

Inter-Rater Reliability ICC(2,1)

Standalone verified Python analysis

Test: Inter-Rater Reliability ICC(2,1)

Design: Two-way random-effects absolute agreement for one randomly selected grader among G1, G2, and G3.

Variables: G1, G2, G3

Null hypothesis: The single-rater absolute-agreement ICC is zero.

Formula: $ICC(2,1) = (MSR - MSE) / [MSR + (k-1)MSE + k(MSC - MSE)/n]$.

Source rows: 649

Calculated metrics:

icc_2_1: 0.858846839218952

f_statistic: 20.245973021325074

p_value: 0.0

targets: 649

raters: 3

Independent reference values:

icc_2_1: 0.858846839218952

f_statistic: 20.245973021325074

p_value: 0.0

targets: 649

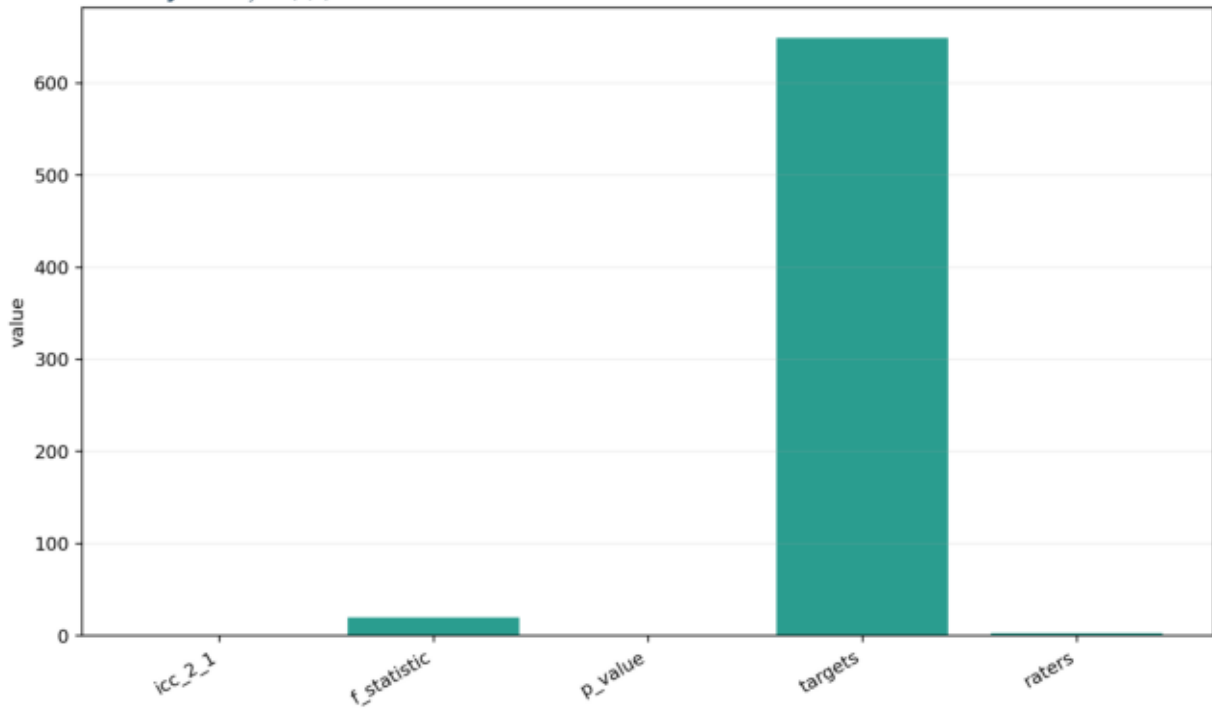
raters: 3

Interpretation:

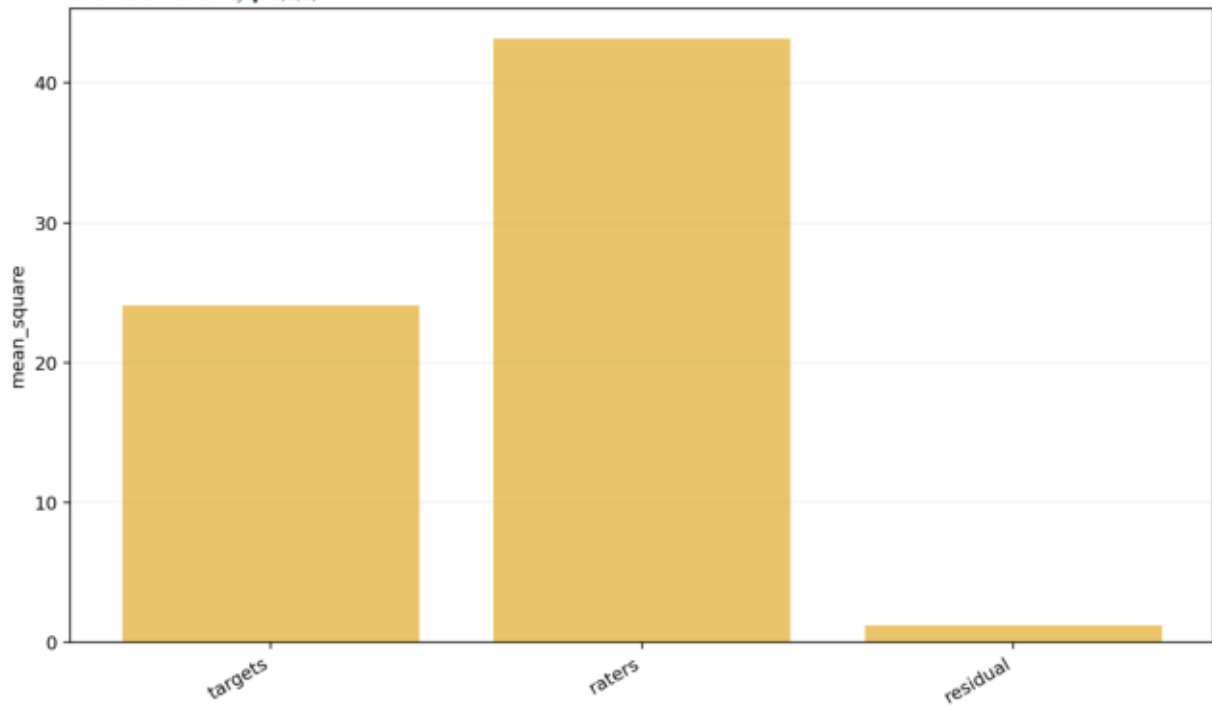
ICC(2,1) treats raters as random and systematic offsets as disagreement.

Single-rater absolute-agreement ICC=0.858847.

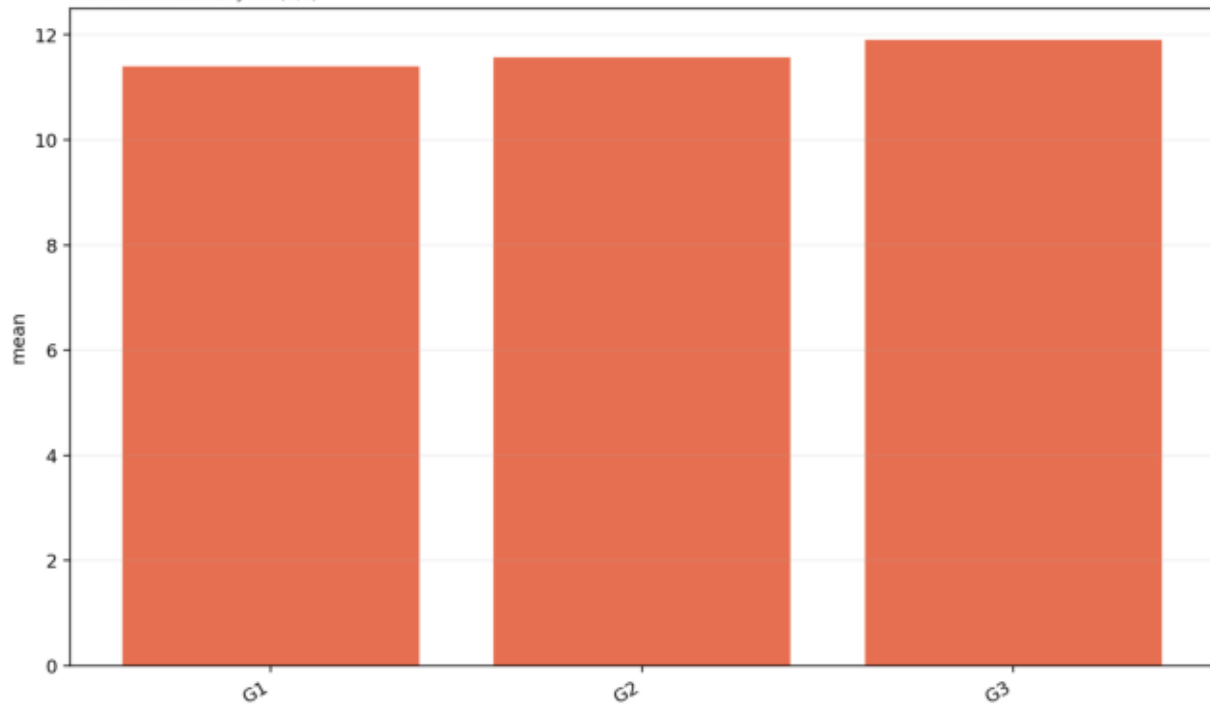
Primary Metrics



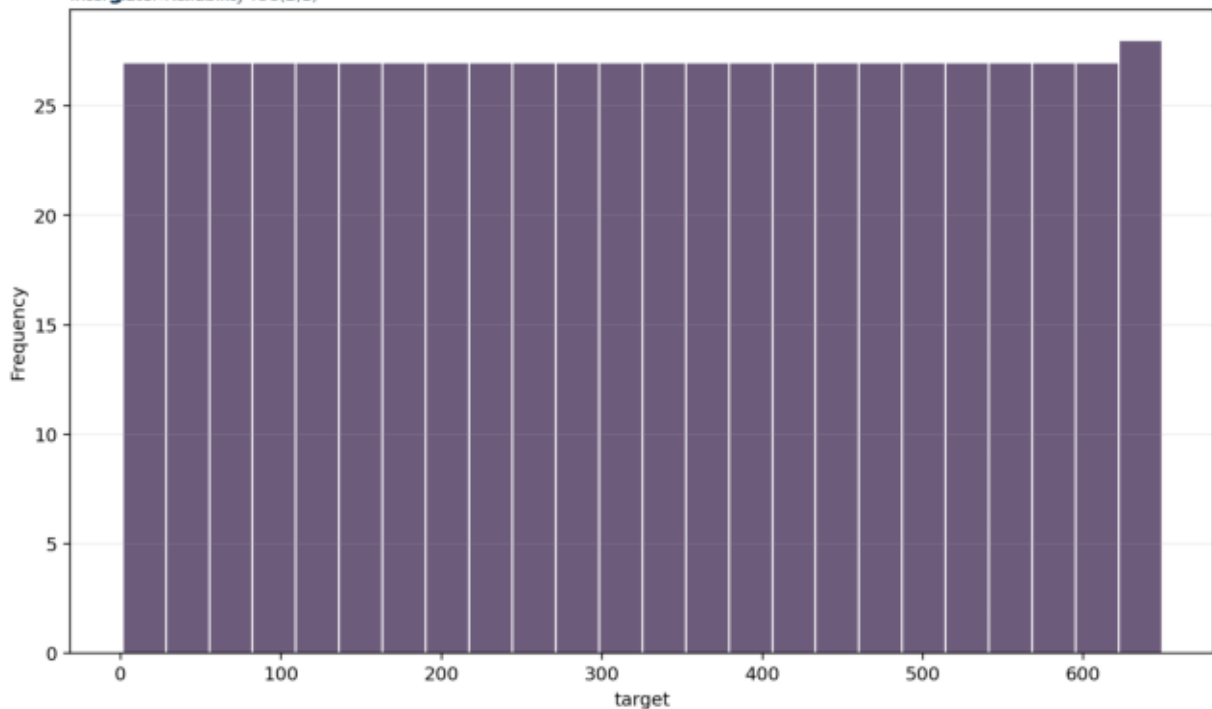
Icc Anova Components



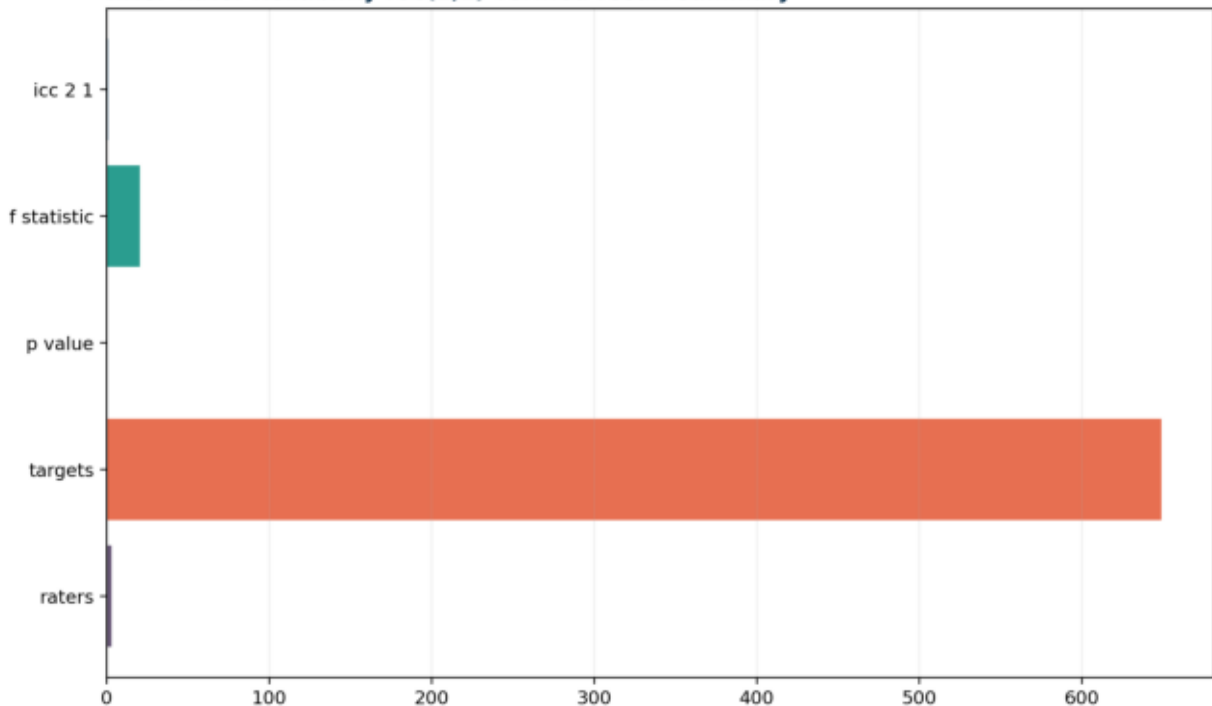
Rater Means
Intraclass Correlation ICC(2,1)



Target Means ICC(2,1)



Inter-Rater Reliability ICC(2,1) verified result summary



Primary Metrics

metric	value
icc_2_1	0.858847
f_statistic	20.246
p_value	0
targets	649
raters	3

Icc Anova Components

source	mean_square
targets	24.0838
raters	43.1654
residual	1.18956

Rater Means

rater	mean
G1	11.3991
G2	11.5701
G3	11.906

Target Means

target	mean_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