THE ESTIMATION OF INTERCENSAL POPULATIONS.*

By J. SPOTTISWOODE CAMERON, M.D., B.Sc.,

Medical Officer of Health for the City of Leeds.

WE have been fortunate in Leeds in having been able to obtain for each intercept† three populations—namely, those for the census periods of 1881, 1891, and 1901, and to calculate by the method of interpolation approximate populations for the middles of the intermediate years.

This method has advantages over either dividing the differences equally, or estimating the populations in the later period as though they had increased at geometrical progression to the population enumerated at the last census. Both methods are open to fallacies, and during the autumn, since the earlier portion of this report was printed, I suggested to my friend Mr. Smith, of Pudsey, who was working at the subject of public health with me at the time, that he should average the results obtained by the methods of interpolation and geometrical progression, and use these figures, after certain corrections, for the accepted intercept populations for the middles of the intermediate years. This, with a large amount of labour, he has been kind enough to do, and I have already referred to and used some of his results.

A few words of explanation may, perhaps, be advisable for the sake of any who are not familiar with the processes used. By the method of geometrical progression it is, of course, understood that the population is supposed to increase from one census to the other at compound interest, the earlier population being looked upon as the principal, the latter as the amount. It is quite easy by the use of logarithms to obtain the rate at which the increase per annum would have taken place, and to assign to any particular period the amount which the population of a district should have reached at that date. This method supposes that the rate of progress has been uniform during the ten years.

In a district, however, in which the population has been increas-

* From Supplement to Annual Report.

† *Intercepts of City Wards in Leeds Union Divisions.*—In these several divisions the subdistricts are cut into by the wards, so that portions of one ward are sometimes in several subdistricts, and even townships are not confined entirely to wards. The township of Leeds, for instance, is divided into three subdistricts, of which the North contains portions of the North Ward, North-East Ward, and Central Ward. Such portions I have for convenience called "intercepts."

ing less rapidly during the 1891-1901 intercensal period than during the 1881-1891 period, it is probable that the increase during the earlier years has been at a greater rate than during the later ones. This is probably what really occurred in the population for the whole city. The rate of increase in Leeds during the 1871-1881 decade, supposing it to be uniform, would be at the rate of about 1.78 per cent. per annum. The rate of increase of the whole population of Leeds between the 1881-1891 censuses was similarly approximately at the rate of 1.75 per cent. per annum, while the population enumerated in 1901, regarded as having increased at uniform geometrical progression from that of 1891, would have necessitated that the annual increase should have been at the rate of a little under 1.56 per cent. Regarding these three periods as three periods with separate rates of increase—1.78, 1.75, 1.56 per cent.—it will be seen that these rates were diminishing, and that the curve describing the populations of each year estimated for the several decades at these three rates of increase would have cusps or points at the censal periods.

THE INTERPOLATION OF POPULATIONS BY THE METHOD OF DIFFERENCES.

If it be supposed that the change from 1.78 to 1.75, and from 1.75 to 1.56, in the rate of annual increase were coming about in a somewhat regular manner, it is possible to restate the population for the intermediate years in such a way that the curve of estimated populations should have a more uniform course. The method usually adopted—I believe at the suggestion of Mr. Noel Humphries, of the Registrar-General's office—is to establish three columns of differences.

The logarithms of the populations for the middles of 1881, 1891, and 1901 are written in the right-hand column, space being left between them for those of the intermediate years. Two other logarithms are found by a simple formula; one is placed in the first, one in the second column.* The logarithm just placed in

* The formulæ for obtaining the logarithms which are to head respectively the first and the second column, and which for convenience may be called (a) and (b), are as follows :

$$\begin{aligned} 100 (a) &= (\log. \text{ pop. } 1901 - 2 \log. \text{ pop. } 1891 + \log. \text{ pop. } 1881). \\ 10 (b) &= (\log. \text{ pop. } 1891 - \log. \text{ pop. } 1881) - 45 (a). \end{aligned}$$

These formulæ will hold good for any three periods ten years apart, whether 1881, 1891, and 1901 as above, or 1871, 1881, and 1891, and the interpolated populations for the intervening years will, of course, be for those years at the same time of the year as the known populations refer to. Understanding by (a) and (b), therefore, the numbers found by these formulæ, and where the known populations are ten years apart, the population for any intermediate year may

the second or middle column is then added to the first logarithm in the right-hand column, and gives the logarithm of the population for the middle of 1882. This logarithm is then written under that for 1881, and is that of the first of the interpolated populations. The logarithm in the first or left-hand column is then added (algebraically) to the logarithm in the second column, and their sum written immediately below. This new logarithm is now added to the logarithm in the third column, which we said was the logarithm of the population for 1882. The result is the logarithm of the population for 1883, which is now written in the third line of the third column.

In the same way the logarithm in the first column is always added to the last line in the second column, and that line then added to the line corresponding to it in the third, and so on until the amount in the tenth line of the second column when added to the tenth line of the third will give the logarithm that we have already written in the eleventh line of the right-hand column; and similarly for the second decade, the logarithm of the twentieth line in the middle column added to the logarithm in the twentieth line of the right-hand column should produce the logarithm already written in the twenty-first line of the right-hand column. It will be noticed that any particular logarithm in the right-hand column—say the eleventh—will consist of the original number at the top of that column plus (say) 10 times the number written in the second column, plus (say) $10 \times 9 \div 2$ times the number in the first column, the co-efficients of the quantities heading the first and second columns being for any particular line in the series, the second and third of the co-efficients of the expansion of the binomial for the power

be found from the binomial co-efficient as described in the text. For instance, the population for the year 1892, using the formulæ above, would be:

$$\begin{aligned}\text{Log. pop. 1892} &= \text{log. pop. 1881} + 11 (b) + 55 (a). \\ \text{But log. pop. 1891} &= \text{log. pop. 1881} + 10 (b) + 45 (a). \\ \therefore \text{log. pop. 1892} &= \text{log. pop. 1891} + (b) + 10 (a).\end{aligned}$$

If it be desired to get the populations for each quarter of each year of the decade, they can be obtained from the census population with a slightly altered formula—namely:

$$\begin{aligned}1,600 (a) &= \text{log. pop. 1901} - 2 \text{ log. pop. 1891} + \text{log. 1881}. \\ 40 (b) &= \text{log. pop. 1891} - \text{log. pop. 1881} - 3,160 (a).\end{aligned}$$

From these formulæ the populations, say, for the end of the second quarter—*i.e.*, the middle—of 1891 can be found thus:

$$\begin{aligned}\text{Log. pop. 1891 (mid.)} &= \text{log. pop. 1881 (cens.)} + 41 (b) + 820 (a). \\ \text{But log. pop. 1891 (cens.)} &= \text{log. pop. 1881 (cens.)} + 40 (b) + 780 (a). \\ \therefore \text{log. pop. 1891 (mid.)} &= \text{log. pop. 1891 (cens.)} + (b) + 40 (a).\end{aligned}$$

This method of getting the mid-year populations involves some very long logarithms, and the shorter method I have employed is described in the text and the foot-note on p. 39.

corresponding to the number of the line above. Thus, to take another example, the co-efficients of (*b*) and (*a*) respectively for the twenty-first line will be 20 and $20 \times 19 \div 2$.

DISADVANTAGES OF THE METHOD.

It is by no means certain that the changes in the rates of increase or decrease of population have been made at all uniformly, and strange results are occasionally obtained by this method when relied upon alone. For instance, the intercept of the North-East Ward in the Potternewton township had a population of 102 at the 1881 census, of 179 at the 1891 census, and of 6,210 at that of 1901. If the curve of population be obtained by interpolation, the population for 1882 would be actually less than that for 1881, and would continue to decrease to 1884, after which it would increase, until in 1889 it had exceeded that of 1881, whence it would rise with a rapidly ascending curve through the census figure of 1891, and reach 6,210 in 1901. This is obviously incorrect, as there is no reason to suppose that the population in this part of the town actually decreased between 1881 and 1891. It was probably nearly stationary, though slightly increasing during that period. So far as the interpolation populations between 1891 and 1901 are concerned, they are not much affected by this comet-like orbit, but on the whole I thought it safer in dealing with intercepts to equalize this irregularity by combining the geometrical estimates with those by interpolation. In dealing with the whole city this difficulty is not likely to occur. Intercepts alter in population, as in the one instanced, by the sudden opening out of land for building purposes and the migration to the area of families from other parts of the town.

COMPARISONS OF ESTIMATED POPULATIONS FOR THE CITY FROM 1890-1901.

On the annexed table will be found populations for the city as a whole arrived at in three different ways. In the second column populations have been estimated as if they had increased at compound interest at a rate calculated from the actual populations at the census of 1881 and that of 1891. The census is taken at the end of the first quarter of the year. The populations were estimated to the middles of the years named in the first column. The supposition on which these estimates were made was, of course, that the population went on increasing during these years at a uniform rate of compound interest obtained by the comparison of these two populations. The estimate of the population for the date of the

1901 census upon this calculation would have been 436,919, but the actual population enumerated was only 428,968.

Table showing Mid-Year Population of Leeds from 1890-1901, estimated in Three Ways, and giving Factors for Correcting Rates as given in the Annual Reports from 1891-1900, to Rates upon the Population estimated by the Two Other Methods.

Year.	Population calculated by Geometrical Progression at Rate of Increase between 1881-1891 (see Table D, Part II.).	Population calculated by Geometrical Progression at Rate of Increase between 1891-1901.	Populations interpolated according to Changing Rates between 1881-1891 and 1891-1901.	Factors to Correct Rates, calculated on the Populations given in Column 2 to Rates on the Corresponding Populations—	
				In Column 3.	In Column 4.
1	2	3	4	5	6
1890 ...	362,768	363,264	363,079	—	0·999102
1891 ..	369,099	368,929	369,034	1·000458	1·000175
1892 ...	375,540	374,678	375,081	1·002298	1·001220
1893 ...	382,094	380,517	381,157	1·004140	1·002456
1894 ...	388,761	386,448	387,259	1·005987	1·003879
1895 ...	395,546	392,470	393,387	1·007838	1·0054907
1896 ...	402,449	398,587	399,535	1·009688	1·0072982
1897 ...	409,472	404,799	405,716	1·011544	1·0092884
1898 ...	416,618	411,108	411,895	1·013404	1·011467
1899 ...	423,889	417,514	418,101	1·015267	1·013842
1900 ...	431,287	424,021	424,322	1·017133	1·0164125
1901 ...	438,813*	430,630	430,555	1·019030	1·019178

In the third column the same method of computation is employed, but the rate of increase is obtained from a comparison of the populations at the censuses in 1891 and 1901, the increase each year, if regarded as uniform, having been, as already stated, at a rate a little less than 1·56 per cent. This method of assuming the population as having grown at a rate obtained from the known results of the censuses near the beginning and the end of a period is probably not very far from correct; but as the rate calculated in the same way for the preceding decade was higher, it is probable that the populations in the earlier years will be a little under- and those in the later years a little overestimated. The error, however, is probably in no case very great. Assuming for the moment that these populations were approximately correct, then the birth and death rates, which were published in the annual reports during the decade, would require to be multiplied in each case by the figure which I have given in the corresponding line of column 5, and

* Used only in earlier reports for 1901.

which have been obtained by dividing the numbers in column 2 by those in column 3.

In the fourth column are given the populations for the several years estimated by the method of interpolation already described, and the last column gives the factors for correcting rates obtained from the populations in the second column, so as to make them correspond to the rates which would have been published had we known and used the populations now printed in the fourth column. Some of the fallacies about the calculation of populations by the method of interpolation have been already mentioned.

It will be noticed that the difference between the factors given in the fifth and sixth columns is in no case very great, and if the corresponding factors in any particular year be divided one by the other, the resulting factor will generally be so small as not to affect the death-rate from all causes in the first place of decimals.

HOW THE MID-YEAR POPULATIONS HAVE BEEN ARRIVED AT.

The method given above presupposed that the populations for the middle of the years 1881, 1891, and 1901 are known. It may be worth while putting on record the method by which these mid-year populations were found. As already said, the census population is taken at the end of the first quarter. The census populations for 1871, 1881, and 1891 were treated by the method of differences already explained, and the corresponding figures for the same time-periods for 1880 and 1882 were obtained from them. With these three data—the populations for the end of the first quarter of 1880, 1881, and 1882—populations were next found by the method of differences for the intermediate quarter periods, including, therefore, that for the middle of 1881.*

In the same way, from the census populations of 1881, 1891, and

* Using again (a) and (b) to represent the head-lines of the first and second columns in the tabulations by the method of difference, the formulæ become :

$$16 (a) = (\log. \text{ pop. } 1882 - 2 \log. \text{ pop. } 1881 + \log. \text{ pop. } 1880).$$

$$4 (b) = (\log. \text{ pop. } 1881 - \log. \text{ pop. } 1880) - 6 (a).$$

The several population logarithms are those found for the ends of the first quarter of the years named. To find the intermediate populations it must be remembered that we write the logarithms of the census period populations for 1880, 1881, and 1882 in the third column, leaving three blank lines between each for the logarithms to be found of the populations for the ends of the second, third, and fourth quarters of 1880 and 1881. The logarithm of the population for 1881 (middle) will therefore fall on the sixth line of the third column. The co-efficients of the second and third terms of the expansion of a binomial to the fifth power are 5 and 10, and to the fourth power 4 and 6.

$$\text{Hence } \log. \text{ pop. } 1881 (\text{mid.}) = \log. \text{ pop. } 1880 (\text{cens.}) + 5 (b) + 10 (a).$$

$$\text{But } \log. \text{ pop. } 1881 (\text{cens.}) = \log. \text{ pop. } 1880 (\text{cens.}) + 4 (b) + 6 (a).$$

$$\text{Hence } \log. \text{ pop. } 1881 (\text{mid.}) = \log. \text{ pop. } 1881 (\text{cens.}) + (b) + 4 (a).$$

1901, populations for the end of the first quarter of 1890 and 1892 were obtained, and from these the intermediate population for the middle of 1891. This method was not available for the population for the middle of 1901, but the formulæ just used were employed to obtain approximate census period populations for 1900 and 1902, and by the same method the population for the middle of 1901.

Having now obtained by this double method the three populations for the middle periods of 1881, 1891, and 1901, the intermediate middle-year periods were interpolated, from these data, by the method already described.

* * * * *

MORTALITY IN 1895-99.*

The mortality for the quinquennium 1895-1899, taken as an average of the rates printed at the time, was 19·5. As, however, the population had not been increasing in quite as large a ratio during the latter half of the decade as it was estimated to have done in the earlier, the rate should probably have been more nearly 19·7.

Leeds Death-Rates, 1895-1899, estimated on Populations.

Years.	Regarded as Terms of a Geometrical Series, whose Ratio was obtained from the Results of the Censuses—		Found by the Method of Differences from the Data of the Censuses 1881, 1891, and 1901.
	1881-1891.	1891-1901.	
1895	20·5	20·7	20·6
1896	18·8	18·9	18·9
1897	19·9	20·1	20·1
1898	19·2	19·5	19·4
1899	19·1	19·4	19·4
Lustrum	19·5	19·7	19·7

This rate is much lower than the rate of any of the six preceding quinquennia for which we have official returns. For the triennium 1900-02 the rate, as the revised population, has been 19·1, the lowest for any three consecutive years in our Leeds record.

* From Annual Report for the year 1902.