Page loading time on Amazon.com with different settings

Tyler Chun, Jaehee Jeong, Yolanda Jin, Harrison DiStefano

1. Introduction

As consumers turn to online shopping amid the COVID-19 pandemic and shocks to global supply chains as well as increases in botting, purchasing high demand items over the internet in time for the holidays has only gotten more difficult recently. One way holiday shoppers can improve their chances on hot products, without resorting to botting themselves, is to have an optimal setup so that shopping web pages load faster, helping them get to the checkout page before stock runs out.

In this paper, we show how we designed a full factorial experiment to identify how choice of browser, internet speed, and web page layout affect website loading times on popular online retailer Amazon.com. The full factorial design we used is outlined in Section 2. In Section 3, we first fit a linear regression model that showed all three main factors were significant as well as the secondary interaction effects of device layout with speed and speed with choice of browser. However, model assumptions weren't met due to non-constant variance in the loading time response variable so we performed a Box Cox transformation on the data and re-fit a linear regression model. This second model maintained that internet speed was the most significant factor, while showing that at slower internet speeds Firefox was the fastest browser. These findings could help consumers with slower internet speeds to reduce their loading times by using Firefox.

2. Experimental design

To test what effects webpage loading times we used a full factorial design with 3 factors with the response variable being full loading time of the webpage in seconds and the three factors being Internet Speed (2.5mpbs-50mbps), Browser (Chrome, Firefox or Opera), and Page Layout (desktop or mobile). The full factorial design allowed us to examine interaction effects between the factors in addition to the primary effects. Each combination of factors was replicated 3 times to try to account for variability in experimental conditions beyond our control such as changes in network conditions, computer software latencies, or to what is loaded on the web page between each run.

For each run, we observed the following procedure:

1. Open a new Private/Incognito browser window–this ensures the browser cache is empty so each webpage element will need to be re-downloaded each time
2. Test the internet speed on https://www.speedtest.net/ to verify it falls in the range
3. Open the inspection console for the browser and navigate to the Network tab
4. Set the page layout to either Desktop or Mobile depending on which is being tested
5. Navigate to https://www.amazon.com and wait for loading to complete
6. Note the final loading time and Close the browser completely

Variations in internet speed were achieved by switching between WiFi, Wired ethernet connection, and various VPNs on the computer we ran the tests on. The results are shown below in Table 1.

Table 1: Experiment Design and Collected Data

				seconds		
Run	Page Layout	Speed	Browser	Replicate 1	Replicate 2	Replicate 3
1	D	1	C	2.95	4.09	2.92
2	D	1	F	3.75	3.08	3.25
3	D	1	O	2.25	2.15	2.52
4	D	2	C	4.46	3.51	3.86
5	D	2	F	3.35	3.61	3.52
6	D	2	O	3.68	3.38	4.57
7	D	3	C	6.7	3.87	3.91
8	D	3	F	10.86	5.69	5.82
9	D	3	O	5.12	5.41	6.74
10	D	4	C	25.35	24.97	25.77
11	D	4	F	18.55	16.67	18.98
12	D	4	O	23.44	24.07	24.43
13	M	1	C	2.45	1.67	1.45
14	M	1	F	5.13	4.84	4.78
15	M	1	O	1.61	1.75	1.72
16	M	2	C	1.8	4.12	2.3
17	M	2	F	5.27	5.13	5.16
18	M	2	O	3.42	3.63	3.28
19	M	3	C	2.99	4.76	2.5
20	M	3	F	5.75	4.84	4.73
21	M	3	O	5.31	5.52	5.17
22	M	4	C	31.93	33.58	34.42
23	M	4	F	15.83	17.80	15.63
24	M	4	O	26.10	30.45	32.54

Page layout is noted as either Desktop (D) Laptop (L). Four levels of internet speed were used with average speeds of 50 mbps (1), 30 mbps (2), 10 mbps (3), and 2.5 mbps (4). The three browsers being tested were Chrome (C), Firefox (F), and Opera (O). Results from each of the three replicates are shown in seconds.

3. Data analysis and results

Upon initial inspection of the effects of the three primary factors on loading time, neither device nor browser seem to have much of an effect on loading times by themselves, while internet speed seems to have a large effect as we would expect. In particular the runs with 2.5 mbps internet speed have drastically slower loading times. This first look at the factors is summarized in Figure 1.

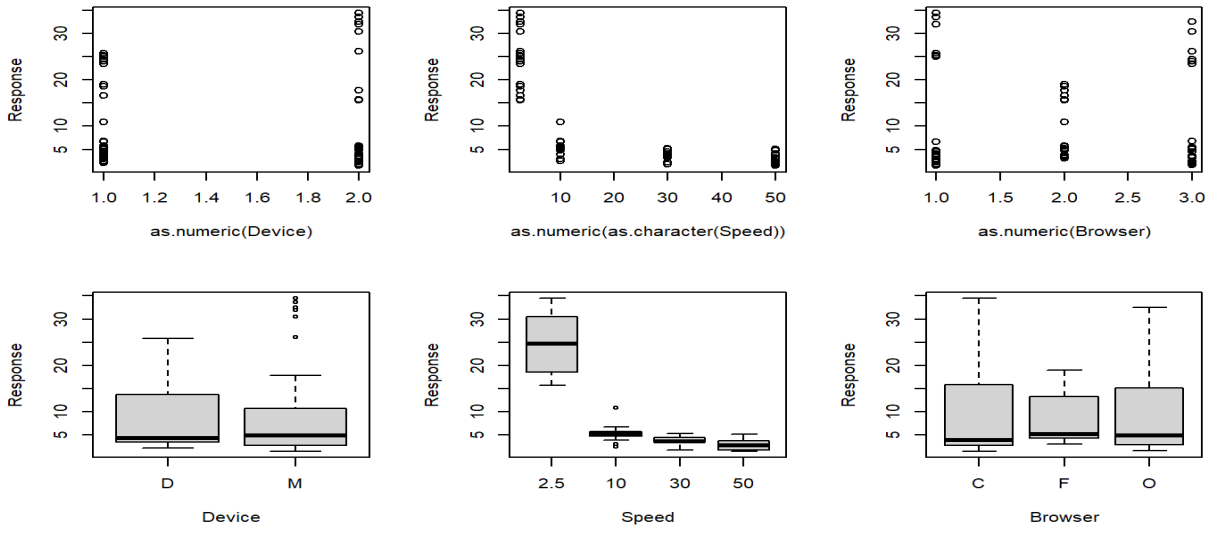

Figure 1: Plots of each primary factor against the page loading time in seconds

3.1. Model fitting - Model 1

We first fitted the data into a regression model with main factors and two factor interactions. In the model summary and anova table, the results were very similar. Our replicate variable is not statistically significant with very high p values which means replicates are not different statistically and it is a good sign. Main Factors - Device, Speed, and Browser were statistically significant. The interaction between Device and browser was not statistically significant. The interaction between Device and Speed and the interaction between Speed and Browser were statistically significant.

Table 2: Model 1 - Regression Summary Table

Coefficients	Estimate	Pr(>|t|)	
(Intercept)	27.12306	< 2e-16	***
DeviceM	4.35889	5.08E-05	***
Speed10	-22.515	< 2e-16	***
Speed30	-24.00167	< 2e-16	***
Speed50	-24.65889	< 2e-16	***
BrowserF	-11.51625	2.30E-14	***
BrowserO	-2.54542	0.025044	*
replicate2	0.0225	0.963798	
replicate3	0.08	0.871808	
DeviceM:Speed10	-5.4	1.70E-05	***
DeviceM:Speed30	-3.98667	0.000966	***
DeviceM:Speed50	-4.17889	0.000576	***
DeviceM:BrowserF	-1.15417	0.247436	
DeviceM:BrowserO	0.09417	0.924334	
Speed10:BrowserF	14.25333	4.97E-14	***
Speed30:BrowserF	13.09167	8.97E-13	***
Speed50:BrowserF	13.64333	2.25E-13	***
Speed10:BrowserO	3.92167	0.006958	**
Speed30:BrowserO	2.81667	0.048703	*
Speed50:BrowserO	1.91	0.176944	

Model 1 - R squared was 0.968.

Table 3: Model 1 - Anova Summary Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Device	1	6.8	6.79	2.3247	0.1333922	
Speed	3	5710.8	1903.59	651.778	< 2.2e-16	***
Browser	2	46.4	23.21	7.946	0.0009745	***
replicate	2	0.1	0.04	0.014	0.9861143	
Device: Speed	3	74.3	24.77	8.4797	0.0001104	***
Device: Browser	2	5.8	2.9	0.9927	0.3774856	
Speed: Browser	6	474.9	79.15	27.102	2.19E-14	***
Residuals	52	151.9	2.92			

We had a very high R-squared of 0.968 and most of our factors were statistically significant. However, we validated our model with the model assumptions and we found that our residual variance is not constant and it's increasing as fitted values increase on Residuals vs. Fitted plot and Normal Q-Q plot shows that there is some discontinuity. We highly suspect outliers may have been caused by cookies and cache differences on the browsers, such as run 8 which is abnormally higher than the other two replicates. Also, we realized that we have an imbalance in response values (The data shows it is centered near Fitted values around 5) and that led us to transform y.

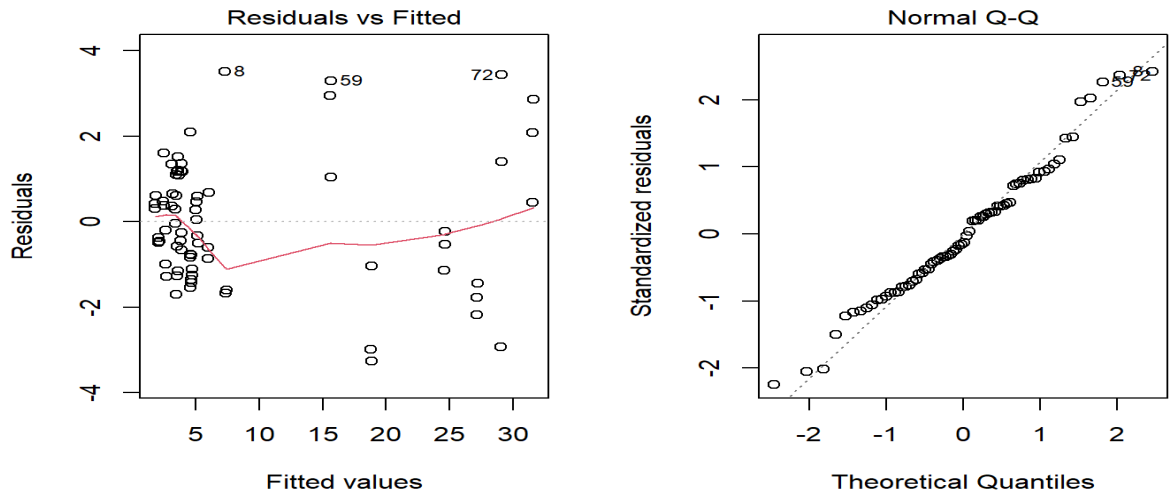

Figure 2: Diagnostic Plots on Model 1

3.2. Model fitting - Model 2

Since we find that the model 1 using a linear regression does not meet the assumptions, we decide to drop the outlier and use box cox transformation creating a linear regression. The lambda of -0.7 was used. Additionally, we change the Speed variable to a numerical value because it could be better since it can also be considered as a quantity even if it was an average value of the speed each run. When we check the model summary table with this model, DeviceM and Speed main effects are significant and DeviceM:BrowserF and Speed:BrowserF interaction effects are significant. In the anova table, the results are very similar, Device, Speed, and Browser main effects are significant and Device:Browser and Speed:Browser interaction effects are significant.

Figure 3

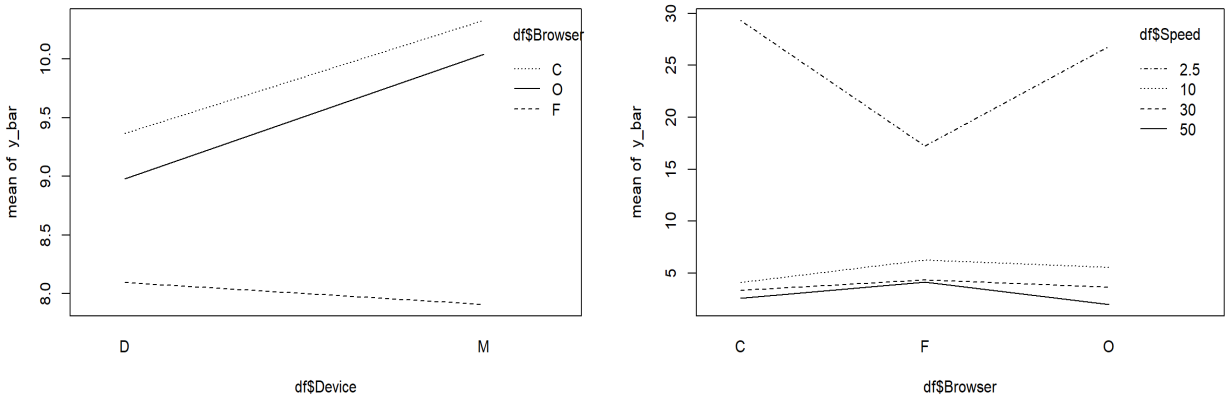

Figure 3: Interaction Plots Between Device vs Browser and Browser vs Speed

The interaction plot above between Device and Browser shows that Firefox is generally faster than Chrome and Opera and mobile layout works better than desktop layout with Firefox, unlike Chrome or Opera. For the interaction plot between Browser and Speed, when the internet speed is fast the loading times were almost the same for all the three browsers, but when the internet speed was slow which was 2.5, Firefox works so much faster than the other browsers.

Figure 4

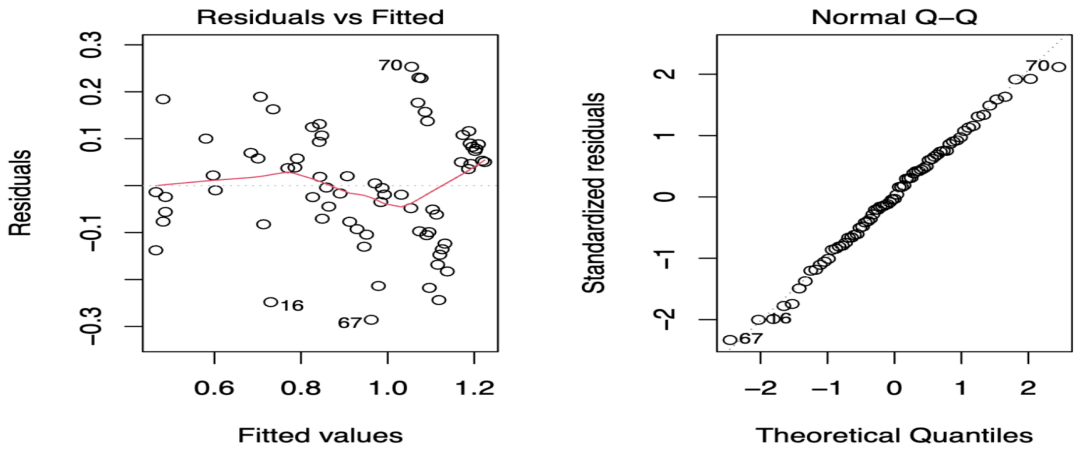

Figure 4: Diagnostic Plots on Model 2

This model performs better than the first model. The R-squared value of 0.76 is smaller than the previous R-squared. Also, when we look at the residual plot above, the red line is almost horizontal which means that the variances are constant. On the normal Q-Q plot, most of the observations are on the dashed line which indicates that the model meets normality assumption.

Table 4: Model 2 - Regression Summary

	Estimate	Pr(>|t|)	
Intercept	1.216009	< 2e-16	***
DeviceM	-0.112954	0.09527	.
Speed	-0.010304	8.45E-08	***
BrowserF	-0.116709	0.12019	
BrowserO	0.035285	0.62593	
replicate2	0.005479	0.88761	
replicate3	-0.016579	0.66909	
DeviceM:Speed	-0.002127	0.21355	
DeviceM:BrowserF	0.218293	0.00639	**
DeviceM:BrowserO	0.103918	0.17826	
Speed:BrowserF	0.00519	0.0152	*
Speed:BrowserO	-0.002781	0.18275	

Table 5: Model 2 - Anova Summary Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Device	1	0.04866	0.04866	2.788	0.0100273	
Speed	1	2.74277	2.74277	157.1355	< 2.2e-16	***
Browser	2	0.18243	0.09121	5.2257	0.008138	**
replicate	2	0.00784	0.00392	0.2247	0.79943	
Device:Speed	1	0.02425	0.02425	1.3893	0.24325	
Device:Browser	2	0.1327	0.06635	3.8013	0.027999	*
Speed:Browser	2	0.26472	0.13236	7.5831	0.001172	**
Residuals	59	1.02983	0.01745			

3.3. Results

Our analysis indicates speed is the most significant factor to reduce web page loading time. The other factors layout type and browsers give mostly similar results. Firefox browser has a strong interaction effect with slow speed. At a low speed of 2.5Mbps, Firefox loads twice as fast as Chrome or Opera at around 17 seconds. At a speed of 10 Mbps or higher, all browsers have a consistent loading time of around 5 seconds. Therefore, at slower internet speed, it's better to use Firefox if the page has many elements such as with a shopping website. Other interactions are not significant.

The residual analysis confirmed that the residual vs fitted has constant variance with no obvious trends. The normality plot shows all the points closely aligning with the normal dotted line.

4. Summary

Our experiment investigates the effects of different settings on page loading time on Amazon.com. Our findings showed that the page loading time can be affected by internet speed, browser type, and page layout (whether desktop layout or mobile). The statistical analysis suggests that internet speed is the most significant factor, and Mobile layout has some effect in improving the loading time. Overall, the three browsers produce similar results at high internet speed, with Firefox browser performing significantly better at low speed of 2.5 Mbps. Chrome and Opera loads faster using desktop layout, and Firefox performs slightly better using Mobile layout.

Due to the current ongoing covid epidemic, many families must stay at home and heavily rely on their home internet setup for work, entertainment, and the upcoming holiday online shopping. Our study introduces a simple way for users to test between different options and select the best setup for their home. In this study, we used a 24-run full factorial design with one factor at 2-level and two factors at 3-level.

5. Future Improvements

We could have tested mobile layout performance on actual mobile phones since it is a more natural setting. Our study result is also dependent on many other factors that we cannot control. Internet Latency is a confounding variable that can affect the loading time. We tried to test the internet speed before each run to ensure the speed is within the expected speed range. Another factor that cannot be controlled is the fluctuating traffic on Amazon.com. We performed three replicates to ensure results are consistent and not greatly influenced by traffic. Furthermore, Device hardware such as ram and CPU setup could have also impacted website loading time.

6. Acknowledgements

Thanks to our professor Xu for feedback to further improve our model.

Appendix

With the first model, we also ran a Tukey test to see the main factors and the result was not exactly the same as the regression model. Device variable is not statistically different between Mobile and Desktop. Most of the possible events in speed are statistically significant, but speed between 50 and 30 Tukey did not find a difference. Tukey also noted that Browsers between Firefox and Chrome are different and Opera and Firefox are different, but there was no difference between Opera and Chrome.

Table A-1: Model 1 - Tukey Test with Main Factors

Factors	Levels	Difference	P-Values
Device	M-D	0.6141667	0.1333922
Speed	10-2.5	-19.1566667	0
Speed	30-2.5	-20.6922222	0
Speed	50-2.5	-21.5638889	0
Speed	30-10	-1.5355556	0.0452395
Speed	50-10	-2.4072222	0.0005442
Speed	50-30	-0.8716667	0.4272319
Browser	F-C	-1.84625	0.0013069
Browser	O-C	-0.33625	0.7752172
Browser	O-F	1.51	0.0096095
Replications	2-1	0.0225	0.9988539
Replications	3-1	0.08	0.9856117
Replications	3-2	0.0575	0.9925397