The Loglinear Model as a Tool for the Analysis of Demographic Data

Mona Khalifa

INTRODUCTION

In demographic practice, two main categories of data are encountered. Variables such as age, income and number of children per family can be measured quantitatively on an interval scale. A unique numerical value is assigned to each individual observation, and so they are called quantitative variables. Other variables such as the level of education, the marital status and the attitude to a specific issue provide a classification of objects into categories which describe the qualities possesed by each object. They are therefore called categorical or qualitataive variables.

Both types can be further classified. The quantitative variables can be either discrete or continuous. For a discrete variable, the set of values that the variable can assume is finite or countabley infinite and the values are ususally integers e.g the number of births per woman, the number of years that elapse before a member of a birth cohort dies. The continuous vaiable is one with an uncountable infinte set of possible valuse contained in a closed interval of the real linei.e integer non-integer values. Examples of such variables are age and income.

On the other hand qualitative variables can be of two types. Some are such that thie values, for a paricular suvey unit are determined in advance by the investigator. For example the researcher may choose a group of women that will attend a family planning educational programme and those who will be used as a control group in a project e.g aiming to examine the effect of a family planning educational programme on the prevelance of contraceptives. Similarly, the distribution of the number of births per woman can be studied seperately for three groups of women : those With no education, primary education, secondary education. In both these cases, the total number of observation for each category are determined

* Departmentof statistics - Cairo University, Khartoum Branch
March, 1990.

x_1	x_2: 1	2	3	4	b	Total
1	n_{11}	n_{12}	n_{13}	....	n_{1b}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	....	n_{2b}	$n_{2.}$
3	.					
.						
.						
.						
a	n_{a1}	n_{a2}	n_{a3}	....	n_{ab}	$n_{a.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	....	$n_{.b}$	n

Thus, we have:

$$n_{i.} = \sum_{j=1}^{b} n_{ij}$$

$$n_{.j} = \sum_{i=1}^{a} n_{ij}$$

$$n = \sum_{i,j} n_{ij}$$

Since every individual in the sample must belong to one, and only one, cell in the contingency table, the simplest analysis would aim at the estimation of the cell probabilities p_{ij} p_{ij} such that $\sum p_{ij} = 1$ under a specific hypothesis.

The Model

The aim of the analysis is to examine how the probability of an observation falling into a given cell depends on the relationship between the variables, a relationship specified by the model. In the example given above, the aim of the analysis would be to understand the relationship between the wife's education and the attitude to family planning because this relationship determines the probability of an observation falling into one of the cells of the table.

The model must imply some hypothesis about p_{ij} and the estimat probability will be denoted by p_{ij} which when multiplied by the number of individuals in the sample will yield the expected frequencies in the cells of the contingency table (u_{ij}) such that :

$$u_{ij} = n\, p_{ij} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The estimates (u_{ij}) are called the fitted values. To test the hypothesis we examine the agreement between the observed cell frequencies (n_{ij}) with the fitted values (u_{ij})

Generally we are interested in the 'null hypothesis' i.e no association between x_1 and x_2. In the above example presented in table(1) we would be interested in testing whether family planning is more acceptable among educated women than among non educated women. Under the null hypothesis the probability of an observatio falling in category i of x_1 is independant of the probability of falling in category j of x_2. This means that the probability of a woman having a favourable attitude to family planning is fixed irrespective of her level of education. Therefore:

$$p_{ij} = p_{i.}\, p_{.j} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

and since $u_{ij} = np_{ij}$, $u_{i.} = np_{i.}$, $u_{.j} = np_{.j}$

therefore $u_{ij} = u_{i.}\, u_{j.} / n \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$

The above equations specify that in the population, the probability of an individual falling in the ij th cell is the product of the marginal probabilities. However, a more convenient model which would correspond to these used for the analysis of variance, is the addetive model. We can creat an additive wodel by tking the natural logarithms of the above equations and we find that,

$$\log p_{ij} = \log p_{i.} + \log p_{.j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

and $$\log u_{ij} = \log u_{i.} + \log u_{.j} - \log n \quad \ldots\ldots\ldots\ldots\ldots (5)$$

Summing (5) over i and dividing by a

$$1/a \sum_i \log u_{ij} = 1/a \sum_i \log u_{i.} + \log u_{.j} - \log n = \theta_j \quad \ldots (6)$$

Summing (5) over j and dividing by b

$$1/b \sum_j \log u_{ij} = \log u_{i.} + 1/b \sum_j \log u_{.j} - \log n = \theta_i \, (7)$$

Summing (5) over i and j and dividing by ab

$$1/ab \sum_i \sum_j \log u_{ij} = 1/a \sum_i \log u_{i.} + 1/b \sum_j \log u_{.j} - \log n = \theta \quad (8)$$

Since we define :

$$\theta = 1/ab \sum_i \sum_j \log u_{ij}$$
$$\theta_i = 1/b \sum_j \log u_{ij}$$
$$\theta_j = 1/a \sum_i \log u_{ij}$$

So that θ is the average of the logarithms of all the expected cell frequencies, θ_i is the average of the logarithms of the expected cell frequencies in the category i of x_1 (b cells) and θ_j is the average of the logarithms of the expected cell frequencies in the category j of x_2 (a cells).

thus,

$$\theta_i - \theta = \log u_{i.} - 1/a \sum_i \log u_{i.} = \alpha_i \quad \ldots\ldots\ldots\ldots\ldots (9)$$

and

$$\theta_j - \theta = \log u_{.j} - 1/b \sum_j \log u_{.j} = \beta_j \quad \ldots\ldots\ldots\ldots (10)$$

Substituting for $\log u_{i.}$ and $\log u_{.j}$ in (5) by (9) and (10) we have,

$$\log u_{ij} = \alpha_i + 1/a \sum_i \log u_{i.} + \beta_j + 1/b \sum_j \log u_{.j} - \log n$$

and using equation (8)

$$\log u_{ij} = \alpha_i + \beta_j + \theta$$

note that $\sum \alpha_i = \sum \beta_j = 0$

or $\log u_{ij} = \Theta + \alpha_i + \beta_j$ ....(11)

This formula is known as the log linear model for the expected frequencies u_{ij} on the hypothesis that x_1 and x_2 are independant.

The Similarity to ANOVA

The log linear model is clearly similar to the problem of the analysis of variance ANOVA. They are both a simple additive model where a dependent variable is split into additive components representing an overall mean and main effects. We can thus think of Θ as the 'overall main effect', while α_i is the main effect of the i th category of the variable x_1 and β_j as the main effect of the j th category of the other variable x_2. It is clear that the main effects parameters (α and β) are measured as deviations of the means (of the marginal totals of the log frequencies of the rows or the columns) from the overall mean Θ. The above model is in terms of the expected frequencies u_{ij} and - as mentioned before- is based on the assumption that x_1 and x_2 are independent. To apply the model we need to estimate these expected frequencies and the parameters in the model. For example the logarithm of the expected frequency u_{12} (the first row and the second column) and the associated parameters Θ, α_1 and β_2. Where Θ is the average of all the logarithms of the expected cell frequencies, α_1 is the difference between the average of the logaithms of the cells of category 1 (for x_1) and Θ, and β_2 is the difference between the average of the logarighms of the cells of category 2 (for x_2) and Θ . The maximum likelihood estimation for log linear models (Birch , 1963) shows that:

$\hat{u}_{ij} = n_{i.}\ n_{.j} / n$ ....(12)

On the other hand, if x_1 and x_2 are not independent, the association between the two variables means that the frequencies in the cells will be affected by the combination of x_1 and x_2 categories i.e (i and j). Again, as in the ANOVA situation, this means that the probability of an individual having a certain attitude to family planning varies according to her level of education. The full model will take the form

$$\log u_{ij} = \theta + \alpha_i + \beta_j + \gamma_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

Where γ_{ij} is the interaction parameter for categories i and j such that $\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$.

The new model contains(ab) unknown parameters[1+(a-1)+(b-1)+(a-1)(b-1)] If the model is fitted to the a*b observed frequencies of the contingency table a perfect fit is expected and the model is called a 'saturated' model.

Subsequently, we need to test the adequacy of the suggested model for the observed data. To test the null hypothesis between $x_1$and x_2 it is enough to test whether $\gamma_{ij} = 0$ for all i and j in(13) since- in this case- the model becomes that expressed in (11).

However, in the context of contingency table analysis - in contrast to ANOVA- the values taken by the 'main effect' parameters are of little concern. The estimation of the interaction effects is more important because it is useful to identify the categories responsible for any departure from independence. In the above example we should be interested to know which level of education is more effective i.e affects the probability of falling in a specific category of the other variable (attitude to family planning). this is a considerable gain in knowledge than to know that the two variables are simply not independent.

It is clear that the aim is to find the model that best fits the data (sample variability taken into consideration). We can- therefore- fit several models where one or more of the parameters

(θ, α, β, γ) are set to zero. Finally we choose the best model using a test for the goodness of fit based on a comparison between the observed frequencies and those expected under the specific model. For each model the number of degrees of freedom is determined by the difference between the number of cells in the table and the number of independent parameters. The goodness of fit of the log linear model can be tested by two common summary statistics. These are the classical Chi square χ^2 statistic and the log likelihood statistic D (Nelder and Wedderburn 1972).

To test a particular term, one should fit a model with and without the relevant term and then examine the resulting change in the goodness of fit. Higher order terms may not be included unless the related lower order terms are included

There are several computer packages that include log linear procedures. For example the LOGLINEAR precedure in SPSSX, models cell frequencies and produces maximum likelihood estimates of the parameters and a very rich output which includes χ^2 and D statistics. The researcher needs to specify the variables to be included in the analysis and to specify the model or models to be fit (see the SPSSX manual). The fitting algorithm aims to find the parametersa that assure that marginal distributions of the observed and the predicted frequencies are almost identical. The iterative fitting procedure is monitored by watching the χ^2 statistic. The LOGLINEAR procedure is thus defined as "a general procedure which does model fitting, hypothesis testing, and parameter estimation for any model that has categorical variables as its major components" (SPSSX User's Guide).

An Application: Life Tables with Covariates

The life table approach is of particular importance in demographic analysis. In addition to its classical value in the study of mortality , it has been recently applied to the study of other demographic evnets e.g the study of birth intervals. This application was introduced because life table analysis is particularly suitable for dealing with censored data (Khalifa, 1987) and therefore both open and closed birth intervals can be included in the analysis of survey data. By forming such life tables for each birth interval, summary statistics can be calculated that provide a description of the process of transition from one parity to the next. B(60) or the quintum (the proportion of women who proceed to the next parity within 60 months from the last) is used as a measure of the quantity of fertility while (T) or the trimean is defined as $T=(T_{25} + 2T_{50} + T_{75})/4$ and is used as a measure of the tempo of transition (Khalifa,1986). The two measures may be calculated for subgroups of the population and may be used for the purpose of comparing fertility behaviour among them e.g by the mother's age, by her education...etc. (Khalifa and Shadad,1986). It is clear that the starting point for such an exercise is the construction of a seperate life table for each of the subgroups of the population.

However, the results remain unsatisfactory for examining the determinants of the concerned probabilities. This is because the population can be divided into a few subgroups before the number of cases become too small to allow for the construction of a seperate life table and for reaching a meaningful analysis. It becomes impossible to consider e.g age, wife's education and husband's education simultaneously. If each of these variables is divided into four categories, the sample will contain 64 subgroups and each will contain a small number of cases. The number of cases will decrease as the number of variables and categories increase and the analysis becomes limited. Since our goal is to discover similarities and differences this limitation becomes a serious disadvantage of the tabular approach. Therefore the attempt to examine the determinants of mortality (or birth interval) by comparing life tables becomes very limited, in the sense that no more than two or three factors can be used at the same time.

A more recent method is now available that permits the simultaneous analysis of life tables with covariates. The method depends on forming a hazard model in which the risk(hazard) of death (or of having a subsequent birth in the case of birth interval analysis) is not the same for all individuals, but is dependent on other factors (covariates). The aim becomes to determine the form of the relationship between the life table probability -as a dependent variable - and the other variables that are believed to affect the probability e.g the age of the mother, her level of education, the type of place of residence...etc. Since the dependent variable is a probability ($0 \leq p \leq 1$) the use of ordinary regression analysis is not valid. This is because linear probability models (mOdels that relate the probability of an event to a series of exogenous factors in a linear fashion) are often unrealistic, and attemps to estimate such models by Ordinary Least Square methods based on individual observations quite generally lead to biased and inconsistent estimates(Hanushek and Jackson, 1977).

At each duration (x) in a bircle interval life table (e.g months since birth), a mother is at a risk of closing the interval by having another birth this risk is A major assumption of the ordinary life table is that all individuals at duration x- are assumed to have the same risk q(x). Life tables with covariates relax this assumption. Instead, the risk is assumed to vary among individuals according to their characteristics even th-ough they belong to the same duration category, and a contingency table situation arises. Thus, the risk of closing during the month (x) for a woman with a characteristic (i) is $q_i(x) = p(x) * C_i(x)$ weher p(x) is the probability of closing after x months shared by all individuals and $C_i(x)$ is a specific multiplier associated with the characteristic i. If $C_i(x)$ is greater than one, the individuals with characterristic i will have a greater risk and if $C_i(x)$ is less than one, there is a lower risk for individuals in group (i). The multiplier $C_i(x)$ is called a proportionality factor.

A more convenient mathmatical frame is a log linear model such that,

$$\log q_i(x) = \log p(x) + \log C_i(x) \quad \text{..........................(14)}$$

In case we consider several covariates simultaneously we can define

$$\log C_i(x) = b_1X_{i1}(x) + b_2X_{i2}(x) + \ldots\ldots + b_nX_{in}(x) \quad \text{.............(15)}$$

where $X_{ij}(x)$ is the value of covariate j at duration (x) for individuals in group (i) and b_j is the coefficient that measures the effect of that covariate on the log of the baseline risk [log p(x)] the model says that the log of the risk is the sum of a constant [log p(x)] and an effect of being an individual with characteristic (i).

The b's (effect estimates) can be interpreted similar to the coefficients in a traditional regression analysis. For example if the covariate is the level of the mother's education and the corresponding b value is negative, then the log risk of closing the interval after x months would decline linearly as the mother' s level of education increases .

The techniques for estimating the parameters of the model are based on finding the parameters that maximize the likelihood (or the log- likelihood function) and therefore maximize the probability of observing the outcomes that did occur (i.e the observed data). In this case we become in a position to determine the preferred model i.e the most suitable model for describing the observed data. As mentioned before several computer packages are available using this approach and yielding the parameter esimates and thus, allowing the formation of the relevant life tables. For a detailed description of the estimation procedure see for example (Trussell and Hammerslough, 1983). In addition to the construction of full life tables, one can use the estimated parameters to assess relative risks of closing(i.e the ratio of two risks). Usually, we would be interested in knowing the relative risk when all other factors are held constant. It is usual in a computer output to provide the parameter estimates in relation to its first category of the variable. For example, in a programme examining the effect of 'mother's education' on the probability of closing the interval within a specific duration e.g 9-12 months and in which 'mother's education' is a factor divided into four categories [none, primary, seccondary, university] we shall find available three parameters that express the risk in each category in relation to that of the grand mean. Suppose the three parameters are -0.142, -0.223 and -0.321. This means that being a woman with primary education, the risk of closing at duration (x = q-12 months) is exp(-0.142) = 0.87 times that of a none educated woman (the grand mean). For a woman with secondary or university education, the risk of closing at duration (q-12 months) is 0.80, 0.73 times that for the none educated woman respectively. All of these estimates are valid as all other variables are considered constant i.e they present the effect of mother's education on the risk of closing a birth interval irrespective of other varibles e.g father's education, type of place of residence ... etc. In this way, the researcher can determine the size of the effect that each variable have on the probability seperately. It becomes also possible to compare variables. The researcher can arrive at answeres to questions such as "which factor has the largest impact on the esimated risk of closing a birth interval . For a detailed application of the procedure see for example (Rodriguez et al., ,1984) . The authers were able to demonstrate that life tables with covariates con be easily estimated with standard computer packages designed for the analysis of contingecy tables.

Example of aloglinear analysis of a birth interval

(i)Objective

The aim of the analysis is to examine the determenants of the fifth birth interval in Sudan i.e the transition of women from the fourth to the fifth birth. This interval is of special interest in Sudan as it is at this stage that some women start to use contraception.

Approach

In a previous paper, the life table approach has been used for the analysis of birth intervals for subgroups of the population (Khalifa and Shadad, 1986). The approach was found useful as it allows for a sequential examination i.e seperate analysis for each birht order. The results showed parity related differences in the speed and quantity of transition as well as differences between socio economic groups of the population. However, these results remain unsatisfactory since it has not been possible to consider several factors simultaneously e.g age, trend, education, place of residence ...etc. This is because the number of cases available for the construction of seperate life tables (for each combination of characteristics) become very small - as the number of variables increase.

The loglinear method is one method that allows for the analysis of life tables with covariates. When all the covariates are categorical in nature, the analysis of contingency tables become suitable and the use of computer programmes such as LOGLINEAR for fitting models become convenient.
The dependant variable is a dichotomous variable reflecting whether the woman had a fifith birth during a specific duration segment or not and the independant variables are those socio economic and intermediate variables that affect fertility.

(ii)The Data

The data used in this example are from the WFS survey of Sudan conducted in 1980. The covariates used are those available in the survey and are related to the process of reproduction. They are all divided into a small number of categories. These covariates are :-

(1)Length of the previous interval

It is believed that the length of an interval is associated with the length of the previous interval as they are both dependant on certain couple specific fertility behaviour. The categories of this variable are : (less than 20 months i.e short),(21 - 30 months i.e medium), (31+ months i.e long).

(2)Period

This a trend variable that referes to the period of time during which the interval was started. The categories are: (1959 - 1963) (1964 - 1968) (1969 - 1974).

(3)Age of mother at the start of the interval

This variable is a major demographic variable associated with fertility behaviour. The categories used are: young(15 - 22) , medium(23 - 26), old(27+).

(4)Survival status of the birth initiating the interval within 12 months of its birth

The death of an infant is known to have an effect on the risk of having a next birth. This variable is categorized into : (dead), (alive).

(5)Place of residence

This variable includes (urban), (rural).

(6)Wife's education

Due to the small number of women in educational subgroups, the variable is divided into : (none), (some)

7) Husband's education

Again this variable is divided into (none), (some). The full description of these variables and their importance in the differential study of birth intervals are available elsewhere (Khalifa Nagieb, 1989) .

(iv) the Model

Having chosen the seven variables and their categories, the next step is to choose the suitble model. The input matrix consists of the records of the women subjected to the risk of having a fifth birth at the start of a specified duration segment of time after the birth of the fourth child. the segments -in months- are specified as (9-18 , 19 - 24 , 25-30, 31 - 45 , 46 - 60). this means that the analysis is repeated five times, and in each time the number of women subject to the risk is reduced by the number that have had a fifth birth within the previous duration segment. The number of women entering each duration segment is presented in table 2.

Table 2 the number of women subject to the risk of having a fifth birht at the begining of each duration segment

Duration	9 - 18	19 - 24	25 -30	31-45	46-60
N	1078	829	585	387	177
percentage	100	77	54	36	16

Several models were tested that include the main effects model and models with first order interactions. The process of selecting a suitable model was done systematically by first specifying the model, then applying the LOGLINEAR procedure of SPSSX. As mentioned before the procedure provides a set of fitted frequencies and measure the goodness of fit, that allow the selection of the best model that fits the data i.e. produces expected frequencies that are close to the observed frequencies. The goodness of fit statistics showed that the main effects model (expressed by equation 11) is a useful model to examine (has the least value of χ^2).

(v)Results from the main effects model

Table 3 presents the results of the analysis of the proportions having a birth within a specific duration segment. These results are provided by the output of the LOGLINEAR programme in the form of parameters. They represent the 'effect' of being in the specific category relative to the baseline category (the grand mean). The positive effects estimates indicates that the risk of closing the interval is higher than that given by the constant term, a negative estimate indicates the opposite. This was explained before in equation 15.

Table 3: The parameter estimates of the main effects model

variable/	Duration Segment				
category	9 - 18	19 - 24	25 - 30	31 - 45	46 - 60
Grand mean	-0.63	-0.524	-0.24	-0.10	-0.24
L short	.25	-.05	-.17	-.17	-.09
L medium	-.24	.18	-1.35	.09	.13
L long	-.01	-.13	1.52	.08	-.04
P 1969-74	.07	-.29	-.07	-.03	-.01
P 1964-69	-.03	.12	—	.09	-.11
P 1959-64	-.04	.17	.07	-.06	.12
A young	.10	-.07	-.01	.19	-.14
A medium	.05	.13	.13	-.12	.17
A old	-.15	-.06	-.12	-.07	-.03
I died	.36	.03	.19	-.15	.03
I alive	-.36	-.03	-.19	.15	-.03
R urban	.11	-.14	-.11	-.05	-.03
R rural	-.11	.14	.11	.05	.03
W some	.03	-.09	.01	-.08	.07
W none	-.03	.09	-.01	.08	-.07
H some	—	-.01	.17	.07	-.07
H nane	—	.01	-.17	-.07	.01

The parameters (coefficients) are interpretable as the log of the odds (expected frequencies of those who experienced a fifth birth / expected frequencies of those who did not). The parameters can be used to calculate the proportion of those who experienced a fifth birth since (proportion = odds/ 1+ odds).

In table 4 the parameter estimates are converted to relative risks (antilogs). The exponentiated effects are easier to understand. they can be interpreted directly as risks relative to the exponentiated grand mean and hence as 'relative risks'. For all factors the risks are presented as relative to the last category of each factor. the 'relative risks' reflect the risk of being in a specific category of the variable while all other variables are held constant.

Entries in this table can be used to calculate any relative ...k required. For example relative to the base category (last category of each factor) women who have had their fourth birth since 25-30 months, whose previous birth interval was short, who are 23-26 years old at the start of the interval in 1964-69, whose fourth child survived to age one, who live in an urban area, are educated and married to educated husbands have a relative of (1.00* 0.84* 1.14* 1.00* 0.90* 1.01* 1.19) = 1.04.

To obtain their estimated absolute risk we multiply by the exponential of the grand mean parameter in table 3 i.e (exp -0.24= 0.783). Then the absolute risk = 1.04*0.783=0.814 compared to the base line category of 0.783. this means that -on the average- the risk of closing a fourth interval is 0.783, but for women of the above characteristics the risk is 0.814. The risk (odds) can be translated into a proportion

(0.814/1.814 = 44.9 %). The expected proportion of women with the specified characteristics who will proceed to the fifth birth within 25-30 months is 44.9%.

Table (4)Relative risks in the main effects model

variable / category	8 - 18	19- 24	25 - 30	31 - 45	46 - 60
			Duration segment		
Grand mean	0.532	0.592	0.783	0.905	0.787
L short	1.28**	0.95	0.84*	0.84	0.91
L medium	0.79**	1.20**	0.26**	1.10	1.14
L long			1.00		
P 1969 -74	1.07	0.75**	0.93	0.97	0.99
P 1964 -69	0.97	1.13	1.00	1.09	0.90
P 1959 -64			1.00		
A young	1.11	0.93	0.99	1.21*	0.87
A medium	1.05	1.14	1.14	0.89	1.19
A old			1.00		
I died	1.44**	1.03	1.21*	0.86	1.03
I alive			1.00		
R urban	1.12*	0.87*	0.90	0.95	0.97
R rural			1.00		
W some	1.03	0.91	1.01	0.92	1.07
W none			1.00		
H some	1.00	0.99	1.19**	1.07	0.99
H none			1.00		

Significance: * at 0.95 , ** at 0.99
L:length of previous interval ; A: age; I: infant mortality
P: period , R: residence; W: wife's education; H: husband's education.

(vi)discussion

The columns of table 3 show the relative importance of a determinant on the probability of experiencing a live birth in a specific segment. the rows show how the effect of each factor varies from one segment to another. It seems that -controlling for other factors- infant mortality and the length of the previous birth interval are important factors affecting the risk of transition to the fifth birth. Also, the place of residence and the husband's education are of some importance. This is indicated by the statistically significant coefficients. These factors act on the risl of having a fifth birth as follows

(1) Length of the previous interval

Having a short fourth interval increases the risk of experiencing a short fifth interval. Also, a medium fourth interval is associated with a short or a medium fifth interval. The strong effect of length of previous interval was observed by other researchers e.g (Trussell et al 1983, Trussell et al 1985). It has been suggested that this effect is related to couple specific behaviour which includes breast feeding practice, contraceptive use, fecundability and coital frequency.

(2) Survival status of the birth initiating the interval within 12 months of its birth.

The results show that the death of the fourth birth before reaching its first birthday is associated with a rapid birth of a fifth child. This relationship reflects the effect of breast feeding on the risk of conception. the early death of the fourth birth causes sudden ending of breastfeeding which leads to a short post partum ammenorrhoea and a rapid conception.. Table 3 shows that the risk of having a fifth birth within 9-18 months after the birth of the fourth child is 44% higher when that child dies in infancy than when it continues to live after age one. As mentioned before this is due to the sudden termination of breast feeding. The difference is also significant at duration 25-30. The reason behind this could be a desire for replacement . Otherwise, it seems that after 30 months, the risk of having a fifth birth is not dependant on the survival status of the fourth birth.

(3)Place of residence

The results show that the risk of having a fifth birth within 9-18 months of the fourth is 12% higher for urban women than rural women. Again, this may be related to breastfeeding practices. It is known that urban women tend to breastfeed for shorter durations than rural women. It seems that contraceptive use is not acting in the expected direction immediately after the fourth birth. However, the effect of early use of contraceptives is evident in the second duration segment, when the risk is lower for urban women. On the other hand there is no significant difference at the higher durations. This is an indication that the use of contraception at this stage , by urban women, is for the prupose of delaying the fifth birth.

(4) Husband's education

The table shows that the risk of having a fifth birth (within 25-30 months of the fourth) is higher (by 19%) for women married to husbands with some education, than for those with no education. The result is indicative to the preference of such families to delay the fifth birth, in contrast with families of none educated husbands.

Generally, having such a small number of statistically significant parameters shows that the factors used are not effective in creating fertility differentials in the Sudan.

Conclusion

The paper has discussed the recent use of loglinear models in the area of demographic analysis. The technique is convenient because it is related to the analysis of contingency tables, for which several simple computer programmes are available. It is of special interest to demographers as it allows the examination of the effect of several determinants simultaneously. This examination has been difficult using other techniques (e.g life tables) because of the small number of cases in each cell when the number of variables become large. However, the contingency table approach is only appropriate for categorical covariates only which may restrict the analysis , if the researcher wishes to include continuous variables.

The illustration we used demonstrates the procedure and shows the ability of the technique to improve our understanding of the mechanism of fertility differentials using individual - instead of aggregate- data. It has shown how each variable affects the process when all other variables are held constant.

References

Birch 'Maximum Likelihood in 3-way Contingency Tables' in Journal of Royal Statistical Society B25,1963.

Bishop et. al.'Discrete Multivariate Analysis' ,1975.

Evirett 'The Analysis of Contingency Tables',1977.

Fienberg 'The Analysis of Cross Classified Categorical Data' 1981.

Hanushek E.A and Jackson J.E 'Statistical Methods for Social Scientists' Academic Press, 1977.

Khalifa, Mona 'A Preliminary Examination of Birth Intervals in North Sudan' the 21st Annual conference in Statistics, Computer Science and Operation Research. December 1986.

Khalifa, Mona 'An Introduction to Multivariate Life Table Analysis of Birth Intervals' The Egyptian Population and Family Planning Review vol21 No.1 June 1987.

Nelder and Wedderburn 'Generalized Linear Models' 'Journal of Royal Statistical Society A135 ,1972.

Rodriguez G.,Hobcraft J., McDonald J., Menken J. and Trussell J. 'A comparative Analysis of Determinants of Birth Intervals' WFS Comparative Studies No.30, May 1984.

Trussell J. and Hammerslough C. ' A Hazards Model Analysis of the Covariates of Infant and Child Mortality in Sri Lanka' Demography 20(1), 1983.

١٠ منى عبد الفتاح خليفه وفاطمه شداد .
دراسة الاختلافات فى الخصوبه استخدام اسلوب جداول الحياه .
تحليل علمى عن السودان " المجلة العشرون العدد الثانى ديسمبر ١٩٨٦م " المجلة السكانيه .

Khalifa, M. & Nagieb 'The effect of wife's education on the tempo of fertility in Egypt and Sudan' , The 24 th annual conference on statistics computer science and operations research ,23-25 December 1989.

Trussell, Martin, Feldman, Palmore, Concepcion and Abu Bakr 'Determinants of Birth Interval Length in the Philippines, Malaysia and Indonesia: A Hazard Model Analysis' Demography vol 22, No.2, 1985 .