`balance-scale` dataset

Minimal & maximal fitness
Median fitness & 25, 75 percentiles
Test successrate using best
Test successrate using worst
Baseline test successrate