SEASONAL VARIATION IN THE INDIAN BIRTH-RATE

By D. D. KOSAMBI and S. RAGHAVACHARI

1. In this note we study the seasonal variation in the Indian birth-rate (or, properly speaking, in the total number of recorded births), and discuss briefly the major factors that cause such variation.

The actual fact of the variation has generally been admitted by all demographers of modern times. A good deal of the classical work will be found cited in the discussion by Havelock Ellis. The European records seem reasonably satisfactory in this respect, but the modern conclusion that there is a maximum in March and perhaps, according to Landry (1949), a small secondary maximum in September (Fig. 1) is certainly not extensible to India. We may even venture to doubt the generality of any such statement for Europe and to suggest that if the data were broken down regionally for the separate economic classes, a considerable inner variation would be found. The purely agrarian regions of eastern Europe would almost certainly demonstrate the necessity of such an analysis when the birth-rate variation there is contrasted with that in highly industrialized

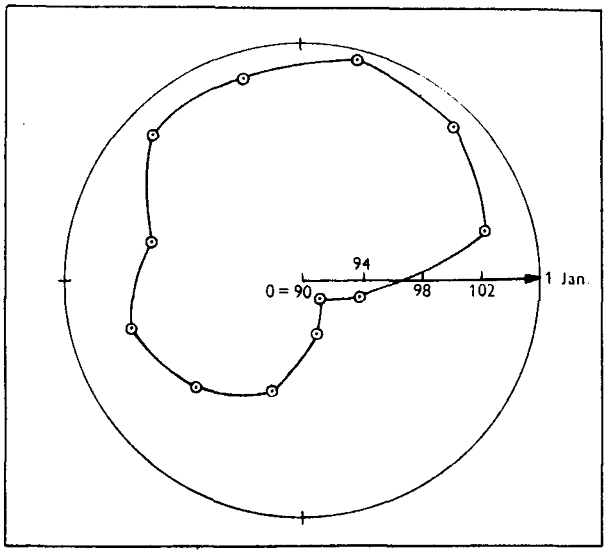

Fig. 1. Seasonal indices of crude birth-rates in France for the years 1932–8 from data given by Landry (1949).

countries like England. In the latter region it seems to have been generally accepted that conceptions are highly correlated with the principal holidays. We shall show that this cannot hold in general for India. One of our major conclusions is that the villages, where the greater part of people in India dwell, show an entirely different seasonal variation from the cities which they surround; so that purely racial or climatological factors cannot be taken as fundamental. On the other hand, it must also be emphasized that the data are singularly incomplete and unreliable for the whole of the country. The only plea that can be made for the validity of any such comparative inquiry as the present is that the unreliability of these data presumably operates in a more or less uniform fashion for all the localities considered. In addition to the untrustworthiness of the data, even access is rather difficult, so that we sometimes wonder how it is that authoritative statements upon Indian vital statistics obtain so much publicity and credence (Chandrasekhar, 1949).

2. In India, deaths as well as births are supposed to be recorded for official purposes. But, whereas no obsequies may be performed without some record of death and a fairly satisfactory cause being assigned to it, birth records are still kept in a remarkably haphazard manner. The earliest known figures, namely, of the first census of India in 1872, still lie unpublished in the Archives of the India Office, London, where they may be consulted by special permission. The earliest published records on vital statistics which we have in India seem to be those of S. A. Hill (1888). Hill attempted an extremely thorough inquiry in his analysis of birth-rates for the United Provinces, actually taking into account the deaths and immigration as well as emigration. We regret to have to express the opinion that the extraordinary accuracy which seems to characterize

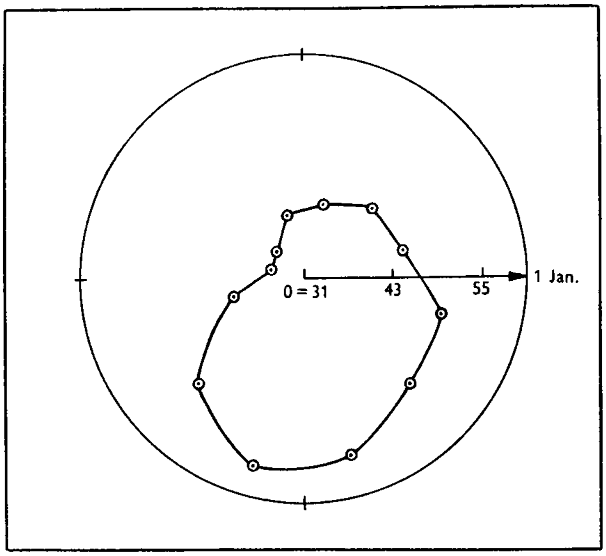

Fig. 2. Monthly average rates of birth per thousand per annum for the years 1878–87 as quoted by Hill (1888).

his survey of the ten years 1878–87 is spurious, in view of the fact that even now returns from the villages are quite inaccurate as far as birth statistics are concerned. Nevertheless, there is no doubt about the qualitative accuracy of Hill's data, and therefore his conclusions may be taken as properly established. The records are illustrated by our Fig. 2. Hill gave an extremely sane objective analysis of provincial conditions, pointing out that the maximum births corresponded to conceptions about the month of December when the conditions of life for the population at large were about the best. Similarly, in his opinion, the minimum agreed with conceptions in the worst month (September) of the year both from the climatological and the economic point of view (p. 250).

'...The minimum falls in June and the maximum in September—dates which point to a maximum of conceptions in December, and a minimum in September. The latter month is near the end of the long and depressing hot season, when malarial influences are rapidly increasing to a maximum, the food supply of the year is nearly exhausted, and there is the greatest tendency to suicide. The births, as well as the deaths, therefore show that at the end of the rains the vitality and energy of the people have reached low-water mark.

'In December, on the other hand, not only is the salubrity of the country greatly increased as shown by the rapid diminution of nearly every cause of death, but food is again cheap and abundant. The crops of millet, on which the poorer classes live, are sown in July and reaped in November. During December and the latter half of November they are threshed out, and then is the season for paying the village functionaries and labourers their share of the produce. Consequently food is more abundant at this time of the year than at any other, and as a result of these conditions we find a large number of births the following September and October.

'It thus appears that among the poorest of the population there is probably still a more or less distinct annual reproductive season, but instead of being determined by the returning warmth of spring, as must have been the case in prehistoric Europe, it follows the annual return of healthy conditions with abundant food supply.'

3. Coming to recent times, a jump which seems justified by the absence of any useful or even first-hand survey for India during the interval, we find a series of papers concerned with birth statistics. M. K. Subramaniam (1933) attempted to correlate temperature with sex determination in man and showed that Heape's argument, namely, 'sudden changes in temperature induce boisterous activity', did not hold, at least for his own data. The data itself cover the five years 1926–30 for Madras city. Subramaniam took his figures from the *Fort St George Gazette*, which we should have been glad to use, had it been possible to obtain copies of the publication in Bombay, even on loan. This deficiency is partly made up by the data published periodically in the *Gazette of India* (discussed later), and surprisingly enough our own findings for the 9 years 1931–9 contrast markedly with those of Subramaniam for the 5 years just preceding. It is not worth while discussing the source of the discrepancy (illustrated by our Fig. 3), as no deep investigation could be based upon 5 years' data, however accurately recorded. Nevertheless, the incompatibility must be pointed out, for the simple reason that authoritative statements are often made on the strength of such inadequate material. Subramaniam neglected one extremely useful tool of the modern statistician, the analysis of variance, though the point is of no great importance for our purpose, because the author's main interest was in the sex-ratio, which we have ignored altogether. The next attempt seems to have been by K. Das & P. C. Mahalanobis (1933–4), whose conclusions are based on 6481 hospital cases for the period 1850–1901. All these cases, however, were taken from the records of just one hospital in Calcutta, and therefore cannot be said to represent anything of interest for our major problem. The authors themselves were primarily interested in maternity deaths and still-births. Their table 3, again unsupported by any analysis of variance, nevertheless, shows by the χ^2 test a significant difference between the months. In this case, the accuracy of the primary data cannot be doubted, only its representative character. Their figures show a maximum in October and a minimum in June, and therefore seem to agree qualitatively with our figures from the *Gazette of India* for the cities in Bengal (Fig. 12). But the authors' reasoning seems odd (p. 218): 'The maximum number of childbirths (643) which occurs in October is 1·7 times the minimum number (376) which occurs in June. This shows that the optimum condition for pregnancy occurs in December–January, which in Bengal is the time of harvest. The season of minimum pregnancy, on the other hand, occurs in August–September, that is, towards the end of the monsoon season.' That harvest conditions should affect childbirth in a large rural area like the United Provinces seems quite reasonable, but just why this should affect Calcutta city is not clear. The maternity statistics in Madras were studied on similar lines by K. R. Nair (1935–6) from the data of the Government Hospital for Women and Children at Egmore, Madras city. This contains the results of 42,206 confinements between the years 1920 and 1933; the analysis of the data is apparently more thorough than in the previous notes. It is of particular interest that table 2 of Nair's note agrees qualitatively with our own graph (Fig. 3). He gives both the total number of births in Madras city extracted from the Administration Reports of the Health Department of the Corporation of Madras and the Egmore Hospital deliveries, showing that his restricted sample is in fair proportion to the total number. Nair's table 3 gives the monthly variation in confinements, which he corrects by the reasonably justified assumption of proportionality, using the percentage of live to total births for the whole city of Madras. Both the number of confinements and the corrected

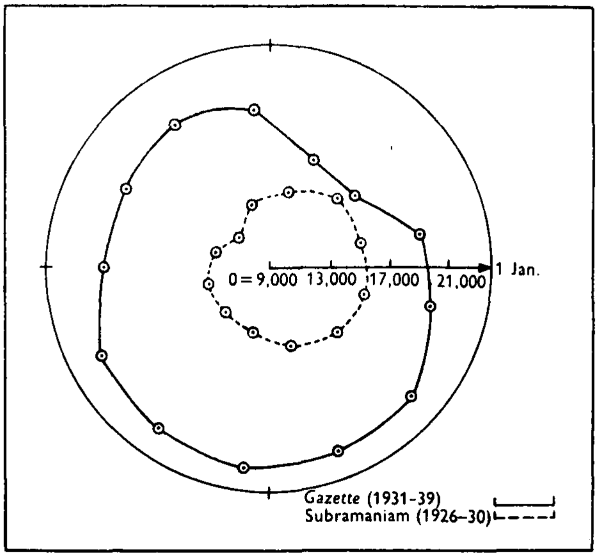

Fig. 3. Total number of births in Madras city as quoted by M. K. Subramaniam (1933) for 12 months for the years 1926–30 and as tabulated by us for thirteen 4-weekly periods for the years 1931–9.

indices compare very well with our data, in contrast with the figures given by Subramaniam. On the other hand, it must be pointed out that the totals calculated from Subramaniam's statements for the years 1926–30 are decidedly in excess of the corresponding figures quoted in Nair's table 2, namely, 1926: 31,307–23,105; 1927: 33,686–26,018; 1928: 31,314–25,050; 1929: 35,419–24,411; 1930: 36,093–26,922. We do not know how to account for the differences, which are not even in strict proportion ($\chi^2 = 126{\cdot}99$, D.F. $= 4$, $P < 0{\cdot}001$). The point is emphasized only to show the unreliability of most published sources of data in this country. In Nair's note is found the statement (p. 159). 'This shows that the optimum condition for pregnancy occurs in December–January and the minimum in April–May. The first belongs to the coldest part of the year in Madras and the second period to the hottest.' This seems quite reasonable.

The last report which we shall cite here is that of N. T. Mathew (1940–1). He also takes his figures for Madras city from the *Fort St George Gazette*. This is the first paper which confines itself deliberately to the influence of seasons upon human reproduction. Mathew reduces the data to a much clearer time-base, namely, successive 4-weekly periods from 1926 to 1938. He goes further to estimate the birth-rate, which means division of the number of births by the population figures as estimated by logarithmic interpolation between the years 1921 and 1931, as well as extrapolation after 1931. This seems to us to vitiate, rather than add to, the accuracy of his analysis. Mathew was primarily interested in the effect of temperature upon conception rate and was greatly struck by the inverse correspondence between the two. He found a high negative correlation between the monthly average temperature and the monthly birth-rate, namely, $-0{\cdot}6971$. The obvious remark on this point is that the fitting of a multiple regression is called for with other factors such as humidity; also, an analysis of variance, if properly carried out, would show the amount of variation not explained by temperature to be of considerable significance. Temperature cannot be of any immediate importance as the direct primary factor for conceptions, for Mathew himself finds the diurnal temperature variation important; and it is known that the probable difference between the day and night as regards conception rates could not primarily be ascribed to the temperature alone. Such definite statements about conception rates are based upon an arbitrarily fixed period of gestation, and the variation in the length of this period neither is nor (with such data) can be allowed for.

Table 1. *List of towns with regions and the respective periods for which data is collected from the* Gazette of India

Province	Region	Towns	Period
Madras	A	Madras city, Salem, Madura	1931–9
	B	Bellary, Adoni	1931–9
	C	Mangalore, Calicut	1931–9
	D	Vizagapatam, Rajahmundry, Cocanada	1931–9
Bombay	A	Poona, Sholapur	1930–9
	B	Surat, Ahmedabad	1930–9
	C	Bombay city	1930–9
United Provinces		Agra, Bareilly, Moradabad, Meerut, Muttra, Farrukkabad, Dehra Dun, Cawnpore, Lucknow, Benares, Allahabad, Gorakhpur	1930–9
Central Provinces and Berar		Raipur, Nagpur city, Nagpur civil station, Jubbalpore	1928–35, 1937, 1938
Bihar		Patna, Gaya, Jamshedpur	1930–9
Sind		Hyderabad, Karachi	1930–4, 1936–9
Punjab		Bhiwani, Lahore, Amritsar, Sialkot, Rawalpindi, Multan, Delhi, New Delhi	1930–9
North-West Frontier Province		Dera Ismail Khan, Peshawar	1923, 1925–7, 1934–9
Orissa		Cuttack, Puri	1930–5, 1937–9
Bengal		Calcutta, Howrah, Bhatpara, Dacca	1930–9

4. The *Gazette of India* provides weekly birth statistics for several cities in India and what is now Pakistan from the year 1921. The natural assumption, that these figures would give a sure foundation for any statistical work, is sadly in error, as is seen by a glance at these *Gazette* reports. The number of cities included in the list increased steadily during the first 10 years. Throughout the entire period, many cities furnish no information at times. In a considerable number of cases, the entry merely means that no information from those particular places was supplied for that particular week; the deficiency is sometimes but not always corrected in some later issue of the same publication. The result is that a considerable number of gaps remain in the data. All vital statistics in India seem to be painfully collected every year, but the greater part of the record disappears without trace. Inquiry from the lowest to the highest sources shows that most local registers vanish after a few years, or at least are not to be found whenever an investigator wishes to trace them. Specific examples are necessarily omitted to avoid unpleasantness.

The fifty-two cities of India for which fairly complete data were available for the period 1930–9 are grouped here into several climatological regions, following for this purpose the accepted

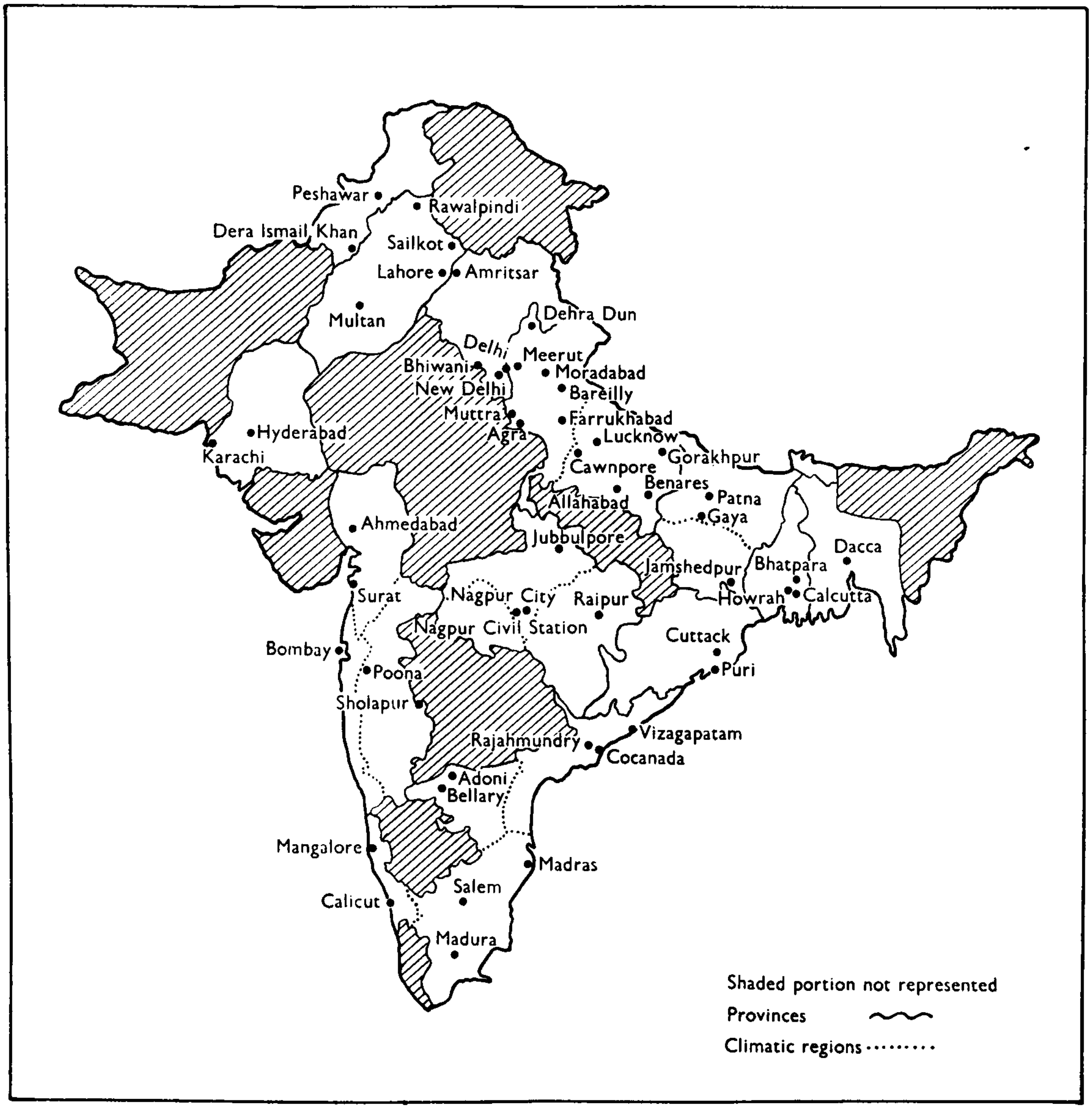

Fig. 4. Map of India showing the provinces, the climatic regions and the fifty-two towns given in Table 1.

regional divisions of the Indian Meteorological Department (information very kindly supplied by Dr S. L. Malurkar of the Colaba Observatory). After 1939 the war caused a sudden increase in the urban population of India, with disturbed wartime economic conditions and the still greater post-war disturbances which culminated in the partition of the country in 1947. Hence the figures for the subsequent years would not give a fair basis of comparison, even had they been available. As a matter of fact the *Gazette* did not publish any figures from October 1940 to June 1946. A list of the towns, regions and the years considered is given in Table 1, and these are also represented in the accompanying map (Fig. 4). Of these figures the gaps, if isolated, were filled in for that particular year and city by means of linear interpolation between adjacent figures. Where many consecutive gaps occur, the year or the city affected had to be omitted altogether, which accounts primarily for the scanty number of years and towns that have been included. In spite of these defects, a comparison of the figures for Poona and Bombay cities (Table 2), as collected directly from the municipal registers and as recorded in the *Gazette of India* for the years 1930–9, brings out the qualitative value of the *Gazette* data on the whole. The extra labour of analysis of variance has been avoided in these cases for reasons explained at the end of §8 and by Tables 5 and 6. The χ^2 values for Bombay and Poona cities are 11·06 and 41·80 respectively in a uniformity

Table 2. *Total 4-weekly number of live births in Bombay and Poona cities as collected from respective municipalities and as found in the* Gazette of India

Bombay city		Poona city	
Municipality (1930–9)	*Gazette* (1930–9)	Municipality (1931–40)	*Gazette* (1930–9)
26,539	27,088	3,548	3,566
24,359	24,349	3,543	3,625
22,470	22,486	3,719	3,882
22,160	22,099	3,556	3,968
22,281	22,567	3,428	3,679
21,616	21,852	3,441	3,446
22,512	22,626	3,513	3,518
23,290	23,065	3,883	4,067
23,608	23,688	4,286	4,215
27,221	26,962	4,528	4,402
27,982	28,167	4,487	4,323
30,238	30,251	4,180	4,228
30,836	30,734	3,927	3,840

test. The degrees of freedom being 12 in each case, the corresponding probabilities are 50 % and less than 0·1 %. The first is not serious, and therefore we have graphed only our own figures from the municipal records. For Poona city however, the most serious discrepancies are located in the fourth and fifth 4-weekly periods which contribute 22·56 and 8·86 to the total χ^2 value; we are unable to account for this serious difference. No allowance has been or can in general be made for the time-lag which is almost certain to exist between the actual date of birth and the date of registration (upon which the reports are based). This will be discussed later on for Bombay. It might be nearer to the truth to assume that the actual maxima and minima occur about a month earlier than in the reported data, which are represented for convenience by means of polar graphs in Figs. 5–12. The sequence is from south to north, recalling that Madras city is already represented in Fig. 3. The general feature of these curves is their dissimilarity. The exception as regards climatological regions is that of the United Provinces, Central Provinces and Bihar, which showed such complete similarity that the data for the three have been combined into one figure (Fig. 8).

Madras A (south-east), made up of the three cities Madras, Salem and Madura, and Madras B (Deccan), comprising Bellari and Adoni, are represented in Fig. 5. The shapes of the curves on the whole are not dissimilar, but it will be noticed that the minimum and maximum occur a month earlier in group B. Madras C (Fig. 6) takes up two cities on the west coast, namely, Mangalore and Calicut, where the curve is of an entirely different shape with two maxima and two minima. Madras D (Fig. 6), consisting of the three cities Vizagapatam, Cocanada and Rajahmundry, presents a still odder shape with one maximum and one minimum 6 months apart. Both maximum and minimum are further advanced than in groups A and B, and the shape is nearer to those two than to group C. This is approximately what would be expected, because these east coast cities are climatically more similar to those of Madras A and B than to the cities of the opposite coast.

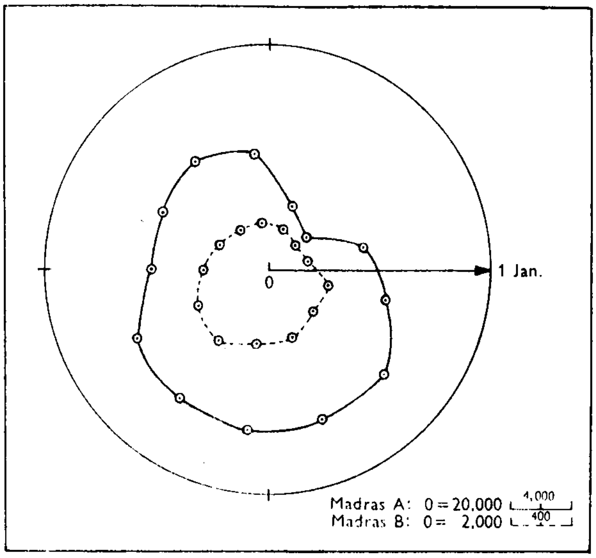

Fig. 5. Total 4-weekly number of live births in Madras A and Madras B for the 9 years 1931–9.

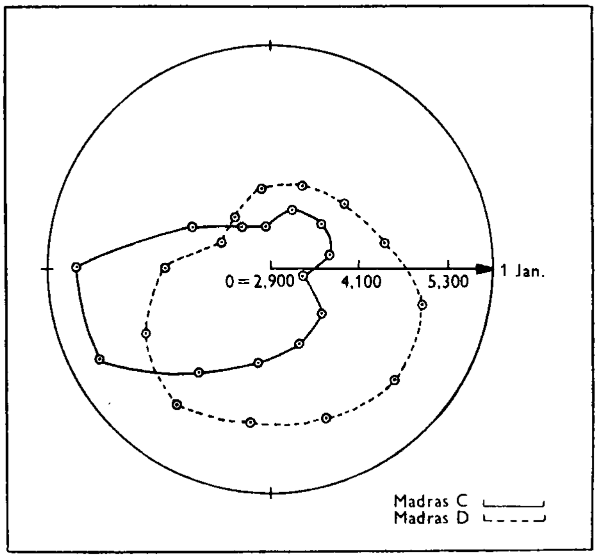

Fig. 6. Total 4-weekly number of live births in Madras C and Madras D for the 9 years 1931–9.

Next in order is Bombay Province, rather feebly represented (apart from Bombay city in Fig. 20) by two cities in each of the two different groups (Fig. 7), group A containing Poona and Sholapur and group B covering Surat and Ahmedabad. The sharp minima of the Madras curves have disappeared and the maximum falls now in September–October. Farther north come United Provinces, Central Provinces and Berar and Bihar (Fig. 8), represented by nineteen cities altogether, where there is just enough reason to suspect the existence of a secondary maximum about March–April besides the major maximum in September–October. But actually the shape of the curve does differ noticeably from those of Bombay Province in the neighbourhood of the secondary maximum.

Farther to the west is Sind (Fig. 9), with the two cities of Hyderabad and Karachi, where the maximum seems to lie in December. To the north of these comes the Punjab (Fig. 10), represented by eight cities, including Delhi and New Delhi, besides Bhiwani, Amritsar, Rawalpindi, Lahore, Sialkot and Multan. This is characterized by a maximum about October–November with a fairly sharp minimum 6 months later. In the same direction we have the data for the North-West Frontier Province (Fig. 11) comprising the two cities of Peshawar and Dera Ismail Khan. The maximum here is in January, and again the curve cannot be derived by a simple rotation of the others.

In eastern longitudes we have two regions, namely, Orissa and Bengal (Fig. 12), of which the former includes Cuttack and Puri while the latter is denoted primarily by Calcutta, and its suburbs Howrah and Bhatpara, with just one other city, Dacca. The Bengal curve is smoother than that for Orissa simply because the monthly figures are far greater, which reduces the irregularities. The Orissa maximum seems to fall in December and the Bengal maximum a month earlier, but there is no similarity in the shapes of these two curves.

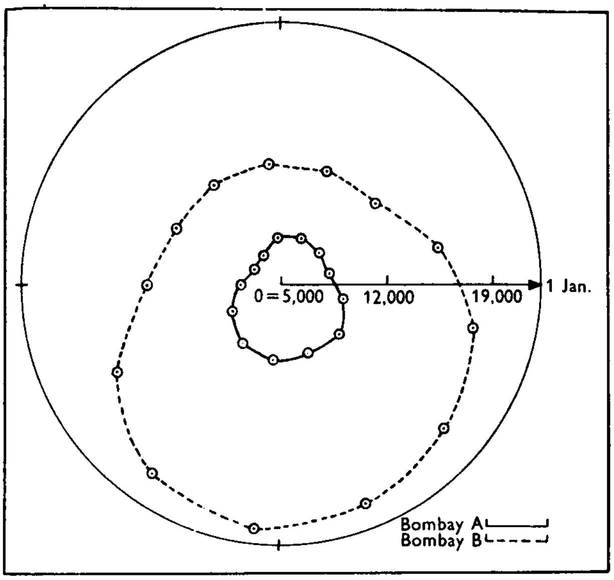

Fig. 7. Total 4-weekly number of live births in Bombay A and Bombay B for the 10 years 1930–9.

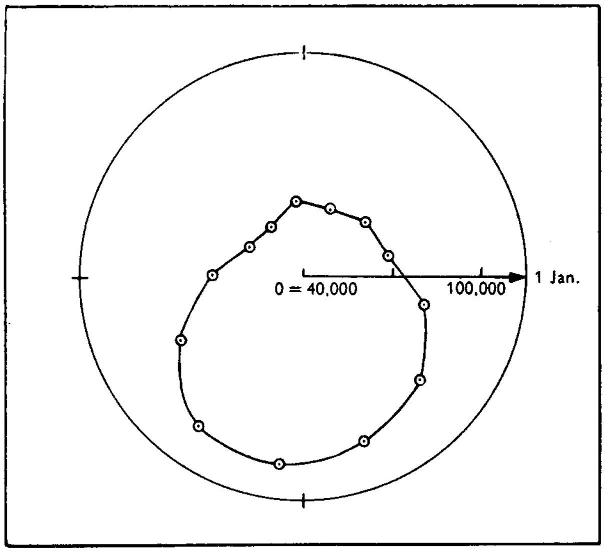

Fig. 8. Total 4-weekly number of live births in United Provinces, Central Provinces and Berar and Bihar for the 10 years 1930–9.

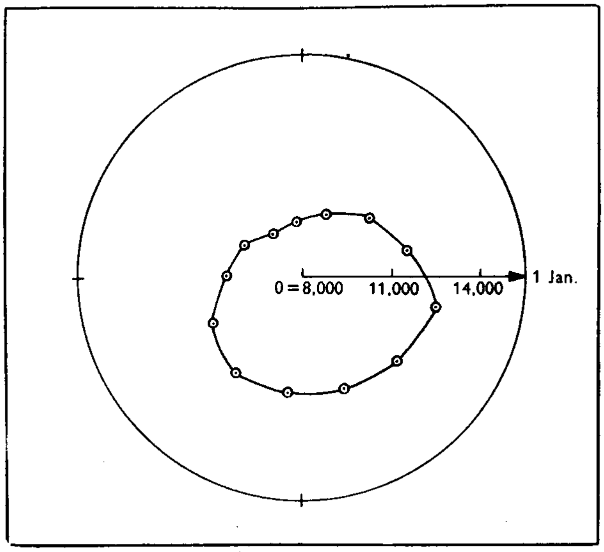

Fig. 9. Total 4-weekly number of live births in Sind for some 9 years between 1930 and 1939.

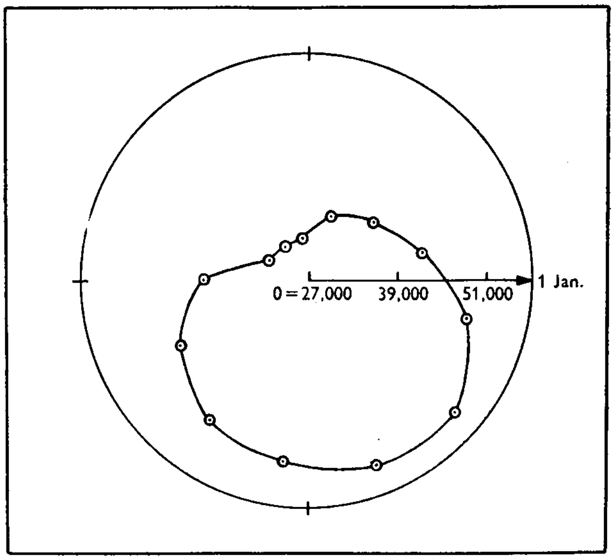

Fig. 10. Total 4-weekly number of live births in Punjab for the 10 years 1930–9.

In discussing these results, the first salient feature is that, judged by the records, there is no such thing as a standard Indian birth-rate curve. This is seen to be true regardless of the differences in origin and scale which exaggerate individual peculiarities of the polar co-ordinate graphs. We suggest that the urban variations are due primarily to changes in the weather, which in India is dominated entirely by the monsoon rainy season. The monsoon starts in the south-west coast by about the beginning of June, and advances steadily in 3–4 weeks over the entire country. The return monsoon begins in the north-east corner by about August and comes down in the opposite direction. Almost the entire rainfall of the year is compressed within approximately 4 months. The season just before the monsoon is usually very unpleasant because of the intense heat, and the onset of the monsoon causes an amelioration which is more marked in the Deccan Plateau than anywhere else in the peninsula. The cessation of the monsoon is followed by another period of excessive heat which generally does not pass away till some time in November, when the best

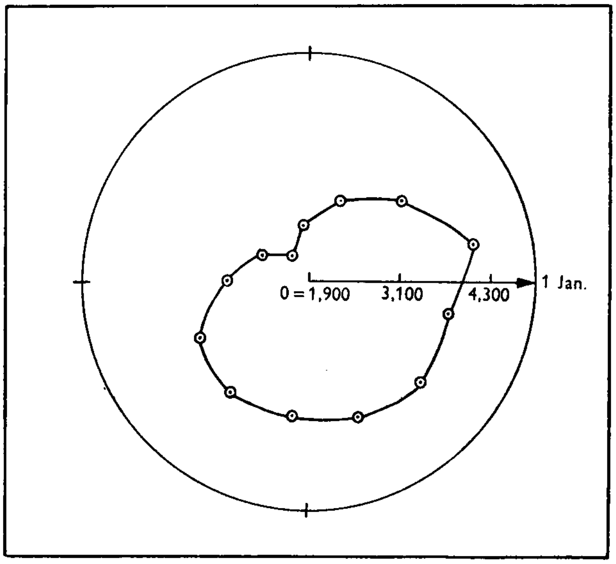

Fig. 11. Total 4-weekly number of live births in North-West Frontier Province for some 10 years between 1923 and 1939.

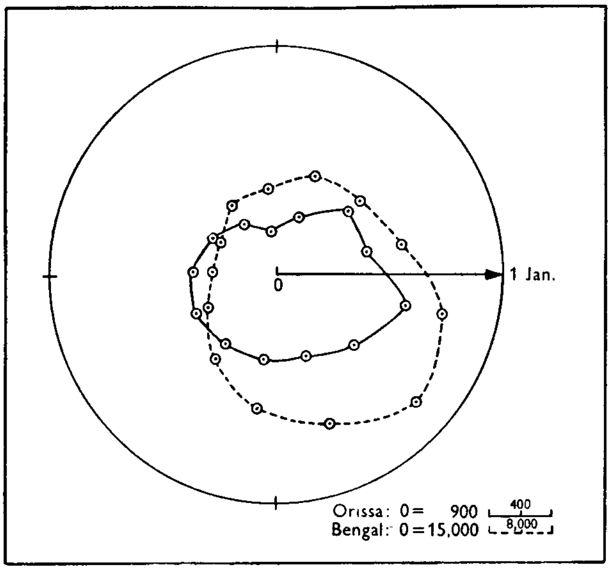

Fig. 12. Total 4-weekly number of live births in Orissa and Bengal for the years 1930–9.

season of the year begins in most cities. The months of December and January are pleasantest in all cities except those of the extreme north, and perhaps for a few weeks those of the United Provinces also. With these exceptions, where the cold may be considered as relatively severe, the season is undoubtedly the pleasantest and presumably the most favourable for conceptions, seeing that the level of employment or unemployment should be comparatively stable in the urban areas right through the year.

5. The general data being clearly unsatisfactory, it was decided to see if intensive rather than extensive work could give more information. Actually this part of the work was begun by Mr A. D. Taskar at Poona in 1947 under the direction of the senior author. Among the records accessible at Poona were the municipal registers and certain other registers in the Public Health Directorate. Inasmuch as the latter registers included data for places beyond Poona city, we may discuss their figures first before coming to Poona city and its suburbs. Bombay Province is divided into three for purposes of the Department of Public Health, the southern division having its headquarters in Poona under an Assistant Director of Public Health. In each district of the division the data are recorded separately for the urban and rural circles. The custom seems to be that municipal reports are sent directly to the corresponding officer of the Health Department, and that the report from each village is received by a clerk assigned to each group of villages. Taskar was able to collect the data relating to town and rural circles of only six districts of the southern division, namely, Ratnagiri, Kolaba, Thana on the coast and Poona, Nasik, Ahmednagar on the Deccan plateau (elevation above sea-level about 1800 ft.). His data relate to the following years:

Thana	1934–6, 1940–2, 1941–6	Poona	1931, 1935–9, 1941–6
Kolaba	1934–6, 1938–42, 1944–6	Nasik	1934, 1935, 1937, 1939–46
Ratnagiri	1934, 1937–46	Ahmednagar	1936–46

An attempt was made through Dr M. D. D. Gilder, Minister for Health, Bombay State, to collect such data for the whole of Bombay State. Unfortunately, his Deputy Secretary was unable to locate any of these figures for any of the divisions; the ultimate purpose and fate of the registers

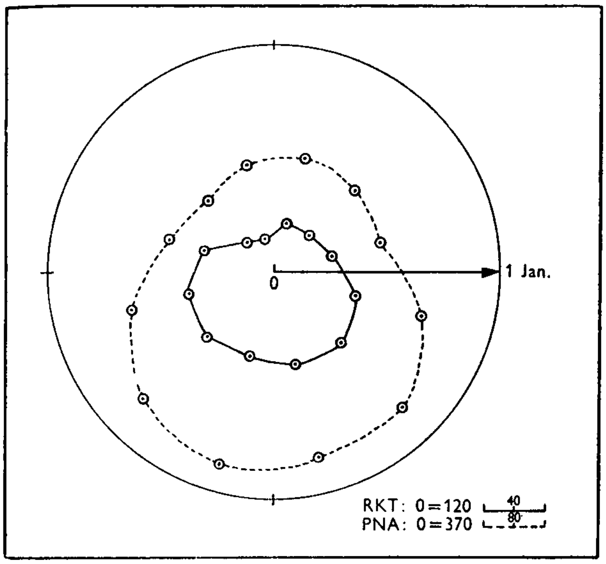

Fig. 13. Normalized number of live births in the town circles of: (i) Ratnagiri, Kolaba and Thana (RKT) and (ii) Poona, Nasik and Ahmednagar (PNA) for some 11 years between 1934 and 1946.

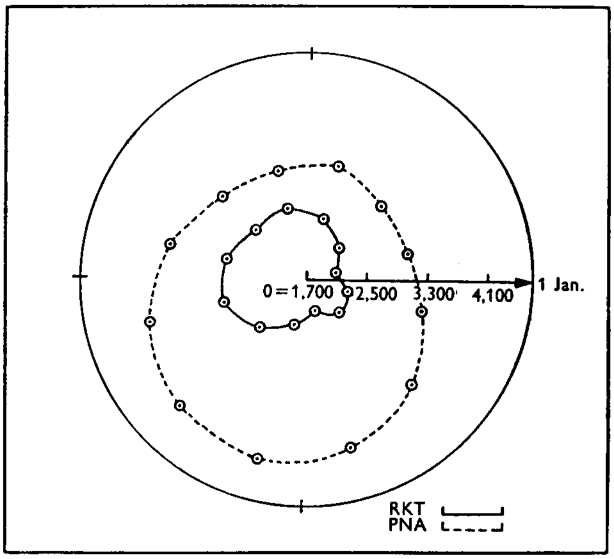

Fig. 14. Normalized number of live births in the rural circles of: (i) Ratnagiri, Kolaba, Thana (RKT) and (ii) Poona, Nasik and Ahmednagar (PNA) for some 11 years between 1934 and 1946.

still remain something of a mystery. Since Taskar's work was done conscientiously and checked, we accept his figures and express our gratitude to him for having made them available to us. The corresponding graphs are in Figs. 13 and 14. Here, at a glance, is seen a sharp and unmistakable contrast. The recorded figures have been normalized (by Taskar) by dividing out the total by the number of days in a month (taking into account leap years for February, etc.). In the town circles (Fig. 13) the maximum for the three plateau districts comes in September and for the three coastal districts about October–November. The reason is fairly clear, namely, that the post-monsoon heat is decidedly worse in the coastal regions than on the plateau, and the favourable season which is optimistically called winter begins somewhat earlier on the Deccan plateau. But much more noticeable is the fact that the plateau districts give a secondary maximum which becomes quite insignificant for the coast, where only a sharp minimum is to be noticed. The reason that we propose to give for this phenomenon is that the onset of the monsoon on the plateau causes a certain fall in temperature, greatly improving the dry but extremely hot and dusty climate of the months just preceding. On the coastal strip, the monsoon is certainly welcome, but does not cause so marked an improvement, for the weather continues to be unpleasantly humid without any great reduction in temperature. Bombay city, for which the figures are the most reliable, shows only a smooth oval with just one observed maximum and minimum.

When we come to the rural circles (Fig. 14) we see an entirely different type of variation. On the plateau the primary maximum is about the same as in Fig. 13, but the secondary maximum has dwindled to a modest peak which, in fact, would not be visible at all but for the preceding minimum which corresponds to the summer heat. For the coastal rural areas the primary maximum has gone as far back as June, and there now appears a secondary maximum in November–December emphasized by two adjacent minima. If weather were the sole essential factor there should have been no reason for differences for the urban and rural circles, nor would there be any such marked change in the character of the secondary maximum.

The only explanation that fits is that climate can be taken as the major influence only in the urban areas. In the rural areas it acts primarily through the harvest seasons. The essential crops are cereals (*Statistical Atlas of Bombay State*, 1950), in comparison with which the cash crops such as sugar-cane and cotton are of minor importance in the districts charted. The 'Kharif' crop is sown at the onset of the monsoon, say the end of June or beginning of July. The 'Rabi' crop is sown between the middle of August and middle of September on the plateau and about December on the coast. On the plateau the kharif crop is harvested in November and the winter crop sometime about the middle of January to the middle of February. This means that December on the plateau is most favourable for conception, from both the climatic and the economic point of view, with the greatest availability of food and least necessity for exertion. On the coast the main cereal is rice, which is harvested from September to November and the later crop is ready for harvest in March. It is seen that these harvest seasons correspond fairly well to the two clear-cut maxima in the birth-rate. It should be noted that the coastal crops are mostly consecutive on the same land, whereas the plateau crops are not. The sharp amelioration caused by the onset of the monsoon in the plateau cannot affect the rural population to the extent that it affects the city dwellers, for the simple reason that most of the peasants, men and women, have to work hard at that time in order to prepare the ground for sowing. The first ploughing is generally completed in summer. Second ploughing, harrowing and sowing have to be done with very primitive implements just after the beginning of the monsoon; which means strenuous exertion for the peasant at a time when the city-dweller a few miles away feels a pleasant relief.

The connexion with major holidays is definitely absent. Throughout the country, the two main Hindu holidays, Holi and Puja, are roughly at the end of March and during October respectively. The former, which is current throughout the north and east of the country, has also homologues for the whole of the peninsula. It is definitely of an orgiastic character, being the ancient spring festival and Saturnalia. If anything, it is in that particular period that one should expect a maximum for conceptions, particularly in the villages. From available data no part of the country, except possibly cities in Orissa, Sind, the North-West Frontier Province (and perhaps rural Konkan), shows a birth maximum in December–January. The Hindu holidays are of negligible importance for Sind and the North-West Frontier Province, and the latter has the most rigorous winter of any Indian province. The conceptions, according to our data, may roughly be taken to reach peak levels at the two solstices, but there are no corresponding solstitial festivals. Under these circumstances we see no alternative to the theory proposed in the earlier paragraphs.

6. The absence of reliable data forces an investigator to consider primary sources as far as possible, provided of course they are of a sufficiently representative character. It was as a step in that direction that data for Poona and Bombay were gathered from the respective municipal registers. In the first case, discussed in detail in this section, the recording was done by Taskar, from the municipal registers of Poona city. In Poona, the civic population came under three separate authorities, of which the city municipality and the much smaller suburban municipality are now combined into a single Corporation. The Cantonment (the legendary Poona of retired British army colonels) still remains separate under its own board with a somewhat anomalous military control. Out of these three boroughs, Taskar's record covers only the first, which, however, accounts for more than half the total population. In the records a certain number of still-births had also been included, and the totals were often erroneous. The necessary corrections were duly made by Taskar. Neither in Bombay nor Poona nor any other city is there any compulsory recording of all births. The primary aim of the records till now seems to be to help the authorities in issuing a legally recognized birth certificate. Though according to law all births must be reported to the authorities within a week after their occurrence, they may be registered at any time within 6 months or even a year of the actual date of birth; the only penalty is the non-issuance of the certificate after a year, whereas not reporting a death would be treated as a crime. It follows from this practice that though the date of birth may be correctly recorded and the birth certificate properly issued, the actual birth of the child is usually reported for statistical purposes as on the date on which the entry was made. For Bombay city, the problem is somewhat further complicated by the fact that the central municipal registers are copied from returns from the ten registration districts, thereby introducing an average time-lag of about a week in transit. For Hindu children the name is given on the twelfth day after birth, which may be another cause for delay, even though most parents decide upon the name earlier and some may register it in anticipation. There is nothing for the vast mass of population in India, whether in cities or in rural areas, corresponding to the parish or local registers of Europe. Clerks are assigned by the Bombay Municipality to various wards with the specific duty of investigating and recording births in the city, and it is gathered from the municipality that most of the births are recorded in this way. But on account of general public indifference and lack of any regular method by which the clerical agent can obtain quick notice of a birth, the time-lag is not materially reduced. Only the total number of recorded births would presumably be nearer to the actual.

The striking feature about the Poona city birth records is the clear-cut secondary maximum. Fig. 15 shows this quite unmistakably. Direct analysis of variance and fitting by orthogonal polynomials confirms this, though it will be seen from the last section of this paper that this procedure would not be justified theoretically. The correct method will be shown to be the fitting of a certain periodic regression on monthly totals made from the logarithms of individual weekly entries and not from the original figures themselves. For this purpose the procedure we followed was first to group the daily figures week by week and then consider the logarithms of the weekly figures. No unit smaller than 1 week has any meaning, in view of the fact that both Poona and Bombay have a weekly rhythm imposed upon them by the Sunday holiday that was introduced under the British régime for all government and municipal offices, schools and major business enterprises, though not the factories. Further, since it is the upper classes that are best represented in these registers, it follows that a fictitious weekly rhythm would necessarily be found in the records. The direct analysis of the data week by week would have given nothing of any special importance, seeing the excessive variation shown by the graphs of the weekly grand totals. The question arose as to the advisability of discussing the variation in terms of bi-weekly or 4-weekly units, and we found, after analysis of variance with subsample groupings, the last to be the most suitable. In any case, the last day of the year, or the last 2 days in a leap year, were discarded, which is a trivial loss. The remaining 364 days were conveniently grouped into 13 months of 4 weeks each for our purpose.

The fitting of a periodic regression to the Poona city data on the basis of 4-weekly sums of logarithms of weekly entries still shows an unmistakable secondary maximum in March (Fig. 16). Some effort had then to be made for tracing the causes of the seasonal variation. We thought

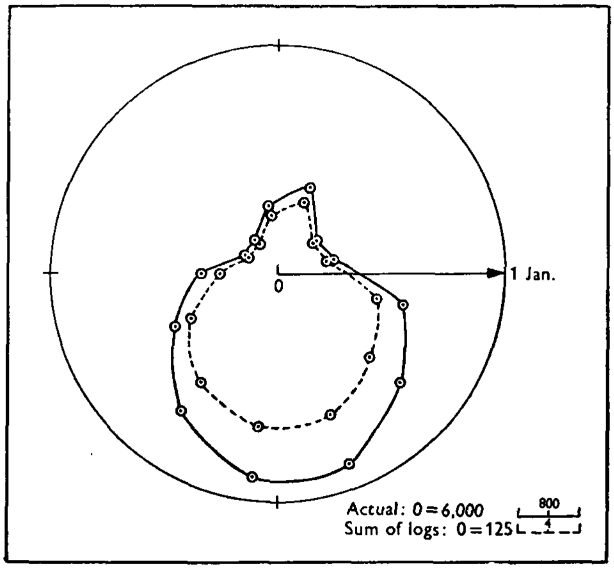

Fig. 15. Total 4-weekly number of live births in Poona city and total 4-weekly sums of logarithms of weekly entries for the years 1931–46.

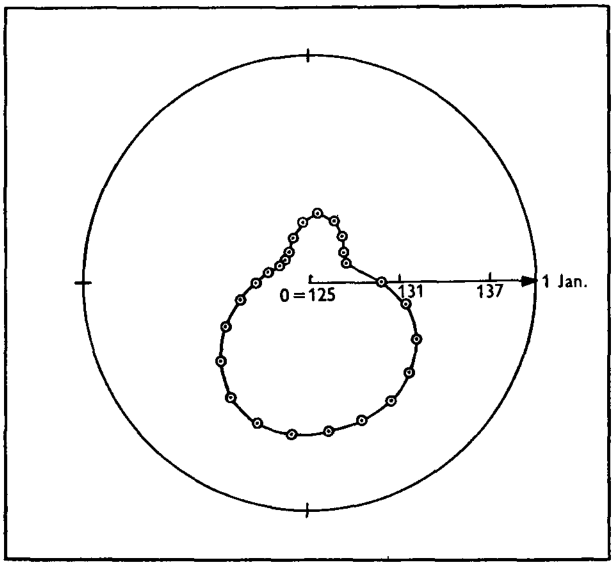

Fig. 16. Total 4-weekly sums of logarithms and thirteen other mid-terms as calculated from the periodic regression equation for Poona city.

that an immediate biological control could be had from birth records in the surrounding villages. These were obtained by special kindness of Mr M. R. Yardi, M.A., I.C.S., the District Magistrate and Collector of Poona District. He had them copied by the village officials (Patil, Talathi) who are responsible for sending the data to the Mamlatdars of the various talukas into which Poona (like every other Indian district) is subdivided for administrative purposes. Again these records clearly show the miserable quality of even the primary data over the greater part of the country. The filthy, badly scribbled and often illegible registers (with totals over three figures generally erroneous) which we saw are identical with the form in which the data are sent on to higher authorities, and seem but a poor foundation for the weighty announcements often made by scholars analysing Indian vital statistics. We took the individual record for each village, for each month of every recorded year, separately and made our own totals. In a few cases, the well-meaning official had made thirteen entries for the year with an additional column for totals, and since no one could explain just what this meant the entire village was omitted. In some cases there were gaps which could only indicate deficient entries and not the absence of births. When these were heavy we were also forced to omit the corresponding villages. In this way, out of the reports for 124 villages in Haveli Taluka (Poona District) we had to discard the data for nineteen villages altogether. The data, which was separately given for ten villages in the suburbs of Poona city, for the years 1920–40 are represented in Fig. 17, which shows a maximum in October. We do not emphasize the low secondary maximum in March–April which may not be significant, considering the paucity of the actual entries. On the other hand, for the 105 villages in the next administrative unit surrounding Poona city (Haveli Taluka) as shown in Fig. 18, both the October maximum and the secondary in March–April are unmistakable. But here the secondary is far less pronounced than in the case of Poona city, where the maximum comes precisely at the sixth fortnight in the year. The minimum seen in December for Haveli Taluka is fictitious, being due to the fact that many of the older village registers tend to drop out an entry or two towards the end of the year.

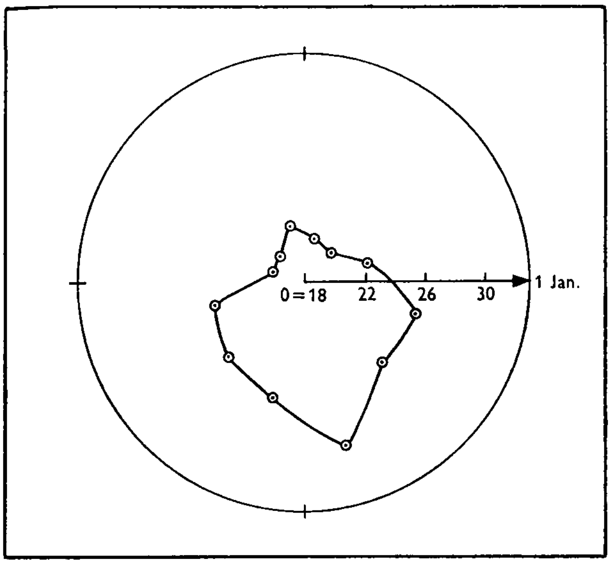

Fig. 17. Normalized total daily number of live births in ten villages round Poona city for the years 1920–40.

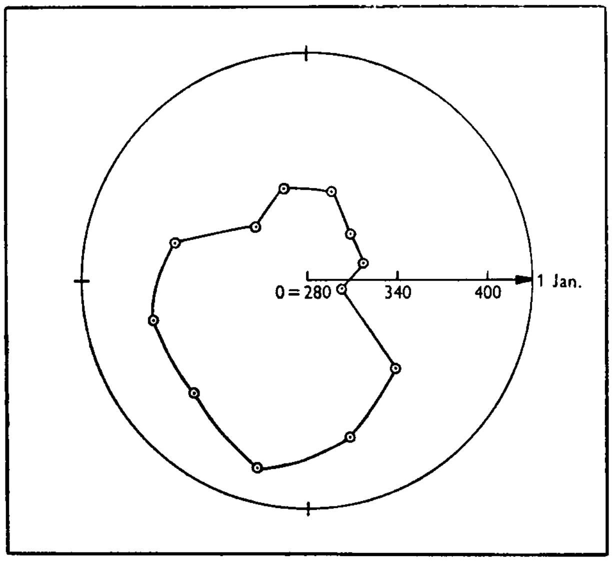

Fig. 18. Normalized total daily number of live births in 105 villages in Haveli Taluka (Poona District) for the years 1900–40.

The probable explanation is that the annual totals to be sent on to some higher official were either made as soon as the December figures were known or perhaps even made without the figure for December which was not entered at all, being thereafter unnecessary! It should be added that there is no monthly record actually kept in the villages except perhaps in an unusually literate one. The records as they exist are made at headquarters, separately for each village, but on the basis of the information supplied orally by the village Patil; and as so much of his attention is taken up by other administrative duties, his figures for the births are from his memory, occasionally several months out. Our major conclusion then is supported in this case; namely, that climate has a greater direct influence when the economic conditions are reasonably steady all the year round, as would be expected in a city. In rural areas the influence of the climate may be counteracted by economic conditions, specifically in the energy spent during sowing and harvesting times, which are controlled in India by the monsoon season. For Poona city and its rural environs, the two maxima occur at about the same time, but the shapes of the curves are markedly different.

7. The problem for Bombay city was mainly that of collecting data. The clerical help that we could command was not sufficient for going over the records in detail and entering each recorded live birth against its proper date. As a preliminary measure, utilizing the part-time services of

Messrs V. H. Bhate and B. N. Nachnolkar (which was all that our resources permitted), all the figures from the extant municipal registers were copied down. The actual records for Bombay city go as far back as 1869, but the entries before 1903 are decidedly unreliable, as may be seen at once from the bad state of the registers. From 1903 onwards the records seem to be quite well kept, except for the fact that entries are made as the information came in, and not against the actual date of birth. Further, a considerable number of mothers go to their parents' home for the lying-in period and the effect of this cannot be assessed. The suburban population (Kurla, Ghatkopar, Bhandup, Bandra, Khar, Jogeshwari, etc.) is about a sixth of the city population according to the last census figures. Some of the mothers from the suburbs also deliver in the Bombay city maternity

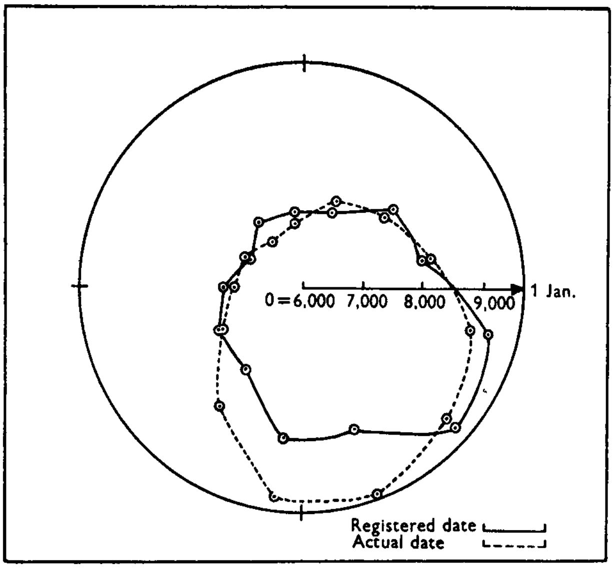

Fig. 19. Comparison of total 4-weekly number of live births in Bombay city for the years 1910, 1920, 1930 and 1940 as tabulated according to: (i) the actual date (A) of birth and (ii) the registered date (R).

homes, but here again the effect is of unknown magnitude. We took four sample years in which (Table 3, Fig. 19), by actual day-to-day analysis of the registers, two sets of data were compiled, one by the recorded date and the other by the actual date of birth. Fig. 19 will show that the two graphs are similar on the whole and that there is only a rotation of the order of 3–4 weeks approximately.

The actual figures are often slightly greater than recorded figures for the same time interval, as would be expected from the steady growth of the population (Fig. 22) and the time-lag introduced by representing the actual births at some earlier period by the recorded figures. But a glance at Table 3 will show that this is not always so. The time-lag is certainly not uniform, for the 5 months or so from the fourth generally show more births recorded than the actual number, even if we allow a month for the interval between birth and registration. The maximum and minimum (in totals) are clearly shifted by a period of about 4 weeks, but otherwise the shift is not uniform and the shape of the curve would be changed to a certain extent, for which, again, no allowance has been made. The split-plot analysis at the lower halves of Tables 3*a* and *b* show that the error term is reduced by shifting the recorded data back by 1 month; but if this alone could account for the difference between actual and the recorded numbers the MT term should also

Table 3. *Four-weekly totals of live births in Bombay city during the* 4 *years* 1910, 1920, 1930 *and* 1940 *grouped according to actual* (A) *and registered* (R) *dates of birth*

1910		1920		1930		1940		Total	
A	R	A	R	A	R	A	R	A	R
1,644	1,542	1,527	1,509	1,911	1,948	3,093	3,058	8,175	8,057
1,443	1,553	1,440	1,509	1,753	1,916	3,151	3,007	7,787	7,985
1,452	1,376	1,323	1,278	1,895	1,779	2,874	2,903	7,544	7,336
1,326	1,420	1,266	1,340	1,861	1,659	2,610	2,814	7,063	7,233
1,338	1,358	1,280	1,326	1,788	1,841	2,486	2,768	6,892	7,293
1,366	1,396	1,314	1,290	1,803	1,727	2,573	2,531	7,056	6,944
1,507	1,443	1,263	1,322	1,861	1,837	2,463	2,702	7,094	7,304
1,597	1,534	1,473	1,436	1,861	1,859	2,549	2,698	7,480	7,527
1,515	1,530	1,582	1,473	2,076	1,939	3,185	2,724	8,358	7,666
1,775	1,650	1,856	1,584	2,284	2,085	3,548	3,188	9,463	8,507
1,767	1,677	1,737	1,584	2,414	2,178	3,663	3,064	9,581	8,503
1,796	1,792	1,748	1,765	2,325	2,345	3,366	3,520	9,235	9,422
1,597	1,711	1,669	1,683	2,190	2,313	3,383	3,487	8,839	9,194
20,123	19,982	19,478	19,099	26,022	25,426	38,944	38,464	104,567	102,971

Table 3 (*a*). *Analysis of variance of data in Table* 3 *according to actual date and registered date of birth*

Y = years. M = months. T = 'treatments' A × R.

Source	D.F.	Sum of squares	Mean square	F
Y	3	37,210,990·8	12,403,663·60	529·759***
M	12	4,313,919·2	359,493·26	15·354***
YM	36	842,896·7	23,413·80	
T	1	24,492·5	—	2·840
YT	3	4,320·5	1,440·17	$(5{\cdot}988)^{-1}$
MT	12	358,110·5	29,842·54	3·460***
YMT	36	310,461·5	8,623·93	—
Total	103	43,065,191·7		

Table 3 (*b*). *Analysis of variance of data in Table* 3 *with registered figures shifted back a month*

Source	D.F.	Sum of squares	Mean square	F
Y	3	37,210,990·8	12,403,663·60	463·425***
M	12	4,414,309·4	367,859·12	13·744***
YM	36	963,547·5	26,765·21	—
T	1	24,492·5	—	4·645***
YT	3	4,320·5	1,440·17	$(3{\cdot}661)^{-1}$
MT	12	257,720·3	21,476·69	4·073***
YMT	36	189,810·6	5,272·51	—
Total	103	43,065,191·7		

have become negligible, instead of which its significance rises. Even so, we were forced to take the recorded figures as they stood for want of sufficient clerical assistance.

The peculiar feature about Bombay city which may be emphasized is that the *fitted* graph (Fig. 21) shows the secondary maximum quite clearly though slightly, whereas there is no reason for suspecting its existence from the graph of the 13-month data (Fig. 20). In the fortnightly and weekly graphs, however, the secondary is visible, but its significance would have seemed rather doubtful because of the increased fluctuations. The theoretical curve is fitted on the basis of the

Table 4. *Analysis of variance for Bombay city for the* 38 *years* 1903–40

Source	D.F.	Sum of squares	Mean square	F
Y	37	2,031,691,121·6	54,910,570·85	384·702***
M	12	486,582,490·5	40,548,540·87	284·082***
YM	444	116,375,390·8	262,106·74	1·836***
Error	1482	211,534,011·5	142,735·50	—
Total	1975	2,846,183,014·4	1,441,105·32	—
Rest (YM + error)	1926	327,909,402·3	170,254·10	—

Table 5. *Analysis of variance for Poona city for the* 16 *years* 1931–46

Source	D.F.	Sum of squares	Mean square	F
Y	15	1,208,438,017·5	80,562,534·50	341·208***
M	12	158,118,060·4	13,176,505·03	55·807***
YM	180	115,438,192·3	641,323·29	2·716***
Error	624	147,332,440·2	236,109·68	—
Total	831	1,629,326,710·4	1,960,681·96	—
Rest (YM + error)	804	262,770,632·5	326,829·91	—

Table 6. *Analysis of variance for Poona city for the* 10 *years* 1931–40

Source	D.F.	Sum of squares	Mean square	F
Y	9	387,761,416·3	43,084,601·81	153·437***
M	12	81,923,858·3	6,826,988·19	24·313***
YM	108	87,264,451·0	808,004·18	2·878***
Error	390	109,510,495·2	280,796·14	—
Total	519	666,460,220·8	1,284,123·74	—

13-month data of logarithms of weekly figures, and the non-significant terms of the periodic regression were omitted, as in the case of Poona. The significance in each case was decided, however, not by YM interaction (D.F. 444 and 180 respectively) only, but on the basis of the 'rest' mean square (D.F. 1926 and 804; Tables 4 and 5) calculated from weekly entries. This, in effect, amounts to a slight overfitting, seeing that the YM interaction gives a greater mean square than that from the remaining degrees of freedom in the total error. Therefore, fitting to the last possible term could only give a theoretical curve which would pass through all the observed points and would give no fundamental information.

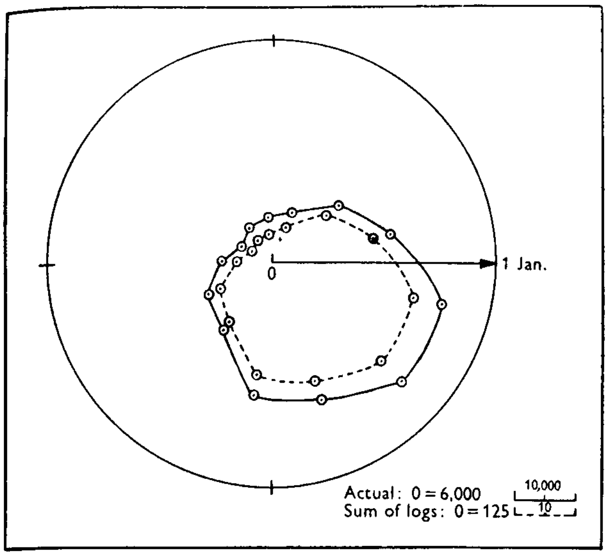

Fig. 20. Total 4-weekly number of live births in Bombay city and total 4-weekly sums of logarithms of weekly entries for the years 1903–40.

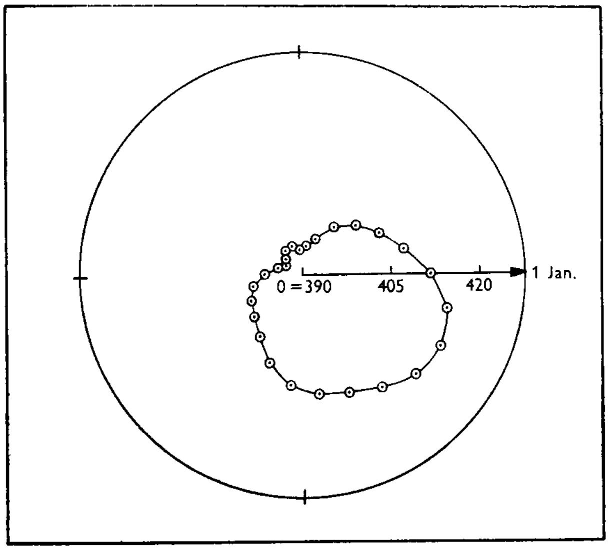

Fig. 21. Total 4-weekly sums of logarithms and thirteen other mid-terms as calculated from the periodic regression equation for Bombay city.

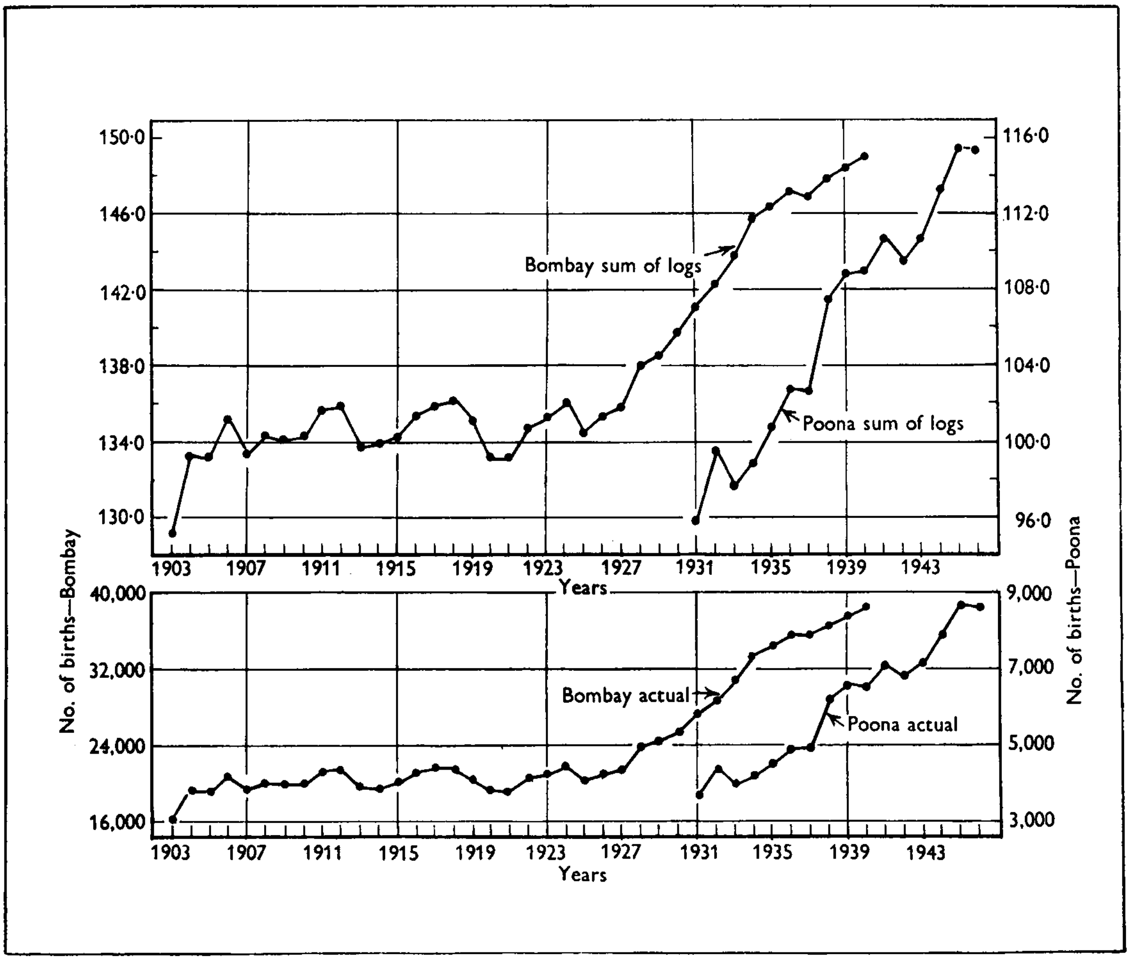

Fig. 22. The general trend in the yearly number of live births and in the yearly sums of logarithms of weekly entries in Bombay and Poona cities for their respective periods.

8. *Theoretical considerations on curve-fitting.* The standard methods of curve-fitting by Chebyshov-Hermite polynomials are motivated by the occurrence of these polynomials in a natural way, in connexion with the normal distributions (Kendall, 1943). They have the additional property of being orthogonal functions for summation over a finite number of suitably fixed points, as well as orthogonality in the usual function-theoretic sense over an infinite range when weighted with the error-function. However, these properties are not restricted only to such polynomials, as may be seen from the following considerations. Given a finite sequence of points $\{x_n\}$ on the real line, monotonic and distinct, $x_{n+1} > x_n$, we propose the following definition:

A set of real functions $\phi_i(x)$, defined over a common interval including $\{x_n\}$ is orthogonal over $\{x_n\}$ if $\sum_r \phi_i(x_r)\,\phi_j(x_r) = 0\ (i \neq j)$; and normal if $\sum_r \phi_i^2(x_r) = 1$; when both properties are satisfied, the set $\{\phi_i(x)\}$ is orthonormal over $\{x_n\}$.

We now proceed to derive a few almost obvious but quite essential theorems. The simplest such orthonormal sequence is obviously $\phi_k(x) = 1(x_k \leqslant x < x_{k+1})$ and $\phi_k(x) = 0$ elsewhere. These are also orthogonal functions for integration, with weight unity. In fact, they are the most commonly used statistical functions for presenting data in the form of a histogram. But their orthogonality properties are of little use, inasmuch as no possibility arises of fitting any smooth curve or of extrapolation by means of such functions. Moreover, the full set is necessary for any representation, when $\{x_n\}$ is fixed in advance; the property of successive approximations over the whole basic interval which allows the fitting to be stopped at a suitable stage does not exist. The theoretical advantages of Haar or Rademacher functions in analysis are also absent, seeing the practical nature of the problem.

Clearly, for the intervals defined by $\{x_n\}$, there can be no unique choice of a 'best' set of functions, as is shown by the following

THEOREM 1. *Any set of $n+1$ real functions $f_n(x)$ of a real variable x, whose values are linearly independent at the points of a given sequence $\{x_n\} = x_0 < x_1 < \ldots < x_k < \ldots$ can be replaced by linear combinations thereof which are orthonormal over $\{x_n\}$.*

Proof. The procedure is analogous to ordinary orthonormalization of integrable functions with integrable square. None of the functions, by assumption of linear independence, can vanish at all points of $\{x_n\}$. So we set $\phi_1 = \lambda f_1$, λ being determined by $\Sigma\phi_1^2(x_n) = 1$, i.e. $\lambda^2 = 1/\Sigma f_1^2(x_n)$, where summations are always over all points of $\{x_n\}$ unless otherwise restricted. If an f_2 is already orthogonal to ϕ_1 it may be normalized at once. If not, take $\phi_2 = \mu(\phi_1 + bf_2)$. Orthogonality conditions determine b by $1 + b\Sigma\phi_1(x_n)f_2(x_n) = 0$, where the summation is not zero since f_2 is not orthogonal to ϕ_1 by hypothesis. The coefficient μ is determined as before to make $\Sigma\phi_2^2(x_n) = 1$, and this is always possible seeing that $\phi_1 + bf_2 = \lambda f_1 + bf_2$ cannot vanish at all points of $\{x_n\}$ by the hypothesis of linear independence, no matter what the values of λ, b (neither being zero by the method followed). Then ϕ_3 can be taken as $\nu(\phi_1 + c\phi_2 + df_3)$, and the three coefficients calculated by $\Sigma\phi_1\phi_3 = \Sigma\phi_2\phi_3 = \Sigma\phi_3^2 - 1 = 0$. This proves the theorem and gives an actual procedure for the orthonormalization.

THEOREM 2. *No matter how the set was obtained, the values of any set of orthonormal functions must necessarily be linearly independent over the points of the basic sequence $\{x_n\}$.*

Proof. For, if not, we would have one function, say ϕ_1, which could be expressed in terms of the rest:

$$\phi_1 = a_2\phi_2 + a_3\phi_3 + \ldots + a_k\phi_k.$$

The equality being valid over all $\{x_n\}$. Multiply both sides by ϕ_1 and sum. The left-hand side gives $\Sigma\phi_1^2 = 1$, and the right vanishes identically, which leads to a contradiction.

THEOREM 3. *If the set of orthonormal functions $\phi_i(x)$ be equal in number to the points of $\{x_n\}$, then*

$$\det |\phi_i(x_j)| = \pm 1.$$

Proof. Multiply the square matrix $\| \phi_i(x_j) \|$ by its transpose. The general term in the product matrix is $\Sigma\phi_i(x_k)\,\phi_j(x_k) = \delta_{ij}$, whence the product is the unit matrix, and the square of the original determinant equal to unity. Such a set of orthonormal functions may be called a *full* set. Clearly, we have

THEOREM 4. *Any real-valued function may be exactly represented at points of $\{x_n\}$ as a linear combination of any full set of orthonormal functions.*

Proof. Let $\phi_i(x)$ be a full set. Then for an arbitrary $f(x)$ we may set

$$f(x) = \Sigma a_i \phi_i(x)$$

and determine the coefficients as usual; by multiplying both sides by $\phi_k(x)$ and summing over all $\{x_n\}$

$$\Sigma f(x_r)\,\phi_k(x_r) = \Sigma\Sigma a_i \phi_i(x_r)\,\phi_k(x_r) = a_k.$$

The actual representation only says that

$$f(x_r) = \sum_i a_i \phi_i(x_r).$$

Since we have $\det |\phi_i(x_r)| = \pm 1$, the equations admit a unique solution for the a_i's in terms of $f(x_r)$, and the determination is exact.

However, orthogonality gives the advantage that the *best* representation in the sense of least squares involves each function separately; in statistical terminology, the *covariance* vanishes for any two orthogonal functions. We have the analogue of Bessel's inequality as follows:

THEOREM 5. *The coefficients $a_k = \Sigma_r f(x_r)\,\phi_k(x_r)$ are precisely those which minimize the least-square approximation $\Sigma(f(x) - \Sigma a_i \phi_i(x))^2$ for ϕ_i orthonormal over $\{x_n\}$.*

This is proved immediately by differentiation with respect to a_i. It is basic here in allowing least squares and the analysis of variance to be used.

The essential is now seen to be the deduction of a set of *continuous* functions which are orthonormal over a basic set of points and thereby allow both fitting of the discrete data, and extrapolation by means of a fitted curve, if necessary. We utilize the preceding almost self-evident theorems for the special conditions of our problem.

Though the data of our problem appears as a time-series, we have a basic assumption, namely, that the figures are collected as samples from a population (of births) which appears in nature in the form $P(x)\,R(x)$, where x is the time and $P(x)$ a periodic factor which represents the seasonal variation in which we are really interested, $R(x)$ being the term which must here represent a steady increase due to parent immigration and the excess of births over deaths. In this form, fitting would be rather difficult, and the analysis of variance could not be applied to the two portions separately. It is essential, therefore, to make the two terms additive by taking logarithms whereby the fundamental regenerative process will appear as $p(x) + r(x)$. This is the reason for taking logarithms of the entries for Bombay and Poona, the unit entry being per week.

We have now to treat the problem on the basis of a periodic and an additive non-periodic portion, the basic period being known. This basic period we call a 'year', and the points of fitting $\{x_n\}$ within the year are designated as 'months', no matter what the actual time-units might be. Finally, the problem is simplified by taking the months as equally spaced within the year, which is in fact the situation for our grouping units from daily records, and approximately true for the rather unevenly adjusted Gregorian calendar months.

Let the data be known for μ months for each of ν years. Then the preceding treatment shows that not more than μ functions would be of any use, and naturally they must be periodic as well as continuous. Therefore *our basic set of periodic functions are given by* 1, $\cos\theta x, \cos 2\theta x, \ldots, \sin\theta x, \sin 2\theta x, \ldots$, where $\theta = 2\pi/\mu$, *there being m cosines and m sines if* $\mu = 2m+1$; *and m cosines,* $m-1$ *sines if* $\mu = 2m$.

Proof. The cosines being even and the sines odd functions, we have orthogonality between the two sets when summing products over the whole period. Properties of the μth roots of unity can be used to prove orthogonality within the sets of cosines and sines.

(1) $\sum_{x=0}^{\mu-1} \cos k\theta x = 0$, for this is (no matter what k fixed other than zero might be) the sum of the real parts of the roots of an equation of type $z^n - 1 = 0$. If k is relatively prime to μ, kx runs through the complete set of residues modulo μ as $x = 0, 1, \ldots, \mu-1$; if not, k being a divisor of μ and less than μ, $k\theta$ is a μ/kth root of unity and the multiples kx run through the complete set of residues modulo μ/k, again giving the same result. The only difference in the argument is that of the exponent n in $z^n - 1 = 0$. Thus the constant term is orthogonal to all the other periodic terms.

$$\sum_{x=0}^{\mu-1} \cos k\theta x \cos l\theta x = \tfrac{1}{2}\Sigma \cos \overline{k+l}\theta x + \tfrac{1}{2}\Sigma \cos \overline{k-l}\theta x = 0, \tag{2}$$

for $k \neq l$, $k+l<\mu$; both sums on the right vanishing by the previous argument. The case $k+l \geqslant \mu$ does not arise by the limitations on our basic set. Therefore the cosine terms are orthogonal.

The formula
$$2\sin ax \sin bx = \cos\overline{a+b}x - \cos\overline{a-b}x$$
shows the same result to hold within the set of sines.

It only remains to normalize the set of periodic functions, for which we need the square roots of divisors
$$\Sigma 1^2 = \mu; \quad \Sigma\cos^2 k\theta = \tfrac{1}{2}\mu, \quad \Sigma\sin^2 k\theta = \tfrac{1}{2}\mu.$$
This completes the analogy with ordinary Fourier series. The orthonormal set is therefore $1/\sqrt{\mu}$, $\sqrt{(2/\mu)}\cos k\theta x$, $\sqrt{(2/\mu)}\sin k\theta x$, where $\theta = 2\pi/\mu$ and $k = 1, 2, \ldots, [\frac{1}{2}\mu]$, the integer μ denoting the number of equally spaced 'months' within the basic interval which we call a 'year'.

The problem does not concern itself with periodogram analysis, as the fundamental period is assumed to be known as one year. Any procedure of standardizing the original data, say by dividing by the successive annual totals, would throw away our most powerful tool, the analysis of variance. Therefore, the main assumption is that the observed figures $b(x)$ *in logarithmic units* can be fitted by means of $p(x)+r(x)$, where the first portion is periodic.

The major effects of interest for the particular problem are the seasonal rhythm, and the trend over the years. The constant term, being taken in the periodic regression, may be omitted from $r(x)$ altogether, and in fact shall be assumed to have been completely eliminated from the fitting by measuring $b(x)$ from the general mean. We further break up $r(x)$ into functions $r_n(x)$ which are orthogonal among themselves. Further, these can all be made orthogonal to the components $p_m(x)$ of $p(x)$ by the simple expedient of taking each $r_n(x)$ to be constant for all the months in any given year, though differing from year to year. Thus, the annual totals or means alone can be fitted by $r(x)$. The total sum square, which is to be assumed minimized by the fitting, is

$$S = \sum_{x=1}^{\mu\nu} [b(x) - \sum_m \alpha_m p_m(x) - \sum_n \beta_n r_n(x)]^2.$$

Here, x is the generic point representing a given month of a given year, $b(x)$ the observed value in log units, $p_m(x)$ the periodic functions with 1-year period, hence the same for a fixed month, no matter what the year; $r_n(x)$ are the annual regression functions which are constant within any fixed year, as explained. Thus, if the points x are displayed in a rectangular array as x_{ij}, where i denotes the year (row) and j the month (column), our regression functions may be regarded as $p_m(x_{.j})$ and $r_n(x_{i.})$, where the dot indicates the suppression of the corresponding index, the variable x then being only a column of a row variable. Our summation then becomes

$$S = \sum_{i=1,}^{\nu} \sum_{j=1}^{\mu} \left\{ b(x_{ij}) - \sum_m \alpha_m p_m(x_{.j}) - \sum_n \beta_n r(x_{i.}) \right\}^2.$$

Let the successive monthly and annual means be denoted by

$$b_{.j} = \frac{1}{\nu} \sum_{c=1}^{\nu} b(x_{ij}), \quad b_{i.} = \frac{1}{\mu} \sum_{j=1}^{\mu} b(x_{ij}).$$

Then we apply the standard formula for reducing a sum of squares to the sum of squared deviations from the mean

$$\sum_{s=1}^{k} y_s^2 = \sum_{s=1}^{k} (y_s - \bar{y})^2 + k\bar{y}^2, \quad \bar{y} = \frac{1}{k} \sum_{s=1}^{k} y_s.$$

Inasmuch as the general mean is zero we get

$$S = \nu \sum_j (b_{.j} - \sum_m \alpha_m p_m(x_{.j}))^2 + \mu \sum_i (b_{i.} - \sum_n \beta_n r_n(x_{i.}))^2 + \sum_i \sum_j (b(x_{ij}) - b_{i.} - b_{.j})^2.$$

The three terms on the right are due respectively to fitting monthly means by $p_m(x)$, annual means by $r_n(x)$, and a residual interaction term. These correspond precisely to analysis of variance into M + Y + YM, with degrees of freedom $(\mu - 1) + (\nu - 1) + (\mu - 1)(\nu - 1) = (\mu\nu - 1)$. We have thus proved

THEOREM 6. *Fitting separate regressions to the monthly and annual means (or totals) is justified provided the YM interaction is negligible.*

Where the YM mean square is itself the measure of error, the data being given by months, our theorem gives the complete result. In the better classes of data (namely, Bombay and Poona cities), we have a separate residual mean square because of weekly grouping. In our two particular cases, the YM term does appear significant as against the residual, but even so the YM mean square is very much smaller than either Y or M, and if the three be compared among themselves, Y and M are very highly significant as against YM. Thus the use of the total residual error to omit any terms from the regression which are not significant is justified, and obviates the danger of overfitting.

For the general case, there is no 'best' choice of functions that presents itself so clearly. If orthogonality to the periodic portion be demanded for functions $r_n(x)$ which are not constant over the months within a year, we shall in general have the periodic terms appearing in the orthogonalization of any set of functions given as a basis for $r_n(x)$ unless they all assume monthly values proportional to some one $p_m(x)$ which may be eliminated from $b(x)$ altogether. We took this as the simplest, namely, the constant term. For the most general type of fitting, it may be convenient to take the origin at the centre of the given set of points. If the total number is even, it is easier to take the periodic functions in terms of odd multiples of $\frac{1}{2}\theta$, instead of all multiples of θ.

The precise role of the analysis of variance is illustrated simultaneously with the necessity of taking a sufficient number of years by Tables 5 and 6. The second of these shows the analysis for

Poona city birth figures (Taskar) for the ten years 1931–40 as against the 16 years 1931–46 in Table 5. The additional 6 years reduce the residual error mean square term by about 16 %, but the really substantial gain is in the reduced magnitude of the YM term which is neglected in fitting Y and M by our two regressions; and even more, in the increased values for sum squares, mean squares,

Table 7. *Significant periodic terms and the corresponding regression equation for Bombay city*

1	2	3	4	5	6	7
No.	Term	Regression sum of squares	Deviation from regression	D.F. (for col. 4)	Mean square	F
1	$\cos\theta x$	237,164,061·1	249,418,429·4	11	22,674,402·67	—
2	$\sin\theta x$	229,192,698·0	20,225,731·4	10	2,022,573·14	—
3	$\cos 2\theta x$	3,399,523·7	16,826,207·7	9	1,869,578·63	—
4	$\sin 2\theta x$	10,393,926·6	6,432,281·1	8	804,035·14	—
5	$\cos 3\theta x$	828,271·4	5,604,009·7	7	800,572·81	—
6	$\sin 3\theta x$	970,729·3	4,633,280·4	6	772,213·40	4·536***
7	$\sin 5\theta x$	3,067,283·8	1,565,996·6	5	313,199·32	1·840
8	$\cos 5\theta x$	696,898·2	869,098·4	4	217,274·60	1·276
	Rest mean square (D.F. 1926)				170,254·10	

Regression equation:

$$Y = 402{,}7747{\cdot}9231 - 74{,}471{\cdot}4063\cos\theta x + 8916{\cdot}0840\cos 2\theta x - 4401{\cdot}0008\cos 3\theta x + 4036{\cdot}9167\cos 5\theta x + 73{,}209{\cdot}1735\sin\theta x - 15{,}590{\cdot}3167\sin 2\theta x + 4764{\cdot}4655\sin 3\theta x + 8469{\cdot}1943\sin 5\theta x.$$

Table 8. *Significant periodic terms and the corresponding regression equation for Poona city*

1	2	3	4	5	6	7
No.	Term	Regression sum of squares	Deviation from regression	D.F. (for col. 4)	Mean square	F
1	$\cos\theta x$	4,162,122·7	153,955,937·7	11	13,995,994·33	—
2	$\sin\theta x$	22,196,606·3	131,759,331·4	10	13,175,933·14	—
3	$\cos 2\theta x$	123,505,893·4	8,253,438·0	9	917,048·67	2·806***
4	$\sin 3\theta x$	4,449,921·8	3,803,516·2	8	475,439·53	1·455
5	$\sin 4\theta x$	1,588,533·3	2,214,982·9	7	316,426·13	$(1{\cdot}033)^{-1}$
	Rest mean square (D.F. 804)				326,829·91	

Regression equation:

$$Y = 130{,}6570{\cdot}5384 - 6401{\cdot}6326\cos\theta x - 14783{\cdot}4773\cos 2\theta x + 34872{\cdot}0236\sin\theta x + 6619{\cdot}2608\sin 3\theta x - 3954{\cdot}8633\sin 4\theta x.$$

and even the F ratios for the main effects Y and M. Though not illustrated here, the same effect appeared for Bombay data for 20 years as against all 38 years. This is one reason why we do not attach more than qualitative importance to the *Gazette* figures for urban areas, which cannot cover more than 10 years' data each, if that much, because of the numerous gaps in the published entries.

Tables 7 and 8 give the significant periodic terms and their corresponding coefficients for Bombay and Poona respectively. The values calculated from these regressions, including the mid-terms, are cited in Table 9.

Table 9. *Four-weekly sums of logarithms of weekly entries for Bombay and Poona cities, including thirteen other mid-terms as calculated from the corresponding periodic regression equations*

Bombay city		Poona city	
Observed	Calculated	Observed	Calculated
407·4951	407·3969	128·3008	128·3706
	404·5108		127·7366
402·0360	402·1631	127·9181	128·0062
	399·4571		128·7789
396·3825	396·5116	130·0303	129·4302
	394·5292		129·5105
394·6984	394·2509	128·7425	128·9886
	394·8610		128·1910
394·3521	394·8636	127·2542	127·5270
	393·8950		127·2371
393·4923	393·1782	127·1595	127·3425
	394·0119		127·7785
396·0203	396·1829	128·7835	128·5385
	398·2826		129·6745
399·7644	399·4856	131·4330	131·1583
	400·5061		132·7622
402·0336	402·5447	133·7967	134·1082
	405·6355		134·8822
409·0733	408·5428	135·0959	135·0419
	410·3135		134·8287
411·1647	411·4215	134·9564	134·5510
	412·9691		134·3110
414·9301	414·9648	133·2626	133·9099
	415·9504		133·0316
414·6295	414·5656	131·8083	131·5686
	411·1499		129·8189

9. In conclusion, we may summarize our findings as follows: It is essential to note that generally the basic data and its usual presentation are both deficient. The fundamental difference in our presentation consists of graphs by polar co-ordinates. All graphs in works cited here are based upon orthogonal Cartesian co-ordinates. Thereby the rhythmic feature of the phenomenon is concealed and similarities or differences between the shapes of the graphs are either lost or become misleading. Finally, the tendency is to fit theoretical curves to data in Cartesian co-ordinates by means of orthogonal polynomials. These make the situation still worse, because January and December then come at opposite ends of the graph instead of being essentially contiguous. Goodness of fit therefore cannot be tested as effectively; we can vouch for this from our personal experience of fitting such graphs from some of the same data presented in our work. It is this which has led to the new theoretical approach in §8. The deficiencies and the contradictions of the recorded figures have been discussed from stage to stage, and it would be seen that for qualitative results at least, the discrepancies are not too serious for the validity of our conclusions regarding seasonal variation. In all likelihood our graphs suffer from a retardation of about a month, and all maxima and minima could be taken as occurring a month earlier. On the other hand, the reliability of the data leaves a great deal to be desired. A few available entries copied by Taskar from the Poona Public Health Department registers had allowed a certain amount of discrimination to be made between the urban and rural phenomena. It is thus clear that the variation for the urban population depends directly upon the climate, which in India is dominated and characterized by the rainy season of 4 months. In the rural areas, the same season, and particularly its beginning and end (i.e. onset and break of the two monsoons), is still the dominating factor acting through local conditions of soil and crops. As regards birth-rate variation we find a marked difference between town and country which can only be explained by the gathering of the harvest or harvests and the strenuous labour attendant upon the hurried field work at the onset of the monsoon.

There is in our data no reason to believe that an absolute standard of living has any direct connexion with an absolute rate of conception; if this were so the poorest people in this country should not be able to breed at all, and in comparison with other countries hardly anybody here should be able to produce children except perhaps the tiny upper middle class. But this contradicts most of the known facts and all published assumptions about the birth-rate in India. The correct statement is that conceptions tend to reach a maximum at a time when there is a decided relative change for the better in the conditions of life. This, of course, must be taken for society as a whole and not for any one particular individual. The worst deficiency in our data is in the absence of an analysis by economic classes or categories, and so far as can be seen there exists very little possibility of it being available, except perhaps in some of the public maternity hospitals in the major cities of India. Bombay Municipality gives the occupations, but this would permit only the roughest sort of separation into categories. With the occupation, some index of employment or unemployment and of current prices of what are considered the necessities by that particular group of people would be necessary.

Connected with this inquiry, though in no way directly approached, is the problem of birth control and checking the growth of a population for which there is apparently no possibility of growing sufficient food in this country after the separation of Pakistan. The reports from China after 1949 seem to show that a redistribution of land would have a considerable effect in increasing the harvests immediately. *For birth control to be effective in India the parents must be induced to believe firmly that some method, other than begging and its concomitants, will be assured for their subsistence in old age, even in the absence of children.* So long as no such effective guarantee is forthcoming birth-control measures and propaganda will also remain ineffective. As things are at present no birth-control measures, however effective, would reduce the number of mouths to feed to within reasonable limits for several years to come. In addition, there is to be considered the effect of birth control—without a revolutionary change in the economy—upon the supply of labour and its complex relationship with the total crop harvested, the prevailing prices of cereals, of finished goods; and thereupon again with the general standard of living.

In any case no authority in this country seems willing to take upon itself the burden of promoting birth control, or even of backing its advocates with emphasis sufficient to provide some real check on the growth of the population. It may be considered that the religious demand for sons to perpetuate the cult of the dead and to provide a place in heaven for ancestors might effectively block any such move on the part of the politicians. This supposition may here be contradicted flatly. The religious question as settled by quotations from the Hindu scriptures really affects a very small percentage of the population, in most places less than 15 %. It happens that Hinduism has come to be synonymous with Brahminism, in spite of the vast internal differences it allows and encourages. Moreover, as is shown for example by the reported prevalence of female infanticide in some barren regions of the country, economic necessity does forcibly readjust local religious superstition to a certain consonance with reality. We have therefore to explain such superstitions as having not only historic but also at present an economic basis. The reasons for the pressing desire for progeny are quite clear if one looks for them. For the upper classes who have property to transmit, or whose mental and sociological orientation give them hopes of rising to that status, children are primarily to be viewed as providing legitimate heirs. The lower middle class and, of course, the vast majority of the population of this country suffer considerable privations whenever the family has to feed an additional mouth. In that case it is clear that there must be some primary economic necessity for children. This necessity may be discovered immediately by asking any villager (not of the landlord or rich peasant classes) just how he expects to live when he becomes too old to support himself by his labour. The answer will be that children are the sole possible guarantee that aged or infirm parents will get anything to eat. In fact, the ancient but still current rite of annual token food-offerings to the manes of departed parents can only be an extension of the practice of feeding senescent elders who could no longer manage for themselves. There is no other regular method for the vast majority of our population, which have no real estate, insurance, unemployment allowances, old age pensions, social security services, or private savings. A few might for a time eke out a painful subsistence by begging, which has long been a recognized occupation in this country for those who can no longer produce enough of the necessities of life, or possess any control over the means of production. High child mortality in the countryside combines with ignorance and superstition to inhibit control after a son or two have been born. Begging has become an extremely uncertain process for warding off starvation in these days of food shortage, which only emphasizes the need for children as a measure of security. We suggest that in the frequent discussions on birth control and India's population, officials and experts might well consider this rather obvious economic factor which seems to have escaped their notice.

Lastly, it is necessary to add a word of caution to prevent misleading reflexions as to the precise bearing of the work in this note. The seasonal variation might naturally lead serious-minded planners to try to devise methods by which conceptions could be controlled and minimized at certain fixed seasons. Such limited contraception will prove futile, as can be seen clearly from the actual differences in the nature of the graphs for the various regions. The greater maximum is never a sharp peak but a gentle rise above adjacent points; its suppression would logically mean nothing more than an increase in the conception rate for the succeeding months, provided the methods of contraception did not cause permanent sterility. It must be kept in mind that pregnancy is the most effective method of preventing conception! Further, the total range of monthly variation in the number of births, of about 15–20 % of the average, hardly indicates the existence of a sterile period in the year. On the contrary, it forces one to assert that the reproductive process runs through the whole year, showing peaks and depressions in periods when the conditions of life change markedly for the better or for the worse. Hence there is no reason to believe that conceptions artificially regulated at an optimum period will control the growth of population at large.

The data consisting of the weekly number of live births in Bombay City and in Poona City for the periods 1903–40 and 1931–46 respectively are placed in the archives of the Galton Laboratory.

REFERENCES

CHANDRASEKHAR, S. (1949). Population problems of India and Pakistan. *Science News*, no. 13, 102–20.

DAS, SIR K. & MAHALANOBIS, P. C. (1933–4). A preliminary note on the rates of maternal deaths and still-births in Calcutta. *Sankhyā*, **1**, 215–30.

ELLIS, H. (1936). *Studies in the Psychology of Sex*, **1**, i, 3, (pp. 122–60). 'They (the sexual climaxes) resemble those periods of atmospheric tension, of storm and wind, which accompany the spring and autumn phases in the earth's rhythm, and they may fairly be regarded as ultimately a physiological reaction to those cosmic influences' (p. 160). The Random House Edition, New York.

Gazette of India (1921–40). Published by the Manager of Publications, Civil Lines, New Delhi.

HILL, S. A. (1888). Life statistics of an Indian province. *Nature, Lond.*, **38**, 245–50 and 565–6.

KENDALL, M. G. (1943). *Advanced Theory of Statistics*, **1**, 145–7. London: Charles Griffin and Co. Ltd.

LANDRY, A. (1949). *Traité de Démographie*, 12th ed., pp. 334–5. Paris: Payot.

MATHEW, N. T. (1940–1). The influence of seasons on human reproduction. *Sankhyā*, **5**, 261–8.

NAIR, K. R. (1935–6). Maternity statistics from Madras. *Sankhyā*, **2**, 155–68.

Statistical Atlas of Bombay State (Provincial part) (1950). Published by the Bureau of Economics and Statistics, Secretariat, Bombay, §2, pp. 17–53.

SUBRAMANIAM, M. K. (1933). Temperature as a factor in the sex determination in man. *J. Asiat. Soc. Beng.*, N.S., **29**, 289–306.