Birth Order, Parental Ages, and Sex of Offspring

E. NOVITSKI AND A. W. KIMBALL

Biology Division and Mathematics Panel, Oak Ridge National Laboratory, Oak Ridge, Tennessee*

INTRODUCTION

VARIATIONS IN THE PROPORTION OF MALE BIRTHS (i.e., the secondary sex ratio) have been analyzed many times as a function of those characteristics of the parents, the sibship, or the environment, for which available data lend themselves to statistical treatment. Among the interesting relationships discovered previously is the observation that the age of the mother apparently has no, or very little, effect on these changes whereas the increasing age of the father is correlated with a decreasing secondary sex ratio (Novitski, 1953). This is a most puzzling result since one can readily imagine a number of ways in which some factor related to maternal age might be responsible, as, for instance, differential mortality of the two sexes before birth, whereas a system related to the age of the father (and relatively independent of the mother's age) is not so easily conceived.

An attempt to assess the relative contributions of the parental ages and birth order showed that there was no simple unifying interpretation of data that gave, on the one hand, the parental ages and sex ratio of offspring and, on the other hand, the maternal age, birth order, and sex ratio (Novitski and Sandler, 1956). It was therefore suggested that a more informative interpretation would depend on the analysis of data that gave, for each birth, the sex of the child, both of the parental ages, and the birth order simultaneously. Through the kindness of Dr. Halbert Dunn and his associate, Dr. Delbert Waggoner, of the National Office of Vital Statistics, such data for the year 1955 were made available to us, and this paper is concerned with the analysis of those data.

METHODS AND RESULTS

This analysis differs from the previous one in several important respects. Because live births for the year 1955 were classified simultaneously according to three criteria—age of father, age of mother, and birth order—it was possible to fit a regression function of sex ratio on all three variables simultaneously. Furthermore, the ORACLE, a high-speed digital computer operated by the Mathematics Panel at Oak Ridge National Laboratory, was available for use in the analysis. As a result we were able to fit a quadratic regression surface to the data (the existence of nonlinear effects had been suggested earlier) and to make several subsidiary calculations that would have required a prohibitive amount of work with desk calculators. In fact, it would have been possible to fit a cubic or higher order regression to the data with almost equal facility, but the data were not considered precise enough for this

Received January 4, 1958.

* Operated by Union Carbide Nuclear Company for the U. S. Atomic Energy Commission.

purpose. The fitted equation can be viewed as a Taylor's series approximation to the true regression function provided that the function admits a convergent infinite series expansion.

The following notation will be used:

x_1 = age of father, coded as in table 1
x_2 = age of mother, coded as in table 1
x_3 = birth order minus one
N_c = number of births in a given cell of table 1
N = total number of births
M_c = number of male births in a given cell
M = total number of male births
y = observed sex ratio in a given cell = M_c/N_c
Y = regression estimate of sex ratio for any x_1, x_2, x_3.

Since the N_c vary considerably in table 1, each y must be assigned an appropriate weight (w) and the regression function estimated by minimizing $\sum w(y - Y)^2$. The variance of y is estimated by $y(1 - y)/N_c$, assuming binomial errors, and ordinarily one would use the reciprocal of this quantity as the weight. However, only negligible errors are introduced if we take $w = N_c/\bar{y}(1 - \bar{y})$, where $\bar{y} = M/N$, because the true sex ratio in any cell in table 1 will differ only slightly from $\bar{y}$.

As a first step in the analysis, the function,

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3,$$

was fitted to the 175 observed sex ratios given in table 1 by the weighted least-squares procedure indicated. The estimated coefficients and their standard errors are shown in table 2; the residual sum of squares is 218.95 with 165 degrees of freedom. Under the null hypothesis that $E(y) = Y$, the residual sum of squares is distributed approximately as χ^2 and corresponds to a significance probability of 0.003. The value of χ^2/d.f. agrees well with the values obtained in analyses referred to above, although the significance level is different because of the greater number of degrees of freedom. Several interpretations of the significant χ^2 are possible. It is well known that the use of weights based on observed data introduces a bias into the computed χ^2. However, since all of the true sex ratios must differ only slightly from one half, this effect is negligible. A second possibility is that sampling errors include a component that is extrabinomial. As a check on this possibility, the 1955 data were compared with data from previous years for corresponding subclasses of the three independent variables. This analysis indicated a relatively uniform difference in sex ratio for most comparable subclasses (possibly because of the inclusion of nonwhite births in the 1955 data) but gave no evidence of extrabinomial variation. If these two possibilities are rejected, the natural conclusion would be that the departure of the computed χ^2 from its expectation is caused primarily by the failure of the quadratic model to explain all of the variation in sex ratio. A better fit might be obtained by including terms of higher order; alternatively, it is entirely possible that factors other than parental ages and birth order have some effect on sex ratio and that

TABLE 1. TOTAL LIVE BIRTHS AND SEX RATIO IN THE U.S.A. BY AGE OF FATHER, AGE OF MOTHER AND BIRTH ORDER: WHITE AND COLORED, 1955
(Source: National Office of Vital Statistics)

Age of father (yr)→				Under 20		20–24		25–29		30–34		35–39		40–44		45–49	
x_1→				0		1		2		3		4		5		6	
Age of mother (yr)	x_2	Birth order	x_3	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y	No. of births N_c	Sex ratio y
Under 20	0	1	0	67,764	.519730	174,493	.516405	40,895	.517814	6,282	.496975	1,655	.502115	483	.476190	227	.590308
		2	1	11,563	.506270	65,066	.509467	19,103	.508088	3,382	.517150	943	.503712	309	.533981	144	.479167
		3	2	1,165	.509013	13,712	.514367	5,115	.498338	1,096	.535584	303	.511551	127	.456693	53	.547170
		4	3	113	.486726	2,140	.495327	1,148	.515679	234	.517094	83	.469880	33	.545455	13	.461538
		5+	4	30	.700000	338	.565089	194	.520619	53	.509434	26	.461538	8	.375000	6	.333333
20–24	1	1	0	7,830	.515837	226,762	.515571	167,747	.514346	31,579	.511289	6,811	.519601	2,021	.518060	647	.511592
		2	1	2,815	.502309	173,484	.512681	184,963	.510756	38,359	.510284	8,412	.519853	2,354	.505098	818	.512225
		3	2	529	.466919	68,661	.510872	104,898	.513508	25,700	.515837	6,024	.508300	1,910	.526702	653	.509954
		4	3	139	.539568	20,546	.512362	44,556	.514072	13,300	.509098	3,361	.506397	1,175	.483404	435	.517241
		5+	4	52	.576923	7,273	.503919	23,102	.507965	9,274	.506038	2,698	.512231	945	.512169	395	.470886
25–29	2	1	0	184	.478261	15,473	.515349	95,187	.514766	51,912	.510499	13,698	.513433	3,718	.490048	1,119	.528150
		2	1	79	.506329	14,019	.521435	158,439	.513819	105,729	.513227	24,820	.507977	5,956	.525353	1,750	.521143
		3	2	54	.444444	7,857	.506427	121,294	.512663	102,507	.510424	24,763	.514598	5,951	.517728	1,745	.490545
		4	3	25	.520000	3,552	.513514	61,786	.512980	60,853	.511758	16,535	.508255	4,434	.509021	1,492	.508043
		5+	4	47	.297872	2,848	.514396	51,092	.499315	62,916	.511746	21,677	.513032	7,019	.507907	2,827	.506898
30–34	3	1	0	12	.583333	880	.498864	10,339	.521134	29,494	.518241	16,928	.514828	6,206	.510313	2,011	.522626
		2	1	11	.454545	856	.511682	16,129	.520863	69,949	.515233	40,401	.512958	12,028	.512388	3,238	.507412
		3	2	8	.625000	681	.516887	13,242	.507401	84,914	.516723	54,315	.517021	14,013	.512239	3,541	.505225
		4	3	6	.333333	459	.488017	8,088	.504080	57,010	.513156	40,820	.513547	11,075	.504199	2,928	.508538
		5+	4	25	.520000	737	.469471	9,157	.508900	67,265	.506727	64,623	.508890	23,672	.511617	7,655	.514958
35–39	4	1	0	6	.666667	94	.478723	1,072	.524254	4,393	.504211	9,454	.510260	6,086	.519717	2,379	.518705
		2	1	3	.0	109	.522936	1,382	.520984	8,015	.504679	21,657	.517015	12,454	.505621	4,061	.503571
		3	2	1	.0	80	.487500	1,252	.496006	8,385	.519141	31,662	.509570	18,538	.513809	4,958	.514320
		4	3	1	1.000000	66	.469697	859	.534342	6,108	.511297	26,899	.515595	17,174	.513916	4,645	.508504
		5+	4	10	.600000	162	.537037	1,582	.498104	8,508	.514104	49,011	.506907	44,387	.511749	15,697	.505319

TABLE 2. FITTED CONSTANTS AND STANDARD ERRORS FOR THE COMPLETE SECOND DEGREE REGRESSION EQUATION

Coefficient	Estimate	Standard error	t - ratio
b_i	$\hat{b}_i \times 10^3$	$s_{\hat{b}_i} \times 10^3$	$t = \mid \hat{b}_i \mid / s_{\hat{b}_i}$
b_0	515.620	1.1310	—
b_1	−0.5195	1.0474	0.496
b_2	0.5184	1.1100	0.467
b_3	−1.9029	0.8380	2.271
b_{11}	−0.2676	0.2513	1.065
b_{22}	0.0077	0.3893	0.020
b_{33}	−0.3208	0.2105	1.524
b_{12}	0.2031	0.4803	0.423
b_{13}	0.8016	0.2908	2.757
b_{23}	−0.3024	0.3516	0.860

these factors are not represented in our analysis. When appropriate data for additional years become available, further elucidation of the point may be possible.

Any regression analysis may be interpreted as a method for explaining the total variability in the data, as measured by the total sum of squares (weighted, if necessary), in terms of the factors by which the data are classified. When these factors can be quantitatively represented, it is customary and often useful to consider the functional relationship between the dependent variable (in our case, sex ratio) and the independent variables representing the factors being studied. Usually, the mathematical form of such relationships is unknown, and one ordinarily begins by assuming a simple linear or quadratic model such as the one described above. When this is done and the regression analysis is completed, it is possible to associate with each term or group of terms in the model a sum of squares which represents that portion of the total variability attributable to the particular term or group of terms being considered. Frequently, we refer to this sum of squares as a reduction (in total variability), since it reduces the amount of variability that is unexplained. The discussion that follows depends on arguments of this nature.

An examination of the t ratios in table 2, based on observed error variances, reveals clearly that certain terms in the regression contribute very little to the reduction in total sum of squares. Since the facilities of the automatic computer were available, we decided to examine the fits for several subsets of the ten fitted constants in an effort to determine the minimum number of terms in the regression equation that would be required to account for the variations in sex ratio. Without high-speed computing equipment, this type of analysis would probably not be feasible.

The results are shown in table 3. The total sum of squares ($\sum wy^2$) is 2640.86 and the reduction attributable to b_0 is 2350.92, which leaves a remainder of 289.94 to be reduced by fitting the other constants. Because every constant involving age of mother (x_2) had small t values in table 2, the first step was to eliminate this independent variable completely from the regression. From table 3 it can be seen that the additional reduction attributable to all terms involving x_2 is only 4.52, and an F test yields a significance level greater than 25 per cent. Thus, as was suggested by earlier work, the direct contribution of age of mother to the variations in sex ratio

TABLE 3. REDUCTIONS IN SUM OF SQUARES DUE TO REGRESSION

Constants fitted	Reduction attributable to constants fitted		Additional reduction attributable to remaining constants	
	Degrees of freedom	Sum of squares	Degrees of freedom	Sum of squares
$b_1, b_2, b_3, b_{11}, b_{22}, b_{33}, b_{12}, b_{13}, b_{23}$	9	70.99	0	0
$b_1, b_3, b_{11}, b_{33}, b_{13}$	5	66.47	4	4.52
$b_3, b_{11}, b_{33}, b_{13}$	4	66.46	5	4.53
b_1, b_3, b_{33}, b_{13}	4	64.73	5	6.26
b_1, b_3, b_{11}, b_{13}	4	61.78	5	9.21
b_3, b_{33}, b_{13}	3	58.70	6	12.29
b_3, b_{11}, b_{13}	3	61.70	6	9.29
b_1, b_3, b_{13}	3	60.38	6	10.61
b_3, b_{13}	2	56.98	7	14.01
b_3, b_{33}	2	54.17	7	16.82
b_3	1	53.63	8	17.36
b_{33}	1	51.77	8	19.22
b_{13}	1	31.23	8	39.76
b_2	1	5.84	8	65.15

is not significant. As a check on this result, chi-square heterogeneity tests were performed among maternal age classes within each of the 35 subclasses represented by combinations of the five birth orders and seven paternal age classes. The number of significant chi-squares found was no greater than one would expect from random binomial variation. This simple calculation provides additional evidence in support of the hypothesis that maternal age has no effect on sex ratio. It also strengthens the previous observation that extrabinomial variation, if present, must be relatively small, and it minimizes the likelihood that any appreciable heterogeneity has been introduced by pooling data for two races.

As the next step in the analysis, subsets of the five remaining constants were examined, and the corresponding additional reductions in sums of squares are shown in table 3. It is immediately apparent that b_1 contributes nothing significant to the regression since eliminating it leaves the reduction in sum of squares virtually unchanged. Consideration was then given to dropping individual constants from the subset (b_0, b_3, b_{11}, b_{33}, b_{13}), but from table 4 it can be seen that each of these constants contributes significantly to the regression. Nevertheless, the successive tests of significance are not independent, and there is some doubt as to whether the quadratic term for birth order should be retained and whether the quadratic term for father's age might be replaced by the linear term. These are borderline decisions, however, that depend primarily on one's interpretation of the P-values in analyses of this sort. In our judgment none of the constants shown in table 4 can be safely omitted from the regression.

The surface described by the final regression equation,

$$Y = b_0 + b_3x_3 + b_{11}x_1^2 + b_{33}x_3^2 + b_{13}x_1x_3 ,$$

can be examined in various ways. In figure 1, contours representing equal sex ratios have been plotted, and the fitted surface can be visualized quite readily. The varia-

TABLE 4. FITTED CONSTANTS AND STANDARD ERRORS FOR FINAL REGRESSION EQUATION

Coefficient	Estimate	Standard error	t - ratio	Significance level
b_i	$\hat{b}_i \times 10^3$	$s_{\hat{b}_i} \times 10^3$	$\lvert\hat{b}_i\rvert/s_{\hat{b}_i}$	P
b_0	515.0620	0.6384	—	—
b_3	−1.8478	0.7788	2.373	.025
b_{11}	−0.1892	0.0778	2.432	.021
b_{33}	−0.3825	0.2010	1.903	.062
b_{13}	0.6757	0.2212	3.055	.003

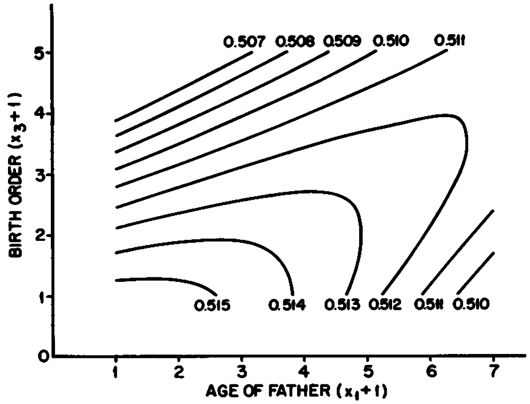

FIG. 1. Contours of equal sex-ratio plotted as a function of birth order and age of father, computed from final regression equation.

tions in Y as a function of either independent variable when the other is held constant can be obtained from figure 1 by examining appropriate cross sections of the surface. Thus, when Y is plotted as a function of x_3 for each value of x_1, (vertical cross sections) all curves are found to be convex downward; i.e., sex ratio seems to decrease more rapidly at higher birth orders, at least for young fathers. But for older fathers there seems to be very little dependence on birth order. In other words, sex ratio seems to depend primarily on birth order, decreasing as birth order increases, but this dependence diminishes rapidly as age of father increases, and accounts for the significance of the interaction term (b_{13}) in the regression equation. Alternatively, Y might be plotted as a function of x_1 for each value of x_3 (horizontal cross sections). This shows a negative regression of sex ratio on age of father for low births and just the opposite for high birth orders. Perhaps it should also be pointed out that by reference to figure 1, successive pairs of values for age of father and birth order can be chosen such that both variables increase but sex ratio remains unchanged. Previously there has been no evidence adduced that would have suggested such a relationship.

DISCUSSION

In assessing the reliability of these results, we must remember that the data used are for only one year, and include about four million births. Ordinarily this would be considered a very large sample, but the magnitude of the effect being studied is only about one per cent and the observations are necessarily clustered along a diagonal in the x_1, x_2, x_3 space owing to the high correlations among these three variables. When more data of this type become available, the study will be continued. It is probable that future analyses will result in some changes in the relationships as they are now conceived, but it is most unlikely that the conclusions with respect to age of mother will be altered.

The apparent lack of any effect of the maternal age, either as a linear, quadratic, or interaction component with any of the other variables must be reemphasized. In the works previously cited, it was considered curious that maternal age showed no relation to changing sex ratio, not only because one might expect such an influence for biological reasons, but also because one might predict the appearance of a relationship in those instances where the data were limited to only parental ages, since the probable correlations of those ages with other factors (as, for instance, birth order) that might themselves have an effect on sex ratio should show up as a spurious relationship. Since we now know that changes in birth order and paternal age are both directly related to changes in the secondary sex ratio, it is clear that birth data must not be analyzed without taking both of these factors into account or, at least, that it is not advisable to analyze such limited data without being aware of the inevitable pitfalls involved. For this reason, we would hesitate to place any great reliance on an apparent influence of maternal age based on a simple breakdown by parental ages only, and would be inclined to question the relevance of such analyses to a biological interpretation of the causes of variations in sex ratio. Specifically, Lejeune and Turpin (1957a, 1957b) have analyzed data from the U. S. Public Health Service, giving the sex ratio of live births as a function of the parental ages for the period from 1946 to 1956.[1] They report that, although the effect of the paternal age is most pronounced, there is nevertheless a significant effect of the maternal age and that this might be interpreted as evidence for a higher incidence of sex-linked lethals in older women than in younger ones. It is our opinion that whereas the conclusion may be correct and, in fact, sounds plausible on genetic grounds, there is no evidence for it at the present time. It might be pointed out in this connection that the signs of the small, insignificant linear and quadratic regression coefficients on the age of the mother are in the wrong direction for an interpretation in terms of sex-linked lethals; that is, they are now positive, whereas those coefficients from analyses that do not take into account all three variables at the same time are negative.

Lest the intent of the preceding argument be misunderstood, it should be pointed out that we place no great reliance on birth order or paternal age as having a direct influence on secondary sex ratio but feel rather that these two items represent easily characterized attributes of any given birth, for which large amounts of data are available, and that they are, in all probability, themselves correlated with other

[1] The tabulations for the years 1951 to 1953 inclusive are based on a 50% sample. The totals given by Novitski and Sandler (1956) which include these years are therefore inaccurate. This inaccuracy does not substantially affect the general conclusions presented in that paper, but any worker wishing to use those totals should be aware that the estimates of error derived from those tables will be somewhat smaller than they should be.

factors more closely related to the true causes of the variations in sex ratio. It is therefore almost certain that there exist other attributes of a birth that, when known and properly available in statistical form and incorporated in an appropriate analysis, will show birth order, or parental age, or both, *as such* to be of no direct influence in altering the secondary sex ratio. The value of an analysis of the sort presented here rests in the demonstration that certain factors formerly considered important may represent spurious statistical correlations. Directing attention to the factors that seem to be most closely correlated with sex-ratio changes leads to the formulation of additional questions that may ultimately reveal the causes of those changes.

Little if anything can be said about the possible biological basis for the changes reported here. The basis for changes that have their expression as a function of the age of the male or of birth order, by themselves, would not be difficult to imagine, but one that depends on both of these factors simultaneously (as represented by the highly significant interaction term b_{13}) is not so easily conceived. There is some possibility, of course, that with the accumulation of more data, the relations will in some way be changed so that they seem more comprehensible in simple biological terms. It is also possible, as implied earlier, that the interaction term is simply a reflection of effects not adequately represented in the model or of the inadequacy of the functional form of the present model in terms of identifiable factors. Nevertheless, it would be a mistake at this stage to dismiss the interaction term as of no real significance purely because no simple biological explanation is at hand.

ACKNOWLEDGEMENT

We are grateful to Mr. G. J. Atta for assistance with the ORACLE computations and to Dr. E. R. Dempster for suggesting the chi-square homogeneity tests within birth order-paternal age subclasses.

SUMMARY

An analysis of U. S. birth data giving simultaneously the parental ages, birth order, and sex of offspring for the year 1955 shows that the variation in secondary sex ratio is independent of the mother's age but is dependent on both birth order and paternal age. It is further shown that the relationships are not linear and cannot be represented adequately without the inclusion of a component that reflects an interaction between birth order and age of father.

REFERENCES

LEJEUNE, J., & TURPIN, R. 1957a. Influence de l'âge des parents sur la masculinité des naissances vivantes. *Compt. rend.* 244: 1833–1835.

LEJEUNE, J., & TURPIN, R. 1957b. Mutations radio induites chez l'homme et dose de doublement. Sur la validité d'une estimation directe. *Compt. rend.* 244: 2425–2426.

NOVITSKI, E. 1953. The dependence of the secondary sex ratio in humans on the age of the father. *Science* 117: 531–533.

NOVITSKI, E., & SANDLER, L. 1956. The relationship between parental age, birth order and the secondary sex ratio in humans. *Ann. Human Genet.* 21: 123–131.