ISSN: 2641-6336
Annals of
Biostatistics & Biometric Applications
DOI: 10.33552/ABBA.2019.01.000521
Iris Publishers

Short Communication

Principal Component Analysis of Factors Affecting Ovulation Interval of Selected Women

Matthew Chukwuma Michael[1]*, Eriobu Nkiru Obioma[2] and Ashinze Mkpuruoma Akudo[3]

[1,3]*Department of Mathematics and Statistics, Nigeria*

[2]*Department of Statistics, Nigeria*

***Corresponding author:** Matthew Chukwuma Michael, Department of Mathematics and Statistics, Delta State Polytechnic, Ogwashi Uku, Delta State, Nigeria.

Received Date: February 21, 2019
Published Date: February 27, 2019

Abstract

The factors affecting ovulation interval, like another biological phenomenon are numerous and interdependent. This paper identified the factors to include Age, Height, Weight, Work Time (Stress), Menstrual Duration, Number of Conceptions, Number of Births and Exposure to Sun. It is the aim of this paper to formulate a smaller number of independent random variables such that minimum information is lost. A total of two hundred (200) women in their reproductive age interval were selected for the study. Questionnaires were used to get the relevant information from them. Eighty eight percent (88%) of the total variation in the ovulation interval was accounted for by six (6) principal components compared to the original eight random variables. It was observed that the most important factors are Stress, Height, Menstrual Duration, Weight and Age, and that only Exposure to Sun did not show any significance in affecting Ovulation Interval.

Keywords: Principal Component; Ovulation; Conception; Factors and multivariate

Introduction

The time between two menstrual flows in which egg is released from the ovary into the fallopian tube in readiness for fertilization is called ovulation [1]. Being able to determine the period of ovulation is one of the major challenges to conception and unwanted/unexpected conception. The general believe is that ovulation occurs halfway into the normal menstrual cycle. There are signs and symptoms of ovulation which can help to notice the occurrence of ovulation [2]. The most important sign according to [2] is virginal mucous. Other signs like dizziness, increased body temperature and feverishness exist. Predicting ovulation; that is, knowing when it will occur and when it has occurred is very important for conception [3]. In human population the average duration of ovulation occurrence is 28 days. However, most women experience deviations from this universal parameter. Ultimately, many women experience variation in the interval of their ovulation.

Like every other biological phenomenon, ovulation is affected by some factors which are also interdependent. To enhance a good prediction of ovulation, this work was designed to perform a principal component analysis of some suspected factors affecting ovulation interval in selected women in order to determine a fewer and independent set of random variables.

Methodology

A total of 200 women in their reproductive age were selected for the study. The women were mostly teachers in primary and secondary schools, health workers in public and private health facilities and literate women in market places. The data collection process took place in a period of six weeks. Questionnaires were administered to the women and they supplied the necessary information under proper supervision. The major factors considered for analysis are Age, Height, Weight, Work Time (Stress), Menstrual Duration, Number of Conceptions, Number of Births and Exposure to Sun.

The method of data analysis is Principal Component Analysis. It is a multivariate Technique used in data reduction and data analysis [4,5]. The aim of principal component analysis is to ascertain if the joint variation in p variables $\left(x_1, x_2, x_3, \ldots . x_p\right)$ can be represented approximately in terms of the joint variation of a fewer number, say

$k < p$, of hypothetical variables without much loss of information. That is $X_{p\times1}$ are replaced by a linear transformation,

$$Y_{p\times1} = A'_{p\times p}X_{p\times1} \quad (1)$$

Where $Y_{p\times1}, k < p$ are significant and approximately chosen from $Y_{p\times1}$. The $Y_{p\times1}$ are interpretable in terms of the original problem and they are independent [5,6].

$$Y_i = \sum_{i,j=1}^{p} a_{ij}x_j \quad i = 1,2,..p \quad (2)$$

The are Principal Components (PC) if they are uncorrelated and their variances are as large as possible.

$$\left.\begin{array}{r} \operatorname{var}(Y) = a_i'\sum a_i \\ \operatorname{cov}(Y_i, Y_k) = a_i'\sum a_k = 0 \\ i,k = 1,2,3...p \;\&\; i \neq k \end{array}\right\} \quad (3)$$

If the mean vector and covariance matrix of are given as and respectively and, has eigen value – eigen vector pairs; $(\lambda_1, e_1),(\lambda_2, e_2),(\lambda_3, e_3),....(\lambda_p, e_p)$ with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq .... \geq \lambda_p \geq 0$, the i^{th} Principal Component Analysis is given by

$$i = 1,2,..p \quad (4)$$

with

$$\left.\begin{array}{r} \operatorname{var}(Y_i) = \lambda_i \\ \operatorname{cov}(Y_i, Y_k) = e_i'\sum e_k = 0 \\ i \neq k \end{array}\right\} \quad (5)$$

The total variance is given as

$$\sum_{i=1}^{p} \operatorname{var}(Y_i) = \sum_{i=0}^{p} \lambda_i \quad (6)$$

Thus, the proportional variance explained by the k^{th} principal component is

$$\zeta_k = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i}; k = 1,2,...p$$

If 80% to 90% of the total population variance for large can be attributed to the first one, two or three components, these components can be used to replace the original variables without much loss of information.

When the variables, $(x_1, x_2, x_3,....x_p)$ are in different units, two or more variables are measured on vastly different ranges or standardized random variables are used instead of the original random variables, the correlation matrix, , is used to calculate Principal Components (PC) instead of Σ [4].

Data Analysis

The eigen values for the eight random variables considered are 1.8989, 1.2740, 1.1409, 0.9970, 0.8840, 0.8225, 0.6101 and 0.3707 with the following respective proportions 0.237, 0.158, 0.143, 0.125, 0.110, 0.103, 0.076nand 0.047 and cumulative proportions 0.237,0.397, 0.539, 0.664, 0.774, 0.877, 0.953 and 1.000. The independent Principal Components are thus represented as,

$$Y_1 = -0.441x_1 - 0.139x_2 - 0.023x_3 + 0.032x_4 - 0.125x_5 - 0.604x_6 - 0.593x_7 - 0.050x_8$$

$$Y_2 = 0.156x_1 + 0.518x_2 + 0.270x_3 - 0.331x_4 + 0.323x_5 - 0.101x_6 - 0.274x_7 - 0.579x_8$$

$$Y_3 = 0.200x_1 - 0.162x_2 + 0.368x_3 - 0.577x_4 - 0.669x_5 - 0.069x_6 + 0.078x_7 + 0.086x_8$$

$$Y_4 = -0.451x_1 + 0.430x_2 + 0.260x_3 - 0.274x_4 + 0.221x_5 - 0.034x_6 + 0.056x_7 + 0.645x_8$$

$$Y_5 = -0.035x_1 + 0.207x_2 - 0.801x_3 - 0.518x_4 - 0.084x_5 - 0.160x_6 + 0.109x_7 - 0.0030x_8$$

$$Y_6 = 0.329x_1 - 0.650x_2 + 0.126x_3 - 0.425x_4 + 0.447x_5 + 0.206x_6 + 0.037x_7 - 0.173x_8$$

Where x_1 =Age, x_2 = Height, x_3 =Weight, x_4 =Work Time, x_5 = Menstrual Duration, x_6 =Number of Conceptions, x_7 =Number of Births and x_8 =Exposure to Sun.

The six Principal Components accounts for 87.7%, approximately 88% of the total variation in the ovulation interval of the selected women. The Principal Components attach more significance to Stress, Height, Menstrual Duration, Weight and Age, and less significance to Exposure to Sun in affecting Ovulation Interval.

Conclusion

The original data which was collected in eight dimensions known to be correlated has been successfully reduced into six principal components which are shown to be uncorrelated. The six principal components account for 88% of the total variation in ovulation interval of the studied population. Hence, further analysis of the data demanding the condition of independence of data set can be done using the six principal components.

Acknowledgement

None.

Conflict of Interest

No conflict of interest.

References

1. Gurevich R (2018) verywellfamily. S Anita (Ed) Retrieved July 03, 2018, from www.verywellfamily.com.
2. Billings J (2002) The Quest Leading to the Discovery of the Billings Ovulation Method. Bulletin of Ovulation Method Research and Reference Centre of Australia 29(1): 18-28.
3. Matthew CM (2011) Application of Multivariate and Time Series Techniques in Preconception Sex Determination. Faculty of Physical Sciences, Statistics. Awka: Nnamdi Azikiwe University, Nigeria.
4. Ogum GE (2002) Inrroduction to Methods of Multivariate Analysis. Aba, Nigeria: Afri Towers Limited, Nigeria.
5. Onyeagu SI (2003) A First Course in Multivariate Statistical Analysis. Awka, Nigeria: Megaconcept, Nigeria
6. Johnson RA (1992) Apllied Multivariate Statistical Analysis. New Jersey: Englewood Cliffs.