

# Statistical Benchmark Report

Generated: 2025-11-15 22:55

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	0.9997
ndcg@k ci95 lo	0.9992
ndcg@k ci95 hi	1.0000
faith mean	0.9852
faith ci95 lo	0.9636
faith ci95 hi	1.0000
ndcg@k median	1.0000
faith median	1.0000
ndcg@k std	0.0006
faith std	0.0303
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.8000
faith mid thr	0.5000

## 2. Correlation and Outlier Analysis

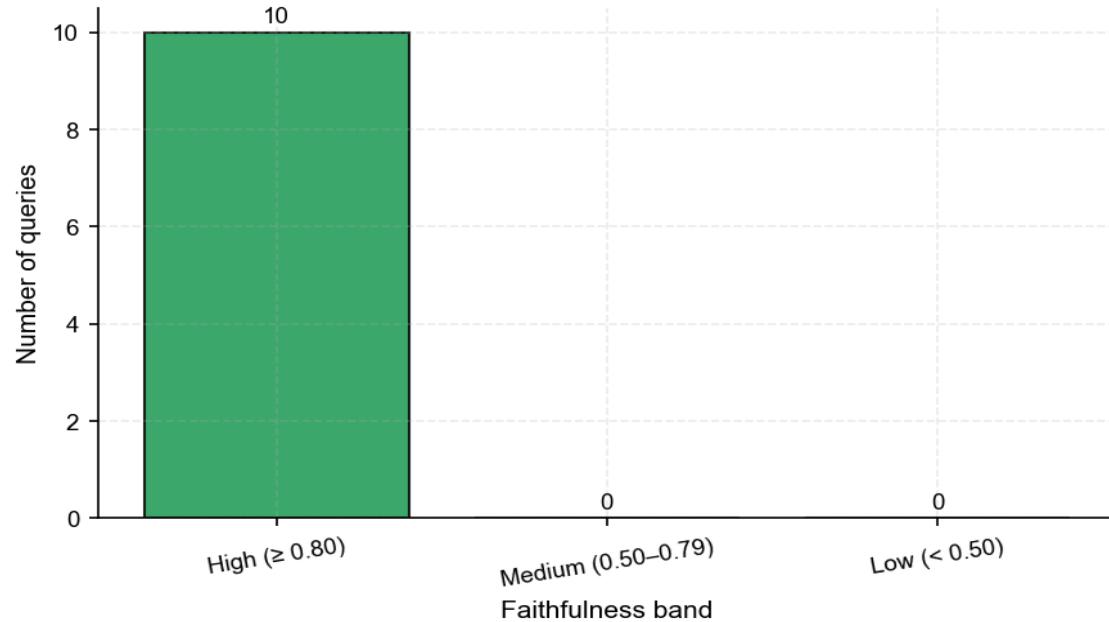
Metric Pair	Correlation	p-value
Pearson r	0.358	3.094e-01
Spearman $\rho$	0.364	3.008e-01

No outliers detected.

### 3. Visual Analytics

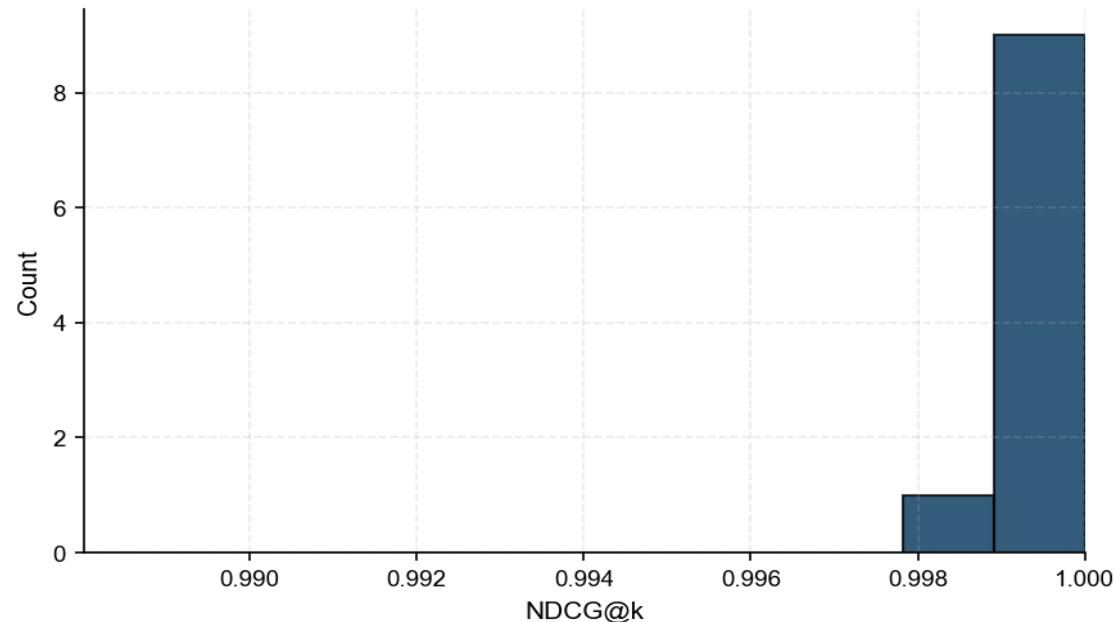
**faithfulness bands global.png**

Figure 2: Faithfulness bands across all queries



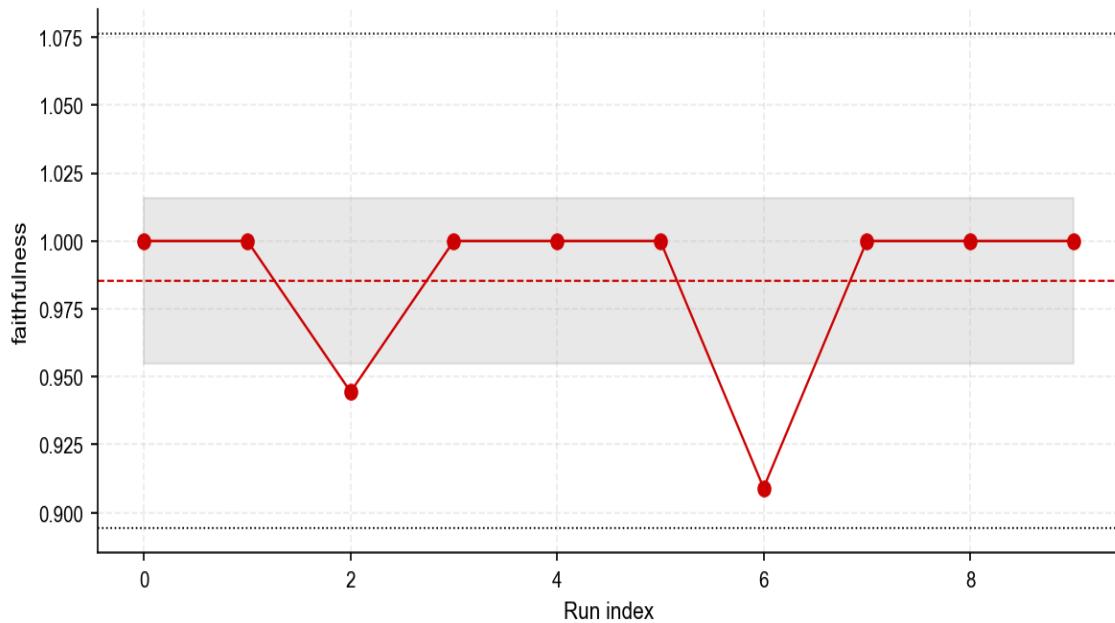
**hist ndcg.png**

Figure 1: Distribution of NDCG@k



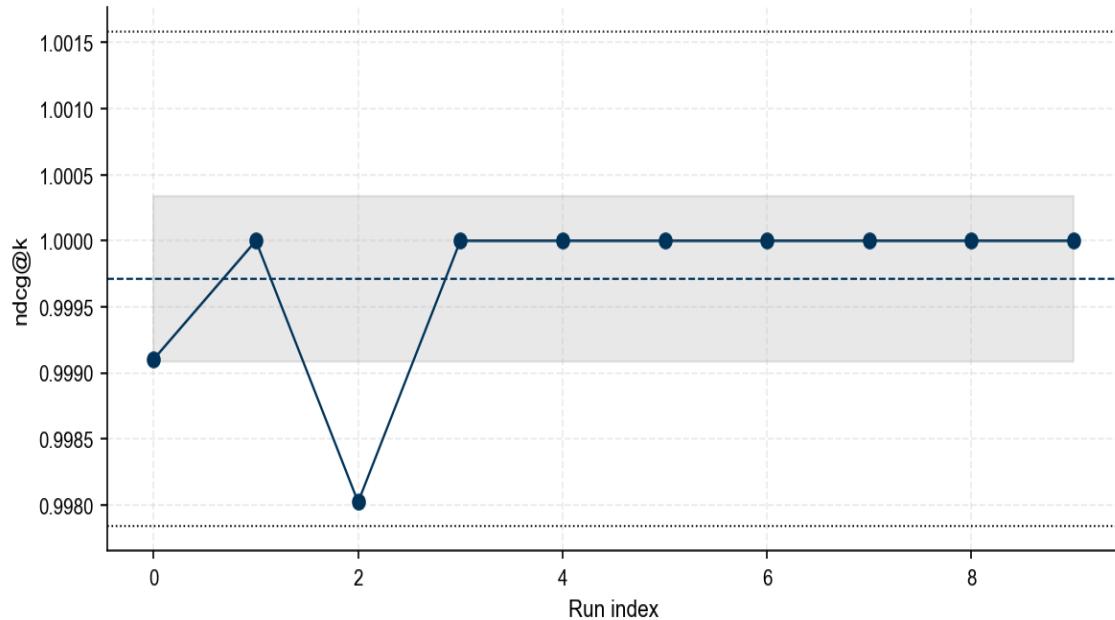
**run order faithfulness.png**

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 ( $r=0.358$ ,  $p=0.364$ )

