Big data safari

Joe Parker

14th November 2018

Task

Today we're going to practise a **really** useful skill in applied statistics, e.g. 'reading and summarising a paper really &^%@ quickly'. This comes in handy ***far*** more often than most researchers would like to admit. More importantly, it gives us a chance to flex our new-found statistical muscles by looking at some real studies and picking out the salient points from their analyses.

You will be split up into small groups. Pick a paper on the list. You have 30-60mins to read the paper together and pick out the statistical techniques the authors have used. We'll then get together again and you'll have 5mins to present a summary of the paper (from a statistical perspective). **Don't** get too hung up on the biology here (although you'll need to keep referring back to the authors' main question) but focus on how their analysis measures up to the principles we've learnt on this module.

You should consider:

- What the basic biological question is
- What the null hypothesis looks like (consider this both as a phenomenon, and graphically - ***can you draw a sketch-graph of the null model?***)
- What the alternative hypothes(es) is
- What factors, treatments, variables and blocks are present
- What the sample size is, and the number of effective replicates (they may not be the same…)
- What model has been fitted
- What alternative models were (or could have been) fitted
- What the model selection criterion was
- What the implicit, and explicit assumptions were
- What transformations, smoothing, pooling or averaging steps were applied, and whether they were justifiable and/or could have affected the results.

Also try to determine:

- Whether all ***in silico*** analyses are fully detailed
- Whether the input data are available
- Whether the study is reproducible - if you had to, could you repeat their analyses in reading week next week?

You will be strictly time-limited to 5mins, so focus on the main points only and be clear which are the most important

---

Papers

The papers are all taken from the journal ***PLoS Comp. Biol.***, e.g. they're open-access:

- Shapiro ***et al.*** (2017) - Temperature and parasites' effects on malaria transmission in mosquitos
- Brady ***et al.*** (2017) - Transcriptomics of hypoxic cancer cells
- Sawala & Gould (2017) - Neuron sex and body growth in ***Drosophila***
- Wang ***et al.*** (2017) - Modelling yeast colony formation
- Lau ***et al.*** (2017) - Spatio-temporal epidemic modelling and ebolavirus
- Parag & Vinnicombe (2017) - Sound and vision*
- Li ***et al.*** (2017) - Neurobiology of the thalamus

---

*Ashes to ashes