Statistical Benchmark Report

Generated: 2025-11-16 17:01

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	1.0000
ndcg@k ci95 lo	1.0000
ndcg@k ci95 hi	1.0000
faith mean	0.5310
faith ci95 lo	0.4388
faith ci95 hi	0.6150
ndcg@k median	1.0000
faith median	0.5786
ndcg@k std	0.0000
faith std	0.1524
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.7000
faith mid thr	0.4000
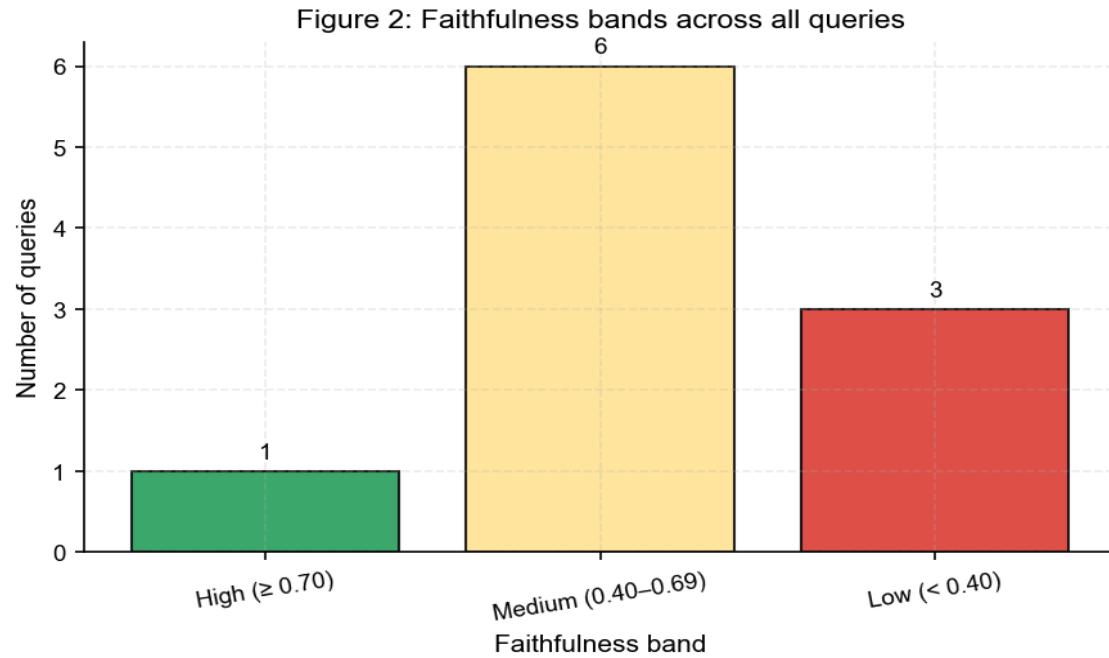
2. Correlation and Outlier Analysis

Metric Pair	Correlation	p-value
Pearson r	nan	nan
Spearman ρ	nan	nan
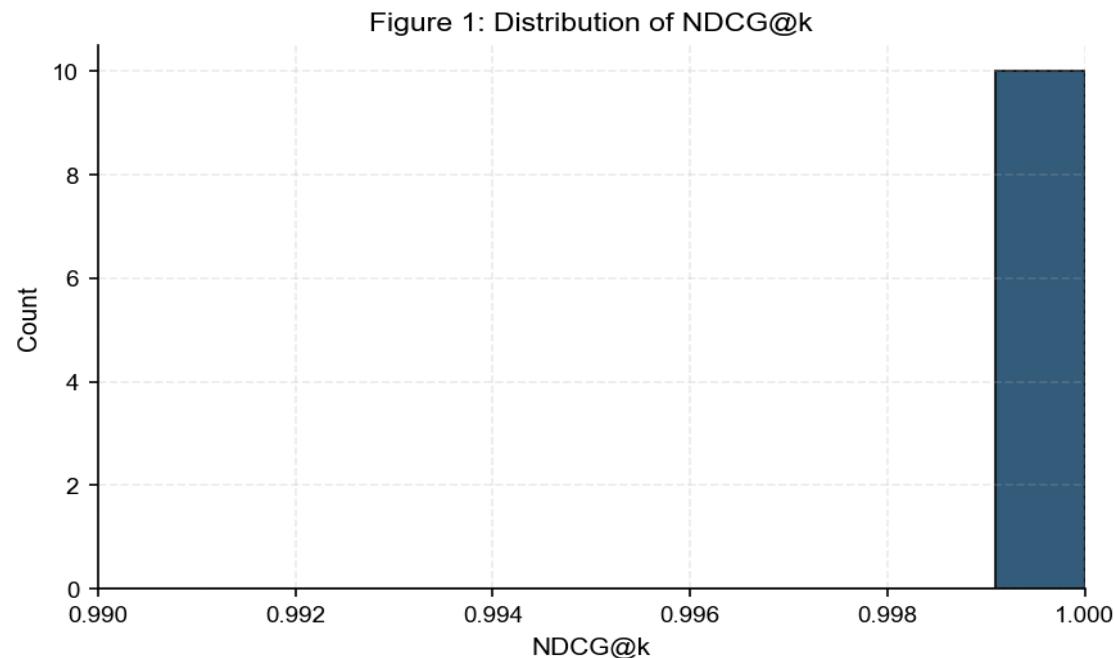
No outliers detected.

3. Visual Analytics

faithfulness bands global.png

Figure 2: Faithfulness bands across all queries

hist ndcg.png

Figure 1: Distribution of NDCG@k
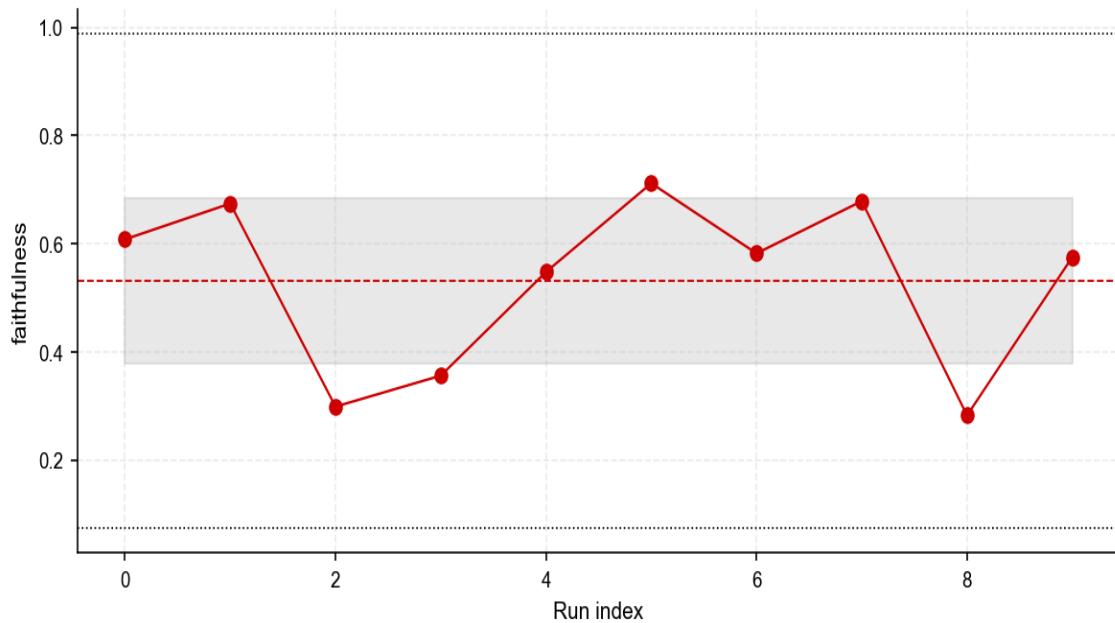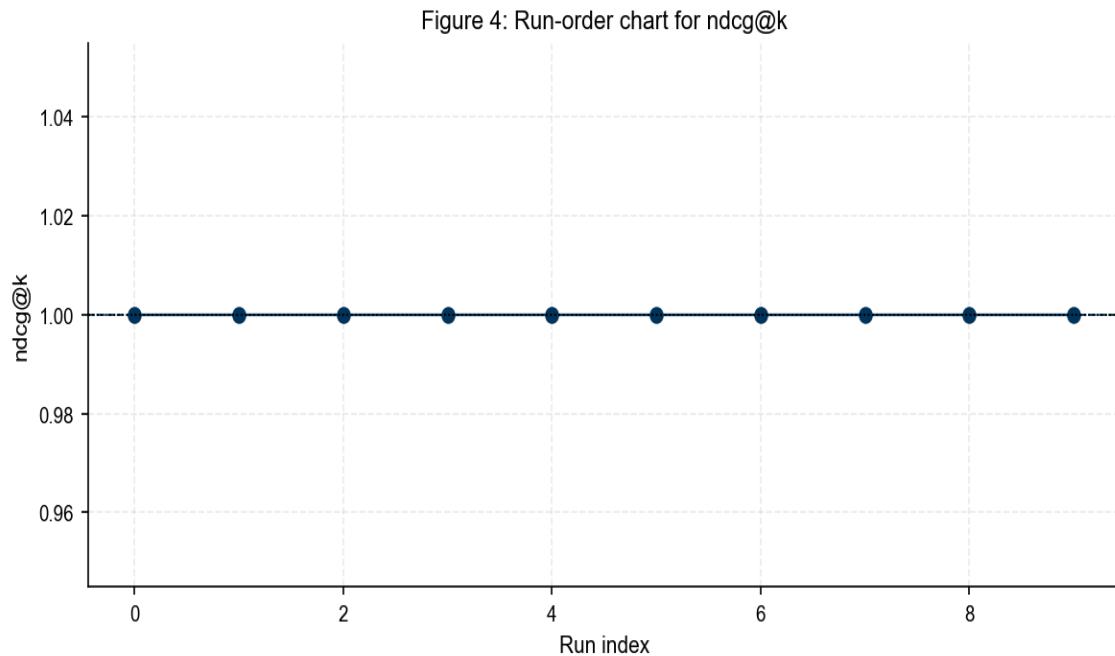
run order faithfulness.png

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=nan, ρ=nan)