AAS 235 : WGAA Splinter Session into the 2020s *: 2020 Jan 5*

The Feigelson List

Vinay Kashyap (CfA)
Mark Weber (CfA)
Aneta Siemiginowska (CfA)

What is the problem?

Some time ago, Eric Feigelson, the first Statistics Editor of the AAS Journals observed that, at the point of submission, too many astronomers — no one in *this* audience! — were doing too many wrong things.

This covered everything from straight-lines fitting to time series analysis to clustering and classification to dealing with Poisson data.

http://hea-www.harvard.edu/AstroStat/aas235/EF_list_mar2017.pdf

It isn't so much lack of sophistication in analysis (understandable), but rather the lack of rigor and careless methodologies that are problematic. E.g.:

- not understanding the error envelope of a linear fit
- not checking whether a fit is good
- over interpreting misspecified models
- over interpreting *p*-values
- insufficient attention paid to residuals
- etc.

My focus here is not about what is *in* the list, but rather that *there exists* such a list

A Steep Learning Curve

To a certain extent, this is just part of the process of doing science — you do something, your peers tell you how you messed it up, and you do better the next time.

There are books, tutorials, blog posts, summer schools, lecture series, YouTube videos, Wikipedia articles, Stack Overflow Q&As, ASAIP Forums, et cetera, that people can refer to and learn from. There are courses in analysis methods that are regularly taught to astro grad students.

These are all necessary, but not sufficient:

- There is a lot of diverse and really high-quality material, but it is hard to keep up with it, and figure out what is the best strategy to deal with a *particular* problem
- Professional astronomers do not have the time to turn into professional statisticians. Collaborations and consultations take too long to mature.
- Lecture series and teaching don't scale, and can quickly go out of date.
- General tutorials address foundations and theory, but are usually targeted to non-astro audiences and do not address the specific problems faced in astro data analysis.
- Most astronomers don't know they need any expert help.
- Pointers for improvements by the Statistics Editor comes too late in the process

we publish with the statistics we know, not the statistics that we wish we knew

- We don't really have an answer.
- Definitely requires a multi-pronged approach, where textbooks, tutorials, grad-level courses, etc. are all needed. But that is not enough. There is plenty of information out there, but it is hard to find because there is a First Mile problem.
- Consider the predicament of an astronomer who realizes they have a gap in their analysis:
  - Even if they know to ask, they might not be able to properly formulate the question
    - Even if they can, they won't know where to look for reliably good information
      - Even if they do, they may (will) find they need more background knowledge
- Two facets to this: getting astronomers to realize there is a blind spot, and getting them to the right resource so they can do something about it

The First Mile

- Needs a large network, both human and internet, that can act as a guide, to make a path through the wilderness.
- A non-exhaustive list:
  - Build long-term relationships between astro and stats communities. We have several astrostatistics and astroinformatics organizations in place now: WGAA at AAS, the Astrostatistics Interest Group (AIG) at the American Statistical Association, the Astrostatistics and Astroinformatics Portal (ASAIP) at the Penn State Center for Astrostatistics, the International Astrostatistics Association (IAA).
    - Maybe these groups should start a journal!
    - Proactive "advertising" of available, reliable, up-to-date resources
    - We must leverage the internet, and build a necessary database of Frequently *Unasked* Questions, frequently updated: with contextual explanatory text, with clear examples demonstrating good and bad methods and consequences thereof. But who will (or can) do it? (And whoever does do it will need a lot of funding!)
    - Different strokes for different folks: someone who needs help with the stats they need to finish a paper is in a different boat than someone who is looking to broaden their toolkit
- Give us your ideas! How can we start building this infrastructure?