75 YEAR ANNIVERSARY
NCSU STATISTICS
EST. 1941

What my experiment died from: common sources of variation

KATHERINE ALLEN (NCSU)

JON STALLINGS (NCSU)

FALL TECHNICAL CONFERENCE 2017

Designing Experiments and Studies

"The statistician… makes his most valuable contribution simply by persuading the investigator to explain why he wishes to do the experiment" - Cox (1950)

- Successful statisticians are able to:
  - Understand the underlying science of a problem at some level
  - Extract the pertinent information
  - Incorporate into the design and analysis

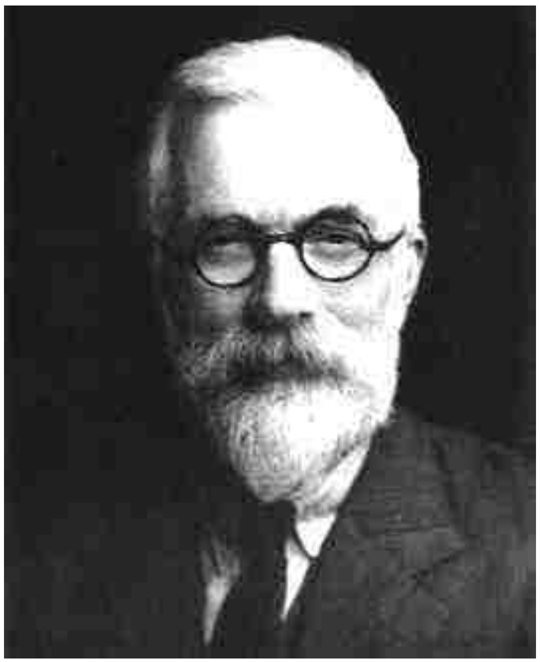
- Issues arise when a statistician lacks DOX training, and/or an ability to effectively communicate

- All statisticians are taught to analyze data but not always to design experiments (DOX)

- Analysis-only training rarely emphasizes the importance of understanding how data were generated/collected
  - DOX and survey sampling highlight how different collection procedures lead to different statistical models

Learning from Statisticians

R. A. Fisher



George Box

Gertrude Cox

Communicating with Non-Statisticians

"…if we aspire to be effective in our collaborations with scientists … we need to act like scientists" - Vining (2013)

- Statisticians wish their clients were more familiar with statistics
- **Clients wish statisticians were more familiar with their area**

- FTC sessions about training data analysts in both technical and communication skills

- Technical skills are only useful if you can effectively communicate ideas and results with your collaborators

Learning from more Consulting Experts

- Experts have been discussing improving collaboration for years
- Bliss (1969): avoid answering any questions on statistics until the scientific background of the problem is ascertained
- Zahn and Isenberg (1983) first part of session is identification of relevant aspects of the problem.
  - Emphasize that statistician should frequently reflect back to the client his/her understanding of the problem
- Hunter (1981): “Be curious. Ask a lot of questions.”
- Coleman and Montgomery (1993): article helps “bridge the gap” between statisticians and engineers in collaboration

Thinking like Scientists

- Ideas presented aren't intuitive; no one is born communicating well
- Are all experiments doomed to die an untimely death while we wait to gain collaborative experience?
- It is one thing to tell people to ask questions and think like a scientist, it is another to know how to implement that
- Ask science questions; evaluate like a statistician
- **Memorable examples might be the key to communication**

Experimental Units and Treatments

- **Treatments:** factor(s) of interest manipulated by experimenter
- **Experimental units (EU)**: what we apply the treatments to
  - Observational units (OU): part of the EUs measured if we have subsampling
- The treatment application process (which people often just call treatment) is ideally a major source of variation
- **All other major sources of variation are realized through the experimental units and the measurement process**

Common Types of Sources of Variation

- DOX textbooks say to first list out all major sources of variation but give little advice on how to figure them out
  - Hinkelman and Kempthorne (2008) discuss this at length and provide examples in their textbook

- Students/professionals learn these indirectly from experience or classroom examples

- Campbell and Stanley (1963) detail these sources of variation for education research but concepts still apply

- Framed in terms of internal and external validity

Internal and External Validity

- **Internal Validity:** how well an experiment was conducted
  - For the units in this experiment, can observed changes – or lack of changes – be attributed to the treatment?

- **External Validity:** how well the conclusions from one experiment apply to another
  - Is the experiment reproducible?
  - Can the results be generalized outside of an experimental setting?

Threats to Internal Validity

- *Treatment Replication Error:* inability to perfectly replicate a treatment application
  - Moral: All EUs should be treat(ment)-ed equally

- *State Error:* EU changes during the experiment in way unrelated to the treatment
  - Moral: Nothing is permanent but (EU) change

Threats to Internal Validity

- *Selection/Sampling Error:* characteristics in one sample of EUs/OUs will differ from another sample, both may differ from population
  - Moral: A sample is like a box of chocolates: you never know what you're going to get

- *Measurement Error*
  - Moral: Do your measurements measure up?

Threats to Internal Validity

- *Dropout*
  - Moral: Don't count your EUs before they hatch (or before you've taken measurements)

- *Selection bias:* sample collected in a way that misrepresents population
  - Moral: To thine own population be true

Threats to External Validity

- *Testing:* awareness of testing may impact outcome (Internal Validity)
- *Reactivity:* results occurred only as an effect of studying the situation (External Validity)
  - Moral: You are what you test
- *Experimenter bias*
  - Moral: You might be your own worst enemy

Threats to External Validity

- *Multiple-treatment interference (crossover designs):* when applying multiple treatments to the same EU, carryover effects may impact generalizability
  - Moral: If two treatments diverge in the woods, test them both

- *Situational Error:* differences between experiments in treatment conditions, time, location, etc.
  - Moral: It is our differences that divide our results.

Modelling

"All models are wrong; some are useful" – Box (1979)

- Data-driven models
- Data-driven models might be prone to overfitting
- Consider science-driven models instead

- Useful to have same statistician see experiment from design phase to inference

- If your experiment has died, the modelling step isn't going to save it

Beyond Morals

"My revered teacher Prof. Whitehead of Cambridge used to say…: 'The essence of applied mathematics is to know what to ignore'"

- Fisher (1938)

- Learn to critique statistical models to improve understanding of when a model is and isn't appropriate

- For every experiment, explain how the analysis from a given statistical model answers the research question(s)

- Encourage statisticians to collaborate outside the discipline and learn how to ask other people about their work
  - Group projects in interdisciplinary statistics courses

- **Must be reiterated throughout undergraduate, graduate, and early-career programs**

Conclusions

- **How to decrease experiment mortality rate:**
- DOX understanding and vocabulary
- A willingness to think outside the box (as Box did)
- An ability to practice communication, just as we practice our technical skills
- Using short, non-technical examples to illustrate potential issues
- **Let's keep our experiments alive!**

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." - Fisher (1938)

References

- Bliss, D. I. (1969),"Communication Between Biologists and Statisticians, A Case Study," *The American Statistician,* 2, 3, pp. 15-20.
- Box, G. E. P. (1976), "Science and Statistics", *Journal of the American Statistical Association*, 71, 356, pp. 791-799.
- —— (1979), "Robustness in the strategy of scientific model building", *Robustness in Statistics,* ed. Launer, R. L., and Wilkinson, G. N. *Academic Press,* pp. 201-235.
- Campbell, D. T., and Stanley, J. C. (1963), "Experimental and Quasi-Experimental Designs for Research." Belmont, CA: Wadsworth.
- Coleman, D. E., and Montgomery, D. C. (1993), "A Systematic Approach to Planning for a Designed Industrial Experiment," *Technometrics*, 35, pp. 1-12.
- Cox, G. (1950), Speech delivered to the Department of Agriculture.
- Fisher, R. A. (1938), Presidential Address. *Sankhyā: The Indian Journal of Statistics,* 4, 1, pp. 14-17.
- Hinkelman, K., and Kempthorne, O. (2008), *Design and Analysis of Experiments Volume 1: Introduction to Experimental Design* (2nd Ed). Hoboken, New Jersey: Wiley.
- Hunter, W. G. (1981), "The Real World is an Idea Whose Time Has Come," *The American Statistician*, Vol. 35, No. 2 pp. 72-76.
- Vining, G. (2013), "George Box: Scholar, Scientist, and Statistician," *Quality Engineering,* 25, 3 pp. 203-205.
- Zahn, D. A., and Isenberg, D. J. (1983), "Nonstatistical Aspects of Statistical Consulting" *The American Statistician*, 37, 4, 1, pp. 297-302.

Thank you!