

# Applied Sequence Notes

Kenneth Tay

## Questions to ask about your data:

- How many data points are there, and how many parameters are there? Are we in a high bias or high variance situation?
- Are there sub-populations with much fewer data points? If so, some borrowing of strength across populations would be good.
- Are there outliers? Might there be reasons for them?
- Are there missing data? Are the data missing at random, or are there reasons why the data are missing?
- Are there correlations between data points? This is always the case for time series.
- Are there correlations between errors? For example, consecutive measurements, or measurements done by the same person, might have correlated/systematic errors.
- Is there a batch effect? If so, the data points are probably correlated.