

Metric	Target types	Forecast formats	Properties
(Continuous) ranked probability score (CRPS)	continuous, discrete	closed-form, samples (approximation)	proper scoring rule, global, stable handling of outliers
Log score (logS)	continuous, (discrete not in scoringutils)	closed-form, samples (approximation)	proper scoring rule, local, unstable for outliers
(Weighted) interval score (WIS)	continuous, discrete	quantile or interval predictions	proper scoring rule, global, stable handling of outliers, converges to crps
Dawid-Sebastiani score (DSS)	continuous, discrete	closed-form, samples (approximation)	proper scoring rule, somewhat global, somewhat stable handling of outliers
Brier score (BS)	binary	binary probabilities	proper scoring rule
Interval coverage	continuous, discrete	interval forecasts (needs matching quantiles)	measure for calibration
Quantile coverage	continuous, discrete	quantile or interval forecasts	measure for calibration
Probability integral transform (PIT)	continuous, discrete, quantile	closed-form, samples, quantile or interval forecasts	assesses calibration
Sharpness	continuous, discrete	closed-form, samples, quantile or interval forecasts	measures sharpness, slightly different depending on forecast format
Bias	continuous, discrete, quantile	closed-form, samples, quantile or interval forecasts	captures tendency to over-or underpredict (aspect of calibration)
Mean score ratio	depends on score	depends on score	compares performance of two models
Relative skill	depends on scored	depends on score	Ranks models based on pairwise comparisons

Metric	Discrete	Continuous	Binary	Closed-form	Samples (approx.)	Quantiles	Properties
(Continuous) ranked probability score (CRPS)	✓	✓		✓	✓		proper scoring rule, global, stable handling of outliers
Log score (logS)	(✓)	✓		✓	✓		proper scoring rule, local, unstable for outliers
(Weighted) interval score (WIS)	✓	✓			THINK	✓	proper scoring rule, global, stable handling of outliers, converges to crps
Dawid-Sebastiani score (DSS)	✓	✓		✓	✓		proper scoring rule, somewhat global, somewhat stable handling of outliers
Brier score (BS)			✓		✓		proper scoring rule
Interval coverage	✓	✓			THINK	✓	measure for calibration
Quantile coverage	✓	✓			THINK	✓	measure for calibration
Probability integral transform (PIT)	✓	✓		✓	✓	✓	assesses calibration
Sharpness	✓	✓		✓	✓	✓	measures forecast dispersions
Bias	✓	✓	✓???	✓	✓	✓	captures tendency to over-or underpredict (aspect of calibration)
Mean score ratio	depends	depends	depends	depends	depends	depends	compares performance of two models
Relative skill	depends	depends	depends	depends	depends	depends	Ranks models based on pairwise comparisons