Statistical Benchmark Report

Generated: 2025-11-17 13:54

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

1. Summary Statistics

files	100.0000
ndcg@k mean	1.0000
ndcg@k ci95 lo	1.0000
ndcg@k ci95 hi	1.0000
faith mean	0.6307
faith ci95 lo	0.5935
faith ci95 hi	0.6657
ndcg@k median	1.0000
faith median	0.6752
ndcg@k std	0.0000
faith std	0.1796
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.5000
faith mid thr	0.2500

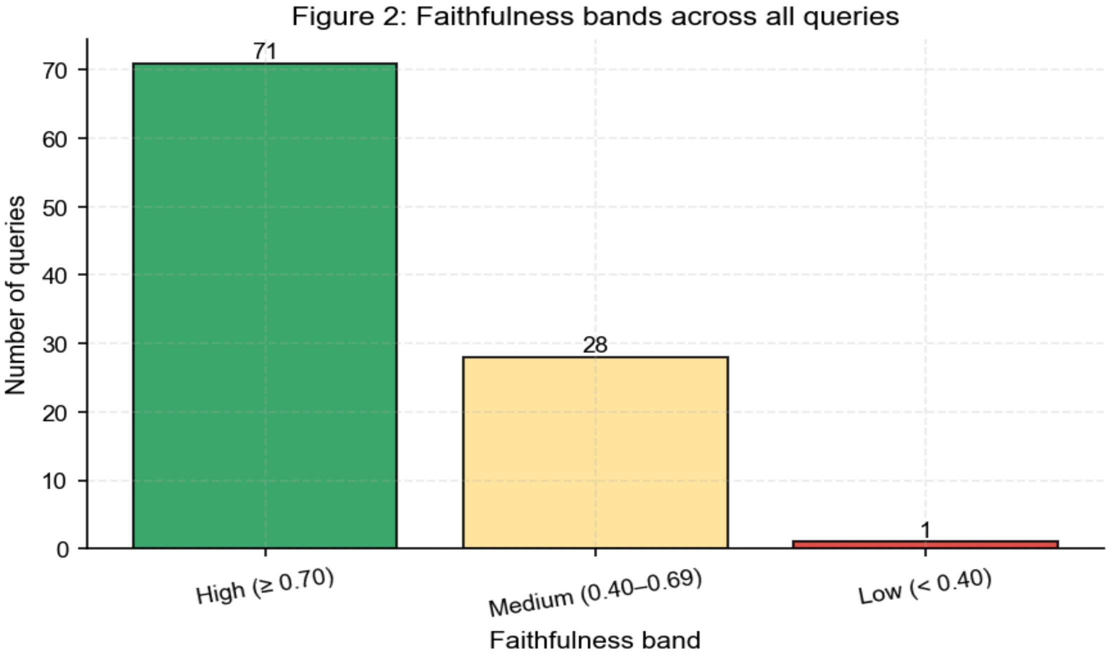
2. Correlation and Outlier Analysis

Metric Pair	Correlation	p-value
Pearson r	nan	nan
Spearman ρ	nan	nan

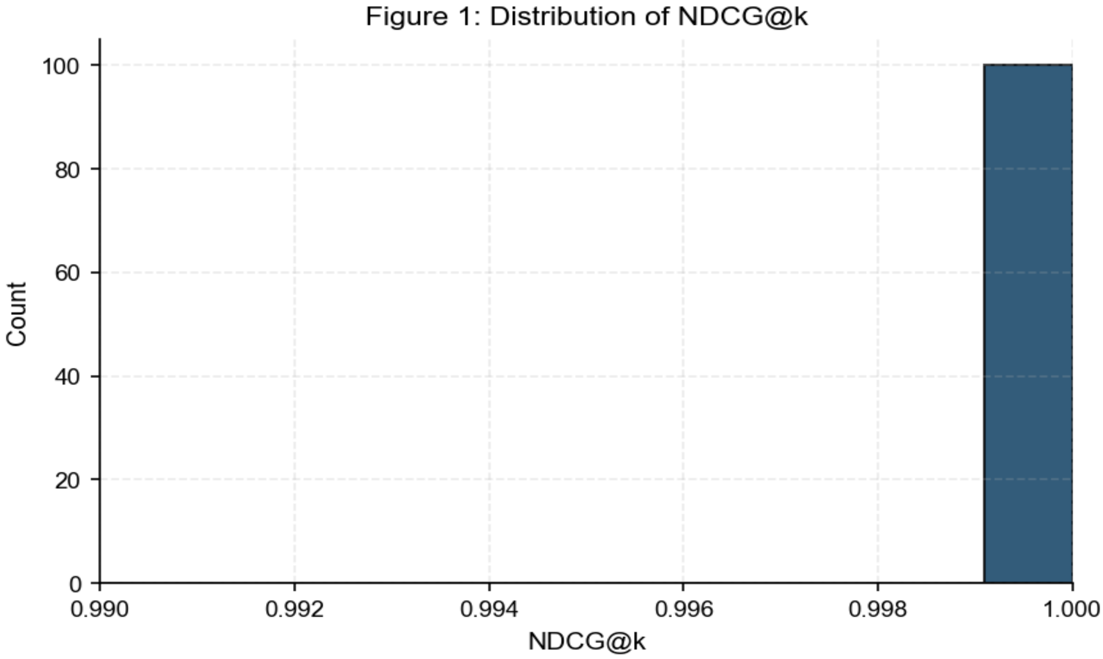
No outliers detected.

3. Visual Analytics

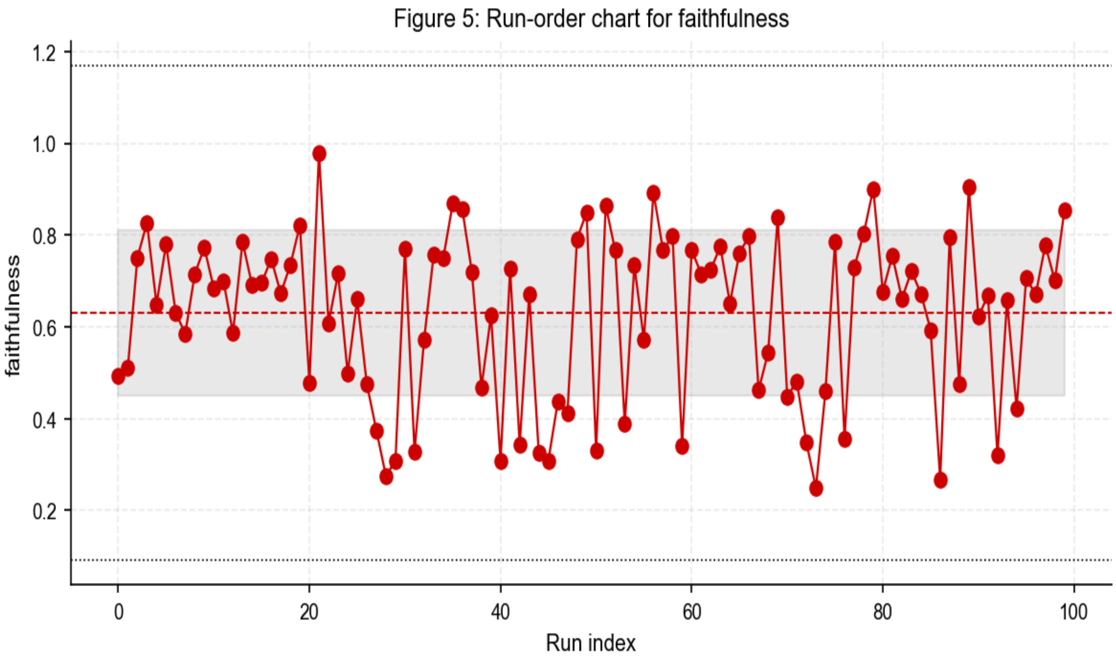
faithfulness bands global.png



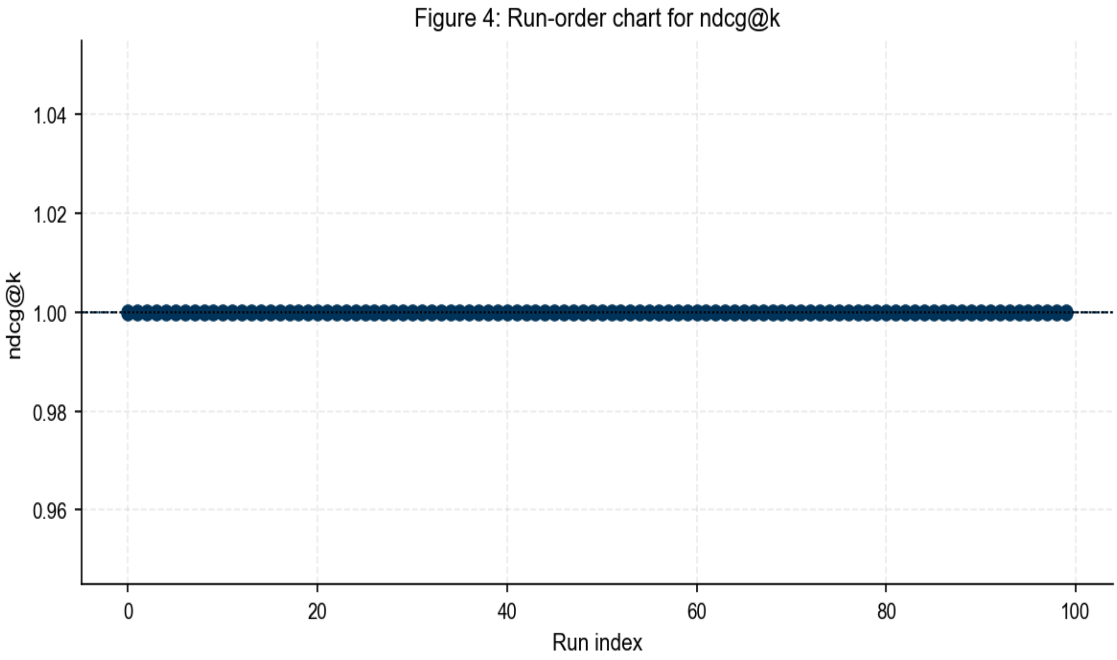
hist ndcg.png



run order faithfulness.png

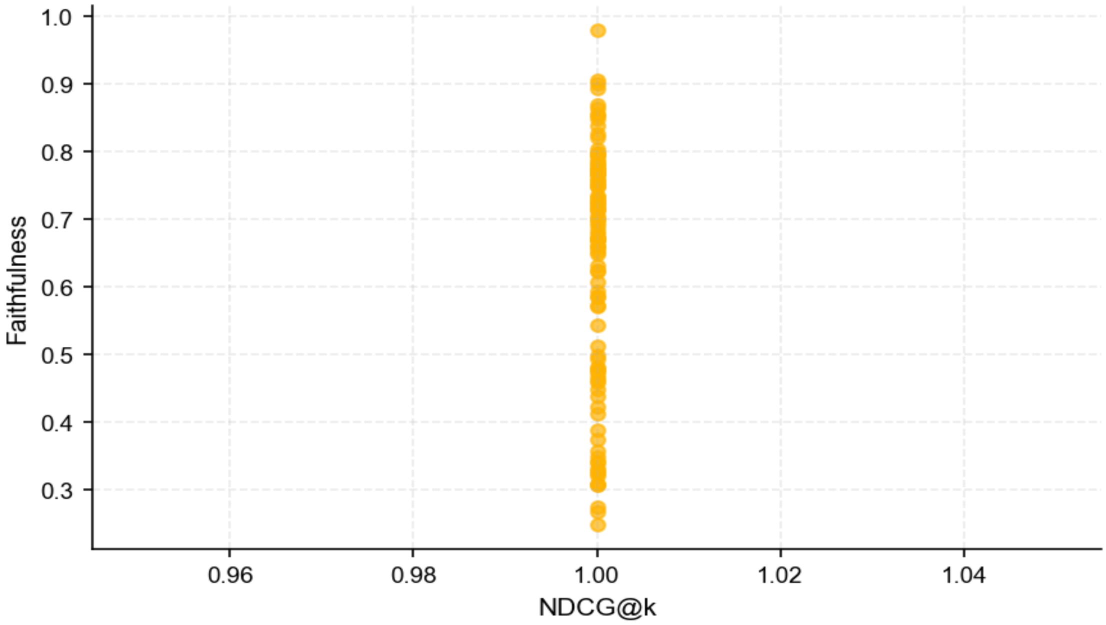


run order ndcg@k.png



scatter ndcg vs faithfulness.png

Figure 3: Scatter NDCG@k vs Faithfulness (r=nan, ρ=nan)


Faithfulness
NDCG@k