Lighthouse Variance One-Pager

Author: Patrick Hulce (patrick.hulce@gmail.com)
Lighthouse PM Contact: esweeny@
Updated: March 2019

Where does variance come from?

The page, the client machine, the browser, and the network are all responsible for variance.

What has the largest impact on score movement?

Differences in the page are typically responsible for the largest performance variations. A/B tests run by the page, random/non-deterministic code branches, and dynamic content all affect performance characteristics. During Lighthouse team's recent analysis, some sites were found to have at least 14 unique versions out of just 25 sequential visits!

Differences in the client environment are also responsible for large performance variations. During Lighthouse team's recent analysis, limited system resources caused a metric variation of almost 300% in some cases!

What if the page doesn't seem like it's changing though?

Even when the page looks identical, the page can be changing behind the scenes. During Lighthouse team's recent analysis, one site had identical above the fold screenshots between two visits but one version issued only 40 network requests while the other issued more than 170! Many pages will also store the state of ongoing tests in a cookie so manual investigation through page refreshing frequently won't be enough to identify this as the culprit.

What can I do to mitigate variance?

Disable non-deterministic features (like A/B testing) during testing. Use wprgo to replay the same network resources if you must test against production sites.

Repeat tests and use the median run. 3 is good, 5 is better. Use pwmetrics to do this automatically for now. A lighthouse-built solution is coming soon.

Run tests on the exact same, dedicated hardware. User machines tend to have unpredictable workloads running alongside Lighthouse tests. Burstable and serverless cloud infrastructure can be highly variable. Use dedicated test hardware/instances with at least 2 cores for the most stable results.

What is the Lighthouse team doing to help with variance?

A few efforts are underway:

1. **More robust simulation techniques to mitigate browser-introduced variance.**
2. **Better tools to automatically repeat and merge the results of multiple runs.**

3. **Re-evaluating the scoring curve for ways to reduce the impact of variance.**

What is a reasonable expectation for variance?

When proper steps are taken to mitigate variance, Lighthouse scores should be within +/- 3 points, and Lighthouse metrics should within +/- 4%. If you've taken all the steps outlined here and are still seeing variability beyond this range, contact the Lighthouse team.

Other resources:

- lighthouse/variability.md
- LH Variability and Accuracy Analysis