

Z-test averaged on 12 pair of good vs bad methods

