# Chapter 3. Model estimation.

## 3.1 Choice between logit and probit

To accurately capture the determinants of the dummy dependent variable the most used approach is the logistic regression, where we can either choose the probit or the logit model. Both models estimate the probability of success and failure (two levels of the dependent variable) as a function of the independent variables, and both models capture the nonlinear relationship between the predictors and the outcome. Their key difference is that logit uses the logistic sigmoid function as a link between linear predictor to the probability, while the probit uses cumulative normal distribution function. Their tails of the distribution also differ as the logit has heavier tails, which makes it more robust to the outliers or extreme values in the predictors. On Plot 1 the difference between the model functions is shown.

Plot 1. Functions used for both models to capture the linear predictor to the probability.

A graph with a line

Description automatically generated

For most scenarios the probit and logit model tend to fit data equally well, but the logit works better when there are extreme values or outliers in the independent variables. Hahn and Soyer in their paper[[1]](#footnote-1) suggested that to properly select between the two approaches, one should use the information criteria like AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). The steps are to evaluate both general models for probit and logit to choose one with the lower criterion value.

The general model can be represented as:

Where Churn is the dependent variable,  is the intercept, and the rest of are the coefficients of independent variables. Those dependent variables represent the model without yet created dummy variables from the categorical columns. For logit we then transform the linear predictor η using logistic link function, and for probit η is transformed using the cumulative distribution function. We can evaluate both models and determine which has smaller AIC. For future variable selection we changed the qualitative variables into binary form, treating one level of each as a base level. Table 1 represents both model estimation with their corresponding information criteria.

Table 1. Comparison of logit and probit general models

=======================================================================

Dependent variable:

----------------------------

Churn

logistic probit

Logit Model Probit Model

(1) (2)

-----------------------------------------------------------------------

Tenure -0.226\*\*\* -0.110\*\*\*

(0.011) (0.005)

WarehouseToHome 0.027\*\*\* 0.014\*\*\*

(0.005) (0.003)

Gender 0.155 0.077

(0.125) (0.067)

HourSpendOnApp -0.027 -0.001

(0.055) (0.030)

NumberOfDeviceRegistered 0.709\*\*\* 0.367\*\*\*

(0.130) (0.069)

SatisfactionScore 0.604\*\*\* 0.306\*\*\*

(0.143) (0.076)

NumberOfAddress 0.243\*\*\* 0.130\*\*\*

(0.019) (0.010)

Complain 1.550\*\*\* 0.819\*\*\*

(0.153) (0.083)

OrderAmountHikeFromlastYear -0.005 -0.0003

(0.012) (0.006)

CouponUsed 0.063\* 0.033\*

(0.034) (0.018)

OrderCount 0.052 0.038\*\*

(0.033) (0.017)

DaySinceLastOrder -0.127\*\*\* -0.070\*\*\*

(0.023) (0.012)

CashbackAmount -0.014\*\*\* -0.008\*\*\*

(0.003) (0.001)

LoginDevice\_MPhone -0.416\*\*\* -0.196\*\*\*

(0.114) (0.062)

LoginDevice\_Phone -0.468\*\*\* -0.207\*\*\*

(0.129) (0.072)

PaymentMode\_CCard -0.691\*\*\* -0.438\*\*\*

(0.166) (0.090)

PaymentMode\_DCard -0.499\*\*\* -0.367\*\*\*

(0.161) (0.087)

PaymentMode\_EWallet 0.087 -0.048

(0.208) (0.113)

PaymentMode\_UPI -0.817\*\*\* -0.520\*\*\*

(0.232) (0.127)

CityTier2 0.841\*\*\* 0.403\*\*\*

(0.234) (0.129)

CityTier3 0.661\*\*\* 0.348\*\*\*

(0.122) (0.066)

OrderCat\_Grocery 0.620\* 0.232

(0.363) (0.188)

OrderCat\_Laptop -1.706\*\*\* -0.952\*\*\*

(0.191) (0.101)

OrderCat\_Mobile -0.831\*\*\* -0.511\*\*\*

(0.229) (0.123)

OrderCat\_Other 2.312\*\*\* 1.161\*\*\*

(0.447) (0.234)

Martial\_Divorced -0.726\*\*\* -0.414\*\*\*

(0.140) (0.076)

Martial\_Married -1.027\*\*\* -0.571\*\*\*

(0.102) (0.055)

Gender:Complain 0.475\*\* 0.264\*\*

(0.194) (0.106)

OrderCount:DaySinceLastOrder 0.010\*\*\* 0.005\*\*\*

(0.003) (0.002)

NumberOfDeviceRegistered:SatisfactionScore -0.086\*\* -0.044\*\*

(0.036) (0.019)

Constant -1.588\*\* -0.668\*

(0.734) (0.393)

-----------------------------------------------------------------------

Observations 5,630 5,630

Log Likelihood -1,541.030 -1,576.153

Akaike Inf. Crit. 3,144.059 3,214.307

=======================================================================

We can clearly see that the AIC for the logit model is lower than the one calculated for probit. Both are relatively high (3144 and more), which indicates a rather poor fit of the models to the data. By ranking the models from the lowest AIC to highest, we will choose to proceed with the analysis for the logit model, as it shows a slightly better fit.

One of the logit model properties is that the coefficients represent the change in the log odds of the outcome for a one unit increase in the predictor variable holding other variables constant. The coefficients indicate only the direction of the relationship between predictors and marginal effects are used to quantify the actual influence of change in the probability of success. In our general model, we would be able to say, that being a Male increases the probability of Churn and nothing more for now. Also, logit model expresses the relationship between the dependent variable and the independent variable as odds[[2]](#footnote-2). The logit is a natural logarithm of the odds:

So, the model can be written as:

## 3.2 Testing general and null model

To check whether the general model is statistically better than the nested model – here is a null model where we estimate the outcome only with an intercept – we can perform a likelihood ratio test (LRT). The null hypothesis of the test states that the simpler model provides an adequate fit to the data, or that adding additional parameters to the model won’t improve its quality. The results of the likelihood ratio test are shown in Table 2.

Table 2. Likelihood ratio test for general and null model

=========================================

LogLik Df Chisq Pr(> Chisq)

-----------------------------------------

1 31 -1,541.030

2 1 -2,552.158 -30 2,022.257 0

As the p-value of the test is 0, we can reject the null hypothesis, and state that the general model is better than the null model only with an intercept.

## 3.3 General to specific variable selection

Knowing that the general model is of better quality than the null one, we can now focus on improving its quality further, by proper variable selection. First, we need to ensure that all insignificant variables in the model are jointly significant. For this purpose, we will evaluate a model with Churn as a dependent variable and all the insignificant variables will be predictors. Then we compare two models using ANOVA test, with null hypothesis stating that the simpler, reduced model fits the data as well as the general model. The p-value equal to 0 provides strong evidence to reject the null hypothesis, which means we will need to take a general to specific approach to reduce the number of variables, keeping the quality of the model. Such approach consists of multiple steps:

1. We find the independent variable with the highest p-value,
2. We evaluate the general model without the chosen variable,
3. We use ANOVA test to determine whether the reduced model is as good as the general model,
4. We continue this process until there are no insignificant variables in the model.

Table 3 Shows the comparison of the general model and the final model.

Table 3. General and Final model comparison

=======================================================================

Dependent variable:

----------------------------

Churn

General Model Final Model

(1) (2)

-----------------------------------------------------------------------

Tenure -0.226\*\*\* -0.225\*\*\*

(0.011) (0.011)

WarehouseToHome 0.027\*\*\* 0.027\*\*\*

(0.005) (0.005)

Gender 0.155 0.149

(0.125) (0.125)

HourSpendOnApp -0.027

(0.055)

NumberOfDeviceRegistered 0.709\*\*\* 0.700\*\*\*

(0.130) (0.129)

SatisfactionScore 0.604\*\*\* 0.608\*\*\*

(0.143) (0.143)

NumberOfAddress 0.243\*\*\* 0.241\*\*\*

(0.019) (0.019)

Complain 1.550\*\*\* 1.541\*\*\*

(0.153) (0.153)

OrderAmountHikeFromlastYear -0.005

(0.012)

CouponUsed 0.063\*

(0.034)

OrderCount 0.052 0.085\*\*\*

(0.033) (0.028)

Complain:Gender 0.475\*\* 0.477\*\*

(0.193)

DaySinceLastOrder:OrderCount 0.010\*\*\*

(0.003)

DaySinceLastOrder -0.127\*\*\* -0.124\*\*\*

(0.023) (0.023)

CashbackAmount -0.014\*\*\* -0.012\*\*\*

(0.003) (0.003)

LoginDevice\_MPhone -0.416\*\*\* -0.436\*\*\*

(0.114) (0.113)

LoginDevice\_Phone -0.468\*\*\* -0.454\*\*\*

(0.129) (0.128)

PaymentMode\_CCard -0.691\*\*\* -0.734\*\*\*

(0.166) (0.137)

PaymentMode\_DCard -0.499\*\*\* -0.542\*\*\*

(0.161) (0.128)

PaymentMode\_EWallet 0.087

(0.208)

PaymentMode\_UPI -0.817\*\*\* -0.851\*\*\*

(0.232) (0.212)

CityTier2 0.841\*\*\* 0.839\*\*\*

(0.234) (0.233)

CityTier3 0.661\*\*\* 0.677\*\*\*

(0.122) (0.115)

OrderCat\_Grocery 0.620\*

(0.363)

OrderCat\_Laptop -1.706\*\*\* -1.705\*\*\*

(0.191) (0.189)

OrderCat\_Mobile -0.831\*\*\* -0.790\*\*\*

(0.229) (0.225)

OrderCat\_Other 2.312\*\*\* 2.065\*\*\*

(0.447) (0.393)

Martial\_Divorced -0.726\*\*\* -0.727\*\*\*

(0.140) (0.140)

Martial\_Married -1.027\*\*\* -1.028\*\*\*

(0.102) (0.102)

Gender:Complain 0.475\*\*

(0.194)

OrderCount:DaySinceLastOrder 0.010\*\*\*

(0.003)

NumberOfDeviceRegistered:SatisfactionScore -0.086\*\* -0.088\*\*

(0.036) (0.036)

Constant -1.588\*\* -1.930\*\*\*

(0.734) (0.699)

-----------------------------------------------------------------------

Observations 5,630 5,630

Log Likelihood -1,541.030 -1,544.271

Akaike Inf. Crit. 3,144.059 3,140.543

=======================================================================

The final improved the information criteria to 3140, which means the final model better fits the data. It now consists of 13 dummy variables, 7 continuous variables and 3 interactions. The approach helped remove irrelevant variables that could lead to overfitting, while still capturing key drivers of the outcome variable.

## 3.4 Model Evaluation

To be certain of the goodness of the model, we must evaluate it based on the following criteria:

1. Pseudo R-Squared – how good the model fits the data (Tjur, McKelvey-Zavoina, Count R-squared, Adjusted Count R-squared),
2. Goodness of fit tests - how good the model fits the data (Hosmer-Lemeshow, Osius-Rojek test),
3. Link test – evaluates the specification of the link function in a regression model,
4. Wald test – assesses the significance of individual predictors in a regression model.

The tests will help us understand how well the specified form of the model fits the data, and the importance of individual predictors in the model. Knowing the AIC for the model is high, we expect that the tests will also reflect the poor fit of the data to the model.

### 3.4.1 Pseudo R-Squared

The logistic regression models don’t have a true R-squared measure because they model the log odds of a binary outcome as a linear combination of the predictors, not directly modeling the outcome variable itself. In linear regression R-squared represents a proportion of variance in the outcome explained by the predictors, which can’t be directly achieved in logistic regression. Also, the logistic model assumes a non-linear relationship between the predictors and the output probability, which violates the linear assumption of the R-squared calculation. The pseudo R-squared statistics, based on the log-likelihood function have been proposed as approximations to quantify the goodness of fit of logistic regression models. All the calculated R-squared measures are shown in Table 4.

Table 4. Different pseudo R-squared statistics

|  |  |  |  |
| --- | --- | --- | --- |
| Tjur | McKelveyZavoina | Count R2 | Adj Count R2 |
| 0.4120765 | 0.6519568 | 0.8912966 | 0.3544304 |

Many pseudo R-squared statistics are not interpretable, but the one listed above has a logical interpretation. The first one shows how well the model discriminates between success and failure, were the value Tjur R-square = 0.412 indicates that the discrimination is not very sufficient.

McKelvey-Zavoina can be interpreted similarly to the R-squared in linear regression models, which indicate the proportion of variance in the latent variable that the model explains. With the value 0.652 we can conclude that the model explains over 65% of the variance in the unobserved tendency of a client to churn.

Count R-squared assesses the accuracy of predictions made by the model in terms of counts or frequency of events. High value of this statistic – 0.891 indicates that the model predicts the observed count well, but it is very likely due to the fact of unbalanced data, where success event occurs only around 17% of the time.

To counteract this imbalance, we can calculate Adjusted Count R-squared. It adjusts the Count R-squared by considering the most frequent outcome to provide a more accurate assessment of the models’ predictive performance. As we can see it dropped significantly to 0.354, therefore we can conclude that the model struggles to predict the success event correctly and creates many false positives.

### 3.4.2 Goodness of fit

A goodness of fit (GOF) test evaluates how accurately a model represents a set of observations. It measures the difference between observed values and the values predicted by the model, summarizing the overall fit.[[3]](#footnote-3) All GOF tests state a null hypothesis that the fitted model is correct in all aspects, so it fits the data well. We performed two different tests to measure the goodness of fit. Their results can be seen in Table 5.

Table 5. Goodness of fit tests results

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Statistic | Value | p-value |
| Hosmer Lemeshow | Chi Square | 151.73 | 0.00000 |
| Osius Rojek | Z | 1.385 | 0.1658 |

The first one in use is Hosmer-Lemeshow which is commonly used due to its simplicity and it is useful for detecting overall lack of fit. It is sensitive to misspecification like incorrect link function, lack of significant predictor interactions or incorrect distribution assumption. Our calculation resulted in a p-value of the test under the significant level, so the null hypothesis can be rejected. The results show that the model lack fit to the data.

Osius Rojekt test tries to assess the fit of the model by comparing the observed and expected frequencies across all possible covariate patterns[[4]](#footnote-4). The test is more powerful than the Hosmer-Lemeshow in determining the misspecifications of the model. A p-value higher than the significance level mean we fail to reject the null hypothesis stating that the model fits the data well. It suggests that the data is not significantly different from what would be expected under the null hypothesis, and that there is 16.58% probability of obtaining a test statistic at least as extreme as the one calculated from the observed data.

### 3.4.3 Link Test

Link test is a diagnostic tool used to check for model specification errors by comparing the predicted values and their squared terms. Its null hypothesis states that the model is properly specified, and the used function is adequate to the specificity of the data. Table 6 shows the output of the link test.

Table 6. Link test results

==========================================================

Estimate Std. Error z value Pr(> Chisq)

----------------------------------------------------------

(Intercept) -0.01025 0.05724 -0.179 0.858

yhat 1.22778 0.04143 29.63 0.000

yhat2 0.07399 0.00607 12.19 0.000

----------------------------------------------------------

Both yhat and yhat2 are statistically significant, which means there is no need to include or omit any other variable from the model. It also indicates that the predicted yhat is very identical to the real y (dependent variable) values.

### 3.4.4 Wald Test

Wald test evaluates whether one or more coefficients in a model are significantly different from a hypothesized value, typically zero. It is usually used to measure if some independent variables contribute significantly to the model’s predictive ability[[5]](#footnote-5). The test was performed as a comparison of a final model, with a final model with one insignificant variable added, that was removed during the general to specific variable selection. The results of the test are shown in Table 7.

Table 7. Wald test results (1 – final model, 2 – model with the omitted variable)

=============================================

Res.Df Df F Pr(> F)

---------------------------------------------

1 5604

2 5603 1 0.5199 0.4709

---------------------------------------------

With the p-value more than the significance level, we fail to reject the null hypothesis, and can conclude that the coefficients of the omitted variables are significantly different from zero.

# Chapter 4. Results and findings

## Hypothesis testing

The logit model analysis provided valuable insights into the factors that drive customer churn in the ecommerce industry. The model’s results the significant customer attributes and order data and quantified their respective impacts into insights for customer retention strategies. Based on the tests results in previous chapter, we can reject the null hypothesis which stated that “Client Churn is not determined by individual characteristics of the client”. Significance of variables in the final model, as well as the proper link function used in the model, and a well fitted model for the data let us state that individual characteristics of the client are in fact significant insights into predicting clients Churn. Table 8 provides all significant predictors of client churn.

Table 8. Model with all significant variables

======================================================================

Dependent variable:

---------------------------

Churn

----------------------------------------------------------------------

Tenure -0.225\*\*\*

(0.011)

WarehouseToHome 0.027\*\*\*

(0.005)

NumberOfDeviceRegistered 0.700\*\*\*

(0.129)

SatisfactionScore 0.608\*\*\*

(0.143)

NumberOfAddress 0.241\*\*\*

(0.019)

Complain 1.541\*\*\*

(0.153)

DaySinceLastOrder -0.124\*\*\*

(0.023)

CashbackAmount -0.012\*\*\*

(0.003)

LoginDevice\_MPhone -0.436\*\*\*

(0.113)

LoginDevice\_Phone -0.454\*\*\*

(0.128)

PaymentMode\_CCard -0.734\*\*\*

(0.137)

PaymentMode\_DCard -0.542\*\*\*

(0.128)

PaymentMode\_UPI -0.851\*\*\*

(0.212)

CityTier2 0.839\*\*\*

(0.233)

CityTier3 0.677\*\*\*

(0.115)

OrderCat\_Laptop -1.705\*\*\*

(0.189)

OrderCat\_Mobile -0.790\*\*\*

(0.225)

OrderCat\_Other 2.065\*\*\*

(0.393)

Martial\_Divorced -0.727\*\*\*

(0.140)

Martial\_Married -1.028\*\*\*

(0.102)

OrderCount 0.085\*\*\*

(0.028)

Complain:Gender 0.477\*\*

(0.193)

DaySinceLastOrder:OrderCount 0.010\*\*\*

(0.003)

NumberOfDeviceRegistered:SatisfactionScore -0.088\*\*

(0.036)

Constant -1.930\*\*\*

(0.699)

----------------------------------------------------------------------

Observations 5,630

Log Likelihood -1,544.271

Akaike Inf. Crit. 3,140.543

======================================================================

What can be interpreted from the table is only the direction of effect each variable has. For example, a positive and statistically significant coefficient of 0.839 for the CityTier2 variable indicates that when a customer lives in a city with second tier, he is more likely to churn, holding all other variables constant.

## 4.2 Marginal Effects

### Marginal Effects for average characteristics

Marginal effects for average characteristics measure a change in the outcome variable after a one-unit increase of a given independent variable. It is an intuitive way to interpret the effects of variables in nonlinear models like logit. The results of the marginal effects for average characteristics are presented in Table 9.

Table 9. Final model marginal effects for average characteristics

dF/dx Std. Err. z P>|z|

Tenure -0.01334864 0.00065422 -20.4040 < 2.2e-16 \*\*\*

WarehouseToHome 0.00172198 0.00035501 4.8506 1.231e-06 \*\*\*

NumberOfDeviceRegistered 0.04459504 0.00830703 5.3683 7.946e-08 \*\*\*

SatisfactionScore 0.03735082 0.00922481 4.0490 5.145e-05 \*\*\*

NumberOfAddress 0.01583138 0.00132828 11.9187 < 2.2e-16 \*\*\*

Complain 0.13034870 0.01702175 7.6578 1.892e-14 \*\*\*

DaySinceLastOrder -0.00815673 0.00145587 -5.6027 2.111e-08 \*\*\*

CashbackAmount -0.00088329 0.00016373 -5.3948 6.860e-08 \*\*\*

LoginDevice\_MPhone -0.02508323 0.00745394 -3.3651 0.0007652 \*\*\*

LoginDevice\_Phone -0.02279651 0.00720861 -3.1624 0.0015647 \*\*

PaymentMode\_CCard -0.04552051 0.00752592 -6.0485 1.462e-09 \*\*\*

PaymentMode\_DCard -0.04030521 0.00803362 -5.0171 5.247e-07 \*\*\*

PaymentMode\_UPI -0.04341924 0.00716754 -6.0578 1.380e-09 \*\*\*

CityTier2 0.06332043 0.02563457 2.4701 0.0135068 \*

CityTier3 0.04539262 0.00929947 4.8812 1.054e-06 \*\*\*

OrderCat\_Laptop -0.09954252 0.01007081 -9.8843 < 2.2e-16 \*\*\*

OrderCat\_Mobile -0.05418730 0.01243931 -4.3561 1.324e-05 \*\*\*

OrderCat\_Other 0.23108282 0.06628527 3.4862 0.0004900 \*\*\*

Martial\_Divorced -0.04005465 0.00617933 -6.4820 9.049e-11 \*\*\*

Martial\_Married -0.07254368 0.00782915 -9.2658 < 2.2e-16 \*\*\*

OrderCount 0.00663156 0.00179257 3.6995 0.0002161 \*\*\*

Complain:Gender 0.03712860 0.01676541 2.2146 0.0267878 \*

DaySinceLastOrder:OrderCount 0.00055291 0.00022252 2.4847 0.0129649 \*

DeviceRegistered:SatScore -0.00539421 0.00230959 -2.3356 0.0195137 \*

As there is plenty of results to interpret I will only highlight some. For example when a customer with average characteristics has complained his probability of churn increases by 13 percentage points which is the highest effect any variable has and which seems intuitive. Also when the same client’s tenure increases by one unit his probability of churn decreases by 1 percentage points.

### 4.2.2 Marginal Effects for user defined characteristics

Previously explained marginal effects were for all the independent variables on the level of average value of that variable. The same effects can be calculated for any other possible mix of the variable levels, and will be interpreted in the exact same way but as marginal effects for given characteristics. Our calculated marginal effects can be viewed in Table 10.

Table 10. Final model marginal effects for user defined characteristics

Marginal effects at X=

Tenure -0.027028036 12

WarehouseToHome 0.003188164 15

NumberOfDeviceRegistered 0.084034855 3

SatisfactionScore 0.073044949 4

NumberOfAddress 0.029005094 2

Complain! 0.361203241 1

DaySinceLastOrder -0.014912818 8

CashbackAmount -0.001424882 50

LoginDevice\_MPhone! -0.074866204 1

LoginDevice\_Phone! -0.077552522 0

PaymentMode\_CCard! -0.115583609 1

PaymentMode\_DCard! -0.090241283 0

PaymentMode\_UPI! -0.129375932 0

CityTier2! 0.188473067 0

CityTier3! 0.148808516 1

OrderCat\_Laptop! -0.199728666 0

OrderCat\_Mobile! -0.122410101 1

OrderCat\_Other! 0.474864949 0

Martial\_Divorced! -0.114735871 0

Martial\_Married! -0.148228703 1

OrderCount 0.010197178 5

Complain:Gender 0.057314831 0

DaySinceLastOrder:OrderCount 0.001155346 40

DeviceRegistered:SatScore -0.010539221 12

Table shows that for example for client with 12 months using the service (tenure = 12), if he would use it for one more month the churn is less likely by 27 percentage points. Also a client who already ordered 5 times, his next order would increase the probability of churn by 1 percentage points.

### 4.2.3 Average Marginal Effects

Table 11. Final model average marginal effects

dF/dx Std. Err. z P>|z|

Tenure -0.01760792 0.00075804 -23.2283 < 2.2e-16 \*\*\*

WarehouseToHome 0.00227143 0.00045762 4.9636 6.921e-07 \*\*\*

NumberOfDeviceRegistered 0.05882442 0.01087280 5.4102 6.294e-08 \*\*\*

SatisfactionScore 0.04926871 0.01211086 4.0681 4.739e-05 \*\*\*

NumberOfAddress 0.02088286 0.00153924 13.5670 < 2.2e-16 \*\*\*

Complain 0.14902299 0.01644709 9.0607 < 2.2e-16 \*\*\*

DaySinceLastOrder -0.01075938 0.00188398 -5.7110 1.123e-08 \*\*\*

CashbackAmount -0.00116513 0.00021426 -5.4380 5.388e-08 \*\*\*

LoginDevice\_MPhone -0.03304347 0.00968342 -3.4124 0.0006440 \*\*\*

LoginDevice\_Phone -0.03173836 0.01055943 -3.0057 0.0026498 \*\*

PaymentMode\_CCard -0.06452035 0.01081820 -5.9641 2.460e-09 \*\*\*

PaymentMode\_DCard -0.05443717 0.01069026 -5.0922 3.539e-07 \*\*\*

PaymentMode\_UPI -0.06968507 0.01388398 -5.0191 5.191e-07 \*\*\*

CityTier2 0.07027926 0.02463199 2.8532 0.0043285 \*\*

CityTier3 0.05599503 0.01058140 5.2918 1.211e-07 \*\*\*

OrderCat\_Laptop -0.14167807 0.01350936 -10.4874 < 2.2e-16 \*\*\*

OrderCat\_Mobile -0.07456757 0.01756982 -4.2441 2.195e-05 \*\*\*

OrderCat\_Other 0.20370147 0.04559982 4.4672 7.927e-06 \*\*\*

Martial\_Divorced -0.06054504 0.01014321 -5.9690 2.387e-09 \*\*\*

Martial\_Married -0.09340771 0.00890035 -10.4948 < 2.2e-16 \*\*\*

OrderCount 0.00874756 0.00236349 3.7011 0.0002147 \*\*\*

Complain:Gender 0.04538606 0.01902041 2.3862 0.0170246 \*

DaySinceLastOrder:OrderCount 0.00072933 0.00029206 2.4972 0.0125187 \*

DeviceRegistered:SatScore -0.00711540 0.00305136 -2.3319 0.0197071 \*

The third way of how the marginal effects can be calculated is the average marginal effects. The measure the marginal effect averaged across all observations in the data, so the average effect of a change in the independent variable on a predicted probability of outcome. An interpretation for example is that on average a one unit increase in number in registered devices is associated with a 5.8 percentage points increase of the Y.

## 4.3 Findings

Churn predictions and its management is a very practical and useful tool for business in every industry. Therefore after building the model it is important to analyze the results in a business context. Some of the results shown in the marginal effects are obvious for the Ecommerce business, such as the fact that with higher tenure the client is less likely to churn, or that when he complains he probably won’t use the service again. Based on the results we can also construct a statement that users who order electronical devices (Laptop / Mobile category) are more probable to stay and use the service once more, which cant be said for those who buy from other categories. On average an user who bought product from “Others” will be 20 percentage points more likely to churn.

We can also interpret the results for payment methods. On average, users who pay in a digital way (Card, UPI) decrease the log-odds of churn compared to using cash (as cash is treated as a reference level). A conclusion might be provided, that in order to keep more clients it should encourage them to use less cash methods.

Lastly the studied ecommerce business should mostly focus on the Divorced and Married people, who show less probability to churn than base status – Single. The model also showed that variables such as hours that users spend on the app, or the number of coupons used does not influence the churn.

1. „Probit and Logit Models: Differences in the Multivariate Realm”. Eugene Hahn, Refik Soyer [↑](#footnote-ref-1)
2. „Logistic Regression: A Brief Primer”. Jill C. Stoltzfus [↑](#footnote-ref-2)
3. „Goodness-of-Tests for Logistic Regression”. Sutan Wu [↑](#footnote-ref-3)
4. „Goodness-of-fit for Logistic Regression: Simulation Results”. D.W. Hosmer, N.L. Hjort [↑](#footnote-ref-4)
5. <https://www.statlect.com/fundamentals-of-statistics/Wald-test> (access date: 22.05.2024) [↑](#footnote-ref-5)