**Drishti Bansal**

Dr. Douglas MacDonald

DATA 5070

***“Logistic and Discriminant Analysis”***

Using data obtained from 276 women who had given birth within one year of time, the factors affecting the women’s childbirth experience (associating with satisfaction with childbirth), measurement of five-item version of ESI-R (Expression of Spirituality Inventory- Revised) are being assessed in this study. The aim of this paper is to run a logistic regression and discriminant analysis.

We can tell that there are missing values in the data, for the education level and there are five parts of what kind of birth delivery is consisted of from the religious factor. The (Case 45, 59, 146, 204, 221, 254) six women who are in undergraduate left blank in values in refaffca, refaffev, relafffsp, refaffno. We can consider those to be either unmedicated births or not pregnant, and thereby removing them would be the ideal for the data quality. The reason for removal is because logistic regression doesn’t have the capability to deal with missing or null values. Case 11 has a missing value in placedes, Case 267 got a missing value in pregprob, we will leave them as be because they have ordinal values which cannot be replaced with average or any function. There are a few cases that contain missing values in age variable, but we can replace them with the average mean. Case 20, 40 in ESI-R variables have been missing, we can replace them with the mean value as they are of scale data type. After data cleaning, the data now consists of 31 variables (6 variables added from replacing the missing data with averages- age\_1, esicosp\_1, esiepdpo\_1, esiewbpo\_1, esiparpo\_1 and esirelpo\_1) [Appendix 1]

Logistic regression is used to estimate the probability between the dependent variable and one or more independent variables. This type of analysis can help you predict the likelihood of a choice being made. In this case, we determine the predictors based on correlation matrix. The better the variables are correlated with satisfy variable, the better they have chance to being predictors. We have used Pearson’s Correlation [Appendix 2]. Pearson’s correlation varies between -1 and +1 which means that -1 is a negative correlation whereas +1 is a perfect positive correlative. We have all ESI-R variables that are positively correlated with each other at the level of significance of 0.01, the standard p-value is 0.05; however, we noticed that with 0.05 p-value, some variables have weakly correlated with each other. Since we know that the 0.01 alpha (in reduction from 0.05) reduces the chance of a false positive (committing a type I error), we will select the variables that have positively correlated at a 0.01 level of significance. We have relaffno (No Religious Affiliations) correlating with relaffca (Catholic Religious Affiliations) and relatend (Attendance to religious/spiritual practices), relafssp (Spiritualist Religious Affiliations) correlating with refaffev (Evangelical Religious Affiliations) and relaffca. Based on those correlations related to religious, we have the best predictor which is relaffno correlating with relatend for .433 level of significance at 0.01 alpha. This tells us that those women who gave birth and had no religious beliefs are likely to attend the religious or spiritual sessions/practices. Those women want to instill the faith, religious beliefs into their kids if possible.

From Figure 2 in Appendix 2, Esirelpo (Religious Beliefs and Behaviours) correlating with esicospo (Spiritual Beliefs) with the best predictor in mind, which tells us that those women who are practicing religious beliefs and acting on religious behaviors have 86.4% chance of adopting the spiritual beliefs.

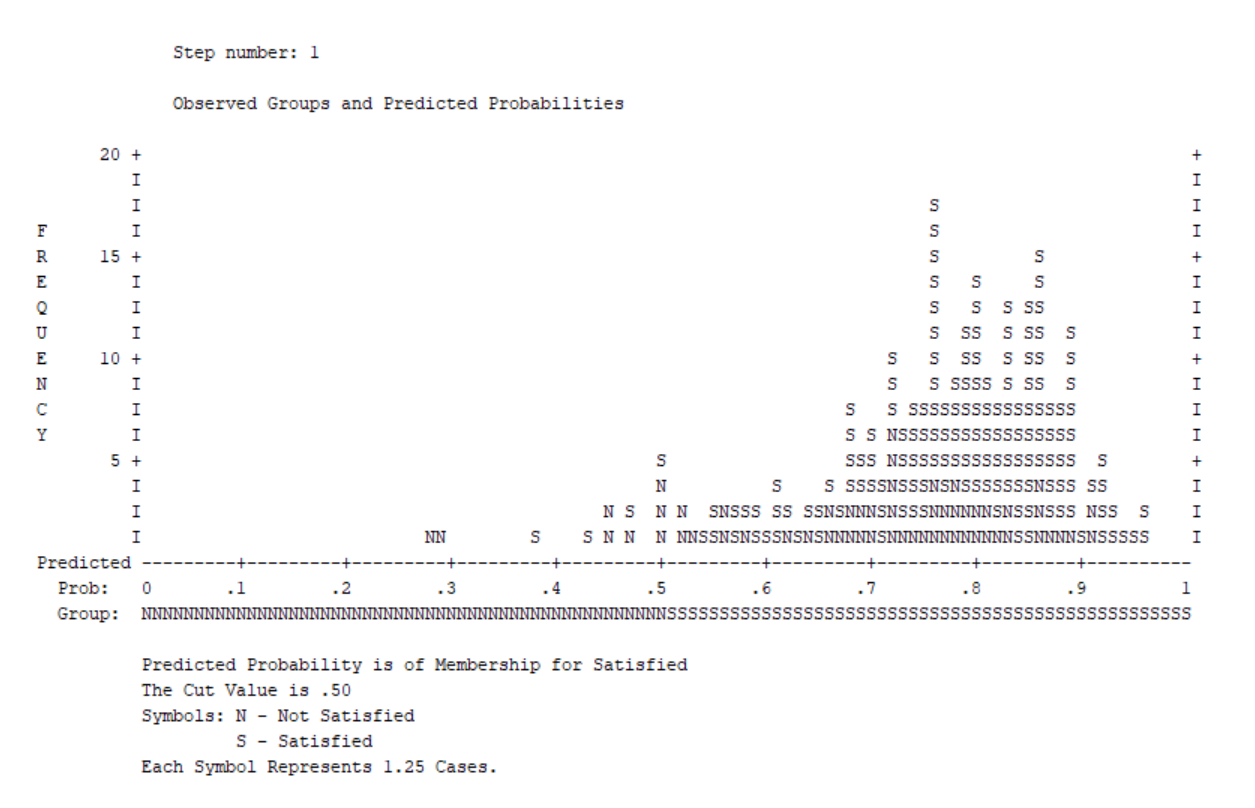
From Figure 3 in Appendix 2, brthtype (Type of childbirth), hmbrth (Home as place of birth), partdad (Father participated in birth), and partmidt (Traditional midwife participating in birth) that are positive correlated with satisfy variable.

Considering all correlations, measurement of five scales of ESI-R variables are best predictors for the logistic regression and discriminant analysis (later below). First, we have to make sure that the assumptions of logistic regression have been met in order to run the analysis. Secondly, we are to interpret the output (*some* which are significant for this paper) and the rest of the output generated are displayed in appendix.

Logistic regression assumptions to be met:

1. Your dependent variable has to be dichotomous scale. This assumption is met because we satisfy variable for grouping variable and it is on dichotomous scale, 0.00 is coded as not satisfied and 1.00 is coded as satisfied.
2. You have one or more independent variables. This assumption is met because we have determined that the predictors are in scale/nominal data types.
3. You should have independence of observations and the dependent variable should have mutually exclusive and exhaustive categories.

Using Satisfaction of childbirth, pregnancy, etc as a grouping variable to run a logistic regression and using the Revised\_ESI variables for independent variables, and the result is below:



As we can see the graph above, the frequency of having satisfaction with childbirth aligning with ESI-R items are higher than not satisfied. Conceptually, you can see the other (classification) table with frequency (Appendix 5).

Table

Description automatically generated

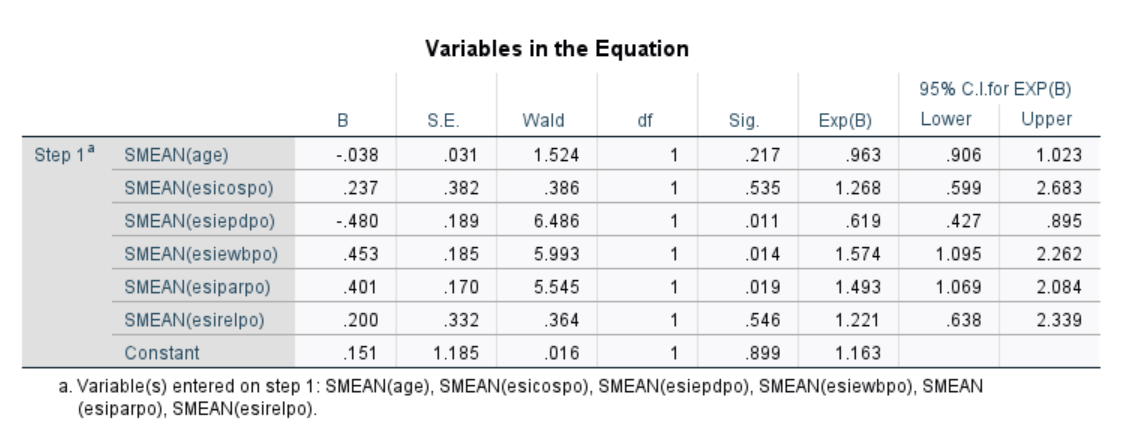
This table above contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. This model summary allows you to specify model within a single regression or, in this case, within multiple regressions. The explained variation in the dependent variable based on our model ranges from 7.10% to 10.6%.

Logistic regression estimates the probability of being satisfied with childbirth based on the predictors. If the estimated probability of the event occurring is greater than or equal to 0.5 (better than even chance), PSPP Statistics classifies the satisfactory of childbirth, otherwise not. It is very common to use binomial logistic regression to predict whether cases can be correctly classified (i.e., predicted) from the independent variables. Therefore, it becomes necessary to have a method to assess the effectiveness of the predicted classification against the actual classification.

Table

Description automatically generated

We have the overall percentage in the classification table of 77.4% which indicates the accuracy of classification. There were 98.0percent that the women had received satisfaction of childbirth. Note that the table has a footnote which states that the cut value is .500 which means if the probability of a case being classified into satisfaction with childbirth category is greater than 500, then that case is classified into the category.



The Wald test ("Wald" column) is used to determine statistical significance for each of the independent variables. The statistical significance of the test is found in the "Sig." column. From these results you can see that esiepdpo (p = .0.011), esiewbpo (p = 0.014), esiparpo (p = 0.19) added significantly to the model/prediction, but age (p = .217), esicospo (p = .535) and esirelpo (p = 0.546) did not add significantly to the model.

Discriminant analysis is used to examine the significance of difference between groups in terms of predictors. It evaluates the accuracy of classification just like logistic regression. Now we run a one-way ANOVA for predictors.

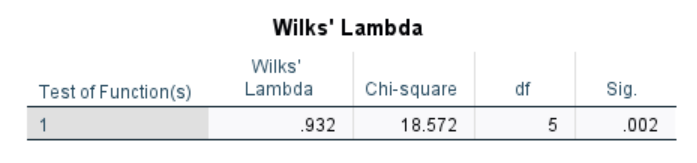
Discriminant analysis assumptions to be met:

1. A sample size that would produce more than 20 degrees of freedom for error in the univariate ANOVA case should ensure robustness with respect to multivariate normality as long as sample size are equal. This assumption is met because we have 275 df for error when running univariate ANOVA in PSPP [Appendix 11].
2. Homogeneity of variance.

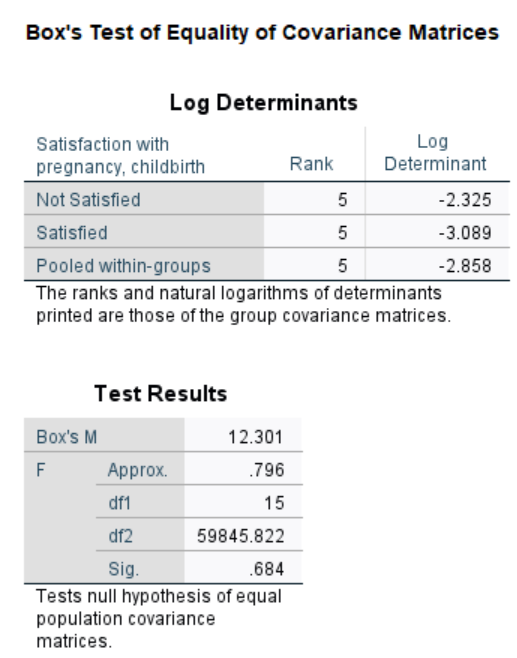
A picture containing text, screenshot, indoor

Description automatically generated

Now, we are running Discriminant Analysis by clicking Classify in Analysis tool in SPSS, then Discriminant, and here’s the result below along with interpretations:



Wilks’ lambda indicates the significance of the discriminant function. This table above indicates a highly significant function which is p < .005 and provides the proportion of total variability not explained, i.e., it is the converse of the squared canonical correlation. Therefore, we have 93.2% unexplained.



As we know that the assumption for ANOVA is that the variances were equal for each group but in discriminant analysis, the basic assumption is that the variance co-variance matrices were equal. Box’s M tests hypothesis that the covariance matrices do not differ between groups formed by the dependent. In the Log Determinant table, it appears that they are similar within the range of 1 or 2, and Box’s M is 12.301 with F = .796 which is highly significant at p = 0.005. It is not considered as significant for the report as they don’t tell us much.

Table

Description automatically generated

The table above is an interpretation of the discriminant coefficients is like that in multiple regression. It provides an index of the importance of each predictor like the standardized regression coefficients (beta’s) did in multiple regression. The sign indicates the direction of the relationship. Esiewbpo score (which is well-being) was the strongest predictor while low esiepdo (which is spiritual experiences) was next in importance as a predictor. These two variables with large coefficients stand out as those that strongly predict allocation to the satisfaction with childbirth or not satisfied group.

Table

Description automatically generated

The table above is considered more accurate than standardized canonical discriminant function coefficient which is previous section because the structure matrix table illustrates the correlations of each variable with each discriminant function. The Pearson coefficients are similar to serving like factor loadings in factor analysis. By identifying the largest loadings for each discriminate function, we get to extract an insight into how to name each function. Here we have esirelpo (Religious beliefs and behaviors) which is low score, this generally indicates that the mothers don’t have religious beliefs which is why their child is unlikely to learn about religious behaviors.

In conclusion, the discriminant analysis was performed to predict whether the patient had a satisfaction with childbirth, pregnancy. Predictor variables were ESI-R items which are measurement of five scales. Significant mean differences were observed for all the predictors on the dependent variables/grouping variables. While the log determinants were similar, Box’M indicated that the assumption of equality of covariance matrices was violated. However, given the decent sample size, this problem can be regarded as not serious which means we need a large sample to have more accurate findings.

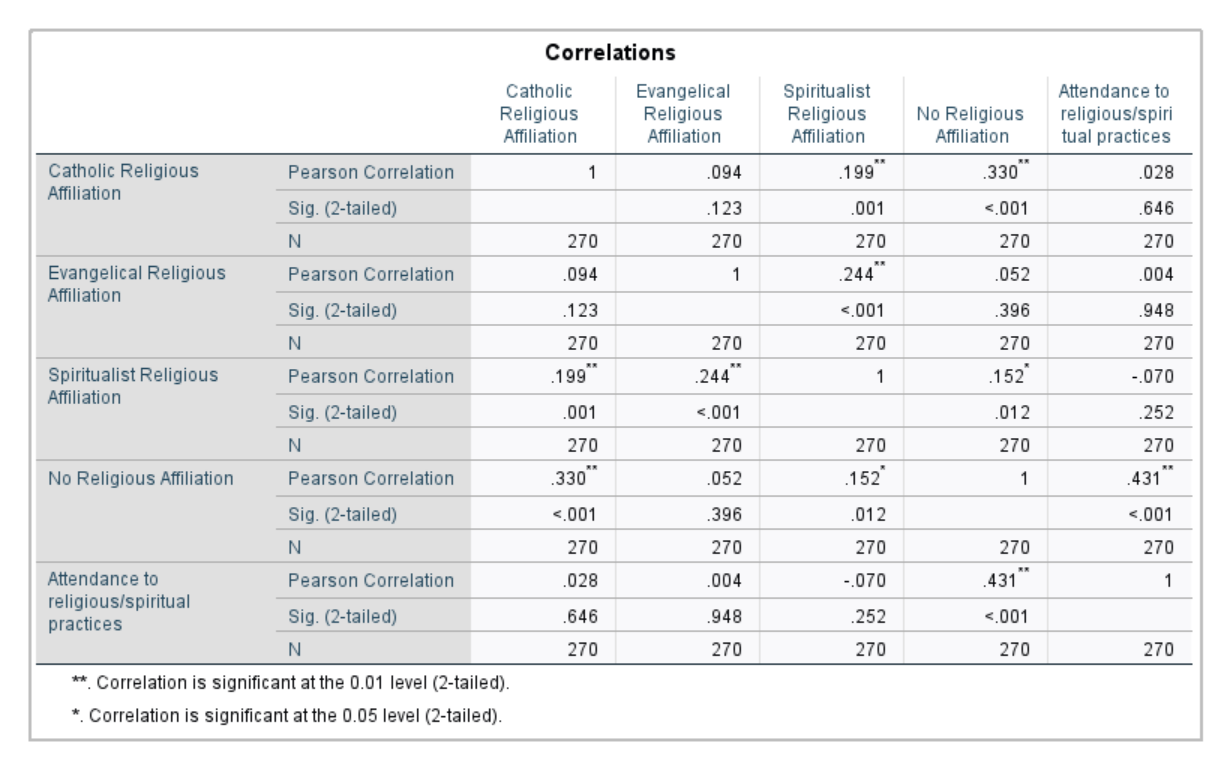
**Appendix**:

1. A sample data from the birthdataassignment3.sav after data cleaning (you can notice that the original variables have missing values and they have been replaced with average and named differently.

A screenshot of a computer

Description automatically generated with low confidence

1. Correlation Matrix for determining the predictors [Figure 1]



[Figure 2]

Table

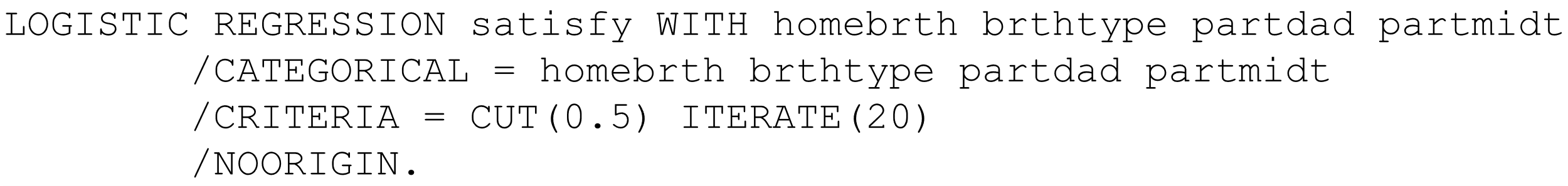
Description automatically generated

[Figure 3]

Table

Description automatically generated

1. the syntax for logistic regression in PSPP:



1. categories variable codings

PSPP output

Table

Description automatically generated

SPSS output

Table

Description automatically generated

1. Classification table (Block 0)

Table

Description automatically generated

1. Variation in the equation (Block 0)

Table

Description automatically generated

1. Hosmer and Lemeshow Test (Block 1)

Table

Description automatically generated

1. Omnibus test of model coefficient (Block 1)

Table

Description automatically generated

1. Case processing summary

Table

Description automatically generated

1. Casewise List

Table

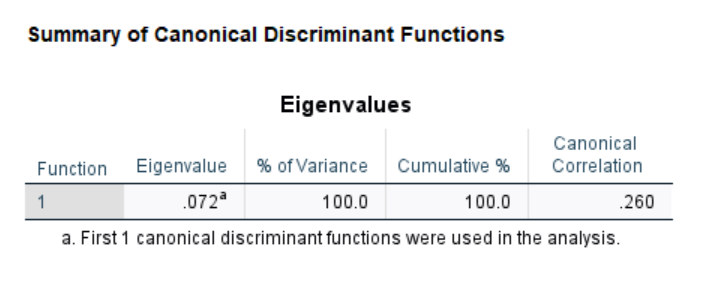
Description automatically generated

1. univariate analysis for DA assumption in PSPP

Graphical user interface, text, application, email

Description automatically generated

1. Eigenvalues



1. Classification Summary for DA

Table

Description automatically generated

1. Classification Result

Table

Description automatically generated

1. Homogeneity of variance

Table

Description automatically generated