EPPS 6323 **James Norcross**
Knowledge Mining **Assignment 1**

4 February 2023

Comparison Article Review:

Breiman, L. (2001). Statistical modeling: The two Cultures. Statistical Science, 16(3), 199-231. https://doi.org/10.1214/ss/1009213726

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289-310. https://doi.org/10.1214/10-STS330

Reading through both articles was pretty insightful and concerning at the same time. Breiman (2001) points out well the overuse of modeling, not only in the statistical sciences (S.S.), but also in other disciplines as well. It seems that S.S. is having a significant influence on the statistical methods being used in academic research, throughout most disciplines. This influence is the overusage of statistical modeling versus other methodologies; to such the extent that it has become more about the model than the data.

Shmueli (2010) take a different approach by indicating that the statistical sciences are influenced by the other academic disciplines in developing statistical models that can be used in research. Shmueli (2010) mentions the influence of as the basis of why S.S. cannot "go its own way" in developing its own unique methods in which the focus is on the data and not the model.

While I found both arguments reasonable, I found Shmueli's much less compelling. As a research scholar in the field of Public Affairs, I am expected to produce publishable research using common statistical methods that interpret the mean of data as compared to relevant theory. However, according to an article by Mergel et al. (2016) the use of "Big Data" in Public Affairs (P.A.) research is often improperly done. Computation methods and models are utilized by P.A. scholars, many times in the wrong way, producing incorrect results from the data (Mergel et al.,

2016). This, from my perspective, hurts Shmueli's argument about other academic fields influencing model and algorithmic progression in the field of Statistical Science. Additionally, why I can see the influence that academic fields may have on one another, there is no excuse for any one or several fields control the research and development of any other field. Independent fields stand alone because of their uniqueness and the fact they contribute to the body of total scholarly knowledge, separating from other fields. If a field is controlled to such a point that it limits its progression, then is that field truly independent or is it just a subspecialty of another field?

While I appreciate both scholar's argument, I find myself agreeing more with Breiman with their perspective that the field of Statistical Science needs to look beyond the basis of a model and look more towards methods that best fit the data at hand, to provide a proper interpretation of its meaning. While I can also see Shmueli's perspective that the field of Statistical Science needs to work cooperative with the family of other academic disciplines, it almost must maintain its own independence to develop itself and the proper means to compute and interpret data.

Reference:

Mergel, I., Rethemeyer, K., & Isett, K. (2016). Big data in Public Affairs. Public Administrative Review, 76(6), 928-937. https://doi.org/10.1111/puar.12625