Fisher and the 5% Level

Stephen Stigler

Surely R. A. Fisher played a major role in the canonization of the 5% level as a criterion for statistical significance, although broader social factors were involved. Fisher needed tables for his 1925 book and, evidently, Karl Pearson would not permit the free reproduction of the *Biometrika* tables, so Fisher computed his own.

Fisher found it convenient to table values in the extremes for levels such as 10%, 5%, 2%, 1%—roughly halving the level with each step. One simple explanation for the format he selected lies in the fact that the book introduced "analysis of variance," or ANOVA. For most readers, this would be their first exposure to ANOVA, and Fisher needed a way to make the new test accessible—essentially the F-test, although he preferred to work in terms of z = log(F). The table here was entirely novel, requiring entry via two parameters: the numerator and denominator degrees of freedom (df).

It would have been impractical to provide a full table of the distribution for each pair of values: With the 10 levels of both dfs he wished to include, 100 tables would have been required if he gave the same level of detail he gave for his normal distribution table, or 10 tables if he gave the reduced level of detail that Gosset gave in his 1908 table for the t-distributions. So, Fisher initially settled on only giving one table for the 5% point. Once that was decided, it is not implausible that Fisher chose (in a book for practical workers) to make the other tables conform to that same simple format. This was not a huge task, and it had the bonus of casting all assessments of significance in the same accessible form.

The first edition (1925) of Fisher's book *Statistical Methods for Research Workers* had six tables:

I. and II. Tables of the inverse cumulative normal distribution (of z in terms of P, where $P = F(-z) + 1 - F(z) = \Pr\{|Z| > z\}$ and Z has a standard normal distribution). He gave this for P = .01 to .99 (increments of .01) and for P = .001, .0001, ..., .000000001.

III. Percent points y, where P= 1–F(y), for chi-square, df = 1, 2, ..., 30, and P = .99, .98, .95, .90, .80, .70, .50, .30, .20, .10, .05, .02, .01.

IV. Percent points for the t-distributions, df = 1, 2, ..., 30, ∞, and P = .9, .8, .7, .6, .5, .4, .3, .2, .1, .05, .02, .01.

V. Percent points for the correlation coefficient r, for n = 1 (1) 20 (5) 50 (10) 100 and for P = .1, .05, .02, .01. He also gave (as Table V (B)) the hyperbolic tangent transformation of r.

VI. Table VI gave only the P = .05 percent points for the distribution of z (the log of the F-statistic) by numerator df and denominator df, for df = 1, 2, 3, 4, 5, 6, 8, 12, 24, ∞. By the third edition (1930), he had added a table giving the 1% points and enlarged the range of denominator df considerably.

Note that only Fisher's Table VI strongly emphasized the 5% point. The others gave varying degrees of extended coverage, especially for the Normal, t, and chi-square distributions, where they gave a pretty good idea of each whole distribution.

Later editions of *Statistical Methods for Research Workers* (from the seventh of 1938) moved all the tables from the end of the book and interspersed them through the text. All these tables and more were given in Fisher and Frank Yates' book, *Statistical Tables for Biological, Agricultural and Medical Research*. There, the table for (essentially) the F-distribution was expanded to include a range of values from the 20% to 0.1% points.

My own view is that while Fisher's initial Table VI (but only that table) fixed attention at the 5% level (rather than, say, 6%, 10%, or 2%), that fixation is largely the result of a social process extending back well before Fisher. Even in the 19th century we find people such as Francis Edgeworth taking values "like" 5%—namely 1.5%, 3.25%, or 7%—as a criterion for how firm evidence should be before considering a matter seriously. Odds of about 20 to 1, then, seem to have been found a useful social compromise with the need to allow some uncertainty, a compromise between (say) .2 and .0001. That is, 5% is arbitrary (as Fisher knew well), but fulfils a general social purpose. People can accept 5% and achieve it in reasonable size samples, as well as have reasonable power to detect effect-sizes that are of interest. In my 1986 book, *The History of Statistics*, I speculate that the lack of such a moderate standard of certainty was among the factors that kept Jacob Bernoulli and Thomas Bayes from publishing. The use of Fisher's tables only served to make the choice more specific.

One may look to Fisher's table for the F-distribution and his use of percentage points as leading to subsequent abuses by others. Or, one may consider the formatting of his tables as a brilliant stroke of simplification that opened the arcane domain of statistical calculation to a world of experimenters and research workers who would begin to bring a statistical measure to their data analyses. There is some truth in both views, but they are inextricably related, and I tend to give more attention to the latter, while blaming Fisher's descendents for the former. After all, a perceptive 1919 article warning of the potential misuse of what we now call statistical significance by the psychologist Edwin G. Boring is ample evidence that the abuse predated Fisher.

Further Reading

Boring, Edwin G. (1919) "Mathematical vs. Scientific Significance." *Psychological Bulletin*, 16:335–338.

Fisher, Ronald A. (1925) *Statistical Methods for Research Workers* (first ed.), Edinburgh: Oliver & Boyd.

Fisher, Ronald A., and Yates, Francis (1938) *Statistical Tables for Biological, Agricultural and Medical Research* (first ed.), London: Oliver & Boyd.

Stigler, Stephen M. (1986) *The History of Statistics: The Measurement of Uncertainty Before 1900*, Cambridge, Mass.: Harvard University Press.