Application of Design of Experiments in Population Analysis

Keshab C. Bhuiyan*

Abstract

One of the components of population analysis is the study of differential fertility among different groups of mothers classified by child mortality level, regional variation and / or socioeconomic status. Among different socioeconomic group of mothers, the levels of fertility or the levels of child mortality are different. The homogeneity of differentials in fertility levels or in child mortality levels can be investigated using the technique of combined analysis of groups of experiments under both homogeneous and heterogeneous conditions. In the present paper, homogeneity of fertility levels by child mortality levels and homogeneity of child mortality levels by number of ever born children recorded from North–Eastern Libya are investigated. The data show that child mortality levels do not vary due to the variation in the number of ever born children of mothers. The level of occupation of mothers changes with the change in their level of education. The study of level of fertility variation according to changes in level of occupation nested within the levels of education is of interest. Again, fertility variation is influenced by duration of marriage, level of education of husband, duration of breastfeeding period etc. Thus, fertility variation should be studied after eliminating the impacts of the above said variables. This can be done using the technique of covariance analysis in nested classification. The method of analysis is explained with real world data.

Introduction

Among the important components of population analysis, the study of fertility and / or mortality differentials of different socio-cultural groups of mothers is of interest to the researchers. The impact of socioeconomic factors on level of fertility varies from one group of mothers to another group when they are classified by socio-cultural factors. The homogeneity of impacts and the impact for all mothers together are of interest in the field of population analysis. These can be done if one applies the technique of combined analysis of groups of experiments under both homogeneous and heterogeneous environment. Keeping this

* Associate Professor of Statistics, School of Science, American International University-Bangladesh (AIUB).

view in mind, Bhuyan's (1985) work has been redesigned to show its application in population analysis.

The fertility level varies with the variation in the levels of socioeconomic status of women. But levels of social status are assumed nested within the levels of education and levels of occupation are nested within the levels of education and so on. In such population analysis, if fertility is the response variable, its Ievel can be studied for each level of social status nested within the levels of education after adjusting the effects of different demographic factors. This can be done by covariance analysis.

Under usual assumptions, the covariance analysis is available for cross-classification. The analysis in nested classification to study the fertility differential has been reported by Budijaja and Bhuyan (1999). The method is reported here in short with a new example.

Methodology

Combined Analysis of Groups of Demographic Data:

Let n mothers be investigated to estimate their child mortality and fertility Ievels. Let these mothers be divided into p categories according to socioeconomic variable and for h –th group of mothers the vector of estimate of any population characteristic is $\hat{F}_h$ (h = 1,2, ……….., p) where $V(\hat{F}_h) = W_h$. Let F_h be the related population value for h-th group of mothers. Let $\hat{W}_1, \hat{W}_2, \hat{W}_3, \ldots\ldots, \hat{W}_p$ be the unbiased estimates of W_1, W_2 , W_3 , ……….., W_p and distributed independently of $\hat{F}_h$ and of each other in Wishart forms with f_1 , f_2 , ……..., f_p d.f., respectively.

The objective is to test the hypothesis

$$H_o \ : F_1 = F_2 = \ldots\ldots = F_p \qquad (2.1)$$

A test statistic to test the significance of the hypothesis is formulated as

$$T = \sum (\hat{F}_h - \hat{F}) W_h^{-1} (\hat{F}_h - \hat{F}) \ldots\ldots\ldots\ldots \qquad \ldots(2.2)$$

[James (1954), Bhuyan (1985)] which is exactly distributed as chi-square with 2 (p – 1) d.f. if W_h's are known or f_h's are large. For small value of f_h, T is to be compared with James suggested approximate χ^2 value, say H. The comhined estimate of population characteristic for all groups of mothers is given by

$$\hat{F} = \hat{W}^{-1} \sum \hat{W}_h \hat{F}_h \text{ where } \hat{W} = \sum \hat{W}_h$$

If F_h contains only estimate of fertility level or mortality level of h – th group of mothers, then the required test statistic is

$$\sum_{h=1}^{p} \hat{W}_h \hat{F}_h^2 - \frac{\left(\sum \hat{W}_h \hat{F}_h\right)^2}{\hat{W}} \quad \text{.......... ....} \quad (2.3)$$

And this statistic is distributed as chi-square with (p – 1) d.f. provided f_h's are large. For f_h not large enough, James (1951) suggested to compare the calculated statistic value of (2.3) with H_1, where

$$H_1 = \chi^2 \left[1 + \frac{3\chi^2 + (p-1)}{2(p^2-1)} \sum \frac{1}{f_h} \left(1 - \frac{\hat{W}_h}{\hat{W}} \right)^2 \right]$$

where χ^2 is the table value of χ^2 at $\alpha\%$ level of significanee with $(p-1)$) d.f.

Covariance Analysis in Nested Classification

Let us assume that the levels of social status (S) of women are nested within the levels of their occupation (O) and the levels of occupation are nested within the levels of education (E). Let y_{ijlm} be the number of ever born children of m-th mother of l-th social status nested within the j-th level of occupation and i-th level of education [i = 1,2 ….., p ; j = 1,2 ….,q_i ; l = 1,2, …..r_{ij} ; m = 1,2, … …, n_{ijl}]. Let us assume that this response is affccted by k concomitant variables x_1,x_2 , … x_k

; viz. duration of marriage, child mortality, duration of breastfeeding etc. We write

$\Sigma q_i = q$, $\Sigma r_{ij} = R_{i.}$, $\Sigma\Sigma r_{ij} = \Sigma R_{i.} = R$, $\Sigma n_{ijl} = N_{ij.}$, $\Sigma\Sigma n_{ijl} = N_{i..}$, $\Sigma\Sigma\Sigma n_{ijl} = n$

The model for the response is

$$y_{ijlm} = \mu+\alpha_i+\beta_{j(i)}+\psi_{l(ij)}+\gamma_1(x_{1ijlm} - \bar{x}_{1....})+\gamma_2(x_{2ijlm} - \bar{x}_{2....})+ \\ + \dots\dots\dots\dots +\gamma_k(x_{kijlm} - \bar{x}_{k....})+e_{ijlm} \tag{2.4}$$

where μ = general mean, α_i = effect of i-th level of E, $\beta_{j(i)}$ = effect of j-th level of O within i-th level of E, $\psi_{l(ij)}$ = effect of l-th level of S within j-th level of O and i-th level of E, γ_s = effect of s-th (1,2, ..., k) concomitant variable on y, e_{ijlm} = random component, x_{sijlm} = the value of s-th concomitant variable corresponding to y. The model can be written as

$$Y = \begin{bmatrix} 1 & A & B & C & V \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \beta \\ \varphi \\ \gamma \end{bmatrix} + U$$

$$= MP + U \tag{2.5}$$

where Y , M , P , and U have been defined in Budijaja and Bhuyan (1999).

The sums of squares of estimates for the model (2.5) is given by $\hat{P}'MY$. It has R + k d.f. and the error sum of squares is $S_1 = Y'Y - \hat{P}'M'Y$ has (n – R – k) d.f.

One of the objeetives of the analysis is to test the hypothesis

$$H_o : \gamma = 0 \tag{2.6}$$

Under this hypothesis the model is

$$Y = M_1 P_1 + U \tag{2.7}$$

where $M_1 = [1\ A\ B\ C]$, $P_1 = [\mu\ \alpha\ \beta\ \psi]'$
Here rank $(M_1) = R$. Therefore, under the restrictions cited above, the estimates of elements in P are

$$\hat{\alpha} = (A'A)^{-1} A'Y - n^{-1}E_p 1'Y, \hat{B} = (B'B)^{-1} B'Y - D(A'A)^{-1} A'Y$$
$$\hat{\psi} = (C'C)^{-1} C'Y - D_1(B'B)^{-1} B'Y, \mu = n^{-1}[1'Y]$$

Now, under the null hypothesis (2.6) , the sum of squares due to $\hat{\gamma}$ is $S_2 - S_1$ having k d.f , where $S_2 = Y'Y - \hat{P}_1' M_1' Y$. This S_2 has (n – R) d.f. Thus, under the usual assumption of analysis of variance the test statistic to test the hypothesis (2.6) is

$$F = \frac{(S_2 - S_1)/k}{S_1/(n - R - k)}$$

where $F \sim F_{k,n-R-k}$
If no significant evidence is concluded against H_o, the parameter γ can be deleted from the model (2.5) and the analysis of the model can be performed by usual technique for the analysis of unbalanced three–stage nested classification [Searle (1979)]. However, the rejection of (2.6) does not mean that all γ's have significant influenee on response variable. In such a situation, one may be interested to test the hypothesis

$$H_o: \gamma_s = 0 \text{ for all } s = 1,2,\ldots\ldots,k \qquad (2.8)$$

The test statistic for this hypothesis is
This test will enable the researcher to drop s–th concomitant variable from further analysis.

$$F = \frac{S_3 - S_1}{S_1/(R + k - 1)}, \text{ where} F \sim F_{1,R+k-1}$$

The main analysis with the retained concomitant variables is to test the hypotheses

$$H_o : \psi = 0 \qquad (2.9)$$
$$H_o : \beta = 0 \qquad (2.10)$$
$$H_o : \alpha = 0 \qquad (2.11)$$

The test statistics for these hypotheses are,

$$F_1 = \frac{(S_4 - S_1)/(R-q)}{S_1/(n-R-k)}$$

$$F_2 = \frac{(S_5 - S_1)/(q-p)}{S_1/(n-R-k)} \qquad F_3 = \frac{(S_6 - S_1)/(p-1)}{S_1/(n-R-k)}$$

Respectively, $F_1 \sim F_{R-q,n-R-k}$, $F_2 \sim F_{q-p,n-R-k}$, $F_3 \sim F_{p-1,n-R-k}$

The values S_3, S_4, S_5, and S_6 are obtained in a similar way as S_1 and S_2 are obtained. The adjusted fertility/child mortality levels corresponding to different socio-demographic factors can be calculated using the technique suggested by Budijaja and Bhuyan (1999).

Data

The secondary data collected by Ghffar (1996) have been utilized to show the application of the technique of combined fertility estimate of mothers. The data have been collected from North–Eastern Libya Stretching from Benghazi to Emsaad. In the said area, there are 27 localities. Seven localities are selected by probability proportional to sizes of number of families living in the localities. From the selected localities 2 per cent couples of childbearing ages are investigated. In all 1252 couples are investigated and the information related to fertility and child mortality are recorded.

Bhuyan et al (1996) have collected data from 16 contiguous villages of both Savar and Dhamrari upazila under Dhaka district. In these villages, there are 4635 households. Twenty percent of these are selected randomly for investigation. However, during investigation information from 890 couples of childbearing ages have been recorded by personal interview. These data have been utilized to show the application of covariance analysis in nested classification. The level of fertility of a mother is recorded by her number of ever born children [y, alive+dead+married]. It has been observed that 95% of the standardized values of fertility lie between –2 to +2 indicating normality of data.

Results

Combined Fertility Estimate:

The investigated mothers are divided into 6 groups according to their level of child mortality. The distribution of mothers and their fertility level by child mortality level is shown in Table 1. The fertility level is measured by average ever born children.

Table 1: Distribution of motbers and tbeir fertility level by child mortality Level.

Level of child mortality	0	1	2	3	4	5	Total
Number of mothers	805	306	93	36	7	5	1252
%	64.3	34.4	7.4	2.9	0.5	0.5	100.0
Level of fertility	5.08	7.40	9.48	9.67	10.29	10.60	6.16
s.d. of fertility	2.91	2.82	1.80	2.00	1.25	0.89	3.20

The variances of the estimates of fertility are found heterogeneous by Bartlett's (1937) χ^2 test where $\chi^2 = 45.53$ with 5 d.f. Therefore, to test the homogeneity of the levels of fertility of different groups of mothers, the test statistic (2.3) is applied, where the calculated value of the statistic is 4.30. This value is compared with $H_1 = 12.37$,where the d.f. of the test statistic is 5 .Since the calculated value of the test statistic is smaller compared to the theoretical value, it is concluded that the levels of fertility of mothers by their child mortality levels are homogeneous. The overall fertility level of mothers is found out as 9.92 per mother.

In a similar way , the homogeneity of child mortality levels of mothers classified into 11 groups according to their level of fertility is studied. The distribution of mothers and their child mortality levels by their fertility levels are shown in Table 2. The level of child mortality is measured by average number of dead children per mother.

Table 2: Distribution of motbers and tbelr child mortality levels by fertility levels.

Level of fertility	1	2	3	4	5	6	7	8	9	10	11	Total
No. of mothers	95	122	99	112	126	125	119	100	93	95	166	1252
%	7.6	9.7	7.9	8.9	10.1	9.9	9.5	7.9	7.4	7.6	13.3	100.0
Level of child mortality	0.05	0.11	0.17	0.25	0.29	0.33	0.39	0.63	0.95	0.82	1.45	0.53
s.d.	0.22	0.34	0.38	0.56	0.54	0.59	0.61	0.81	1.06	0.85	1.30	0.87

The variances of estimates of mortality levels are heterogeneous as is observed by Bartlett's (1937) χ^2 - test, where $\chi^2 = 591.01$ with 10 d.f. Thus, the statistic (2.3) is used to test the homogeneity of the levels of child mortality of mothers. The calculated statistic is 2.94 which is to be compared with the theoretical value 18.07 with 10 d.f. Since the calculated value is much less than the theoretical value, the homogeneity of the child mortality levels is concluded. The overall estimated child mortality level is 0.21. However, increasing trend in child mortality level is observed with the increase in the number of ever born children.

Covariance Analysis:

The investigated mothers are classified by their employment status assuming that employment levels are nested within their levels of education. The number of mothers according to different classes along with their fertility levels are presented in Table 3.

The levels of fertility are influenced by duration of marriage, breastfeeding period, husband education, sociocconomic condition etc. [Majumder and Bhuyan (1993), Bhuyan et al (1996)]. Therefore, attempts have been made to estimate the levels of fertility after eliminating the impacts of the above said variables. This is done by covariance analysis in two-stage nested classification. The adjusted fertility levels are shown in the column heading A [U for unadjusted fertility level.] in Table 3.

All the adjusted number of ever born children are found less than the unadjusted ones. This has happened due to the negative impact of education of husband socio-economic condition and breastfeeding period. Only duration of marriage has a positive impact on number of ever born children.

Table 3: Distribution of Mothers Along with their Fertility Level according to the Levels of Occupation Nested within the Levels of Education.

Level of Education	Level of Occupation	No. of mothers	%	Fertility level by adoption behavior			
				Yes		No	
				U	A	U	A
Illiterate	House wife	316	69.0	3.60	3.21	3.95	3.52
	Work Outside	142	31.0	2.92	2.14	3.42	2.48
	Sub Total	458	51.5	3.39	2.95	3.86	3.14
Primary	House wife	125	65.8	3.41	2.96	3.54	3.33
	Work Outside	65	34.2	2.62	1.96	2.96	2.52
	Sub Total	190	21.3	3.11	2.86	3.08	2.48
Secondary and above	House wife	160	66.1	2.38	1.89	2.75	2.51
	Work Outside	82	33.9	2.05	1.88	2.42	2.12
	Subtotal	242	27.9	2.08	1.80	2.60	2.42
Total	House wife	601	67.5	2.64	2.32	3.43	3.06
	Work outside	289	32.5	2.89	2.51	3.48	3.11

Conclnsion

The analytical results show that with the increase in child mortality level of Libyan mothers there is an increase in the level of fertility. However, the levels of fertility of different groups of mothers are homogeneous. The overall fertility level of mothers is found out as 9.92 per mother.

The fertility level of Bangladeshi mothers are influenced by duration of marriage, breastfeeding period, husband's education, socioeconomic condition, etc. The adjusted fertility level of mothers are found out after eliminating the influence of the above said variables. This has been done by covariance analysis and adjusted fertility levels are observed smaller compared to that of unadjusted ones. The results indicate that fertility level can further be decreased, even among educated mothers, if duration of marriage is reduced and duration of breastfeeding is increased.

References

Bartlett, M.S. (1937): *"Properties of Sufficiency and Statistical Test"*, Proc. Roy. Soc., Ser., A, 160, 268 – 282.

Bhuyan, K.C. (1985): *"Combined Analysis of a Group of Experiments with a Heteroskedastic Model"*. Sankhya, B, 47, 224 – 232.

Bhuyan, K.C., Majumder, A.K. and Chaudhury, K. (1996): "Differential Fertility among Adopter and Non-adopter Couples Differing Social Status and Mobility Pattern in Rural Bangladesh", *Journal of Family Welfare*, 42.

Budijaja, K.Y. and Bhuyan, K.C. (1999): *"On Covariance Analysis in Nested Classification to Study the Fertility Differential in North- Eastern Libya*, Cal. Stats. Assoc. Bulletin, 49, 193-194, 101 – 109.

Ghffar, A.F. (1996): *Socioeconomic Factors Influencing Child Mortality in North Eastern Libya.* Unpublished M.Sc. Thesis, Garyounis University, Libya.

James, G.S. (1951): "The Comparison of Several Groups of Observations When the Ratios of Population Variances are Unknown", *Biometrika*, 38, 324 – 329.

James, G.S. (1954): "Test of Linear Hypothesis in Univariate Analysis when the Ratios of the Population Variances are Unknown", *Biometrika*, 41, 19 – 43.

Majumder, A.K. and Bhuyan, K.C. (1993): "Fertility Differentials According to Levels of Education and Family Planning Practices : A Case Study in Bangladesh", *Jahangirnagar Review*, 17, 159 – 170.

Searle ,S.R. (1979): *"Alternative Covariance Analysis Models for the Two-Way Crossed Classification"*, Commun. Statist. Ser. A, 8, 799 – 818.