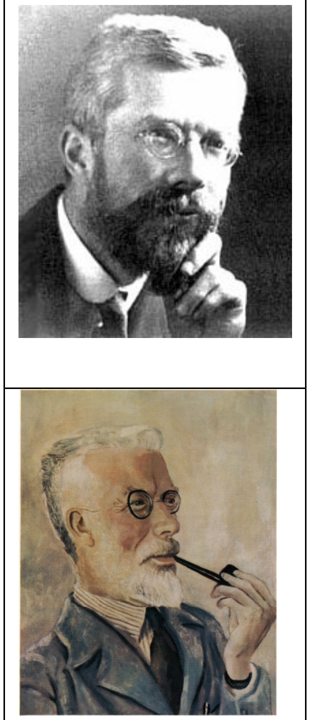
History of Statistics 8. Analysis of Variance and the Design of Experiments. R. A. Fisher (1890-1962)



R. A. Fisher

In the first decades of the twentieth century, the Englishman, Ronald Aylmer Fisher, who always published as **R. A. Fisher,** radically changed the use of statistics in research. He invented the technique called **Analysis of Variance** and founded the entire modern field called **Design of Experiments.** During his active years he was acknowledged as the greatest statistician in the world. He was knighted by Queen Elizabeth in 1952.

Fisher's graduate work in mathematics focused on statistics in the physical sciences, including astronomy. That's not as surprising as it might seem. Almost all of modern statistical theory is based on fundamentals originally developed for use in astronomy. The progression from astronomy to the life sciences is a major thread in the history of statistics. Two major areas that were integral to the widening application of statistics were agriculture and genetics, and Fisher made major contributions to both.

The same basic issue appears in all these areas of research: how to account for the variability in a set of observations. In astronomy the variability is caused mainly by *measurement erro*r, as when the position of a planet is obtained from different observers using different instruments of unequal reliability. It is assumed, for example, that a planet has a specific well-defined orbit; it's just that our observations are "off" for various reasons. In biology the variability in a set of observations is not the result of measurement difficulty but rather by uncountable impacts on the living organisms themselves. You can measure heights of plants in a plot accurately and precisely. The variability in the list of their heights is not the fault of the tools of the observers. The idea of a "true" height for such plants is a more abstract concept. We may be less interested in the average height than in what is causing some plants to be taller and some to be shorter. In astronomy, differences from the mean are called "errors." In Fisher's biological work they are called "**deviations.**"

In this quotation from an early essay by Fisher called *The Causes Of Human Variability* you can see his adaptation of the statistical methods of astronomy, particularly the decision to measure variability by the mean of *squared* deviations. What he calls "the law of errors," we call "the normal curve."

> **The practical method of measuring the variability of a population is extremely simple, although explanations of it are apt to appear complicated at a first reading. If a measurement of stature, for example, is made upon a sufficiently large sample -I,OOO to 10,000, let us say, of a population- the measurements are usually found to be grouped symmetrically about a mean value, the average stature of the sample. The deviations from this average follow very closely in their distribution what is known as the law of errors, that is to say the frequencies with which deviations of different magnitudes occur are related in the same way as the frequencies of errors of observation. Consequently the amount of variability may be measured, as errors of observation are habitually measured, by the mean of the squares of the deviations of different individuals from the mean, this mean square deviation being strictly comparable to the mean square error used in astronomy and in all physical science.**

Analysis of variance (ANOVA)

In scientific reasoning, “analysis” is the opposite of “synthesis.” Analysis implies breaking apart a large concept into smaller components, while synthesis means combining smaller parts into a larger whole. The goal of analysis of variance is to split the overall variability of a set of measurements into component parts. For example, if you grow a field of plants, their heights will have a certain variability and you may want to estimate how much of that is due to effect of sunshine and how much to the quality of the soil. A useful explanatory model needs to account for as much of the variability as possible, leaving only a little unexplained, what we now call “residual error.” It was one of Fisher’s great early contributions to show mathematically how the effects of various contributors could be assessed and compared. His method of analysis involves calculating the ratios of variances, and Fisher was able to derive the probabity curves for such ratios. The American statistician, George Snedecor, named these curves ***F* distributions** in Fisher’s honor.

Starting as early as 1912 when he was in his early twenties and continuing into the 1960s Fisher published well over a hundred mathematical papers that provided the theoretical basis for much statistical analysis. In contrast to such deeply mathematical papers, Fisher’s first text book, ***Statistical Methods for Research Workers**,* published in 1925, is not highly theoretical. The methods demonstrated in this book, such as ANOVA, are presented in the context of specific agricultural experiments. *Statistical Methods for Research Workers* quickly became the go-to guidebook for researchers world-wide, and is arguably the most influential statistics text written in the last 100 years. It went through 13 editions in his lifetime. The 14th was finished by a colleague after Fisher’s death and published in 1970. His intention to provide a straightforward guidebook is clear in the introduction.

> **The prime object of this book is to put into the hands of research workers, and especially of biologists, the means of applying statistical tests accurately to numerical data accumulated in their own laboratories or available in the literature. … The mathematical complexity of these** [distribution] **problems has made it seem undesirable to do more than (i.) to indicate the kind of problem in question, (ii.) to give numerical illustrations by which the whole process may be checked, (iii.) to provide numerical tables by means of which the tests may be made without the evaluation of complicated algebraical expressions . (page 16)**

Even though the book is not full of mathematical proofs, at times it can be difficult to follow. In a praiseful appreciation of Fisher’s work, another honored statistician, Maurice Kendall, wrote of *Statistical Methods for Research Workers*, “Somebody once said that no student should attempt to read it unless he had read it before.”

Here are some key statistical concepts that were first illustrated for a general research audience in *Statistical Methods.*

- **Hypothesis tests for statistical significance**

> **The idea of an infinite population distributed in a frequency distribution in respect of one or more characters is fundamental to all statistical work. From a limited experience, for example, of individuals of a species, or of the weather of a locality, we may obtain some idea of the infinite hypothetical population from which our sample is drawn, and so of the probable nature of future samples to which our conclusions are to be applied. If a second sample belies this expectation we infer that it is, in the language of statistics, drawn from a different population; that the treatment to which the second sample of organisms had been exposed did in fact make a material difference, or that the climate (or the methods of measuring it) had materially altered. Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first. (page 41, 14th ed)**

- ***p*-value as a measure of statistical significance.** Note that Fisher suggested .05 as “convenient,” and not a fixed rule.

  > **"The value for which P = .05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not." (page 44, 14th ed)**

- **Degrees of freedom** (The concept – but not the name - first appears in the astronomical work of Gauss.) Degrees of freedom is a crucial idea in *t*-tests and chi-square tests. The value that you use for *df* is one of the parameters of the curve you use to find *p*-values. When you change *df* you change the shape of the probability curve. Fisher showed how to calculate *df* to get the correct shape for your particular analysis. It is interesting to note that in Fisher’s theoretical papers degrees of freedom is a *geometrical* concept. It is the dimension of a vector space. He treated a set of *n* measurements of plant height, for example, as a vector in an *n*-dimensional space, just as you can think of a set of two observations, say, 4 and 7, as the pair (4,7) that you can represent as a vector on ordinary 2-dimensional graph paper. Fisher was able to make significant breakthroughs in statistical theory precisely because he was able to work in *n*-dimensional geometry.

- **Analysis of Variance**

  > [Some data] **may be more usefully and accurately treated by the analysis of variance, that is by the separation of the variance ascribable to one group of causes, from the variance ascribable to other groups. (page 213, 14th ed)**

Design of experiments

For fourteen years (from ages 29 to 43) Fisher was the primary statistician at the Rothamsted Experimental Station, one of the oldest agricultural research institutions in the world. He began working there in 1919, charged with making sense of the huge amount of data collected in field experiments dating back to 1843. Based on his agricultural research, Fisher published ***The Design of Experiments*** in 1935**.** In this textbook he described a now famous *randomized experiment* called ***The Lady Tasting Tea***. A woman (in real life, a colleague at Rothamsted, Muriel Bristol) claimed she could tell just by tasting whether the tea or the milk had been poured first into a cup. Fisher gives the design of an experiment to test the hypothesis that she was just guessing. This is the first use of the phrase “**null hypothesis.”**

The design called for the lady to taste (in random order) 4 cups of each type. He calculated that the chances of her getting all 8 cups identified correctly was 1/70. If she succeeded, he was prepared to reject the null hypothesis and grant she was not just lucky. In the textbook there are no data, but it has been reported that in real life, Muriel did get them all correct. (It has also been reported that Fisher never actually performed this experiment.) At the conclusion of the section where he describes this experiment he gives this now famous warning –

> **"...the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis." (page 16, 14th ed)**

Before Fisher, the preferred way to tease out the effect of an intervention was to compare an experimental group to an intervention group where the ONLY difference between them was the treatment given to the intervention group. Fisher worked mainly in agriculture – so his “groups” were plants. Perhaps one exposed to a fertilizer and one not. But suppose you thought that sunlight was also important. That would mean that you needed two separate experiments. One where the only difference between the groups was fertilizer, and one where the only difference between groups was amount of sunlight. As the number of potential causal factors increased, the number of experiments increased even more. This approach is clearly inefficient. Fisher suggested designing experiments where separate plots contained *different* ***combinations*** *of treatments.* Then he developed the statistical tools to analyze the data to reveal the relative contributions of each treatment, and also to evaluate any significant interactions among them.

Here is an example of how he might lay out an agricultural experiment with three treatments. Each block contains plants exposed to the treatments – the treatments being assigned randomly within the block. The blocks themselves provide four different soil conditions. This type of experiment, unsurprisingly called a *randomized block design,* is the fundamental design for comparing the effects of different treatments, not just in agriculture, but in virtually all fields of research. For example, in comparing three educational interventions, you could use four different instructors as "blocks," and each instructor could teach three sections of a course using different interventions in each section.

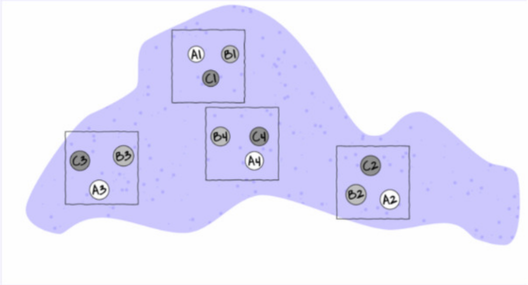


Figure rretrieved April 23, 2016, from http://www.flutterbys.com.au/stats/tut/tut9.3a.html

In 1933 Fisher left Rothamsted when Karl Pearson, perhaps England's most well-known statistician, retired as chair of the Applied Mathematics Department at University College in London. That department was split into a new department of Eugenics, with Fisher as head, and the department of Statistics, with Pearson's son Egon as head. In 1943 Fisher moved to Cambridge University as professor of Genetics until he retired in 1957 and moved to Australia. Ironically, this "greatest statistician" never held a position in a statistics department.

Exercise: Show that the probability of randomly guessing which of four cups of tea out of eight has the milk is 1/70.

Sources:

Kendall, M. G. "Ronald Aylmer Fisher, 1890-1962", *Biometrika*, (Jun., 1963)
Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, Freeman, 2001
Stigler, Stephan. *The Seven Pillars of Statistical Wisdom*, Harvard (2016)
Weisberg, Herbert. *Willful Ignorance,* Wiley*, 2014*
Lehmann, Erich. *Fisher, Neyman, and the Creation of Classical Statistics*, Springer; 2011
Fisher Images: