

Статистический анализ факторов, оказывающих влияние на качество предоставляемых услуг населению Оренбургской области

© 2010 Н.Ю. Цыпина
Оренбургский государственный университет
E-mail: cippka@mail.ru

Автором разработана и реализована методика выборочного обследования основных видов услуг, предоставляемых жителям областного центра, г. Оренбурга.

Ключевые слова: статистика, эконометрика, фиктивные переменные, регрессионные модели, качество услуг.

Статистическая информация, собираемая и обобщаемая территориальным органом Федеральной службы государственной статистики, недостаточна для выявления факторов, оказывающих влияние на качество предоставляемых населению услуг, и степени удовлетворенности ими потребителей. В связи с этим в исследовании нами была разработана и реализована методика выборочного обследования основных видов услуг предоставляемых жителям областного центра, г. Оренбурга. В опросных листах выделялись наиболее значимые услуги, оцененные по наибольшему удельным весам в общем объеме потребления населения, это услуги системы жилищно-коммунального хозяйства, транспорта и связи.

Получаемая информация носит как количественный, так и качественный характер, в силу чего не представляется возможным проведение классического регрессионного анализа, что привело нас к «кодированию» нечисловой информации, переводу ее в количественную и использованию фиктивных переменных.

Предложенный вариант, на наш взгляд, обладает двумя преимуществами:

- во-первых, легко проверить, является ли воздействие качественного фактора значимым;
- во-вторых, при условии выполнения определенных предположений регрессионные оценки оказываются более эффективными.

Обычно фиктивная переменная отражает два противоположных состояния качественного фактора и может выражаться в двоичной форме:

$$D = \begin{cases} 0, & \text{фактор не действует,} \\ 1, & \text{фактор действует.} \end{cases}$$

Переменная D является фиктивной (искусственной, двоичной) переменной (индикатором).

Следует отметить, что в эконометрике фиктивные переменные используются для моделирования взаимосвязей как в пространстве, так и

во времени, при этом они могут находиться как в правой, так и в левой части уравнения.

Учитывая особенности имеющихся в нашем распоряжении данных, остановимся на рассмотрении класса моделей, основанных на пространственных данных с фиктивной переменной в левой части уравнения. Таковыми являются: пробит-, логит-модели и модели множественного выбора.

Пробит (probit)-модель основана на законе нормального распределения $N(0,1)$:

$$F(z) = f(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \quad (1)$$

Пробит-модель для бинарных данных:

$$P(y_i = 1 | X_i^T) = \int_{-\infty}^{x_i^T \beta} \varphi(t) dt = f(X_i^T \beta), \quad (2)$$

$$P(y_i = 0 | X_i^T) = 1 - f(X_i^T \beta),$$

где y_i - дискретная зависимая переменная;

X_i - вектор независимых переменных;

$\varphi(\cdot)$ - функция плотности вероятностей стандартного нормального закона распределения;

$f(\cdot)$ - функция распределения нормального закона.

Логит (logit)-модель основывается на логистическом законе распределения вероятностей. Функция распределения вероятностей логистического закона:

$$\Lambda(z) = \frac{e^z}{1 + e^z}. \quad (3)$$

Логит-модель для бинарных данных:

$$P(y_i = 1 | X_i) = \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} = \Lambda(X_i^T \beta), \quad (4)$$

$$P(y_i = 0 | X_i) = \frac{1}{1 - e^{X_i^T \beta}} = 1 - \Lambda(X_i^T \beta).$$

Графики функции распределения нормального и логистического распределения при соответствующей нормировке достаточно близки. На интервале $z \in [-1, 2; 1, 2]$ они практически одинаковы. Однако логистическая функция имеет более “тяжелые хвосты”, т. е. медленнее стремится к нулю при $z \rightarrow -\infty$ или единице при $z \rightarrow \infty$. Поэтому логит- и пробит-модели дают похожий результат, если только изучаемая вероятность не слишком близка к нулю или к единице.

Во многих экономических задачах количество альтернатив может быть больше двух. В этих случаях уместно использование класса моделей множественного выбора. Данные модели позволяют описать вероятность каждой из альтернатив как функцию наблюдаемых характеристик объекта. При этом вероятности должны лежать в интервале от 0 до 1, а сумма вероятностей по всем альтернативам должна быть равна единице.

В рамках данной группы моделей, согласно Н.П. Тихомирову¹, В.С. Мхитаряну², выделяют:

Модели множественного выбора. В данном направлении наибольшее распространение получили множественные логит- и пробит-модели

(*multinomial logit model, multinomial probit model*), гнездовые множественные логит-модели (*nested multinomial logit model*).

Модели упорядоченного выбора. Наиболее часто в качестве распределения вероятностей выбирают логистический и нормальный законы. Соответствующие модели получили название “порядковые логит и пробит” (*ordered logit, ordered probit*). Их успешно применяют при моделировании результатов опросов общественного мнения, выбора уровня сложности работы, типов страховых полисов, когда альтернативы могут быть естественным образом упорядочены.

Для некоторых параметров, например, числа поездок в супермаркет, посещений врача, количества полученных патентов и пр., в качестве закона распределения ошибки используется закон Пуассона или отрицательный биномиальный закон. Эти модели получили название модели счетных данных (*count data*).

Модели многовариантного бинарного выбора.

Обобщая рассмотренный теоретический материал, можно сформулировать следующую классификацию фиктивных переменных (см. рисунок).

Применительно к теме исследования представляется возможным использование моделей множественного выбора, это объясняется тем, что



Рис. Фиктивные переменные в регрессионных моделях

результатирующая переменная (качество услуг, оцененных потребителем) представлена в виде упорядоченной переменной 0 - наихудшее качество, ..., 5 - наилучшее качество.

Кроме того, возможно упростить задачу, потеряв часть точности информации, и свести имеющуюся переменную к бинарной:

- 0 - ответы респондентов 0,1
- и 2 - неудовлетворительное качество,
- 1 - ответы респондентов 3,4
- и 5 - удовлетворительное качество.

Пользуясь имеющимися данными о качестве оказываемых услуг, последовательно оценим logit-модель для транспорта, связи и ЖКХ, в разрезе 2008 г. и 2009 г. При этом расчеты проведем в пакете STATISTICA 6.1 в модуле "Нелинейное оценивание".

На первом этапе проверим влияние факторов на качество услуг связи, при этом необходимо отметить, что среднее значение показателя в 2008 г. составляет 3,8 балла, а в 2009 г. - 4,0 балла. Получаем в результате перекодирования в бинарную переменную удовлетворительных оценок - 92,1% (2008) и 94,3% (2009), а неудовлетворительных - 7,9% (2008) и 5,7% (2009).

В качестве факторов, оказывающих влияние на результирующую переменную, выберем следующие (табл. 1).

Выбор переменных объясняется следующими соображениями:

- во-первых, услуга связи является разновидностью продукта, поэтому возможно различное восприятие мужчинами и женщинами качества данной услуги, так как считается, что женщины более требовательные к качеству;
- во-вторых, в связи с размерами сети, локализацией и финансовыми возможностями различные операторы связи предоставляют разноразличные по качеству услуги;
- в-третьих, несоответствие цены и качества негативным образом влияет на мнение респондентов относительно качества услуги конкретного оператора.

В результате оценки модели получаем следующие результаты (табл. 2).

Можно сделать вывод о наличии влияния на оценку качества пола человека (модель значима по χ^2 и *t*-критерию Стьюдента), другие модели характеризуются статистической значимостью свободного члена и общей незначимостью, соответственно, выделенные факторы не оказывают влияния на результат.

Таблица 1. Факторы, оказывающие влияние на качество услуг связи

Фактор	Фиктивная переменная
Пол человека	$D_1^{пол} = \begin{cases} 1-мужчины \\ 0-женщины \end{cases}$
Оператор связи	$D_2^{оператор} = \begin{cases} 1-Мегафон \\ 0-остальные операторы \end{cases}$
	$D_3^{оператор} = \begin{cases} 1-МТС \\ 0-остальные операторы \end{cases}$
Соотношение цены и качества услуги	$D_4^{цена-кач.} = \begin{cases} 1-удовлетворены \\ 0-не удовлетворены \end{cases}$

Таблица 2. Результаты оценки моделей влияния факторов на качество связи в 2009 г.*

Модель	χ^2	Уровень значимости
$\hat{y}_i = -3,296 + 1,027D_1^{пол}$ (-4,577) (5,091)	3,221	0,026
$\hat{y}_i = -3,296 + 0,288D_2^{оператор} + 0,736D_3^{оператор}$ (-3,114) (0,230) (0,584)	0,383	0,825
$\hat{y}_i = -2,521 - 0,809D_4^{цена-кач.}$ (-3,028) (-0,852)	0,681	0,409

* В скобках приведены фактические значения *t*-критерия Стьюдента.

Таблица 3. Факторы, оказывающие влияние на качество услуг транспорта

Фактор	Фиктивная переменная
Пол человека	$D_1^{\text{пол}} = \begin{cases} 1 - \text{мужчины} \\ 0 - \text{женщины} \end{cases}$
Частота пользования общественным транспортом	$D_2^{\text{частота}} = \begin{cases} 1 - \text{часто} \\ 0 - \text{в остальных случаях} \end{cases}$
	$D_3^{\text{частота}} = \begin{cases} 1 - \text{периодически} \\ 0 - \text{в остальных случаях} \end{cases}$

Таблица 4. Результаты оценки моделей влияния факторов на качество услуг транспорта в 2009 г.*

Модель	χ^2	Уровень значимости
$\hat{y}_i = 0,793 + 0,816D_1^{\text{пол}}$ (2,869) (1,330)	1,953	0,162
$\hat{y}_i = 1,386 - 0,457D_2^{\text{частота}} - 0,511D_3^{\text{частота}}$ (2,148) (-3,640) (-4,611)	3,491	0,025

* В скобках приведены фактические значения t-критерия Стьюдента.

Таблица 5. Факторы, оказывающие влияние на качество услуг ЖКХ

Фактор	Фиктивная переменная
Частота уборки в подъезде	$D_1^{\text{уборка}} = \begin{cases} 1 - \text{раз в неделю} \\ 0 - \text{в остальных случаях} \end{cases}$
	$D_2^{\text{уборка}} = \begin{cases} 1 - \text{раз в день} \\ 0 - \text{в остальных случаях} \end{cases}$
Перебои в поставке воды, электричества, газа и т.д.	$D_3^{\text{перебои}} = \begin{cases} 1 - \text{не бывает или редко} \\ 0 - \text{в остальных случаях} \end{cases}$
Наличие текущего ремонта дома	$D_4^{\text{тек.ремонт}} = \begin{cases} 1 - \text{периодически} \\ 0 - \text{в остальных случаях} \end{cases}$
Дополнительные денежные затраты при вызове работников ЖКХ	$D_5^{\text{доп.затраты}} = \begin{cases} 1 - \text{да} \\ 0 - \text{в остальных случаях} \end{cases}$
	$D_6^{\text{доп.затраты}} = \begin{cases} 1 - \text{в зависимости от причины} \\ 0 - \text{в остальных случаях} \end{cases}$

Далее рассмотрим модели влияния факторов на качество услуг транспорта, при этом выделим следующие факторы (табл. 3).

Рассмотрение пола обусловлено тем, что общественным транспортом пользуются в основном женщины, которые более требовательны к обслуживанию (вежливость, наполняемость, безопасность, чистота и т.д.).

Включение второго показателя объясняется, что чем чаще респондент пользуется услугой, тем выше вероятность столкнуться с негативными факторами и тем ниже оценка качества услуги.

В результате построения моделей получен следующий результат (табл. 4).

Можно утверждать, что пол человека не оказывает влияния на оценку качества услуги, а второй фактор оказывает (модель значима по χ^2 и t-критерию Стьюдента).

На заключительном этапе данного исследования рассмотрим влияние факторов на качество

услуг жилищного и коммунального хозяйства (табл. 5), при этом выберем следующие показатели:

1. Частота уборки в подъезде - чем чаще убирают, тем выше оценка качества услуг ЖКХ.

2. Перебои в поставке воды, электричества, газа и т.д. - частые перебои негативно влияют на оценку респондентами качества услуг.

3. Наличие текущего ремонта дома - ремонт, проводимый в недалеком прошлом, благотворно влияет на оценку качества ремонта.

4. Дополнительные денежные затраты при вызове работников ЖКХ - ожидается негативная реакция при требовании их от домохозяйств за дополнительные услуги от работников ЖКХ, так как обслуживание входит в ежемесячную оплату.

В результате оценки моделей получаем следующие результаты (табл. 6).

Таким образом, на мнение респондентов, относительно качества услуг ЖКХ, оказывает

Таблица 6. Результаты оценки моделей влияния факторов на качество услуг жилищной и коммунальной сферы в 2009 г.*

Модель	χ^2	Уровень значимости
$\hat{y}_i = -1,466 + 0,198D_1^{\text{уборка}} + 0,640D_2^{\text{уборка}}$ (-2,289) (0,266) (0,815)	0,834	0,656
$\hat{y}_i = -0,693 - 0,600D_3^{\text{перебои}}$ (-1,265) (-0,959)	0,883	0,347
$\hat{y}_i = -0,693 - 1,576D_4^{\text{тек.ремонт}}$ (-2,264) (-2,319)	6,691	0,010
$\hat{y}_i = -23,555 + 22,734D_5^{\text{доп.затраты}} + 22,20D_6^{\text{доп.затраты}}$ (-2,001) (0,236) (0,113)	3,814	0,149

* В скобках приведены фактические значения t-критерия Стьюдента.

большее и статистически значимое влияние наличие ремонта дома.

В заключение можно сделать следующие выводы:

1) предложенное методическое обеспечение дает возможность применения регрессионных моделей к нечисловой информации;

2) использование моделей множественного выбора позволяет получать большее количество статистически значимых моделей, так как использование logit- и probit-моделей вынуждает пре-

образовывать исходную информацию, что снижает ее качество;

3) использование logit-модели позволило выявить влияние на качество связи пола респондента, частоты пользования общественным транспортом, наличия текущего ремонта дома.

¹ Тихомиров Н.П., Дорохина Е.Ю. Эконометрика: учебник. М., 2003. С. 403-416.

² Эконометрика: учебник / под ред. В.С. Мхитаряна. М., 2009. С. 233-236.

Поступила в редакцию 07.08.2010 г.