A DETAILED ANALYSIS ON DIGITAL BANKING USAGE

Shramana Bhattacharya

MSc. Biostatistics, International Institute for Population Sciences, Mumbai (Intern at United Bank of India, Lake Gardens branch)

Abstract

With the advent of technology, banking has become easier, faster and hassle free. Banks provide digital facilities such as internet banking and others, like sending account related messages to mobile phones of customers (mobile banking facilities). This project is done by collecting data from the United Bank of India (UBI), Lake Gardens branch. Various kinds of people open accounts in banks. Basic information regarding the account holders are collected by banks to keep record. Certain demographic variables like age, gender, marital status, profession, education, income, etc. might affect the choice of facilities by the customers that are provided by the bank. The purpose of this report is to statistically analyse whether there is any impact of demographic variables like gender, profession, marital status and age on choice of opting for digital banking. The methods used are Chi-Square test of Independence, Odds Ratio and Binary Logistic Regression. In this study, Chi-Square test for independence showed association between certain demographic variables (independent variables) and dependent variable (internet banking/mobile banking facility) and for those variables, odds ratio explains the measure of association between them. By using binary logistic regression, a customer's behaviour of whether he/she would use internet banking or mobile banking facility provided to them by UBI, Lake Gardens branch; based on their demographic information, have been predicted.

Keywords : internet banking, mobile banking facility, demographic variables, Chi-square test for independence, Odds Ratio, Binary Logistic Regression

1. Introduction

Technology has touched every aspect of our lives. Online banking, also known as net banking and e-banking, is the facility provided by banks and financial institutions which allows customers to use banking services via internet. There are scores of services like online money transfer, account opening, bill payment, tracking account activity, etc., which are made available to customers with the help of online banking**.** Many banks have mobile banking facility which provides the customers, services such as information relating to Account/s, details about transactions and such other services as may be provided on the mobile phone number by bank, from time to time. With the rapid development of internet, internet banking and use of mobile banking facilities have become an important phenomenon. Various studies have been performed to increase the diffusion of internet banking. There are various advantages of net banking. It saves time spent in banks, provides scope of international banking, provides banking throughout the year 24/7 days from any place having internet access, etc. It has, hence, become important to analyze whether these aforementioned demographic variables affect usage of such net banking or mobile banking facilities. Use of these facilities, help banks to be more efficient and user friendly for the customers. On the other hand, banks can collect more data and use them effectively and aptly for the betterment of their services if they can attract more customers on using such technological facilities. This study aims at understanding whether there exists any association between gender, profession, marital status, age and usage/ non usage of such facilities. If there exists an association, then what is the measure of such an association; for example, if an association between gender and internet banking exists, then which gender (male/female) is more inclined to use internet banking or mobile banking facility. Here, a logistic regression model is established which can help to predict the behaviour of any future customer regarding his/her choice of availing such facilities. In this project, data from UBI, Lake Gardens branch on gender, profession, marital status, age usage/non usage of internet banking and opting/not opting for mobile banking facilities like receiving SMS; have been taken.

2. Methodology

A sample of 200 savings account holders have been taken and their demographic details as mentioned before are recorded. Among them, 28.50 % were internet banking users and 71.50% were non users, 95% opted for mobile banking facilities and only 5% did not. 46% of them are females and 54% are males.

Ages between 18-45 have been combined into group named "Young" and the ages 46 and above into another group named "Elderly". There are 37% young and 63% elderly in the sample.

For variable profession, categories business, public sector, private sector and self-employed have been combined to form group "Employed" and the rest (Students, unemployed, retired, homemaker) have been considered into the group "Unemployed". 51.50% are employed and the remaining are unemployed.

For variable marital status, only married people are considered in group "Married" and the rest (Single, divorced, widow/widower) are included in the group "Others". 60.50% are married.

2.1. Chi-Square Test for independence

The Chi-Square test for independence is used to determine if there is a significant relationship between two nominal (categorical) variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable.

The following hypotheses have been considered as the null hypotheses.

1. a. H1: Gender and internet banking usage have no significant relationship.

b. H2: Gender and mobile banking facility usage have no significant relationship.

2. a. H3: Profession and internet banking usage have no significant relationship.

b. H4: Profession and mobile banking facility usage have no significant relationship.

3. a. H5: Marital status and internet banking usage have no significant relationship.

b. H6: Marital status and mobile banking facility usage have no significant relationship.

4. a. H7: Age and internet banking usage have no significant relationship.

b.H8: Age and mobile banking facility usage have no significant relationship.

Level of significance : 0.05

If the p value obtained after testing the null hypothesis is greater than 0.05, then we accept the null hypothesis. Otherwise, we reject the null hypothesis and accept the alternative hypothesis.

Table 1:

TEST	P Value	Decision
1a.	0.485	Accept H1. They have no significant relationship.
1b.	0.027	Reject H2. They have significant relationship.
2a.	0.017	Reject H3. They have significant relationship.
2b.	0.455	Accept H4. They have no significant relationship.
3a.	0.038	Reject H5. They have significant relationship.
3b.	0.528	Accept H6. They have no significant relationship.
4a.	0.000	Reject H7. They have significant relationship.
4b.	0.840	Accept H8. They have no significant relationship.

2.2. Odds Ratio

Odds ratio is a statistic that quantifies the strength of association between two events, A and B, say. The odds ratio is defined as the ratio of the odds of A in the presence of B and the odds of A in the absence of B, or equivalently (due to symmetry), the ratio of the odds of B in the presence

of A and the odds of B in the absence of A. Two events are independent if and only if the OR equals 1: the odds of one event are the same in either the presence or absence of the other event. If the OR is greater than 1, then A and B are associated (correlated) in the sense that, compared to the absence of B, the presence of B raises the odds of A, and symmetrically the presence of A raises the odds of B. Conversely, if the OR is less than 1, then A and B are negatively correlated, and the presence of one event reduces the odds of the other event.

We know, odds ratio is a measure of association and chi-square test is a statistical test of association between two categorical variables.

Chi-square test showed no significant association between gender and internet banking. It also showed no significant association for variables profession, marital status and age in case of mobile banking facility. Therefore, odds ratio is calculated for profession, marital status and age in case of internet banking usage and only for gender in case of mobile banking facility usage.

Table 2: *Odds Ratio for impact of demographic variables on internet banking usage*

Odds Ratio of -	**Odds Ratio**	**Interpretation**
Profession	2.16	Employed are 2.16 times more likely to use internet banking than unemployed.
Marital Status	1.93	Others are 1.93 times more likely to use internet banking than married.
Age	6.30	Young are 6.30 times more likely to use internet banking than elderly.

Table 3: *Odds Ratio for impact of gender on mobile banking facility usage*

Odds Ratio of -	Odds Ratio	Interpretation
Gender	5.05	Males are 5.05 times more likely to use mobile banking facility than females.

2.3. Logistic Regression

Binary logistic regression has been used to analyse the impact of gender, profession, marital status and age on usage of internet banking and mobile banking facility respectively. The dependent variable is a dichotomous variable coded as 0 -- non user of internet banking/ non user of mobile banking facility and 1 -- user of internet banking/user of mobile banking facility. The independent variables are demographic variables i.e gender, profession, marital status and age which are coded as 0 and 1 into two groups as mentioned before.

We consider our level of significance to be 0.05. If p value is less than 0.05, then our model is significant, otherwise the model is not significant.

This model is used to assess the effect of these demographic variables on the dependent variable and hence we can get an idea regarding their impact. It also predicts the dependent variable for a set of values given for the independent variables.

The binary logistic regression is performed by dividing the data set into two parts: train and test. Most of the data is used for training and a smaller portion is used for testing. Training set is used to process the logistic regression model and then the model is tested by making predictions against the test set. The last 12 points have been kept for testing set and the remaining are considered for training.

2.3.1. Internet Banking Usage

P value: 0.0000 < 0.05

Hence, we consider that the model is significant.

2.3.2. Mobile banking facility usage

P value: 0.2276 > 0.05

Hence, we consider that the model is not significant. So, I will not further use this for prediction.

2.3.3. Prediction

Now, predictions with the logistic regression model in case of internet banking usage are obtained.

188 data points are in the training set and 12 are in testing set. The two groups for each independent demographic variable are as follows:

Gender: Male & Female

Profession: Employed and Unemployed

Marital Status: Married and Others

Age: Young and Elderly

After performing prediction on the test set, accuracy of logistic regression classifier on test set is 0.83.

3. Limitations

The results obtained are purely based on the sample of 200 data which has been collected from the bank records. Errors due to sampling are present in this report since the whole population of all savings account holders is not considered. Moreover, data on only savings account records have been analysed.

4. Conclusion

This project studies if there is any impact of demographic variables gender, profession, marital status and age on choice of opting for digital banking using Chi-square test, Odds Ratio and Logistic Regression based on the sample of 200 savings account holders of UBI, Lake Gardens branch.

The results of Chi-square test show that profession, marital status and age are significantly associated with internet banking usage while gender is not. On the other hand, only gender is associated significantly with mobile banking facility usage.

Odds Ratio is calculated for each independent variable having significant association with the dependent variable.

For internet banking, employed people are 2.16 times more inclined to internet banking than those who are students, unemployed, retired or homemaker.

Single, widow/widower and divorced people (others category) are 1.93 times more likely to use internet banking than those who are married. Young people aged between 18-45 are 6.30 times more using internet banking than those who are elderly, aged between 46-70.

For mobile banking facility, males are 5.05 times more likely to use mobile banking facility than females.

Next, Binary Logistic Regression results are considered.

Binary logistic regression model is significant for internet banking usage while the model is not significant for mobile banking facility usage.

Prediction has been obtained for the test set in case of internet banking usage with an accuracy of 83 %. By using this model, we will be able to predict a person's behaviour on using/ not using internet banking based on the values of given demographic variables.

5. Appendix: Python Code of Binary Logistic Regression for internet banking usage

```
import pandas as pd

from sklearn import preprocessing

from sklearn.linear_model import LogisticRegression

from sklearn.model_selection import train_test_split


data=pd.read_excel(r'C:\Users\DELL\OneDrive\Desktop\UBI\pythondatagrouped.xlsx')


X=data.iloc[0:200, 0:4]

y=data.iloc[0:200, 4]


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.06, random_state=0)

logreg = LogisticRegression()

logreg.fit(X_train, y_train)


y_pred = logreg.predict(X_test)


print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))
```

6. References

1. Berna S. Statistical Analysis of Internet Banking Usage with logistic regression
Procedia Computer Science 2016; **102**: *648 – 653*

2. Celik H. What determines Turkish customers' acceptance of internet banking? *International Journal of Bank Marketing* 2008; **26**:353-370.

3. Zhao AL, Lewis NK, Lloyd SH, Ward P. Adoption of internet banking services in China? Is it all about trust? *International Journal of Bank Marketing* 2010; **28**: 7-26.

4. Chao-Ying J.P., Kuk L.L., Gary M. I. An Introduction to Logistic Regression Analysis and Reporting

5. Serener B. Effects of demographic characteristics and internet usage on risk perceptions of internet banking in Northern Cyprus. Paper presented at the 5th World Conference on Business, Economics and Management 12-14 May 2016, Kemer, Antalya, Turkey.

6. Kolodinsky J M, Hogarth J M, Hilgert MA. The adoption of electronic banking technologies by US consumers. *International Journal of Bank Marketing* 2004; **22**:238-259.