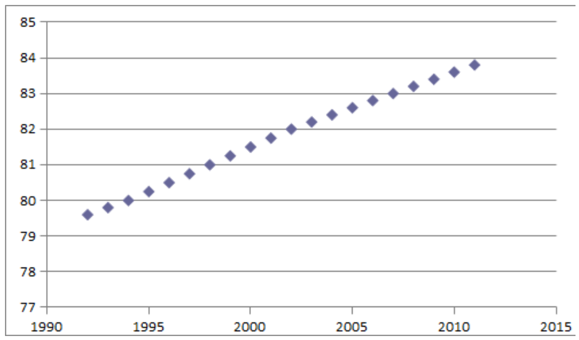
Gabriel Hackworth
Algebra II #1

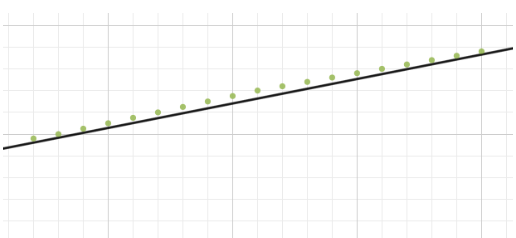
A Study of Correlation Between Average Income and Lifespan in Japan

For my project, I chose to study the relationship between income and lifespan. Japan, being both a highly industrialized country, and also having many of the world's oldest people, seemed a suitable subject. This analysis is relevant to nearly anybody that works in healthcare or the economy, as it was intended to provide insight in how the two fields are related (or, in this case, unrelated).

To the right is a scatterplot showing the average lifespan of people from Japan. (Note: due to the fact that age is measured to the closest year, the data has been adjusted to better reflect a linear increase in lifespan.) As we can see, the average lifespan increases by about 1 year every 5 years that pass.



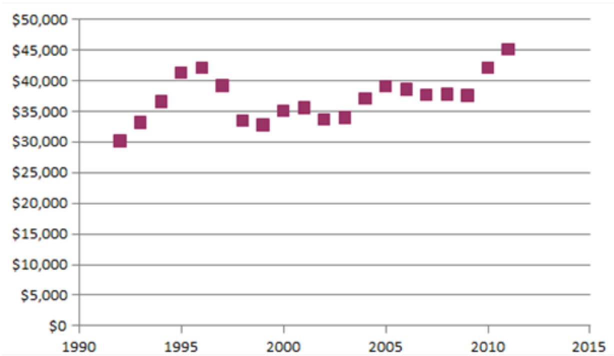
To better understand the scatterplot, it is necessary to know several things. A regression value is a number that will tell you how closely given numbers match an equation. Numbers closest to 1 and -1 show a very strong correlation between the numbers and the function, and numbers closer to zero show a very weak correlation between them. For example, the image to the left shows the data from above, as well as the equation y=0.2248x-368.191 (a linear equation stemming from the parent y=mx+b), which is the black line on the graph.

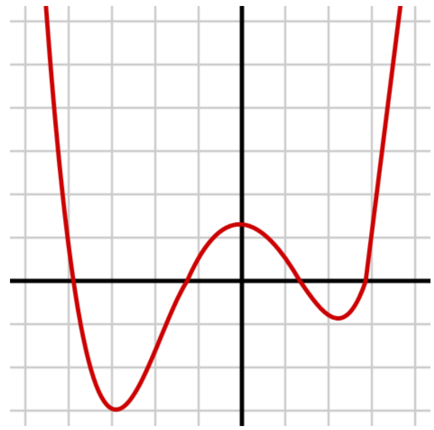


If we further break down the equation y=0.2248x-368.191, we can better understand the correlation between the variables. For example, the number "0.2248" is the *slope* of the linear equation, meaning that every time the x value is increased by 1, the y value goes up by 0.2248. This is very close to an increase of 1 year of lifespan per 5 years (which means that it increases by roughly .2 years each year that passes, very close to the 0.2248).

As we can see, the data points (showing lifespan) are very near the linear equation, showing a very strong (linear) regression. This is also shown by the fact that, when calculated, this data gives a regression value of r=0.9988, which is very close to 1. There is a very strong correlation between the *year* and average lifespan in Japan. Although lifespan is measured in whole years, the data collected can be used to infer the average lifespan for other years as well. For example, if we used the number 1991 (substituting the year for the variable x), we can see that the average lifespan for 1991 would be 79.38, or roughly 79 years and 5 months. Also, it could be estimated that the average lifespan as of the year 1994.5 would be approximately 80.1726, or about 80 years and 2 months (though age is measured in whole years, partial values can also be substituted for x in the equation and plotted accurately). It could also be estimated that the average lifespan for the year 2012 (which hasn't actually been calculated yet due to the recency of the year) would be around 84.1066, or 84 years and 1 month.

Although there is very strong correlation between years and the average lifespan, there is little correlation between *income* and lifespan. As we can see below, the average income of Japan has fluctuated wildly over the past two decades. Changing from anywhere between $30k and $45k seemingly randomly, it appears that average income does not follow any sort of function. The type of function and/or regression that the data most closely follows is a *quartic function*, a function that follows the format of $y=ax^4+bx^3+cx^2+dx+e$. Due to the magnitude of the data points collected, the coefficients of several variables are extremely large (which is also why there is no graphed equation - most calculators cannot graph an equation using the number twenty-seven *trillion* in it quickly, if they can at all). The image below is an example of a quartic function for reference. The equation that most closely resembles the data collected for average income is as follows:





$y=-1.7x^4+13619.599x^3-40924041.86x^2+5.4652513*10^{10}x-2.737*10^{13}$

As you can see, the equation is quite unwieldy, and difficult to draw conclusions from. What we *can* deduce is that the equation itself is of a very large scale, as the "$-2.737*10^{13}$" shows that we are having to reduce the output values by a very significant amount.

The relationship between year and average income is quite poor. The r^2 value (similar to *r,* only for quartic functions) is only *0.6406*. While still closer to 1 than it is to 0, a value near 0.6 shows relatively poor correlation between the values. Perhaps a sextic or septic equation would be a closer match, though it is more difficult to calculate such large equations with such large numbers. We can still calculate theoretical values just as we did for the lifespan data, though it is difficult due to the magnitude of the numbers involved.

What is important to recognize is that there is poor correlation between lifespan and average income. Not only is it evident that lifespan will continue to increase regardless of income (or is at the very least unrelated to income), but it is also clear that it is near impossible to accurately estimate the average yearly income of an entire country. Average income depends upon the economic state of the country, and there are simply too many variables that affect it to make any sort of accurate prediction.

Works Cited

"Data | The World Bank." *Data | The World Bank*. N.p., n.d. Web. 30 Sept. 2013.

File:Polynomialdeg4.svg. Digital image. *File:Polynomialdeg4.svg*. N.p., n.d. Web. 30 Sept. 2013.

"Oldest People." *Wikipedia*. Wikimedia Foundation, 28 Sept. 2013. Web. 01 Oct. 2013.