

МОДЕЛИРОВАНИЕ ПОТРЕБИТЕЛЬСКОГО ВЫБОРА НА РЫНКЕ ЭЛЕКТРОЩИТОВОГО ОБОРУДОВАНИЯ С ПОМОЩЬЮ IBM SPSS STATISTICS

УДК 338.27

Сергей Владимирович Мхитарян,
д.э.н., профессор, профессор кафедры
маркетинга и коммерции МЭСИ
Тел.: 8 (903) 770-31-59
Эл. почта: smhitarian@mesi.ru

Моделирование потребительского выбора на рынке электрощитового оборудования позволит предприятиям-производителям повышать эффективность проектной и маркетинговой деятельности за счет снижения финансовых и трудовых потерь, связанных с предварительной обработкой заказов. Для разработки модели потребительского выбора может применяться логистическая регрессия.

Ключевые слова: модель потребительского выбора, рынок электрощитового оборудования, логистическая регрессия, классификация с откликом, качество прогноза.

Sergey V. Mkhitarian,
PhD in Economics, Professor, Professor of the Department of Marketing and Commerce MESI
Tel.: 8 (903) 770-31-59
E-mail: smhitarian@mesi.ru

MODELLING CONSUMER CHOICE IN THE MARKET SWITCHBOARD EQUIPMENT USING IBM SPSS STATISTICS

Modelling consumer choice in the market switch equipment will allow manufacturing enterprises to improve the efficiency of design and marketing activities by reducing the financial and human losses associated with pre-treatment orders. To develop a model of consumer choice can be used logistic regression.

Keywords: model of consumer choice, market switchboard equipment, logistic regression, classification of the response, the quality of the forecast.

1. Введение

Рынок электрощитового оборудования характеризуется высоким уровнем конкуренции. Широкое распространение коммерческой информации привело к тому, что потребители этой продукции размещают заказы одновременно у нескольких поставщиков. Обработка заказов по электрощитовому оборудованию является достаточно трудоемким процессом, поскольку требует участия не только коммерческого, но и инженерного персонала для формирования спецификаций щитов. Вследствие этих факторов появляется потребность в предварительной оценке вероятности успешного завершения сделки, поскольку, существует проблема не оплаты клиентами предварительно обработанных заказов, что приводит к финансовым и трудовым потерям фирмы. В этой связи актуальна задача моделирования потребительского выбора на рынке электрощитового оборудования. Полученные модели могут использоваться для определения на предварительном этапе с достаточной вероятностью: будет ли оплачен заказ или нет. Наличие такой информации позволит компании сосредоточить основное внимание на заказах с предполагаемым положительным результатом. Для построения подобной модели можно использовать параметры заказа и данные о клиенте, имеющиеся в базе данных компании.

Предлагается разработать модель с помощью логистической регрессии.

2. Применение логистической регрессии для моделирования потребительского выбора

В множественной логистической регрессии непосредственно оценивается вероятность события (рис. 1) по формуле:

$$p = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}},$$

где $z = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m$, b_i – коэффициенты регрессии, при x_i -й объясняющие переменные, где $i = 1, 2, \dots, m$.

$$p = \frac{e^z}{1 + e^z} \Rightarrow e^z = \frac{p}{1-p}, z = \ln\left(\frac{p}{1-p}\right),$$

где $\frac{p}{1-p}$ – шансы, $z = \ln\left(\frac{p}{1-p}\right)$ – логит

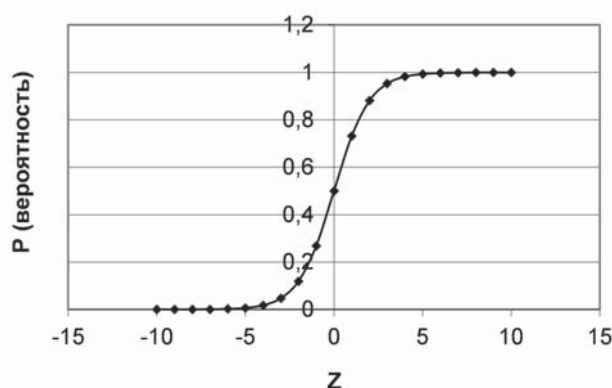


Рис. 1. Логистическая кривая

Вероятность успеха сделки (оплаты счета) возрастает с усилением влияния совокупности факторов, характеризующих как саму сделку, так и уровень взаимодействия покупателя с компанией-поставщиком. Данная функция не линейная, а S-образная. Небольшое изменение в побуждении может иметь или не иметь эффекта в зависимости от того, где расположена отправная точка. Однако для предсказывающей переменной, являющейся дихотомической или имеющей ограниченную шкалу, переход может происходить намного быстрее (более крутой подъем кривой на графике). В логистической регрессии для оценки параметров используется метод максимального правдоподобия и итерационный алгоритм (из-за нелинейности).

Преимущество применения логистической регрессии для моделирования потребительского выбора позволяет не только определить предполагаемый бинарный отклик клиента, но и рассчитать вероятность заключения сделки. Данный нюанс чрезвычайно полезен в прикладном аспекте решения поставленной задачи.

2. Моделирование потребительского выбора на рынке электрощитовой продукции с помощью логистической регрессии с линейным логитом

При построении модели в качестве отклика будет выступать факт оплаты заказа. Неоплаченная сделка является несостоявшейся.

Логит-модель, используя логистическую функцию для аппроксимации, позволяет определить оценки результирующего признака, т.е. в данном случае – вероятности совершения сделки (благоприятного выбора покупателя для компании-поставщика).

В качестве инструмента для проведения анализ и построения модели воспользуемся пакетом IBM SPSS Statistcs.

Рассмотрим влияние на оплату или неоплату счета, выставленного фирмой заказчику, 5-и факторов:

- выписанная сумма счета, тыс. руб. (x_1),

- количество заказов, ранее оплаченных заказчиком (x_2),
- количество месяцев, в течение которых заказчик обращался на фирму (x_3),
- сумма премиум-товара в заказе (нестандартного электрощитового оборудования), тыс. руб. (x_4),
- количество наименования продукции в счете (x_5).

Анализ проводился данным внутренней отчетности компании, производящей электрощитовое оборудование, число наблюдений $n = 231$.

Воспользуемся пошаговым алгоритмом включения переменных.

Для построения логистической регрессии проводилось взвешивание исходных данных, поскольку количество наблюдений для различных уровней признака оплаты непропорционально. Для выявления значимых предикторов использовался пошаговый алгоритм построения модели. В результате расчетов получены оценки параметров логистической модели (табл. 1).

Оценка логистического регрессионного уравнения вероятности покупки имеет вид:

$$\hat{p} = \frac{1}{1 + e^{-\hat{z}}}$$

где $\hat{z} = 0,034 - 0,003x_1 + 0,690x_3$

Следовательно, при увеличении суммы заказа вероятность покупки

уменьшается, но чем больше месяцев клиент сотрудничает с фирмой, тем выше вероятность покупки.

Полученное уравнение можно использовать для классификации новых заказов и оценки вероятности их оплаты. Для оценки качества модели построена таблица классификации (табл. 2), показывающая количество, процент и характер ошибок. Ее недостаток – не показывается распределения вероятностей, кроме того, непонятно, как будет себя вести модель с новыми данными (для этого есть возможность в модуле логистической регрессии отобразить часть имеющихся данных для контроля).

Общий процент корректно предсказанных значений составляет 66,5%, следовательно, можно говорить об удовлетворительном качестве модели. Формальный признак разделения наблюдений на группы следующий: если предсказанная вероятность покупки превышает 0,5 то предполагается, что счет будет оплачен, если меньше 0,5 – то оплачен не будет.

Как видно из таблицы 2 доля корректно предсказанных результатов неоплаты и оплаты составляю 64,5% и 69,6% соответственно. Для повышения этих показателей предлагается сформировать переменную отклика не бинарную, а имеющую три категории и соответствующую

Таблица 1

Характеристики логистической регрессии вероятности оплаты заказа электротехнической компании

Переменные, j	b_j	Средняя квадрат. ошибка	Статистика Вальда	Степень свободы	Уровень значимости	e^{b_j}
x_1	-0,003	0,001	7,835	1	0,005	0,997
x_3	0,069	0,023	8,889	1	0,003	1,072
Const	0,034	0,268	0,016	1	0,900	1,034

Таблица 2

Классификация оплаты/неоплаты заказов электротехнической компании

Наблюдаемые данные	Предсказанные данные		Итого
	Неоплата	Оплата	
Признак оплаты, частота, %	Неоплата	85 64,5%	30 30,4%
	Оплата	47 35,5%	69 69,6%
Итого	132 100,0%	99 100,0%	231 100,0%

Таблица 3

Классификация оплаты/неоплаты заказов электротехнической компании (3 категории предсказания)

Наблюдаемые данные		Предсказанные данные			Итого
		Неоплата	Неизвестно	Оплата	
Признак оплаты, частота, %	Неоплата	30 76,9%	70 51,9%	15 26,3%	115 49,8%
	Оплата	9 23,1%	65 48,1%	42 73,7%	116 50,2%
Итого		39 100,0%	135 100,0%	57 100,0%	231 100,0%

Таблица 4

Характеристики полиномиальной модели логистической регрессии вероятности оплаты заказа электротехнической компании

Переменные, j	b_j	Средняя квадрат. ошибка	Статистика Вальда	Степень свободы	Уровень значимости	e^{b_j}
x_1	-0,004	0,001	7,835	1	0,006	0,996
x_3	0,230	0,057	16,373	1	0,000	1,259
x_2^2	-0,007	0,002	11,406	1	0,001	0,993
x_2^3	5E-5	0,000	9,701	1	0,002	1,000
Const	0,034	0,336	3,916	1	0,048	0,514

Таблица 5

Классификация оплаты/неоплаты заказов полиномиальной модели логистической регрессии электротехнической компании

Наблюдаемые данные		Предсказанные данные		Итого
		Неоплата	Оплата	
Признак оплаты, частота, %	Неоплата	78 65,0%	37 22,4%	115 49,8%
	Оплата	42 36,0%	74 66,7%	116 50,2%
Итого		120 100,0%	111 100,0%	231 100,0%

различным уровням вероятности оплаты: *Неоплата* ($p < 0,4$); *Не определено* ($0,4 \leq p < 0,6$); *Оплата* ($p \geq 0,6$). В результате применения данного подхода была получена следующая таблица классификации (табл. 3).

В соответствии с новым решением доля корректно предсказанных результатов неоплаты и оплаты составляю 76,9% и 73,7% соответственно, что улучшает качество прогноза.

Уровни вероятности и количество категорий отклика могут варьироваться в зависимости от поставленной задачи и регрессионной модели. Расширение границ среднего диапазона (зоны неопределенности) будет повышать качество прогноза, но в то же время увеличивать число сделок, по которым результата предсказания не будет однозначным.

3. Моделирование потребительского выбора на рынке электрощитовой продукции с помощью логистической регрессии с нелинейным логитом

Представленная выше модель основывалась на предположении о линейном характере влияния предикторов на логит. Рассмотрим гипотезу о нелинейной зависимости регрессоров и логита. В качестве основы воспользуемся полиномиальной моделью 3-го порядка.

$$z = b_0 + b_{11}x_1 + b_{12}x_2 + \dots + b_{1m}x_{1m} + b_{21}x_1^2 + b_{22}x_2^2 + \dots + b_{2m}x_m^2 + b_{31}x_1^3 + b_{32}x_2^3 + \dots + b_{3m}x_m^3$$

Для построения модели воспользуемся методами линеаризации (приведения уравнения к линейному виду): в качестве дополнительных переменных выступают предикторы второй и третьей степени.

В результате применения пошагового алгоритма получены оценки параметров новой логистической модели (табл. 4).

Оценка логистического регрессионного уравнения вероятности покупки имеет вид:

$$\hat{p} = \frac{1}{1 + e^{-\hat{z}}}$$

где $\hat{z} = 0,034 - 0,004x_1 + 0,230x_2 - 0,007x_2^2 + 5 \cdot 10^{-5}x_2^3$

Следовательно, при увеличении суммы заказа вероятность покупки уменьшается, но чем больше количество заказов было ранее оплачено клиентом, тем выше вероятность покупки. Причем, последняя зависимость носит явно нелинейный характер и характеризуется полиномом 3-ей степени.

Полученное уравнение также можно использовать для классификации новых заказов и оценки вероятности их оплаты. Для оценки качества модели построена таблица классификации (табл. 5), показывающая количество, процент и характер ошибок.

Общий процент корректно предсказанных значений составляет 65,5%, следовательно, можно говорить об удовлетворительном качестве модели. Данное значение на 1% ниже, чем у предыдущей модели, но последняя модель лучше предсказывает неоплату счета (64,7% против 61,9%). Формальный признак разделения наблюдений на группы следующий: если предсказанная вероятность покупки превышает 0,5 то предполагается, что счет будет оплачен, если меньше 0,5 – то оплачен не будет.

Как видно из таблицы 2 доля корректно предсказанных результатов неоплаты и оплаты составляю 65,0% и 66,7% соответственно, что несколько хуже первоначальной мо-

Таблица 6

Классификация оплаты/неоплаты заказов полиномиальной модели логистической регрессии электротехнической компании (3 категории предсказания)

Наблюдаемые данные		Предсказанные данные			Итого
		Неоплата	Неизвестно	Оплата	
Признак оплаты, частота, %	Неоплата	46 70,8%	48 60,0%	21 26,3%	115 49,8%
	Оплата	19 29,2%	32 40,0%	65 75,6%	116 50,2%
Итого		65 100,0%	80 100,0%	86 100,0%	231 100,0%

дели, но соотношение оплаченных и неоплаченных заказов оказывается ближе к наблюдаемому соотношению у второй модели. Для повышения качества предсказания воспользуемся приемом, примененным в первой модели, и сформируем переменную отклика не бинарную, а имеющую три категории и соответствующую различным уровням вероятности оплаты: *Неплата* ($p < 0,36$); *Не определено* ($0,36 \leq p < 0,6$); *Оплата* ($p \geq 0,6$). В результате применения данного подхода была получена следующая таблица классификации (табл. 6).

В соответствии с новым решением доля корректно предсказанных результатов неоплаты и оплаты составляют 70,8% и 75,6% соответственно, что улучшает качество прогноза. По сравнению с предыдущей моделью по категории «неоплата» качество снизилось на 6,1%, по категории «оплата» – повысилось на

1,9%. Но существенное преимущество второй модели заключается в том, что в категорию «неизвестно» существенно снизилась доля наблюдений: с 58% до 35%, что существенно снизило неопределенность прогноза.

4. Заключение

Можно сделать вывод, что для моделирования потребительского выбора на рынке электрошитового оборудования может быть применена логистическая регрессия. Целесообразно рассматривать варианты нелинейного логита, а также не две, а три категории предсказания. Данный подход позволяет существенно повысить качество модели и снизить ее неопределенность.

Литература

1. Мхитарян С.В. Отраслевой маркетинг. М.: Эксмо, 2006.

2. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. М.: Юнити, 2001.

3. Гасимова Г.М., Дейнекин Т.В., Зубин С.И., Казанцева С.Ю., Каптюхин Р.В., Красюк И.Н., Серeda О.В., Мхитарян С.В., Невоструев П.Ю., Панюкова В.В., Попова И.М., Романов А.А., Тультаев Т.А., Тультаева И.В., Шкляр Т.В. Влияние инновационных маркетинговых технологий на развитие коммерческой деятельности в России и за рубежом. Коллективная монография профессорско-преподавательского состава кафедры Маркетинга и коммерции Института Менеджмента МЭСИ – М.: МЭСИ, 2013.

References

1. Mkhitarian S.V. Field marketing. M.: Eksmo, 2006.

2. Ayvazian S.A., Mkhitarian V.S. Probability Theory and Applied Statistics. M.: Unity, 2001.

3. Gasimova G.M., Deinekin T.V., Zubin S.I., Kazantseva S.Yu., Kaptyuhin R.V., Krasjuk I.N., Sereda O.V., Mkhitarian S.V., Nevostruev P.Yu., Panyukova V.V., Popova I.M., Romanov A.A., Tultaeв T.A., Tultaeвa I.V., Shkliar T.V. Influence of innovative marketing technologies to the development of commercial activities in Russia and abroad. Collective monograph of the faculty of the Department of Marketing and Commerce Institute of Management MESI – M.: MESI, 2013.