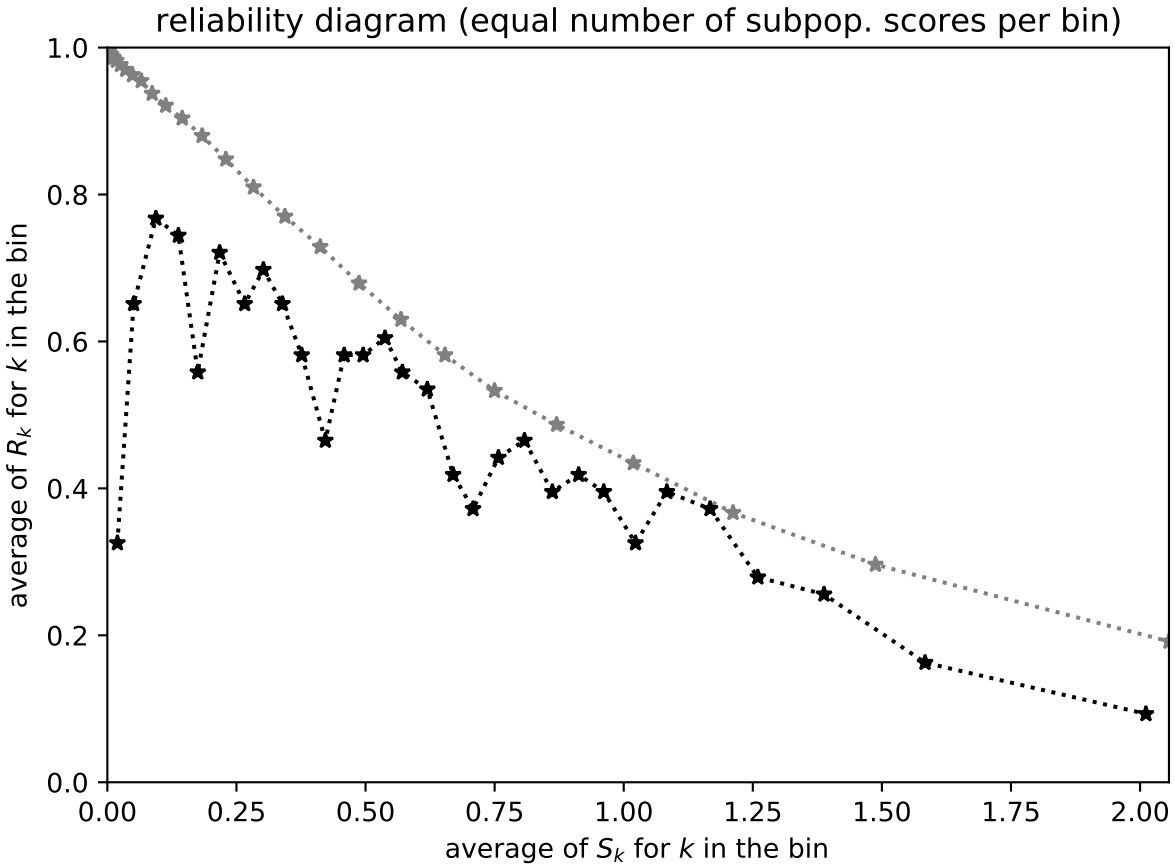
reliability diagram (equal number of subpop. scores per bin)
average of R_k for k in the bin
average of S_k for k in the bin
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00