

Binary Kappa Statistic with Prevalence and Bias Diagnostics

Verified Python analysis transcript

Test: Binary Kappa Statistic with Prevalence and Bias Diagnostics

Design: Binary agreement between G2 and G3 pass/fail classifications.

Variables: G2 >= 10, G3 >= 10

Null hypothesis: The two binary classifications agree only at the level expected by their margins.

Primary formula: $\text{kappa} = (P_o - P_e) / (1 - P_e)$; $\text{PABAK} = 2P_o - 1$

Valid source rows: 649

Verified results:

binary_kappa: 0.6655282489135034

p_value: 8.34189307285141e-68

observed_agreement: 0.8967642526964561

pabak: 0.7935285053929122

prevalence_index: 0.6224961479198767

bias_index: 0.06933744221879815

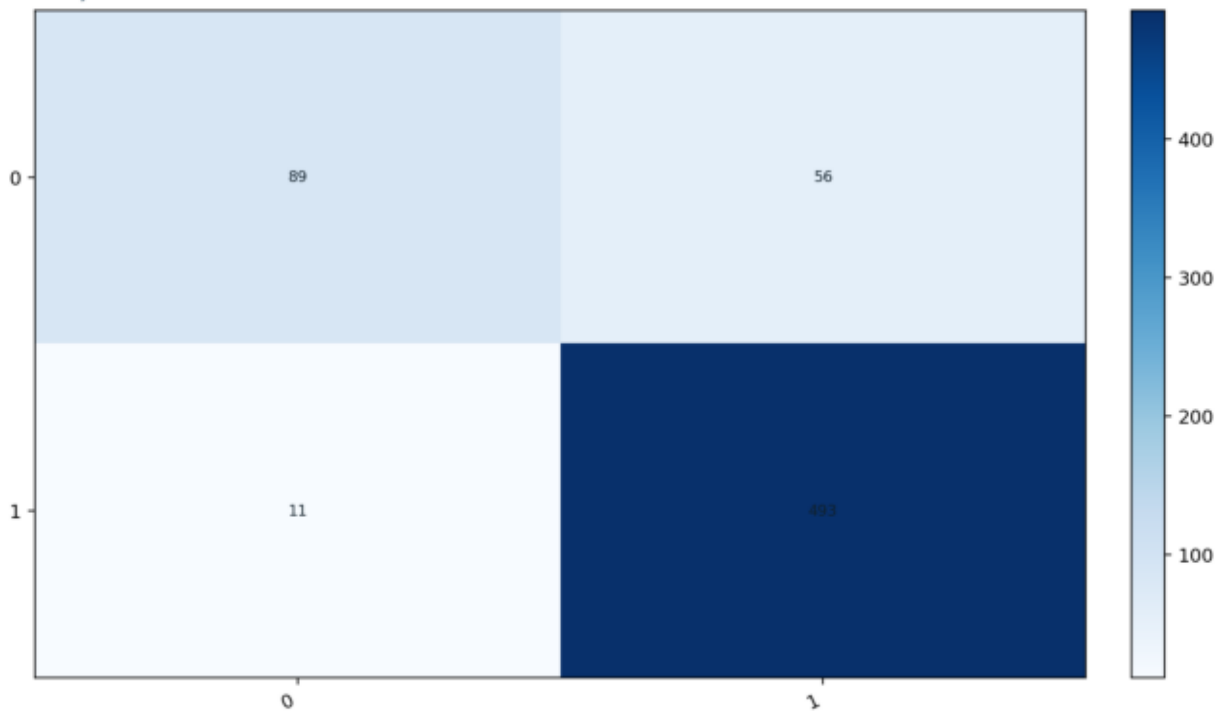
Interpretation:

Binary kappa = 0.665528, p = 8.34189e-68; PABAK = 0.793529.

The prevalence and bias indices are reported because high marginal imbalance can depress kappa despite high observed agreement.

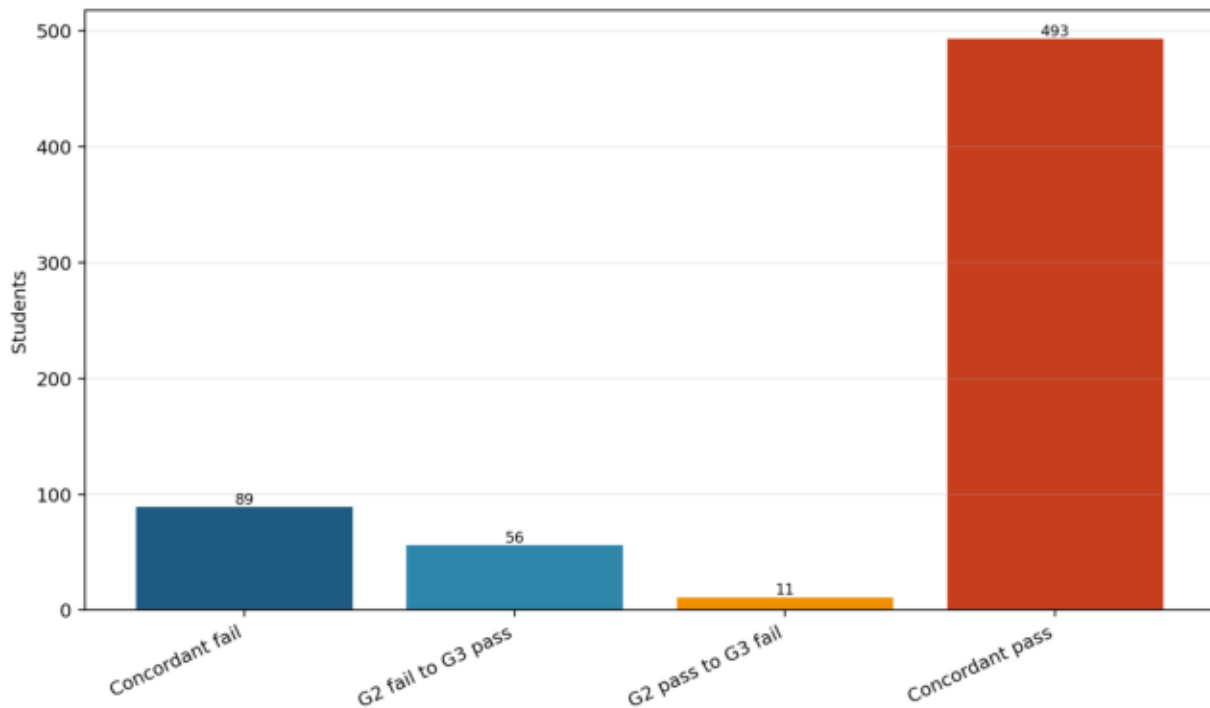
G2 and G3 pass/fail agreement

Binary classification table



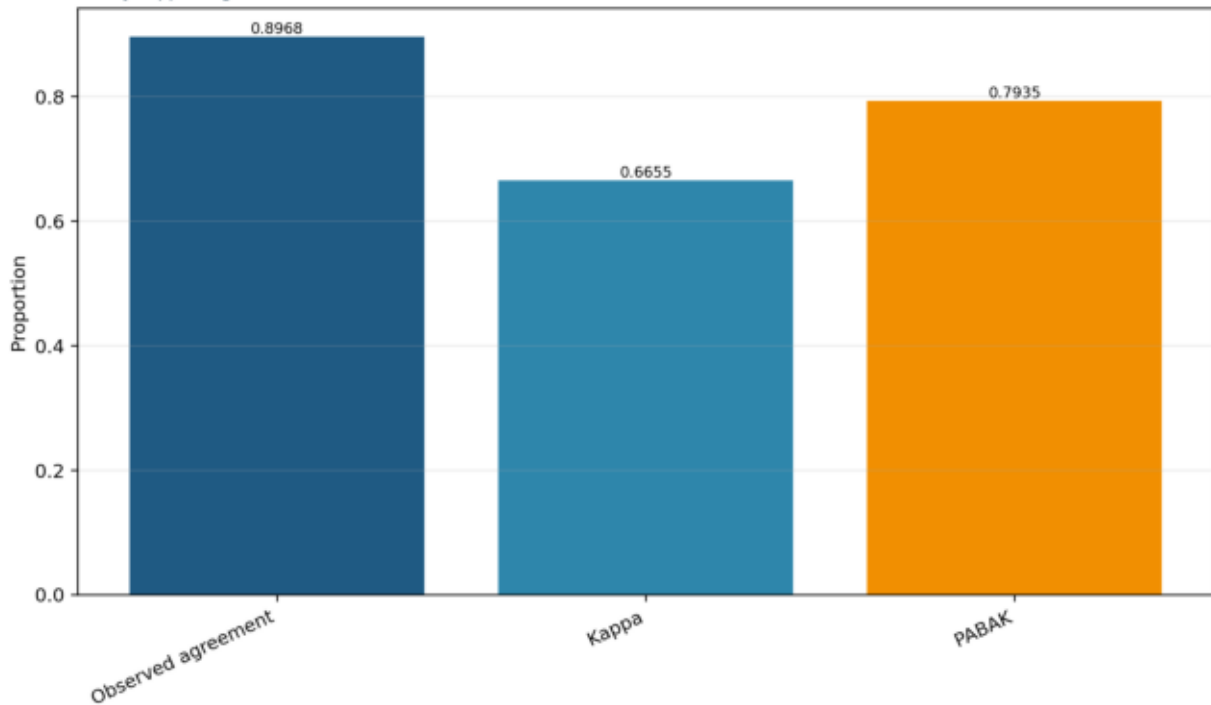
Binary agreement cell counts

Concordance and directional discordance



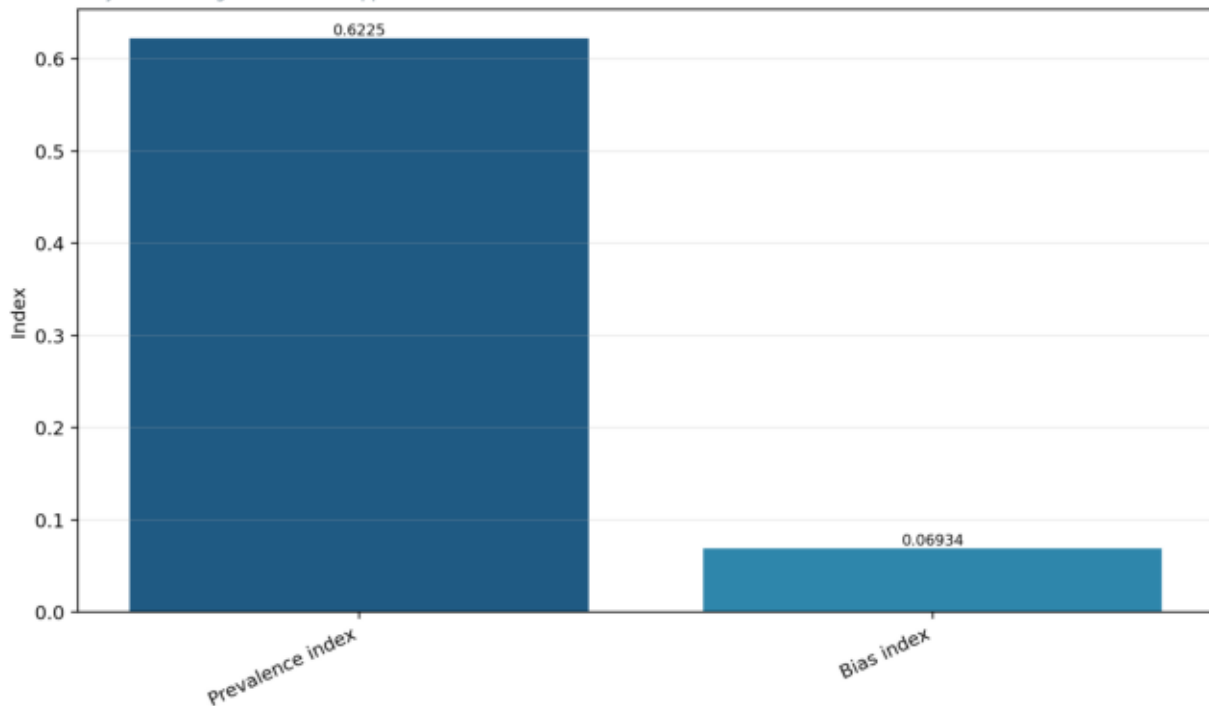
Chance and prevalence-adjusted agreement

Binary kappa diagnostics



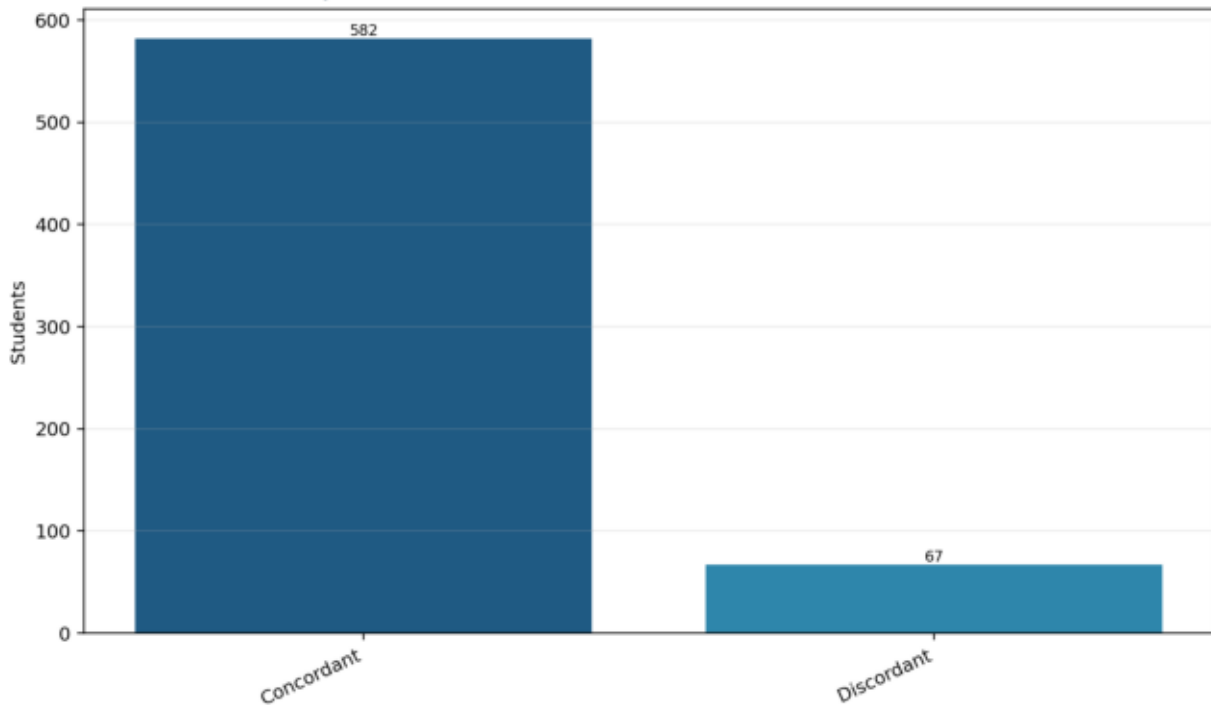
Prevalence and bias diagnostics

Why observed agreement and kappa differ



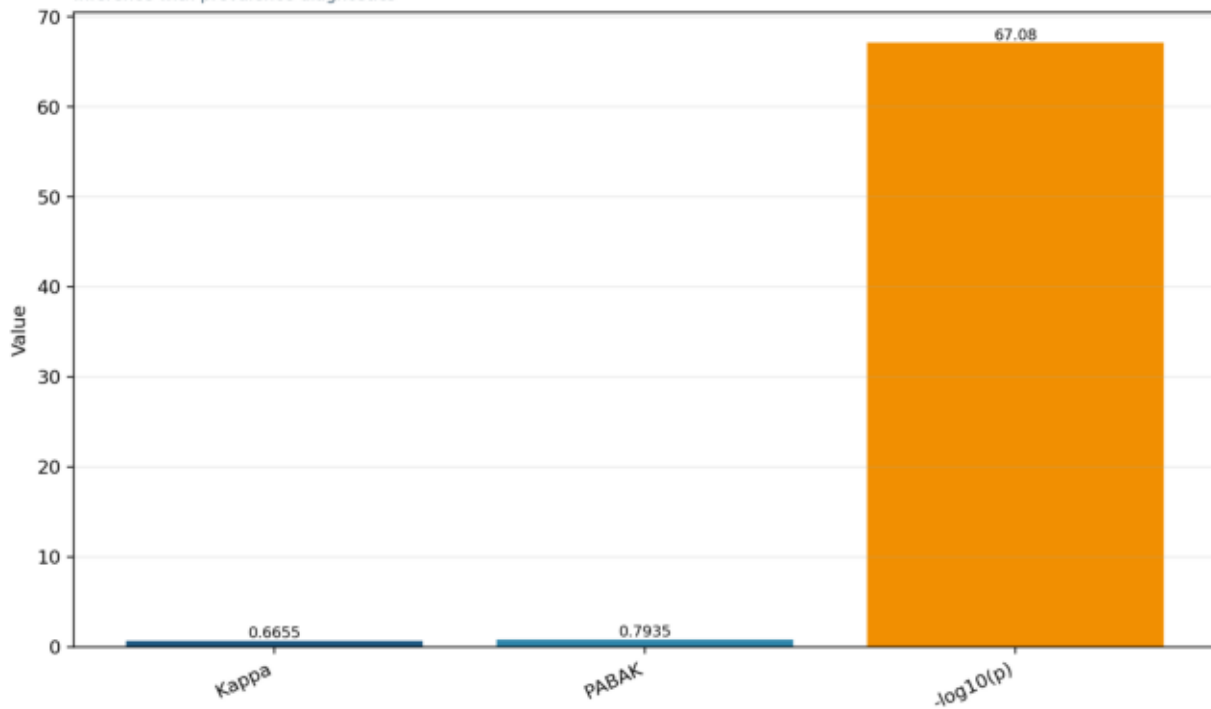
Concordance balance

Overall classification stability



Binary kappa result

Inference with prevalence diagnostics



Binary Agreement Table

0	1
89	56
11	493

Diagnostic Indices

value
0.896764
0.665528
0.793529
0.622496
0.0693374