTRAPPING TO VERIFY NORMAL DISTRIBUTION

The *parametric* bootstrapping, as performed here, generates random correlations around the *estimated* correlation matrix R^N by repeated simulation (B times). This is done using standard error formulas (and optionally the Fisher-z transformation)^[3]. The goal is to determine whether the correlations in R^N appear normally distributed in order to justify the application of the $\pm 2\sigma$ threshold.

General Formulation (Parametric Approach)

In contrast to the *non-parametric* approach (resampling of raw data), here we assume that each correlation r fluctuates around a mean $\hat{r}$ and that the estimation error is approximately normally distributed. Alternatively, one can add noise on the Fisher-z scale to ensure that the correlations remain in the allowed range $[-1, 1]$.

Variant A: Direct Normal Approximation

$$r_b^*(f_n, f_m) = \hat{r}(f_n, f_m) + \varepsilon b, \quad \varepsilon b \sim N(0, SE_r) \quad \dots(5)$$

This variant can lead to $\hat{r}^* > 1$ or $\hat{r}^* < -1$ when ε_b is large. For moderate $|\hat{r}|$ and large N , this problem is often negligible.

Variant B: Fisher-z Transformation

A more robust variant uses;

$$z = \frac{1}{2} \left(\frac{1+r}{1-r} \right), \text{ and } z^* = z + \eta_b, \quad \eta_b \sim N(0, \sigma_z), \quad \dots(6)$$

and then transforms back using;

$$\hat{r}^* = \frac{e^{2z^*} - 1}{e^{2z^*} + 1} \quad \dots(7)$$

In this way, the simulated correlations cannot fall outside of $[-1, 1]$, and the approximation is generally better^[3].

Calculation of the Standard Error of the Correlation

Regardless of whether Variant A or B is used, an approximation for the standard error SE_r is needed. A common formula is:

$$SE_r \approx \sqrt{\frac{1-r^2}{N-1}}, \quad \dots(8)$$

where N is the sample size and r is the estimated correlation. For the Fisher-z scale, one typically uses (for large N):

$$\sigma_z \approx \frac{1}{\sqrt{N-3}}. \quad \dots(9)$$

Bootstrapping Process

1. Simulate B correlations (e.g., 10,000) for each pair of facets, either using $\hat{r} + N(0, SE_r)$ or via Fisher-z transformation^[2].
2. This yields an empirical distribution $\{r_b^*\}_{b=1}^B$ for each pair.
3. From this, one can calculate an empirical 95% confidence interval (e.g., the 2.5% and 97.5%-percentiles) and an estimate of the standard deviation.
4. To test the assumption that $\hat{r}$ (or $\hat{z}$) is normally distributed, one may perform a normality test (e.g., Shapiro–Wilk)^[4].

Diagnostic Threshold and Multiple Testing

If the normality assumption for most facet pairs is not violated, the threshold $\pm 2\sigma$ (or equivalently a $\pm 2 \cdot SE_r$ interval) can be used to identify unusual deviations. Alternatively, one may use the 95% confidence interval from the bootstrapping distribution.

It should be noted, however, that with 30 facets there are $\binom{30}{2} = 435$ pairs. A *pairwise* test without correction could lead to an increased number of seemingly “significant” deviations (multiple testing problem). In practice, one should consider an appropriate correction or a global test^{[5][8]}.

PARAMETRIC BOOTSTRAPPING IN PYTHON (EXAMPLE)

Below is an example code that applies *Variant A* (simple normal approximation per correlation). In addition, a second function is provided that uses the Fisher-z transformation. Finally, an example is given on how to determine a 95% confidence interval and test subject values R^p against it.

```

import numpy as np
import scipy . stats as stats

#####
# Example normative correlation matrix (3 x3) -- in practice 30 x30:
#####
R_N = np. array ([
    [1.0 , 0.5 , 0.3] ,
    [0.5 , 1.0 , 0.2] ,
    [0.3 , 0.2 , 1.0]
])

N = 727      # Sample size of the normative sample
B = 10000    # Number of bootstrapping iterations

#####
# Standard error of the correlation ( approximation ):
#####
def standard_error_correlation (r, n):
    # Check that  $|r| < 1$ :
    return np. sqrt ((1 - r **2) / (n - 1))

#####
# Parametric Bootstrap ( Variant A:  $r + N(0, SE_r)$  ):
#####
def bootstrap_correlations_normal (R_N , N, B):
    num_facets = R_N. shape [0]
    bootstrapped_correlations = np. zeros ((B, num_facets , num_facets ))

    # Upper triangular indices ( without diagonal )
    i_idx , j_idx = np. triu_indices ( num_facets , k=1)

    for b in range (B):
        # Copy of the original matrix ( Diagonal = 1)

```

```

matrix_b = np. copy (R_N)
# For each pair i,j in the upper triangular matrix :
for i, j in zip(i_idx , j_idx ):
    r_ij = R_N [i, j]
    SE_r = standard_error_correlation (r_ij , N)

    # Normally distributed disturbance :
    r_star = r_ij + np. random . normal (0, SE_r )
    # Note : r_star can be >1 or < -1 ->
    # can happen with larger SE_r .

    # Symmetric entry :
    matrix_b [i, j] = r_star
    matrix_b [j, i] = r_star

# Keep diagonal = 1.0:
np. fill_diagonal ( matrix_b , 1.0)

bootstrapped_correlations [b] = matrix_b

return bootstrapped_correlations

# #####
# BETTER Variant B: Fisher -z Transformation ( always in ( -1 ,1)):
# #####
def fisher_z (r):
    # Consider edge cases :
    if np.abs (r) >= 0.999999:
        r = 0.999999 * np. sign (r)
    return 0.5 * np.log ((1 + r)/(1 - r))

def inv_fisher_z (z):
    return (np.exp (2* z) - 1) / (np.exp (2* z) + 1)

```

```

def bootstrap_correlations_fisher (R_N , N, B):
    num_facets = R_N. shape [0]
    bootstrapped_correlations = np. zeros ((B, num_facets , num_facets ))
    i_idx , j_idx = np. triu_indices ( num_facets , k=1)

    # SE on the z- scale
    # ( Rule of thumb : 1/ sqrt (N -3) , depending on the exact formula .)
    se_z = 1.0 / np. sqrt (N - 3)

    for b in range (B):
        matrix_b = np. copy (R_N)
        for i, j in zip(i_idx , j_idx ):
            r_ij = R_N [i, j]
            # Convert to z- scale :
            z_ij = fisher_z ( r_ij )

            # Normally distributed disturbance on the z- scale :
            z_star = z_ij + np. random . normal (0, se_z )

            # Back - transformation :
            r_star = inv_fisher_z ( z_star )

            matrix_b [i, j] = r_star
            matrix_b [j, i] = r_star

        np. fill_diagonal ( matrix_b , 1.0)
        bootstrapped_correlations [b] = matrix_b

    return bootstrapped_correlations

# #####
# 95% - Confidence Interval from Bootstrap :
# #####

```

```
def compute_confidence_intervals ( bootstrapped_matrices , alpha =0.05) :
    lower_bound = np. percentile ( bootstrapped_matrices , 100*( alpha /2) ,
    axis =0)
    upper_bound = np. percentile ( bootstrapped_matrices , 100*(1 - alpha /2) ,
    axis =0)
    return lower_bound , upper_bound

#####
# Example subject correlation matrix R^P:

#####
R_P = np. array ([
    110 [1.0 , 0.4 , 0.1] ,
    111 [0.4 , 1.0 , 0.5] ,
    112 [0.1 , 0.5 , 1.0]
])
#####
# Function to evaluate significant deviations
# ( without multiple test correction ):
#####
def evaluate_significance (R_P , lower_ci , upper_ci ):
    return (R_P < lower_ci ) | (R_P > upper_ci )

#####
# DEMO RUN: Here using simple normal approximation ( Variant A)
#####
bootstrapped_matrices_normal = bootstrap_correlations_normal (R_N , N, B)
lower_ci_normal , upper_ci_normal = compute_confidence_intervals (
    bootstrapped_matrices_normal )
signif_normal = evaluate_significance (R_P , lower_ci_normal ,
    upper_ci_normal )

print (“ Significant deviations ( Variant A, Normal approximation ):”)
print ( signif_normal )
```



```
# #####
# DEMO RUN: Fisher -z Transformation ( Variant B)
# #####

bootstrapped_matrices_fisher = bootstrap_correlations_fisher (R_N , N, B)
lower_ci_fisher , upper_ci_fisher = compute_confidence_intervals (
    bootstrapped_matrices_fisher )
signif_fisher = evaluate_significance (R_P , lower_ci_fisher ,
    upper_ci_fisher )
print (" Significant deviations ( Variant B, Fisher -z):")
print ( signif_fisher )

# #####
# Shapiro - Wilk Test for Normality ( only as an example on ONE pair ):
# #####

def test_normality ( bootstrapped_matrices , i=0, j=1):
    sample_boot_corrs = bootstrapped_matrices[:, i, j]
    stat , p_value = stats . shapiro ( sample_boot_corrs )
    return stat , p_value

shapiro_stat , shapiro_p = test_normality ( bootstrapped_matrices_fisher ,
    0, 1)
print (f" Shapiro - Wilk ( Pair (0 ,1)), W={ shapiro_stat :.4f}, p={ shapiro_p
    :.4 g}")
if shapiro_p > 0.05:
    print ("No significant deviation from normality according to Shapiro
    - Wilk .")
else :
    print (" Significant deviation (p < 0.05) - Not normally distributed .
```

Listing 1: Parametric Bootstrapping in Python

Explanations

1. **Normal Approximation vs. Fisher-z:** The function bootstrap correlations normal shows a quick, simple variant, which can also produce $\hat{r}^* \notin [-1, 1]$.
bootstrap correlations fisher uses the Fisher-z transformation, thus guaranteeing correct values within $[-1, 1]$.
2. **Confidence Intervals:** The percentiles (2.5% / 97.5%) are used as estimators for the 95% interval. If a subject correlation $\hat{r}^P$ lies outside this interval, it is marked as “significantly deviating.”
3. **Normality Test:** The Shapiro-Wilk test is applied as an example only on the simulated values for a single correlation pair $((0, 1))$. For many pairs (e.g., 435 in a 30x30 matrix without diagonal), multiple tests would need to be performed, which might require appropriate correction.

Mathematical Explanation of the $\pm 2SD$ Threshold

In the preceding explanations, we interpret deviations of the subject correlation $r^P(f_i, f_j)$ from the normative correlation $\hat{r}^N(f_i, f_j)$ as *significant* once;

$$|r^P - \hat{r}^N| > 2 \cdot \sigma(\hat{r}^N) \quad \dots(10)$$

Here, $\sigma(\hat{r}^N)$ denotes the *standard deviation* (in the sense of the *standard error*) of the estimated normative correlation $\hat{r}^N$. From a sampling theoretical perspective, $\hat{r}^N$ can be regarded as a random variable with an expected value (the *true* correlation) and a dispersion $\sigma(\hat{r}^N)$. The latter is determined either:

1. via bootstrapping methods (empirical approximation), or
2. using the known formula for standard error estimation

$$SE_r \approx \sqrt{\frac{1 - r^2}{N - 1}}, \quad \dots(11)$$

where in Equation (11) r stands for the correlation estimator and N for the size of the normative sample^[1].

The choice of the threshold $\pm 2\sigma(\hat{r}^N)$ corresponds to a classic rule of thumb from the normal distribution approximation. Under sufficient conditions, it holds that:

$$P(\hat{r}^N - 2\sigma(\hat{r}^N) \leq r_{true} \leq \hat{r}^N + 2\sigma(\hat{r}^N)) \approx 0.95 \quad \dots(12)$$

where r_{true} denotes the unknown *true* (population) correlation. If the subject correlation r^P lies *outside* this interval, it is classified as “significantly deviating^[1].” Thus, it is clear that the used unit $\pm 2SD$ does not originate “out of thin air” but is directly based on the estimated *variability* of the correlation coefficient from the normative sample.

CONCLUSION

Based on the presented (parametric) bootstrapping methods, it can be empirically verified whether a normal distribution approximation for the correlations of the normative matrix $\hat{R}^N$ is sufficiently tenable to apply the diagnostic threshold;

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95 \quad \dots(13)$$

Otherwise, a percentile-based approach (e.g., 2.5%- and 97.5%-percentiles) is preferred^[2].

In practice, heatmaps or overviews can be created that show for which facet pairs a subject's $\hat{R}^P$ significantly deviates from the norm (i.e., lies outside the confidence interval). Diagnostically, this allows a refined consideration of individual psychological abnormalities and their relevance in the context of the norm. However, with many facet pairs, multiple testing corrections should be considered in order to control the error probability^[5].

REFERENCES

1. Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
2. Efron, B. and Tibshirani, R.J. 1993. *An introduction to the bootstrap*. Chapman & Hall/CRC.
3. Fisher, R.A. 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, **1**: 3–32.
4. Shapiro, S.S. and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3–4), 591–611.
5. Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, **57**(1): 289–300.
6. Rodgers, J.L. and Nicewander, W.A. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician*, **42**(1): 59–66.
7. Cumming, G. and Finch, S. 2001. Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, **56**(3): 182–189.
8. Westfall, P.H. and Young, S.S. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Wiley.
9. Dunn, O.J. and Clark, V.A. 1969. Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, **64**(325): 366–377.

