

Экономико-статистический анализ поведения потребителей в сфере электронного бизнеса (на примере Казахстана)

Родионов А. Ю. (rodionov@au.ru)

Казахский Экономический Университет им. Турара Рыскулова

На протяжении последнего десятилетия социологи и экономисты значительное внимание уделяют изучению потребительского поведения аудитории Интернета и ее коммуникативного окружения. В теоретических построениях, лежащих в основе данного исследования, автор рассматривает потребительский выбор в рамках понятий теории поведенческих решений — направления *кардиналистского* (количественного) подхода к анализу выбора потребителей, что соответствует подходу американской школы *социального выбора*. Данная теория явилась естественным результатом развития институциональной экономической теории /1/ (или шире — институционально-социологического течения экономической мысли) Запада¹. Взамен парадигмы «экономического человека» неоклассиков была предложена парадигма «человека Кахенмана-Тверского» /2/, в которой поведение человека при выборе может быть описано как *процесс принятия решения*, который формируется восприятием и верой, основанными на доступной информации, и под влиянием переживаний, отношений, мотивов и предпочтений.

Теория социального выбора имеют радикальные отличия от классической экономической теории *ординалистской* (порядковой) полезности в представлениях о процессе принятия решения. Во-первых, исходное направление институционально-социологических исследований нацелено на понимание природы компонентов решения о выборе, как они устанавливаются и изменяются опытом, а также и как они устанавливают ценности того или иного выбора. В фокусе экономистов находится прорисовка кратчайших маршрутов от входящей информации к итоговому выбору. Предпочтения или полезность могут раскрываться в большинстве экономических исследований в результате довольно примитивного анализа, а сам процесс решения скорее всего будет представлен как «черный ящик». Получивший широкое хождение вследствие этого афоризм «экономисты знают *цену* всему, а *полезность* ничему» не может верно характеризовать научные приоритеты дисциплины.

Во вторых, институционально-социологические представления о процессе принятия решения находятся во власти идей, что поведение конкретно, адаптивно, обучаемо, изменчиво, зависит от социально-экономического и этнополитического контекста и подвержено влиянию сложных взаимодействий восприятия, мотивов, отношений и переживаний. В противоположность этому *стандартная модель* в экономике описывает потребителей ведущих себя так, как будто информация обработана, чтобы сформировать восприятие и веру, используя строгие Байесовские статистические принципы (*рациональность восприятия*), предпочтения в этом случае

¹ В последние десятилетия среди магистральных направлений экономической мысли выделяют: а) *классическую* (или *неоклассическую*) экономическую теорию, занимающуюся вопросами, прежде всего, микроэкономики, б) *кейнсианство*, создавшее основы теории макрорегулирования, и в) *институционально-социологическое* направление, пытающееся выяснить влияние на экономическое поведение экономических субъектов конкретных социальных институтов.

рассматриваются примитивными, последовательными и неизменяемыми (*рациональность предпочтения*), а познавательный процесс лишь максимизация предпочтений, дающая рыночные константы (*рациональность процесса*).

Математические модели, которые основываются на вышеизложенных психометрических подходах, относят к классу *моделей дискретного выбора*, применяемых в теории случайной полезности (разделе теории социального выбора). Данные модели нашли в последние тридцать лет широкое применение в таких прикладных областях как размещение домовладельцев, транспортные потоки, привлечение рабочей силы, миграция народонаселения, природопользование и многих других, см.: /3/.

Исследование поведения казахстанских потребителей электронного бизнеса, изложенное в данной работе, было проведено в два этапа.

Первый этап — моделирование зависимости обращения в Интернет от социально-экономических и поведенческих характеристик индивидуумов и домохозяйств при помощи бинарной логистической модели (Binary Logit Model), помимо этого производился отбор факторных признаков и их редуцирование (сокращение), направленный на повышение качества статистической модели и способствовавшей более понятной экономической интерпретации полученных результатов.

На втором этапе применялась более совершенная методика построения множественной логистической модели (Multinomial Logit Model) для изучения интенсивности применения электронной коммерции, давших результаты, толкование которых, позволило сформулировать рекомендации по развитию электронной коммерции в Казахстане.

Ряд исследований поведения потребителей, связанных с выходом в Интернет и участием в электронной коммерции, проведенных в развитых странах, выявили сильную положительную корреляцию между интенсивностью обращения к Интернету, ведением электронного бизнеса и склонностью к активному использованию инноваций, особенно если пользователи предполагают, что они сэкономят их усилия и время. Маркетологи из американских университетов У. Моэ (W. Моэ) и П. Фадер (P. Fader) полагают, что особенности, имеющиеся в поведении пользователей серверов в глобальной паутине, дают серьезные основания для выделения их из круга обычных потребителей телекоммуникационных услуг и проведения специального изучения, которое позволит существенно повысить эффективность маркетинговой деятельности /4/.

Г. Лозе (G. Lohse), С. Белмен (S. Bellman) и Э. Джонсон (E. Johnson) в результате анализа панельных данных опроса американских пользователей Интернета выделяют четыре группы по интенсивности использования электронной коммерции: «никогда не покупающих», «случайных покупателей», «новичков» и «постоянных покупателей» /5/. Исследователи отмечают, что для новичков и постоянных покупателей характерно более продолжительное знакомство с Интернетом, интенсивное применение электронной почты, активное использование Сети для получения информации по бизнесу, путешествиям, значительный отклик на баннерную рекламу, а также более высокий уровень доходов, чем у не покупающих и случайных покупателей интернет-магазинов.

В серии научных работ, выполненных по программе изучения будущих медиасистем, осуществляемой в Лондонской школе бизнеса, было установлено, что использование Интернета предполагает достаточно тесное общение с лицами, приме-

няющими современные информационно–телекоммуникационные технологии (эффект влияния окружения) /6/. К ним, в первую очередь, относятся персональная мобильная связь и подписка на различного рода информационные услуги. Наряду с этим отмечается стремление к использованию данных технологий для бизнеса и в личных интересах.

Затрагивая различные аспекты мотивации применения сетевых технологий, европейские исследователи полагают, что подобные процессы прежде всего побуждаются возрастающими потребностями западного постмодернистского общества к использованию связи и получения разнообразной информации. Указывают, что интенсивное обращение к Интернету наблюдается в отношении пользователей, давно посещающих Сеть. У активных пользователей повсеместно прослеживается наличие широких контактов посредством электронной почты, как в личных, так и в деловых целях. Активное применение Интернета ведет к цепи действий, способных вызвать повышение экономического, а в условиях развитого рыночного общества, и социального статуса человека. Такая особенность характерна в первую очередь для молодого поколения /7/.

Английские экономисты К. Хамонд (K. Hammond), П. Тернер (P. Turner) и М. Бэйн (M. Bain) утверждают, что важнейшим фактором, обуславливающим выход в Интернет британских потребителей, являются склонность потенциальных пользователей Интернета к включению в свою жизнь инноваций, которые способны по их мнению сберечь усилия и время, предоставить оперативный доступ к различным источникам информации и дать ощущение принадлежности к слою более социально значимых людей. Пользователи Интернета в сравнении с лицами вне зоны Сети с большей вероятностью предпочтут информационные телевизионные программы, чем газетные и журнальные публикации /8/.

В качестве анализируемого массива данных были взяты материалы опроса взрослых жителей крупнейших городов Казахстана по вопросам применения национальной информационно–телекоммуникационной инфраструктуры в 2001 г. Дефиниции применяемых понятий, методика сбора информации (*офлайн формат*), демографические характеристики выборочной совокупности приведены в специальной работе /9/.

После того как на стадии априорного анализа произведен отбор факторов, влияющих на применение Интернета, а на основе теории случайной полезности определена логистическая форма зависимости, была собрана и проанализирована исходная статистическая информация. Это позволило перейти к построению модели применения Интернета в Казахстане. Моделирование применения Интернета производилось по четырнадцати факторам, отобранным автором.

На массиве с данными на однородной совокупности, состоящей из 1200 наблюдений ($N = 1200$), была построена матрица коэффициентов непараметрической корреляции. Исследование матрицы выявило достаточно существенную связь между исследуемыми факторными показателями и результативным порядковым признаком — интенсивностью обращения к Интернету.

Особенно интересными представляются показатели компьютерной грамотности; количества знакомых регулярно посещающих Интернет; возраст и пол респондента; логарифм дохода на одного члена домохозяйства (в тенге), а также принадлежность к некоторым социальным группам: предприниматели, руководители, спе-

циалисты и студенты. Симметричная матрица сформирована при помощи коэффициента связи Кендэлла τ , используемого для измерения взаимосвязи между качественными и количественными признаками /10/.

Ряд качественных признаков ($X_1, X_2, X_3, X_{10}, X_{11}, X_{12}$ и X_{13}) преобразованы в индикаторные дихотомические переменные. При этом значение последней категории назначается контрастной переменной, а значение ее логистического коэффициента принимается равным 0 и не выводится в таблице оценки параметров. Например, для переменной X_{12} («Социальный статус респондента») дихотомический признак «студент/учащийся» является контрастной переменной. Интерпретация коэффициентов для переменных, например, «домохозяйка» и «специалист» заключается в сравнении со значением контрастной переменной «студент/учащийся».

Логистическое уравнение вероятности того, что индивидуум выберет альтернативу, в данном случае применение Интернета произойдет, можно записать, базировавшись на модели дискретного выбора, как

$$Y_{\text{Prob}(\text{Internet})} = \frac{1}{(1 + \exp(-Z))}, \quad (1)$$

где *Internet* — результативный дихотомический признак, определяющий принадлежность опрошенного к аудитории Интернета, а Z из (1) есть линейная комбинация:

$$Z = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_p * X_p. \quad (2)$$

К сожалению, для оценки максимального правдоподобия, используемого для нахождения коэффициентов β уравнения (2), не имеется точного аналога R^2 , применяемого для метода наименьших квадратов в линейной регрессионной модели. Для построения бинарной логистической модели использовался многошаговый регрессионный анализ, основанный на исключении из модели несущественных факторов по тесту Вальда (аналогу t -критерию Стьюдента), нашедший реализацию в модуле регрессионного анализа программного обеспечения SPSS /11, pp. 293–297/. В принципе, статистики Вальда при малом числе наблюдений может давать заниженные оценки наблюдаемой значимости коэффициентов, но в рассматриваемых выборках эта особенность не критична.

По критерию Вальда проверяется гипотеза: существенно ли отличен от нуля коэффициент регрессии β_j при некотором заданном уровне теста, который показывает вероятность отвергнуть правильную гипотезу /12/. При этом чем меньше уровень значимости, тем выше указанная вероятность отвержения гипотезы. В нашем исследовании принимаем уровень отсечения по тесту Вальда = 1,75.

Дополнительно в модель включили логически важный факторный показатель X_9 («возраст респондента»). Результаты многошагового регрессионного анализа при построении модели применения Интернета помещены в таблицу А.1. приложения А.

Первый шаг в интерпретации результатов должен состоять в проверке и интерпретации качества приближения модели (goodness of fit of the model). В подходах, принятых при оценке максимального правдоподобия, чаще всего исследуется различие между остатками модели при ограничении, когда все коэффициенты регрессии нулевые, а присутствует только константа и остатками, полученными по модели с оцениваемыми коэффициентами. Модель только с одной константой является выро-

жденным случаем логистического уравнения, например, когда все параметры равны нулю, данную модель используют для сравнения с расширенной моделью с параметрами. Снижение «некачественности приближения» в результате освобождения параметров для каждого факторного показателя X может быть проверено по статистике χ^2 со степенями свободы (DF – degree of freedom) равными количеству квантифицируемых факторов.

В оцениваемой логистической модели, $-2 \text{ Log Likelihood}$ (удвоенный логарифм функция правдоподобия со знаком минус) из уравнения только с константой (начальный $-2LL$) – 799. В модели с константой и шестью факторными признаками конечное $-2LL$ снижается до 417. Различие, или «качество приближения», составляет 382 со степенью свободы 20, существенное при $p < 0,0000$.

Поскольку, как отмечалось выше, для оценки максимального правдоподобия не имеется аналога R^2 , применяемого для метода наименьших квадратов, применяется семейство из нескольких коэффициентов. Один очень простой подход состоит в том, чтобы исследовать величину χ^2 качества приближения в сравнении с χ^2 модели только с константой. Данный коэффициент псевдо- R^2 (23), предложенный Д. Мак-Фадденом (D. McFadden)¹, иногда называемый индексом отношения правдоподобия (likelihood ratio index — LRI):

$$LRI = 1 - \frac{LL_{\text{model}}}{LL_0}, \quad (3)$$

где LL_{model} — логарифм функции правдоподобия оцениваемой модели,

LL_0 — логарифм функции правдоподобия модели только с константой.

Индекс отношения правдоподобия является скалярной мерой, которая варьируется от 0 до 1, подобно коэффициенту R^2 , применяемому для оценки качества линейной регрессии. Для оцениваемой логистической модели этот показатель равен 0,4779 или приблизительно 47,8%. При больших средних значениях данный подход к оценке качества приближения может ввести в заблуждение, поскольку величина χ^2 зависит от размеров выборки.

Кроме описанного выше индекса отношения правдоподобия существует целая группа подобных коэффициентов для анализа оценки качества приближения модели, например, псевдо-коэффициент детерминации Cox and Snell /13/, его дальнейшая модификация Nagelkerke /14/, Darlington's псевдо- R^2 /15/ и ряд других. Для оцениваемой логистической модели коэффициент детерминации Cox and Snell равен 0,273, а Nagelkerke's псевдо- R^2 – 0,561. Следует признать, что относительно высокий уровень этих коэффициентов свидетельствует о достаточном качестве приближения построенной модели.

Вторым шагом в квантификации качества модели должен стать расчет коэффициента согласованной переквалификации. Общий коэффициент согласованной переквалификации модели равен 92,64%, что является очень высоким показателем, но коэффициент согласованной переквалификации Интернет аудитории существен-

¹ Даниэль Мак-Фадден (1937 г. р.) американский экономист, существенно расширивший сферу использования экономической теории и статистических моделей, применив их в анализе способов принятия индивидуумами жизненно важных решений. Профессор Калифорнийского Университета Беркли с 1990 г. Лауреат премии Банка Швеции в области экономических наук имени Альфреда Нобеля за 2000 г.

но ниже 46,36%. Причинами возникновения подобной ситуации могут служить ряд факторов. Во-первых, это объясняется малым удельным весом группы в выборочной совокупности. Во-вторых, нужно учесть значительный объем самой выборки, содержащей некоторую долю аномальных наблюдений. В-третьих, сигнализирует о том, что в модель необходимо включить ненаблюдаемые признаки альтернатив и ненаблюдаемые признаки индивидуумов (так называемые «ненаблюдаемые вариации вкуса») /16/.

На основе модели логистической регрессии можно строить предсказание: произойдет или не произойдет событие $\{Y = 1\}$. Если $Y > 0,50$ считается, что событие произойдет; $Y \leq 0,50$ считается, что событие не произойдет. Это правило оптимально с точки зрения минимизации числа ошибок, но очень грубо с точки зрения исследования связи. Зачастую оказывается, что вероятность события $\text{Prob}\{Y = 1\}$ мала (значительно меньше 0,5) или велика (значительно больше 0,5), поэтому оказывается, что все имеющиеся в выборке сочетания X предсказывают событие или все предсказывают противоположное событие.

Поскольку в моделируемом явлении вероятность попадания в аудиторию мала (10,4%), то необходима иная переклассификация, демонстрирующая связь между результативной и факторными переменными. Для этого необходимо отнести к предсказываемому классу, наблюдения для которых применение Интернета ожидается с меньшей вероятностью, чем в среднем, а остальные — к противоположному классу. В нашем случае к классу предсказанных значений отнесены наблюдения со значением $Y > 0,10$. Результаты этих двух классификаций представлены в классификационной таблице (см. Таблица 1).

Таблица 1

Классификационные таблицы логистической регрессионной модели принадлежности к аудитории Интернета

		Пороговое значение $Y = 0,50$				Пороговое значение $Y = 0,10$		
		предсказано				предсказано		
		0	1	корректный процент		0	1	корректный процент
А	Б	1	2	3	В	4	5	6
Наблюдается	0	1054	22	98,00%	0	921	160	85,20%
	1	67	58	46,36%	1	13	106	89,08%
<i>Всего</i>		<i>1121</i>	<i>80</i>	92,64%		<i>934</i>	<i>266</i>	85,58%

Из таблицы видно, что использование более реалистического порога, приблизительно равного доли аудитории во взрослом населении крупнейших городов (0,10) дало существенное улучшение качества предсказания попадания наблюдений в данный класс (с 46,38% до 89,08%). При этом общий коэффициент согласованной переклассификации оцениваемой модели остался весьма высоким – 85,58%.

Однако для экономико-статистического анализа крайне важно не только оценить, но и проинтерпретировать частные коэффициенты логистической регрессионного уравнения. При этом подходы к интерпретации нелинейной модели существенно отличаются от линейной модели регрессии /17/. Параметры уравнения бинарной

логистической регрессии модели принадлежности к аудитории Интернета приведены в таблице А.1. приложения А, что позволяет их интерпретировать непосредственно (гр. 1).

С одной стороны, при этом следует говорить, что предсказанный логарифм отношения шанса того, что человек является пользователем Интернета против шанса, что он мужчина равен 0,5713, при условии, что человек имеет нулевой возраст, не учитывая опыт пользования компьютеров, наличие знакомых или родственников, регулярно пользующихся Интернетом, его социальный статус. Каждый прожитый год уменьшает логарифм отношения шанса того, что человек будет пользоваться Интернетом на 0,0201 или 2,0%, обладание навыками работы на компьютере увеличивает логарифм отношения шансов, что человек входит в аудиторию Сети с $-4,5568$ («нет навыков») до 0 («программист»).

Анализ статистики Вальда (гр. 2) позволяет утверждать, что наибольшую значимости в оцениваемой модели играют факторы обладания индивидуумом навыками компьютерной грамотности (61,68), обширность контактов с лицами, регулярно пользующихся Интернетом (47,03), а также социальный статус (20,29). Первоначально включенные в модель признаки: логарифм дохода на одного члена домохозяйства (0,07); наличие домашнего телефона (1,97), мобильной связи (0,20), уровня образования (2,52), национальность (13,95), семейное положение (1,81) и т. п. не являются в Казахстане значимыми для определения вероятности попадания в аудиторию Интернета.

В таблице коэффициентов почти все переменные значимы на уровне значимости 5% (гр. 3). Пожалуй, только коэффициент категории «опытный пользователь» признака компьютерной грамотности не обладает значимостью. Универсальность статистики Вальда позволяет оценить значимость не только отдельных переменных, но и в целом значимость категориальных переменных, несмотря на то, что они дезагрегированы на дихотомические переменные /18/. Хотя показатель социального статуса имеет высокую значимость, будучи развернутым в набор дихотомических признаков, значимость ряда категорий (домохозяйки, служащие силовых ведомств, рабочие и руководители) утрачивается.

Альтернативный способ интерпретации параметров логистической модели заключается в объяснении воздействия на отношение шансов, а не воздействия на логарифм такого отношения. Большинство людей имеет интуитивное понимание отношения шансов. Напомним, что шанс (риск) это отношение между частотой выбора одной альтернативы и частотой ее невыбора. Отношение шансов (Odds Ratio) предсказывает, во сколько раз вероятность выбора одной альтернативы больше (меньше) чем выбора другой альтернативы /19/.

Чтобы перейти непосредственно к уравнению (1) проэкспоненцируем значения коэффициентов регрессии (гр. 4) и 95% доверительные границы (графы 5 и 6). Теперь уже можно говорить о том, что отношение шансов того, что человек, в возрасте 18 лет (минимальный возраст респондента при опросе в формате офлайн) и без учета всех остальных факторов модели пользуется Интернетом – от 0,062 до 1,000 (влияние константы регрессионного уравнения). С увеличением возраста на один год, это отношение должно быть умножено на 0,9801 (или, отношение уменьшено примерно на 2%) или точнее в 0,9496–1,0115 раза. У мужчин отношение шанса того, что он посещает Интернет, будет умножено на 1,771 (1,0173–3,0814 раза). Это, более

понятный для большинства людей способ объяснения, относительно направления и величины воздействия факторов, включенных в уравнении регрессии.

Однако можно идти далее, и говорить о воздействии изменения на одну единицу измерений в каждом факторе (X) на вероятность принадлежности аудитории Интернета. Это будет еще наиболее ясный и очевидный способ описания воздействия на результативный показатель. Однако, поскольку логистическая модель нелинейна в измерении вероятностей, мы должны выбрать контрольную точку для вычисления воздействия изменений единицы в X /20/. Наиболее общий подход при выборе такой точки — среднестатистический человек в исследуемой выборке. Чтобы вычислить вероятность того, что такой человек является интернетчиком, в полученное регрессионное уравнение подставляются средние значения (среднее или мода) и вычисляется предсказанный логарифм отношения шансов. Уравнение (4) дает возможность рассчитать предсказанную вероятность Y для конкретного вектора значений факторов X .

$$Y = \frac{1}{(1 + e^{-(-0,4023 - 2,1135 * 1 + 0,0 * 1 + 0,0574 * 0 + 0,5713 * 1 - 0,0201 * 41 - 0,5586 * 1)})} \quad (4)$$

В рассматриваемой модели эта предсказанная величина равна 0,0346. Следовательно, человек, который является наиболее типичным в данной выборке по возрасту (41 год) – специалист-мужчина, начинающий пользователь без домашнего компьютера, не имеющий знакомых, регулярно посещающих Интернет, имеет предсказанную вероятность того, что он принадлежит к аудитории Интернета, равна 0,0346, или приблизительно 3,5%.

Это полезный подход для понимания полученной модели, поскольку теперь мы можем определять насколько изменится вероятность в том случае, если человек будет младше на один год (в сравнении с типичным в данной выборке), или иметь в коммуникативном окружении одного или двух регулярных пользователей Интернета. Простая замена значений в нелинейном регрессионном уравнении позволит определить искомую вероятность.

В оцениваемой ситуации, мы находим, что сокращение возраста человека на один год в сравнении со средним в выборке повышает вероятность того, что респондент будет посещать Интернет на 0,0007, или 0,07%. Появление в коммуникативном окружении типичного представителя выборки хотя бы одного интернетчика увеличивает вероятность того, что он также будет в аудитории Сети сразу на 0,1116 или 11,16%. Предприниматель с такими же типичными для выборки значениями факторных признаков будет иметь вероятность на 0,0665 или 6,65% выше, чем специалист.

Мы могли также оценить изменение вероятности наступления некоторого события путем увеличения каждого факторного показателя на величину одного стандартного отклонения от среднего значения в выборке. В этом случае, для среднестатистического человека в данной выборке уменьшение возраста на одно стандартное отклонение (15,81 лет) увеличивает вероятность того, что он будет интернетчиком на 0,0124, или 1,24%.

Наконец, для наглядной демонстрации воздействия факторных показателей можно построить график. В этом случае на оси ординат откладываются легко вычисляемые предсказанные вероятности результативного показателя Y_1 («принадлежность аудитории Интернета») по всему диапазону значений каждого факторного показателя X , при использовании остальных X равных средним значениям или дру-

гим константам (например, при исследовании конкретной группы, класса или кластера) /21/. Результаты лучше всего будут показаны в виде кривой функции кумулятивной вероятности выбора альтернатив лицом, принимающим решение.

Рисунок 1 изображает два графика кумулятивной вероятности принадлежности к аудитории Интернета специалиста–мужчины 41 года без домашнего компьютера.

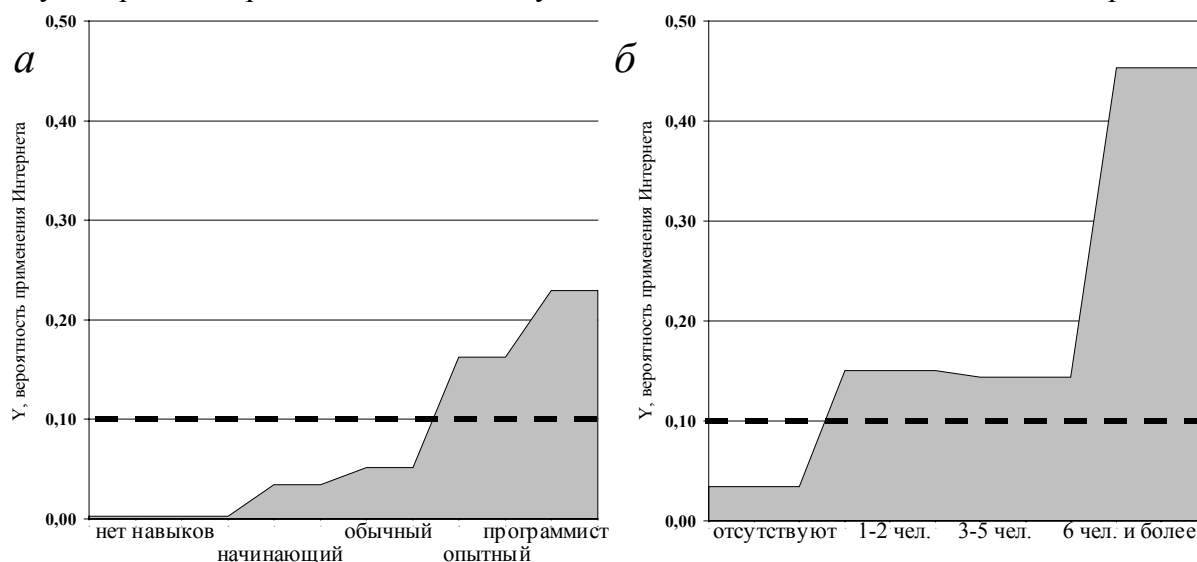


Рисунок 1 – Кривые кумулятивной вероятности принадлежности к аудитории Интернета: а) по уровню компьютерной грамотности и б) количеству знакомых, регулярно посещающих Интернет.

Очевидно, что повышение уровня навыков применения вычислительной техники и соответственно доступ к ней (коллинеарный признак) существенно повышает предсказанную вероятность выше порога 0,10, т. е. попадания лица в аудиторию Интернета. Сходным образом воздействуют на предсказанную вероятность Y_1 наличие в коммуникативном окружении человека хотя бы одного человека, регулярно посещающего Интернет.

Графический метод анализа позволил наглядно диагностировать отклонения оцениваемой логистической модели. В качестве наблюдаемых характеристик, по примеру американского статистика Л. Хамильтона (L. Hamilton), рассматриваются характеристики недостаточности приближения модели (Poorness-of-fit statistic) $\Delta\chi^2 - V^2(2)$, взвешенные на аналог статистики влияния Кука (Cook's influence statistic) – D, используемой для оценки аномальных явлений /22/.

$$V_i^2 = \frac{G_i^2}{(1 - h_i)}, \tag{5}$$

где G_i – значение уклонений (Deviance value), статистика логарифма функции правдоподобия модели для данного наблюдения;

h_i – значение балансировки (Leverage value), измеряющее относительное влияние каждого наблюдения на качество приближения модели /11, pp. 302–304/.

Статистика расстояния Кука отражает степень влияния соответствующего наблюдения на уравнение регрессии¹. Эта величина показывает разницу между вычисленными β – коэффициентами и значениями, которые получились бы после исключения соответствующего наблюдения. В адекватной модели все расстояния Кука должны быть примерно одинаковыми; если это не так, то имеются основания считать, что соответствующее наблюдение (или наблюдения) смещает оценки коэффициентов регрессии /24/.

На рисунке А.1 в приложении А приведена пузырьковая диаграмма. На ней по оси ординат отложены значения показателя недостаточности приближения модели – V^2 , определяемого по уравнению (5), а по оси абсцисс значения предсказанной вероятности результирующего показателя (принадлежность аудитории Интернета). Окружности, соответствующие каждому наблюдению, имеют радиус прямо пропорциональный величине значения статистики влияния Кука (D). На диаграмме можно легко выделить семнадцать аномальных наблюдений, у которых тесты V^2 и D превышают сумму среднего и трех значений среднего квадратичного отклонения («правило трех сигм»), приведенных в таблице А.2 приложения А: «аномальные интернетчики» – 12 наблюдений (1,00% выборочной совокупности) и «аномальные неинтернетчики» – 5 наблюдений (0,42%).

Яркий представитель первого класса пенсионерка с высшим образованием из г. Рудного (ID = 1199), считающая, что она не обладает даже начальными навыками применения вычислительной техники, но при этом регулярно использует Интернет у своих знакомых, утверждая при этом, что у неё нет знакомых регулярно посещающих Интернет ($\text{Prob}_{(\text{Internet})} = 0,0007$). Остальные наблюдения данного класса также соответствуют респондентам, не обладающим компьютерной грамотностью, но имевшим единственный опыт общения с Интернетом (V^2 от 4,23 до 15,06).

Во втором классе выделяется молодой программист с высшим образованием (ID = 235), работающий специалистом в финансовом секторе г. Алма-Аты, имеющий дома компьютер и телефон, общающийся с четырьмя регулярными пользователями Интернета, но никогда не бывавший в Интернете ($\text{Prob}_{(\text{Internet})} = 0,84$). Подобные ему опытные пользователи вычислительной техники молодого возраста, общающиеся со многими регулярными пользователями Интернета, но никогда сами в нем не бывавшие, имеют значения V^2 от 3,65 до 4,74. Исключение из массива данных лишь 17 аномальных наблюдений (1,42% от выборочной совокупности) позволяет существенно улучшить качество оцениваемой модели: индекс отношения правдоподобия увеличился с 47,8% до 61,8%, псевдо-коэффициент детерминации Cox and Snell с 0,273 до 0,316, его дальнейшая модификация Nagelkerke- R^2 с 0,561 до 0,688.

В заключение первого этапа моделирования, произведена оценка устойчивости (робастности²) и приемлемости логистической модели к новым (тестовым) данным. С 1960–х гг. теорию робастности разрабатывали П. Хубер (P. Huber) /25/, Ф. Хампель (F. Hampel) /26/ и многие другие. Из монографий на русском языке, трактующих о робастности и устойчивости статистических процедур, самой ранней и наиболее общей была книга /27/, следующей – монография /28/.

¹ Вышеприведенные переменные, сохраняемые для каждого наблюдения, рассчитываются в модуле логистической регрессии статистического пакета SPSS версии 9.0 /23/.

² От английского слова *robust* – крепкий, трудоемкий, здравый.

Специально вопрос получения несмещенных моделей в логистической регрессии исследован в диссертационном исследовании Г. Гонга (*G. Gong*)¹ в 1982 г. /29/, а его основные выводы опубликованы им позже в обобщающей статье /30/.

Как правило, для оценки устойчивости логистической регрессии исследователи применяют следующие методы: кросс–проверки, группировочной кросс–проверки, скользящего экзамена, бутстреп с различными модификациями. Метод, именуемый в англоязычной литературе *jack-knife* (методы «складного ножа»), получил в литературе на русском языке наименование «скользящего экзамена». Данный метод является одним из самых жестких способов проверки на устойчивость при небольшой вариации обучающих данных. Исходная работа, в которой изложена методология этой меры проверки робастности была опубликована британским статистиком М. Х. Куэнулла (*M. H. Quenouille*) в 1949 г. /31/.

Суть метода заключается в том, что из всей выборки случайным образом извлекается один элемент, а по оставшимся рассчитываются коэффициенты регрессии, потом по полученной логит–модели производится отнесение этого исключенного элемента как нового. Затем элемент возвращается, удаляется другой элемент выборки. Данная процедура повторяется заданное число раз (например, до полного перебора всех элементов выборки). В результате таких действий подсчитываются случаи правильного и неправильного отнесения².

Такой метод оценки устойчивости является наиболее предпочтительным, так как дает меньшую дисперсию оценки вероятности ошибки. Однако метод скользящего экзамена является трудоемким, так как требует многократного построения коэффициентов модели, поэтому его не часто применяют в практике экономико–статистического анализа /32/.

Таблица 2

Проверка устойчивости логистической регрессионной модели принадлежности к аудитории Интернета методом скользящего экзамена

		Исходное				После скользящего экзамена		
		предсказано				предсказано		
		0	1	корректный процент		0	1	корректный процент
А	Б	4	5	6	В	4	5	6
Наблюдается	0	921	160	85,20%	0	924	157	85,48%
	1	13	106	89,08%	1	10	109	91,60%
Всего		934	266	85,58%		934	266	86,08%

¹ Г. Гонг — ученик профессора Б. Эфрона (*B. Efron*), создателя метода *бутстреп* (метод «шнуровки», размножение выборок).

² SPSS 9.0 предоставляет возможность сохранить в файле данных отклонение в значениях коэффициентов логистической регрессии β , при помощи параметра *DFBETA* в опции *SAVE* команды *LOGISTIC REGRESSION*. Данные изменения последуют при исключении специфического элемента выборки. Эти значения вычисляются для каждого параметра в модели, включая константу.

Проценты ошибочных классификаций (когда опрашиваемый, фактически, принадлежит к одной группе, а по логистической модели он отнесен к другой), по методу скользящего экзамена, составили для классов «Аудитория Интернета» и «Вне аудитории Интернета» и в целом по модели, подвергнутой скользящему экзамену, по 14,5, 8,4 и 13,9%, соответственно. Соответствующий процент верных решений — 85,5, 91,60 и 86,08% (см. Таблица 2). Из этих оценок устойчивости может заметить, что построенная бинарная логистическая регрессионная модель обладает вполне приемлемой чувствительностью и достаточной устойчивости при небольшой вариации обучающих данных.

В целом, оцениваемая логистическая регрессионная модель принадлежности к аудитории Интернета обладает хорошим качеством и устойчивостью. Мы выявили основные факторные показатели, влияющие на применение Интернета в Казахстане. При проведении маркетинговой деятельности компаний, предоставляющих услуги доступа к Интернету, на эти факторы следует сосредоточить внимание в первую очередь. Их изменение (за исключением факторного показателя X_9 – возраст респондента) во многом зависит от решений в области отраслевого маркетинга, а также государственного управления и не связано со значительными финансовыми вложениями в национальную информационную инфраструктуру.

Следующим этапом статистического анализа электронного бизнеса в Казахстане является построение регрессионной модели для изучения потребительского выбора электронной коммерции в Казахстане. В отличие от традиционного бинарного подхода мы рассматриваем интенсивность применения электронной коммерции, т. е. регулярность или желание сделать покупки в интернет-магазинах либо заказы, осуществляемые через Сеть, не как множественный или дихотомический качественный признак, а ранжированную (порядковую) качественную переменную. Порядковый признак обычно кодируется как 1, 2, 3, 4 и так далее. Это представление отражает только ранжирование; при этом не известно, в какой степени переход категории от 1 до 2 отличается от 3 до 4. Для такого порядкового результативного признака, используется разновидность множественного логита – порядковая логистическая модель (Ordered Logit Model). Информационной базой для построения модели берется массив с данными интерактивного опроса казахстанских представителей активной аудитории Интернета старше 12 лет, проведенный в ноябре 2000 г. (онлайн-формат), включавший 750 наблюдений ($N = 750$). Краткое описание методов проведения опроса и его демографические характеристики даны в специальной работе /33/.

Корреляционный анализ был применен на предварительном этапе и послужил основой для выбора факторных признаков для дальнейшей экономико-статистической обработки панельных данных. Исследование корреляционной матрицы указало на сильную взаимосвязь отбираемых показателей, что вызвало необходимость снижения размерности признакового пространства. Имеется, по меньшей мере, три основных типа принципиальных предпосылок, обуславливающих возможность практически безболезненного перехода от большого числа исходных показателей состояния анализируемого объекта к существенно меньшему числу наиболее информативных переменных /34/. Это 1) дублирование информации, доставляемой сильно взаимосвязанными показателями; 2) неинформативность показателей, мало меняющихся при переходе от одного объекта к другому (малая вариабельность показателя), что свойственно дихотомическим признакам; 3) возможность агрегирова-

ния, т.е. простого или взвешенного суммирования некоторых физически однотипных показателей.

Одним из наиболее распространенных методов снижения размерности исследуемого признакового пространства является метод *главных компонент*, который позволяет перейти от исходного набора показателей к небольшому числу вспомогательных переменных, по которым он впоследствии мог бы достаточно точно воспроизвести интересующие свойства анализируемого массива данных. Такой подход дает возможность также классифицировать объекты в пространстве первых главных компонент.

Редуцированию признакового пространства были подвергнуты два набора дихотомических признаков – многоальтернативных вопросов: «*возможности работы в Интернете*» и «*интересы в Интернете для личных целей*». Массив подвергся процедуре выделения главных компонент и варимаксного вращения с нормализацией Кайзера – метода ортогонального вращения факторов, который минимизирует число переменных, имеющих большие нагрузки для каждого фактора, что существенно упрощает интерпретацию факторов.

При рассмотрении каждого столбца матрицы нагрузок, полученной при реализации метода главных компонент, была выявлена сущность признаков, влияющих на формирование компоненты, что позволило охарактеризовать эти новые обобщенные показатели. При исследовании каждой строки матрицы нагрузок признак можно рассматривать в пространстве главных компонент, причем координатами признака будут соответствующие ему элементы матрицы нагрузок. Такой подход позволил проанализировать близость признаков в пространстве главных компонент.

Перейдем к содержательной интерпретации полученных главных компонент. При сокращении размерности многоальтернативного вопроса «*интересы в Интернете для личных целей*» в **первой компоненте** выделяется интерес индивидуума к вопросам ведения бизнеса, финансам, информации о товарах и услугах, а также тематике недвижимости. Этот фактор мы можем назвать «*деловое использование Интернета*». **Вторая компонента** отражает интерес к науке, образованию, технике, программному обеспечению, а также к разнообразной справочной информации. В **третьей компоненте** наибольшие нагрузки имеют признаки характеризующие интерес к информации о странах, спорте и туризме. **Четвертая компонента** носит ярко выраженный развлекательный характер (литературы, музыка, анекдоты, игры и эротика), в **пятой** – доминирует заинтересованность в поисковых системах, каталогах ресурсов Интернета и применении электронной почты. И, наконец, **шестая компонента** объединила медицинскую тематику и вопросы трудоустройства. Первые шесть компонент описывают 49,5% дисперсии вопроса.

Редуцирование признакового пространства многоальтернативного вопроса «*возможности работы в Интернете*» позволяет свести его к трем главным компонентам, объясняющим 51,6% дисперсии. В **первую компоненту** с наибольшими нагрузками вошли наличие опыта опрошенного по созданию вэб–страниц, осуществления различных настроек на серверах и вэб–браузере. Этот фактор своего рода индикатор «*квалификации посетителя Интернета*» в программных средствах и возможностях создания ресурсов в Интернете. Во **второй компоненте** наибольшие нагрузки имеют группа признаков, описывающих разнообразные способы и средства общения и получения информации в Интернете, что позволяет назвать выделенный

фактор «коммуникабельностью посетителя Интернета». Третья компонента объединяет тесно связанную между собой практику обучения на специальных семинарах, по вопросам применения Интернета и электронного бизнеса и покупку книг по данной тематике – фактор «стремления к обучению посетителя Интернету».

Такая интерпретация дает важную информацию для понимания процессов осуществления выбора представителями активной аудитории Интернета, далеко выходящую за рамки тех подходов, которые нашли свое воплощение при проведении серии американских научных исследований GVU's WWW User Survey в конце 1990–х гг. /35/. Полученные нами факторы наглядно представляют в соответствии с теорией социального выбора процесс принятия решения о выборе как познавательный процесс, в котором активно взаимодействуют, играющие каждый свою роль, восприятие, вера, отношения, предпочтения и мотивы. Знаменательно, что при факторном анализе, проведенном автором, панельные данных десятой волны опросов упомянутой выше американской программы, в составе анкеты которой также был многоальтернативный вопрос «возможности работы в Интернете», хорошо выделяются те же самые три главные компоненты – квалификации, коммуникативности и стремлению к обучению, проявляющиеся у посетителей Интернета во всем мире.

Подобный подход позволит отказаться от ряда факторных признаков, которые весьма опосредованным способом объясняют процесс принятия решения индивидуумом об участии в электронной коммерции. Таких, например, как часто применяемый в англоязычных научных исследованиях признак «год первого выхода в Интернет», простое суммирование видов работ, практикуемых пользователем и т. п. В свою очередь данный шаг даст возможность включить в логистическую модель более информативные факторные показатели.

После того как на стадии априорного анализа произведен отбор факторов, влияющих на практику применения электронной коммерции, и на основе теории случайной полезности определена порядковая логистическая форма зависимости и собрана исходная статистическая информация, можно перейти непосредственно к построению модели практики участия в электронной коммерции в Казахстане. Описательная статистика качественного порядкового признака «Покупки и заказы в Интернете» приведена в таблице А.3 приложения А. Моделирование применения Интернета производилось по четырнадцати факторам.

На массиве с данными опроса в онлайн формате на однородной совокупности была построена матрица коэффициентов непараметрической корреляции. Особенно интересными представляются такие атрибуты индивидуума как показатели «год первого выхода в Интернет», «практика выхода в Интернет на работе», «регулярность посещения Сети» и «интенсивность использования электронной почты». К числу тесно связанных с электронной коммерцией следует отнести четыре признака, полученные методом главных компонент: квалификация, коммуникабельность, стремление к обучению и деловое использование Интернета пользователем.

Как и предыдущем случае построения регрессионной модели при отборе факторных признаков учитывалась их независимость между собой /36, pp. 308–329/. Рассмотрение матрицы непараметрической корреляции дает основания предположить о наличии некоторой взаимосвязанности между признаками «год выхода в Интернет» и группой признаков отражающих интенсивность применения телекоммуникационных технологий: «выход в Интернет на работе», «регулярность посеще-

ния *Интернета*» и «использование электронной почты» (τ от $-0,287$ до $-0,201$). Значимая отрицательная связь, между этими показателями логически легко объясняется. Поэтому во избежание проявления мультиколлинеарности признак «год выхода в Интернет» был исключен из числа факторных переменных модели.

Функцию вероятности выбора индивидуумом альтернативы – неучастия или различной степени участия в электронной коммерции, описываемой порядковой логистической моделью (иногда называемой в литературе кумулятивной логистической регрессией) можно выразить уравнением отношения шансов /37/:

$$\ln\left(\frac{\text{Prob}(Y > j)}{\text{Prob}(Y \leq j)}\right) = \alpha_j + \sum_{k=1}^K \beta_k X_k, \quad \forall j \in J, \quad (6)$$

Уравнение (6) оценивает параметры константы α_j в количестве $J-1$ для результирующего признака с J категориями и только один коэффициент регрессии β_k для каждого факторного признака X_k . Эти коэффициенты описывают вероятности того, что результирующий признак представлен категорией большей чем j против категории меньше чем или равной категории j из множества J . Константы задают «совокупный логит» для ситуации, когда все факторные переменные нулевые.

Результаты многошагового регрессионного анализа¹ при построении порядковой логистической модели применения Интернета приведены в таблице А.4. приложения А. Интерпретация данных результатов во многом аналогична случаю бинарной логистической регрессии. В оцениваемой модели, удвоенный логарифм функция правдоподобия со знаком минус из уравнения только с константой (начальный $-2LL$) равен 1616. Модель с константой и семью факторными признаками (конечный $-2LL$) имеет -1253 . Различие, или «качество приближения» составляет 363 со степенью свободы 11, существенными при $p < 0,0000$. Значение коэффициент псевдо- R^2 равняется 44,26%, что необходимо признать достаточно высоким.

Таблица 3

Классификационные таблицы логистической регрессионной модели практики применения электронной коммерции

Варианты признака		Предсказано				Всего	корректный процент
		Не собираюсь	Хочу попробовать	Нерегулярно	Регулярно		
А		1	2	3	4	5	6
Фактически	Не собираюсь	2	78	1	0	81	2,47
	Хочу попробовать	4	420	16	0	440	95,45
	Нерегулярно	0	83	83	12	178	46,63
	Регулярно	0	16	21	14	51	27,45
	<i>Всего</i>	6	597	121	26	750	69,20

Результаты классификации результирующего признака представлены в класси-

¹ К сожалению, в применявшейся автором 9-ой версии программного продукта SPSS не реализована порядковая логистическая регрессия, появившаяся только в 10-ой версии (команда PLUM). Поэтому для анализа использовался макрос OLOGIT для SPSS версии 5 и выше. Макрос любезно предоставлен Стефаном Куехнелем (Steffen Kuehnel), профессором университета Гессена (ФПГ). Адрес в Интернете – <http://baserv.uci.kun.nl/~johnh> [Февраль 2002].

фикационной таблице (см. Таблица 3). Общий коэффициент согласованной пере-классификации – 69,20%. Однако из таблицы видно, что реалистично оцениваемая модель описывает наиболее представительные варианты участия в электронной коммерции: «*желание попробовать электронную коммерцию*» (удельный вес 58,67% в выборочной совокупности) – коэффициент конкордации равен 95,45% и «*нерегулярное применение электронной коммерции*» (удельный вес 23,73%) – 46,63%.

При этом оцениваемая модель плохо описывает «младший» вариант – «*неучастие в электронной коммерции*» (коэффициент конкордации – 2,47%), а также менее весомый «старший» вариант – «*регулярное применение электронной коммерции*» (27,45%), такая особенность является характерной для данного класса логистических моделей /36, pp. 288–307/.

Параметры уравнения оцениваемой порядковой логистической модели, в которые включены семь независимых признаков, приведены в таблице А.4. приложения А. Все факторные признаки являются значимыми по тесту Вальда (гр. 3). Анализ оценок значимых параметров регрессионной модели дает возможность соотнести между собой влияние отдельных факторных признаков. Надо отметить, что дихотомический признак «*доступа к Сети дома*» и ряд категорий множественных признаков «*образование пользователя Интернета*» и «*использование электронной почты*» представляются менее значимыми, чем признаки, полученные методом главных компонент. При этом отмечается, что коэффициенты регрессии индикаторных переменных являются отрицательными. Это свидетельствует о том, что в сравнении с контрастным вариантом признака практики уровня образования – «*высшее/незаконченное высшее*» и интенсивности электронной почты – «*более 10 писем в день*», проявление всех остальных вариантов характеризуется меньшим уровнем использования электронной коммерции.

Сопоставление параметров регрессии $\exp(\beta)$ (гр. 5) дает детальную информацию о влиянии факторных признаков на результативный. Следует отметить, что на практику применения электронной коммерции не сказывается наличие доступа к Интернету дома (X_6) — коэффициент непараметрической корреляции 0,178, но в логит-модели получивший незначимый параметр регрессии близкий нулю. В отличие от него признак наличия выхода дома (X_5) проявляется более существенным ($\exp(\beta) = 1.270$), хотя его коэффициент Кендэлла был мал – 0,103. Данный фактор в проявляется на важном пороговом переходе «*хочу попробовать–нерегулярно*», поскольку константа при варианте «*хочу попробовать*» близка нулю (–0,819).

В несколько большей мере в модели доминирует признак «*образование посетителя Интернета*» (X_2) – от 0,236 у лиц с образованием ниже среднего до 1,000 для получивших полное или незаконченное высшее образование. Этот показатель также имеет довольно низкий коэффициент связи¹ с практикой покупки и заказов в Интернете (0,125). Фактор интенсивности использования электронной почты (X_8) не демонстрирует однозначного воздействия на практику электронной коммерции, а один из параметров (вариант «*6-10 писем в день*») даже незначим.

Наиболее важными значащими факторными признаками выступают квалификация (X_{11}), стремление к обучению (X_{13}) и настрой на деловое использование Ин-

¹ Подробнее о взаимосвязи множественных качественных признаков см., например /38/.

тернета (X_{14}), а также в некоторой степени коммуникабельность пользователя (X_{12}). Поскольку, вышеперечисленные признаки получены методом главных компонент, полезно сравнить максимальные, минимальные значения их произведения на экспоненцированные параметры регрессии (см. таблицу А.5 приложения А).

Очевидно, что наибольшее влияние на отношение шансов выбора альтернативы имеет квалификация индивидуума (размах между произведениями максимального и минимального значения фактора на экспоненту $\beta - 11,236$) и настрой на деловое использование Интернета ($R - 12,658$). Несколько меньшее доминирование проявляется у признака «стремления к обучению Интернету» (10,38) и совсем мала вариация практики электронной коммерции от влияния коммуникабельности посетителей Интернета (0,939).

Наконец, сравним значения вероятностей различных альтернатив лицом, принимающим решение, при нулевых величинах факторных, когда проявляются константы различных вариантов порядкового логистического регрессионного уравнения: 0,04 («неучастие»), 0,63 («желание участвовать»), 0,30 («нерегулярное участие») и 0,03 («регулярное участие»). Сопоставление полученных вероятностей с частотами вариантов результативного признака (см. таблицу А.3 приложения А) обнаруживает относительно слабую степень качества предсказания неучастия в электронной коммерции за счет большей вероятности выпадения шансов «хочу попробовать» и «нерегулярно».

Однако, принимая во внимание сложность восприятия сумм вероятностей выбора альтернатив, описываемых порядковым показателем, обусловленных нелинейностью логит-модели, мы прибегнем к наглядному способу интерпретации оцениваемой модели по графикам кривых кумулятивной вероятности вариантов практики применения электронной коммерции (см. Рисунок 2).

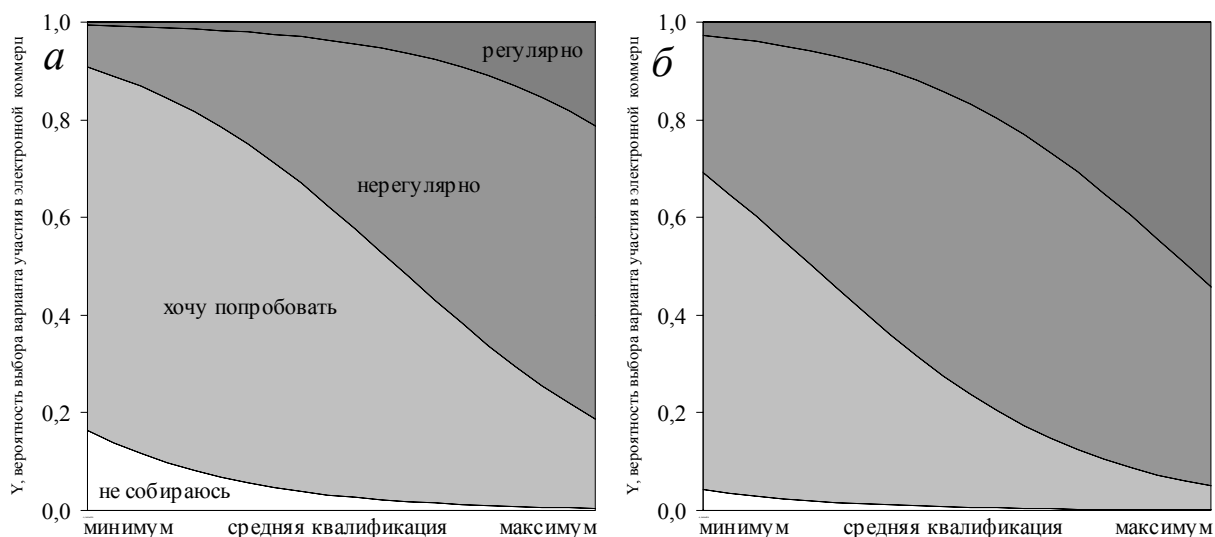


Рисунок 2 – Кривые кумулятивной вероятности практики применения электронной коммерции от квалификации посетителя Интернета при: а) средней деловой направленности применения Сети и б) высокой направленности применения Сети.

Оцениваемая модель была подвергнута процедуре «испытания» устойчивости способом группировочной кросс-проверки. Данный способ оценки робастности вы-

бран из-за его относительно меньшей трудоемкости. Кросс-проверка представляет собой процедуру оценки точности прогнозирования с помощью данных из специальной *тестовой выборки* (используется также термин *кросс-проверочная выборка*) путем сравнения точности прогноза с той, что достигается на *обучающей выборке*, то есть на выборке, по которой собственно строилась модель. Итак, модель строится по обучающей выборке, а точность прогноза на основании модели оценивается по тестовой выборке. В идеале, когда имеется достаточно большая выборка, включающая тысячи объектов, часть наблюдений (например, половину или две трети) можно использовать для обучающей выборки, а оставшиеся наблюдения - для тестовой. Если на тестовой выборке модель дает результаты того же качества, что и на обучающей выборке, то говорят, что модель хорошо прошла кросс-проверку /39/.

В нашем случае выборочная совокупность была разбита на 10 групп (кросс-проверочных выборок) по 75 случайно отобранных наблюдений в каждой. Затем было произведено построение 10 порядковых логистических моделей на обучающей выборке, которая включала всю выборочную совокупность за исключением наблюдений одной из тестовых выборок. Потом каждая из этих моделей была испытана на тестовой группе, которая не была включена в массив данных при ее построении. Результаты проверки робастности модели по результатам классификации приведены ниже (см. Таблица 4).

Таблица 4

Проверка устойчивости логистической регрессионной модели практики применения электронной коммерции методом группировочной кросс-проверки

Варианты признака		Предсказано на кросс-проверочных выборка				Всего	корректный процент
		Не собираюсь	Хочу попробовать	Нерегулярно	Регулярно		
А		1	2	3	4	5	6
Фактически	Не собираюсь	0	27	50	4	81	0,00
	Хочу попробовать	0	82	321	37	440	18,64
	Нерегулярно	0	6	84	88	178	47,19
	Регулярно	0	0	16	35	51	68,63
	<i>Всего</i>	<i>0</i>	<i>115</i>	<i>471</i>	<i>164</i>	<i>750</i>	26,80

Проценты ошибочных классификаций (когда опрашиваемый, фактически, принадлежит к одной группе, а по логистической модели он отнесен к другой), по методу с группировочной кросс-проверки, составили для классов «*неучастие в электронной коммерции*», «*желание попробовать электронную коммерцию*», «*нерегулярное применение электронной коммерции*» и «*регулярное применение электронной коммерции*» и в целом по модели по 100,0, 81,4, 52,8, 31,4 и 73,2%, соответственно. Соответствующий процент верных решений — 0,0, 18,6, 47,2, 68,6 и 26,8%.

Обращает внимание, что при уменьшении величины общего коэффициента конкордации модели, в сравнении с оцениваемой моделью (см. Таблица 3), удельный вес корректно классифицированных вариантов класса «*нерегулярное применение электронной коммерции*» немного вырос (с 46,6 до 47,2%), а качество классификации класса «*регулярное применение электронной коммерции*» увеличилось суще-

ственно (с 27,5 до 68,6%). Эти два класса наиболее интересны для целей маркетинга электронного бизнеса.

Из вышеприведенных оценок устойчивости может заметить, что построенная порядковая логистическая регрессионная модель обладает вполне приемлемой чувствительностью, но недостаточной устойчивостью при большой вариации обучающих данных. В связи с этим применение оцениваемой порядковой логистической модели может носить качественный, описательный характер, что вполне достаточно для того, чтобы сделать адекватные выводы, логично вытекающие из статистического анализа, и принять должные решения, которые лягут в основу маркетинговой деятельности по развитию электронной коммерции в Казахстане.

К началу XXI в. опыт использования интернет-магазинов имели 40 тыс. казахстанцев (55% активной аудитории в ноябре 2000 г.). Около 9 тыс. использовали Сеть для совершения покупок регулярно за последние три месяца. Наибольшая популярность электронной коммерции наблюдалась у потребителей в возрасте до 25 лет, а также специалистов и менеджеров среднего звена, последние составляли 50% посетителей интернет-магазинов. Чаще всего источником информации о товарах и заказах в Сети являются поисковые системы и каталоги, за ними следуют баннеры и ссылки с других серверов.

По популярности формы оплаты у всей аудитории, прибегавшей к электронной коммерции, лидируют наличные по факту доставки. По степени удобства оплаты эта часть аудитории предпочитает наличные (34%) и пластиковые карты (30%). Также наблюдается превышение сторонников электронных платежей (15%) над банковскими переводами (12%).

Основной причиной неиспользования розничной электронной коммерции активной аудиторией является отсутствие на сайте достаточной информации о предлагаемом товаре или отсутствие компьютера дома. Большая часть посетителей не пользуются интернет-магазинами, потому что они не довольны предлагаемой системой оплаты.

Главным плюсом, по мнению всей аудитории интернет-магазинов, является фактор экономии времени. Самой актуальной проблемой для активной аудитории интернет-магазинов является недостаток информации о товарах. Самым же важным параметром для интернет-магазина является удобство навигации и визуального восприятия. Рекомендуется для продвижения услуг электронной коммерции шире использовать возможности прямой адресной рассылки по электронной почте. Нацеливание данной рекламы следует производить на целевые группы, интересующиеся вопросами ведения бизнеса, финансов, информации о товарах и услугах, а также тематики недвижимости. При этом следует учитывать степень квалификации потенциального участника, что позволит повысить уровень и продолжительность взаимодействия субъектов электронного бизнеса.

Выводы

Если при рассмотрении поведения потребителей применять широко распространенную в маркетинговой деятельности развитых стран методологию исследования потребления, базирующуюся на принципах логического позитивизма, на основе которой была разработана модель поведения потребителей, предлагаемая в трудах Д. Ф. Энджела и Р. Д. Блэкуэлла /40/ и по ее критике в работах Хоукинса Д., то сле-

дует признать, что потребление услуг электронного бизнеса необходимо отнести к тому виду потребления, понять который можно на путях реализации «кросс-культурных и глобальных маркетинговых стратегий» при условии «снятии преград на пути межкультурных маркетинговых коммуникаций».

Исходя из известного тезиса о том, «что интеграция стран в мировую рыночную экономику достигла такой степени плотности, что нет необходимости рассматривать потребителя в рамках одной ‘субкультуры’» /41/ логично провести сравнение результатов поведения казахстанской аудитории Интернета и выводов по эмпирическому изучению опыта западных интернетчиков, изложенных в работах /4, 5, 6, 7 и 8/.

Результаты подобного компаративного исследования, дают основания, с некоторой поправкой на ряд местных факторов, для формулирования следующих заключений. Во-первых, следует внимательно осваивать международный опыт организации и ведения электронного бизнеса с целью осуществления комплекса работ по обеспечения инновационной деятельности, создания эффективного инструментария, позволяющего в полной мере задействовать имеющейся потенциала электронного бизнеса. Поэтому поведение потребителя в данном секторе народного хозяйства следует рассматривать как деятельность, производимую на глобальном рынке.

Во-вторых, давая обобщенную оценку процесса принятия решения о выходе в Интернет, т. е. приобщения к одному из видов электронного бизнеса, следует рассматривать процесс выбора в рамках модели случайной полезности. Проведенный экономико-статистический анализ позволил установить, что в процессе выбора лица, принимающим решение о выходе в Интернет проявляется ряд закономерностей, представленных в форме решающих правил.

В-третьих, главную роль в описанном процессе играют навыки работы на компьютере (вклад в логит-модель – 44%), обширность коммуникативного окружения, прибегающего к услугам всемирной сети (вклад – 33%), а также доступность индивидууму компьютера на работе или по месту учебы. Социальный статус, т. е. по сути, важнейший атрибут лица, принимающего решение, выступает существенным фактором (14%) в данном процессе. В первую очередь этот вывод относится к руководителям, предпринимателям, специалистам и студентам. Пол, возраст и обладание домашним компьютером следует рассматривать в качестве вспомогательных атрибутов индивидуума в оцениваемой модели.

В-четвертых, обширная группа признаков, которые, как казалось, могли быть включены в состав решающих правил модели случайной полезности выхода в Интернет, по результатам эмпирического анализа оказались незначимыми. Уровень образования, национальность, имущественное положение семьи посетителя Интернета, наличие мобильных средств связи и банковских платежных карточек, а также регион проживания практически не оказывают какого-либо влияния на выбор лица, принимающего решение об участии в электронном бизнесе в Казахстане.

Следовательно, в маркетинговой политике интернет-провайдеров и при проведении иных мероприятий по совершенствованию деятельности субъектов электронного предпринимательства необходимо учитывать, выявленные в ходе исследования поведения потребителей услуг отмеченные выше особенности процесса выбора.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Всемирная история экономической мысли: в 6 т./ Под ред. Ю. Я. Ольсевича. – М.: Мысль, 1994. Т. 5: Теоретические и прикладные концепции развитых стран Запада (послевоенный период). – С. 571.
2. *Kahneman D., Tversky A.* Prospect Theory: An Analysis of Decisions Under Risk // *Econometrica*. – 1979. – Vol. 47, pp. 263–291.
3. Structural Analysis of Discrete Data and Econometric Applications / *Charles F. Manski and Daniel L. McFadden* (Eds.). – Cambridge: The MIT Press, 1981. – p. xxvii. Адрес в Интернете – <http://emlab.berkeley.edu/users/mcfadden/discrete.html> [Февраль 2002].
4. *Moe W., Fader P.* Capturing Evolving Visit Behavior in Clickstream Data // Working Paper, The Wharton School, University of Pennsylvania. – January 2001. Адрес в Интернете – <http://www-marketing.wharton.upenn.edu/sitesearch/ideas/pdf/00-003.pdf> [Февраль 2002].
5. *Lohse Gerald L., Bellman Steven, Johnson Eric J.* Consumer buying behavior on the Internet: Findings from panel data // *Journal of Interactive Marketing*. – 2000. – Vol. 14. – Issue 1. pp. 15–29. Адрес в Интернете – <http://ecom.gsb.columbia.edu/Papers/99wvtm2.pdf> [Февраль 2002].
6. *Emmanouilides C., Hammond K.* Internet usage: Predictors of active users and frequency of use // *Journal of Interactive Marketing*. – 2000. – Vol. 14. – Issue 2. pp. 17–32. Адрес в Интернете – <http://www3.interscience.wiley.com/cgi-bin/abstract/72001011/START> [Февраль 2002].
7. *Xenikou A., Hammond K. A., Svennevig M.* Attributions and Motivations to use the Web: the Role of Perceived Stability // *Future Media Working Papers of London Business School*, 2000. Адрес в Интернете – [http://forum.london.edu/lbsfacpubs.nsf/workingPapersTableView/8BB2298C3A79F1AA80256A24003DF62B/\\$File/00-802.pdf](http://forum.london.edu/lbsfacpubs.nsf/workingPapersTableView/8BB2298C3A79F1AA80256A24003DF62B/$File/00-802.pdf) [Февраль 2002].
8. *Hammond K. A., Turner P., Bain M.* Internet Users versus Non-users: Drivers in Internet Uptake // *International Journal of Advertising*. – 2000. – Vol. 19. – No. 5. pp. 665–681. Адрес в Интернете – <http://www.warc.com/open/publications/IJoA.asp> [Февраль 2002].
9. Аудитория и окружение Интернета в Казахстане, май 2001. Алматы.: Акстис Системс Азия, 2001. С. 7. Адрес в Интернете - <http://www.actis.kz/experience/RMKI0105.Short.pdf> [Февраль 2002].
10. *Кендэлл М. Дж., Стюард А.* Статистические выводы и связи / Пер. с англ. – М.: Главная редакция физ.-мат. литературы, 1973.
11. *Neruis M.* SPSS 7.5 Statistical Algorithms. – SPSS, 1996. Адрес в Интернете – <http://www.spss.com/tech/stat/algorithms/logregre.pdf> [Февраль 2002].
12. Прикладная статистика и основы эконометрики: Учебник для вузов Изд. 2-е, испр.: В 2 тт: Т. 2: *Айвазян С. А.* Основы эконометрики. – М.: Юнити-Дана, 2001. – С. 187–193.
13. *Cox D. R., Snell E. J.* The analysis of binary data, 2nd ed. – London: Chapman and Hall. 1989.
14. *Nagelkerke, N. J. D.* A note on a general definition of the coefficient of determination // *Biometrika*. – 1991. – Vol. 78. – p. 691–692.

15. *Darlington R. B.* Regression and linear models. – New York: McGraw-Hill, 1990. – p. 449.
16. *Chatterjee S., Hadi A. S., Price B.* Regression Analysis by Example / Third Edition. – New York: John Wiley & Sons, 2000. – pp. 328–330. Адрес в Интернете – <http://www.ilr.cornell.edu/~hadi/RABE> [Февраль 2002].
17. *Сошникова Л. А., Тамашевич В. Н., Уебе Г., Шефер М.* Многомерный статистический анализ в экономике: Учеб. пособие для вузов / Под ред. проф. В.Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – С. 312–315.
18. Анализ социологических данных с применением статистического пакета SPSS. – Новосибирск, НГУ, 2001. Адрес в Интернете – http://www.ieie.nsc.ru/~meta-nsk/docs/Rostovtsev/book_datan/Content.htm [Февраль 2002].
19. *Garson D. G.* Statnotes: An Online Textbook. – Raleigh, NC State University, 2000. Адрес в Интернете – <http://www2.chass.ncsu.edu/garson/pa765/logistic.htm> [Февраль 2002].
20. *Long J. S.* Regression Models for Categorical and Limited Dependent Variables / Advanced Quantitative Techniques in the Social Sciences No 7. – Thousand Oaks: Sage Publications, 1997. – pp. 79-82. Адрес в Интернете – <http://www.indiana.edu/~jsl650/cldvtoc.pdf> [Февраль 2002].
21. *Fox J.* Applied Regression Analysis, Linear Models, and Related Methods. – Sage Publications, 1997. – pp. 481–482. Адрес в Интернете – <http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression> [Февраль 2002].
22. *Hamilton L.* Regression with Graphics. – Wadsworth Publishing Company, 1995. – pp. 218–242. Адрес в Интернете – <http://www.stata.com/bookstore/rwg.html> [Февраль 2002].
23. SPSS Base 9.0 Syntax Reference Guide. – Prentice Hall, 1999. – p. 32-33.
24. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. В 2-х книгах – 2-е изд. Кн. 1. – М.: Финансы и статистика, 1987. – 351 С.
25. *Хьюбер П.* Робастность в статистике. – М.: Мир, 1984. – 304 с.
26. *Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В.* Робастность в статистике. Подход на основе функций влияния. – М.: Мир, 1989. – 512 с.
27. *Орлов А. И.* Устойчивость в социально-экономических моделях. – М.: Наука, 1979. – 296 с.
28. *Смоляк С. А., Титаренко Б. П.* Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. – М: Статистика, 1980. – 208 с.
29. *Gong, G.* Cross validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression, PhD thesis, Stanford University, 1982. 3781-1982G.
30. *Gong G.* Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression // Journal of the American Statistical Association. – 1986. – Vol. 81. No. 393. – May. – pp. 108–113.
31. *Quenouille M. H.* Approximate tests of correlation in time series // Journal of the Royal Statistical Society Series B, 1949. – Vol. 11, p. 18-84.
Quenouille M. H. Notes on bias in estimation // Biometrika Soc. Ser. B. – 1956, Vol. 43, p. 353–360.

32. Степанов В., Заяц А. Анализ состояния банка // Банковские технологии. – 1996. – № 8. Адрес в Интернете – <http://www.bizcom.ru/rus/bt/1996/nr8/12.htm> [Февраль 2002].
33. Описание отчета о состоянии и перспективах развития аудитории Интернета в Казахстане. – Алматы: Актис Системс Азия, 2000. С. 5–7.
34. Дубров А. М., Мхитарян В. С., Трошин Л. И. Многомерные статистические методы: Для экономистов и менеджеров. – М: Финансы и статистика, 1998. С. 134-170.
35. Graphic, Visualization, and Usability Center at Georgia Tech. GVU's 10th WWW User Survey. – 1998. Адрес в Интернете – http://www.gvu.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/graphs.html [Февраль 2002].
36. Hosmer D., Lemeshow L. Applied Logistic Regression / 2nd Edition. – NY: Wiley & Sons, 2000. Адрес в Интернете – <http://www.amazon.com/exec/obidos/ASIN/0471356328> [Февраль 2002].
37. Powers D., Xie Y. Statistical Methods for Categorical Data Analysis. – Academic Press, 1999. Адрес в Интернете – <http://www.la.utexas.edu/research/faculty/dpowers/book> [Февраль 2002].
38. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1983.
39. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. - Springer-Verlag, 2001. – pp. 214-216. Адрес в Интернете – <http://www-stat-class.stanford.edu/~tibs/ElemStatLearn/> [Февраль 2002].
40. Engel James F., Blackwell Roger D., Miniard Paul W. Consumer behavior / Eighth Edition. – The Dryden Press, Harcourt Brace College Publishers, 1995. / Русский перевод 1997 г., Энджел Д. Ф., Блэкуэлл Р. Д., Миниард П. У. Поведение потребителей / Пер. с англ. – Спб.: Питер, 1999. – 768 с.
41. Алешина И. В. Поведение потребителей: Учебное пособие для вузов. – М.: Фаир-Пресс, 2000. – 384 с.

ПРИЛОЖЕНИЕ А

Таблица А.1

Оценки параметров логистической регрессионной модели принадлежности к аудитории Интернета 95%

Факторные показатели	β – регрессионный коэффициент	Статистика Вальда	Значимость β	exp(β)	уровень значимости для exp(β)	
					нижняя	верхняя
А	1	2	3	4	5	6
Навыки пользования компьютером (X_1)		61,6751	,0000			
нет навыков	-4,5568	26,1502	,0000	0,0105	0,0018	0,0602
начинающий пользователь	-2,1135	6,6082	,0102	0,1208	0,0241	0,6053
обычный пользователь	-1,6963	4,2115	,0402	0,1834	0,0363	0,9266
опытный пользователь	-0,4254	,2446	,6209	0,6535	0,1211	3,5266
Кол-во знакомых, пользующихся Интернетом (X_3)		47,0346	,0000			
1-2 человек	1,5974	11,2576	,0008	4,9402	1,9431	12,56
3-5 человек	1,5423	9,2949	,0023	4,6754	1,7346	12,6016
6 и более человек	3,1386	39,5509	,0000	23,072	8,6752	61,3614
Наличие домашнего компьютера (X_4)	1,0574	6,9314	,0085	2,8788	1,3102	6,3252
Респондент мужчина (X_8)	0,5713	4,0831	,0433	1,7705	1,0173	3,0814
Возраст респондента, лет (X_9)	-0,0201	1,5577	,2120	0,9801	0,9496	1,0115
Социальный статус респондента (X_{12})		20,2904	,0266			
военнослужащий	-7,1557	0,1426	0,7057	0,0008	0	∞
другое	-2,3784	0,0016	0,9681	0,0927	0	∞
пенсионер/инвалид	-1,1161	0,8196	0,3653	0,3276	0,0292	3,6701
безработный, временно не-работающий	-1,5898	6,7819	0,0092	0,2040	0,0617	0,6748
домохозяйка	-0,0833	0,0210	0,8847	0,9201	0,2986	2,8351
рабочий	-0,3205	0,4375	0,5083	0,7258	0,2808	1,8761
руководитель (и их заместители)	0,2847	0,1543	0,6945	1,3294	0,3211	5,5039
служащий	-1,1626	5,6783	0,0172	0,3127	0,1202	0,8135
предприниматель	0,5838	1,1382	0,2860	1,7928	0,6134	5,2392
специалист	-0,5586	1,3709	0,2417	0,5720	0,2246	1,4571
<i>Константа</i>	-0,4023	0,1569	0,6921			
Initial -2 Log Likelihood = 799,1 Model -2 Log Likelihood = 416,8 Goodness of Fit = 2158,0 Cox & Snell - R^2 = 0,273 Nagelkerke - R^2 = 0,561 Chisqu. DF Prob. 382,302 20 ,000 R-Square (%) = 44,26						

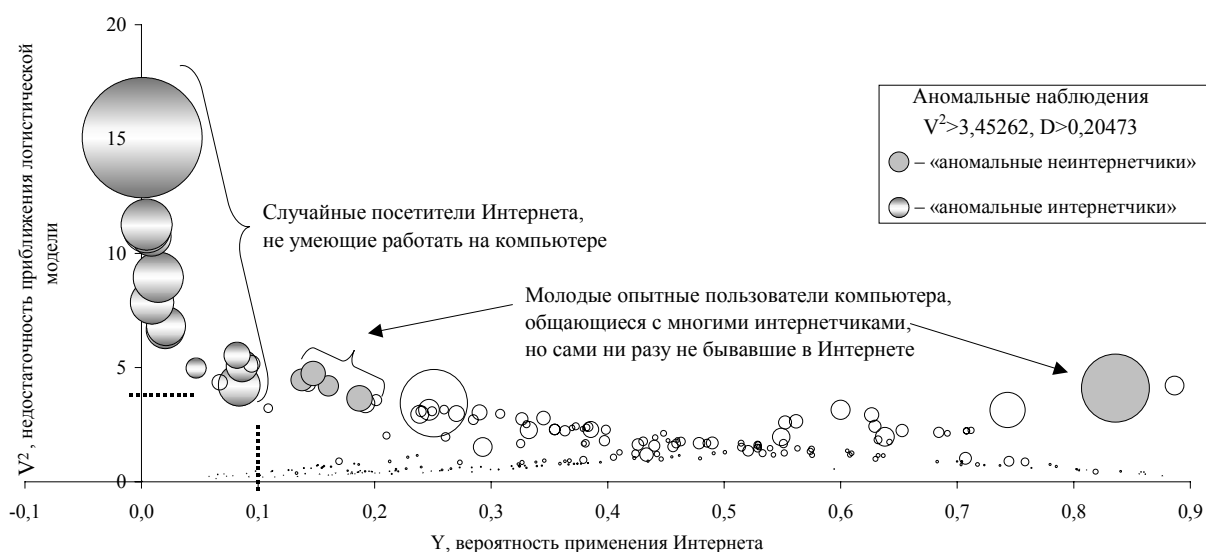


Рисунок А.1 – Недостаточность приближения модели (V^2), взвешенная на статистику влияния Кука (D) против предсказанной вероятности принадлежности к аудитории Интернета (Y).

Примечание: радиус окружностей прямо пропорционален величине значения статистики влияния Кука (D).

Таблица А.2

Описательная статистика диагностики отклонений логистической регрессионной модели применения Интернета

Наименование показателя	Минимум	Максимум	Среднее арифметическое (\bar{X})	Среднее квадратичное отклонение (σ)	Медиана	$\bar{X} + 3\sigma$ (гр. 3 + 3 * гр. 4)
А	1	2	3	4	5	
Недостаточность приближения модели (V^2)	0,0000 4	15,057 84	0,36307	1,02985	0,0141 4	3,45262
Статистика влияния Кука (D)	0,0000 0	1,2146 8	0,01630	0,06281	0,0000 3	0,20473

Таблица А.3

Описательная статистика признака «Покупки и заказы в Интернете» по данным в формате онлайн за 2000 г.

Варианты признака	Частоты	Частоты, %	Среднее арифметическое (\bar{X})	Среднее квадратичное отклонение (σ)	Медиана	Мода
А	1	2	3	4	5	6
Не собираюсь (1)	81	10,80	2,27	0,74	2,00	2
Хочу попробовать (2)	440	58,67				
Нерегулярно (3)	178	23,73				
Регулярно (4)	51	6,80				
Итого	750	100,00				

Таблица А.4

Оценки параметров логистической регрессионной модели практики применения электронной коммерции по данным в формате онлайн за 2000 г.

Факторные показатели	β – регрессионный коэффициент	Стандартная ошибка ($\bar{\varepsilon}$)	Статистика Вальда	Значимость β	$\exp(\beta)$
А	1	2	3	4	5
Образование посетителя (X_2)					
Ниже среднего (до 9 класса)	-1,443381	0,567823	6,461540	0,011023	0,236128
Среднее/среднее профессиональное	-0,323267	0,221669	2,126732	0,144748	0,723780
Выход в Интернет дома (X_5)	0,239220	0,163867	2,131140	0,144333	1,270258
Использование e-mail (X_8)					
не пользуюсь	-0,330665	0,467494	0,500293	0,479371	0,718446
до 2 писем в день	-0,679104	0,250378	7,356653	0,006681	0,507071
2-5 писем в день	-0,241198	0,231972	1,081126	0,298444	0,785686
6-10 писем в день	-0,167962	0,285493	0,346124	0,556316	0,845386
Квалификация посетителя (X_{11})	0,911811	0,090546	101,407743	0,000000	2,488825
Коммуникабельность посетителя Интернета (X_{12})	0,660619	0,087144	57,468140	0,000000	1,935990
Стремление к обучению в Интернете (X_{13})	0,637968	0,080635	62,596581	0,000000	1,892631
Деловое использование (X_{14})	0,402475	0,081517	24,377066	0,000001	1,495522
Константа					
не собираюсь	3,127174	0,246340	161,151451	0,000000	22,809433
Хочу попробовать	-0,819436	0,222745	13,533625	0,000234	0,440680
Нерегулярно	-3,590747	0,278482	166,255245	0,000000	0,027578
-2 Log-Likelihood of Model with Constants only = 1616,086, -2 Log-Likelihood of full Model = 1252,780 Chisqu. DF Prob. %-Reduct 363,307 11,000 ,000 ,225 R-Square (%) = 44,26					

Таблица А.5

Анализ значений отношения шансов модели практики применения электронной коммерции от факторных признаков, полученных методом главных компонент по данным в формате онлайн за 2000 г.

Наименование факторных признаков	$\exp(\beta)$	$x_{\min} * \exp(\beta)$	$x_{\max} * \exp(\beta)$	размах, R (гр. 3 – гр. 2)
А	1	2	3	4
Квалификация посетителя Интернета	2,489	6,677	17,914	11,236
Коммуникабельность посетителя Интернета	1,936	2,628	3,567	0,939
Стремление к обучению Интернету	1,893	6,164	16,537	10,373
Деловое использование Интернета	1,496	5,162	17,820	12,658