strike action of junior doctors in their fight for a fair salary. Little or no help or leadership was given to a small group of junior physicians who were being discriminated against in attempting to maintain a strong ethical position.

It would seem that if the medical profession in Britain were to maintain the highest possible medical ethical principles in its practice, then it would have stronger bargaining power with the Government than it has at the present time. Vocal protests concerning ethics arise seemingly only when money is an issue, thus the Government and public may look, with some fair degree of scepticism, on the possible motives of the profession.

It is recognised that the National Health Service has the duty to provide equal and efficient care for all. Nevertheless, by its interference in the freedom to practise medicine by removing physicians who it considers are upsetting the efficient operation of the service, the state infringes on the principle of the total autonomy of the physician and the rights of women in general. The practice of medicine in Britain will suffer by this intervention, for a sameness of practice will develop which will stifle further thought and progress. The freedom that one group of women have gained through the introduction of the 1967 Abortion Act has been lost by another group of women by their inability in the future to be able to consult a physician whose

method of practice is based on a profound respect for life.

Fortunately, it has been possible to come to a country whose health care delivery system still maintains a high degree of ethical standards. The ability to practise a specialty is judged purely on qualifications and experience and not on religious background. One hopes most sincerely that this situation will continue.

What a sad world it is when a physician is unable to continue to practise because he has a profound respect for life. The declaration of Geneva was never more pertinent: "I will practise my profession with conscience and dignity. I will maintain the utmost respect for human life, from the time of conception; even under threat."⁵

References

- Walley, R L, *British Medical Journal*, 1972, 4, 234.
- Committee on the Working of the Abortion Act, *Report*, Cmnd 5538. London, HMSO, 1974.
- Haring, B, *Medical Ethics*. Notre Dame, Indiana, Fides Publications, 1973.
- Royal College of Obstetricians and Gynaecologists. *Evidence to the Enquiry into the Working of the Abortion Act (1967)*. London, 1973.
- General Assembly of the World Medical Association, *The Declaration of Geneva*. New York, WMA, 1948.

Statistics at Square One

IV—Standard deviation (concluded)

T D V SWINSCOW

British Medical Journal, 1976, 1, 1458-1459

Standard deviation from grouped data

Often the standard deviation must be calculated on such a large number of data that they need to be grouped for convenient handling. We have already met this necessity with the calculation of the mean. When Dr Green had only 15 readings for concentra-

tion of lead in the urine he could keep them separate in an array. But when he collected 140 readings he compiled a frequency distribution to make them manageable (Part II, table 2.1).

The calculation of the standard deviation from data grouped in a frequency distribution is similar to the calculation from ungrouped data, but one important point needs watching. As with the calculation of the mean from grouped data, the *midpoint* in each class is taken as the reading.

As an example, Dr Green's data are set out in table 4.1, which is simply an extension of table 2.1, with two additional columns. Just as in calculating the standard deviation from ungrouped data (Part III) so here we do not need to measure the actual differences between the observations and their mean. Instead we make use of the identity that we used in Part III:

(sum of squares of differences between observations and mean) equals (sum of squares of observations) minus (sum of observations)² ÷ number of observations, or

$$\sum(x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}$$

It is important to remember that the "observations" in this case are the midpoints in the frequency distribution.

The sum of the observations, $\sum x$, was calculated in table 2.1 to find the mean and is now repeated in table 4.1, col (4), where the midpoint of each class of lead concentration is multiplied by the number of children in the class. Then, just as when calculating the standard deviation from the ungrouped series we squared each of the observations in turn, so now we take each of the midpoints of the observation classes shown in col (3) and square them, as in col (5). These correspond to the squares of the observations in the ungrouped series. But since we have

TABLE 4.1—Calculation of standard deviation from grouped data (continuous variable)

(1) Lead concentration μmol/24h	(2) Number of children	(3) Midpoints of col (1)	(4) Col (2) × col (3)	(5) Midpoints squared	(6) Col (2) × col (5)
0 —	2	0.2	0.4	0.04	0.08
0.4 —	7	0.6	4.2	0.36	2.52
0.8 —	10	1.0	10.2	1.0	10.0
1.2 —	16	1.4	22.4	1.96	31.36
1.6 —	23	1.8	41.4	3.24	74.52
2.0 —	28	2.2	61.6	4.84	135.52
2.4 —	19	2.6	49.4	6.76	128.44
2.8 —	16	3.0	48.0	9.0	144.0
3.2 —	11	3.4	37.4	11.56	127.16
3.6 —	7	3.8	26.6	14.44	101.08
4.0 —	1	4.2	4.2	17.64	17.64
4.4 —					
Total	140		305.6		772.32

British Medical Journal

T D V SWINSCOW, MSC, MB, deputy editor

several children in each class, shown in col (2), each squared midpoint in col (5) must be multiplied by the corresponding number of children. The result is shown in col (6), and the sum of the squares is at the foot of the column, namely, 772.32.

Continuing the procedure as for ungrouped data, we take the sum of the squares of the observations and subtract from it the sum of the observations squared divided by the number of observations. Dr Green's figures now look like this:

$$772.32 - \frac{305.6^2}{140} \quad \text{This equals } 105.239.$$

To obtain the variance we divide by $n-1$, which is 139, and get: 0.76. The square root of that gives us the standard deviation = 0.87.

The procedure may be summarised as follows:

Set out the classes of observations in a column. Put the numbers in each class against the corresponding numbers in it. Set out the class midpoints in a column (table 4.1). Multiply the midpoint in each class by the number in the class and add them Σx

Square this total $(\Sigma x)^2$

Divide by the number of observations in all classes $\frac{(\Sigma x)^2}{n}$ (1)

Square each class midpoint x^2

Multiply each squared midpoint by the number of observations in the corresponding class and add them Σx^2 (2)

Subtract (1) from (2) $\Sigma x^2 - \frac{(\Sigma x)^2}{n}$

Divide by the number of observations minus 1 $\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}$

Take the square root $\sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$

This is the standard deviation.

When using a calculator it is probably best to set out at least the first three columns of table 4.1—namely, the class intervals, the numbers in each class, and the midpoints of each class. The multiplication of the midpoints by the number in each class can then proceed, the resulting products being accumulated in the memory and brought out on to the display screen when complete, to give the total at the foot of column (4).

The operations in columns (5) and (6) need not be written down. They can be carried out successively for each class, and summed in the memory to give the total at the foot of column (6). For example, in the first row of table 4.1, 0.2 is squared (=0.04 in col (5)), the square is multiplied by 2 (col (2)), and the product entered in the memory. The products derived thus from each row are accumulated in the memory, and the sum is finally obtained as 772.32, but left in the memory.

The sum of the observations is squared, 305.6², and divided by the number of observations, 140, to give 667.08. This is then subtracted from the memory and the result brought on to the display screen = 105.239. This is divided by $n-1$ (=139), making 0.757, and the square root taken, to give 0.87.

Continuous and discrete variables

The readings obtained of urinary concentration of lead are described as "continuous" in contrast to "discrete" or "discontinuous". This is because each reading can be any value within the possible range, the value depending on the amount of lead present and the sensitivity of the apparatus measuring it. A discrete variable is a numerical value attached to an event or finding that stands on its own and cannot take intermediate values. For example, the amount of fluid measured from a

syringe can be any amount within the capacity of the syringe, and so is a continuous variable. But the number of pills taken from a bottle must be a whole number of pills, a discrete variable. Though this may seem to labour the obvious, the difference needs to be kept specially clear in the statistical treatment of data.

In tables 2.1 and 4.1 the readings are grouped in classes. Each class represents a range covering 0.4 $\mu\text{mol}/24 \text{ h}$ of a continuous variable. In the calculations made on the data the midpoints of each class were taken to represent the whole range of the class. But when discrete data are used, the procedure is slightly different and simpler.

For example, as well as studying the lead concentration in the urine of 140 children Dr Green asked how often each of them had been examined by a doctor at home or in his surgery during 1975. After collecting this information he tabulated the data shown in table 4.2, cols (1) and (2). He went on to calculate the mean number of visits and the standard deviation.

Clearly there can be no midpoint in the classes listed in col (1). Therefore these numbers themselves are multiplied by the numbers of children in each class to produce the figures in col (3). The total number of visits, 455, is then divided by the number of children, 140, to give the mean number of visits, 3.25.

Likewise the standard deviation would be calculated by the method shown in table 4.1 but by using the actual numbers of visits listed in 4.2, since the discrete classes have no midpoints.

TABLE 4.2—Calculation of standard deviation from grouped data (discrete variable)

(1) Number of visits to or from doctor	(2) Number of children	(3) Col (2) × Col (1)	(4) Col (1) squared	(5) Col (2) × Col (4)
0	2	0	0	0
1	8	8	1	8
2	27	54	4	108
3	45	135	9	405
4	38	152	16	608
5	15	75	25	375
6	4	24	36	144
7	1	7	49	49
Total	140	455		1697

Mean number of visits = $455/140 = 3.25$.

Standard deviation = $\sqrt{\frac{1697 - \frac{455^2}{140}}{139}} = 1.25$.

Exercise 4. In the campaign against smallpox a doctor inquired into the number of times 150 people aged 16 and over in an Ethiopian village had been vaccinated. He obtained the following figures: never, 12 people; once, 24; twice, 42; three times, 38; four times, 30; five times, 4. What is the mean number of times those people had been vaccinated and what is the standard deviation? *Answer:* Mean = 2.41, SD = 1.27.

A middle-aged woman has gallstones. She has had several attacks of biliary colic coinciding with the onset of menstruation. Is this a recognised phenomenon?

This association is not recognised, but the wide-ranging actions of the sex steroids suggest that they do act on the musculature of the biliary tract and on hepatic function. Exactly what these effects are is not known. There is muscular atony of the gall-bladder in pregnancy.¹ At menstruation the level of sex steroids is low but has rapidly changed. An investigation of 300 young women treated by cholecystectomy showed that three-quarters of them had had their initial symptoms during pregnancy.²

¹ Gerdes, M M, and Boyden, E A, *Surgery, Gynecology and Obstetrics with International Abstracts of Surgery*, 1938, **66**, 145.

² Glenn, F, and McSherry, C K M, *Surgery, Gynecology and Obstetrics with International Abstracts of Surgery*, 1968, **127**, 1067.