Fisher, Statistics, and Randomization

Qingyuan Zhao

University of Cambridge

April 22, 2022
Fisher in the 21st Century Conference
Gonville & Caius College, Cambridge

Fisher's work in this talk

Books

- R. A. Fisher, *Statistical methods for research workers*, (Oliver and Boyd, Edinburgh, 1925). [*SMRW*]
- R. A. Fisher, *The design of experiments*, (Oliver and Boyd, Edinburgh, 1935). [*DOE*]

Fisher's work in this talk

Articles

- R. A. Fisher, "Frequency distribution of the values of the correlation coeffients in samples from an indefinitely large population", Biometrika **10**, 507–521 (1915).
- R. A. Fisher, "On the mathematical foundations of theoretical statistics", Philosophical Transactions of the Royal Society of London. Series A **222**, 309–368 (1922).
- R. A. Fisher and W. A. Mackenzie, "Studies in crop variation. II. the manurial response of different potato varieties", The Journal of Agricultural Science **13**, 311–320 (1923).
- R. A. Fisher, "The arrangement of field experiments", Journal of the Ministry of Agriculture **33**, 503–513 (1926).
- R. A. Fisher, "Development of the theory of experimental design", in *Proceedings of the international statistical conferences*, Vol. 3 (Washington, 1947), pp. 434–439.
- R. A. Fisher, "Statistical methods in genetics", Heredity **6**, 1–12 (1952).
- R. A. Fisher, "Statistical methods and scientific induction", Journal of the Royal Statistical Society: Series B (Methodological) **17**, 69–78 (1955).
- R. A. Fisher, "Cancer and smoking", Nature **182**, 596–596 (1958).

Outline

Randomization

- Randomization is Fisher's first principle of experimental design, profoundly changing how modern science is being done.
- *SMRW* (1925) $\rightarrow$ Fisher (1926) $\rightarrow$ *DOE* (1935).
- Nowadays we take this idea for granted. But this was not the case even decades after *DOE*.
- For example, W S Gosset ("Student") repeatedly disagreed with Fisher.

  I do not expect to convince you but I do not agree with your controlled randomness. You would want a large lunatic asylum for the operators who are apt to make mistakes enough even at present.

  (Gosset proofreading SMRW, 1924)

The multifaceted role of randomization (Cox, 2009)[1]

1. Avoid selection and other biases and do so in a publically convincing way.
2. Provide a unified approach to the analysis of many standard designs.
3. Provide a basis for exact tests of significance and related interval estimates.

[1] D. R. Cox, "Randomization in the design of experiments", International Statistical Review **77**, 415–429 (2009).

Avoid bias

- Before Fisher's work, randomized experiments had been used in psychological experiments. This prehistory is nicely reviewed by Hacking (1988).[2]
- Peirce and Jastrow (1884)[3] is believed to be the first randomized experiment.[4]
- To test whether there is a threshold in our sensation of pressure, experimental subjects first experienced a weight of 1kg and then a second weight either slightly heavier or slightly lighter than the first, which was determined by well shuffled decks of cards.
- C Richet (1880s): Can we deteck weak powers of telepathy? He used a long sequence of trials in which an "agent" drew a playing card at random and concentrated upon it for a short time, after which a "reagent" guessed the suit of the card.
- J E Coover (starting from 1912): Not only randomize the card, but also whether the trial would be regular or control (in which the agent did not look at the card at all).

[2] I. Hacking, "Telepathy: origins of randomization in experimental design", Isis **79**, 427–451 (1988).
[3] C. S. Peirce and J. Jastrow, "On small differences in sensation", Memoirs of the National Academy of Sciences **3** (1884).
[4] S. M. Stigler, "Mathematical statistics in the early states", The Annals of Statistics **6**, 239–265 (1978).

Psychophysics research

- After Fisher' work, Bradford Hill argued forcefully for randomization in the context of clinical trials.
- The psychologists and Hill emphasized on how randomization eliminates personal idiosyncracies and confounding bias.
- Fisher surely knew and clearly expressed this point.

  Randomisation properly carried out ... relieves the experimenter from the anxiety of considering and estimating the magnitude of **the innumerable causes by which his data may be disturbed***.*

  (DOE, p. 44).

- Hacking (1988)'s conclusion: Fisher was well aware of psychophysics research, but Fisherian randomization involves a very different level of sophistication.

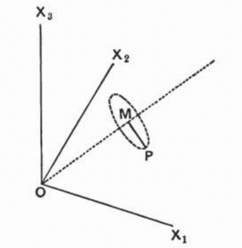
Small sample statistics

- Fisher started to correspond with Gosset (aka Student) in 1912 and worked on the statistics of small samples. He first gave a mathematical proof of Gosset's result.

  This result, although arrived at by empirical methods, was established almost beyond reasonable doubt...[but] the form **establishes itself instantly when the distribution of the sample is viewed geometrically**.

  (Fisher, 1915)

- Fisher's geometric argument involved representing n observations as a point P in the n-dimensional space. "For, given $\bar{x}$ and μ_1, P must lie on a sphere in $n-1$ dimensions."
- But the actual math leading to the t-distribution is not as straightforward as Fisher indicated.
- Fisher repeatedly used this intuition in the next decade.



Randomization and analysis of variance

- After being appointed as a statistician at Rothamsted Experimental Station in 1919, Fisher started to connect this geometric idea with randomization.

  If all the plots are undifferentiated, as if **the numbers had been mixed up and written down in random order**, *the average value of each of the two parts [sum of squares] is proportional to the number of degrees of freedom [, respectively].*

  (Fisher and Mackenzie, 1923)

  One way of making sure that a valid estimate of error will be obtained is to **arrange the plots deliberately at random**, *so that no distinction can creep in between pairs of plots treated alike and pairs treated differently; in such a case an estimate of error, derived in the usual way from the variations of sets of plots treated alike, may be applied to test the significance of the observed difference between the averages of plots treated differently.*

  (Fisher, 1926)

- It is clear that Fisher's emphasis is on correctly quantifying variance—the statistical error. This clearly distinguishes him from the pshychologists.

Fisher's geometric intuition

- The last statement can be directly justified by assuming normality and independence. But Fisher was convinced that those assumptions are not needed with randomization.

  His confidence in the result, however, depended on the geometric representation that was by then second nature to him. He could picture the distribution of n results as a pattern in n-dimensional space, and he could see that **randomization would produce a symmetry in that pattern rather like that produced by a kaleidoscope**, *and which approximated the required spherical symmetry available, in particular, from standard normal theory assumptions.*

  (Joan Fisher Box, 1980)[5]

[5] J. F. Box, "R. A. Fisher and the design of experiments, 1922-1926", The American Statistician **34**, 1–7 (1980).

A mystery

However, what Fisher precisely had in mind is still a mystery (at least to me).

- In the standard theory for analysis of variance, a key enabling result is that $\|\boldsymbol{P}\boldsymbol{Y}\|^2 \sim \chi^2_{\mathrm{tr}(\boldsymbol{P})}$, where $\boldsymbol{P}$ is a projection matrix, $\boldsymbol{Y} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$ is standard normal, and $\mathrm{tr}(\boldsymbol{P})$ is the trace of $\boldsymbol{P}$.
- Fisher must have seen that such a result is approximately true if $\boldsymbol{Y}$ is in some sense randomized. But some conditions on the randomization seem necessary.
- For example, $\mathbb{E}[\|\boldsymbol{P}\boldsymbol{Y}\|^2] = \mathrm{tr}\{\boldsymbol{P}\,\mathrm{Cov}(\boldsymbol{Y})\}$. So it seems necessary that $\mathrm{Cov}(\boldsymbol{Y}) \propto \boldsymbol{I}$ (identity matrix) in the randomization distribution. This would almost, but not precisely, hold for the permutation distribution.
- The connection between the normal theory distribution and the randomization distribution was not formally established until 1950s by requiring "additivity" (treatment effect is homogeneous).[6]

[6] O. Kempthorne, "The randomization theory of experimental inference", Journal of the American Statistical Association **50**, 946–967 (1955).

Randomization tests

There has ... been a tendancy for theoretical statisticians, not closely in touch with the requirements of experimental data, to stress the element of normality, ..., as though it were a serious limitation to the test applied. In these discussions it seems to have escaped recognition that **the physical act of randomisation, ..., affords the means of examining the wider hypothesis** *in which no normality of distribution is implied.*

(DOE, p. 45)

- Fisher then demonstrated how randomization can be used to construct a test using a pariwise randomized experiment.

TABLE 3

Differences in eighths of an inch between cross- and self-fertilised plants of the same pair

49	23	56
−67	28	24
8	41	75
16	14	60
6	29	−48

Their sum, taking account of the two negative signs which have actually occurred, is 314, and we may ask how many of the 2^{15} numbers, which may be formed by giving each component alternatively a positive and a negative sign, exceed this value. *(DOE, p. 45-46)*

Randomization p-value

TABLE 4

Number of combinations of differences, positive or negative, which exceed or fall short of the total observed

Number of negative values.	>314	= 314	<314	Total.
0 . . .	1	...	...	1
1 . . .	15	...	...	15
2 . . .	94	1	10	105
3 . . .	263	3	189	455
4 . . .	302	11	1,052	1,365
5 . . .	138	12	2,853	3,003
6 . . .	22	1	4,982	5,005
7 or more . .	...	...	22,819	22,819
Total . . .	835	28	31,905	32,768

- Fisher then found that the randomization p-value is $(835 + 28)/32768 \times 2 \approx 0.053$, which is very close to the p-value of a t-test.
- Notice that the validity of a randomization test entirely rests on the physical act of randomization. No modelling assumption is needed.

How Fisher's view shifted in his own words (Fisher, 1947)

- Fisher was brought in Rothamsted as a "data analyst/scientist" to examine a great number of files of records. And that was also his expectation.

  At the time, ... nothing was further from my thoughts, or from those of my contemporaries, than that the art of experimental design would ever come to be, as it now surely is, an integral part of the subject. The horizon was indeed filled with problems of ... estimation.

- Then he developed the theory of the maximum likelihood estimator.

  It was rigorously demonstrated that the amount of information extracted in the process of estimation could never exceed the quantity supplied by the data.

- Then what is the role of a statistician?

  This finding shifted the moral balance. His business became much less like that of a conjuror who is expected to work wonders, and more like that of a chemist who undertakes to assay the proportion of gold in a sample. **The weight of his responsibility was thrown back on to the process by which the data had come into existence.**

Blocking

Blocking is one of Fisher's principles for experimental design.

- The link to agricultural experiment is clear. Observations of the fertility of adjacent plots were known to be highly correlated.
- Intuitively, a more "balanced" design—assigning treatments equally often in each block—should be favoured.
- Fisher showed that blocking allows one to isolate and eliminate the variance due to block differences. This greatly improves the efficiency of the analysis of variance.

The Latin square design

A	B	C	D	E
E	A	B	C	D
D	E	A	B	C
C	D	E	A	B
B	C	D	E	A

(*DOE*, p. 77)

- Each row and column is a permutation of all the treatments. Two factors are blocked simultaneously.
- The Latin square design is extremely elegant, demonstrating Fisher's talent and contribution in combinatorics, geometry, and statistics.

Interplay between genetics and statistics

- Fisher's interests in genetics and statistics interacted strongly. The most notable were the the use of "variance" and the development of factorial experimental designs.[7]

  It is therefore desirable in analysing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quality the **Variance** *of the normal population to which it refers, and we may now ascribe to the constituent causes fractions or percentages of the total variance which they together produce.*

  (Fisher, 1918)

  And here I may mention a connection between our two subjects which seems not to be altogether accidental, namely that the **"factorial" method of experimentation***, now of lively concern so far afield as the psychologists or the industrial chemists, derives its structure, and its name, from the simultaneous inheritance of* **Mendelian factors***.*

  (Fisher, 1952)

[7] A nice review of this is given by W. Bodmer, "R A Fisher, statistician and geneticist extraordinary: a personal view", International Journal of Epidemiology **32**, 938–942 (2003).

Probability, genetics, and randomization

- The origin of Fisher's idea about randomization is still a mystery. Some believe it is related to Fisher's geometric intuition. Could it also be seeded in Fisher's understanding of Mendelian genetics?

  Mendel's theory was the first in natural science to **express its experimental predictions in terms of exact frequency probabilities**—*for example that the probability that a pea from a certain type of cross will be round rather than wrinkled is 3/4-not approximately 0.75, but 3/4. ... But the idea that nature itself could behave like a perfect gambling machine was revolutionary.*

  (Barnard, 1990)[8]

  Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. **The different genotypes possible from the same mating have been beautifully randomised by the meiotic process.** *A more perfect control of conditions is scarcely possible, than that of different genotypes appearing in the same litter.*

  (Fisher, 1952)

[8] G. Barnard, "Fisher: a retrospective", Chance **3**, 22–28 (1990).

Fisher and lung cancer

- Over 1950s, several studies revealed a large and robust association between smoking and lung cancer. In 1957, the *British Medical Journal* declared that "the most reasonable interpretation of this evidence is that the relationship is one of direct cause and effect."
- Fisher wrote several letters and articles to voice his objections. His main concern is the following.

  Such results suggest that **an error has been made, of an old kind, in arguing from correlation to causation***, and that the possibility should be explored that the different smoking classes, ..., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly to be assumed to be equivalent in their genotypic composition.*

  (Fisher, 1958)

- Fisher provided some data showing that monozygotic ("identical") twins are more likely than dizogytic ("fraternal") twins to have the same smoking habits. This suggests that smoking is (at least partly) heritable.

Criticisms of Fisher

- Fisher was criticized on several fronts, both at the time and later.

  [Fisher's analysis was] incomplete and [used] highly selected data (or no data but much speculation), with scant attempts to weigh the evidence or reveal the obvioius deficiencies in his data. Fisher soundes like a man with "an axe to grind". ... Fisher was ... unwilling to seriously examine the data and review all the evidence before him to try to reach a judicious conclusion.

  (Stolley, 1991)[9]

- Several things need to be considered when we evaluate this controversy.
  1. Fisher never denied the possibility that smoking could cause lung cancer.
  2. Fisher accepted invitation of the tobacco manufacturers standing committee to be their scientific consultant in 1956. Fisher believed receiving payment for so doing should not influence the nature of his advice.
  3. To date, there is still no "caught red-handed" evidence. I am only aware of two randomized clinical trials in 1970s-1980s on smoking cessation. None of them provided a definitive conclusion.

[9] P. D. Stolley, "When genius errs: R. A. Fisher and the lung cancer controversy", American Journal of Epidemiology **133**, 416–425 (1991).

Fisher and non-randomized studies

- I think Fisher's ultimate mistake in this controversy was that he did not recognize the need for non-randomized studies of causality in many fields.

  About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: **"Make your theories elaborate."** *The reply puzzled me at first, since by Occam's razor, the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible, and plan observational studies to discover whether each of these consequences is found to hold.*

  (Cochran, 1965)[10]

- However, Fisher did not provide us a concrete methodology.
- Formal analyses of observational studies did not start until the 1980s. And we still don't have a good framework to synthesize observational studies.

[10] W. G. Cochran and S. P. Chambers, "The planning of observational studies of human populations", Journal of the Royal Statistical Society. Series A (General) **128**, 234–266 (1965).

Start of the Fisher-Neyman controversy

- Neyman was a celebrated Polish statistician, best known for the Neyman-Pearson theory of hypothesis testing and confidence interval.
- Neyman and Fisher were on fairly good terms until Neyman published an article in 1935, suggesting the analysis of variance is invalid for analyzing Latin squares.[11] This sparked Fisher's legendary temper.

  Professor R. A. Fisher, in opening the discussion, said he had hoped that Dr. Neyman's paper would be on a subject with which the author was fully acquainted, and on which he could speak with authority. ... Since seeing the paper, he had come to the conclusion that Dr. Neyman had been somewhat unwise in his choice of topics.

  (Fisher in discussion of Neyman (1935))

[11] J. Neyman, "Statistical problems in agricultural experimentation", Supplement to the Journal of the Royal Statistical Society **2**, 107–180 (1935).

Fisher's attacks on Neyman

> *"And he said to me that he and I are in the same building and he is going to this meeting... And then I said, 'Do you mean that if I am here, I should just lecture using your book?' And then he gave an affirmative answer. ... And I said, 'Sorry, no. I cannot promise that.' And then he said, 'Well, if so, then from now on I shall oppose you in all my capacities.' ... Then he left. Banged the door."* *(C Reid, 1982)*[12]

- Fisher levelled up his attacks later WWII, suggesting Neyman was "importing from Eastern Europe his misconceptions as to the nature of scientific research".

Some contexts:

- A recent article found a crucial algebraic mistake in the calculations in Neyman (1935).[13] And Fisher's dicussion indeed hinted at the error.
- Fisher had a "history" with the Soviets.[14]

---

[12] C. Reid, *Neyman—from life*, (Springer-Verlag, New York, 1982), pp. ii+298.

[13] A. Sabbaghi and D. B. Rubin, "Comments on the Neyman-Fisher controversy and its consequences", Statistical Science **29**, 267–284 (2014).

[14] J. Lenhard, "Models and statistical inference: the controversy between Fisher and Neyman-Pearson", The British Journal for the Philosophy of Science **57**, 69–91 (2006).

Fisher and the Lysenkoism

Fisher was among a number of prominent Anglo-American geneticists who had been attacked by the Soviets in the early 1930s for their elaboration of "bourgeois genetics". Eighteen years later, rumours that Lysenko had a hand in the death of a scientific critic, Nicolai Vavilov, drew him to the Lysenko debate.

In 1948, Fisher, along with JBS Haldane and CD Darlington, did a BBC broadcast about the controversy. Fisher argued that Lysenko was using his political influence with Stalin to intimidate his scientific opponents. Where intimidation was insufficient, Fisher charged, Lysenko helped to see that 'many Russian geneticists' were 'put to death either with or without pre-treatment in a concentration camp.' Fisher devoted the bulk of this inflammatory speech to explaining Lysenko's actions as the product of a system where scientific judgements were made by political leaders.

(Marks, 2003)[15]

[15] H. M. Marks, "Rigorous uncertainty: why R A Fisher is important", International Journal of Epidemiology **32**, 932–937 (2003).

Fisher's ideological differences with the Neyman-Pearson-Wald theory

Russians are made familiar with the ideal that research in pure science can and should be geared to technological performance, in the comprehensive organized effort of a five-year plan for the nation. ... it may be safer ... to regard one's scientific work simply as a contributary element in a great machine, and **to conceal rather than to advertise the selfish and perhaps heretical aim of understanding for oneself the scientific situation.**

In the U.S. also the great importance of organized technology has I think made it easy to confuse the process appropriate for drawing correct conclusions, with those aimed rather at, let us say, speeding production, or saving money. There is therefore something to be gained by at least being able to **think of our scientific problems in a language distinct from that of technological efficiency***.*

(Fisher, 1955)

Fisher's objections to Neyman-Pearson

- Fisher (1955) criticized three concepts in the Neyman-Pearson theory.
    (i) "Repeated sampling from the same population",
    (ii) Errors of the "second kind",
    (iii) "Inductive behaviour".
- Both Fisher and Neyman considered the distinction between "inductive inference" and "inductive behaviour" to lie at the center of their disagreement.
- As a statistician, I find Fisher's first objection quite sobering.

  The postulate of randomness thus resolves itself into the question, 'Of what population is this a random sample?' which must frequently be asked by every practical statistician.

  (Fisher, 1922)
- In most of Fisher's practical work, he did not need to worry about this. because there is either artificial randomization (in agricultural experiments) or natural randomization (in genetics).

Topics discussed in this talk

- **Randomize to avoid hidden bias**: Fisher was not the first, but others struggled to give a mathematical justification.
- **Randomize to justify statistical inference**: Fisher saw this through his extraordinary geometric intuition and championed it through advocating "scientific induction".
- **Randomize to give a model-free test**: Fisher developed the basic ideas. Randomization is a bridge for him, connecting mathematics to the real world.
- **Design of experiments**: Fisher elevated statisticians' role in scientific collborations. The Latin square design is an epitome of his talent and insight.
- **Interplay between genetics and statistics**: Fisher's interdisciplinary knowledge and work made him an extraordinary scientist.
- **Smoking and lung cancer**: Usually seen as a stain in his legacy. He was wrong about the substance, but he was mainly criticizing the logic. Fisher did not recognize the need for non-randomized studies of causality.
- **Controversy with Neyman**: Started rather unfortunately. Even though he did have some good points, Fisher's ideological attacks are disturbing.

Some reflections

- Fisher's scientific ideas are fascinating, but his personality is complicated.
- Fisher's ideal of science is free and pure. Maybe that explains why he often had strong views.
- Fisher did not always work out his genius ideas fully. DeFinetti commented that Fisher "handles with mastery in individual problems but rather cavalierly in conceptual matters, thus exposing himself to clear and sometimes heavy criticism."
- We should all be respectful, but more personality in academic publications can be enlightening and entertaining.
- Design is a difficult concept to grasp for my generation of statisticians. "Repeated sampling from a distribution" is the default setup for mathematical statistics, but not for the practice.
- Randomization without Fisher? We might be talking about decades of delay.
- More reading demystified Fisher but made it ever more difficult to summarize Fisher.