Comparing the interval scores of the two models at different level