

Statistical Benchmark Report

Generated: 2025-11-19 00:41

This report provides quantitative evaluation results for NDCG@k and faithfulness, correlation diagnostics, outlier detection, and statistical visualizations.

1. Summary Statistics

files	10.0000
ndcg@k mean	0.9644
ndcg@k ci95 lo	0.9426
ndcg@k ci95 hi	0.9849
faith mean	0.5042
faith ci95 lo	0.4030
faith ci95 hi	0.6068
ndcg@k median	0.9710
faith median	0.4771
ndcg@k std	0.0343
faith std	0.1683
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.5000
faith mid thr	0.2500

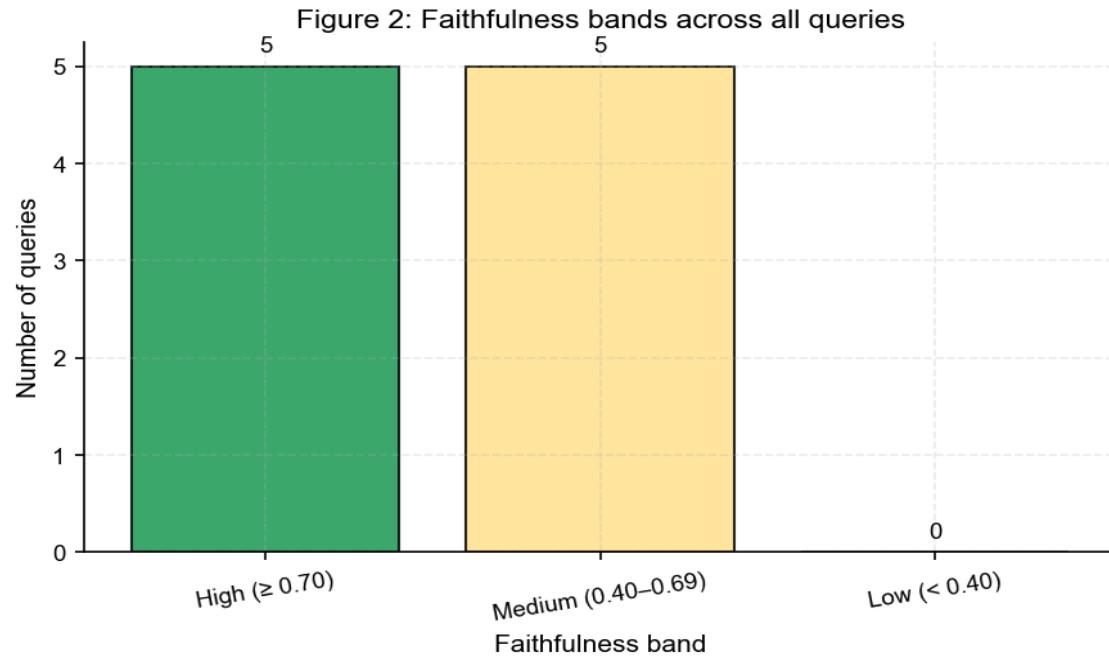
2. Correlation and Outlier Analysis

Metric Pair	Correlation	p-value
Pearson r	-0.205	5.702e-01
Spearman ρ	0.129	7.228e-01

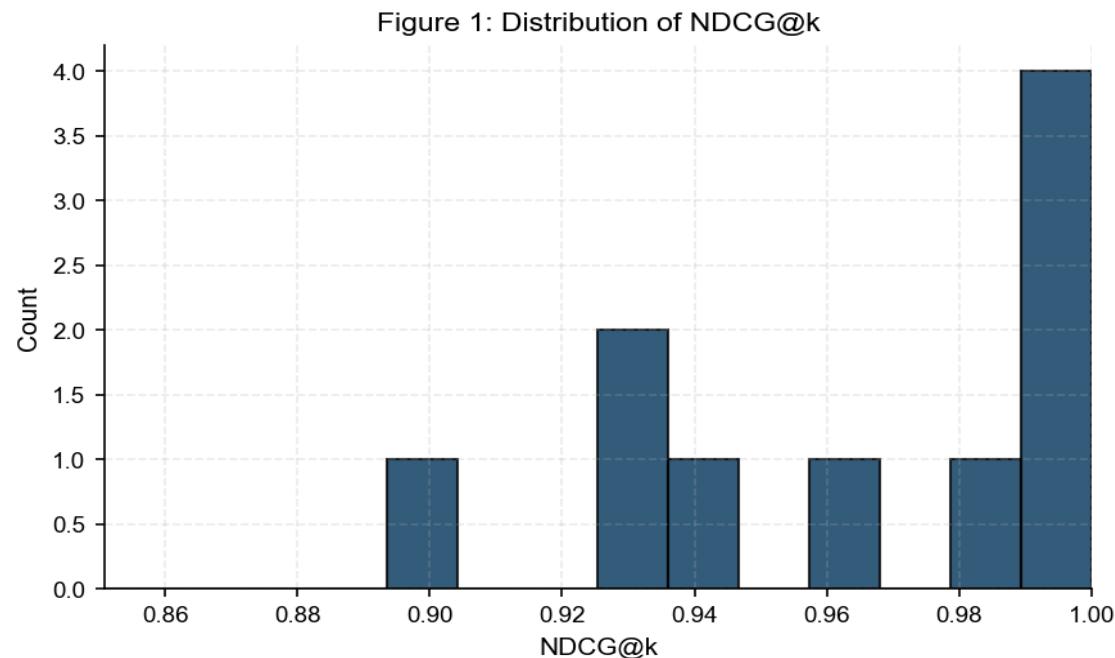
No outliers detected.

3. Visual Analytics

faithfulness bands global.png

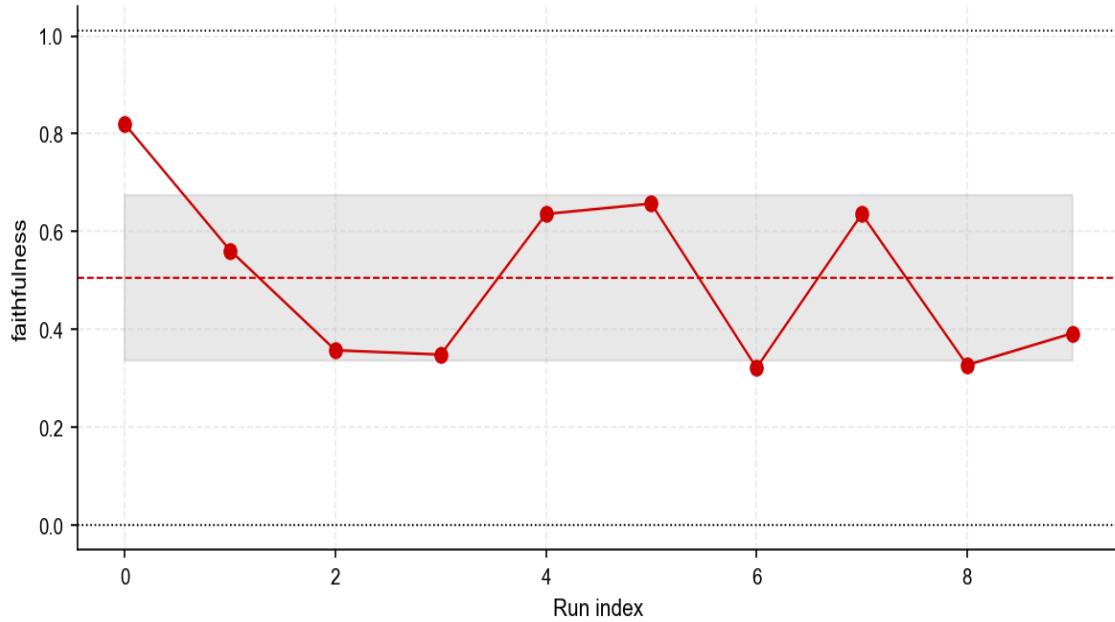


hist ndcg.png



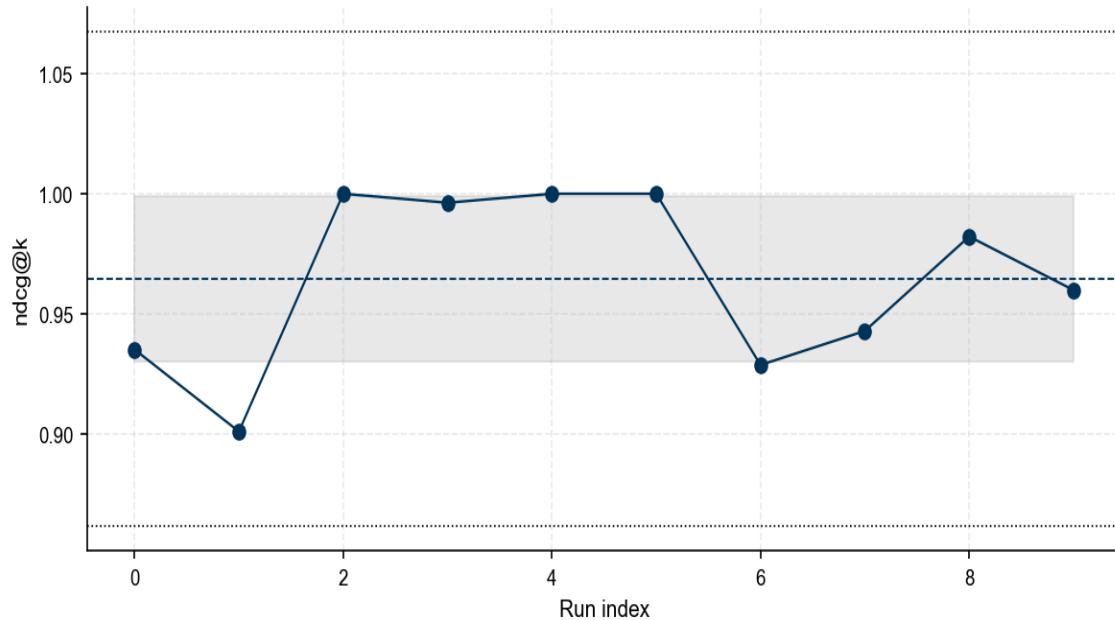
run order faithfulness.png

Figure 5: Run-order chart for faithfulness



run order ndcg@k.png

Figure 4: Run-order chart for ndcg@k



scatter ndcg vs faithfulness.png

Figure 3: Scatter NDCG@k vs Faithfulness ($r=-0.205$, $p=0.129$)

