Two-Sample Kolmogorov-Smirnov Test

Standalone verified Python analysis

Test: Two-Sample Kolmogorov-Smirnov Test

Design: Maximum ECDF separation in absence counts between female and male students.

Variables: absences, sex

Null hypothesis: The two complete absence distributions are identical.

Formula: D = sup_x |F_F(x)-F_M(x)|; tied-count significance uses fixed-size label permutations.

Source rows: 649

Calculated metrics:

d_statistic: 0.050187479141718505

p_value: 0.5056494350564944

permutation_null_mean: 0.05217586426902732

permutations: 10000

seed: 20260716

asymptotic_p_scipy: 0.8049827570841168

statistic_location: 3.0

statistic_sign: 1

Independent reference values:

d_statistic: 0.050187479141718505

p_value: 0.5056494350564944

permutation_null_mean: 0.05217586426902732

permutations: 10000

seed: 20260716

asymptotic_p_scipy: 0.8049827570841168

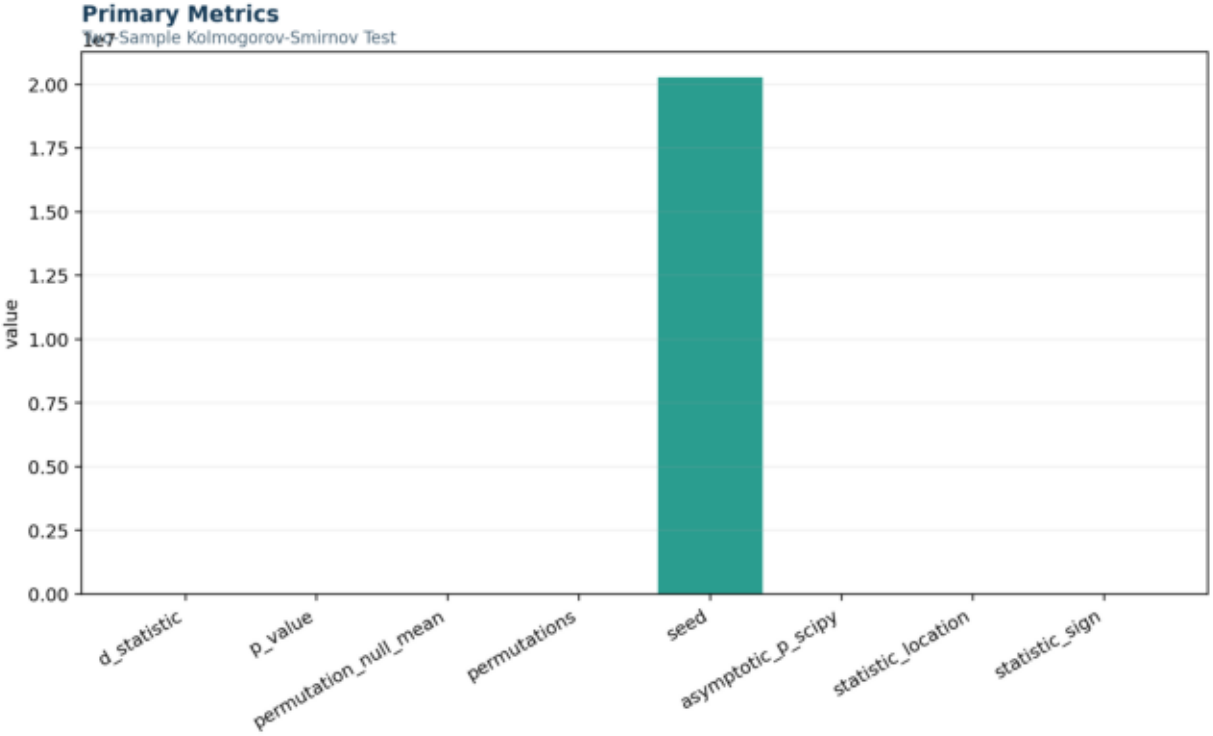
statistic_location: 3.0

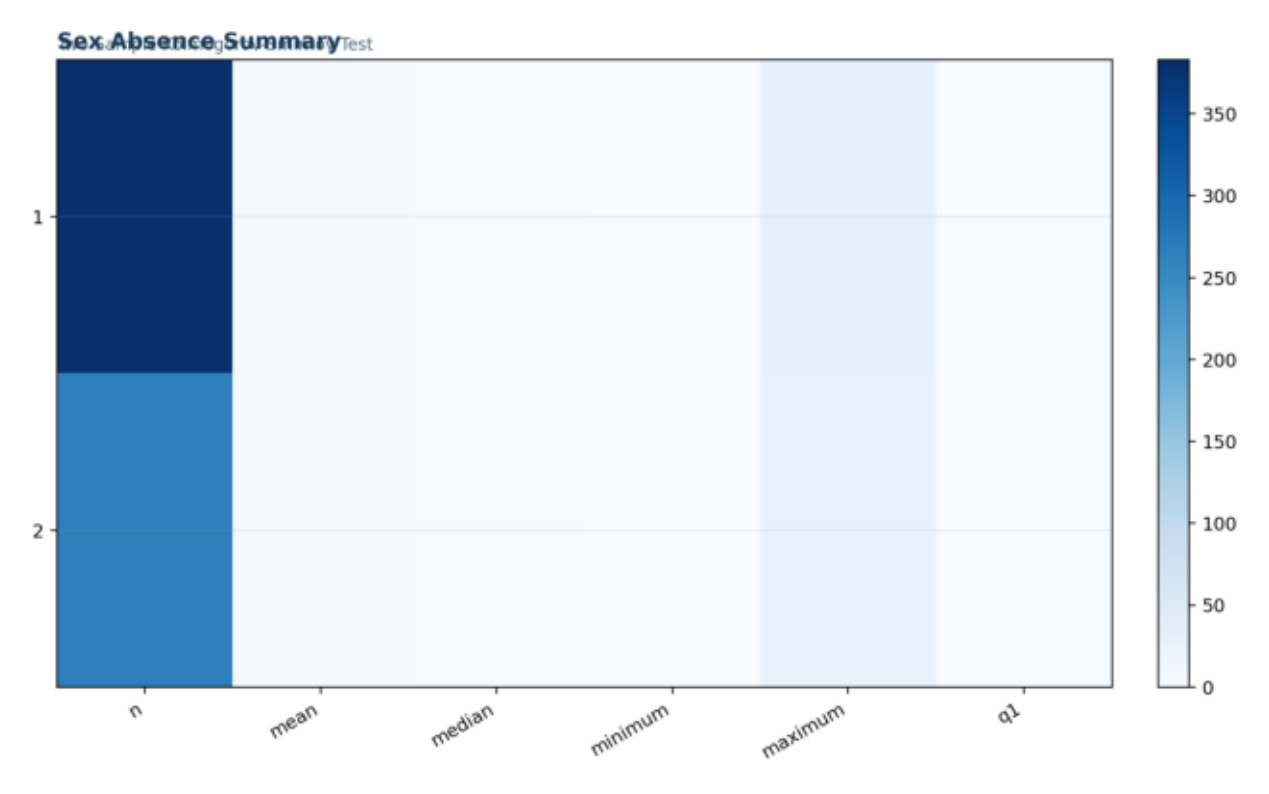
statistic_sign: 1

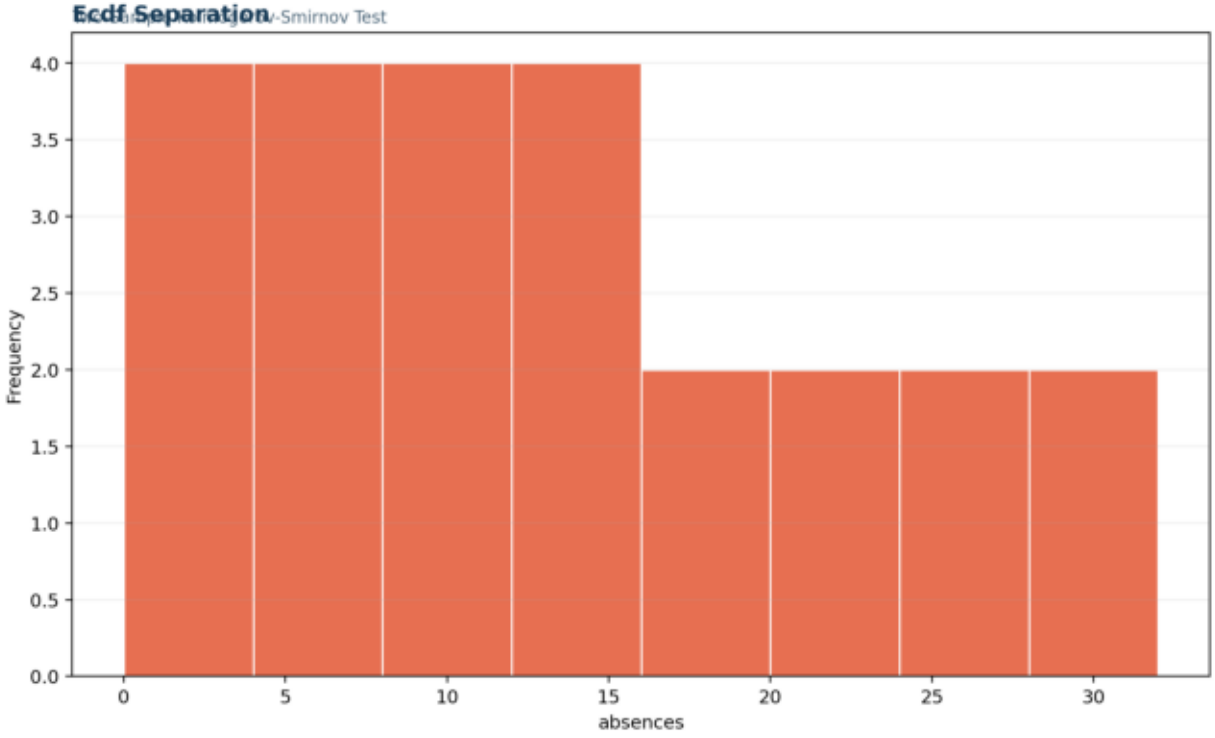
Interpretation:

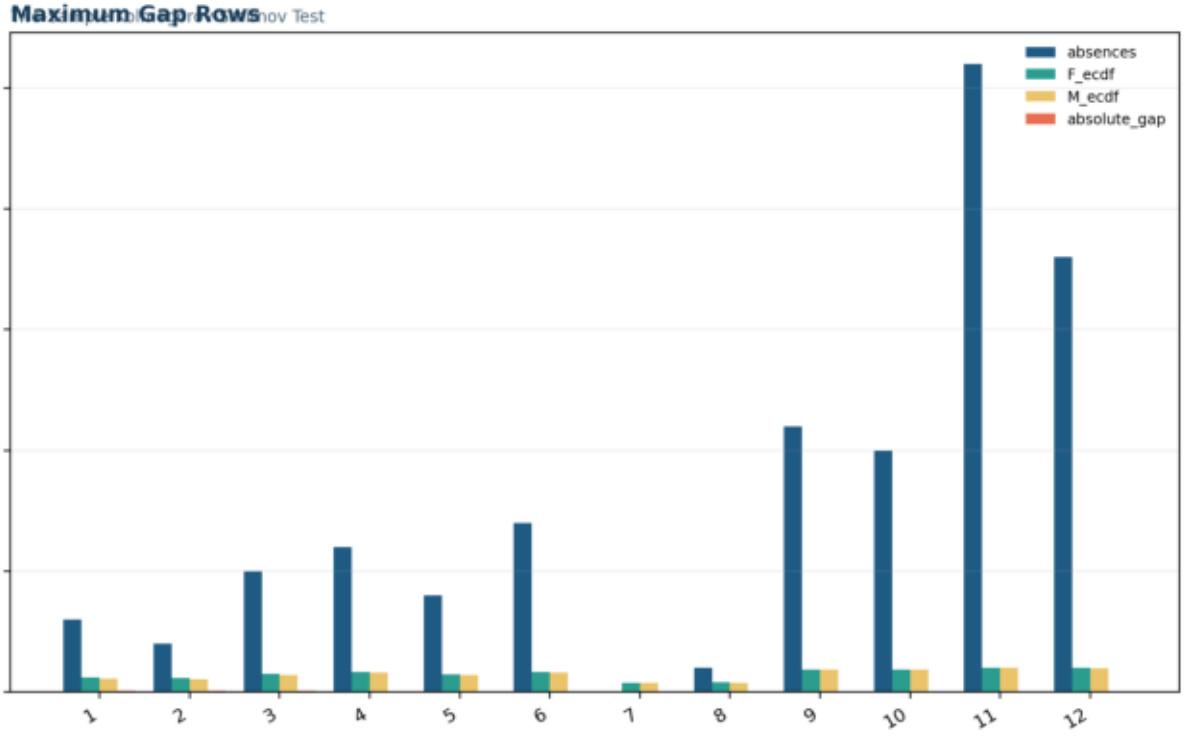
The statistic is the maximum vertical gap between the two empirical CDFs.

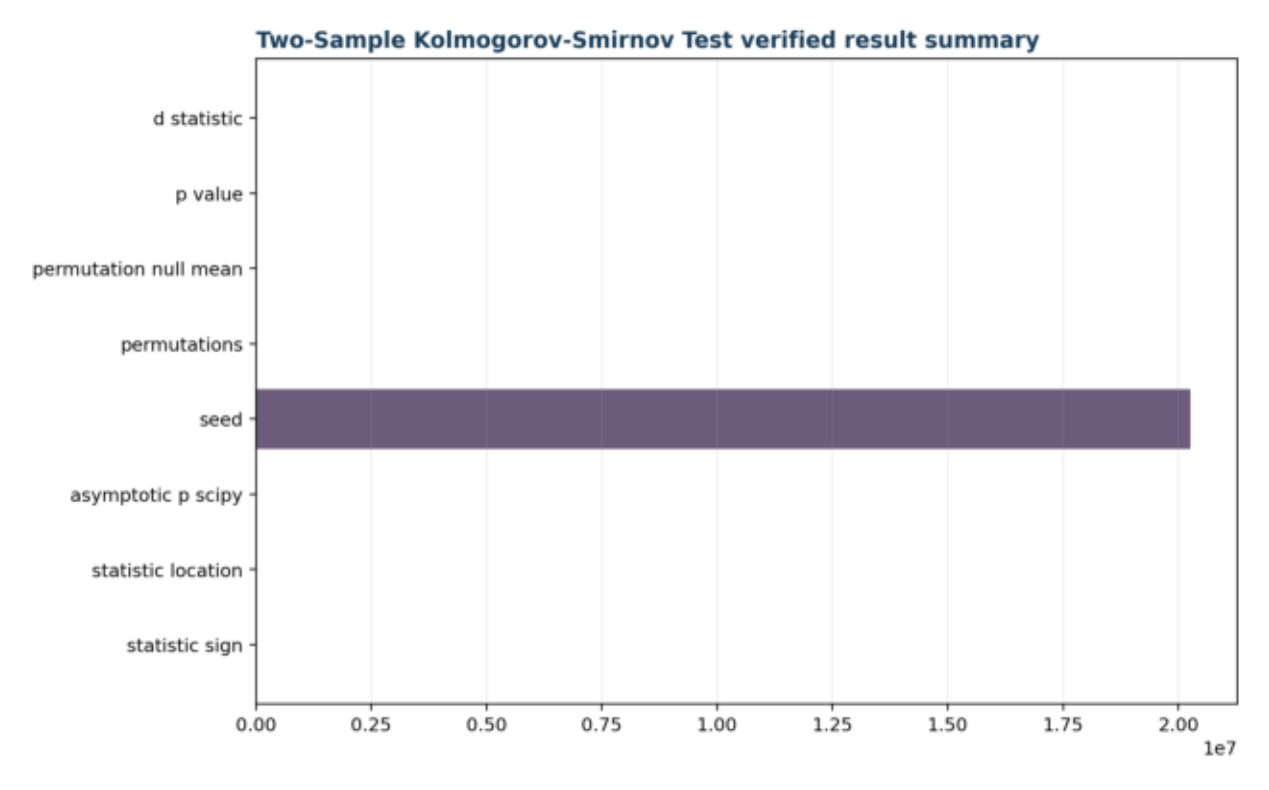
Permutation inference is primary because absence counts contain extensive ties.


Primary Metrics
2-Sample Kolmogorov-Smirnov Test
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
value
d_statistic
p_value
permutation_null_mean
permutations
seed
asymptotic_p_scipy
statistic_location
statistic_sign


Sex Absence Summary
Test
1
2
n
mean
median
minimum
maximum
q1
350
300
250
200
150
100
50
0


Ecdf Separation
Kolmogorov-Smirnov Test
Frequency
absences
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
5
10
15
20
25
30


Maximum Gap Rows
absences
F_ecdf
M_ecdf
absolute_gap
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12


Two-Sample Kolmogorov-Smirnov Test verified result summary
d statistic
p value
permutation null mean
permutations
seed
asymptotic p scipy
statistic location
statistic sign
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
1e7

Primary Metrics

metric	value
d_statistic	0.0501875
p_value	0.505649
permutation_null_mean	0.0521759
permutations	10000
seed	2.02607e+07
asymptotic_p_scipy	0.804983
statistic_location	3
statistic_sign	1

Sex Absence Summary

sex	n	mean	median	minimum	maximum	q1	q3
F	383	3.57702	2	0	32	0	5
M	266	3.7782	2	0	26	0	6

Ecdf Separation

absences	F_ecdf	M_ecdf	absolute_gap
0	0.370757	0.383459	0.0127015
1	0.399478	0.387218	0.0122598
2	0.582245	0.537594	0.0446514
3	0.5953	0.545113	0.0501875
4	0.72846	0.703008	0.025452
5	0.754569	0.710526	0.0440429
6	0.825065	0.793233	0.0318322
7	0.825065	0.804511	0.020554
8	0.882507	0.879699	0.00280728
9	0.890339	0.894737	0.00439742
10	0.926893	0.921053	0.00584032
11	0.929504	0.93609	0.00658631
12	0.950392	0.951128	0.000736175
13	0.950392	0.954887	0.00449557
14	0.966057	0.962406	0.00365143
15	0.968668	0.966165	0.00250299
16	0.984334	0.981203	0.0031312
18	0.989556	0.984962	0.00459373
21	0.992167	0.988722	0.0034453
22	0.992167	0.996241	0.0040735
24	0.994778	0.996241	0.00146253
26	0.994778	1	0.00522193
30	0.997389	1	0.00261097
32	1	1	0

Maximum Gap Rows

absences	F_ecdf	M_ecdf	absolute_gap
3	0.5953	0.545113	0.0501875
2	0.582245	0.537594	0.0446514
5	0.754569	0.710526	0.0440429
6	0.825065	0.793233	0.0318322
4	0.72846	0.703008	0.025452
7	0.825065	0.804511	0.020554
0	0.370757	0.383459	0.0127015
1	0.399478	0.387218	0.0122598
11	0.929504	0.93609	0.00658631
10	0.926893	0.921053	0.00584032
26	0.994778	1	0.00522193
18	0.989556	0.984962	0.00459373