DETERMINANTS OF LABOR FORCE PARTICIPATION OF MARRIED WOMEN: A CASE STUDY OF DISTRICT GUJRAT

Mirza Rizwan Sajid
Department of Statistics,
University of Gujrat,
PAKISTAN
mirzarizwansajid@gmail.com

Fauzia Maqsood
Department of Sociology,
University of Gujrat,
PAKISTAN
fauziamaqsood@gmail.com

Sidra Maqsood
Department of Sociology,
University of Gujrat,
PAKISTAN
sidramaqsood@gmail.com

Iram Afzal
Department of Statistics,
University of Gujrat,
PAKISTAN
iram.afzal99@yahoo.com

ABSTRACT

This paper examines determinants of labor force participation of married women in District Gujrat, Pakistan. Although in Pakistan women are participating in labor force but their proportion is very low. According to Labor force Survey 2009-10 female labor force contribution is 15.45% which is very low as compared to other South Asian Countries. In this study an attempt is made to find out the determinants which affect female labor force participation. The study was conducted with a sample of 301 married women by using stratified two stage cluster sampling method. Multilayer Perceptron Neural Network Model was used for analysis purpose. The study showed that literacy status, area of residence, family system, family size, husband's education were most important determinants in predicting female labor force participation.

Key words: Female Labor force participation, Determinants, Multilayer Perceptron Neural Network analysis

INTRODUCTION

Although women in Pakistan are participating in labor force but their proportion is significantly low i.e. 15.45% (2009-10) as compared to other South Asian countries. Studies indicate that in South Asian countries, labor force participation rate is 42 % in Bangladesh, 41% in Nepal, 32 % in India and Bhutan, 37% in Sri Lanka (World Bank, 2002). Pakistan's low labor force contribution is due to the lesser percentage of women in labor force market. Consequently this is a most important issue regarding the improvement of Pakistan. (Ejaz 2007).

Cultural milieu of different regions could be important factor for variation in female labor force participation. Similarly in Pakistani society, cultural values, normative structure and relatively conservative profile of society provide ample justification for women in general and married women in particular, not to enter in labor force. Various studies have identified different factors like husband education, family size, locality, husband employment status, education of women which affect women's participation in labor force (Azid et al 2001, Kohara 2008)

Within the cultural context of Pakistani society, various factors are thought to be considered contributing to low level of labor force participation of married women. It is a common observation that after marriage family responsibilities including raising children and managing household chores are considered most important for married women. This responsibility may limit their capacity in terms of time and energy to participate in labor force. Type of family system might be another important element affecting women's labor force participation. For instance in nuclear family system there are more chances for a woman to take part in labor force because of the freedom to take decisions. In joint family system, woman might not be as independent to take decisions because of the influence of other family members.

Similarly, literacy level of women might affect their participation in labor force. With higher level of education women might get more liberty to enter labor market because of the respectability attached with education and respective job (Lisaniler 2001). Area of residence could be another factor affecting women's participation in labor force. For instance women living in rural area have lesser chances to participate in labor force (Ferdaus 2006). Studies also indicate that husband's education could have an impact on woman's labor force participation. With the higher level of education of husband women's chances to enter labor force increase (Ejaz and Khan 2009). The present study is focused to identify the factors which significantly affect women's labor force participation in District Gujrat.

Objectives of the study

a. To find out the demographic profile of married employed and unemployed women.
b. To build a Neural Network model to find out the factors affecting married women's labor force participation
c. To determine the importance of each factor in predicting the married women labor force participation

MATERIAL AND METHODS

It was a cross sectional exploratory study to explore the determinants of Labor Force Participation of Married Women in District Gujrat. It was conducted within the jurisdiction of District Gujrat. A married woman living in Tehsil Gujrat having age 16-48 years would be considered as sampling unit of the study. All the married women having age 16-48 years living in District Gujrat comprised target population of the study. Tehsil Gujrat was the sampled population of the study. Stratified 2-Stage cluster sampling was used in this study. There are three Tehsils of District Gujrat. Firstly the sampled population (Tehsil Gujrat) was stratified into rural and urban areas because there was an element of variability according to the women labor force participation (female labor force participation is different in rural and urban areas). In the first stage of cluster sampling every rural and urban union council was considered as clusters. One urban and one rural union council were selected randomly. In the second stage one village and one block was selected randomly further from selected rural and urban union councils. Every married woman aged 16-48 years present in the selected areas was a respondent of the study. 301 married women who were living in the selected rural and urban blocks and met the defined criteria were considered as a sample size of the study. Data was collected by using a well structured questionnaire which contained the demographic and socio-economic information of the respondents.

Data Analysis Technique

Different statistical techniques were used to analyze the data. For description of variables frequency distribution was used. Neural network very advanced statistical technique which was used to make a prediction of categories of binary dependent variable and relative importance of determinants in prediction process.

Neural Network:

The term neural network applies to a loosely related family of models, characterized by a large parameter space and flexible structure, descending from studies of brain functioning. As the family grew, most of the new models were designed for non biological applications, though much of the associated terminology reflects its origin. Neural networks are the preferred tool for many predictive data mining applications because of their power, flexibility, and ease of use.

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

i. Knowledge is acquired by the network through a learning process.
ii. Interneuron connection strengths known as synaptic weights are used to store the knowledge.

The Multilayer Perceptron Neural Network Model:

Multilayer Perceptron Neural Network known as feed forward architecture because the connection in the network flow forward from the input layer to the output layer without any feedback loops. A multilayer feed forward neural network is an interconnection of perceptrons in which data and calculations flow in a single direction, from the input data to the outputs. The number of layers in a neural network is the number of layers of perceptrons. This network has an input layer (on the left), one hidden layer (in the middle) and an output layer (on the right). There is one neuron in the input layer for each predictor variable ($x_1 \ldots x_p$).

Activation Functions**.** The activation function "links" the weighted sums of units in a layer to the values of units in the succeeding layer.

a) Softmax: This function has the form: $\gamma(c_k) = \exp(c_k)/\Sigma_j \exp(c_j)$. It takes a vector of real-valued arguments and transforms it to a vector whose elements fall in the range (0, 1) and sum to 1. Softmax is available only if all dependent variables are categorical. When automatic architecture selection is used, this is the activation function for units in the output layer if all dependent variables are categorical.

b) Hyperbolic tangent: This function has the form: $\gamma(c) = \tanh(c) = (e^c - e^{-c})/(e^c + e^{-c})$.

c) It takes real-valued arguments and transforms them to the range (–1, 1). (Sajid R.M 2011)

RESULTS AND DISCUSSION

Descriptive Statistics

In this section basic description of demographic variables has been given. For qualitative variables frequencies and for quantitative variables basic Descriptive statistics has been computed. A question regarding literacy status of females is very important in this study. Table -1 shows that 76.7% females were literate. They knew the basic reading and writing of daily routine life. Only 22.9% females said that they didn't attend even single day schooling.

Table: 1 Frequency and Percentage of Literacy Status

	Literacy status	
Categories	Frequency	Percent
Yes	231	76.7
No	69	22.9
Total	300	99.7
Missing	1	.3
Total	301	100.0

Studies have shown that women labor force participation is not much high in Pakistan. Table 2 also shows that female labor force participation is low. Only 17.6% women were engaged in any type of paid labor at the time of study.

Table: 2 Frequency and Percentage of paid labor

Are you engaged in any paid labor at this time		
Categories	Frequency	Percent
Yes	53	17.6
No	248	82.4
Total	301	100.0

Table 3 shows that 45.2% females were from rural area and remaining 54.5% from urban area of tehsil Gujrat. Women labor force participation differs in Pakistan rural and urban community.

Table: 3 Frequency and Percentage of Area of residence

	Area of residence	
Categories	Frequency	Percent
Rural	136	45.2
Urban	164	54.5
Total	300	99.7
Missing	1	.3
Total	301	100.0

Table 4 gives an idea about the type of family system which the respondent had at the time of study. 58.8% of the female respondents living in the nuclear family system and remaining in joint family system. Family system matters a lot in the female labor force participation.

Table: 4 Frequency and Percentage of Family system

	Family system	
Categories	Frequency	Percent
Nuclear	177	58.8
Joint	121	40.2
Total	298	99.0
Missing	3	1.0
Total	301	100.0

Table 5 provides the description for quantitative variables Family size has Minimum value is 1, Maximum value is 25, mean value is 6.79 and Standard deviation is 3.410 and the variable education that your husband has completed Minimum value is 0, Maximum value is 25, mean value is 11.63 and Standard deviation is 3.125.

Table: 5 Descriptive statistics for Family Size and husband's Education:

Variable	**N**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Family Size	293	1	25	6.79	3.410
Education that your husband has completed	233	0	20	11.63	3.125

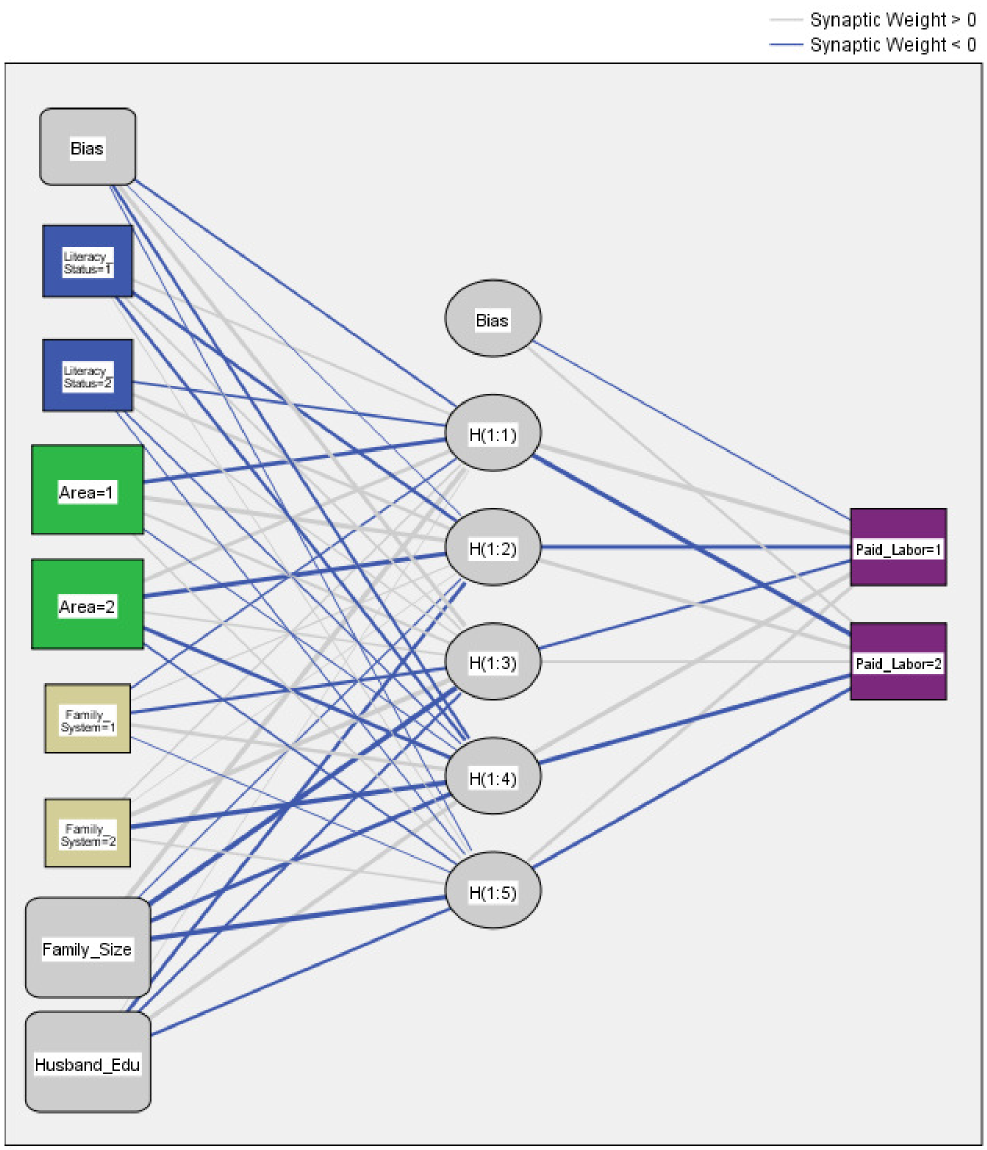
Multilayer Perceptron Neural Network (MLPNN)

Multilayer perceptron method was used in which the dependent variable is paid labor and the independent variables are literacy status, area of residence, family system, family size, husband education.

Table 6: Case processing summary

		N	Percent
Sample	Training	148	65.2%
	Holdout	79	34.8%
Valid		227	100.0%
Excluded		74	
Total		301	

Table 6 shows the model summary of Multilayer perceptron. The model partitions the data into two samples Training and Holdout. Total cases are 301 in which 148 cases are chosen in training samples and 79 in holdout samples.74 cases are excluded from the model at this stage.



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

Figure 1: MLPNN Architecture

Figure-1 and Table 7 shows the structure of Multilayer Perceptron Neural Network known as feed forward architecture because the connection in the network flow forward from the input layer to the output layer without any feedback loops. The number of layers in a neural network is the number of layers of perceptrons. This figure is an aggregation of input layers, hidden layers and output layers. The input layer contains the predictors (literacy status, area of residence, family system, family size, husband education). The hidden layers contain unobservable nodes, or units. The output layer contains the responses. Since paid labor is a categorical variable with two categories, it is recoded as two indicator variables. The value of each hidden unit is some function of predictors and each output unit is some function of hidden units. Behind the figure the hidden layers and the output layers uses some mathematical activation functions. The hidden layer activation function is hyperbolic tangent and output activation function is soft max.

Number of units in the input layer are 8. Figure -1 and table- 7 depicts that Ist unit of input layer has positive relationship with 2 and 4 unit of hidden layer and negative relationship with 1, 3 and 5 unit of hidden layer. 2nd unit of input layer has positive relationship with 2 and 3 units of hidden layer and negative relationship with 1, 3 and 5 unit of hidden layer. 3rd unit of input layer has positive relationship with 1 and 4 unit of hidden layer and negative relationship with 2, 3 and 5 unit of hidden layer. 4th unit of input layer has positive relationship with 1 and 3 unit of hidden layer and negative relationship with 2, 4 and 5 unit of hidden layer and so on.

Table 7: Parameter Estimate

Parameter Estimates

Predictor		Predicted						
		Hidden Layer 1					Output Layer	
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	[Paid_ Labor=1]	[Paid_ Labor=2]
Input Layer	(Bias)	-.807	-.139	1.952	-1.345	-.343		
	[Literacy_Status=1]	.946	-1.527	.798	-1.374	.257		
	[Literacy_Status=2]	-.990	1.396	.670	-.653	-.440		
	[Area=1]	-2.047	2.396	1.286	-.404	.720		
	[Area=2]	1.784	-2.465	.761	-1.805	-.743		
	[Family_System=1]	-.723	.468	-1.378	1.693	-.264		
	[Family_System=2]	.679	.114	2.999	-3.349	.783		
	Family_Size	3.245	-.528	-4.946	-2.274	-3.712		
	Husband_Edu	.054	-1.536	-1.234	2.205	-1.363		
Hidden Layer 1	(Bias)						-.648	1.159
	H(1:1)						2.731	-2.720
	H(1:2)						-2.058	1.907
	H(1:3)						-1.315	1.027
	H(1:4)						2.696	-2.384
	H(1:5)						1.631	-1.534

1st category of response variable has strong positive relationships with 1st unit of hidden layer and strong negative relationships with 2nd unit of hidden layer. Category 2nd has strong positive relationships with 2nd unit of hidden layer and strong negative relationships with 1st and 4th units of hidden layer. Grey lines show positive weights and blue lines show negative weights. Thickness of lines shows the strength of weights.

The classification matrix provides a comprehensive picture of the classification performance of the model. The ideal classification matrix is the one in which the sum of diagonal is equal to the number of samples. On cells diagonal of cross-classification are correct classification and off cells diagonal of the cross classification are incorrect classification.

Table-8 shows the classification results of analysis. In which 18 out of 25 who actually belong to group 1 are classified correctly in training sample. 120 out of 123 who actually belong to the group 2 are classified correctly in training. Overall 93.2% of the training cases are correctly classified. 8 out of

17 who actually belong to group 1 are correctly classified in holdout sample. 52 out of 62 who actually belong to group 2 are correctly classified in holdout sample. Overall 75.9% of the holdout cases are correctly classified. It means the estimated model provide better prediction results.

Table 8:- Classification of paid labor for training and holdout sample

Sample	Observed	Predicted		
		Yes	No	Percent correct
Training	Yes	18	7	72.0%
	No	3	120	97.6%
	Overall percent	14.2%	85.8%	93.2%
holdout	Yes	8	9	47.1%
	No	10	52	83.9%
	Overall percent	22.8%	77.2%	75.9%

Dependent Variable: Are you engaged in any paid labor at this time?

Figure 2 shows box plot of predicted pseudo probabilities. The chart shows clustered box of predicted pseudo probabilities for categorical dependent variable for both testing and training samples. The x-axis crossponds to observe response categories and y-axis crossponds to predicted categories. The portion of the box plot above 0.5 on the y-axis shows correct predictions. The portion below the 0.5 shows incorrect predictions. The 1st box plot of predicted-by-observed chart provides the same information as the first diagonal of the classification table. The 2nd box plot shows for cases that have observed category 1st the predicted pseudo-probability of category 2nd. It is representing incorrect classification because box plot is below the 0.5 mark and the case above the box plot are misclassified. The first box plot of second category shows for cases that have observed second category, but predicted pseudo-probability of 1st category. The second box plot of second category shows those cases that have observed category second but the predicted pseudo-probability of second category. In this case, it can say that the second box-plot of first category and first box-plot of second category shows misclassifications.

Figure 2: Predicted Pseudo Probability

Figure-3 shows the Receiver Operating Characteristic (ROC) curve. It provides a chart which represents the sensitivity and specificity for all possible cut offs in a single plot. ROC curve shows the trade-off between the true positive rate or sensitivity (portion of positive cases that are properly identified) and the false negative rate (portion of negative cases that are improperly identified as positive) for a given model. The chart shown in the figure displays two curves, one for the each category of the dependent variable .Both curves are close to upper left corner; which shows that fitted model is good.

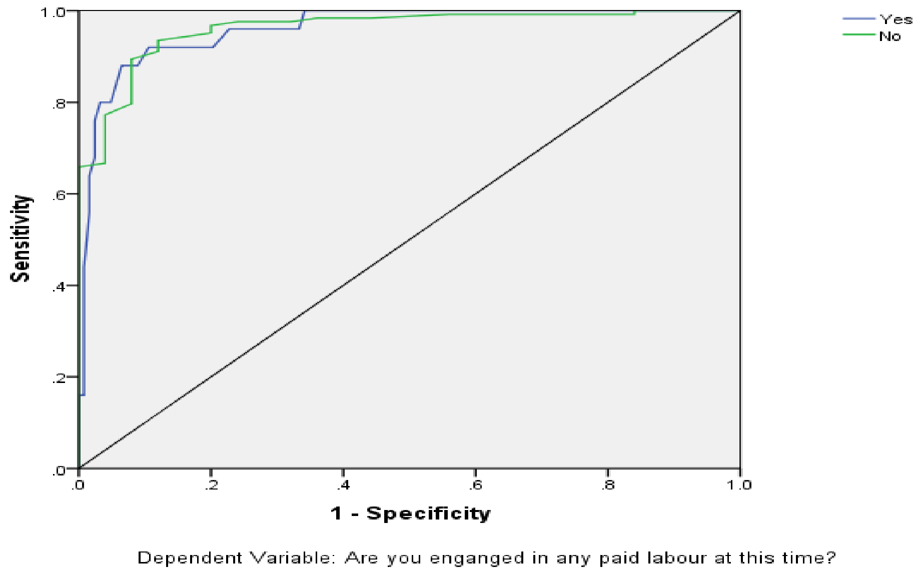


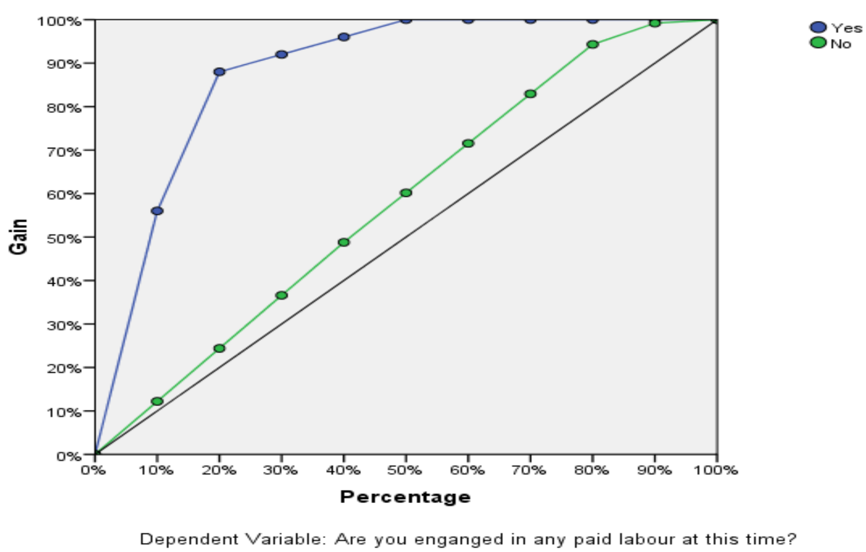
Figure 3: ROC Curve

To assess the accuracy of the model, area under the curve can be measured. Table-9 depicts the area under curve. If the area is 1 the test is an ideal test, because it achieves both 100% sensitivity and 100% specificity. If the area is 0.5, then it has 50% sensitivity and 50% specificity. If the area is close to 1, then the test is better. And if the area is closer to 0.5, then the test is worse. Since this model has an area close to 1 for the two categories of response variable, so our predicted model is accurate.

Table 9: Area under the curve

Variable	Category	Area
Are you engaged in any Paid labor at this time?	Yes	.960
	No	.960

Figure-4 shows the Cumulative Gains chart that provides the percentage of the overall number of cases in a given category "gained" by targeting a percentage of the total number of cases. Cumulative



ISSN: 2223-9553
Volume 1, Issue 3, November 2011

Figure 4: Gain Chart

gains are used to predict the model performance. It contains a lift curve and a baseline. The model is considered better when the area between the lift curve and the baseline is greater. The first point on the curve for the 1^{st} category is at (10%, 55%), meaning that if we score a dataset with the network and sort all the cases by predicted pseudo-probability of 1^{st} category, it should expect the top 10% to contain approximately 55% of all the cases that actually take the category 1^{st} . Likewise, the top 20% would contain approximately 89% of the category 1^{st} the top 30% would contain approximately 91% of the category 1^{st} and so on. If the first point covers more percentage of randomly selected cases then more percentage of correct classification can be obtained. Similarly percentages for other categories can interpret.

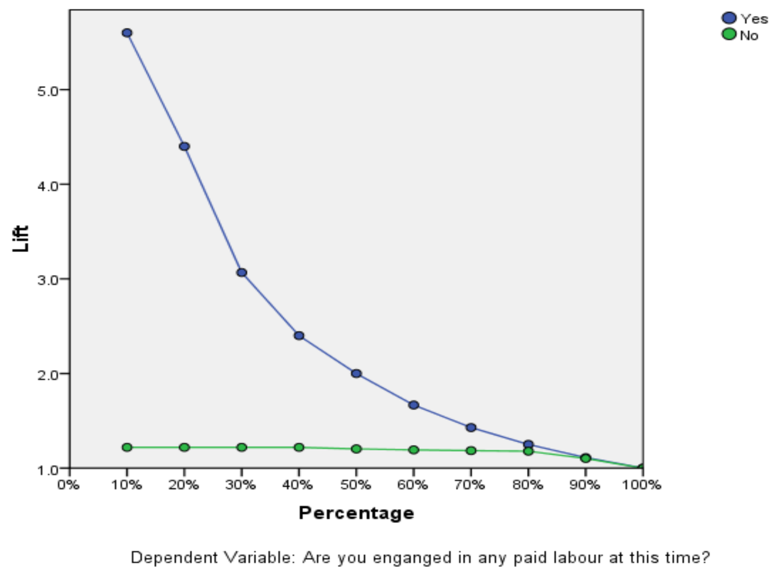


Figure 5: Lift Chart

Figure-5 shows Lift Chart for two categories of dependent variable. Lift is a measure of the efficiency of an analytical model calculated as the ratio between the results obtained with and without the predictive model. The lift chart is resultant of the cumulative gain chart; the values on the y-axis correspond to the ratio of the cumulative gain for each curve to the baseline. Thus the lift at 10% for the category 1^{st} is 55%/10% = 5.5. Similarly, the lift at 20% for the category 1^{st} is 4.4; the lift at 30% for the category 1^{st} is 3.0 and so on. Similarly the lift chart for other categories can interpret.

Table 10:Independent Variable Importance

Independent Variable	Importance	Normalized Importance
Literacy Status	.190	74.4%
Area of Residence	.189	74.0%
Family System	.124	48.6%
Family Size	.256	100.0%
Education that your husband has completed	.241	94.1%

Table-10 and Figure-6 show the independent variables importance. The independent variable Family size has 100% normalized importance, Education that your husband has completed has 94.1%, Literacy Status has 74.4%, Area of Residence has 74.0%, and Family System has 48.6% normalized importance. Different studies like Ejaz and Khan (2009), Kohara (2008), Ferdous (2006), Naqvi and Shehbaz (2002) and Azid (2001) also identified the same factors and determinants.

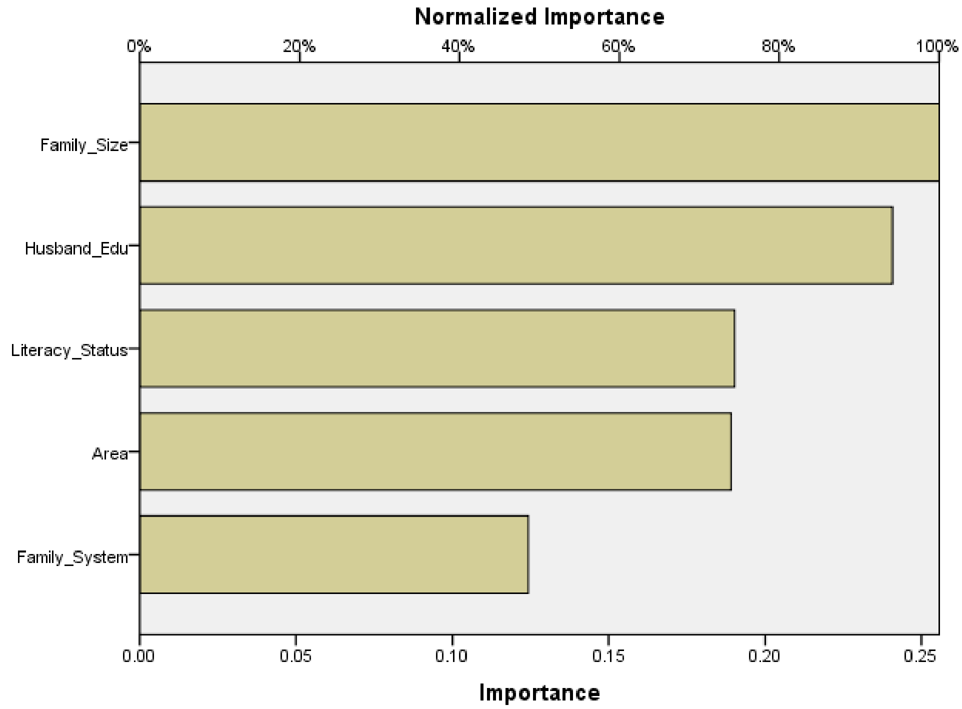


Figure 6: Normalized Importance

Percentage given in table No. 5 highlighted the importance and significance of variables in the process of predicting the women labor force participation in the job market. It indicates that higher the percentage higher will be the importance in prediction process.

CONCLUSION

Women's role in the economy of a country cannot be denied. But In Pakistani society Female labor force participation is very low as compared to other countries. So it may be hindrance in the growth of Pakistan. The aim of this study is to find out determinants or factors which effect the female labor force participation. According the results of the study it can be said that literacy status, area of residence (rural, urban), family system, family size and husband education are the major factors which effect and on which basis one can predict the female labor force participation.

REFERENCES

Azid, T., Aslam, M. and Chaudary, M. 2001. "Poverty, Female Labor Force Participation,and Cottage Industry: A case study of cloth embroidery in Rural Multan" *The Pakistan Development Review*, 40(4), 1105-1118.

Ejaz, R and Khan, T. 2009. "Labor Force Participation of Married Women in Punjab (Pakistan)". *Journal of Economic and Social Research*, 11(2), 77-106

Ejaz, M 2007. "Determinants of female labor force participation in Pakistan An empirical analysis of PSLM (2004-2005) Microdata". *The Lahore journal of economics, SI, 203-235*

Ferdoos, A. 2006. “Social Status of Rural and Urban Working Women in Pakistan - A Comparative Study” PhD *Dissertation.*

Kohara, M. 2008. “*The Response of Wives’ Labor Supply to Husbands’ Job Loss*” OSIPP Discussion Paper: DP-2008-E-007

Lisaniler J F and Bhatti F 2001. Determinants of female labor force participation: A study of north Cyprus. *Review of Social, Economic and Business Studies, 5(6), 209-226*

Sajid R.M, Maqsood F and Rani M ,2011. Determinants of Fertility: A Neural Network Approach. *Middel east Journal of Scientific research 8(2):440-449*

The World Bank (2002). *The World Bank Annual Report.*