Supplementary Material S1

Statistical Methodology for Golden Ratio Deviation Analysis

Supplementary Material for:
"Phi and Alzheimer's Disease: Is the Tree Drawing Test for diagnosing cognitive impairment an inner view of the golden proportion?"

Michelangelo Stanzani Maserati[1], Fabiana Zama[2]

Author Affiliations:

[1]IRCCS Istituto delle Scienze Neurologiche, Bologna, Italy
[2]Department of Mathematics, University of Bologna, Italy

Corresponding Author:
Fabiana Zama
Department of Mathematics
Email: fabiana.zama@unibo.it

Abstract

This supplementary material provides technical details for the statistical methodology employed in the analysis of golden ratio indices: $H/\varphi - T$, $C - H/\varphi^2$, $C - H/\varphi$, $H/\varphi^2 - T$, Δ_{unified}. The document describes computational implementation, normality assessment procedures, multiple comparison corrections, and complete statistical results for all tested indices. These methods support the main findings regarding the discriminative power of trunk-based golden ratio measures in cognitive impairment assessment using the Tree Drawing Test.

Contents

1 Detailed Statistical Methodology for Golden Ratio Deviation Analysis **2**
1.1 Computational Implementation 2
1.2 Statistical Analysis Workflow 2

2 Statistical Results **3**
2.1 Results for sex(f) group 4
2.2 Results for sex(m) group 5
2.3 Results for Education ≤ 5 years 5
2.4 Results for Education > 5 years 6

1 Detailed Statistical Methodology for Golden Ratio Deviation Analysis

1.1 Computational Implementation

All statistical analyses were implemented using MATLAB R2024b [1] with functions from the Statistics and Machine Learning Toolbox. The analysis framework employed the following built-in functions:

- `kstest()`: Kolmogorov-Smirnov normality testing [2] for each diagnostic group.
- `anova1()`: One-way analysis of variance for normally distributed data, returning F-statistics and ANOVA tables [3].
- `kruskalwallis()`: Non-parametric Kruskal-Wallis test [4] for non-normal distributions, providing H-statistics

Custom MATLAB functions and data processing scripts, is available upon request.

1.2 Statistical Analysis Workflow

Test Selection Protocol

Prior to hypothesis testing, distributional assumptions were evaluated using the Kolmogorov-Smirnov test ($\alpha = 0.05$) applied to each diagnostic group separately. While the Shapiro-Wilk test offers greater statistical power for smaller samples, the Kolmogorov-Smirnov test was selected for consistency across varying group sizes in demographic stratifications.

The following decision protocol was implemented:

- If normality satisfied across ALL groups ($p > 0.05$): One-way ANOVA using `anova1()`;
- If any group violated normality: Kruskal-Wallis test using `kruskalwallis()`.

This group-wise approach ensures that parametric assumptions are verified at the appropriate level for valid statistical inference.

Effect Size Calculation Effect sizes were quantified using eta-squared (η^2), which represents the proportion of total variance in the dependent variable explained by group membership [6]. This measure provides crucial information about the practical significance of observed differences, indicating not only statistical differences but also substantial effect magnitudes [7].

For normally distributed data analyzed with ANOVA, eta-squared was calculated using the classical formula:

$$\eta^2 = \frac{SS_{between}}{SS_{total}} \tag{1}$$

where $SS_{between}$ represents the sum of squares between groups and SS_{total} represents the total sum of squares. This formulation directly quantifies the proportion of total variance attributable to diagnostic group differences [6].

For non-normally distributed data analyzed with the Kruskal-Wallis test, eta-squared was approximated using the formula proposed by [8]:

$$\eta^2 \approx \frac{H - k + 1}{n - k} \tag{2}$$

where H is the Kruskal-Wallis test statistic, k is the number of groups (3 in our case), and n is the total sample size. As noted in the statistical literature, such approximations may have limitations regarding range validity and sensitivity to data characteristics [9], though validation confirmed all computed values remained within the theoretical $[0, 1]$ bounds.

This approximation provides a meaningful effect size measure for rank-based analyses while maintaining interpretability consistent with parametric eta-squared values.

The interpretative framework follows established conventions:

- **Small effect**: $\eta^2 = 0.01$ (1% of variance explained)
- **Medium effect**: $\eta^2 = 0.06$ (6% of variance explained)
- **Large effect**: $\eta^2 = 0.14$ (14% of variance explained)

Comparison Correction Implementation of False Discovery Rate (FDR) control using the Benjamini-Hochberg procedure. Given the simultaneous evaluation of five distinct golden ratio deviation indices, multiple testing correction was essential to control the familywise error rate. The False Discovery Rate (FDR) method with Benjamini-Hochberg correction [5] was selected because it offers superior balance between protection against false positives and preservation of statistical power.

The FDR procedure controls the expected proportion of false discoveries among rejected hypotheses at level $\alpha = 0.05$:

$$\text{FDR} = E\left[\frac{V}{R}\right] \leq \alpha \tag{3}$$

where V represents the number of false discoveries and R the total number of rejected hypotheses.

The Benjamini-Hochberg procedure operates by:

1. Ordering p-values from smallest to largest: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
2. Finding the largest k such that $p_{(k)} \leq \frac{k}{m}\alpha$
3. Rejecting hypotheses $H_{(1)}, H_{(2)}, \ldots, H_{(k)}$

2 Statistical Results

Table 1 presents statistical results for each golden ratio index across the complete dataset.

Golden Ratio Index	Test Statistic	p-value (raw)	p-value (adj)	Effect Size (η^2)
$H/\varphi - T$	F = 512.869	< 0.001	< 0.001	0.371*
$C - H/\varphi^2$	F = 519.477	< 0.001	< 0.001	0.376*
$C - H/\varphi$	F = 59.819	< 0.001	< 0.001	0.042*
$H/\varphi^2 - T$	F = 49.407	< 0.001	< 0.001	0.034*
Δ_{unified}	F = 163.578	< 0.001	< 0.001	0.117*

Table 1: **Statistical Analysis Results for Golden Ratio Deviation Measures.** Test statistics, significance levels, and effect sizes for five golden ratio-based indices across diagnostic groups (AD, MCI, CNTRL). All measures underwent normality testing using the Kolmogorov-Smirnov test; F-statistics are reported as all distributions satisfied parametric assumptions ($p > 0.05$ for normality). Raw p-values represent uncorrected significance levels; adjusted p-values incorporate False Discovery Rate correction using the Benjamini-Hochberg method ($\alpha = 0.05$). Effect sizes (η^2) quantify the proportion of total variance explained by diagnostic group membership. Significance levels after FDR correction for multiple comparisons: * p < 0.05.

Trunk-Based Measures: Superior Performance The indices $H/\varphi - T$ and $C - H/\varphi^2$ worked extremely well at distinguishing between the different patient groups, with effect sizes (0.371 and 0.376, respectively) substantially exceeding the large effect threshold ($\eta^2 = 0.14$). These measures explain approximately 37% of the total variance in Tree Drawing Test parameters across diagnostic groups.

The high F-statistics (512.869 and 519.477) indicate substantial between-group differences relative to within-group variability, providing robust evidence for genuine diagnostic differences rather than measurement noise.

Crown-Based Measures: Limited Discrimination In marked contrast, crown-based indices ($C - H/\varphi$ and $H/\varphi^2 - T$) demonstrate significantly reduced discriminative power despite achieving statistical significance. Their small effect sizes (0.042 and 0.034) suggest limited practical clinical utility, explaining less than 5% of total variance in the Tree Drawing Test parameters.

Unified Measure: Intermediate Performance The composite measure Δ_{unified} achieves a medium effect size (0.117), representing a balanced approach that incorporates multiple golden ratio relationships.

The remaining part of the text reports the tables related to classification performance and statistical analysis for sex and education stratification.

2.1 Results for sex(f) group

In this case the number of elements and percentage in each diagnostic group is the following:

$$\text{AD } n = 452(73.7\%), \quad \text{MCI } n = 184(56.1\%), \quad \text{CNTRL } n = 261(59.6\%)$$

	Mean ± SD			Classification Metrics			
	AD	**MCI**	**CNTRL**	**DDR**	**OC**	**FR**	**CV**
$H/\varphi - T$	5.50±11.10 [4.48, 6.53]	12.90±10.71 [11.34, 14.46]	34.06±26.90 [30.77, 37.34]	0.551	0.449	0.576	1.212
$C - H/\varphi^2$	5.15±10.88 [4.15, 6.16]	11.20±11.10 [9.58, 12.81]	34.07±26.57 [30.83, 37.31]	0.552	0.448	0.599	1.294
$C - H/\varphi$	10.91±11.78 [9.82, 11.99]	8.75±10.38 [7.24, 10.26]	2.65±20.05 [0.207, 5.60]	0.186	0.814	0.073	3.277
$H/\varphi^2 - T$	10.55±11.90 [9.45, 11.66]	7.05±9.59 [5.65, 8.44]	2.65±20.53 [0.153, 5.16]	0.184	0.816	0.067	3.407
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	5.00±4.98 [4.54, 5.46]	6.41±5.04 [5.68, 7.15]	11.78±11.72 [10.35, 13.21]	0.289	0.711	0.167	0.926

Table 2: Golden ratio deviation measures for Females (Sex(f)) group. DDR = Distance-to-Diameter Ratio; OC = Overlap Coefficient; FR = Fisher Ratio; CV = Coefficient of Variation. Values in brackets represent 95% confidence intervals for group means.

Index	F/H	p-raw	p-adj	η^2
$H/\varphi - T$	328.269	< 0.001	< 0.001	0.365*
$C - H/\varphi^2$	329.611	< 0.001	< 0.001	0.366*
$C - H/\varphi$	36.140	< 0.001	< 0.001	0.038*
$H/\varphi^2 - T$	32.427	< 0.001	< 0.001	0.034*
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	104.549	< 0.001	< 0.001	0.115*

Table 3: Statistical Test Results for Golden Ratio Deviation Measures for Females (Sex(f)) group. (* $p < 0.05$ after FDR correction for multiple comparisons).

Results in tables 2,3 confirm that Trunk-based measures ($H/\varphi - T$, $C - H/\varphi^2$) demonstrate superior discriminative power in Females with large effect sizes ($\eta^2 > 0.36$), while Crown-based measures show minimal utility ($\eta^2 < 0.04$). Pattern consistent with main analysis.

2.2 Results for sex(m) group

Sample size and percentage for each diagnostic group:

$$\text{AD } n = 161(26.3\%), \quad \text{MCI } n = 144(43.9\%), \text{CNTRL } n = 177(40.4\%)$$

	Mean ± SD			Classification Metrics			
	AD	**MCI**	**CNTRL**	**DDR**	**OC**	**FR**	**CV**
$H/\varphi - T$	4.51±9.35 [3.06, 5.96]	13.04±12.08 [11.05, 15.03]	33.12±27.89 [28.98, 37.26]	0.556	0.444	0.565	1.280
$C - H/\varphi^2$	4.12±9.52 [2.64, 5.61]	11.08±12.02 [9.10, 13.06]	33.12±27.89 [28.98, 37.26]	0.550	0.450	0.567	1.411
$C - H/\varphi$	11.63±11.83 [9.79, 13.48]	10.97±11.62 [9.06, 12.89]	0.84±21.30 [-2.32, 4.00]	0.221	0.779	0.124	9.111
$H/\varphi^2 - T$	11.25±11.79 [9.41, 13.08]	9.01±11.62 [7.10, 10.93]	0.84±21.30 [-2.32, 4.00]	0.219	0.781	0.105	9.199
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	5.33±4.93 [4.56, 6.10]	7.29±5.57 [6.37, 8.21]	13.26±13.11 [11.31, 15.20]	0.315	0.685	0.188	0.893

Table 4: Golden ratio deviation measures for Sex(m) group. DDR = Distance-to-Diameter Ratio; OC = Overlap Coefficient; FR = Fisher Ratio; CV = Coefficient of Variation. Values in brackets represent 95% confidence intervals for group means.

Index	F/H	p-raw	p-adj	η^2
$H/\varphi - T$	177.618	< 0.001	< 0.001	0.367*
$C - H/\varphi^2$	184.403	< 0.001	< 0.001	0.381*
$C - H/\varphi$	26.402	< 0.001	< 0.001	0.051*
$H/\varphi^2 - T$	19.555	< 0.001	< 0.001	0.037*
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	52.245	< 0.001	< 0.001	0.105*

Table 5: Sex(m). Statistical Test Results for Golden Ratio Deviation Measures. (* $p < 0.05$ after FDR correction for multiple comparisons).

Results in tables 4,5 show that Male subgroup has similar performance to Females, with trunk-based indices achieving large effect sizes ($\eta^2 > 0.36$) and superior classification metrics. Crown-based measures remain limited in discriminative capacity. Sex-independent classification utility confirmed.

2.3 Results for Education ≤ 5 years

Sample size and percentage for each diagnostic group:

$$\text{AD } n = 370(60.4\%), \quad \text{MCI } n = 175(53.4\%), \quad \text{CNTRL } n = 68(15.5\%)$$

Lower education group maintains significant discrimination for trunk-based measures ($\eta^2 = 0.23 - 0.26$), though with reduced magnitude compared to overall sample (Tables 6,7). Crown-based measures show very small effects ($\eta^2 < 0.02$), indicating limited utility in this population.

	Mean ± SD			Classification Metrics			
	AD	**MCI**	**CNTRL**	**DDR**	**OC**	**FR**	**CV**
$H/\varphi - T$	4.81±8.40 [3.95, 5.67]	12.40±10.97 [10.76, 14.03]	32.57±27.14 [25.10, 39.14]	0.568	0.432	0.577	1.155
$C - H/\varphi^2$	4.26±8.31 [3.41, 5.11]	9.40±10.85 [7.78, 11.02]	32.57±27.14 [25.10, 39.14]	0.559	0.441	0.588	1.313
$C - H/\varphi$	9.40±10.16 [8.36, 10.44]	9.64±10.77 [8.03, 11.25]	0.20±19.71 [-4.57, 4.97]	0.210	0.790	0.116	33.527
$H/\varphi^2 - T$	8.85±10.26 [7.80, 9.90]	6.64±9.95 [5.15, 8.12]	0.20±19.71 [-4.57, 4.97]	0.205	0.795	0.087	33.681
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	4.20±3.89 [3.80, 4.60]	6.62±5.26 [5.84, 7.41]	12.89±13.60 [9.60, 16.18]	0.365	0.635	0.233	0.925

Table 6: Golden ratio deviation measures for Education ≤ 5 years. DDR = Distance-to-Diameter Ratio; OC = Overlap Coefficient; FR = Fisher Ratio; CV = Coefficient of Variation. Values in brackets represent 95% confidence intervals for group means.

Index	F/H	p-raw	p-adj	η^2
$H/\varphi - T$	159.368	< 0.001	< 0.001	0.258*
$C - H/\varphi^2$	140.666	< 0.001	< 0.001	0.227*
$C - H/\varphi$	11.531	0.003	0.004	0.016*
$H/\varphi^2 - T$	8.878	0.012	0.012	0.011*
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	67.494	< 0.001	< 0.001	0.107*

Table 7: Education ≤ 5 years. Statistical Test Results for Golden Ratio Deviation Measures. (* $p < 0.05$ after FDR correction for multiple comparisons).

2.4 Results for Education > 5 years

Sample size and percentage for each diagnostic group:

$$\text{AD } n = 243(39.6\%), \quad \text{MCI } n = 153(46.6\%), \quad \text{CNTRL } n = 370(84.5\%)$$

	Mean ± SD			Classification Metrics			
	AD	**MCI**	**CNTRL**	**DDR**	**OC**	**FR**	**CV**
$H/\varphi - T$	5.91±13.40 [4.22, 7.60]	13.61±11.70 [11.74, 15.47]	33.89±27.33 [31.09, 36.68]	0.504	0.496	0.499	1.311
$C - H/\varphi^2$	5.83±13.19 [4.16, 7.50]	13.15±11.91 [11.25, 15.05]	33.89±27.10 [31.12, 36.66]	0.506	0.494	0.509	1.323
$C - H/\varphi$	13.68±13.53 [11.97, 15.39]	9.82±11.25 [8.02, 11.62]	2.24±20.72 [0.12, 4.36]	0.242	0.758	0.122	3.800
$H/\varphi^2 - T$	13.61±13.50 [11.90, 15.31]	9.36±11.06 [7.60, 11.13]	2.24±21.05 [0.09, 4.39]	0.241	0.759	0.118	3.858
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	6.44±6.03 [5.67, 7.20]	7.00±5.33 [6.15, 7.85]	12.28±12.07 [11.05, 13.52]	0.225	0.775	0.118	0.894

Table 8: Golden ratio deviation measures for Education > 5 years. DDR = Distance-to-Diameter Ratio; OC = Overlap Coefficient; FR = Fisher Ratio; CV = Coefficient of Variation. Values in brackets represent 95% confidence intervals for group means.

In Tables 8,9 we observe enhanced discrimination across all measures, with trunk-based indices achieving large effect sizes ($\eta^2 > 0.34$).

Stratified analyses confirm the robustness of trunk-based golden ratio measures across demographic subgroups. The ($H/\varphi - T$, $C - H/\varphi^2$) indices maintain superior discriminative power regardless of sex or education level.

Index	F/H	p-raw	p-adj	η^2
$H/\varphi - T$	266.711	< 0.001	< 0.001	0.347^*
$C - H/\varphi^2$	270.243	< 0.001	< 0.001	0.352^*
$C - H/\varphi$	59.493	< 0.001	< 0.001	0.075^*
$H/\varphi^2 - T$	56.087	< 0.001	< 0.001	0.071^*
$min(\|C - H/\varphi\|, \|H/\varphi - T\|)$	55.551	< 0.001	< 0.001	0.070^*

Table 9: Education > 5 years. Statistical Test Results for Golden Ratio Deviation Measures.(* $p < 0.05$ after FDR correction for multiple comparisons.)

References

[1] The MathWorks Inc. (2024). *MATLAB R2024b*. Natick, Massachusetts: The MathWorks Inc.

[2] Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68-78.

[3] Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Sage Publications, Thousand Oaks, CA.

[4] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.

[5] Benjamini, Y., & 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] Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.

[7] Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2011.02.001

[8] Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, *1*(21), 19–25.

[9] Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338