World Journal of Probability and Statistics Research
Vol. 2, No. 1, January 2016, pp. 1- 12
Available online at http://wjmcr.com/

Research article

Preconception Sex Selection Using Proper Ovulation Timing

Osuji, G.A[1], Obubu, M.[1*], Nwosu, C.A[2]

[1] Department of Statistics, Nnamdi Azikiwe University, P.M.B 5025, Awka, Nigeria.

[2] Department of Statistics, Chukwuemeka Odumegwu University, Uli, Nigeria.

[1*]E-mail: maxonstatistics@yahoo.com

* Corresponding author

Abstract

Preconception sex selection refers to any procedure attempting to influence the sex of offspring before pregnancy. Over the past decade, investigation technique called Preimplantation Genetic Diagnosis (PGD) has been seen as the most effective method of sex selection, however, this paper introduces the use of Multivariate Techniques such as Principal Component Analysis and Discriminant Analysis in preconception sex determination using proper ovulation timing. The result showed that all the variables but average time of exposure to sun rays are significant, Ovulation intervals for 300 women flunctuate between short, normal and long interval, also, Age and Number of Birth having the highest correlation coefficient (r = .839) implying that they are the most common factors of change in Ovulation.

Keywords: preconception sex determination, principal component analysis, discriminant analysis, ovulation interval, number of births

1. Introduction

For centuries there has been an interest in trying to determine factors that affect the sex of offspring. The sex is determined by the genetic component of the sperm, but it is unclear if there are any factors that decide whether a sperm bearing a Y or an X chromosome will fertilize the egg. Overall, the ratio of male to female offspring in the general population is slightly greater than unity (1.06). Several studies have tried to assess whether the

interval between intercourse and ovulation has an effect on the sex ratio. It was believed that sperm carrying a Y chromosome had higher motility and therefore had a better chance to fertilize the egg if intercourse occurred around the time of ovulation. On the other hand, sperm bearing an X chromosome were believed to be more resistant and therefore were thought to be more likely to lead to fertilization if intercourse occurred at a time more remote from the time of ovulation. To the contrary, a slightly higher incidence of female offspring was reported for cycles during which intercourse (probable time of fertilization) occurred close to the time of the ovulation. Similarly, the incidence of female offspring was also higher in cycles during which ovulation induction and insemination were used. However, the results of these studies are limited by the fact that various indirect methods (cervical mucous changes, basal body temperature, urinary luteinizing hormone) were used to try to predict the time of ovulation. Maternal age, paternal age, and parity have also been found to have a slight effect on the sex ratio. The ratio was lower (i.e. more females) with increased paternal age and parity. Other studies have evaluated the effect of douching on the sex ratio. These studies were based on theories that sperm bearing the X or Y chromosome favored different vaginal pH. The findings of these reports were inconsistent, however. Although some of the above-discussed sex ratio differences were statistically significant, their impact on the overall sex ratio was marginal. In addition, there is no biologically plausible explanation for such differences. None of these "natural methods" for producing offspring of a particular sex are considered reliable. However, laboratory methods were developed to separate sperm on the basis of the slight size difference owing to the higher DNA content -- about 3% -- in sperm carrying the X chromosome. The older techniques (e.g., modified swim-up, electrophoresis, immunologic separation, albumin gradient, Percoll gradient) were associated with an approximately 70% to 80% accuracy. This accuracy, although higher than that of the "natural methods," is still not sufficient. The latest technique, fluorescence-activated cell sorter, is reported to be able to select out sperm bearing the X chromosome with close to 90% accuracy. The limitations of this technique are that the sperm count is greatly reduced during the process (necessitating assisted reproduction technology [ART]), and long-term follow-up results are not yet available on large numbers of offspring conceived following this method. Pre-implantation genetic diagnosis offers the only sure way of determining the sex of the offspring. This method is expensive and also requires the use of ART. Even if there were a reliable, cheap method that could increase the chance of having a male or female offspring, important ethical questions remain to be answered about the role of sex selection. Currently, the most common reasons to request sex selection is family balancing and in case of certain genetic diseases (X-linked diseases). Therefore in this work, we intend to; Determine the best set of variables that describes the variation in ovulation intervals for the selected women, determine a good discriminant function for the selected number of women, and test the hypothesis that all the sampled women ovulate normally.

2. Methodology

2.1 Principal Component Analysis

Principal component analysis is a multivariate technique for transforming a set of related (correlated) variables into a set of unrelated (uncorrelated) variables that account for decreasing proportions of the variation of the original observations (Rencher, 2002). The rationale behind the method is an attempt to reduce the complexity of the data by decreasing the number of variables that need to be considered. If the first few of the derived variables (the principal components) among them account for a large proportion of the total variance of the

observed variables, they can be used to provide a convenient summary of the data and to simplify subsequent analysis. Algebraically, principal component are particular linear combinations of the p random variables X_1 , X_2, . . . , X_p. Geometrically, these linear combination represents the selection of new coordinate system obtained by rotating the original system with X_1 , X_2, . . . , X_p as the coordinate axes. The new axes represents the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure. Principal components depend solely on the covariance matrix $\sum$ (or the correlation matrix ρ) of X_1 , X_2, . . . , X_p . Their development does not require a multivariate normal assumption.

let the random vector $X^{l} = [X_1 , X_2, . . . , X_p]$ have the covariance matrix $\sum$ with eigenvalues

$\lambda_1 \geq \lambda_2 \geq . . . \geq \lambda_p \geq 0.$

Consider the linear combination

$$\begin{aligned} Y_1 = \mathbf{a}_1^{l}\mathbf{X} &= a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ Y_2 = \mathbf{a}_2^{l}\mathbf{X} &= a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\ &\vdots \\ Y_p = \mathbf{a}_p^{l}\mathbf{X} &= a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p \end{aligned}$$

Then,

$Var(Y_i) = \mathbf{a}_i^{l} \sum \mathbf{a}_i \quad i = 1 , 2 , . . . , p$

$Cov(Y_i , Y_k) = = \mathbf{a}_i^{l} \sum \mathbf{a}_k \quad i, k = 1 , 2 , . . . , p$

Note;

First principle component = linear combination $\mathbf{a}_1^{l}\mathbf{X}$ that maximizes $Var(\mathbf{a}_1^{l}\mathbf{X})$ subject to $\mathbf{a}_1^{l}\mathbf{a}_1 = 1$

Second principle component = linear combination $\mathbf{a}_2^{l}\mathbf{X}$ that maximizes $Var(\mathbf{a}_2^{l}\mathbf{X})$

subject to $\mathbf{a}_2^{l}\mathbf{a}_2 = 1$ and $Cov(\mathbf{a}_1^{l}\mathbf{X} , \mathbf{a}_2^{l}\mathbf{X}) = 0$

At the *i*th step,

*i*th principle component = linear combination $\mathbf{a}_i^{l}\mathbf{X}$ that maximizes $Var(\mathbf{a}_i^{l}\mathbf{X})$

subject to $\mathbf{a}_i^{l}\mathbf{a}_i = 1$ and $Cov(\mathbf{a}_i^{l}\mathbf{X} , \mathbf{a}_k^{l}\mathbf{X}) = 0$ for $k < i$.

Consider the covariance matrix of a Bivariate data

$$\sum = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{11} & \delta_{12} \end{pmatrix}$$

and the derived correlation matrix

$$\rho = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}$$

The proportion of the total variance explained by the first principal component is

$$\Psi_{X_1} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

which is larger than that explained by the second principal component when the X's are not standardized.

In determining the number of Principal Components to retain, the amount of total variance explained, the relative sizes of the eigenvalues, { Joliffe (2002) } suggests using a cutoff on the eigenvalue of 0.7 when correlation matrices are analyzed., and a visual inspection of the scree plots are of prior importance.

2.2 Discriminant Analysis

Discriminant analysis is a classification problem, where two or more groups or clusters or populations are known a priori and one or more new observations are classified into one of the known populations based on the measured characteristics. it is divided into two components;

- Discrimination
- Classification

2.2.1 Fishers Linear Discriminant Function (Common Sense Approach)

Let us classify X into π_1 if X is closer to U_1 than to U_2, the distance between X and U_1 is

$$d_s^2(X, U_1) = (X - U_1)' \sum^{-1} (X - U_1),$$

the distance between X and U_2 is

$$d_s^2(X, U_2) = (X - U_2)' \sum^{-1} (X - U_2).$$

The rule is : Classify into π_1 if ;

$$d_s^2(X, U_1) < d_s^2(X, U_2)$$

and to π_2 otherwise.

Therefore,

$$(X - U_1)' \sum^{-1} (X - U_1) < (X - U_2)' \sum^{-1} (X - U_2).$$

$$X' \sum^{-1} X - X' \sum^{-1} U_1 - X \sum^{-1} U_1 + U_1' \sum^{-1} U_1 < X' \sum^{-1} X - X' \sum^{-1} U_2 - X \sum^{-1} U_2' + U_2' \sum^{-1} U_2$$

$$-2 X' \sum^{-1} U_1 + U_1' \sum^{-1} U_1 < -2 X' \sum^{-1} U_2 + U_2' \sum^{-1} U_2$$

$$-2 X' \sum^{-1} U_1 + 2 X' \sum^{-1} U_2 < U_2' \sum^{-1} U_2 - U_1' \sum^{-1} U_1 \quad \text{.....(1)}$$

Multiplying equation (1) by -1, we have that

$$2 X' \sum^{-1} U_1 - 2 X' \sum^{-1} U_2 > U_1' \sum^{-1} U_1 - U_2' \sum^{-1} U_2$$

$$2 X' \sum^{-1} (U_1 - U_2) > U_1' \sum^{-1} U_1 - U_2' \sum^{-1} U_2$$

$$X' \sum^{-1} (U_1 - U_2) > {}^1/_2 (U_1' \sum^{-1} U_1 - U_2' \sum^{-1} U_2)$$

$$X' \sum^{-1} (U_1 - U_2) > {}^1/_2 (U_1 + U_2)' \sum^{-1} (U_1 - U_2)$$

Where $X' \sum^{-1} (U_1 - U_2)$ is called Fisher's Linear Discriminant Function. Since U_1, U_2 and $\sum$ are unknown, we estimate them from the samples.

Suppose X_{1j}, j = 1, 2 , . . . , n_1 is a random sample from π_1, calculate $\bar{X}_1$ and S_1, Suppose X_{2j}, j = 1, 2 , . . . , n_2 is a random sample from π_1, calculate $\bar{X}_2$ and S_2.

As in the univariate statistics, we collect estimate $\sum$ by;

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

World Journal of Probability and Statistics Research
Vol. 2, No. 1, January 2016, pp. 1- 12
Available online at http://wjmcr.com/

Replacing U_1 by $\bar{X}_1$, U_2 by $\bar{X}_2$, and $\sum$ by S_p we have;

$X^{l} S_p^{-l} (\bar{X}_1 - \bar{X}_2)$ which is called Fishers Sample Linear Discriminant Function.

2.2.2 Evaluating Discriminant Function

One way of evaluating discriminant function is to evaluate the Apparent Error Rate (APER). This is the fraction of observations that are misclassified using the sample classification function.

Researcher's Decision	True Membership		
	π_1	π_2	π_3
π_1	a_{11}	b_{12}	b_{13}
π_2	b_{21}	a_{22}	b_{23}
π_3	b_{31}	b_{32}	a_{33}
	n_1	n_2	n_3

where,

n_1, n_2 and n_3 are sample sizes

π_1, π_2, π_3 are groups

a_{ij} represents numbers of objects i,j correctly classified

b_{ij} represents numbers of objects i,j incorrectly classified

Let P_1, P_2, P_3 be the probability of misclassification into group π_1, π_2, π_3 respectively and $\dot{P}$ be the total probability of misclassification.

$$P_1 = P(\pi_1, \pi_2 / \pi_3) = \frac{b_{13}}{n_3} + \frac{b_{23}}{n_3}$$

$$P_2 = P(\pi_1, \pi_3 / \pi_2) = \frac{b_{12}}{n_2} + \frac{b_{32}}{n_2}$$

$$P_3 = P(\pi_2, \pi_3 / \pi_1) = \frac{b_{21}}{n_1} + \frac{b_{31}}{n_1}$$

$$\dot{P} = \frac{\text{Total number of misclassified objects}}{\text{Total sample size}}$$

2.2.3 Mahalanobi's Squared Distance

$$D_{i,j}^2 = (\bar{Y}_i - \bar{Y}_j)^{l} S^{-l} (\bar{Y}_i - \bar{Y}_j)$$

probability of misclassification assuming the data is normally distributed is;

$$1 - \phi \left(\frac{(D_{ij}^2)^{1/2}}{2} \right) = \phi \left(\frac{-(D_{ij}^2)^{1/2}}{2} \right)$$

2.2.3.1 Using the Mahalanobi's distance to conduct the test between group differences

$H_o : \pi_i = \pi_j$

$H_1 : \pi_i \neq \pi_j$

Decision Rule: Reject H_o if

$$F = \frac{n_i n_j (n_i + n_j - v - 1) D_{ij}^2}{(n_i + n_j)(n_i + n_j - 2) v} > F_{v,\, n_i + n_j - v - 1,\, \alpha}$$

3. Data Analysis and Result

Table 1: *KMO And Bartlett's Test*

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.885
Bartlett's Test Square Sphericity	Approx. Chi-	43.805
	df	21
	Sig.	.001

The null hypothesis that the correlation matrix is an identity matrix was rejected at 5% level of significance (Bartlett's test of Sphericity; χ^2 = 43.805, p-value = .001), this implies that the correlation in the dataset are appropriate for factor analysis. Also, "Kaiser-Meyer-Olkin statistic = .885" revealed that adequate sampling is being used for this analysis.

Table 2: *Correlation Matrix*

Correlation	Age(yrs)	Weight(kg)	Exposure to Sun(hrs/day)	Menstrual Duration(days)	Height(m)	Work Time(hrs/day)	Number of Birth
Age(yrs)	1.000	.040	-.521	.046	-.192	.179	.839
Weight(kg)	.040	1.000	-.258	.100	.047	-.161	.280
Exposure to Sun(hrs/day)	-.521	-.258	1.000	-.379	-.204	.294	-.645
Menstrual Duration(days)	.046	.100	-.379	1.000	.149	-.120	.094
Height(m)	-.192	.047	-.204	.149	1.000	-.385	-.169
Work Time(hrs/day)	.179	-.161	.294	-.120	-.385	1.000	.033
Number of Birth	.839	.280	-.645	.094	-.169	.033	1.000
Sig.(1-tailed)							
Age(yrs)		.434	.009	.424	.209	.225	.000
Weight(kg)	.434		.136	.338	.423	.249	.116
Exposure to Sun(hrs/day)	.009	.136		.050	.195	.104	.001
Menstrual Duration(days)	.424	.338	.050		.265	.307	.346
Height(m)	.209	.423	.195	.265		.047	.238
Work Time(hrs/day)	.225	.249	.104	.307	.047		.446
Number of Birth	.000	.116	.001	.346	.238	.446	

Table 2 reveals that all the factors affecting Ovulation interval considered in this study are relatively important. Significant correlations exists between Age and Exposure to sun (r = -.521, p-value = .009), Age and Number of Birth(r = .839, p-value = .000), Exposure to sun and Number of Birth(r = -.645, p-value = .001), and, Height and Work Time(r = -.385, p-value = .047), The table shows Age and Number of Birth having the highest correlation coefficient (r = .839) implying that they are the most common factors of change in Ovulation.

Table 3: *Principal Components: Eigenvalues, % of Variance Explained, Cumulative % and Eigenvectors*

λ_i	2.494	1.731	.941	.860	.596	.261	.115
% of Variance	35.632	24.732	13.447	12.292	8.517	3.731	1.650
Cumulative %	35.632	60.364	73.811	86.103	94.620	98.350	100.000
Age(yrs)	.849	.412	-.127	-.218	.066	.216	-.208
Weight(kg)	.373	-.264	.761	.421	.174	.036	-.056
Exposure to Sun(hrs/day)	-.820	.311	.124	.069	.035	.398	.052
Menstrual Duration(days)	.330	-.419	-.525	.647	-.065	.134	.010
Height(m)	-.021	-.757	-.138	-.324	.542	.085	.025
Work Time(hrs/day)	-.110	.763	-.171	.306	.512	-.141	.017
Number of Birth	.917	.253	.083	-.115	.002	.098	.256

The independent principal components are thus represented as;

$Y_1 = .849x_1 + .373x_2 - .820x_3 + .330x_4 - .021x_5 - .110x_6 + .917x_7$

$Y_2 = .412x_1 - .264x_2 + .311x_3 - .419x_4 - .757x_5 + .763x_6 + .253x_7$

$Y_3 = -.127x_1 + .761x_2 + .124x_3 - .525x_4 - .138x_5 - .171x_6 + .083x_7$

$Y_4 = -.218x_1 + .421x_2 + .069x_3 + .647x_4 - .324x_5 + .306x_6 - .115x_7$

$Y_5 = .066x_1 + .174x_2 + .035x_3 - .065x_4 + .542x_5 + .512x_6 + .002x_7$

$Y_6 = .216x_1 + .036x_2 + .398x_3 + .134x_4 + .085x_5 - .141x_6 + .098x_7$

$Y_7 = -.208x_1 - .056x_2 + .052x_3 + .010x_4 + .025x_5 + .017x_6 + .256x_7$

where x_1 = Age, x_2 =Weight, x_3 = Exposure to Sun, x_4 = Menstrual Duration, x_5 = Height, x_6 = Work Time, x_7 = Number of Birth.

The first principal component Y_1 attaches more to Number of Birth, Age, Weight, and Menstrual Duration. The second principal component Y_2 attaches more to Work Time, Age and Exposure to sun. The Third principal component Y_3 attaches more to Weight and Number of Birth. The fourth principal component Y_4 attaches more to Menstrual Duration, Weight, Work Time and Exposure to sun. The fifth principal component Y_5 attaches more to Height, Work Time, Weight, Age, Exposure to sun and Number of Birth. The sixth principal component Y_6 attaches more to Exposure to sun, Age, Menstrual Duration, Number of Birth, Height and Weight. The seventh principal component Y_7 attaches more to Number of Birth, Exposure to sun, Height, Work Time, and Menstrual duration in that order.

3.1 Discriminant Analysis.

Table 4: *Descriptive Statistics for Women Experiencing Short, Normal and Long Cycle*

Ovulation Interval		**Mean**	**Std. Deviation**
Short*(< 20days)*	Age(yrs)	34.2000	7.64061
	Height(m)	1.4310	.58002
	Weight(kg)	59.4005	18.26744
	Work Time (hrs/day)	10.6595	4.58516
	Menstrual Duration(days)	4.3825	1.38147
	No of Conceptions	3.2500	1.88833
	No of Births	2.5500	1.84890
	Exposure to Sun	4.2615	2.69643
Normal (*20-34days)*	Age(yrs)	34.3000	7.63027
	Height(m)	1.6115	.06769
	Weight(kg)	60.8500	13.44492
	Work Time (hrs/day)	9.5250	3.37746
	Menstrual Duration(days)	4.3250	.73045
	No of Conceptions	2.6500	1.84320
	No of Births	2.3000	1.89459
	Exposure to Sun	4.4250	2.51979
Long *(35^{+})*	Age(yrs)	4.4250	2.51979
	Height(m)	36.6500	8.02152
	Weight(kg)	1.6690	.42302
	Work Time (hrs/day)	70.7700	17.95735
	Menstrual Duration(days)	7.8880	4.38660
	No of Conceptions	3.8010	1.03187
	No of Births	3.3000	2.36421
	Exposure to Sun	3.4690	1.79322
Total	Age(yrs)	2.4000	1.81804
	Height(m)	3.4690	1.79322
	Weight(kg)	35.0500	7.71840
	Work Time (hrs/day)	1.5705	.42178
	Menstrual Duration(days)	63.6735	17.18963
	No of Conceptions	9.3575	4.23753
	No of Births	4.1695	1.09494
	Exposure to Sun	4.0518	2.36621

Variance-Covariance Matrix for Short Ovulation Interval.

58.379	1.233	2.606	4.467	-2.265	3.684	2.779	-.959
1.233	.336	-3.524	.640	-.163	-.083	-.083	.039
2.606	-3.524	333.699	-39.349	2.941	3.996	3.465	-12.269
4.467	.640	-39.349	21.024	.793	1.499	2.452	.289
-2.265	-.163	2.941	.793	1.908	1.163	.689	-.909
3.684	-.083	3.996	1.499	1.163	3.566	3.118	-1.422
2.779	-.083	3.465	2.452	.689	3.118	3.418	-1.087
-.959	.039	-12.269	.289	-.909	-1.422	-1.087	7.271

Variance-Covariance Matrix for Normal Ovulation Interval.

58.221	-.159	2.311	-.745	1.082	9.005	12.642	-2.265
-.159	.005	.147	-.023	.001	-.012	-.005	3.684
2.311	.147	180.766	-11.996	2.657	1.366	5.574	2.779
-.745	-.023	-11.996	11.407	-.864	-.596	.229	-.959
1.082	.001	2.657	-.864	.534	.409	.187	58.221
9.005	-.012	1.366	-.596	.409	3.397	2.268	-.554
12.642	-.005	5.574	.229	.187	2.268	3.589	.129
-.345	-.031	.172	-2.709	-.198	-.554	.129	6.349

Variance-Covariance Matrix for Long Ovulation Interval.

64.345	-.470	43.334	4.162	.352	10.584	10.779	.406
-.470	.179	1.314	.568	.047	.128	.005	.093
43.334	1.314	322.466	3.252	2.913	5.792	11.184	7.770
4.162	.568	3.252	19.242	2.862	2.416	.692	1.517
.352	.047	2.913	2.862	1.065	.455	-.017	.481
10.584	.128	5.792	2.416	.455	5.589	3.558	.807
10.779	.005	11.184	.692	-.017	3.558	3.305	.447
.406	.093	7.770	1.517	.481	.807	.447	3.216

Pooled Estimate of the Common Variance-Covariance Matrix.

59.574	.278	21.336	1.325	-.569	7.675	8.419	-.760
.278	.178	-.265	.273	-.057	.004	-.035	.008
21.336	-.265	295.483	-21.069	1.397	4.286	6.391	-3.456
1.325	.273	-21.069	17.957	1.185	1.010	1.147	.113
-.569	-.057	1.397	1.185	1.199	.615	.282	-.094
7.675	.004	4.286	1.010	.615	4.131	2.904	-.462
8.419	-.035	6.391	1.147	.282	2.904	3.332	-.167
-.760	.008	-3.456	.113	-.094	-.462	-.167	5.599

Inverse of the Common Variance-Covariance Matrix.

0.0278643	-0.04651	-0.0006558	0.000369	0.03244	-0.012670	-0.06141	0.001104
-0.0465128	6.12524	-0.0091979	-0.142865	0.44232	-0.234252	0.41454	-0.017398
-0.0006558	-0.00920	0.0041470	0.006301	-0.01199	0.006849	-0.01340	0.002321
0.0003691	-0.14286	0.0063012	0.072998	-0.08702	0.028032	-0.05662	0.001834
0.0324402	0.44232	-0.0119872	-0.087021	1.08244	-0.275767	0.12422	-0.002748
-0.0126696	-0.23425	0.0068493	0.028032	-0.27577	0.709319	-0.58616	0.038693
-0.0614108	0.41454	-0.0134015	-0.056622	0.12422	-0.586163	1.00357	-0.032406
0.0011040	-0.01740	0.0023209	0.001834	-0.00275	0.038693	-0.03241	0.182354

The two linearly independent discriminant functions are Y_{12} and Y_{13}

$Y_{12} = -0.0091x_1 - 0.0192x_2 -0.0099x_3 - 0.0568x_4 + 0.0033x_5 + 0.0679x_6 + 0.0489x_7 + 0.3569x_8$

$Y_{12} = -0.0142x_1 - 0.0197x_2 -0.0109x_3 - 0.0811x_4 + 0.0423x_5 + 0.0620x_6 + 0.0412x_7 + 0.3547x_8$

3.2 Classification Rule

Assign an individual with measurement X into π_1 if $Y_{12} > 0.6219$ and $Y_{13} > 0.5421$. Assign an individual to π_2 if $Y_{21} > -0.6219$ and $Y_{23} > -0.1109$. Assign an individual to π_3 if $Y_{31} > -0.5421$ and $Y_{32} > 0.1109$.

Confusion Matrix

Researcher's Decision	True Membership		
	π_1	π_2	π_3
π_1	0	0	6
π_2	38	42	70
π_3	11	29	6
	49	71	82

Therefore,

$P(\pi_1, \pi_2 / \pi_3)$ = 76/82 = 0.927

Interpretation: Out of every 1000 women who actually experience long ovulation intervals, 927 will be misclassified as experiencing short or normal ovulation interval.

$P(\pi_1, \pi_3 / \pi_2)$ = 29/71 = 0.409

Interpretation: Out of every 1000 women who actually experience normal ovulation intervals, 409 will be misclassified as experiencing short or long ovulation interval.

$P(\pi_2, \pi_3 / \pi_1)$ = 49/49 = 1

Interpretation: Out of every 1000 women who actually experience short ovulation intervals, all of them will be misclassified as experiencing normal or long ovulation interval.

P =Total probability of Misclassification = (6 + 38 + 70 + 11 + 29) / 202 = 154/202 = 0.762

Interpretation: Out of every 1000 women who actually experience short, normal or long ovulation intervals, 762 will be misclassified.

Test of Significance of the difference between the three groups(Short, Normal, Long Ovulation Interval) using mahalanobis distance.

$D_{12}^2 = 0.0889$

$D_{13}^2 = 0.1448$

$D_{23}^2 = 0.0898$

3.3 Testing between group one and group two.

Ho : $\pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$

Decision Rule: Reject H_o if

$$F = \frac{n_i n_j (n_i + n_j - v - 1) D_{ij}^2}{(n_i + n_j)(n_i + n_j - 2) v} > F_{v,\, n_i + n_j - v - 1,\, \alpha}$$

$$F_{12} = \frac{49 * 71 (49+71-10)*0.0889}{(49 + 71)(49+71-2)*9} = 0.267$$

$F_{tab} = F_{9,\,(49+71-9-1),\ 0.05} = F_{9,110,\,0.05} \approx 2.75$

Since $F_{12} < F_{tab}$, we accept H_o and conclude that there is no significance difference between the two groups π_1 and π_2 ($\pi_1 = \pi_2$).

3.4 Testing between group one and group three

Ho : $\pi_1 = \pi_3$

$H_1 : \pi_1 \neq \pi_3$

Decision Rule: Reject H_o if

$$F = \frac{n_i n_j (n_i + n_j - v - 1) D_{ij}^2}{(n_i + n_j)(n_i + n_j - 2) v} > F_{v,\, n_i + n_j - v - 1,\, \alpha}$$

$$F_{13} = \frac{49 * 82 (49+82-10)*0.1448}{(49 + 82)(49+82-2)*9} = 0.463$$

$F_{tab} = F_{9,\,(49+82-9-1),\ 0.05} = F_{9,121,\,0.05} \approx 2.75,$

Since $F_{13} < F_{tab}$, we accept H_o and conclude that there is no significance difference between the two groups π_1 and π_3 ($\pi_1 = \pi_3$)

3.5 Testing between group two and group three

Ho : $\pi_2 = \pi_3$

$H_1 : \pi_2 \neq \pi_3$

Decision Rule: Reject H_o if

$$F = \frac{n_i n_j (n_i + n_j - v - 1) D_{ij}^2}{(n_i + n_j)(n_i + n_j - 2) v} > F_{v,\, n_i + n_j - v - 1,\, \alpha}$$

$$F_{23} = \frac{71 * 82 (71+82-10)*0.0898}{(71 + 82)(71+82-2)*9} = 0.360$$

$F_{tab} = F_{9,\,(71+82-9-1),\ 0.05} = F_{9,143,\,0.05} \approx 2.75,$

Since $F_{23} < F_{tab}$, we accept H_o and conclude that there is no significance difference between the two groups π_2 and π_3 ($\pi_2 = \pi_3$).

4. Conclusion and Recommendation

Age and Number of births are the most common factor of change in Ovulation Interval, the large number of principal components accounting for relatively small proportion of the total variation in ovulation interval suggests that there could be other significant factors that were not considered in this research, the discriminant

function is associated with high probability of misclassification implying that a woman's ovulation interval can fluctuate through the three groups(short, normal, long) equally. We hereby recommend that women keep accurate data about length of their ovulation intervals and the factors affecting its variation. owning to the fact that almost all methods of Preconception Sex Determination are centered on ovulation timing.

5. References

[1] Arnold, Fred, Kishor, Sunita, and Roy, T.K. "Sex-Selective Abortions in India". Population and Development Review 2002; 28(4): 759-785.

[2] Bartlett, M.S "An Inverse Matrix Adjustment Arising in Discriminant Analysis" Annals of Mathematical Statistics, 1951; p.107-111.

[3] Fisher, R.A. "The use of Multivariate Measurements in Taxonomic Problems" Annals of Eugenics, 1936; p. 179-188.

[4] Gray, RH. "Natural family planning and sex selection: fact or fiction?" Am J Obstet Gynecol 1991; 165:1982-4.

[5] Jha, P., Kumar, R., Visa, P., Dhingra, N., Thiruchelvam, D., Moinedin, R. "Low female to male sex ratio of children born in India: national survey of 1.1 million households." Lancet 2006; 367:211-8.

[6] Sex selection and pre-implantation genetic diagnosis. Ethics Committee of the American Society for Reproductive medicine. Fertil Steril 2004;82 (suppl): S254-8.

[7] Stanfield, W.D. "Theory and Problems of Genetics", third edition Schaum Outline Series. The McGraw Hills Company Inc. New York, 1991.

[8] United Nations Gender equality, equity and empowerment of women. In: Population and Development; programme of action adopted at the international conference on Population and Development, Cairo, 5-13 September 1994. New York (NY): UN; 1995. p. 17-21.