

Intraclass Correlation Coefficient ICC(2,k)

Standalone verified Python analysis

Test: Intraclass Correlation Coefficient ICC(2,k)

Design: Two-way random-effects absolute agreement for the mean of all three grade ratings.

Variables: G1, G2, G3

Null hypothesis: The average-measure absolute-agreement ICC is zero.

Formula: $ICC(2,k) = (MSR - MSE) / [MSR + (MSC - MSE) / n]$.

Source rows: 649

Calculated metrics:

icc_2_k: 0.9480614162291533

f_statistic: 20.245973021325074

p_value: 0.0

targets: 649

raters: 3

Independent reference values:

icc_2_k: 0.9480614162291533

f_statistic: 20.245973021325074

p_value: 0.0

targets: 649

raters: 3

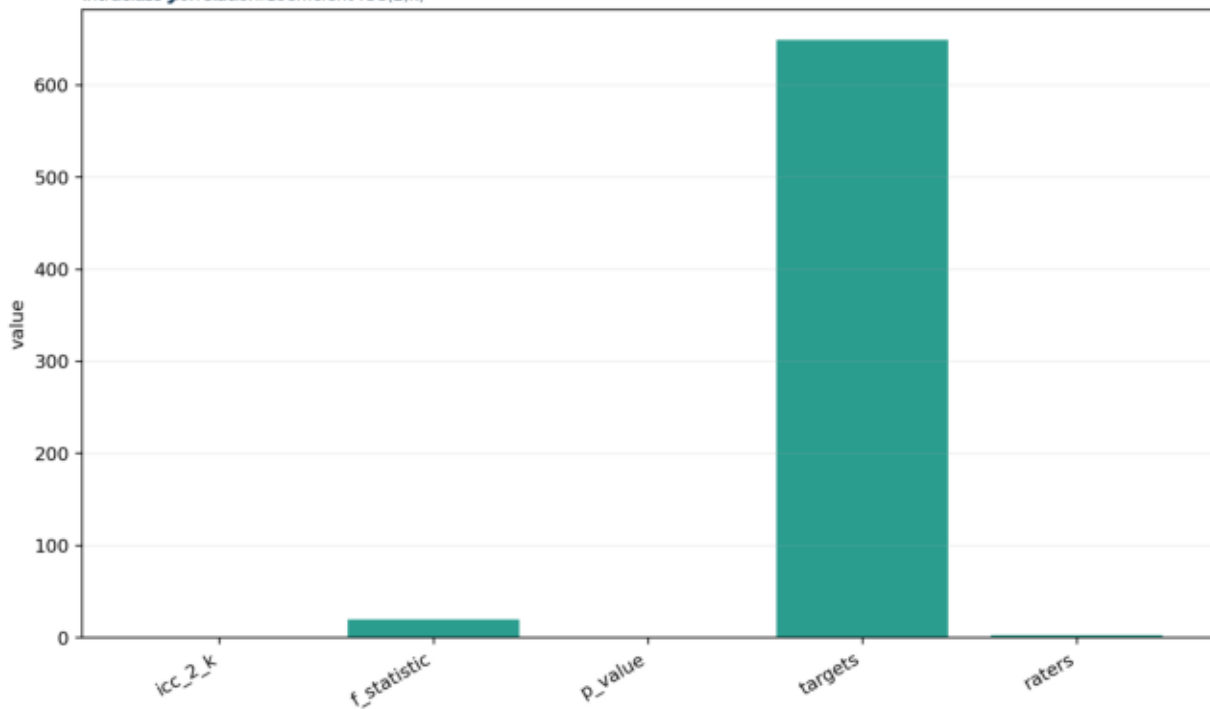
Interpretation:

ICC(2,k) is reliability of the three-rater average, not one rater.

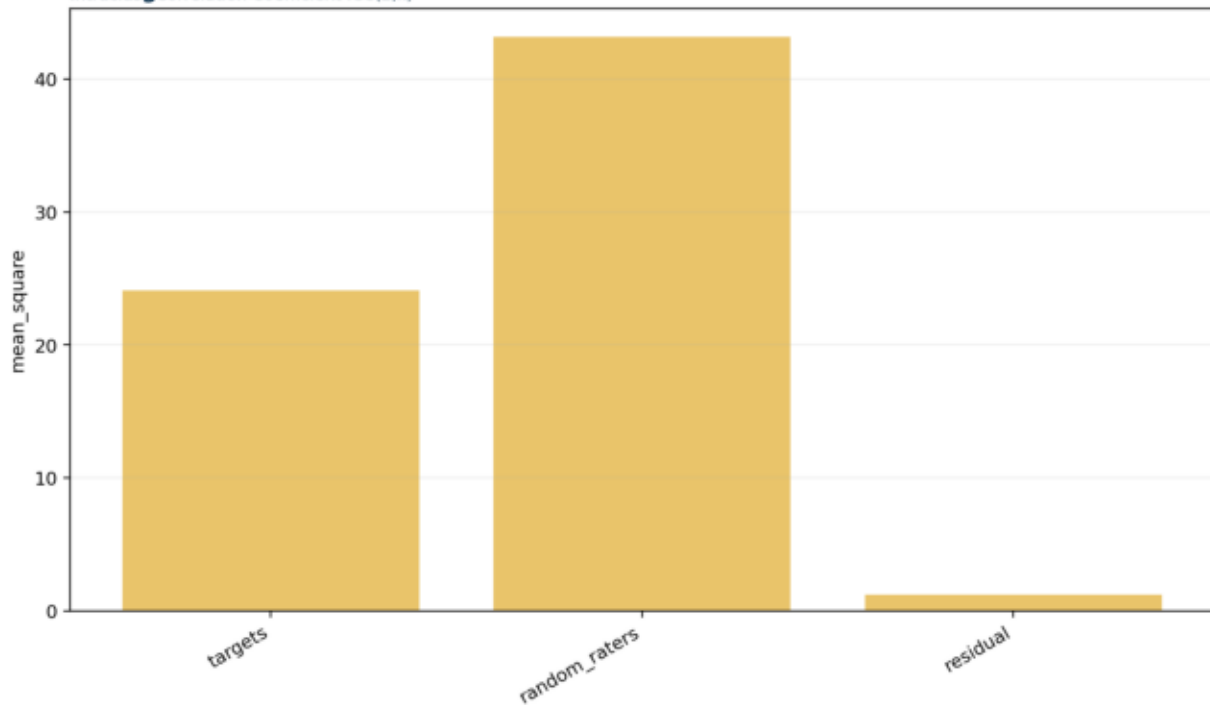
Average-measure absolute-agreement ICC=0.948061.

Primary Metrics

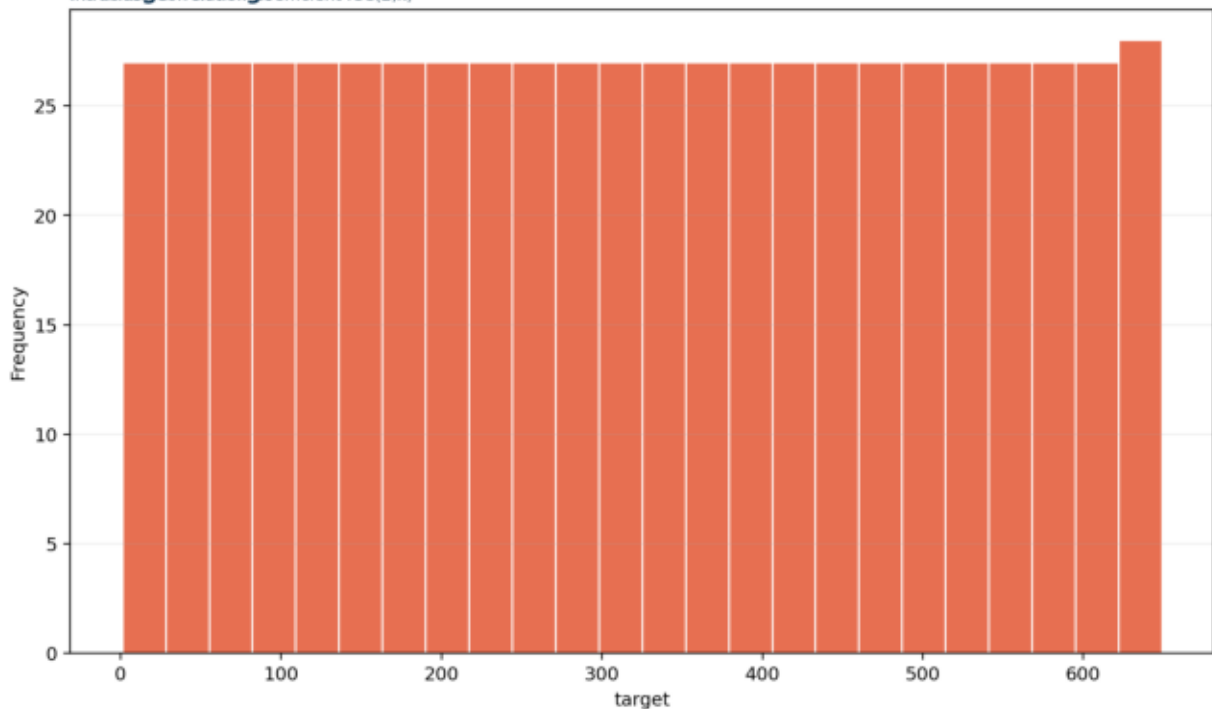
Intraclass Correlation Coefficient ICC(2,k)



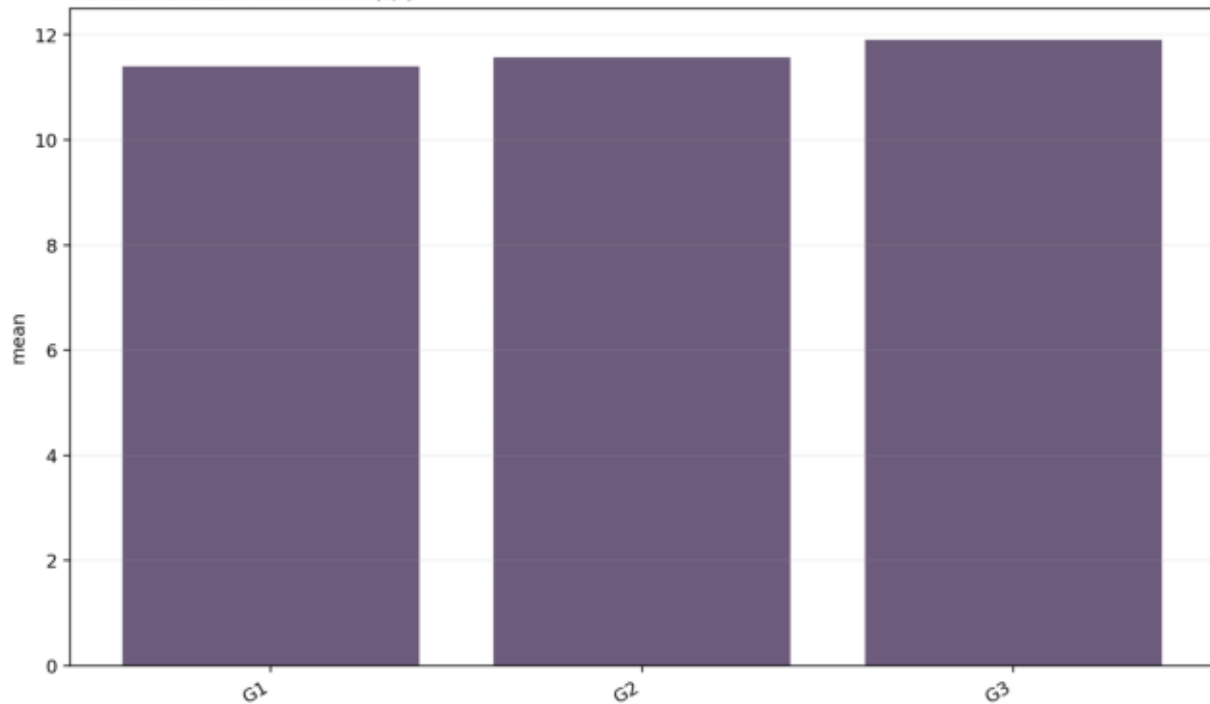
Average Measure Anova



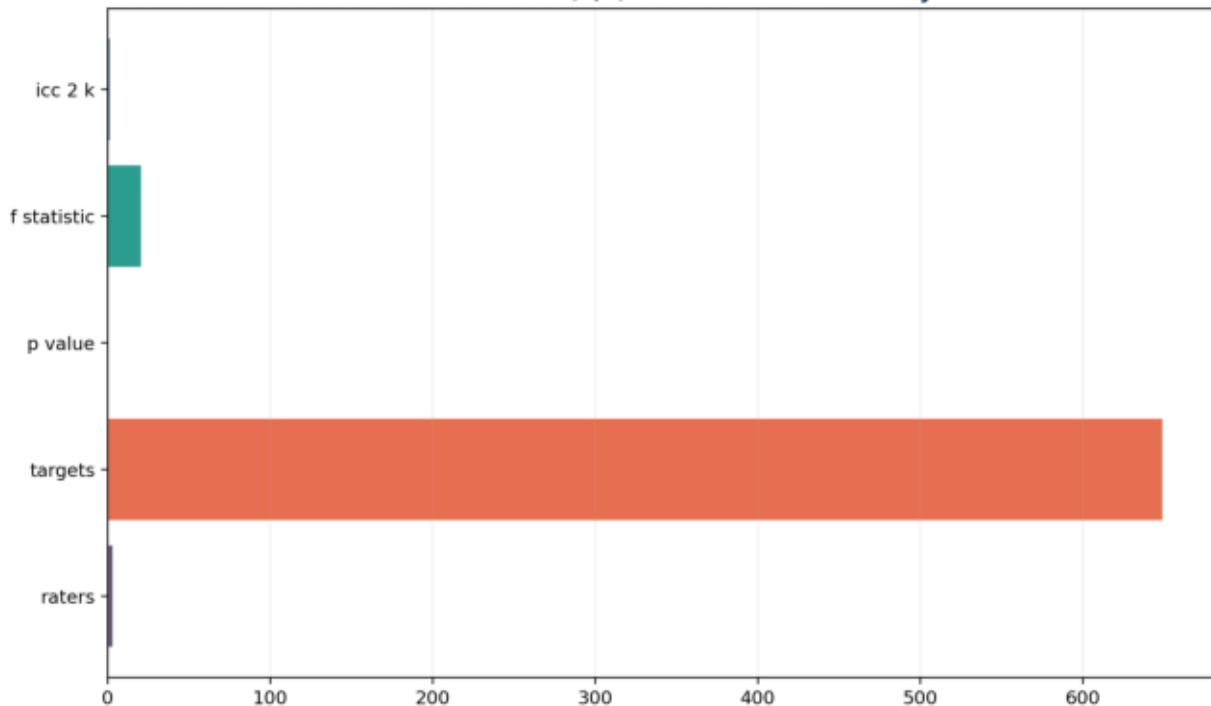
Average Ratings efficient ICC(2,k)



Rater Means on Coefficient ICC(2,k)



Intraclass Correlation Coefficient ICC(2,k) verified result summary



Primary Metrics

metric	value
icc_2_k	0.948061
f_statistic	20.246
p_value	0
targets	649
raters	3

Average Measure Anova

source	mean_square
targets	24.0838
random_raters	43.1654
residual	1.18956

Average Ratings

target	average_rating
1	7.33333
2	10.3333
3	12.3333
4	14
5	12.3333
6	12.3333
7	12.6667
8	12
9	16
10	12.3333
11	14
12	11.6667
13	12.3333
14	12.3333
15	14.3333
16	17
17	13.3333
18	13.6667
19	7.66667
20	12
21	13
22	11.6667
23	13
24	10
25	10.3333
26	11

Rater Means

rater	mean
G1	11.3991
G2	11.5701
G3	11.906