MyStats 2017

5th Malaysia Statistics Conference | 2017

29 November 2017
Sasana Kijang, Bank Negara Malaysia

From Data to Knowledge : The Journey

Statistics in the new cutting edge environment: Journey, opportunities and challenges

Learning for the future

Helen MacGillivray
President, ISI
Editor, Teaching Statistics

Education/preparation of future statisticians, future and current users

- Learning from the past
- What's new, what's not
- Big data, data science, data literacy
- What's needed, what's special about statistics

From ISI President’s message October 2017

“**An article on employment in workplaces increasingly driven by ‘big data’ and ‘big data analytics’.**

In commenting on the various ‘hybrids’ of skills and backgrounds needed, not once did the word statistics appear, but the only workplace person quoted was a statistician, who was also explicitly identified as a statistician! The statistician stressed the need for ability to **analyse** and **communicate** as well as technical skills, and it was very clear that the emphasis was on the **key skills that statistics professionals and educators have been highlighting for decades, including collaboration, communication, and interpretation of data in context**.”

Not new

- Advice for decades to job-seeking graduates: *look for skills in ads, look for ‘analyst’*.
- Explicit identification of skills for students and awareness of broad & technical skills
- Two decades ago, I set up double degree in maths/stats and IT.
  - Those graduates went everywhere
  - Feedback included:
    - tackle anything; foundation for further learning
    - value of statistical learning which reflects ***practice*** of statistics

Statistical investigating at the heart of the science, profession and use of statistics

Cameron (2009) considers

- desirable key components of university-based training
- consults what many "*wide and experienced*" statisticians have written (e.g. Box, 1976, Chambers,1993)
- identifies
  - formulating a problem so that it can be tackled statistically
  - preparing data (including planning, collecting, organising and validating)
  - analysing data
  - presenting information from data
  - researching the interplay of observation, experiment and theory.
- comments that *such training is an appropriate foundation for most statisticians wherever they may be employed.*

“important to take part in collection of data, or at least have the opportunity to watch data being collected or generated.”

Kenett & Thyregod (2005)

- describe the 5 steps in statistical consulting
  - problem elicitation
  - data collection and/or aggregation
  - data analysis using statistical methods
  - formulation of findings & consequences
  - presentation of findings and conclusions/recommendations.
- *“Our long-term objective is to encourage academic courses to cover the full 1–5 cycle....especially steps 1, 2 and 5”*

Same foundation for all – future statisticians & users

- Advocacy for no division at introductory tertiary; same foundation. For example, Wild (2006), MacGillivray (1998, 2005a), Cameron (2009)
- Statistical thinking, understanding & whole investigation process
  - **Identification of issues, what's needed to investigate issues**
  - **Sourcing, handling, understanding, visualising, managing data**
  - Modelling & analysis; **identify & evaluate assumptions**
  - **Interpretation & communication in context**
- Real contexts and data. Complex data but easily-understood contexts
- Student ownership of learning – beware of case studies

Learning from the past – what's gone wrong?

- Some great work internationally, nationally and locally but insufficient penetration and problems persist.
- New ways of teaching old & old sequencing
- Training for research: statistics & other disciplines
- Needs in technology resources & training
- Not enough real, complex, many-variable datasets; toy datasets
- Not enough reform in teaching real probabilistic thinking.
- Domination of 1 and 2 variables
- Not enough visualisation; evaluation of assumptions
- Lack of coherent development and statistical story
- 'The' question & 'the' answer
- Assessment is for learning
- Too much of psychology thinking e.g. analysing understanding

Big data, data science, data literacy

- Descriptions can be constructive but definitions are not
- Discussion enlightening but diagrammatic representations are not
- Big can mean many variables, many observations or both
- What can be learnt for data literacy from decades of promoting and efforts to enable statistical literacy?
- What can be learnt for data science from experiences with statistical sciences?

Some descriptions of statistical literacy

- *Good "statistical citizens": able to consume information that they are inundated with on a daily basis, think critically about it, and make good decisions.* Rumsey (2002)
- *People's ability to interpret and critically evaluate statistical information and data-based arguments appearing in diverse media channels, and their ability to discuss their opinions regarding such statistical information* (Gal 2000)
- *Develop the skills you need to:*
  - *look behind the data with which you are presented,*
  - *ask why these data are being presented in those forms,*
  - *ask what questions can be answered or what arguments are being made with these data.*
- *Become much more critical about the way data are produced, the way data are presented and the way data are interpreted.*

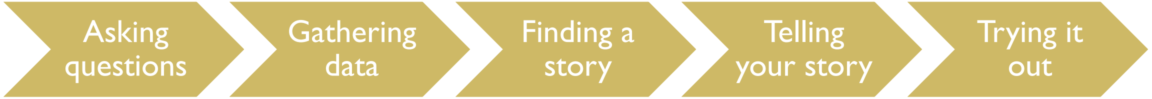
Some recent descriptions of data literacy

- *Data literacy is the ability to read, create and communicate data as information and has been formally described in varying ways.*

- *The desire and ability to constructively engage in society through and about data* http://datapopalliance.org/item/what-is-data-literacy/

- *Data literacy is the ability to interpret, evaluate, and communicate statistical information…how statistical information is created, encompassing data production*
- *Data management …. belongs to the data production phase … perhaps one aspect of data literacy that can be reserved for the specialists.* ???

- *Figure below: "Opportunities for engagement" in data literacy*


Asking questions
Gathering data
Finding a story
Telling your story
Trying it out

????????

Statistics is the science of questioning data, variation, assumptions, models, interpretations

- Maths is the servant of statistics
- Coding is the servant of data science?
- Real, large contexts and data: simple within complex
- Authentic learning experiences; authentic assessment
- Foundational understanding for future learning

Some comments about postgraduate training

- *“funding for doctoral training is primarily about ensuring a growing supply of well-trained researchers to help exploit the potential benefits of the new knowledge economy.”*
- In some countries as few as five percent of PhD graduates find permanent academic positions.
- Many PhD graduates find themselves in non-academic, non-research positions.
- More attention to the more generic and transferrable skills and knowledge that research students develop and the need to pay more explicit attention to their development.

Some comments about postgraduate training

- Reports on HDR training emphasise general research skills, analytic and critical thinking skills and many increasingly highlight the importance of statistical and data analysis skills.
  - *transferable skills … closely linked to the process of research training yet valuable to a range of other professions (for example, critical thinking, project management and statistical analysis).*
- *Research Skills for an Innovative Future: Business Views and Needs* (2012) states value of strong analytical and critical thinking skills in HDR graduates &
  - *Skill sets in data analysis, predictive modelling and decision-making are also highly sought after and there was consensus that this demand is expected to continue to increase.*
- Some reports specify data visualisation & analysis techniques without explicit reference to 'statistics' or 'data analysis', but also emphasize findings that:
  - *knowledge about designing and undertaking research, and about analysing information or data played a significantly larger role than did knowledge of their PhD disciplinary area.*

Fundamentals needed in other disciplines

- Example: Tragic case of Sally Clark included
  - Lack of identification of issues and context
  - Inappropriate data
  - Misunderstanding of conditional probabilities and incorrect multiplication of probabilities
  - More misunderstanding of conditional probabilities - ‘Prosecutor’s fallacy’
  - Withholding of (pathology) data/information

- Incorrect use of types of data
  - Ordinal variable as response in GLM’s
- Multiple testing and overuse of t
- Lack of identification and questioning of assumptions

What's needed

- Real, large contexts and data: simple within complex
- Technological and data systems know-how
  - Maths & coding are servants
- Professionals need to get involved in the nitty gritty
- Observe, listen, communicate; working with other disciplines
- Enable coherent & authentic development of fundamentals
  - Variables, variation, visualisation
  - Coherent development built up around types of variables
  - Authentic full statistical data investigations; student ownership
  - Real & data-linked probability & conditional probability
  - Authentic assessment & real communication
- Collaboration & sharing

Thank you and here's to statistics!