Original Article

George Udny Yule: Statistical Scientist

By

Richard H. Williams

Abstract: George Udny Yule was born February 18, 1871 in Beech Hill near Haddington, Scotland and died June 26, 1951 in Cambridge, England. He was a member of an established Scottish family composed of army officers, civil servants, scholars, and administrators. Both his father and a nephew were knighted. At the age of 16, he began the formal study of engineering at University College, London. George later moved to Bonn, Germany, where he studied under the famous scientist Heinrich Herz. A great influence in Yule's academic life was the well-known statistician, Karl Pearson, who lured him back to London, awarding him a directorship. George Udny Yule was prolific in journal and book publications and in activities related to the Royal Statistical Society, the highlight of his publications being perhaps the book, *Introduction to the Theory of Statistics*, which went through fourteen editions. F. Yates culminated his 1952 obituary of Yule by saying: "To summarize we may, I think, justly conclude that though Yule did not fully develop any completely new branches of statistical theory, he took the first steps in many directions which were later to prove fruitful lines for further progress… He can indeed rightly claim to be one of the pioneers of modern statistics" (Yates, 1952, p. 320).

Introduction

George Udny Yule initially studied to be a civil engineer, but through the influence of the famous statistician, Karl Pearson, he turned his attention to theoretical and inferential statistics. He encountered Pearson, who he viewed as an inspiring teacher, at University College, London.

Familial Background

George was born on February 18, 1871 in Scotland and died on June 26, 1951 in England. He came from a prestigious family. Both his father, who was also named George Udny Yule, and a nephew, were knighted. The Yule family had an impressive reputation for scholarship and administration.

In 1899, George Yule married May Winifred. Unfortunately, the marriage was not successful and was annulled in 1912, there having been no children (Yates, 1952).

Early Vocational Choice

In 1892, while studying engineering and physics, George travelled to Bonn, Germany to study under Heinrich Herz. This research bore fruit as he published four papers on research of electric waves. Nevertheless, he turned away from experimental physics and, at the age of 22, in 1893, returned to London, and was awarded a demonstratorship from Karl Pearson, who was at that time Professor of Applied Mathematics.

Is Normality Realistic?

Yule's first publication in statistics, "On the correlation of pauperism with proportion of out-relief," appeared in *Economic Journal* (1885). Among other things, the paper developed procedures for applying correlational techniques to two-dimensional contingency tables. It should be mentioned that Yule was always suspicious of correlational methods that depended on assumptions involving normality. He felt that important data in economics and many other fields were unlikely to satisfy assumptions of normality (Johnson & Kotz, 1997, pp.168-169). In his own words, "The only theory of correlation at present available for practical use is based on the normal law of frequency, but, unfortunately, this law is not valid in a great many cases which are both common and important. It does not hold good, to take examples from biology, for statistics of fertility in man, for measurements on flowers, or for weight measurements even on adults. In economic statistics, on the other hand, normal distributions appear to be highly exceptional; variation of wages, prices, valuations, pauperisms, and so forth, are always skew. In cases like these we have at present no means of measuring the correlation by one or more "correlation coefficients" such are afforded by the normal theory." (Yule, 1897, p.812)

Professional Organizations

At the rather young age of 25, Yule was elected a Fellow of the Royal Statistical Society and served with this organization under several capacities for the next 56 Years. He served as Honorary Secretary, which apparently was a somewhat thankless task, was President of the Society, and was awarded the prestigious Guy Medal in Gold. He was also active in a number of other professional societies.

Yule's Textbook

Very early in the twentieth century Udny Yule published a number of pa-

pers on regression and correlation. His work paved the way for multivariate developments by R. A. Fisher and Harold Hotelling. And while lecturing in statistics from 1902 to 1909 he produced notes which eventually led to the publication *Introduction to the Theory of Statistics* (1911).This book was readable and comprehensive and widely used---the only textbook of its kind, particularly used in economics and sociology. The first edition of this widely respected text appeared in 1911 and it went through 14 editions, the last three co-authored by the notable statistician, Maurice Kendall (Kendall, 1952). The book was translated into Czech, Polish, Spanish, and Portuguese.

Yule's Regression Approach

George Yule's major contributions in theoretical statistics dealt with correlation and regression, focusing especially on 2X2 contingency tables, time series, Mendelian genetics, and epidemiology. George tended to prefer a regression approach to these problems. It was Yule who first recognized that applications of the chi square statistic to contingency tables are dependent on degrees of freedom (df), where df = (r-1)(c-1), where r = #rows and c = #cols. He wasn't able to prove this mathematically, but demonstrated it through simulation. Later on the famous statistician, R. A. Fisher, provided the mathematical proof of Yule's contention. Unfortunately, Karl Pearson disagreed with Yule on this notion and hard feelings arose. Initially these two men were great friends, even spending vacations together.

Karl Pearson wasn't easy to get along with on academic matters. As George Udny Yule once wrote of Karl Pearson: "[His] temper in controversy was the more remarkable because there was no such temper in relation to anything but matters intellectual" (Reid, 1982, p. 115).

Yule was also adamant about possible causal relations among variables, indicating that there could always exist hidden variables that in reality are responsible for the supposed causal links. Fifty years later George Box made a systematic study of what he called unmeasured "lurking variables." Stigler (1986, p. 357) said, "With more than eighty years accumulated wisdom we might think Yule naively optimistic in his belief in the power of his methods. Yet in many ways [his approach] stands the test of time quite well. Indeed, an embarrassingly high percentage of modern work is not up to Yule's standard."

Charles Spearman developed the well-known rank-order correlation coefficient which is used to examine relations between variables which are ordinal. At the time, Karl Pearson attacked Spearman contending that the method was invalid. On a footnote of page three of Spearman's *British Journal of Psychology* paper titled "Correlation calculated from faulty data," (1910) Spearman wrote, "As regards the general question of ranks and measurements, it is pleasing to find Udny Yule totally disagrees with Pearson's adverse comments and, on the contrary, finds my proposal of ranks a very important step in the simplification of methods dealing with non-measurable characters."

This is yet another matter in which Yule and Pearson differed radically in their statistical views.

Classical Mental Test Theory

It is little known that George Udny Yule made some pioneer statements in the classical theory of mental tests. The fundamental equation in the theory is that $O = T + E$, where O = observed score, T = true score, and E = error of measurement. T and E cannot be observed directly, but the researcher is able to deal empirically with O. Starting with this equation, one can mathematically derive equations linkings means and variances, and expressions for test reliability. Methods designed to reduce E will improve reliability and therefore elevate test validity. For example, changing the format of a measuring device from true-false to multiple-choice is helpful. Legend has it that George Udny Yule communicated the above equation to Charles Spearman. Spearman used this idea to develop the basic ideas of the classical theory of mental tests (Williams, Zimmerman, Zumbo, & Ross, 2003), and these were completely worked out by Harold Gulliksen in his well-known book, *Theory of Mental Tests* (1950). Lumsden's negative view of the classical theory led him to say, "Yule's evil suggestion has yielded a meager fruit and has devoured the energy and skill of some of the finest minds in psychology." (Lumsden, 1976, p. 265)

Yule also "gave a new and much simpler proof of [Spearman's correction for attenuation], putting its validity beyond further cavil, but showing in the plainest light the assumptions on which it is based." (Spearman, 1910, p. 272) Errors of measurement tend to depress the correlation between two variables and Spearman's correction gives an estimate of the true correlation. For a further discussion of attenuation, see Zimmerman & Williams (1997). Yule's proof of the correction formula can be found in the appendix of Spearman's 1910 journal publication.

Yule's Correlation Coefficients for 2x2 Contingency Tables

There are a number of research terms in the literature named after George Udny Yule. Some of them are Yule Process, Yule Distribution, Yule Correlogram, Yule Autoregressive Series, Yule Coefficient of Colligation, and Yule Q. We will discuss the last named statistic, which is a measure of association applicable to 2x2 contingency tables (Kotz & Johnson, 1988). $Q = [f(1,1)f(2,2)-f(2,1)f(1,2)]/[f(1,1)f(2,2)+f(1,2)f(2,1)]$, where Q can range from -1 to $+1$ and 0 indicates no correlation. This statistic can also be expressed as $Q = (\text{alpha} - 1)/(\text{alpha} + 1)$, where alpha $= [f(1,1)/f(2,1)]/[f(1,2)/f(2,2)]$ is the odds ratio; $f(i,j)$ is the observed frequency lying in row i, column j.

Yule's Desire To Fly

George Udny Yule retired from his academic work at the age of 60, thinking the mathematics of his discipline had passed him by. He always had a desire to speed in cars, however, so he decided to learn to fly a plane. But then, being 60, he was an unacceptable risk as a pilot and thus he couldn't obtain insurance. So he bought his own plane and acquired a pilot's license. Although he received his

flying license in 1932, he began to have serious heart problems and became a quasi-invalid for the rest of his life. Ten years later Maurice Kendall and Yule were sitting together during a bombing raid and Yule began laughing and said that he was licensed to fly every one of those planes but hadn't any notion about the controls of any of them (Kendall, 1952, p. 157).

Statistical Aspects of Literary Style

As Yule neared retirement, he again took up the study of Latin and would sometimes write verses in Latin which would deal with topics such as the modern theory of small samples as well as literary topics. The philological interests and abilities of his Scottish lineage began to surface, and as he aged he turned to statistical studies of literature, focusing initially on sentence length. "His main work, however, related to the occurrence of words (principally nouns), and his researches found expression in his last book [Yule,1944] on *The Statistical Study of Literary Vocabulary* (Kendall, 1952, p. 158). He also studied errors accruing in copying manuscripts (Yule, 1946), the word distribution , showing how many words are used once, twice, thrice, etc., means and standard deviation of number of nouns, illustrative effect of a limited vocabulary, tables of Zipf, and general notions on sampling. Yule tried to answer questions such as the following: Did Thomas à Kempis really write that little volume which passes under the title of its first chapter, the *De Imitatione Christi*? Did Shakespeare write the plays that are generally attributed to him? Did St. Paul write the Epistle to the Ephesians? What is the probable chronological order of Plato's works? (Yule, 1944)

A Demonstration of Yule's Research Versatility

We now display the staggering set of substantive variables investigated by George Udny Yule in his many years of research. That is, studies dealing with pure statistical notions will be excluded from this list: anti-vaccination statistics; small-pox and vaccination; pauperism; duration of life and number of offspring; Mendel's Laws and their probable relations to intra-racial heredity; variation in the number of sepals in *anemone nemorosa*; marriage and birth rates; sex-ratios of births in registration districts; Census of Production Act; inheritance of quantitative compound characters on the basis of Mendel's Laws; social and economic statistics; distribution of deaths with age; family size and order of birth; sampling in Mendelian Ratios; statistics of British feeding trials and the starch equivalent theory; anti-typhoid and anti-cholera inoculations; crop production and price; some bacteriological methods employed in water analysis; the fall of birth rate; evolution and geographical distribution in plants and animals; a cross between yellow-wrinkled and green-rounded seeded peas; life table approximations; a mathematical theory of evolution; the growth of population and the factors which control it; time series; yield trials; sunspot numbers; note on the number of Jews in Germany; exposure to silica dust; occupational mortality; Notes and Records of the Royal Society; sentence length in prose; interference phenomena in electrical waves; harmonic analyser; and the statistical theory of accidents.

Concluding Comments

George Udny Yule came from a good Scottish stock, noted for its scholarship and administration. He contributed to regression and correlation theory, distribution theory, stochastic processes, statistics of literary vocabulary, and time series. He made longstanding contributions to the Royal Statistical Society and produced a noteworthy statistical textbook, *An Introduction to the Theory of Statistics*, which went through 14 editions and was published in five languages. Yule's contributions were also displayed in his friendly and humorous relations with his students and colleagues and the consulting he made available to all who approached him. "In character he was kindly, gentle, and genial. His wide knowledge of many subjects and his love of an apposite story made him the best of companions. His correspondence was a delightful mixture of shop, anecdote, and commentary on things in general, as the following extract will show: 'I began to keep a commonplace book many years ago, filled with rude things people say about statistics. I gave it up as they became less and less imagination'" (Kendall, 1952, p. 159).

————————————

Richard H. Williams, Ph.D., received a doctorate in educational psychology from Indiana University. He has taught courses in statistics, psychometrics, measurement and evaluation, and experimental design over the years at Appalachian State College, Boone, North Carolina, Indiana University School of Education, Bloomington, Indiana, East Carolina University, Greenville, North Carolina, and mostly at the University of Miami, Coral Gables, Florida. He served as a member of many doctoral dissertation committees. He has published in the *Journal of Modern Literature* and in two dozen other academic journals, mostly in education and psychology, often co-authoring publications with Professor Donald W. Zimmerman and sometimes with Professor Bruno D. Zumbo. He co-authored the book *Modern Elementary Statistics with Theoretical Supplement and BASIC Programming* with Zimmerman. He served as consulting editor of the *Journal of Experimental Education* for 15 years and was a reviewer for *The Mathematics Teacher, Applied Psychological Measurement, Perceptual and Motor Skills, Social Indicators Research,* and *Journal of Educational and Behavioral Statistics.* He has made presentations at many professional conferences. He also has placed short stories in *Naked Poetry, Blue Rose Bouquet, Drinking Stories, Demensions, Indite Circle, Another Night and Day Alliance,* and *The Harrow* and has recently had stories accepted for publication in *A Taste for Flesh* and *Avant Garde Times.* In addition to pursuing his research in statistics and measurement, he is currently studying Spanish, art, and creative writing.

References

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Johnson, N. L., & Kotz, S. (Eds.) (1997). George Udny Yule. In *Leading personalities in the statistical sciences from the seventeenth century to the present*, pp. 168-169, New York: Wiley.

Kendall, M. G. (1952). Obituary. George Udny Yule. *Journal of the Royal Statistical Society*. Series A (General), 115, Issue 1. 156-161.

Kotz, S., & Johnson, N. L. (Eds.) (1988). Yule, George Udny. *Encyclopedia of Statistical Sciences*, Vol. 9, pp. 664-666. New York: Wiley.

Lumsden, J. (1976). Test theory. *Annual Review of Psychology*, 27, pp. 251-280. Palo Alto, CA: Annual Reviews.

Reid, C. (1982). *Neyman…from life*. New York: Springer-Verlag.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900* (chapter 10-Pearson and Yule-pp. 326-361). Cambridge, MA: Harvard University Press.

Stuart, A., & Kendall, M. G. (Ed.). (1971). *Statistical papers of George Udny Yule.* New York: Hafner.

Williams, R. H., Zimmerman, D. W., Zumbo, B. D., & Ross, D. (2003). Charles Spearman: British Behavioral Scientist. *Human Nature Review*, 3, 114-118.

Yates, F. (1952). George Udny Yule. *Obituary Notices of Fellows of the Royal Statistical Society of London*, 8, 309-323.

Yule, G. U. (1895). On the correlates of total pauperism with proportion of total out-relief. *Economics Journal*, 5, 477-489.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, 60, 812-821.

Yule, G. U. (1911). *An introduction of the theory of statistics*. Charles Griffin & Co.: Ltd., London.

Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge University Press.

Yule, G. U. (1946). Cumulative sampling: a speculation as to what happens in copying manuscripts. *Journal of the Royal Statistical Society*, 109, 44-50.

Zimmerman, D. W., & Williams, R. H. (1997). Properties of the Spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement*, 21, 253-270.