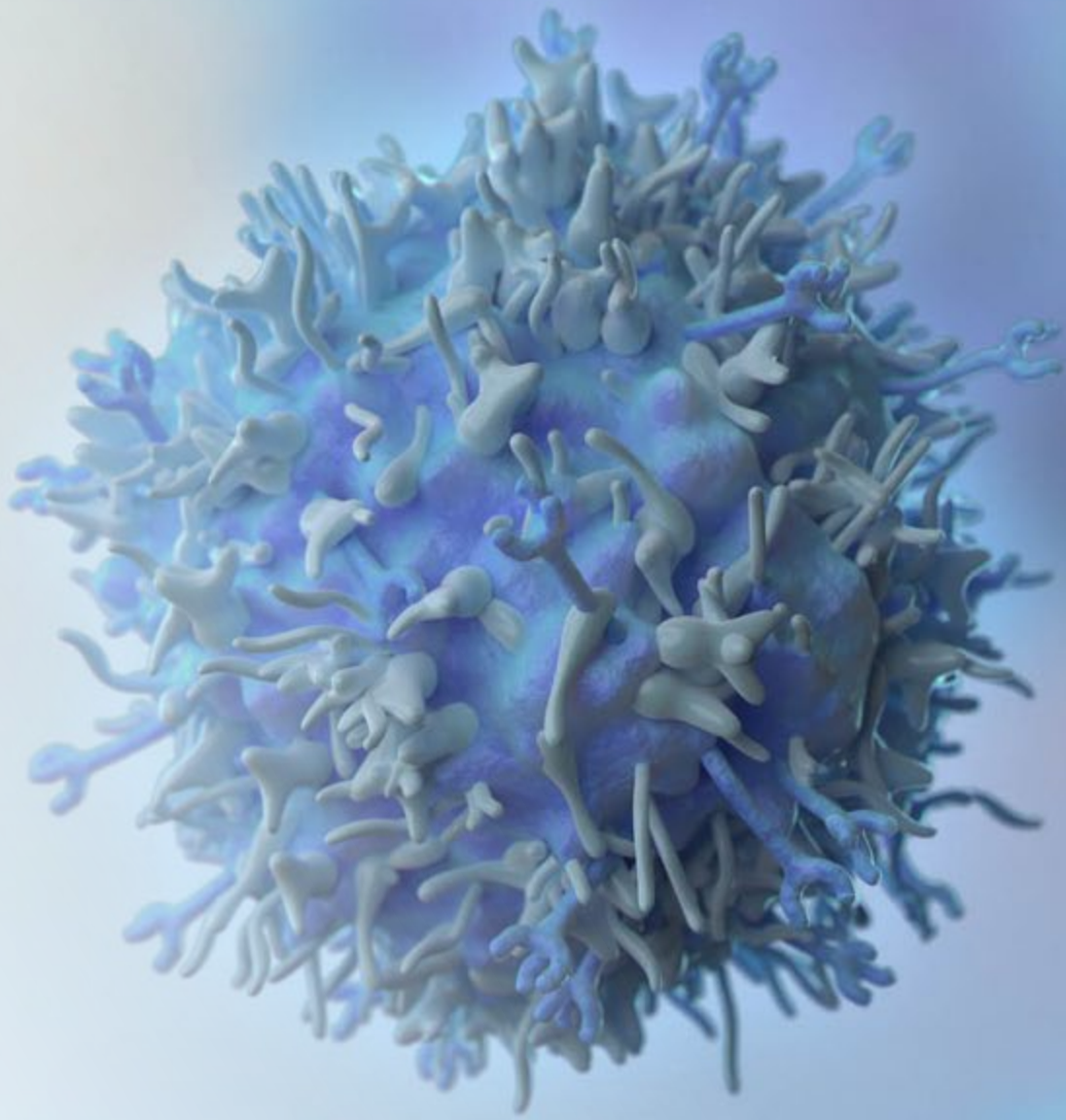


Statistical Science & Statistical Engineering

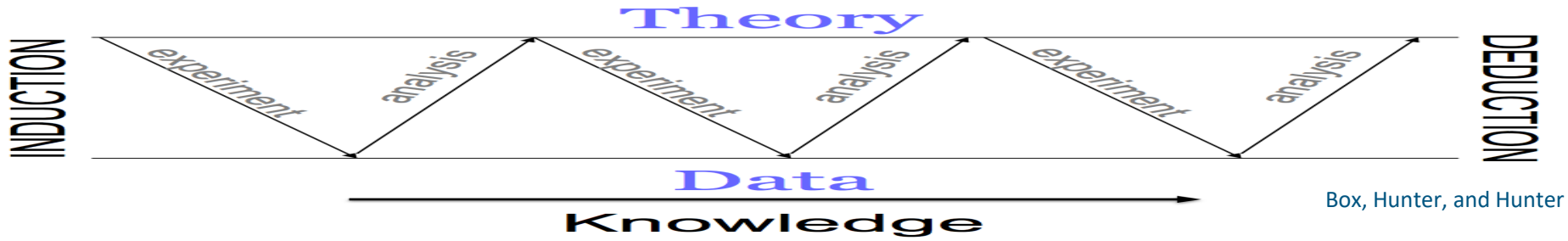
José Ramírez

IABS November 13, 2024

Science and Engineering

- "*Scientists discover the world that exists; engineers create the world that never was.*"

  —Theodore von Kármán



Box, Hunter, and Hunter

- "*Statistics used as a catalyst to engineering creation will, I believe, always result in the fastest and most economical progress.*"

  —Prof. George E.P. Box

- *Statistics as a catalyst for better science and engineering.*
  - *Knowledge generation via the interplay of deduction and induction*
  - *Scientific inference supported by statistical inference*

Statistical Science

- Statistics is *the* original data science.
- Sir David Cox argues,
  - "*That statistical thinking provides a unifying set of general ideas and specific methods relevant whenever appreciable natural variation is present. To be most fruitful these ideas should merge seamlessly with subject-matter considerations.*"(1)
- Unfortunately, the perception, and sometimes training, of statisticians is that we are in a service and consultant role, rather than as a scientist and collaborator.
- Sometimes statisticians are even asked to bless an analysis…
  - "*By contrast, there is sometimes a temptation to regard formal statistical analysis as a ritual to be added after the serious work has been done, a ritual to satisfy convention, referees, and regulatory agencies.*" (1)

(1) Cox, D.R. Statistical science: a grammar for research. *Eur J Epidemiol* 32, 465–471 (2017).

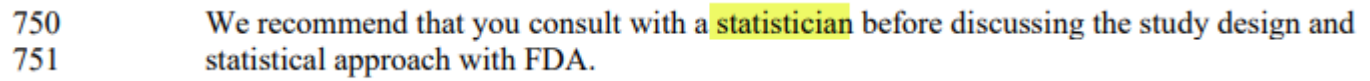
Statistical Science

This is not a new phenomena.

- In his 1938 Presidential Address to the First Indian Statistical Congress, Fisher lamented: "*To consult the statistician after an experiment is finished is often merely to ask him to conduct a postmortem examination. He can perhaps say what the experiment died of.*"

- In 1978 in his Presidential Address to the 138th meeting of the American Statistical Association, Prof. Box said:
  "*By invention of the concept of Experimental Design, Fisher promoted the statistician from a curator of dusty relics to a valued member of a scientific team, responsible for planning and taking part in the conduct of an investigation.*"

- Forty years later, then ASA President Lisa LaVange echoed Fisher's sentiment in her 2018 presidential address when she decried the misuse of statistics in pharmaceutical science, saying:
  "*Statisticians to the rescue, please!*"



FDA draft guidance (July2023) Manufacturing Changes and Comparability for Human Cellular and Gene Therapy Products.

Statistical Science & Art

- "*Mathematistry is characterized by development of theory for theory's sake, which since it seldom touches down with practice, has a tendency to redefine the problem rather than solve it.*" –Prof. George E.P. Box (1976)
- "*Much fine work in statistics involves minimal mathematics; some bad work in statistics gets by because of its apparent mathematical content.*"–Prof. David Cox (1981)
- "*What we want are neither mere theorem provers nor mere users of a cookbook. A proper balance of theory and practice is needed and, most important, statisticians must learn how to be good scientists; a talent which has to be acquired by experience and example.*"

  –Prof. George E.P. Box (1976)

- The art and science of making sense of data.
  - "*Studia prima la scienza, e poi seguita la pratica nata da essa scienza.*"
  - "*First study the science, then follow the practice born from that science.*"

–Leonardo Da Vinci

George E. P. Box. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.
David Cox (1981), Theory and general principle in statistics, JRSS(A), 144, pp. 289-297.

Scientific Inference Supported by Statistical Inference

Goodman, S. N. (2016), "Aligning Statistical and Scientific Reasoning," Science, 352, 1180–1181.

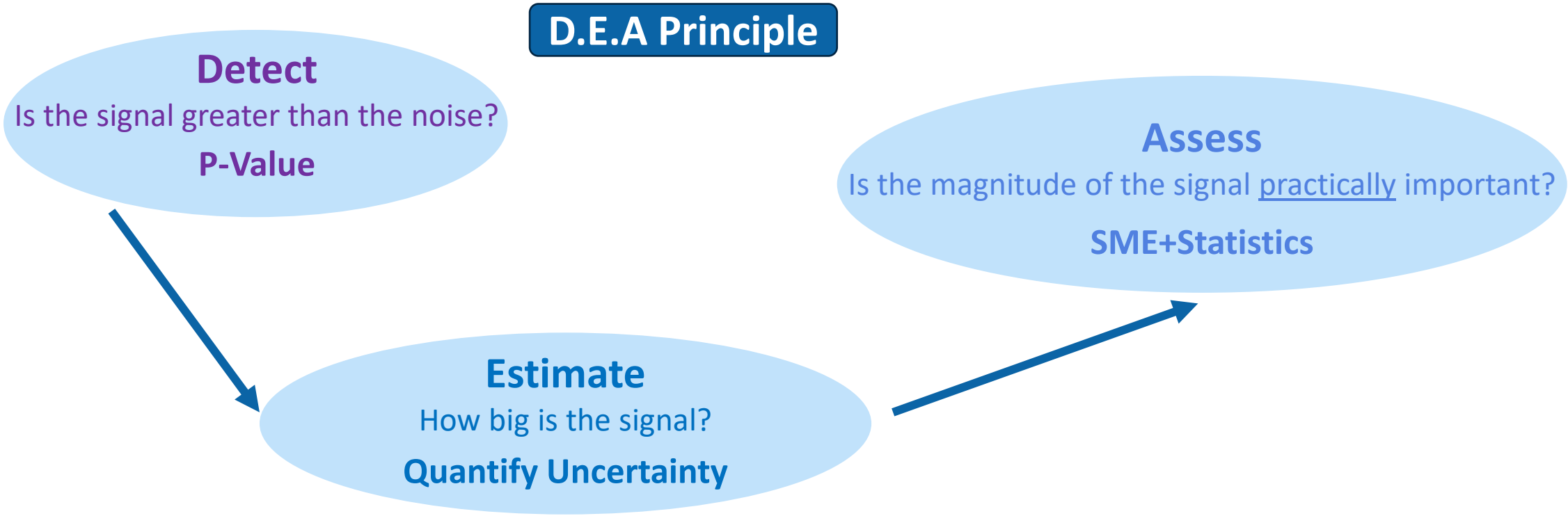
"What scientists want is a measure of the credibility of their conclusions, based on observed data."

- Sir Ronald Fisher, FRS
  - "*formalized an approach to inference involving P values and assessment of significance, based on the frequentist notion of probability, defined in terms of verifiable frequencies of repeatable events.*"
  - "*used 'significance' merely to indicate that an observation was worth following up, with refutation of the null hypothesis justified only if further experiments "rarely failed" to achieve significance.*"
- Jerzy Neyman and Egon Pearson
  - Error rates hypothesis testing: Null vs. Alternative — reject the "null" vs. accept the "alternative"
  - "*Once these error rates are set, scientific reasoning is effectively out of the picture.*"
  - "*Judgement ideally enters through customization of the alternative hypothesis and the error rates, contingent on the seriousness of each kind of error.*"
  - "*Without foundational justification, this created the illusion that quantitative inference could be automated.*"

Scientific Inference Supported by Statistical Inference

Statistical & Practical Significance

- *"No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he [examines] each particular case in the light of evidence and ideas."* (1)



(1) R. A. Fisher, Statistical Methods and Scientific Inference (Hafner, New York, ed. 1, 1956).

Statistical Engineering

- "*The study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results.*" (1)
- Core Principles of Statistical Engineering (Hoerl 2018)
  - Understanding of the problem context
  - Development of a problem-solving strategy
  - Consideration of the data pedigree
    - Data generating mechanism ⇨ *appropriate distribution*
    - Models that can be supported by the data at hand.
  - Integration of sound subject matter theory (domain knowledge)
    - Collaboration
  - Use of sequential approaches
    - Scientific method

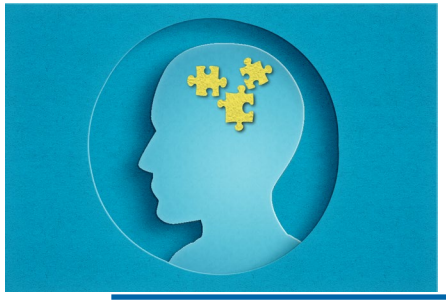
(1) Hoerl , R. W. , Snee , R. D. (2010). Closing the gap: Statistical engineering can bridge statistical thinking with methods and tools . Quality Progress , 43 (5): 52 – 53 .

Statistical Engineering

Long history…

- 1946: NBS (now NIST) Statistical Engineering Division
  - Develops and applies statistical and probabilistic methods and techniques supporting research in measurement science, technology, and the production of standard reference materials.
- 1945 – 1957: Statistical engineer William, Bill, Gore, founder of W.L. Gore and Associates.
  - In 1952 published *Statistical Methods for Chemical Experimentation.*
  - Prof. Stu Hunter said: “Bill Gore was a remarkable guy in his appreciation of the importance of statistics, which was rare in the 1950s. ” “I owe him for a lot of the experimental design work I did within chemical engineering.”
- 1954: Dr. W. Edwards Deming.
  - “*For all four groups of people [management, statistical administration, research, front-line workers], the statistical method is more than an array of techniques. **It is a mode of thought-sharpened thinking**. It helps anyone in the four groups, be he a machine operator or an executive, to make better decisions, and to do his work better, than he could do otherwise.*” (1)

(1) Deming, W.E. (1954). On the teaching of statistical principles and techniques to people in industry. Bulletin of the International Statistical Institute.



Statistical Thinking

A mode of thought-sharpened thinking…

Statistical thinking is a philosophy of learning and action based on three fundamental principles:

- All work occurs in a system of interconnected processes.
- All processes exhibit variation.
- Understanding, quantifying and reducing this variation is key to a successful business.

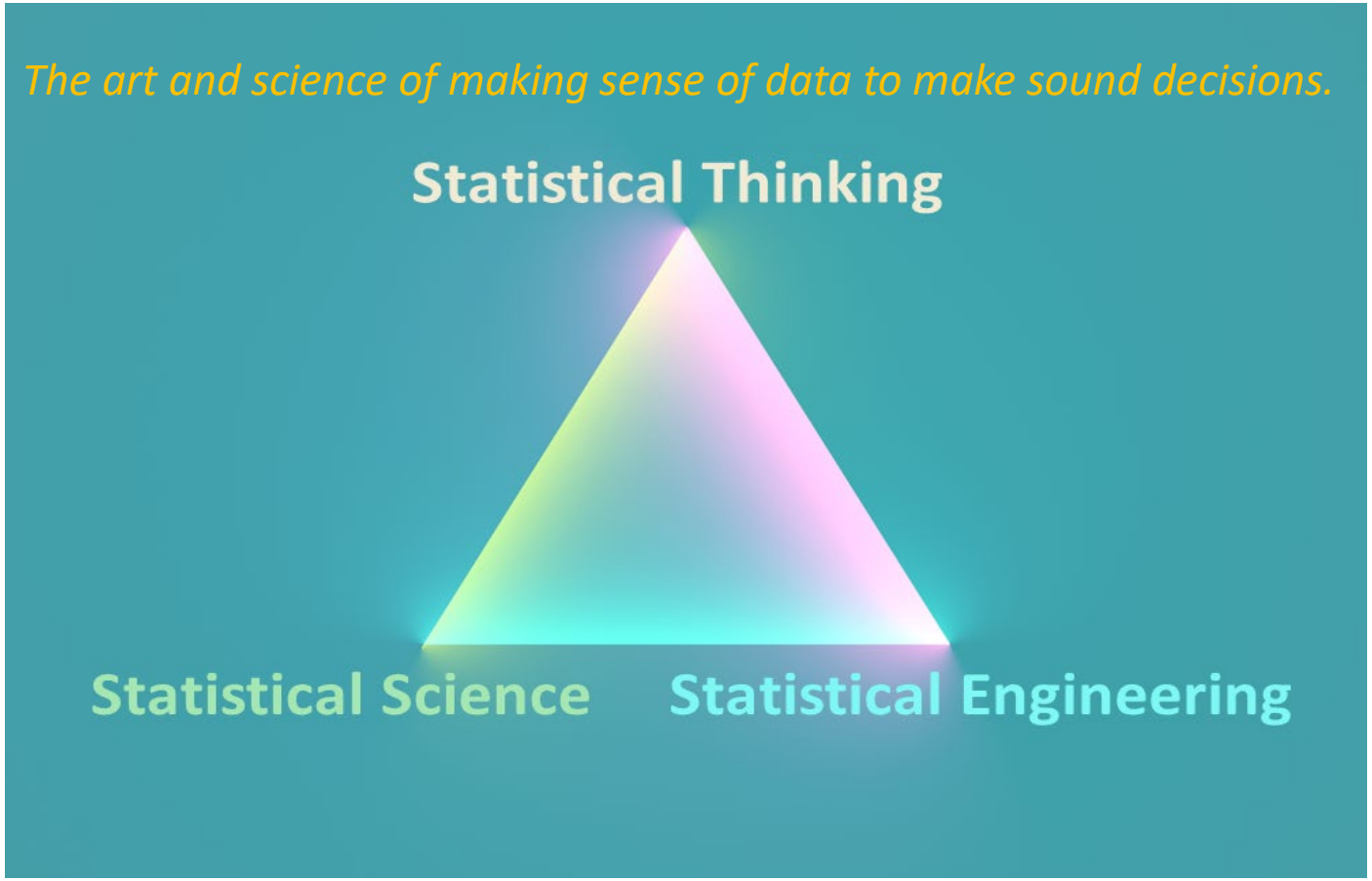
- "Statistical thinking provides a unifying set of general ideas and specific methods relevant whenever appreciable natural variation is present. To be most fruitful these ideas should merge seamlessly with subject-matter considerations." (1)
- Detect — Estimate — Assess
- Does the data contain enough information on the generating mechanism for it to be recovered?

—Dr. Fang Chen

(1) Cox, D.R. Statistical science: a grammar for research. *Eur J Epidemiol* 32, 465–471 (2017).

Kite
A GILEAD Company

Statistics as a Catalyst for Better Science and Engineering.



Inspirations

Many of the most successful and impactful statisticians started out in science or engineering.

Florence Nightingale: Passionate Statistician

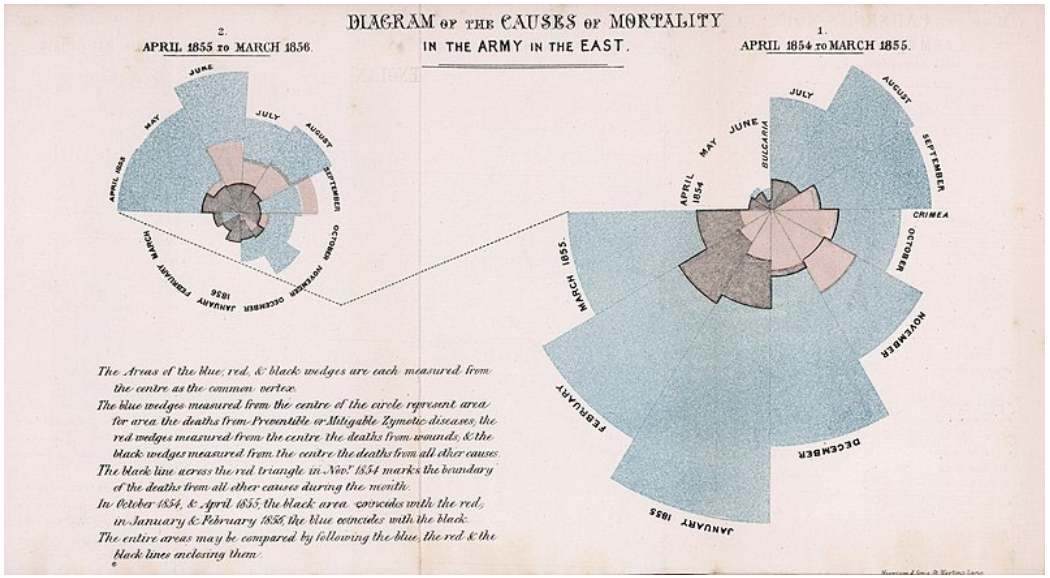
12 May 1820 – 13 August 1910

- English social reformer and statistician, and the founder of modern nursing.
- Elected the first female member of the Royal Statistical Society in 1859.
- Honorary member of the American Statistical Association in 1874.
- Graphical representation of social and medical data.

Kopf EW. (1978). Florence Nightingale as statistician. Res Nurs Health. Oct;1(3):93-102

- "She gave a succinct analysis of the data, pioneering the use of "coxcombs" (area charts) for the graphical presentation of mortality data. Government statistics would never be the same as a result, for her example of providing easy-to-visualize bar charts and pie charts was taken up in subsequent routine publications of Census and other data."

https://cwfn.uoguelph.ca/short-papers-excerpts/passionate-statistician/



Wikipedia Commons: Example of polar area diagram by Florence Nightingale (1820–1910).

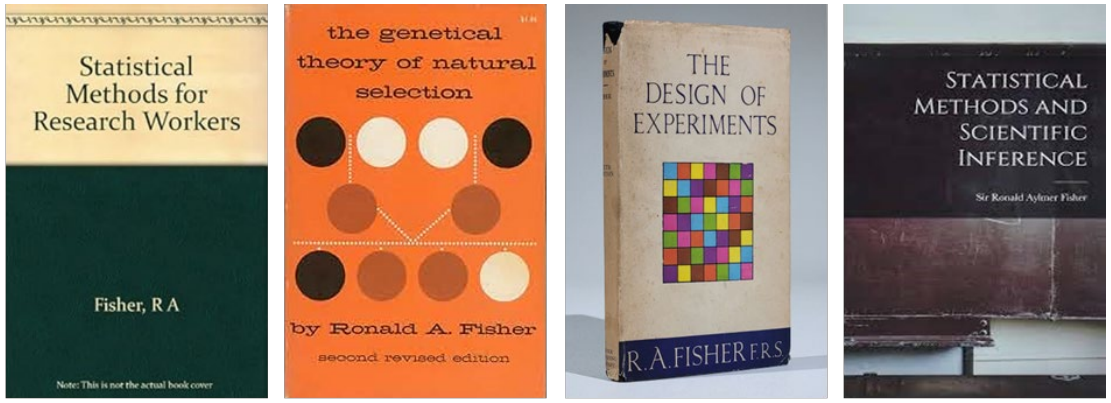
Sir Ronald Fisher, FRS

17 February 1890 – 29 July 1962

A statistician is an "expert in the process of reasoning".

A mode of thought-sharpened thinking…

- "The greatest biologist since Darwin."
- "Not only was he the most original and constructive of the architects of the neo-Darwinian synthesis, Fisher also was the father of modern statistics and experimental design."
  —Richard Dawkins
- Mathematical foundations of theoretical statistics
  - Maximum likelihood estimation (1912, 22 years old)
  - Consistency, sufficiency, efficiency, information.
- Analysis of variance
- The Lady Tasting Tea — Design of experiments
- Regard for the structure of the problem.
  - "the particular peculiarities of the actual body of data to be examined"



14 editions, many reprints.

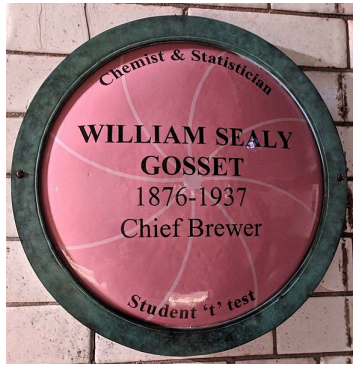
- Was he a Biologist/geneticist?
- Was he an applied statistician?
- Was he a mathematical statistician?
- Was he a data analyst?
- Was he a designer of investigations?

Kite
A GILEAD Company

William Sealy Gossett

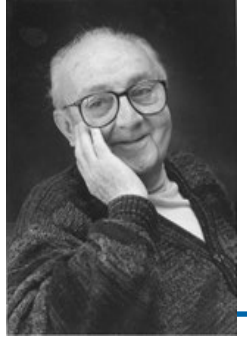
13 June 1876 – 16 October 1937

- a.k.a 'Student'
- Chemist, statistician, brewer.
- Chief Brewer at Guinness Brewing Company.
- In 1907 he published "On the Error of Counting with a Haemacytometer".
  - Poisson distribution
- In 1908 published "The Probable Error of a Mean"
  - the "t-test" paper!
- By asking the right questions, statistical thinking, he was able to derive the t distribution, changing the way we deal with small samples.



Wikimedia Commons

To many in the statistical world "Student" was regarded as a statistical advisor to Guinness's brewery, to others he appeared to be a brewer devoting his spare time to statistics.

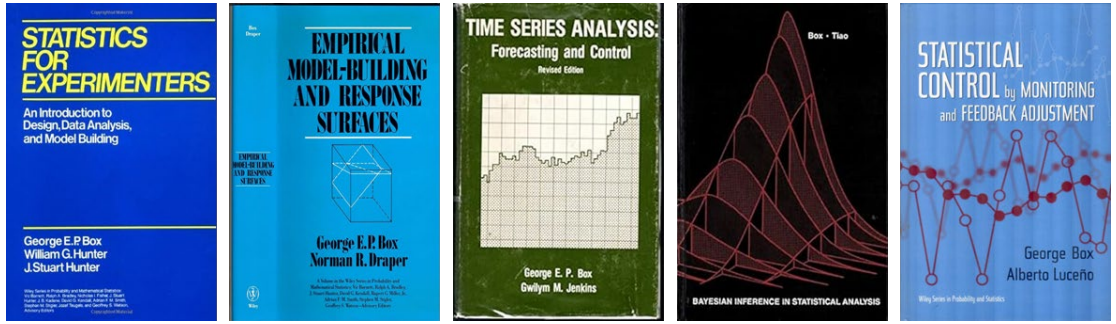


Prof. George E.P. Box, FRS

18 October 1919 – 28 March 2013

- Chemist: First paper when he was 19
- Worked at Imperial Chemical Industries (ICI)
- Director of the Statistical Techniques Research Group at Princeton University
- Founded the UW-Madison department of statistics
- President of ASA

- Box-Cox transformations
- ARIMA time series models (Box-Jenkins)
- Box-Tiao prior for variance components
- Box-Wetz goodness-of-prediction criterion
- Box-Behnken designs
- Central composite designs (CCD)



- Response Surface Methodology (RSM)
  - Box, G.E.P. and Wilson, K.B. (1951)
  - Increase efficiency of lab and full-scale experiments.
  - Adaption of Fisher's ideas to industrial experimentation.
  - A great example of scientific collaboration
  - "It was a success, … After this I could do no wrong."

**"To help them design *effective* experiments,
I had to *know the details* of the process and testing methods."**

Kite
A GILEAD Company

So many more…

- Prof. J Stuart Hunter
  - Electrical engineer, statistician, educator, author, and editor
    - *Statistics for Experimenters*
- Prof. William Hunter
  - Chemical engineer, statistician, educator, author
    - *Statistics for Experimenters*
- Prof. Norman Breslow
  - Statistician, medical researcher, educator, author, Honorary Fellow of the Royal Statistical Society.
    - *Statistical Methods in Cancer Research*

- -------

Thank you

The opinions presented are my own and not necessarily representative of Kite Pharma or Gilead.

13 November 2024