The sentiment expressed by Chris Anderson in 2008 is a popular notion in the Big Data community. “Causality is dead,” say the priests of analytics and machine learning. They argue that given enough statistical evidence, it’s no longer necessary to understand why things happen – we need only know what things happen together. Discuss the circumstances when it makes sense to act on a correlation and when one should not act.

Many predictive models are based on correlated variables. How should such models be developed and used so the predictions can be acted upon?  Some of these types of models are very complex.  Do we need to understand why they work?

I would agree with the assessment by Chris Andersen. More often than not, correlation is used as sufficient information to prove out a hypothesis. These models should be developed with unbiased scientists working to solve a problem. I think that when the scientist has a motive, it’ll be easier for them to take a position that may not accurately reflect the causality and not just the correlations. In order to build the more complex models, having an outline of the experimental design is key. The readers of the study should be aware of all independent variables, treatment variables, and classification variables. In addition, we’d need to understand how the data was generated. If the survey methodology is flawed, then this will also impact your results.

I found a really great article that relates to Chris Anderson from the Harvard Business Review. It talks about when to make the decision that something is not just correlation but indeed causation. The article speaks to the question, “Can I take action based on the correlation findings?”. The answer being it depends. There are two factors that must be considered when taking this approach1:

* Confidence that the correlation will reliably recur in the future. The higher that confidence level, the more reasonable it is to take action in response.
* The tradeoff between the risk and reward of acting. If the risk of acting and being wrong is extremely high, for example, acting on even a strong correlation may be a mistake.

The article outlines several examples of when this trade-off is okay and when it needs to be assessed further.

1. Ritter, D. (2014, November 01). When to Act on a Correlation, and When Not To. Retrieved May 18, 2017, from <https://hbr.org/2014/03/when-to-act-on-a-correlation-and-when-not-to>