Jonathan Lam
Mr. Ecsedy
H. Alg. 2 per. 8
11 / 1 / 15

The Browser Wars: The Competition between IE and Chrome

Linear Regression Project Report

With programming being my favorite pastime, especially web development, the so-called "Browser Wars" have always intrigued me. From Mosaic to Chrome, a few top browsers have been competing for the highest percentage of global browser usage. What is most interesting is the downfall of the '90s giant Microsoft's Internet Explorer and the rise of Google Chrome, now by far the dominating force with 65% of browser usage.

On a more mathematical viewpoint, I predicted that the increase in Chrome's popularity would *have* to decrease Internet Explorer's usage, or else they would own over 100% (an impossibility) of the browser usage worldwide, because they are both incredibly popular browsers that (at some point) had over half (50%) of the world's usage. This would mean a negative correlation between the percentages: an increase in one variable leading to a decrease in the other. But what would make it more interesting is that they are not the only two major players in the browser market: the decline of IE's competitor Netscape Navigator and the three other browsers — Opera, Mozilla Firefox, and Apple Safari — means that the correlation would not be so simple. In other words, a decrease in one Internet Explorer usage percent does not directly indicate a percent increase in Chrome's, and vice versa. Therefore, I also predict that the relationship will be strong, but not very close to perfect.

I collected the data from W3Schools.com, a reputable web-development site. It is very up-to-date with browser statistics, being so involved with the tiniest details of programming for those browsers. I took the percentage statistics for Google Chrome and Internet Explorer for every month since the public release of Chrome in September 2008, a total of 85 data points. The table and graph of the data are attached.

The first analytical part I looked to was the line of best fit: its slope (m) is -0.59 and its y-intercept (b) is 42.88. In context, this means that for every increase in percent of browser usage by Google Chrome (x), Internet Explorer's usage (y) went down by the slope of around a half a point — a negative correlation, as predicted. The interesting part about the slope is that for every percent Google Chrome takes, it only takes *half* a point from Internet Explorer — this means that Google Chrome is stealing the other half a point of users from other browsers and not feeding directly off of IE. The y-intercept (when x is zero) means that when Google Chrome had zero percent of the global usage percent (before it was created), Internet Explorer is estimated to have had approximately 43% of the global browser usage. Similarly, the x-intercept (when y equals zero) means that when Internet Explorer has zero percent of the global usage (if and when it dies off to its successor, Microsoft Edge), Google Chrome is estimated to have around 73.17% of the global usage.

Looking at the graph, especially with the line of best fit graphed, shows a very strong negative correlation: the correlation coefficient (r) is -0.99 for the data points, almost a perfect fit. Therefore, to make any interpolations or extrapolations from the acceptable range and domain of zero to one hundred (because percents of a whole cannot be negative or go above one hundred), I would be pretty confident with my answer. To add onto this, so many data points means more data for a more reliable average. However, there are many factors that might affect the relationship that may make it not reliable. One such factor is the idea of the "browser wars": Google Chrome and Internet Explorer are not the only two browsers in the world. Therefore, their percentages do not (and never will) add up to the 100%, so their relationship cannot be totally linear and predicted. Especially in the rapid digital innovation of today, companies may fall or rise unexpectedly, drastically changing Google Chrome or Internet Explorer's hold on the global usage. Another reason why the line may not be the most accurate is because it is virtually impossible to have such popular browsers down to zero percent, a hundred percent, or anywhere close. As you can tell by the graph, the line is slightly curved, and may (slightly) better be represented by an exponential function: as x or y approaches zero or one hundred, the points deviate from the line away from the axes. For example, (3.1, 49.0) and (65.9, 7.2) are at the ends of the data set I collected, and they are far above the regression line that fits so many other points so well.

Using the regression line, some interesting predictions can be made. Using the intercepts (both x- and y-intercepts) shows interesting the interesting predictions for the percentages of global browser usage of one variable (browser) if the other variable (browser) has 0% of the usage. Another interesting interpolation is to find when they are equal: when does the graph predict that the Google Chrome's percentage of usage *equal* Internet Explorer's? This calculated to be an estimated 27.03%. Of the data points given, x and y were closest together (closest to being equal) at around 25%, which is very close to the estimated value. However, because the realistic range is so small, with data points filling most of it already, there are not many options for other interesting inter- or extrapolations. We can, however, check some other values of x or y and find their residuals just to check the strength of the line. For example, when x equals 20%, then y equals 31.08%. This is very close to the data point (20.5, 28.6) (November 2010).

The data and regression line show that my original hypothesis that the global usage percentage of Internet Explorer (y) as a function of that of Google Chrome (x) would have a negative correlation with a slope other than negative one (in other words, it is not a direct relationship of exactly *one* positive Google Chrome percentage point for *one* Internet Explorer percentage point), probably because of the popularity of other browsers. What I learned is that it is a slightly non-linear relationship (although a linear regression line fits the data very well in a realistic range). What would be interesting to analyze are the relationships between the

popularities of other browsers, which would probably be very different because Chrome and Internet Explorer are unique in their rapid rise and fall, respectively.

This was an interesting project that, above all, showed a real-world application a data set and its regression line. By analyzing a correlated set of data and its graph to see its correlation strength, intercepts, and adding meaning to a set of monotonous percentages, I learned to better understand the numbers and components of a correlated graph and line of best fit. In summary, this project was an excellent way to learn to obtain meaning and discover interesting conclusions from ordinary, everyday data.

Data Table

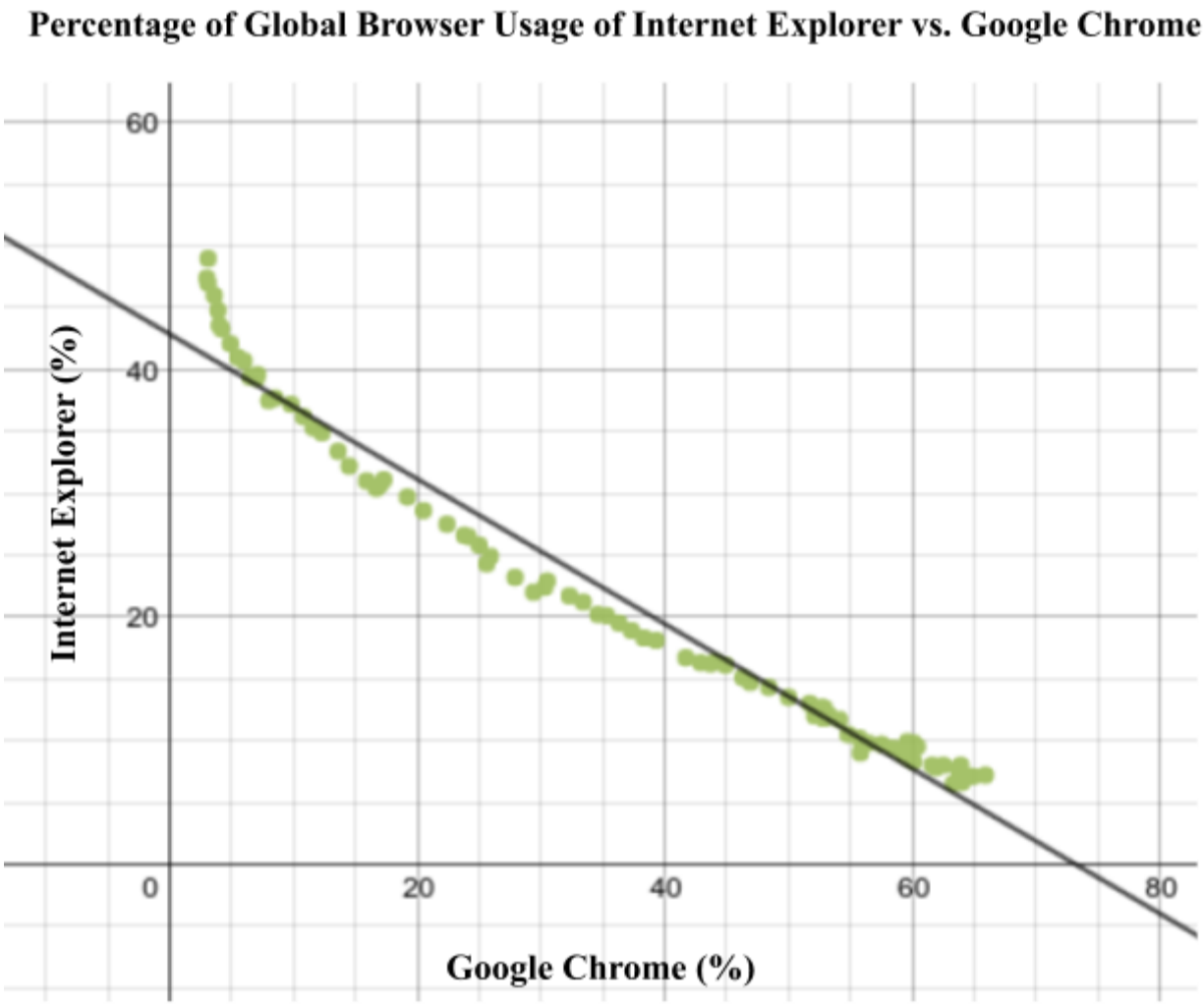
Month	**Chrome (x)**	**IE (y)**
September 2015 **last month**	65.9 %	7.2 %
August 2015	64.0 %	6.6 %
July 2015	63.3 %	6.5 %
June 2015	64.8 %	7.1 %
May 2015	64.9 %	7.1 %
April 2015	63.9 %	8.0 %
March 2015	63.7 %	7.7 %
February 2015	62.5 %	8.0 %
January 2015	61.9 %	7.8 %
December 2014	61.6 %	8.0 %
November 2014	60.1 %	9.8 %
October 2014	60.4 %	9.5 %
September 2014	59.6 %	9.9 %
August 2014	60.1 %	8.3 %
July 2014	59.8 %	8.5 %
June 2014	59.3 %	8.8 %
May 2014	59.2 %	8.9 %
April 2014	58.4 %	9.4 %
March 2014	57.5 %	9.7 %
February 2014	56.4 %	9.8 %
January 2014	55.7 %	10.2 %
December 2013	55.8 %	9.0 %

November 2013	54.8 %	10.5 %
October 2013	54.1 %	11.7 %
September 2013	53.2 %	12.1 %
August 2013	52.9 %	11.8 %
July 2013	52.8 %	11.8 %
June 2013	52.1 %	12.0 %
May 2013	52.9 %	12.6 %
April 2013	52.7 %	12.7 %
March 2013	51.7 %	13.0 %
February 2013	50.0 %	13.5 %
January 2013	48.4 %	14.3 %
December 2012	46.9 %	14.7 %
November 2012	46.3 %	15.1 %
October 2012	44.9 %	16.1 %
September 2012	44.1 %	16.4 %
August 2012	43.7 %	16.2 %
July 2012	42.9 %	16.3 %
June 2012	41.7 %	16.7 %
May 2012	39.3 %	18.1 %
April 2012	38.3 %	18.3 %
March 2012	37.3 %	18.9 %
February 2012	36.3 %	19.5 %
January 2012	35.3 %	20.1 %
December 2011	34.6 %	20.2 %
November 2011	33.4 %	21.2 %
October 2011	32.3 %	21.7 %
September 2011	30.5 %	22.9 %
August 2011	30.3 %	22.4 %
July 2011	29.4 %	22.0 %
June 2011	27.9 %	23.2 %
May 2011	25.9 %	24.9 %
April 2011	25.6 %	24.3 %
March 2011	**25.0 %**	**25.8 %**

x* is the closest to *y		
February 2011	24.1 %	26.5 %
January 2011	23.8 %	26.6 %
December 2010	22.4 %	27.5 %
November 2010	20.5 %	28.6 %
October 2010	19.2 %	29.7 %
September 2010	17.3 %	31.1 %
August 2010	17.0 %	30.7 %
July 2010	16.7 %	30.4 %
June 2010	15.9 %	31.0 %
May 2010	14.5 %	32.2 %
April 2010	13.6 %	33.4 %
March 2010	12.3 %	34.9 %
February 2010	11.6 %	35.3 %
January 2010	10.8 %	36.2 %
December 2009	9.8 %	37.2 %
November 2009	8.5 %	37.7 %
October 2009	8.0 %	37.5 %
September 2009	7.1 %	39.6 %
August 2009	7.0 %	39.3 %
July 2009	6.5 %	39.4 %
June 2009	6.0 %	40.7 %
May 2009	5.5 %	41.0 %
April 2009	4.9 %	42.1 %
March 2009	4.2 %	43.3 %
February 2009	4.0 %	43.6 %
January 2009	3.9 %	44.8 %
December 2008	3.6 %	46.0 %
November 2008	3.1 %	47.0 %
October 2008	3.0 %	47.4 %
September 2008 **public release of Chrome**	3.1 %	49.0 %

"Browser Statistics." *W3Schools*. Refsnes Data, n.d. Web. 01 Nov. 2015. <http://www.w3schools.com/browsers/browsers_stats.asp>.

Graph



Independent Variable (x): Google Chrome's percentage of global browser usage
Dependent Variable (y): Internet Explorer's percentage of global browser usage
LOBF: $y = -0.59x + 42.87$
Intercepts: y: 42.88, x: 73.17
Correlation Coefficient (r): -0.98
Link to Desmos graph: https://www.desmos.com/calculator/9i0miatfer

Calculations

Finding the x-intercept:
$0 = -0.59x + 42.88$

0 **- 42.88** = -0.59x ~~+ 42.88~~ **~~- 42.88~~**
-42.88 / **-0.59** = ~~-0.59x /~~ **~~-0.59~~**
x = 73.17

Checking a value of x:
y = **-0.59(20)** + 42.88
y = **-11.8 + 42.88**
y = 31.08

Solving for when Chrome's percent of usage (x) equals that of IE (y)

$$\begin{cases} x = y \\ y = -0.59x + 42.88 \end{cases}$$

x = -0.59x + 42.88
x + **0.59*x*** = ~~-0.59x + 0.59*x*~~ + 42.88
~~1.59~~x / ~~**1.59**~~ = 42.88 / **1.59**
x = 27.03 = ***y***

1) The introduction is written effectively and the hypothesis and reasoning for it are clearly and convincingly presented (10 pts) 10 points

2) The data table and graph are presented clearly and effectively. (10 pts) 10 points -- good to put in which month you are referring to.

3) The slope and intercepts are interpreted accurately and completely with respect to the data. (15 pts) 15 points

4) The correlation coefficient is used in conjunction with the fit of the line to the data to discuss the strength and linearity of the model. (15 pts) 15 points

5) The regression is used to make predictions for both the dependent variable (for a given value of the independent variable) and the independent variable (for a given value of the dependent variable). The predictions are explained accurately and completely in context of the variables with work shown algebraically. (15 pts) **Except for finding the x-intercept, you didn't really find a corresponding value of x for a particular value of y. You kind of made up for it by finding the point when the two browsers were equal in popularity. 13 points**

6) The conclusion/ending reflection effectively and completely discusses the results, what you have learned, and whether or not the results support or refute the hypothesis. (10 pts) 10 points

Total 73 points. Mostly ACCOMPLISHED to EXEMPLARY.