AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Diagnosing some important factors that affect spontaneous abortions using the linear discriminant model

Osama abdulazeez kadhim Al-Quraishi

Ministry of Education Iraq

ousamastat@gmail.com

Orcid:000-0002-1701-2679

http://dx.doi.org/10.29124/kjeas.1654.4

Abstract:

Discriminant analysis is one of the statistical tools for multivariate descriptive analysis that has been widely used in medical studies. The research problem was the increase in spontaneous abortions by 10% to 50% of pregnancies with clear clinical miscarriages. This percentage is considered high, and its accompanying negative repercussions affect pregnancy. It was represented by introversion, an intense feeling of sadness and loss, irritability, self-rejection and guilt, preoccupation with thinking about getting pregnant again, anxiety and nightmares, eating disorders, thoughts or attempts at suicide...etc , The research aimed to use discriminant analysis to determine the most important factors and their relative importance influencing spontaneous abortion through the extent to which each factor contributes to distinguishing between the two groups. The research also aims to predict the possibility of spontaneous abortion for pregnant women based on the influencing factors, leading to early diagnosis and avoiding spontaneous abortion, A sample size of 100 pregnant was used, including 50 spontaneous abortions, 50 live births, and 14 independent variables. It appeared that the variables that had a significant effect on spontaneous abortion were (age x_1, first pregnancy x_8, diabetes in the pregnant woman x_{13}, blood type matching between spouses x_4, Educational level of the pregnant woman x_3, previous miscarriages x_9, high blood pressure in the pregnant woman x_{11}, previous dead births x_{14}). The contribution rate of variables in distinguishing and classifying between the two groups was (6.54%)

.key words : Discriminatory model , Point of distinction, Classification error

the introduction:

Statistical analysis plays an important role in analyzing various life phenomena, as it is one of the scientific research tools that are used when studying certain phenomena (health, social, economic,..., etc.), Discriminant analysis is one of the statistical tools for descriptive multivariate analysis that is commonly used in medical studies , Discriminant analysis works to differentiate o between two or more groups of things or individuals, and to determine the relative importance of the independent variables affecting the discrimination process by creating a function called the differentiation function that works to increase the differences between the average of the groups and classify the new item's belonging to the groups that have been defined, The research included two groups, the first group represented live births, and a live birth or live birth means the exit of the fetus, regardless of its gestational age, from the mother's body, with the appearance of evidence indicating life such as voluntary movements, heartbeats, and umbilical cord pulsation, . The second group represented spontaneous abortion, which is the expulsion of the fetus from the uterus due to accidental trauma or natural causes and usually occurs before the twenty-second week of pregnancy, The research problem was an increase in spontaneous abortions by 10% to 50% of pregnancies with clear clinical miscarriages, and this percentage is considered high, and the accompanying negative repercussions that affect women after miscarriage, represented by introversion, an acute feeling of sadness and loss, irritability, self-rejection, and feelings of guilt. Preoccupation with thinking about getting pregnant again, anxiety and nightmares, eating disorders, thoughts or attempts of suicide...etc., The research aimed to use discriminant analysis to determine the most important factors and their relative importance influencing spontaneous abortion through the extent to which each factor contributes to distinguishing between the two groups. The research also aims to predict the possibility of spontaneous abortion for pregnant women based on the influencing factors, leading to early diagnosis and avoiding spontaneous abortion, From previous studies on discriminant analysis, Al-Marsoumi, Aseel Mubadar Dawoud, and Al-Hashemi[7], (2019), the research aims to build a probabilistic model to distinguish between the variables that cause heart disease and classify them into two groups (affected and unaffected), in addition to classifying the variables and items that will be studied in the future, as well as identifying the factors. The most important factor affecting heart disease. Finally, the research reached the efficiency of the model used in the research in terms of the low percentage of error in misclassification. The variable blood pressure came first, then smoking, followed by weight, and finally age in terms of relative importance in its impact on the disease, Al-Demerdash, Hani Muhammad Ali[4], (2022), the research aimed to compare the discriminant analysis method and the multivariate analysis of variance method, to determine the most important factors associated with the presence of the phenomenon of stagflation in the Egyptian economy for the period (1980-2021), and the statistical results proved that discriminant analysis performs the analysis tasks. Multiple variance increases the ability to predict and classify, Al-Hamzi, Amat Al-Latif Muhammad Ali and Al-Hamzi, Amani[5], (2023), the study aimed to employ the use of discriminant function analysis to classify data on new premature infants into live or dead premature infants, and the study found the most important independent variables in classifying the group of premature infants into one of the following: The two groups, the first important of which is the variable of the pregnant mother's smoking, followed in importance by the variable of the uterine age of the premature baby in weeks, then the variable of taking nutritional supplements during pregnancy, then the rest of the six variables adopted in building the discrimination function , In the same year, Mustafa , Anwar Al-Zein Babakir[12] presented a research aimed at identifying the most important factors influencing acute and acute leukemia using discriminant analysis. The study found the most important factors influencing the classification of acute and chronic cancer (smoking, obesity, blood disorders, kinship).

1- **the theoretical side** :

1-1: Discriminant analysis :[1] [5]

The discriminant analysis method is one of the important statistical methods in multivariate statistical analysis that is used to process descriptive data, which is concerned with the issue of distinguishing between two or more groups that are similar in many characteristics on the basis of several independent variables. Its knowledge is summarized through the creation of the linear discriminant function by the scientist Fisher. The year 1936, which is defined as a linear combination of a group of independent variables, which predicts or classifies the sample items into different groups by classifying the new sample items into one of the groups under study with the lowest possible classification error, by increasing the difference between the average of the groups and whenever there is The greater the distance between the average groups, the more efficient the discrimination and the smaller the classification error, meaning that the discriminant function works to increase the degree of homogeneity between the items of one group and reduce the homogeneity between the two groups..

1-2: Objectives of discriminant analysis :[4][12]

There are several goals unique to discriminant analysis, as follows:

1- Construct discriminant functions to distinguish or differentiate between groups of the dependent variable.
2- Discriminant functions work to maximize the difference between groups of the dependent variable
3- The discriminant function determines the relative importance of the independent variables that participate in explaining the differences or distinction between groups of the dependent variable.
4- The discriminatory function classifies the new observations into one of the groups of the approved variable to reach the lowest percentage of error in the description - the accuracy of the classification is evaluated as a percentage.

1-3: Types of discriminant functions:[7]

1- Linear discriminant function: It is called the Fisher function after the scientist who derived it and is used when the relationship is linear between variables.
2- Nonlinear discriminant function: It is used when the relationship between variables is non-linear, quadratic or more.

1-4: Types of discriminant analysis:[3][6]

1- Direct discriminant analysis: The variables enter the analysis in one package without giving importance to any variable.
2- Hierarchical discriminant analysis: Variables are entered into it according to the researcher’s opinion
3- Stepwise discriminant analysis: Entering variables for analysis based on a statistical criterion that determines the priority of entering variables into the model, and variables are added to the discriminant functions one after the other until we conclude that adding variables does not give better discrimination.

1-5 : Assumptions of discriminant analysis :[2][8]

1- Randomly selecting the sample under study
2- All study variables are distributed normally .
3- There are no outliers or extreme values because their presence takes the data distribution away from the normal distribution
4- There is a linear relationship between the variables
5- The means of the study groups are not equal.
6- The variance and covariance matrix is equal between the study groups.
7- There is no problem of multicollinearity between independent variables.

1-6: Determine the number of discriminant functions:[5]

The number of discriminant functions is determined based on the number of groups of the dependent variable minus one, and since the groups of the dependent variable under study are two groups, the first group is miscarriage and the second group is live births, we have only one discriminant function, and using this function we can test the unit and determine its return. To any group.

1-7: The steps to calculate the linear discriminant model between two sets are as follows:[1][11][9]

Firstly : Calculating the average of the independent variables in each group:

a- The first group :

$$\bar{x}_i^{(1)} = \begin{bmatrix} \bar{x}_1^{(1)} \\ \bar{x}_2^{(1)} \\ \vdots \\ \bar{x}_n^{(1)} \end{bmatrix} \quad \ldots (1)$$

Whereas:

$$\bar{x}_1^{(1)} = \sum_{i=1}^{n_1^{(1)}} \frac{x_{1i}^{(1)}}{n_1^{(1)}} \quad \ldots (2)$$

$$\bar{x}_2^{(1)} = \sum_{i=1}^{n_2^{(1)}} \frac{x_{2i}^{(1)}}{n_2^{(1)}} \quad \ldots (3)$$

$$\bar{x}_n^{(1)} = \sum_{i=1}^{n_n^{(1)}} \frac{x_{ni}^{(1)}}{n_n^{(1)}} \quad \ldots (4)$$

b- The second group :

$$\bar{\boldsymbol{x}}_{\boldsymbol{i}}^{(2)} = \begin{bmatrix} \bar{\boldsymbol{x}}_{\boldsymbol{1}}^{(2)} \\ \bar{\boldsymbol{x}}_{\boldsymbol{2}}^{(2)} \\ \vdots \\ \bar{\boldsymbol{x}}_{\boldsymbol{n}}^{(2)} \end{bmatrix} \qquad \dots (\ \mathbf{5}\ )$$

Whereas:

$$\bar{\boldsymbol{x}}_{\boldsymbol{1}}^{(2)} = \sum_{i=1}^{n_1^{(2)}} \frac{\boldsymbol{x}_{\boldsymbol{1i}}^{(2)}}{\boldsymbol{n}_{\boldsymbol{1}}^{(2)}} \qquad \dots (\ \mathbf{6})$$

$$\bar{\boldsymbol{x}}_{\boldsymbol{2}}^{(2)} = \sum_{i=1}^{n_2^{(2)}} \frac{\boldsymbol{x}_{\boldsymbol{2i}}^{(2)}}{\boldsymbol{n}_{\boldsymbol{2}}^{(2)}} \qquad \dots (\mathbf{7})$$

$$\bar{\boldsymbol{x}}_{\boldsymbol{n}}^{(2)} = \sum_{i=1}^{n_n^{(2)}} \frac{\boldsymbol{x}_{\boldsymbol{ni}}^{(2)}}{\boldsymbol{n}_{\boldsymbol{n}}^{(2)}} \qquad \dots (\mathbf{8})$$

Second: Calculating the distance between the means of each two independent variables in the two groups:

$$d_i = \bar{x}_i^{(1)} - \bar{x}_i^{(2)} = \begin{bmatrix} \bar{x}_1^{(1)} - \bar{x}_1^{(2)} \\ \bar{x}_2^{(1)} - \bar{x}_2^{(2)} \\ \vdots \\ \bar{x}_n^{(1)} - \bar{x}_n^{(2)} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \qquad \dots (9)$$

Third: Calculate the variance for all independent variables in each group :

a- The first group :

$$\boldsymbol{s}_{\boldsymbol{i}}^{\boldsymbol{2}} = \sum_{i=1}^{n_1} \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{2}} - \frac{(\sum_{i=1}^{n} \boldsymbol{x}_{\boldsymbol{i}})^2}{\boldsymbol{n}_{\boldsymbol{1}}} \qquad \dots (\mathbf{10})$$

b- The second group:

$$\boldsymbol{s_i}^{\boldsymbol{2}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n_2}} \boldsymbol{x_i}^{\boldsymbol{2}} - \frac{(\sum_{i=1}^{n} \boldsymbol{x_i})^2}{\boldsymbol{n_2}} \qquad ...(\boldsymbol{11})$$

Fourth: Calculate the product of each two variables (variance and covariance) within each group:

a- The first group :

$$\boldsymbol{s_{ij}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n_1}} \boldsymbol{x_i}\,\boldsymbol{x_j} - \frac{\sum_{i=1}^{n_1} \boldsymbol{x_i} \sum_{i=1}^{n_1} \boldsymbol{x_j}}{\boldsymbol{n_1}} \qquad ...(12)$$

b- The second group:

$$\boldsymbol{s_{ij}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n_2}} \boldsymbol{x_i}\,\boldsymbol{x_j} - \frac{\sum_{i=1}^{n_2} \boldsymbol{x_i} \sum_{i=1}^{n_2} \boldsymbol{x_j}}{\boldsymbol{n_2}} \qquad ...(\boldsymbol{13})$$

Fifth: Calculate the variance and covariance matrix between the two groups:

a- Variation within groups

$$\boldsymbol{v_{ii}} = \frac{\boldsymbol{s_i}^{2^{(1)}} + \boldsymbol{s_i}^{2^{(2)}}}{\boldsymbol{n_1} + \boldsymbol{n_2} - \boldsymbol{2}} \qquad ...(14)$$

b - Covariance

$$\boldsymbol{v_{ij}} = \frac{\boldsymbol{s_{ij}}^{2^{(1)}} + \boldsymbol{s_{ij}}^{2^{(2)}}}{\boldsymbol{n_1} + \boldsymbol{n_2} - \boldsymbol{2}} \qquad ...(\boldsymbol{15})$$

From Equation No. (14) and (15), we form the symmetric square variance and covariance matrix. The main diagonal represents the variances within the groups and the other elements represent the covariances, agencies:

$$v = \begin{bmatrix} v_{11} & v_{12}\cdots & v_{1k} \\ v_{21} & v_{22}\cdots & v_{2k} \\ \vdots & \vdots & \\ v_{k1} & v_{k2}\cdots & v_{kk} \end{bmatrix} \quad ...(16)$$

Sixth: Finding the linear discriminant function with standard coefficients:

$$\hat{\boldsymbol{L}} = \widehat{\boldsymbol{\alpha_1}}\,\boldsymbol{x_1} + \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}\boldsymbol{x_2} + \cdots + \hat{\boldsymbol{\alpha}}_{\boldsymbol{n}}\boldsymbol{x_n} \qquad ....(\ \boldsymbol{17})$$

Whereas:

$\hat{\boldsymbol{L}}$`: They represent standard discriminant values.

$\boldsymbol{x_n}$: They represent standard discriminant variables .

$\boldsymbol{\hat{\alpha}_n}$

: $The\ standard\ discrimination\ coefficients\ are\ represented\ and\ calculated\ as\ follows$:

$\hat{\alpha} = dv^{-1}$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \cdots & v_{1k} \\ v_{21} & v_{22} \cdots & v_{2k} \\ \vdots & \vdots & \\ v_{k1} & v_{k2} \cdots & v_{kk} \end{bmatrix}^{-1} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{bmatrix} \qquad \ldots (18)$$

Seventh: The relative importance of independent variables:

The task of discriminant analysis is to determine the relative importance of the independent variables that affect the differentiation process through the following equation:

$$\alpha^*_i = \alpha_i \sqrt{v_{ii}} \qquad \ldots (19)$$

The absolute values of (α^*_i) are compared, the largest absolute value corresponding to the independent variable(x_i) means that the independent variable (x_i) is the most important variable that has the ability to differentiate or discriminate between the two groups, The independent variable (x_i), which corresponds to the second largest absolute value of (α^*_i), the second most important variable has the ability to distinguish, and so on until the last variable has the ability to distinguish or differentiate between the two groups.

1-7: Some tests of the discriminant model:[10][12]

There are several tests related to the discriminant model, some of which we mention as follows:

1- **Wilks test**:

It is used to test the following hypothesis:

$\mathbf{H_0}$: There is no linear relationship between the independent variables.

$\mathbf{H_1}$: There is linear relationship between the independent variables.

The test statistic is written as follows:

$$\Lambda = \frac{|W|}{|\mathrm{T}|} = \frac{|W|}{|\mathrm{W} + \mathrm{B}|} \qquad \ldots (20)$$

Whereas:

W: The variance matrix represents the variance and covariance within the groups.

B: The variance matrix represents the variance and covariance between the groups.

T: The variance matrix represents the total covariance of the groups.

The test value (Λ) ranges between zero and one. If the test value appears equal or close to one, this means that there is no discrimination between the groups, meaning that the group means are equal. When the test value approaches zero, this indicates the power of discrimination and that the group means are not equal.

The test statistic is distributed approximately(χ^2) with a degree of freedom p(k-1). If the P-value appears less than(0.05), we do not accept Null hypothesis, meaning there is a linear relationship between the variables, and the discriminant function has the ability to distinguish or differentiate.

2- ***Eigen value***

To determine the extent of the function’s ability to discriminate between groups, we use the eigenvalues (distinctive roots). When the value of the distinctive roots is high, it is an indicator of the function’s ability to distinguish between groups.

3- ***Canonical correlation***

The value of the canonical correlation coefficient is equal to the square of the coefficient of determination, the goodness of fit of the discriminating function, and the higher the value of the canonical correlation coefficient is an indicator of the high quality of the discriminating function in its ability to discriminate and classify.

The correlation coefficient can be found (by dividing the sum of squares of variances between groups) by (the square root of the total sum of squares of variances).

4- ***Box′s M Test***.

This test uses the following hypothesis

$\mathbf{H_0}: \Sigma_1 = \Sigma_2 =. \ldots . = \Sigma_K$

$\mathbf{H_1}$: At least two groups have unequal variances

The test statistic $Box\acute{}s$ M is written as follows:

$$Box\acute{}s\ M = MC^{-1} \sim \chi^2{}_{\frac{1}{2}(\mathrm{K}-1)(P-1)} \qquad \ldots (21)$$

Whereas:

$$M = \ Bartlett's\ test = Ln|S| \sum_{i=1}^{K} n_i - \sum_{i=1}^{K} Ln|S_i|$$

$S = \frac{\sum_{i=1}^{K} n_i S_i}{\sum_{i=1}^{K} n_i}$ = Estimated variance and covariance matrix.

$\boldsymbol{S_i}$: Sample variance, which is an unbiased estimate of (Σ_i)as (i=1,2,….k)

$\boldsymbol{n_i}$: Degree of freedom for sample i.

K : Number of groups.

$$C^{-1} = \text{ constant} = 1 - \frac{2P^2 + 3(P-1)}{6(\mathrm{P}+1)(\mathrm{K}-1)} \left[\sum_{i=1}^{K} \frac{1}{n_i} - \frac{1}{\Sigma n_i} \right]$$

P : Number of independent variables.

1-8: *Classification of sample observations* :[7][9][10]

The classification process is summarized as follows:

1- **Differentiation point**:

It is the first step in classification and is defined mathematically as follows:

$$\bar{\bar{L}} = \frac{\bar{L}^{(1)} + \bar{L}^{(2)}}{2} \qquad \ldots (22)$$

Wereas:

$\bar{\bar{L}}$: Point of differentiation between groups.

$\bar{L}^{(1)}$: Average discriminant values for the first group.

$\bar{L}^{(2)}$: Average discriminant values for the second group.

2- ***observations* classification process**:

It is the second step in classification and comes after the formation of the differentiation point, and its work is summarized as follows:

a- Observations are classified into the first group if $\hat{L} > \bar{\bar{L}}$.

b- Observations are classified into the second group if $\hat{L} < \bar{\bar{L}}$.

c- Observations is randomly classified into one of two groups if it is $\hat{L} = \bar{\bar{L}}$.

1-9: Possibility of classification error :[2][11]

The probability of classification error is divided into two parts as follows:

1- Classification error p_{12}: It is the probability that the Observation will be classified as The second group is originally related to the first group.

2- Classification error p_{21} : It is the probability that the Observations will be classified as The first group originally belongs to the second group.

The classification probability is estimated as follows:

$$p_{12} = p_{21} = \emptyset\left({}^{-D}/_{2}\right) \qquad \ldots(23\ )$$

Whereas:

$\emptyset$: Standard normal distribution function

D: The root of the scale of MahaLanobis D^2 .

2- The applied aspect :

2-1: Description of the sample, research variables, and data collection:

Real data were adopted from social care records for pregnant women in health centers in Baghdad Governorate, Al-Hurriya District for the year 2023. The research sample included 100 pregnant women (spontaneous abortion = 50, natural birth = 50), and the statistical program (spss.29) was used to analyze the effect of variables, which are known as With the following:

The dependent variable (natural birth = 0, spontaneous abortion = 1)

The independent variables are defined as follows:

1- Age of the pregnant woman: (younger than 16 years, 16-35, older than 35 years)
2- Pregnant woman’s occupation: (housewife = 1, earner = 2, employee = 3)
3- Educational level of the pregnant woman: (cannot read or write = 1, primary = 2, intermediate = 3, preparatory = 4, university = 5)
4- Matching blood type between a man and his wife (matched = 1, not matched = 2)
5- Family kinship between a man and his wife (relative = 1, not relative = 2)
6- Smoking by pregnant women (smoker = 1, non-smoker = 2)
7- 7- Body mass index (normal=1 , non-normal)
8- Pregnant woman having her first pregnancy (yes = 1, no = 2).
9- Previous miscarriages of the pregnant woman (there is = 1, nothing = 2)
10- Ectopic pregnancy (there is =1, nothing=2)
11- Pregnant high blood pressure (Infected=1, not infected = 2)
12- Diabetes in pregnant women (Infected=1, not infected = 2)
13- Previous stillbirths for pregnant women (There is = 1, nothing = 2)

2-2: Test for normal distribution of data:

Since the sample size is greater than or equal to 30 observations according to the central purpose theory, the data follows a normal distribution and there is no need to conduct a normality test.

2-3: Testing for the absence of multicollinearity between independent variables:

By looking at Table (1), we see that the VIF value for all variables is less than(5), meaning there is no problem of multicollinearity between the independent variables..

Table (1) Collinearity Statistics

Variables	Collinearity Statistics	
	Tolerance	VIF
X1	.925	1.081
X2	.932	1.073
X3	.900	1.111

X4	.931	1.075
X5	.935	1.069
X6	.871	1.148
X7	.954	1.048
X8	.887	1.127
X9	.873	1.145
X10	.871	1.148
X11	.912	1.096
X12	.940	1.063
X13	.918	1.090
X14	.867	1.154

2-4: Testing that there are no outliers or extreme values:

Looking at all the values of Table No. (2), we see that any of these values is less than the tabular value of χ^2 at a degree of freedom (15-1)(2-1) and at a significance level of 0.95, which equals (23.68), and therefore we do not reject it. The null hypothesis that there are no outliers among all data related to all independent variables.

$\mathbf{H_0}$: There are no outliers or extreme values

$\mathbf{H_1}$: There are outliers or extreme values

Table (2) Mahanlops test values

MAH	MAH	MAH	MAH	MAH	MAH	MAH	MAH	MAH	MAH
14.73774	12.81274	11.84171	13.14082	13.13446	11.38321	11.21893	14.68292	13.36001	17.87840
13.23020	12.78174	16.51583	13.06018	11.19745	11.36692	13.18313	10.97172	14.00336	16.28666
13.24107	9.83237	15.44457	12.31801	12.69548	14.36886	14.90440	12.82023	13.44295	15.94396
12.85250	13.74054	16.18668	13.62740	14.33446	16.56049	10.28936	16.12048	10.78359	13.44290
15.21779	14.87870	16.52924	14.39136	13.44612	15.76873	15.83932	13.67492	14.67650	12.97856
13.92171	10.99680	14.84174	10.78655	12.01783	14.34466	12.08428	14.68082	9.51880	18.35910
11.03906	17.83701	18.04122	11.76648	12.51412	15.13938	13.91034	14.63654	11.91677	13.22630
13.57265	12.60825	13.19273	13.65652	14.82488	17.19736	15.31415	16.09671	14.18882	13.80843
14.70638	12.30795	14.74176	15.54808	12.54725	16.96996	14.30586	13.62860	11.42325	15.80896
14.51664	9.97796	14.94977	16.43445	10.32955	15.16201	14.29042	14.68292	12.52222	18.22574

2-5: Testing the condition that the variance and covariance matrix are equal between the two groups:

To determine the homogeneity of the members of the two groups, we use the (**Box's M**) test to test the following hypothesis:

$\mathbf{H_0}: \Sigma_1 = \Sigma_2$

$\mathbf{H_1}: \Sigma_1 \neq \Sigma_2$

Looking at Table (3), we see that the log determinant values are equal for the groups, which indicates the homogeneity of the members of the two groups, and from Table (4) the value of (Box’s M = 36.707) appeared with a probability value of(.053) and this indicates the homogeneity of the variance between the groups.

Table (3) Log Determinants

Log Determinants		
Y	**Rank**	**Log Determinant**
spontaneous abortion	8	-3.411-
natural birth	8	-4.307-
Pooled within-groups	8	-3.494-

Table (4) Box's M

Box's M		36.707
F	Approx	2.313
	Df1	15
	Df2	38531.248
	Sig.	.053

2-6 : Testing the condition that the means of the two groups are not equal:

$\mathbf{H_0}: \mu_1 = \mu_2$

$\mathbf{H_1}: \mu_1 \neq \mu_2$

The above hypothesis was tested in the next paragraph (2-8) and it was found that the function has the ability to discriminate, meaning we do not accept the null hypothesis

2-7: Steps of discriminant analysis:

1- Testing the significance of independent variables in the discriminant model:
Testing the significance of the study variables is to determine the importance of each variable individually and the extent of its influence in building the linear discriminant

function. Looking at table (5), we see that the independent variables (10) only have a significant effect in the process of differentiating between the two groups, as the probability values appeared smaller than (0.5).

Table(5)Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
(X_1)	.963	3.722	1	98	.000
(X_2)	.569	74.237	1	98	.047
(X_3)	.945	5.734	1	98	.019
(X_4)	.824	21.000	1	98	.000
(X_5)	1.000	1.100	1	98	.997
(X_6)	.981	1.927	1	98	.168
(X_7)	.999	.076	1	98	.054
(X_8)	.885	12.772	1	98	.001
(X_9)	.770	29.306	1	98	.000
(X_{10})	.840	18.622	1	98	.000
(X_{11})	1.000	.000	1	98	.003
(X_{12})	.998	.184	1	98	.669
(X_{13})	1.000	.036	1	98	.000
(X_{14})	.998	.160	1	98	.033

2- Knowing the entered of independent variables that make up the discriminative model

The stepwise selection method was used to select the variables that have an impact on the discriminant model. Looking at Table No. (7), we see that two significant variables were excluded (the pregnant woman’s occupation variable, the high blood pressure variable) and 8 significant variables were left that have a higher ability to differentiate between the two groups and that have The highest F value and the lowest wilk'ˢs lambda value in each stepwise of the gradual selection, as shown in the following figure:

Table (6) Wilks' Lambda

Wilks' Lambda									
Step	**entered of Variables**	**Lambda**	**df 1**	**df 2**	**Df 3**	**Exact F**			
						Statistic	**df1**	**df2**	**Sig.**
1	X_1	.477	1	1	98	381.730	1	98.000	.000
2	X_9	.442	2	1	98	219.867	2	97.000	.000
3	X_4	.416	3	1	98	165.445	3	96.000	.000
4	X_3	. 401	4	1	98	130.234	4	95.000	.000
5	X_8	.391	5	1	98	109.568	5	94.000	.000
6	X_{11}	.380	6	1	98	98.223	6	93.000	.000
7	X_{13}	.371	7	1	98	84.087	7	92.000	.000
8	X_{14}	.364	8	1	98	75.345	8	91.000	.000

3- Discriminant model with standard coefficients :

Table (7)

Standardized Canonical Discriminate Function Coefficients	
	Function
	1
X_1	-.811-
X_3	-.314-

X_4	-.399-	
X_8	.536	
X_9	.271	
X_{11}	.254	
X_{13}	.489	
X_{14}	.189	

Using table (7), the standard discriminant model can be written as follows:

$$\hat{\boldsymbol{L}} = -.811\boldsymbol{x_1} + -.314\boldsymbol{x_3} + .399\boldsymbol{x_4} + .536\boldsymbol{x_8} + .271\boldsymbol{x_9} + .254\boldsymbol{x_{11}} + .489\boldsymbol{x_{13}} + .189\boldsymbol{x_{14}} \quad \ldots(24)$$

Considering the discriminatory model, the relative importance of the factors affecting the process of separation and discrimination between the two groups and the case of spontaneous abortion can be explained as follows:

1- The independent variable, the age of the pregnant woman (X_1) is the most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient appeared between the discriminant model and the variable age (.811), and an inverse relationship. The lower the age of the pregnant woman, the greater the chance of spontaneous abortion. As for importance Relatively, age contributes more to the process of distinguishing between the two groups (81.1%).

2- The independent variable, pregnancy for the first time (X8) is the second most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminant model and the variable pregnancy for the first time (X8) appears to be equal to (.536), and the relationship is direct. The earlier the pregnancy, the greater the chance of spontaneous miscarriage. In terms of relative importance, the variable of getting pregnant for the first time comes in second place and contributes to the process of distinguishing between the two groups by (53.6%).

3- The independent variable, diabetes (X_{13}), is the third most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminant model and the diabetes variable (X_{13}) appears equal to (.489), and the relationship is direct. The greater the number of pregnant women with diabetes, the greater the incidence of spontaneous miscarriages. In terms of relative importance, the diabetes variable comes in third place and contributes to the process of distinguishing between the two groups by (48.9%).

4- The independent variable, the level of blood type matching between the pregnant woman and her husband (x_4), is considered the fourth most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminating model and the variable (x_4) appears equal to (.399), and the relationship is inversely related to blood type matching. Between pregnant men and women, the lower the chances of miscarriage, and in terms of relative importance, the variable (x_4) comes in fourth place and contributes to the process of distinguishing between the two groups by (39.9%).

5- The independent variable is the educational level of the pregnant woman (x_3) as the fifth most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminant model and the variable (x_3) appears equal to (.314), and the relationship is inverse. The lower the educational level of a pregnant woman, the greater the chance of miscarriage. In terms of relative importance, the variable (x_3) comes in Fifth place and contributes to the process of distinguishing between the two groups by (31.4%).

6- The independent variable: previous miscarriages of the pregnant woman (x9) is the sixth most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminating model and the variable (x9) appears equal to (.271), and the relationship is direct. The more miscarriages a pregnant woman has had, the greater the chance of miscarriage. In terms of relative importance, the variable (x8) comes in sixth place and contributes to the process of distinguishing between the two groups by (27.1%).

7- The independent variable, the pregnant woman’s high blood pressure (x11), is the seventh most important variable in the occurrence of spontaneous miscarriage, as the value of the absolute standard discrimination coefficient between the discriminant model and the variable (x11) appears equal to (.254), and the relationship is direct. The more cases of high blood pressure a pregnant woman has, the greater the chance of miscarriage. In terms of relative importance, the variable (x11) comes in seventh place and contributes to the process of distinguishing between the two groups by (25.4%).

8- The independent variable: previous stillbirths for pregnant women (x_{14}) is the last most important variable in the occurrence of spontaneous abortion, as the value of the absolute standard discrimination coefficient between the discriminant model and the variable (x_{14})

appears equal to (.189), and the relationship is direct. The more cases of stillbirth pregnant women have, the greater the chance of miscarriage. In terms of relative importance, the variable (x_{14}) comes in Last place and contributes to the process of distinguishing between the two groups by (18.9%).

2-8: Testing the ability of the discriminative model to discriminate:

For the purpose of knowing which discriminatory model has the ability to differentiate between the two groups, we test the following hypothesis:

$\mathbf{H_0}: \mu_1 = \mu_2$

$\mathbf{H_1}: \mu_1 \neq \mu_2$

There are several tests to test the above hypothesis, as follows:

1- ***Wilks' Lambda*****:** The significance of the differences between the means of the two groups was tested according to the formula (20) and the test results were included in the table (8):

Table (8) Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.345	101.635	5	.000

Looking at Table (8), we see the Wilks Lambda value of .345, which is closer to zero. This is an indication that there are significant differences between the two means and that the discriminating model has the ability to discriminate. We also see that (SIG = .000), and therefore we reject the null hypothesis and accept the alternative hypothesis: Average. Societies are unequal and therefore the discriminatory model has the ability to distinguish and differentiate between the two groups.

2- Eigenvalues:

Looking at the terms (10), we see the Eigen values for the discriminative model A (1.899), which indicates that the discriminative model has a high ability to discriminate, as the Eigen values are greater than the correct one. What confirms that 100% of the variance was explained. As for the Canonical Correlation, it reached (0.809), and this indicates the good fit of the discriminative model, as the coefficient of determination reached (65.45%).

Table (9)Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.899[a]	100.0	100.0	.809

2-8: Linear discriminant model classification:

Looking at the table (10), we see the accuracy of the final results of the classification, as it was found that 44 cases from the first group, at a rate of 89.8%, were classified correctly, and the remaining 5 cases were classified incorrectly, at a rate of 10.2%.

Regarding group 2, it was found that 45 cases were classified correctly, at a rate of (88.2%), and the remaining 6 cases were classified incorrectly, at a rate of (11.8%), As a general result, the results indicated that a percentage of 89% of cases in both groups were classified correctly, and this indicates high quality in the classification results.

Table (11) Classification Results[a]

			Predicted Group Membership		
			0	1	total
Original	Count	0	44	5	49
		1	6	45	51
	%	0	89.8	10.2	100.0
		1	11.8	88.2	100.0

a. 89.0% of original grouped cases correctly classified.

3- Conclusions:

After conducting discriminant analysis of the study sample data using the spss26 statistical program, the researcher reached the following conclusions:

a- The linear discriminant analysis method is appropriate for spontaneous abortion data after the conditions for the discriminant analysis method mentioned in Paragraph (1-5) are met. It can be used to distinguish and classify the new independent variables into one of two groups, either live births or spontaneous abortions.

b- The discriminating model between the two groups with standard coefficients is:

$$\hat{L} = -.811x_1 - .314x_3 + .399x_4 + .536x_8 + .271x_9 + .254x_{11} + .489x_{13} + .189x_{14}$$

c- The factors most affecting the incidence of spontaneous miscarriage, respectively, are (age x_1, first pregnancy x_8, diabetes x_{13}, compatibility of the couple's blood type x_4, educational level of the pregnant woman x_3, previous miscarriages x_9, high blood pressure x_{11}, previous stillbirths x_{14})

d- The factors contributing most to the distinction between the two groups, in percentages respectively, are[(x_1 =81.1%),(x_8 =53.6%),(x_{13} =48.9%), ($x_4 = 39.9\%$), ($x_3 = 31.4\%$), ($x_9 = 27.1\%$), ($x_{11} = 25.4\%$), ($x_{14} = 18.9\%$)

e- The coefficient of determination is equal to (.6545) and this means that the factors affecting the incidence of spontaneous miscarriage contribute (65.45%) to the discrimination and classification between the two groups.

f- The discriminative model has a highly efficient classification rate (89%)

Recommendations:

a- Paying attention to statistical analysis and highlighting its important role in the medical aspect.

b- Using the discriminant function model in early diagnosis.

c- The study sheds light on the moral factors that affect the chances of spontaneous miscarriages, deals with them realistically, and prepares educational programs and seminars for pregnant women to avoid spontaneous miscarriages.

d- Use other multivariate analysis methods and compare the results with discriminant analysis

Sources:

1- Abdul Razzaq, Aseel,(2019) ," Using the linear discriminant function to classify thalassemia patients in Kirkuk General Hospital, Published research, Journal of Management and Economics, Volume (11),Issue 78, Al-Mustansiriya University.

2- Al.Katib, Muhammad Osama,(2018) " The linear characteristic function in the case of more than two sets, Published research", Journal of Education and Science, Volume (23), Issue (4),

3- Al-Ammar, Radwan and others.(2021)," Using Discriminant Analysis To Classify Banks According To Credit Risk(An Applied Study On Private Traditional Commercial Banks Operating In Syria During The Period 2008-2018" Tishreen University Journal of Economic and Legal Sciences, Volume 43, Issue 1

4- Al-Demerdash, Hani Muhammad Ali, (2022)," Discriminant Analysis vs. Multivariate Analysis of Variance :An Empirical Study to Identify the Determinants of Stagflation in Egypt 1980 -2021" Port Said University, Journal of Financial and Business Research, Volume 23, Issue 4

5- Al-Hamzi, Amat Al-Latif Muhammad Ali and Al-Hamzi, Amani, (2023), " The use of discriminant analysis and its application to the birth of a premature baby in the Seventy-Year Hospital for Maternity and Childhood, Sanaa, the capital" Electronic Journal of University of Aden for Basic and Applied Sciences" Volume4, No1.

6- Al-Jaouni, Fredo Ghanem, Adnan,(2018)," Multivariate statistical analysis (discriminant analysis) in the characterization and distribution of families within the socio-economic structure of society, Damascus University Journal of Economic and Legal Sciences, Volume (23), Issue (2).

7- Al-Marsoumi, Aseel Mubadar Dawoud, and Al-Hashemi, (2019), " Diagnosing factors affecting heart disease using discriminant analysis" Journal of Management and Economics, Volume 42, Issue120.

8- Fadl Almul, Suleiman, Ali Absher,(2019) " Comparison between discriminant analysis, binary logistic model and neural network models in classifying observations, PhD thesis in statistics", College of Graduate Studies, Sudan University of Science and Technology.

9- Hunaiti, Dokhi & eta,(2017)," Distinguishing the poor from the non-poor families in the remote areas of the Southern Jordan region", Journal of Economic Development and Policies, Volume (7), Issue (1).

10- Jaafar, Raed Salman ,(2022)," Predicting of surface water quality index using Discriminant analysis and multiple linear regression" Hama University Journal, Volume 5, Issue 20.

11- Matrood , Daham Owaid and Al-Quraishi , (2023),"The use of linear discriminant analysis to determine the variables affecting cancer diseases"University of Kirkuk Journal For Administrative and Economic Science, Volume 13 Issue 1

12- Mustafa ,Anwar Al-Zein Babakir , (2023),"Using Linear Discriminant Analysis to Determine the Most Important Factors Affecting the Classification of Leukemia in Sudan(A case study of Al-Tawrah Hospital, Khartoum, during the year 2022 AD)" International Journal of Humanities and Social Sciences" Issue 44