Full responses 26

Incomplete responses 32

Total responses 58

The mean of 'Q02' (experience) is: 4.83

The standard deviation of 'Q02' is: 1.47

The mean of 'Q04' (age) is: 3.04

The standard deviation of 'Q04' is: 0.77

Mean exp ≈ 17.98 years

Std dev exp ≈ 8.37 years

Mean age ≈ 35.43 years

Std dev age ≈ 7.67 years

- - - b\_histo - - -

Total participants: 23

High agreement (τ ≥ 0.7) : 10 (43.48%)

Medium agreement (0.5–0.7): 6 (26.09%)

Low agreement (τ < 0.5) : 7 (30.43%)

Mean: 0.62

SD: 0.22

- - - p\_histo - - -

Total participants: 23

High agreement (τ ≥ 0.7) : 1 (4.35%)

Medium agreement (0.5–0.7): 6 (26.09%)

Low agreement (τ < 0.5) : 16 (69.57%)

Mean: 0.33

SD: 0.27

- - - m\_histo - - -

Total participants: 23

High agreement (τ ≥ 0.7) : 8 (34.78%)

Medium agreement (0.5–0.7): 7 (30.43%)

Low agreement (τ < 0.5) : 8 (34.78%)

Mean: 0.53

SD: 0.31

**W is the Shapiro–Wilk statistic (close to 1 if normal).**

**p is the p‑value; if p < 0.05 you reject normality at the 5% level.**

Method B: W = 0.907, p = 0.035

Method P: W = 0.954, p = 0.362

Method M: W = 0.849, p = 0.003

The “omnibus” test is simply the global test that asks “Is there any difference across all three methods?” before you drill down into pairwise comparisons. Depending on your data you can choose:

If your omnibus p < 0.05, follow up with pairwise tests (Wilcoxon or paired‑t) with a correction (Bonferroni, Holm, etc.) to locate which methods differ.

#### **How to choose**

* **Friedman + Wilcoxon** is simpler and robust if you can’t assume normality.
* **ANOVA + paired‑t** can be more powerful if your data meet the normality assumption.

Either way, you first look for a significant omnibus test, then drill down with pairwise comparisons (with Bonferroni, Holm–Bonferroni, or similar) to see *which* methods differ from each other.

Friedman χ² = 12.667, p = 0.002

→ Significant differences exist across the three methods

B vs P: W = 21.5, raw p = 0.002, p₍adj₎ = 0.005 → ✓

B vs M: W = 64.0, raw p = 0.125, p₍adj₎ = 0.376 → n.s.

P vs M: W = 64.0, raw p = 0.042, p₍adj₎ = 0.127 → n.s.

Anova

====================================

F Value Num DF Den DF Pr > F

------------------------------------

method 9.1322 2.0000 44.0000 0.0005

====================================

Test Multiple Comparison ttest\_rel

FWER=0.05 method=bonf

alphacSidak=0.02, alphacBonf=0.017

============================================

group1 group2 stat pval pval\_corr reject

--------------------------------------------

B M 1.8368 0.0798 0.2394 False

B P 4.1281 0.0004 0.0013 True

M P 2.3648 0.0273 0.0818 False

--------------------------------------------

**RESPOSTA DO PEKKA REJEITADA**

Full responses 27

Incomplete responses 32

Total responses 59

The mean of 'Q02' (experience) is: 4.88

The standard deviation of 'Q02' is: 1.45

The mean of 'Q04' (age) is: 3.21

The standard deviation of 'Q04' is: 1.10

- - - - -

Mean exp ≈ 18.27 years

Std dev exp ≈ 8.31 years

Mean age ≈ 37.08 years

Std dev age ≈ 11.03 years

- - - b\_histo - - -

Total participants: 24

High agreement (τ ≥ 0.7) : 10 (41.67%)

Medium agreement (0.5–0.7): 6 (25.00%)

Low agreement (τ < 0.5) : 8 (33.33%)

Mean: 0.61

SD: 0.22

- - - p\_histo - - -

Total participants: 24

High agreement (τ ≥ 0.7) : 1 (4.17%)

Medium agreement (0.5–0.7): 6 (25.00%)

Low agreement (τ < 0.5) : 17 (70.83%)

Mean: 0.32

SD: 0.26

- - - m\_histo - - -

Total participants: 24

High agreement (τ ≥ 0.7) : 9 (37.50%)

Medium agreement (0.5–0.7): 7 (29.17%)

Low agreement (τ < 0.5) : 8 (33.33%)

Mean: 0.54

SD: 0.31

OLD v1

- - - b\_ - - -

Total participants: 21

High agreement (τ ≥ 0.7) : 8 (38.10%)

Medium agreement (0.5–0.7): 6 (28.57%)

Low agreement (τ < 0.5) : 7 (33.33%)

Mean: 0.61

S.D.: 0.22

- - - p\_ - - -

Total participants: 21

High agreement (τ ≥ 0.7) : 1 (4.76%)

Medium agreement (0.5–0.7): 6 (28.57%)

Low agreement (τ < 0.5) : 14 (66.67%)

Mean: 0.32

S.D.: 0.28

- - - m\_ - - -

Total participants: 21

High agreement (τ ≥ 0.7) : 6 (28.57%)

Medium agreement (0.5–0.7): 7 (33.33%)

Low agreement (τ < 0.5) : 8 (38.10%)

Mean: 0.51

S.D.: 0.32