Benchmarking Crimes

What is benchmarking?

Benchmarking refers to testing and analysis of softwares/systems performance in its intended environment. Benchmarking can be performed by manually running the system/software with certain inputs and workloads or by using intended benchmarking software taht can automate most of the tasks.

Why is benchmarking important?

Benchmarking tells whether system is working correctly within the environment or if there is something wrong. Lackluster perfromance in benchmarking tests indicates that system can not use hardware optimally and something is needed to be done. Benchamrking results can also indicate where the problem lies within the system.

How would you represent the results described in this comic strip?

In the comic candidates support has plummeted 19 percentage points, so from 20% to 1%. Before candidates statements about stuff fifth of voters were supporting him, but after only 1/100 of voters support him. But when expressed with percentages reader could get an idea that 19% plummet is only from candidates own supporters and not from all voters.

The article by Gernot Heiser mentions another research article on the incorrect usage of arithmetic mean. The following contains a set of tables from that article. Assuming that E, F, .. K are different benchmarks and R, M, and Z are different processors. What can you infer from Table 1 and Table 2?

In table 1 the differences in means are quite neglitable and doesn't offer clear cut solution, which processor is better. In table 2 differences in arithmetic means are much greater and choosing the better processor is much easier. But the measurements of different benchmarks don't change between tables. It is easy to portray something as better if normalized differently.

How would you use the data in Table 1 and Table 2 to compare the performance of the processors

Infering only from numbers is quite hard and in most cases not wanted.

If there is a way to put a weight on different benchmarks, I would calculate a weighted average. If for example we know that procesor that is chosen peforms mostly task that benchmark H measures, I would put more weight on that performance.