CHANCE

ISSN: 0933-2480 (Print) 1867-2280 (Online) Journal homepage: www.tandfonline.com/journals/ucha20

Lionel Penrose's Statistical Consultant and Lessons from the Statistical "Sudoku" They Left Us

James A. Hanley & Supratik Roy

To cite this article: James A. Hanley & Supratik Roy (2025) Lionel Penrose's Statistical Consultant and Lessons from the Statistical "Sudoku" They Left Us, CHANCE, 38:3, 20-27, DOI: 10.1080/09332480.2025.2560277

To link to this article: https://doi.org/10.1080/09332480.2025.2560277

Published online: 03 Oct 2025.

Lionel Penrose's Statistical Consultant and Lessons from the Statistical "Sudoku" They Left Us

James A. Hanley and Supratik Roy

Lionel Penrose's 1933 study was one of the first to establish that "the father's age is not a significant factor in the risk of Down's syndrome, while the mother's age is to be regarded as very important." We identify the statistical consultant who helped Penrose improve on the prevailing regression methods for handling binary responses. Using a form of "statistical sudoku," we explain how to create complete data sets from the summary data Penrose worked with and shared, and we describe the surprises we got when we applied modern methods to these data sets.

The story is a mix of epidemiology, genetics, statistics, optimization, history, and statistical archaeology. It is a chance to reflect on how far we have come in statistical methods and computing in the last 90 years but also an important lesson on what we have forgotten or missed along the way.

The 1933 Analyses and the Until-Now-Unknown Statistical Consultant

Penrose, a physician, geneticist, and statistician, told us where his data came from:

> [...] 150 families of the human species containing Down's syndrome among the children. Every family included was visited personally and, among other things, the ages of the parents at the birth of each child was [sic] carefully recorded: miscarriages and all individuals in whom a diagnosis of normality or Down's syndrome could not be made with certainty were excluded. No obvious disparity was observed between the ages of the parental pairs, which were distributed in a manner resembling that found by pooling all married couples in the general population [...]

From what we can tell, Penrose did all of his analyses using three sets of marginal (2-D) frequencies, which he provided in a single table titled "Summary of Data."

It summarized the data on 154 cases of Down syndrome and 573 normal[1] children from these same 150 families. With just the last row label and the last column label altered to reflect modern-day terminology, we have reproduced that table in Figure 1, and on our webpage (*https://jhanley.biostat.mcgill.ca/StatisticalSudoku*). It contained the three (2-D) marginal distributions but stopped short of giving the full 3-D distribution. Penrose omitted the cell-specific distributions of *N* and *D*'s.

Penrose's challenge was how to deal with the tricky statistical issue that today we call confounding. The strong correlation of the parent's ages means that when the father's age or the mother's age is considered on its own, the probability that a child is affected by Down's syndrome will seem to be strongly related to that parent's age. So, how to disentangle the separate contributions?

Penrose's First Method of Analysis

Penrose's first approach was borrowed from an eminent American geneticist. A decade earlier, in research on essential factors in the occurrence of congenital defects in guinea pigs, Sewall Wright had used partial correlations to show that once one removed (eliminated) the effect of the mother's age, there was no effect of the father's age. "Attempt[ing] a similar treatment," Penrose calculated the three crude and two partial correlations shown in Table 1.

[1] Today, one might refer to children who are or are not "affected by Down Syndrome." We will sometimes use the original and sometimes the modern, and, to accommodate international readers, we will use both Down and Down's syndrome.

Summary of data

150 Families

N = Normal; D = Down's

Maternal Age (q)

Paternal age (p)	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	N	D
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
19	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
20	.	.	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
21	.	.	1	2	2	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.
22	.	.	.	1	4	3	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
23	.	.	2	1	1	2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
24	.	.	.	.	2	3	4	5	2	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16	2
25	.	.	.	1	2	.	4	6	7	4	4	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	27	3
26	.	.	.	3	2	2	3	2	3	5	1	3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	1
27	.	.	.	.	.	1	3	3	2	5	7	1	1	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	3
28	.	.	.	.	.	1	3	5	.	4	8	2	3	4	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	28	6
29	1	.	.	.	.	.	2	1	3	1	3	3	6	2	1	2	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	25	2
30	.	1	.	.	.	.	.	2	3	4	2	5	4	5	1	1	1	2	1	.	.	.	.	.	1	.	.	.	.	.	.	31	2
31	.	.	1	.	1	.	1	.	4	3	4	3	3	6	7	3	4	1	2	.	.	.	.	.	.	.	.	.	.	.	.	39	4
32	.	.	.	1	.	.	.	1	.	1	1	2	1	5	4	6	2	2	1	.	.	.	1	.	.	.	.	.	.	.	.	26	2
33	.	.	.	.	1	.	1	1	.	.	2	3	5	4	3	1	5	1	1	2	.	1	.	.	.	.	.	.	.	.	.	28	3
34	.	.	.	.	1	.	.	.	.	1	.	.	.	5	.	5	8	5	2	1	1	2	.	.	1	.	.	.	.	.	.	30	2
35	.	.	.	.	.	.	.	.	.	1	.	.	1	1	3	3	4	6	10	4	3	3	1	.	.	.	.	.	.	.	.	33	7
36	.	.	.	.	.	.	1	2	.	.	2	.	.	.	6	2	3	3	4	5	1	3	.	2	.	.	.	.	.	.	.	32	2
37	.	.	.	.	.	.	.	1	.	.	.	.	1	.	1	2	6	2	5	2	9	2	2	1	1	1	.	.	.	.	.	27	9
38	.	.	.	.	.	.	.	.	1	1	.	.	2	.	.	1	2	2	2	4	5	4	1	2	2	.	.	.	.	.	.	21	8
39	.	.	.	.	.	.	.	1	.	2	1	.	.	.	1	.	2	4	6	2	2	2	7	1	.	1	1	.	.	.	.	25	8
40	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	1	.	1	3	1	.	2	3	4	3	2	1	1	.	.	.	17	7
41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	2	.	.	3	4	4	1	1	.	.	.	.	10	8
42	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	2	.	.	2	6	2	3	3	2	.	1	.	1	.	9	15
43	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	1	3	1	3	3	4	3	2	.	.	.	11	11
44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	1	2	3	5	1	3	1	2	.	.	.	11	9
45	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	2	1	.	1	1	2	.	1	1	1	.	3	3	3	2	.	15	7
46	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1	1	1	2	1	1	.	4	4
47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	2	.	.	.	.	.	1	5	.	.	2	1	.	7	6
48	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	1	.	.	2	1	1	1	1	.	.	2	1	6	6
49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	1	.	2	4	1
50	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	1	1	.	.	.	.	4	.
51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	1	.	.	.	1	.	3	2
52	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	1	1	.	2	2
53	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	1	.	.	.	.	.	.	.	.	1	2
54	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	2	.	.	3	1
55	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	1	1
56	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	1	1
57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1	.
58	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	2
59	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.
N	1	2	5	8	13	11	22	31	25	31	31	22	30	36	32	32	35	27	35	21	21	22	21	16	10	9	6	5	8	4	3	573	
D	.	.	2	1	3	2	3	1	.	3	4	2	1	.	2	2	6	7	9	5	9	11	10	13	15	16	8	7	4	5	1		154

Figure 1. Reproduction of Penrose's Table I. The abbreviation "*N*" (for "Normal") used in the penultimate row, and column labels remain unchanged, but the letter "*D*" (abbreviation for "Down Syndrome") used in the ultimate row and column labels has been selected to reflect modern-day terminology. As is discussed in the next section, the last two rows contain two errors: One needs to change the 4:5 at maternal age 46 to a 3:6 split and the 3:1 at maternal age 47 to a 2:2 split.

The two "crude" correlations involving the binary indicator *D* (affected or not) are calculated from the two corresponding marginal frequencies (i.e., they are collapsed over (i.e., they ignore) the age of the other parent). They are both strong. Not surprisingly, the correlation between the parents' ages is very high: 0.829. But when Penrose, as social scientists like to say, "partialed out" the effect of the age of that other parent, the effect of the mother's age remained, but the effect of the father's age disappeared. As he went on to say in his paper, paternal age is not a significant factor, while maternal age is to be regarded as very important.

Today the idea of computing a Pearson correlation between a binary (0/1) *Y* and a quantitative *X* seems strange, but it was common back then. However, Penrose did anticipate readers' discomfort with correlations involving a binary all-or-none) variable, so he also used a crisper alternative analysis.

Penrose's Second Method of Analysis

As many people still do with data from "case-control" studies, Penrose first compared the mean parental ages of the Down syndrome and

Table 1—Correlations, Crude and Partial, Obtained from Data Summarized in Penrose's Table, where "Downs" is an Indicator (0/1) of Down's Syndrome

	Crude	
Downs, Father'sAge $+0.294 \pm 0.034$*		Downs, Mother'sAge $+0.362 \pm 0.032$
	Father'sAge, Mother'sAge $+0.829 \pm 0.012$	
	Partial	
Downs, Father'sAge eliminating Mother's Age -0.011 ± 0.04		Downs, Mother'sAge eliminating Father's Age $+0.221 \pm 0.04$
$\left[\frac{0.294\ -\ 0.362 \times 0.829}{\sqrt{(1-0.362^2)(1-0.829^2)}}\right]$		$\left[\frac{0.362\ -\ 0.294 \times 0.829}{\sqrt{(1-0.294^2)(1-0.829^2)}}\right]$

* the values following the $\pm$ signs are probable errors.

Table 2—Mean Ages of Parents of Down's Syndrome and Normal Children, with Expected/Predicted Means

	Father			Mother		
	Actual		Expected	Actual		Expected
Down's	39.383		39.471	37.253		35.712
Normal	33.830		33.803	31.249		31.680
Difference	5.553		5.668	6.004		4.032
		-0.115 (SE 0.392)			+1.972 (SE 0.341)	

$$\widehat{\text{Age of Father}} \Leftarrow 4.304 + 0.944 \times \text{Age of Mother}$$

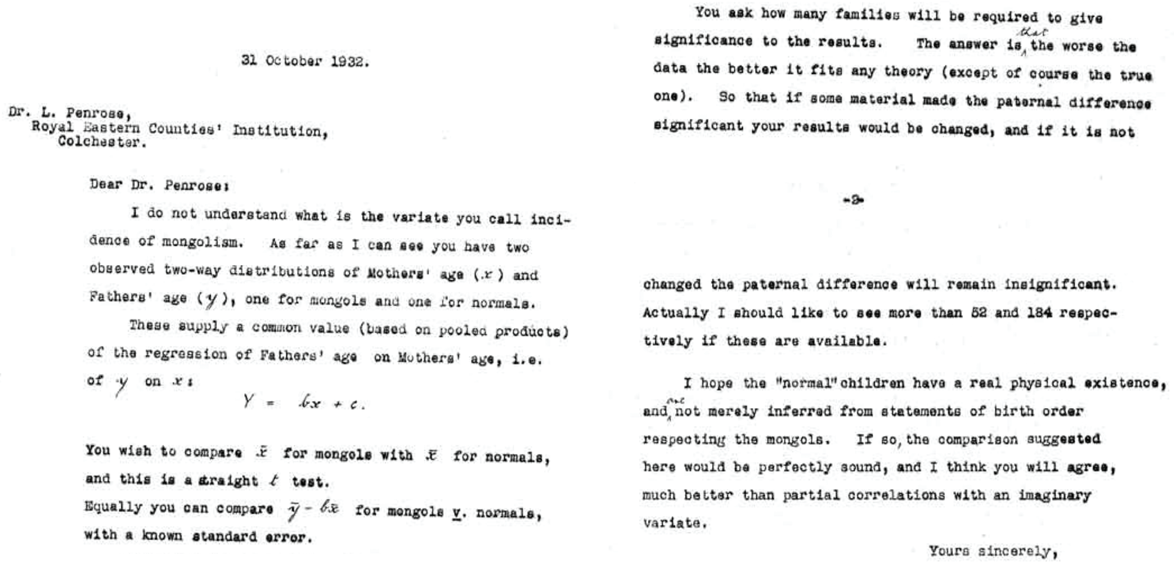
$$7.120 + 0.726 \times \text{Age of Father} \Rightarrow \widehat{\text{Age of Mother}}$$

non-Down syndrome children. In the two parent-specific comparisons (means 39.3 vs. 33.8 and 37.2 vs. 31.2, shown under "Actual" in Table 2), the parents of children with Down syndrome are decidedly older.

But the differences leave open whether there is just one culprit or two joint culprits. To settle this, Penrose used the two fitted equations to compute the expected (predicted) age of fathers, given the mothers' ages, and vice versa. When he substituted the observed mothers' ages (here 35.7 and 31.6) into the first equation to get the expected ages of the fathers, he found that they were very close to the observed ones. Conversely, when he proceeded in the other direction, and substituted the observed fathers' ages into the second equation, he got two expected mothers' ages that don't match very well with those he observed. The discrepancy of almost two years left Penrose in no doubt as to which parent's age drives the risk.

Penrose's Never-Mentioned Consultant, 1932

In 2007, in a fresh look at that second method, Oliver Penrose brought out some of the subtleties that underlay the apparent simplicity of the "beautiful method of analysis" in his father's 1933 paper. Oliver, born in 1929, still remembers that "when I was about four years old[,] I would sometimes venture into my father's study, to find him doing what I



31 October 1932.

Dr. L. Penrose,
Royal Eastern Counties' Institution,
Colchester.

Dear Dr. Penrose:

I do not understand what is the variate you call incidence of mongolism. As far as I can see you have two observed two-way distributions of Mothers' age (x) and Fathers' age (y), one for mongols and one for normals.

These supply a common value (based on pooled products) of the regression of Fathers' age on Mothers' age, i.e. of y on x:

$$Y = bx + c.$$

You wish to compare $\bar{x}$ for mongols with $\bar{x}$ for normals, and this is a straight t test.
Equally you can compare $\bar{y} - b\bar{x}$ for mongols v. normals, with a known standard error.

You ask how many families will be required to give significance to the results. The answer is that the worse the data the better it fits any theory (except of course the true one). So that if some material made the paternal difference significant your results would be changed, and if it is not

-2-

changed the paternal difference will remain insignificant. Actually I should like to see more than 52 and 184 respectively if these are available.

I hope the "normal" children have a real physical existence, and are not merely inferred from statements of birth order respecting the mongols. If so, the comparison suggested here would be perfectly sound, and I think you will agree, much better than partial correlations with an imaginary variate.

Yours sincerely,

Figure 2. Reply from R.A. Fisher to Penrose's October 29 letter. Some of their terminology is no longer used. Image courtesy of University of Adelaide.

described at the time as 'red and blue busywork.'"

Sixty years later he realized, on looking up the 1933 paper, that the work his father was doing with his red and blue pencils may well have been the computations for that paper. He described the collection of this data—"a *tour de force* of dedicated field work by his father and his two assistants working from the Royal Eastern Counties Institution in Colchester"—as a group effort: "But it was my father on his own who invented the method of analysis that induced this confusing jumble of data to give up its secret." Until author Hanley contacted him in 2014, Oliver Penrose was not aware of some relevant correspondence in the Penrose Papers at University College London (*https://archives.ucl.ac.uk/CalmView*). This archival material has since been digitized and put online, and it reveals a hitherto-unknown consultation regarding the statistical methods used in that paper.

On October 29, 1932, Penrose wrote to "Dr. Fisher," telling him he had been "working out some statistical results" in connection with mother's age [M] and father's age [F] and their relative affects in Down's syndrome. Using the term "incidence of Down's syndrome" for what today we would call the 0/1 indicator variable, he shared with Fisher the following three crude correlations calculated from the first 50 families available: M and incidence of Down's syndrome 0.44; F and incidence of Down's syndrome 0.34; M and F (for every child in each family) 0.73, along with the two partial correlations: between M and incidence of Down's syndrome (eliminating F) 0.30; and between F and incidence of Down's syndrome (eliminating M) 0.03. He then got to "[t]he question I want to ask you":

> How many families will be required to give significance to these results, which appear to show that maternal age is of importance but paternal age insignificant? I have available nearly 200 families and have every reason to suppose that similar figures will be obtained from the whole group to those I have found in the first 50 families. The question of significance is rather troublesome because I am not sure what the effect is of having several of the children in the same family and am doubtful whether the ordinary methods of finding the standard deviations of these coefficients are applicable here. If you think that about 180 families would give a significant result, it seems to me that it is an elegant method of demonstrating what has been rather a vexed question for some time.
>
> —from the Penrose archives at UCL.

The UCL archives do not contain Fisher's reply, but a carbon (and now digitized) copy of it is available online at the University of Adelaide. It bears out the present

Table 3—Some Developments in Regression-Type Models for Binary Responses, 1933–1973

Year	Author(s)	Development
1936	Fisher	The Use of Multiple Measurements in Taxonomic Problems. (LDA) Ann. Eugenics.
1952	Duncan, Rhodes	Multiple [Probit] Regression with a Quantal Response (Meeting Abstract)
1955	Berkson	Max. likelihood and min. χ^2 estimates of the logistic function (JASA)
1958	Cox	The regression analysis of binary sequences (JRSS-B)
1962	Cornfield	Joint dependence of risk of CHD on ... : a discriminant function analysis (Fed. Proc)
1966	Cox	Some procedures connected with the logistic qualitative response curve. (Essays)
1967	Cornfield, Kannel	Multivariate Analysis of Risk of Coronary Heart Disease in Framingham (JChronicDis)
1967	Walker, Duncan	Estimation of Probability of an Event as Function of Several Independent Variables (B'ka)
1972	Nelder, Wedderburn	Generalized Linear Models (JRSS-A)
1973	McFadden	Conditional logit analysis of qualitative choice behavior (Book Chapter, Economics)

authors' observation that often, it is not the original question (here to do with sample size and standard errors when one has correlated family data) one puts to the consultant that matters most. Instead, as in the reply shown in Figure 2, it is the entirely different way of looking at the data that an outside consultant suggests. A lot has been written about Fisher's disdain for some of Karl Pearson's work, and there is a hint of it in the first sentence of his reply. But this is immediately followed by a very insightful and transparent approach to settling the issue of whether the father's age matters: use the fathers' ages corrected for mothers' ages.

Fisher's next paragraph addressed the question of sample size, and then he ends with what all statisticians should also worry about, the quality of the data. And he can't resist taking a parting shot at the idea of using correlations using *imaginary* ("binary") variates—his adversary Pearson often used such variates in correlations.

2023: Reconstructed Data, Modern Analyses, and a Surprise

Listed in Table 3 are some of the statistical tools that have been developed since then to specifically handle all-or-none outcomes represented by "imaginary" (binary) variates. And so, a decade ago (notwithstanding the "outcome-based" selection of families and the fact that the data is not segregated by family) we asked students in a graduate data-analysis course to apply them to the data in Penrose's table.

However, they weren't able to convert these frequencies into a contemporary data frame with 727 rows (one per child) and three columns: mother's age, father's age, and the Bernoulli variate affected or not—or, equivalently, 325 rows, one per cell, of binomial splits—that would allow them to apply these modern day regression techniques. Their main complaint was that they were missing the cell-specific splits of the frequencies of affected and unaffected children. And, to add to their frustration, they noticed that Penrose's data-tables had at least one error that might not be resolvable, or might only after a considerable amount of trial and error.

In an effort to reconstruct the complete data, JH put Penrose's table online and offered a prize for splitting (unpacking) the frequencies. He didn't get much interest in this challenge—which he called a "Statistical Sudoku"—until the summer of 2023 when he visited his alma mater and met his co-author, SR.

Reconstruction

SR's first idea was to try Bayesian methods, which are often used to reconstruct joint distributions from marginal ones. But we effectively had only one sample or view. In computed tomography imaging, for example, one reconstructs the tissue densities in a cross-section using as many marginal views as there are rotations of the radiation source. The traditional techniques, such as those used in compressive sampling/sensing, need a sufficient number of samples to estimate the parametric part of the model. Moreover, any such reconstruction would have to be model-based.

Instead, he noted that the challenge was similar in structure to a network flow problem, known as the "optimal transport" problem in the optimization literature. Thus it can be seen as an integer linear programming problem. However, when he coded it, the `lp` function in the `lpSolve` package in `R` could not find a feasible solution. He traced the imbalance in the system back to errors in Penrose's table: whereas the row-specific totals for the different father's ages (at the rightmost margins of Table 1) summed to the overall numbers of 573 normals and 154 cases, the column-specific totals for the different mother's ages summed to 575 and 152. When he changed the 27:7 marginal

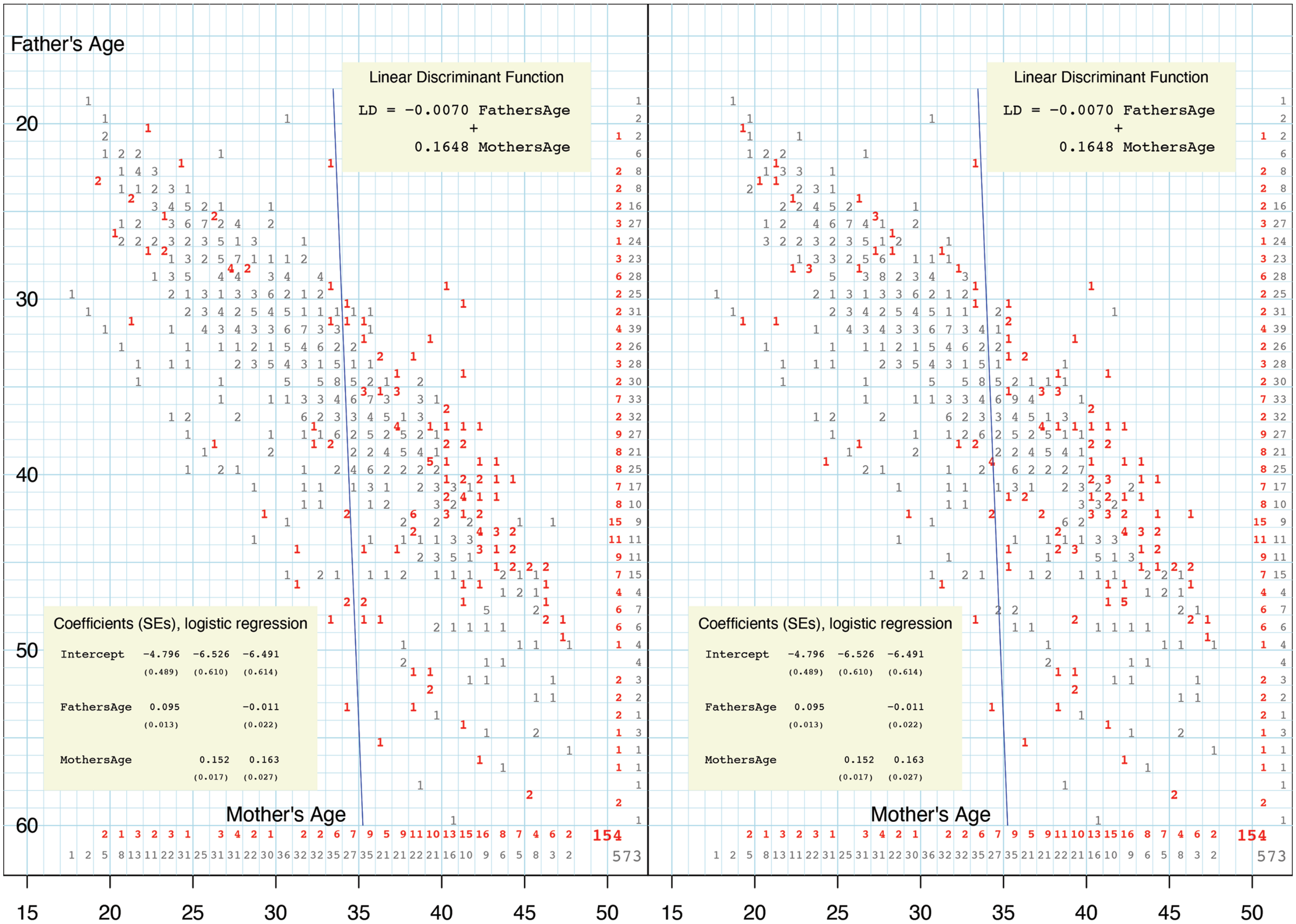


Figure 3. Two "Sudoku" solutions (one per panel), along with the linear discriminant and logistic regression functions fitted to the 3-D data corresponding to each of these two solutions. (See text for two minor pre-processing corrections to the printed Table 1 in Penrose, 1933.) Within each panel, the imputed numbers of Downs and normal children within each "parental-ages" cell are shown in red and gray respectively. The parent-specific marginal (2-D) distributions are shown along the bottom and rightmost margins, and the overall numbers in a slightly larger font at the bottom right corner. The boundary implied by each linear discriminant function is shown in blue.

frequencies in the mother-aged-34 category to a 25:9 split, the resulting mean maternal ages were lower than those reported in the text of Penrose's article. However, by proceeding systematically, we found that changing the 4:5 split at age 46 to a 3:6 one, and the 3:1 split at age 47 to a 2:2 one gave mean ages—and a crude correlation—that numerically matched those reported in the article. Thus we are confident that the errors were in the typesetting of the table and that Penrose did not use the typeset version in his computations.

Once these corrections were made, SR was able to use the `lp` function to find a feasible solution. Our website gives the technical details (*https://jhanley.biostat.mcgill.ca/StatisticalSudoku*).

Our First Reanalyses

Once he had obtained a solution (leftmost panel of Figure 3), we revisited Penrose's conclusion using two approaches developed since then. First we fitted a linear discriminant function of the parental ages—a technique that Fisher published in 1936, just three years after advising Penrose on how to avoid calculating correlations involving an "imaginary" *Y* variate. As can be seen from that panel, the weight (coefficient) for the mother's age is more than 20 times the one for the fathers' age, so the "dividing line" between Down's syndrome and normal children is almost vertical (i.e., almost entirely determined by the age of the mother).

Our second approach was to use a logistic regression model, a technique with several historical roots. A lesser-known root is the discriminant function that Cornfield derived from data collected in the Framingham Heart Study and then converted into a logistic-shape risk (probability) function of the scores. The panel inset shows that, just as in Penrose's analyses, of the two one-parent logistic regressions, the one involving the mother is the stronger one—but that in the model that considers them jointly, the coefficient for the mother's age dominates. Indeed, given the connection that Cornfield exploited, it is not surprising that the two coefficients in the logistic regression (0.162 and –0.011) are close to their counterparts in the linear discriminant function (0.163 and –0.070).

Multiple Solution Sets—and a Surprise

JH had contacted Oliver Penrose in 2014 and asked where we might find the raw data his father had used, but he had no suggestions beyond those locations where we had already looked. But he did point out that "[our] 'sudoku' problem appears to have multiple solutions in general" and provided a toy example. At that time, JH would have been happy to have even *one*. But now, just by altering the order in which the 398 constraints on the 325 unknowns appear in the (quite sparse) constraints matrix, we can use the `lp` software to generate as many separate solutions as we wish. Two such solutions are shown in Figure 3.

Before arriving at these solutions, JH had loosely likened the multiple possible solutions to the different data set "copies" used in multiple imputation when one has missing values. Thus, after fitting a multiple logistic regression to each of the (say *m*) solution-sets, he envisaged first computing the average variance-covariance of the *m* vectors of fitted coefficients, and then, employing "Rubin's Rule," adding to it the between-copy variance-covariance of the *m* vectors.

We were surprised to learn that—as is hinted at in Figure 3—every solution set yields exactly the same vector of fitted coefficients. The invariance also held when we fitted a linear discriminant function of the parental ages and when we computed the old-fashioned partial correlations that Wright and Penrose had used.

After some reflection, the reason for this invariance became clear to us. The marginal totals serve as *sufficient* statistics. Nowadays, tedious calculations are carried out by a machine rather than a human whose job-description was "computer;" and so the labor-saving property of *sufficient* statistics is less appreciated/used. Our website shows the form of the sufficient statistics in a more general form of a multiple logistic regression where the parental age (*M* and *F*) effects are represented as two separate splines; with no product terms involving both ages. In the special/simplest case where they are just linear, and additive, the sufficient statistics for the three parameters consist of three numbers: the number of cases of Down's syndrome: 154; the sum of the ages of their 154 mothers: 5,736 years; and their 154 fathers: 6,065 years. Equating the partial derivatives of the log-likelihood to zero results in three balancing equations that equate these three sufficient statistics to their three expected/fitted values. Finding this balance requires an iterative search, but within any given parental age "cell" one does not need to know how many of the children in the cell were affected and how many were not.

Penrose's calculations involved just one pass though the data. His computing equipment consisted of a "handle-powered desk calculating machine called the Brunsviga." He exploited the labor-saving "sufficiency principle" by processing the 325 cell frequencies to compute the marginal relationship between the parents' ages, the 31 × 2 frequencies to arrive at the marginal relationship between the mother's ages and Down's syndrome, and the 42 × 2 frequencies to arrive at the marginal relationship between the fathers' ages and Down's syndrome.

In the 1970s, when non-human computers were making the concept of sufficiency less relevant in applied statistics, Stephen Stigler, writing in *Biometrika*, advanced another reason: The concept "depends so strongly on the assumed form of the population distribution." Since then, the growth of computational methods in Bayesian statistics has largely restored the central role of sufficient statistics, but also brought model fit/adequacy to the forefront again. In that same cautious spirit that Stigler alludes to, the additional analyses shown on our website are limited to form(model)-free measures that align with Penrose's focus on null-hypothesis testing. We leave it to interested readers and teachers to explore model-based approaches, while recognizing the limited resolving power provided by the large age correlations and the relatively small numbers of events in the data set.

Discussion

Even though Penrose's "response" variable was binary, partial correlations were all that were available to him at the time. However, as the archives now document, it was not Penrose who "on his own invented [...] the [alternative] beautiful method of analysis" his son subsequently extolled. Rather, it was Ronald Fisher who showed him an alternative to calculating correlations involving an "imaginary" Y variate.

This was not an isolated fix. Just a few years later, Fisher generalized this method of exchanging the role of the response and predictor variables when he published the linear discriminant function for use in taxonomic problems.

It would take several decades before the fitting of generalized linear models for binary response data became practical. But by 2014, when we gave Penrose's data to the students in our data analysis course, the R software for fitting a logistic/probit regression model was widely and freely accessible. All they had to do was to create the data-frame. However, because Penrose seemed to "withhold" some of his data, they were unable to do so. They did not have the background training in mathematical statistics to go back to first principles and see their way around this difficulty. Their teachers could have, but they preferred a brute force approach.

At first, we were quite proud of our 2023 use of integer linear programming to provide thousands of data-frames with individual-level data, all in the blink of an eye. But we were soon brought back down to earth when we (or, rather, the computer!) fitted a binary regression model to these different solutions, all with the same marginal totals as Penrose worked with. No matter the within-cell binomial frequencies, the fitted coefficients did not change. Penrose's marginal totals acted as sufficient statistics.

In 1933, the concept of "sufficient statistics" had a very important application. It reduced human computational labor. In 2023, sufficient statistics (and exponential families) are concepts that are of practical use in intensive Bayesian computations, and in data-sharing policies that encourage data privacy. However, for the most part they are limited to courses in mathematical statistics and inference, where unfortunately the teaching of them consists mainly of exercises in calculus. Today's computer-processing speeds make it harder to motivate reducing the data to sufficient statistics. In line with the pleas of Nicholas Horton (Amherst College) and others to make such courses more relevant, today's teachers could encourage their students to appreciate the statistical importance of the concept. They might supplement these exercises by tasks involving reverse engineering or data reconstruction from public data sets, such as the one have described here.

Understanding of the phenomenon Penrose had been studying improved considerably after the late 1950s, when it was found to be a genetic disorder caused by the presence of all/part of a third copy of chromosome 21 ("trisomy 21"). The terminology has also improved: what we today call *Down* (the preferred North American term) or *Down's* syndrome goes back to the name of a 19th century English physician who first accurately described its phenotype. Penrose played a key role in having this terminology adopted.

As the *various* genotypes behind Down's syndrome became better understood, the overall risk function also came to be understood as a mix of a flat line, independent of mother's age, that dominates at younger ages, and a steeply rising curve that dominates at the older ages.

However, these improved understandings of the biology should not deter teachers from introducing the now-unpacked 1933 data set (or the more complete but also challenging one on maternal age and birth order from 1934) in their teaching of study design and regression models. As we have documented, the various approaches to analyzing this data set show how statistical knowledge and statistical methods evolve one study/step at a time. Nor should we be so confident as to think that we ever truly and fully understand the theory that underlies our modern statistical procedures. Some of the best understanding comes from examining our misunderstandings and our faulty intuition. C

Further Reading

Hanley J.A. 2025. Statistical sudoku. *https://jhanley.biostat.mcgill.ca/StatisticalSudoku*.

Harris H. 1973. Lionel Sharples Penrose. *Biographical Memoirs of Fellows of the Royal Society* 19:521–561.

Penrose L.S. 1933. The relative effects of paternal and maternal age in Down's syndrome. *Journal of Genetics* 27:219–224.

Penrose O. 2007. A beautiful method of analysis. In *Fifty years of human genetics: a Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser*, ed. O. Mayo and C. Leach, 434–451. Adelaide, Australia: Wakefield Press.

Wellcome Library. 2024. Birth order and Down's syndrome correspondence. *http://wellcomelibrary.org/player/b20222087*.

About the Authors

James A. Hanley is affiliated with the department of epidemiology, biostatistics and occupational health at McGill University.

Supratik Roy is affiliated with the department of statistics at the University College Cork.