An Introduction to Multivariate Life Table Analysis of Birth Intervals

Introduction

The concept of using the life-table in the analysis of birth intervals has been rapidly developing over the last decade. This is because of the power of the technique in dealing with censored data of the type provided by cross-sectional surveys. A wealth of such information has been collected for several countries which participated in the WFS (World Fertility Survey) program. This information includes birth histories, and marriage histories both of which are detailed records of events of demographic interest. The events reported in a birth history include those that determine the span of the woman's reproductive life i.e. the first and the last birth as well as those that describe the speed of reproduction i.e. the birth intervals or the time between successive live births. The speed by which women move from one parity to the next is determined by several biological, cultural and socio-economic factors. These factors act through the 'intermediate fertility variables' on the length of the birth intervals which, in turn, affect the number of births a woman may have.

However, the recent focus of attention on the analysis of birth intervals is not only due to the availability of WFS birth history data, but also for other reasons. The levels and trends of fertility are the final results of changes in the intermediate variables acting in different directions. These effects may vary at the different stages of the family-building process. This calls for considering fertility as a sequential process in which the rate of transition from one parity to the next is determined by parity specific factors. These factors may be behavourial e.g. contraception or unintentional e.g. age (Bulatao, 1981). The analysis of birth intervals allows for the study of these factors at in traditional behaviour related to fertility e.g. breast feeding patterns associated with the length of the past-partum non-susceptible period. These changes affect the demographic transition as much as it is affected by the adoption of new patterns of behaviour e.g. contraception. Bongaarts' intensive effort (Bongaarts, 1978) to identify the intermediate fertility variables has increased the attention given to breastfeeding in addition to the other factors (marriage and contraception) that have long been used in fertility research (Davis and Blake, 1956). The assesment of the contribution of each of these factors to the fertility transition is possible by looking into the birth intervals rather than looking at the final outcome of the process measured by indices such as the total fertility rate.

Nevertheless, the study of birth intervals faces several methodological problems. The problem of reporting errors in retrospective birth histories is known to have serious effects on the results. Both types of reporting error (omissions and timing) are obstacles to event history analysis in general as it depends on the computations of durations of time between successive events. Other problems of design are also important e.g. censoring and selectivity. The reproductive experiencece of a woman, who has not ended her reproductive career, is 'censored' by the interview. This means that her reported birth history is not complete as it only incudes the births that occured before the date of the interview and ignores the births that this woman may have at a later time. Therefore, the interview can occur before the termination of an interval, resulting in a measurement of a an 'open' interval, which may be 'closed' at a later date. Unless the function governing the childspacing process is known, it is not possible to determine the rates at which open intervals will be terminated by a birth (Marken and Shaps, 1970). Demographers usually use the life table approach to deal with these situations since no specific function is assumed for the closing of the interval. The methcd deals with the probabilities that women, at risk of births of a certain order, have such a birth within a specific duration of time e.g. three months. As in conventional life tables both closed and open intervals are included in the analysis. A woman observed at the time of the interview with an open interval of length x months will be included in the life table as being exposed to the risk of having the next birth for a period of x months. The rate of transition to the next parity is calculated by dividing the number of events (births of the next order) by the total number of months of exposures (closed as well as open intervals). It is obvious that a calculation of the mean length of the intervals for closed intervals only results in a biased result because only women who reached the next parity before the time of the interview are included.

Another design problem is selectivity. Many women in a young age group e.g. 25-29 at the time of the survey will not yet have experienced a fourth birth. Therefore, they will not be considered as exposed to the event of birth of a fifth child. On the other hand, the few women who will be included are those who have had four births at a fast pace. Those are women of early and fast fertilty who usually end up by having higher total fertility and shorter birth intervals than those who had their fourth birth at a higher age. The magnitude of the bias reslting from selectivity is shown to be large (Hobcraft and Rodriguez, 1980). Hobcraft and Rodriguez also show that selectivity problems can be overcome by the introduction of simple, demographic controls especially for the age at the initiation of the interval. Thus, coventional life tables with suitable controls were used for a number of years to analyse birth intervals for the purpose of determining the average time until the next birth occurs (e.g. Hobcraft and McDonald, 1984). When the purpose of the research was the investigation of the effect of certain factors on the speed of reproduction or the examination of the variation across subgroups of the population separate life tables were contstructed, for the different categories of the variables of interest (e.g. Abdel Aziz 1983). It is clear that the number of women left exposed to the risk of a next birth becomes smaller at the longer durations of the risk. Even with large samples, the cell-sizes decrease rapidly especially when several controls are imposed simultaneously (Rodriguez and Hobcraft, 1980, Rindfuss and Bumpass, 1977).

To overcome this difficulty, demographers considered the possibility of the construction of multivariate life tables. The question was also applicable to biostaticians interested in the factors acting from the onset or treatment of a disease. Two approaches were developed. The first is the life tables with covariates suggested by Cox (Cox, 1972) in which time is considered in its continuous form and uses hazards models. Through these models the covariates determine the risk or 'hazard' that a woman, exposed to the risk of having a birth of a specific order, will do so in a certain month after initiating the interval (e.g. Hobcraft and Casterline, 1985). The second approach considers time in its discrete form. Here, the interest is whether a woman, exposed to the risk, has a birth within a specific duration segment. The problem then is that of a regression analysis with a binary (0,1) dependent variable. Regression models have been used for this purpose e.g. ordinary least square OLS, logit or probit models and loglinear contingency table analysis (e.g. Akin et al 1981, Bumpass et al 1982).

The purpose of this paper is to introduce the reader to the two approaches of multivariate life tables. Although, our discusssion is focussed on birth interval analysis, the approaches are also applicabe to event history analysis in general e.g. infant mortaility, marriage dissolution, and contraceptive efficacy. The interval starts by the birth of the child, the marriage or the use of a contracpetive method respectively and is terminated by the infant's death, the dissolution of marriage or pregnancy. Both approaches have been discussed in a statistical style elsewhere (e.g. Cox and Oakes 1984, Tuma and Hannan, 1984, Miller, 1981). However, our aim is to present the techniques to the demographically oriented researcher and to review recent advances in the methodology of the analysis of birth intervals.

1. Hazard-model Analysis

The birth interval is initiated by a birth of a specific order. At each duration d (months since the initiation), a woman is a risk $\mu(d)$ of having another birth. In the traditional life table it is assumed that at a certain duration d, all women are at the same risk, i.e. $\mu(d)$ is the same for all exposed women. Alternatively, in life tables with covariates, it is assumed that the risk varies among women according to their individual socio-economic or demographic characteristics. The aim of the analysis becomes to examine how the covariates affect the risk. The first step is to define a functional form for the relationship between the covariates and the risk. The second step is to find the parameters that maximize the likelihood of observing the actual experience of the population, i.e. the maximum likelihood estimators.

Consider that data for a birth interval (of a specific order) are available together with information on two co-variates, one with j categories and the other with k categories. The first may be the mother's education with three categories (none, primary, above primary) and the other may be the type of place of current residence with two categories (urban, rural). The risk of transition (to the next birth) at duration (d) is given by

$$\mu_{jk}(d) = c \cdot \mu(d) \cdot c_j \cdot c_k$$

where $\mu_{jk}(d)$ is the risk at duration d for women in education category (j) and residence type (k). 'c' is a constant reflecting a risk common to all exposed women, and $\mu(d)$ is the effect of being at duration (d) which is common to all women exposed to the risk of transition at duration d (equivalent to age in the mortaility life table). c_j and c_k are multiplying factors specific for women of characteristics j and k. A multiplying factor which is greater than one will increase the risk and a factor less than one will decrease it. As it is more covenienent to work with additive models we use the log of the risk, i.e. a log-linear model.

$$\log \mu_{jk} = \log c + \log\mu(d) + \log c_j + \log c_k$$
$$R_{jk} = C + E(d) + E_j + E_k$$

The effects E(d), E_j and E_k are positive in cases when the risk increases by duration, education and type of residence and negative when it decreases. Several standard computer packages are available for estimating the parameters of log-linear models. A simple method is the contingency table analysis in which all of the used variables must be categorical (discrete). Within a given category individuals share the same risk. In our example we must, therefore, further treat time duration (d) as categorical. Thus the birth interval is divided into segments (i) (e.g. 5 segments) and within each duration segment, the risk is assumed constant. It is thus assumed that all individuals with the same co-variates and at a given duration segment have the same risk. These types of models are call 'hazards models'. Our model becomes:-

$$R_{ijk} = C + E_i + E_j + E_k \qquad (i = 1,2,\ldots,5;\ j = 1,2,3;\ k = 1,2)$$

where i is the duration segment (e.g. 9-18, 19-24, 25-30, 31-45, 46-60 months measured from the start of the interval), and E_i is the effect of being in that segment.

In this simple model, we considered only the main effect of each of the co-variates and duration on the risk. Also, this model is a 'proportional hazards' model because there is a fixed underlying hazard schedule (exp E_i) which is changed by the covariates (raised or lowered) by the same amount (E_j or E_k) at every duration. However, it is possible to consider a hazard model with interaction effects between duration segment and the two-covariates, i.e. assume that a covariate exhibits a different effect at the different duration segments (in this case it is not a proportional hazard model). Also, there may be interactions between the two covariates (e.g. the effect on the hazard of having no education in the urban areas is different from having no education while living in a rural area) and in this case it is not a main effect model.

To obtain the life table probabilities for the hazards model, consider the following types of respondents: a woman who is observed at the time of the interview with an open interval of length d, i.e. a women who has survived the risk of having a next birth for a duration of (d) moths and is censored by the interview. The probability of surviving the risk is

$$P_{jk}(d) = e^{-\int_0^d \mu_{ijk}(t)dt}$$

and

$$-\int_0^d \mu_{ijk}dt = -[\int_0^{w_1} \mu_{1jk}dt + \int_{w_1}^{w_1+w_2} \mu_{2jk}dt +$$

$$\int_{w_1+w_2}^{w_1+w_2+w_3} \mu_{3jk}dt + \ldots.]$$

$$= -[w_1\mu_{1jk} + w_2\mu_{2jk} + \ldots.]$$

where w_i = length of the duration segment

and $\Sigma w_i = d$.

On the other hand there are women who experience the event (birth of the next child) i.e. fail the risk at duration d. The probability of having a birth at duration d is a conditional probability q(d) and is the product of the probability of not having the birth before duration d and the probability of having it at exact duartion d (which is constant within the segment i).

$$q_{jk}(d) = P_{jk}(d) \cdot \mu_{jk}(d)$$

$$= P_{jk}(d)\, \mu_{ijk}$$

To obtain the likelihood function we multiply the probabiltiies P_{jk} over all women who were censored and the probabilities q_{jk} over all women who experience the event.

$$LF = \underbrace{\Pi P_{jk}(d)\, \mu_{ijk}}_{\text{had a birth}} \cdot \underbrace{\Pi P_{jk}(d)}_{\text{did not have a birth}}$$

$$= \underbrace{\Pi\, \mu_{ijk}}_{\text{had a birth}} \cdot \underbrace{\Pi P_{jk}(d)}_{\text{all exposed}}$$

$$\ln\ LF = \sum_i \sum_j \sum_k \left(B_{ijk} R_{ijk} - \mu_{ijk} w_i N_{ijk}\right)$$

where B_{ijk} is the number of women of characteristics j and k who experienced a birth in duration segment i.

N_{ijk} is the number of person-segments of exposure.

Thus $w_i N_{ijk}$ is the total months of exposure in the duration segment i (for those with closed and open intervals) and denoted E_{ijk}

$$\ln LF = \sum_i \sum_j \sum_k \left(B_{ijk} R_{ijk} - e\ R_{ijk} E_{ijk}\right)$$

To obtain the maximum likelihood estimators i.e. the parameters of the model that maximise the probability of observing the outcomes as they occurred, we take the derivative of the log likelihood function with respect to each parameter to be set at zero.

$$\frac{\delta \ln LF}{\delta R_{ijk}} = B_{ijk} - e^{R_{ijk}} E_{ijk}$$

and

$$B_{ijk} = e^{R_{ijk}} E_{ijk}$$

$$e^{R_{ijk}} = \frac{B_{ijk}}{E_{ijk}}$$

The estimated risk for each combination of i, j and k is the number of births divided by the total exposure and is exactly equal to that of a life table. This is a 'fully saturated' model because it reproduces the data exactly, i.e. the expected number of birth $\hat{B}_{ijk} = R_{ijk} \times E_{ijk} = B_{ijk}$. the number of parameters to be estimated is I x J x K. In our example we have 5x3x2=30 parameters to be estimated.

A simpler model for which we estimate a fewer number of parameters is the 'main effects' model. In this model the effect of each category of each covariate is the same for all duration segments and categories of the other covariates.

$$\ln \mu_{ijk} = R_{ijk} = C + E_i + E_j + E_k.$$

The MLE's of the parameters are obtained by the above procedure such that the number of expected events for all intervals i equals the actual number. Also, the expected number of events for individuals with a characteristic j or k must be equal to the actual number. This is the same restriction as the matching of margins criteria used in contingency-table analysis and justifies the use of the standard computer packages designed for loglinear analysis. We use the observed number of events and expousre in each cell as MLE to obtain life table functions q(x)'s. In a fully saturated model the MLE are the observed rates, otherwise the colutions are obtained by iterative numerical techniques. For the main effects model in our example we estimate 11 parameters only (one for the constant, 5 for duration segments, 3 for education categories and 2 for the residence categories). The estimated parameters are used for the construction of complete life tables. They may also be used directly to express relative risks. A woman in duration segment i=1 who has characteristics j=k=1 can be compared to that of a woman in duration segment i=2 and characteristics j=k=2. The relative risk is

$$\exp(\bar{E}_{i=1} - \bar{E}_{i=2} + \bar{E}_{j=1} - \bar{E}_{j=2} + \bar{E}_{k=1} - \bar{E}_{k=2})$$

It is often that we wish to examine the relative risk for one covariate when all other factors are held constant. If we wish - for example - to know the relative risk related to the category of k=1 (i.e. of living in an urban area), the relative risk is equal to $\exp(\acute{E}_{k=1} - \acute{E}_{k=2})$.

In the 'main effects' model an exponentiated parameter estimate is a relative risk when compared with the overall risk exp (c). A positive effect estimate indicates that the risk is higher for individuals in this category than that given by the constant and a negative estimate implies the opposite. It thus becomes sufficient to estimate a constant term and I-1 parameters for the duration segments, (J-1) and (K-1) parameters for the covariate categories. In our example we need only to estimate 8 parameters.

It is clear that the estimation of a fully saturated model with (IxJxK) parameters is not useful as it is a reproduction of the observed rates. The aim of the investigator is to define a model with relatively few parameters that will describe the process. The choice of the model becomes of greater importance as the cost of computer time becomes higher with the increase in the number of iterations (which is related to the number of cells). Usually, the chioce is made between a 'main effect model' or a main effect model with a few significant first order interactions. The investigaotor is then faced with the problem of determining the goodness of fit of the suggested models. If his chosen model is a 'restricted model' i.e. it is a subset of another 'unrestricted' model e.g. the fully saturated model, a χ^2 text can be used. When the difference is not significantly different from zero, the restricted model is selected i.e.

A number of books and papers are available for demographers dealing with life tables with covariates (Cox and Oakes, 1984; Tuma and Hannan, 1984; Trussell and Hammerslough, 1983; Breslow, 1984; Kalbfleisch and Prentice, 1980; Holford, 1980; Laird and Oliver, 1981 and Miller, 1981). The technique has been applied in several substantive research papers. We mention here a few (e.g. Rodriguez et al, 1984; Hobcraft et al 1985; Martin et al, 1982; Menken et al, 1981, 1982; Trussell et al, 1985)

2. Multivariate Regression

This approach was developed by empirical research and used by demographers who wish to avoid restrictive assumptions. The birth interval (of specific order) is divided into segments. For each segment a dependent variable is assigned whether or not the birth (the event of interest) occured during the segment, i.e. it is a dichotomous variable. It is necessary to use a conditional transition approach where - for each segment - the number of women exposed to the risk are those who did not experience the event in a previous segment. The number of events is the number of births that occured to the exposed women within the duration segment. Censored experiences i.e. intervals terminated by the interview during a segment (open intervals) are assumed to have contributed half of the segment's duration. The independent variables are those socio-economic and demographic characteristics of the respondent likely to have an effect on the length of the birth interval. A multivariate regression analysis is used to determine the effect of the different characteristics on the probability of having a birth within a specific segment. For example, consider the third birth interval i.e. the transition of women from the second to the third live birth. The interval may be divided into five duration segments 8-18, 19-24, 25-30, 31-45, 46-60 months. For the first segment the exposed women are all those who had a second birth. For the second duration segment, the exposed women are those who had a second birth and did not have a third birth in the first segment and so on. In each we observe whether or not the third birth occurs within the segment. The dependent variable takes a value of 1 if the event happens and zero if it does not.

The use of discrete variables in regression analysis has become increasingly important as individual level data have become more available. However, it is recognized that attempts to estimate linear probability models based on individual observations are generally unrealistic and lead to biased estimates. It is unrealistic because predicted probabilities outside the range (0,1) have no meaning. The estimators are biased because unlike simple regression using a continuous dependent variable when it is common to start with a linear additive model; there is reason to believe that the linear functional form may not describe the relationship adequately for a probability model. Empirical observation suggests that nonlinear S-shaped functions are more reasonable. This means that changes in the probability are difficult, as it gets closer to the limits (0,1). Estimating a linear probability model when the time distribution is s-shaped introduces a systematic error in the model. The errors are not independent of the exogenous variable. Also, the nature of the dependent variable adds to the problem of estimation. As we do not observe a probability, we only observe the outcome of the event which can only be 0 or 1, the error terms can only be 0 or 1 and will not be independent of the values of the dependent variable. In this case, the ordinary least square estimation will yield inaccurate results since one of its major assumptions is violated.

The difficulties described above can be reduced by the use of aggregate data. The dependent variable may be the proportion of the women who experience the event (the next birth) within a specific segment, rather than a dichotomous variable observed at the individual level. In this case (OLS) regression procedures can be used cautiously. However, when individual level data (such as that collected by WFS surveys) are available, they are preferred because of their richness and ability in clarifying the relationship between the variables as they do not suppress variations in the independent variables across individuals (Hanushek and Jackson, 1977). The problem then is the choice of a suitable model to describe the probabilities of the events. One of the most frequently assumed underlying distribution is the logistic distribution. In the following paragraphs we describe both methods.

(a) Probability linear models

As mentioned before, we are dealing with a contingency table that presents the joint distribution of counts and a set of factors (categorical regressors). It is necessary to define a response variable and to consider variations in the conditional distribution of that variable over the other variables. Our response variable is a dichotomous variable (whether the next birth takes place or not within a specific segment). For each combination of the factors we form the proportion (i.e. the number of cases for the response variable = 1) so that its conditional distribution given the characteristics j and k can be studied using a two way table of proportions. Because each proportion is also the mean of the response variable (because it is a dichotomous variable taking the values 0,1 only) we now have a table of means, cross classified by the factors expected to influence the deviations from the grand mean. Although there are some theoretical limitations of this procedure, they may be ignored when the proportions lie between 20 and 80% (Little and Pullum, 1979). The analysis of cross tabulations of means decomposes the observed values into fitted values and residuals, and aims at developing a parsimonious description of the data which leaves residuals which are relatively free of explainable patterns. This includes methods of direct standardization (a simple technique of controlling categorical predictors) and multiple classification analysis (MCA) (Little, 1980).

MCA is a method which has the advantage of dealing with predictor variables as weak as nominal measurements (Ogawa, 1980). It finds a best fitting additive structure in which the sum of squares of the average residuals weighted by the sample size in each cell is minimized.

$$SS = \sum_{j=1}^{J} \sum_{k=1}^{K} n_{jk} (\bar{y}_{jk} - \hat{\mu}_{jk})^2$$

where n_{jk} is the number of cases in each cell.

$\bar{y}_{jk}$ is the sample mean in cell jk (proportion having a next birth)

$\hat{\mu}_{jk}$ is the fitted mean in cell jk which takes the additive form calculated by MCA and $\hat{\mu}_{jk} = m + r_j + c_k$ where m is the weighted average of the adjusted means for either factors.

r_j and c_k are the adjusted effects of the factors. MCA calculates fitted values ($\hat{\mu}_{jk}$) that minimize (SS).

An important property of MCA is that it weights the sqaured residuals by the sample size in each cell (n_{jk}). Empty cells are therefore ignored because they are given the weight zero. Also when the sample size of a particular cell is small, it is given less weight which implies that the observed means are allowed to deviate more from the observed means i.e. subject to large variance. This makes MCA an approprite method for calculating the adjusted effects when the cell sample size is small.

Several computer packages provide MCA in the form of an MCA table where the effects (r_j and c_k) are presented as deviations from the overall mean. The table contains the sample size for each cell of the cross tabulation and the unadjusted deviations from the overall mean. It also contains the deviations for each factor, adjusted for the other factors. Some programmes also provide some summary measures of association (eta^2 and beta^2) which are correlation ratios interpreted as the proportions of the total variance explained by the unadjusted and adjusted effects for each factor respectively.

It is customary to wish to summarize a set of effects, indicating the overall magnitude of the differences between the category means (proportions). This is in order to test for the significance of the effects. The method used is the analysis of variance ANOVA which is based on the same additive model of MCA. MCA has been used in several demographic studies (e.g. Rindfuss and Bumpass, 1977; Jain, 1969; Jain and Bongaarts, 1981). It was also used by Bumpass et al (Bumpass et al, 1982) in an analysis that compares the effects of contraception and infant mortality on the pace of fertility in Korea and the Philippines. They examine the transition probabilities within successive duration segments within an interval. This methodological choice is based on the feeling that breast feeding should have greater effect at the early segments of exposure than at the later segments, a pattern that may not be applicable to other intermediate variables e.g. contraception (i.e. the proportionality assumption). This method allows for the multivariate analysis within the structure of the life table as it uses the probabilities q(x)'s as the response variable.

(b) Logistic Regression

As mentioned before, the basis for many experimental designs is to estimate a probability by the relative frequency outcome, given a large sample size. A similar approach can be used for nonexperimental research when many observations for each value of x (a factor) are available. The observed frequency of occurance of an event within each category is a good estimate of the true probability. The next step becomes the specification of the functional form F(x,B) which relates these probabilities to the values of the explanatory variables. These variables are categorized $(X_j)(1,2,\ldots,j)$ and in each category there are N_j observation. The estimated probabilities for category j is

$$\hat{P}_j = \frac{i}{N_j} \sum_{i=1}^{N_j} Y_{ij}$$

Assuming $F(X,\beta)$, the cumulative function; that describes the relationship between the probabilities (estimates of q(x)'s) and the demographic and socio-economic factors is a logistic distribution;

$$P_j = F(X_t\beta) = \text{Prob } (Y_t = 1)$$

$$P = \frac{1}{1+e^{-X\beta}}$$

$$\text{and } 1-P = \frac{e^{-X\beta}}{1+e^{-X\beta}} = \frac{1}{1+e^{X\beta}}$$

Define L (the logit) to be the log of the odds ratio

$$L = \log \frac{p}{1-p} = \log p - \log(1-p)$$

$$= - \log(1+e^{-X\beta}) - [\log e^{-X\beta} - \log(1+e^{-X\beta})]$$

$$= - \log e^{-X\beta} = XB$$

$$Lj = X_j B$$

As $x\beta$ ranges from $-\alpha$ to α, the probability p is bounded (0,1) but L is not and it ranges from $-\alpha$, α. This shows that L_j (the logit) is a linear function of the factors and can be estimated by $\hat{p}_j$ using least squares. The equation may be written as a series of dummy variables. Suppose we have two explanatory variables (j equals education and k type of place of current residence). The first with three categories (none, primary, above primary) and the second with two (urban, rural). This results in a table with six cells where the entries are the relative observed frequencies. The model considers the logits (log of the odds of these probabilities) to be a linear function of the categorical variables.

$$L_{jk} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 = X\beta$$

where $X_1 = 1$ for all cells, $X_2 = 1$ for cells with variable j = '2' primary, 0 otherwise, $X_3 = 1$ for cells with variable j = '3' above primary, 0 otherwise, X_4 for cells with variable K = '2' rural, 0 otherwise. Note that each cell is denoted by a unique combination of values for the dummies. The logit for each cell is obtained from the coefficients (β) for the appropriate combination. Thus the logit for one variable is not affected by the values of the other factors. For example the difference $\beta_3 - \beta_2$ is the expected difference in the logit between observations in category 3 'above primary' and 2 'primary' regardless of the type of place of residence

$L_{11} = \beta_1$ (none, urban) $\qquad$ $L_{12} = \beta_1 + \beta_4$ (none, rural)

$L_{21} = \beta_1 + \beta_2$ (primary, urban) $\qquad$ $L_{22} = \beta_1 + \beta_2 + \beta_4$ (primary, rural)

$L_{31} = \beta_1 - \beta_3$ (above, urban) $\qquad$ $L_{32} = \beta_1 + \beta_3 + \beta_4$ (above, rural)

$$\beta_3 - \beta_2 = L_{31} - L_{21} = L_{32} - L_{22}$$

The next step is to estimate the above equations. Assuming observations in each cell are independent, F_{jk} (the observed frequencies) are bi normally distributed about the true probability P_{jk}

$$E(F_{jk}) = P_{jk} \text{ and } Var(F_{jk}) = \frac{P_{jk}(1-P_{jk})}{N_{jk}}$$

and

$$\hat{L}_{jk} = \log \frac{F_{jk}}{1-F_{jk}}$$

$$\hat{L}_{jk} = X_{jk}\beta + (\hat{L}_{jk} - L_{jk}) = X_{jk}\beta + U_{jk}$$

The error terms (U_{jk}) asymptotically have a zero mean and a variance inversely related to $P_{jk}(1-P_{jk})$ and N_{jk} (Hanushek and Jackson, 1977). Therefore generalised least squares estimation procedure is used where the cells are weighted by the reciprocal of the variances.

When the above logit analysis procedure is repeated for each duration segment, the result is a number of coefficients. If the birth interval is divided into i segments and the number of independent (explanatory) variables is j, the result is a matrix of coefficients of the dimension jxi. The columns of this matrix show the structure of the determinants of the probability of experiencing a next birth within a segment. The rows of coefficients show how the effect of a specific variable changes across the segments. For each coefficent there is a test of significance in addition to a vector of χ^2 coefficents that indicate the goodness of fit of the model in each duration segment. The coefficients can be used to predict the probabilities of transition (q_x) and thus to create a life table for individuals with a given set of characteristics i.e. q_x.

Logistic models have been recently used for event history analysis in several researches. We mention a few such as e.g. Rindfuss et al 1985, Bumpass et al, 1986.

Summary and Conclusions

In this paper two approaches were presented for dealing with multivariate analysis of life tables. The need for this type of analysis became obvious after individual leveld ata became available by the recent WFS surveys and because of the growing concern for regarding fertility as a sequential process rather than looking at its final outcome. The analysis has become possible by the availability of several computer packages designed for loglinear analysis.

The first approach uses a hazard model in which the risk of transition to a higher parity is allowed to vary among women of different characteristics at different durations. The second treats every duration segment separately and uses a loglinear model to describe the relationship between the proportion of women having a birth within the segment and the women's characteristics.

Several references are given to the statistical background material on the methods as well as to some substantive demographic research in which these methods were applied. The choice of a particular method depends on the type of the available data, and - to a large extent - on the availability of suitable computer software.

References

- Abdel Aziz, Abdallah 'A Study of Birth Intervals in Jordan' WFS Scientific Reports No 46, 1983

- Akim et al The Determinants of Breastfeeding in Sri Lanka Demography 18 (3), 1981

- Breslow, N 'Methods for identifying mortality risk factors in longitudional studies'. In Methodologies for the Collection and Analysis of Mortality Data by Vallin et al (eds)

- Bongaarts, J 'A Framework for Analysing the Proximate Determinants of Fertility', Population and Development Review 4 (1), 1978

- Bulatao, R 'Values and Disvalues of Children in Successive Child-Bearing Decisions' Demography 18 (1), 1981

- Bumpass, L et al 'Intermediate Variables and Educational Differentials in Fertility in Korea and the Philippines' Demography 19 (2), 1982

- Bumpass L., Rindfuss R. and Palmore J. 'Determinants of Korean Birth Intervals' The Confrontation of Theory and Data' Population Studies 40, 1986

- Cox, D.R. 'Regression Models and Life-tables' Journal of the Royal Statistical Society B 34 (2), 1972

- Cox, D.R. and Oakes, D 'Analysis of Survival Data' London, New York: Chapman and Hall, 1984

- Davis, K. and Blake, J 'Social Structure and Fertility: An Analytic Framework' Economic Development and Cultural Change 4 (4), 1956

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

- Hobcraft, J. and Rodriguez, G. 'Methodological Issues in Life Table Analysis of Birth Histories' Seminar on the analysis of Maternity Histories, IUSSP and CCAF, London 1980

- Hobcraft, J. and McDonald, J. 'Birth Intervals' WFS Comparative Studies No 28, 1984

- Hobcraft, J. and Casterline, J. 'Speed of Reproduction' WFS Comparative Studies No 25, 1983

- Hobcraft, J., McDonald, J. and Rutstein, S. 'Demographic determinants of infant and early child mortality: A comparative analysis' Population Studies 39 (3), 1985

- Holford, T. 'The Analysis of Rates and Survivorship using log-linear models' Biometrics 36, 1980

- Jain, J. 'Socio-economic correlates of fecundability in a sample of Taiwanese women' Deomgraphy 6 (1), 1969

- Kalbfleisch, J. and Prentics, R. 'The Statistical Analysis of Failure Time Data', NY John Wiley and Sons, 1980

- Laird, N. and Oliver, D. 'Covariance Analysis of censored survival data using log-linear analysis techniques' Journal of the American Statistical Association 76, 1981

- Little, R. and Pullum, T. 'The General Linear Model and Direct Standardization: A Comparison' WFS Occasional Papers No 20, 1979

- Little, R. 'Linear Models for WFS data' WFS Technical Bulletins No 9 1980

- Martin, L. et al 'Co-variates of child mortaility in the Philippines Indonesia and Pakistan: An analysis based on hazard methods' Population Studies 37 (3), 1983

- Menken, J. and Sheps, M. 'On Relationships Between Longitudional Characteristics and Cross-Sectional Data' American Journal of Public Health 60 (8), 1970

- Menken, J. et al 'Proportional hazards life table models: an illustrative analysis of socio-demographic influence on marriage dissolution in the United States' Demography 18 (2), 1981

- Miller, R. 'Survival Analysis' NY, Wiley, 1981

- Ogawa, N. 'Multiple Classification Analysis and its Application to the 1974 Fiji Fertility Survey' in Regional Workshops on Techniques of Analysis of World Fertility Survey Data. WFS Occasional Paper No 22, 1980

- Rindfuss, R. and Bumpass, L. 'Fertility During Marital Disruptions' Journal of Marriage and the Family 39 (3), 1977

- Rindfuss et al 'Child spacing in Asia: Similarities and Differences WFS Comparative Studies No 29, 1984

- Rodriguez, G. and Hobcraft, J. 'Illustrative Analysis: Life Table Analysis of Birth Intervals in Colombia' WFS Scientific Reports No 16, 1980

- Rodriguez, G. et al 'A Comparative Analysis of Determinants of Birth Intervals' WFS Comparative Studies No 30, 1984

- Trussell, J. and Hammerslaugh, C. 'A Hazard Model Analysis of the Covariates of Infant and Child Mortaility in Sri Lanka' Demography 20 (1), 1983

- Trussell, J. et al 'Determinants of Birth Interval Length in the Philippines, Malaysia and Indonesia: A Hazard Model Analysis' Demography 22 (2), 1985

- Tuma, N. and Hannan, M. 'Social Dynamics: Models and Methods' Academic Press, 1984

جدول رقم (٢) النتائج النهائية لاسلوب الحذف التتابعي لبيانات الاردن .

```
Backward Elimination for DESIGN 1 with generating class
  V48*V44*V49*REG
Likelihood ratio chi square =    0.0       DF = 0  P = 1.000

If Deleted Simple Effect is                 DF  L.R. Chisq  Change  Prob   Iter

 V48*V44*V49*VREG                            9               7.779  .4553    6
Step 1
  The best model has generating class
      V48*V44*V49
      V48*V44*REG
      V48*V49*REG
      V44*V49*REG
__Likelihood_ratio_chi_square__=__7.77934__DF = 8__P =_.455__________
Page 45                              SPSS/PC+                       1/1/80
If Deleted Simple Effect is                 DF  L.R. Chisq  Change  Prob   Iter
  V48*V44*V49                               10              18.807  .0428    6
  V48*V44*REG                                9              27.105  .0013    4
  V48*V49*REG                                3               9.157  .0273    4
  V44*V49*REG                                3              10.906  .0122    5

Step 2

  The best model has generating class

      V48*V44*V49
      V48*V44*REG
      V48*V49*REG
      V44*V49*REG

  Likelihood ratio chi square  =   7.77934      DF = 8   P =   .455
-------------------------------------------------------------------

  The final model has generating class

      V48*V44*V49
      V48*V44*REG
      V48*V49*REG
      V44*V49*REG


  The Iterative Proportional Fit algorithm converged at iteration 0.
  The maximum difference between observed and fitted marginal total
is    .202  and the convergence criterion is   .250
```

جدول رقم (1) النتائج النهائية لأسلوب التحدد التتبعي لبيانات مصر.

```
Backward Elimination for DESIGN 1 with generation class
  V48*V44*V49*REG

Likelihood ratio chi square  =     0.0          DF = 0   P = 1.000

If Deleted Simple Effect is                 DF  L.R. Chisq Change  Prob   Iter
  V48*V44*V49*REG                            7            10.498  .1621     7

Step 1

  The best model has generating class

      V48*V44*V49
      V48*V44*REG
      V48*V49*REG
      V44*V49*REG

  Likelihood ratio chi square =   10.49809   DF = 7   P = .162
________________________________________________________________________________

Page 34                           SPSS/PC+                            6/12/88

If Deleted Simple Effect is                 DF  L.R. Chisq Change  Prob   Iter

   V48*V44*V49                               9           100.383  .0000     6
   V48*V44*REG                              10            17.362  .0667     8
   V48*V49*REG                               3            12.499  .0059     4
   V44*V49*REG                               3            14.054  .0028     6

Step 2
   The best model has generating class

      V48*V44*V49
      V48*V49*REG
      V44*V49*REG

Likelihood ratio chi square =   27.86026    DF =' 17   P = 0.47

If Deleted Simple Effect is                 DF  L.R. Chisq Change  Prob   Iter


   V48*V44*V49                              10           101.223  .0000     6
   V48*V49*REG                               3            13.423  .0038     6
   V44*V49*REG                               3            16.321  .0010     8
________________________________________________________________________________
Page 35                           SPSS/PC+                          6/12/88
Step 3
   The best model has generating class
   V48*V44*V49
   V48*V49*REG
   V44*V49*REG
Likelihood ratio chi square =  27.86026   DF = 17   P = .047
________________________________________________________________________________
Page 36                           SPSS/CP+                            6/12/88
The final model has generating class
   V48*V44*V49
   V48*V49*REG
   V44*V49*REG
The Iterative Proportional Fit algorithm converted at iteration 0.
The maximum difference between observed and fitted marginal totals is    .215
and the convergence criterion is     .025
```