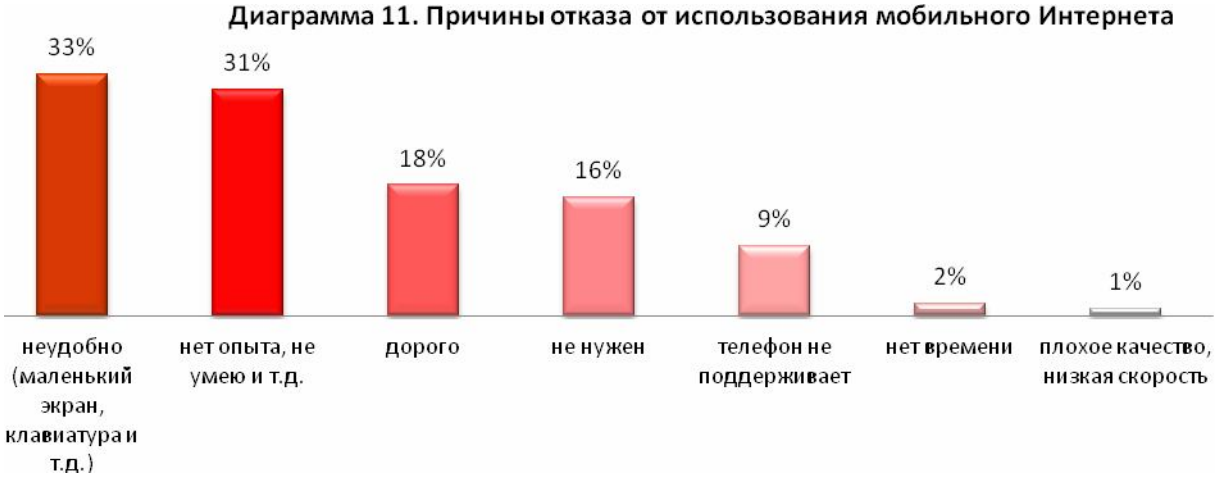
Основания для отказа от использования Мобильного Интернета

Почему же не все пользователи мобильных телефонов используют услугу Мобильного Интернета? Оказывается, для 33 % это неудобно в силу различных причин, связанных с телефоном (маленький экран, клавиатура т.д.), 31 % просто не умеет выходить в Интернет с мобильного устройства, для 18 % это дорого, 16 % эта функция не нужна и т.д.



Выводы

Таким образом, нам удалось нарисовать примерный портрет основного потребителя Мобильного Интернета: преимущественно мужчины до 44 и женщины до 35 лет, студенты, владельцы собственного бизнеса, наёмные сотрудники, работающие в СМИ, ИТ, сфере услуг и здравоохранении со средним и высоким доходами. Это социально активные люди, которые пользуются в Интернете, в первую очередь, поисковиками, почтовыми сайтами, а уже заодно с ними и социальными сетями, просматривая так же и новостные сайты.

Этот сегмент более чем реален, и по мере дальнейшего проникновения Интернета в жизнь общества будет расти, особенно учитывая постоянное снижение цен на мобильный доступ в Интернет и обновление парка мобильных телефонов. Учитывая, что примерно пятая часть отказов от использования Мобильного Интернета происходит из-за дороговизны услуги, можно смело предположить, что при снижении тарифов количество пользователей Мобильного Интернета увеличится. То же самое можно сказать и о тех, кто отвергает использование мобильного Интернета из-за неудобства использования или отсутствия этой функции в телефоне: нельзя игнорировать отчётливый тренд на рынке мобильных телефонов в сторону выпуска всё более многофункциональных устройств, использование Интернета в которых скоро станет второй по использованию опцией после звонков.

Всё вышесказанное говорит о невозможности и далее не замечать стремительно формирующийся сегмент потребителей мобильного Интернета. Данный метод более оперативно, чем существующие методики опросов, получать репрезентативные результаты.

Метод множественного восстановления данных

Кутлалиев Асхат Хасянович
Международный институт маркетинговых и социальных исследований «ГфК-Русь»

Введение

Наиболее распространенным инструментом социальных исследований является массовый опрос. При этом крайне редко можно встретить количественное исследование, база которого не содержит пропущенных данных. Неполнота массива данных происходит вследствие целого комплекса причин: недостижимости отдельных респондентов, необходимых для поддержания репрезентативности в выборке, пропущенных ответов некоторых респондентов на отдельные вопросы вследствие усталости, недостатка знаний, социальной чувствительности вопросов.

Кроме того, с распространением компьютерных технологий сбора информации, особенно CATI и CAWI, исследователи всё чаще сталкиваются с ростом доли незавершенных интервью, особенно при проведении сложных комплексных опросов. Всё вышеперечисленные причины ведут к смещению выборочного массива, неполноте получаемой информации, недостоверности результатов, что является достаточно серьезными проблемами. Исследователи вынуждены искать способы учета и работы с пропущенными данными.

К сожалению, не существует достаточно простых и общепринятых правил как это делать. В каких случаях можно не обращать внимания на пропуски информации, в каких случаях исключать из дальнейшего анализа наблюдения, содержащие пропуски данных и в каких – стараться тем или иным способом заполнять пропуски. Например, часть пропусков заполняется в процессе редактирования данных после их сбора и ввода. Но таким образом можно достоверно заполнять пропуски лишь в тех вопросах, которые могут быть логически проверены ответами на другие вопросы. При больших выборках с небольшим (не более 5 %) количеством наблюдений с пропущенными данными неполные наблюдения (даже при наличии всего лишь одного пропуска) часто исключают из анализа. Наблюдения с пропусками в данных в широко распространенном статистическом пакете SPSS удаляются по умолчанию в большинстве процедур регрессионного и многомерного анализа. Даже при большей доли наблюдений с пропусками в данных многие исследователи предпочитают исключать наблюдения из анализа целиком, полагая, что заполнение пропусков ведет к искажению регрессионных коэффициентов, мер связи и смещению статистических оценок. И для таких предпочтений в отдельных случаях есть основания. Но более правильным кажется альтернативный путь: выявить природу пропусков, попытаться восстановить пропущенную информацию и сравнить результаты с теми, которые получены при анализе массива полных наблюдений.

В ряде случаев процедура восстановления пропусков в данных является обязательным элементом анализа, например, если исследователь сознательно планирует контролируемые пропуски в данных. Такой дизайн исследования может включать в себя опрос по отдельным блокам вопросов только части выборки, опрос различных групп людей из одного выборочного массива с помощью частично совпадающих анкет, охват мониторингом различных групп респондентов в разные периоды времени и/или разной тематикой.

Типы пропусков в данных и методы их восстановления

Согласно классификации Литтла и Рубина[1], в данных могут встречаться:

1) полностью случайные пропуски (missing completely at random (MCAR));
2) случайные пропуски (missing at random (MAR));
3) неигнорируемые пропуски (non-ignorable missingness).

Последний тип пропусков, наиболее сложный для анализа и восстановления, мы в данной работе не рассматриваем, а сосредоточиваемся на первых двух типах пропусков данных.

С полностью случайными пропусками (MCAR) мы имеем дело тогда, когда пропуски случайно распределены в массиве данных по всем переменным. Наличие MCAR можно проверить статистически t-тестом или хи-квадрат тестом. В модуле SPSS Missing Values Analysis (MVA) есть опция Little's MCAR test, которая на базе статистики хи–квадрат проверяет данные на MCAR. Если в тесте наблюдается *незначимый* уровень критерия, то мы имеем дело с полностью случайными пропусками данных.

Случайные пропуски (MAR) данных встречаются тогда, когда пропуски в массиве данных случайно распределены не по всем переменным, а только внутри каких-либо определенных подгрупп переменных. Например, пропуски чаще встречаются среди ответов мужчин, чем ответов женщин, но внутри подгрупп распределены случайны. Такое распределение пропусков в данных случается гораздо чаще, чем MCAR.

Для борьбы с этими двумя видами пропусков применяют восемь основных классов методов:

1) анализ полных наблюдений (listwise deletion);
2) методы, использующие доступную информацию (pairwise deletion);

[1] Little R.J.A., Rubin D.B. Statistical Analysis with Missing Data. New York: John Wiley & Sons, 1987.

3) подстановка среднего по выборке (mean substitution);
4) метод хот-дек (hot deck);
5) регрессионный анализ (regression);
6) оценка с помощью максимизации правдоподобия (maximum likelihood estimation);
7) подстановка с помощью факторного анализа (factor analysis substitution);
8) модель множественного восстановления данных (multiple imputations method).

Два первых метода широко распространены в исследовательской практике.

Смысл первого метода заключается в удалении всех респондентов с пропущенными данными (listwise deletion). При вычислении сложных моделей с большим числом переменных этот метод может привести к большому объему исключенных наблюдений, в ряде случаев до двух третей исходного массива данных. Во втором методе переменные для наблюдений с пропущенными значениями не включаются в проводимый анализ по необходимости (pairwise deletion). Например, корреляции между каждой парой переменных вычисляются из наблюдений, которые для этих переменных содержат допустимые значения, что позволяет использовать больший объем доступной информации по сравнению с анализом полных наблюдений. Применение этих способов методически правомерно, если пропуски в данных распределены полностью случайно (MCAR).

Методы с 3-го по 7-ой используют принцип однократной подстановки восстановленных тем или иным способом данных и могут использоваться, если пропуски распределены случайно (MAR). В основе методов лежит принцип вычисления и подстановки взамен каждого пропущенного значения одного нового значения. Кратко рассмотрим эти методы.

Суть метода **подстановки среднего по выборке** понятна из его названия: вместо пропусков подставляется среднее значение по данной переменной. Метод видоизменяет изначальное распределение, делая его более сконцентрированным около среднего значения и уменьшая дисперсию. Как следствие, из-за подстановки среднего по выборке систематически недооцениваются ковариации переменных. Этот метод дает лучшие результаты, чем методы, основанные на исключении наблюдений, однако он также ведет к смещенным результатам.

Метод хот-дек (hot deck) имеет ряд модификаций. Наиболее простой вариант — сортировка респондентов по ключевым переменным, тогда респонденты со схожими ответами находятся рядом друг с другом. В качестве ключевых чаще всего выступают социально-демографические переменные, но можно использовать и другие переменные, имеющие корреляции с переменной с пропущенными данными. При восстановлении пропущенные значения переменной заимствуются из предыдущего наблюдения. В отдельных модификациях хот-дек респонденты сортируются для каждого признака при помощи разных наборов переменных для предупреждения взаимозависимостей между восстанавливаемыми значениями. Наиболее продвинутой модификацией считается хот-дек со случайным отбором значений из подгруппы схожих респондентов. Респонденты делятся на подгруппы по ключевым переменным. Внутри каждой подгруппы данные для замены пропущенных ответов отбираются случайным образом, но с тем же распределением, что и все ответы в подгруппе. Главным недостатком метода хот-дек является наличие взаимосвязей между восстановленными значениями и занижение значений дисперсии. Снизить уровень зависимости до незначимых величин возможно лишь при очень большом числе подгрупп, что подразумевает большие объемы выборок. Тем не менее, метод широко применяется в работе национальных статистических организаций многих стран, включая страны Евросоюза, Россию и США.

Регрессионный анализ. В зависимости от типа данных используются либо множественная линейная, либо логистическая регрессия. На базе наблюдений, не содержащих пропущенных данных, вычисляются коэффициенты регрессии, и далее с их помощью восстанавливается пропущенное значение зависимой переменной. Если не говорить об обычных проблемах регрессии, таких как мультиколлинеарность, гомоскедастичность и т.д., то можно обозначить две проблемы, связанные с её использованием для восстановления пропущенных данных. Во-первых, из-за самой природы регрессии мы полностью исключаем случайные вариации. Это, например, проявляется в том, что при одинаковом наборе значений независимых переменных мы будем подставлять в разные наблюдения одни и те же восстановленные значения. Это ведет к тому,

что при большой доле пропущенных значений становится очень заметным смещение результатов по направлению к средним оценкам. Для борьбы с этим используется метод случайной подстановки, при котором к вычисленному значению прибавляются случайные величины, например, из набора остатков уравнений регрессии на полных наборах данных. Во-вторых, используя в уравнение регрессии слишком большой набор независимых переменных, мы рискуем моделировать шум вместо каких-то осмысленных значений переменных.

Метод **максимизации правдоподобия (maximum likelihood estimation, MLE)** предназначен специально для работы с недостающей информацией на больших выборках. Другое название метода, под которым он известен пользователям SPSS, — Expectation Maximization algorithm, алгоритм максимизации математического ожидания.

Идея метода основана на допущении: то, что «случилось» в исследовании, то и должно было произойти, то есть реализовались события, наиболее вероятные в исследуемой системе и, одновременно, наиболее соответствующие применяемому инструменту исследования. Поэтому, все неизвестные данные, которые мы пытаемся восстановить, надо искать таким образом, чтобы они как можно лучше согласовывались с уже имеющимися данными в базе данными. Тогда оценки пропущенных данных и будут «наиболее правдоподобными».

Процесс итераций состоит из двух шагов. На шаге Expectation вычисляется ожидаемое значение логистической вероятности математического ожидания для всех данных на основе условного распределения пропущенных значений относительно наблюдаемой информации и текущих оценок интересующего нас параметра из предыдущей итерации. На шаге Maximization максимизируется конечная функция для получения новых оценок параметра. После определенного набора итераций этих двух шагов решение с некоторой точностью приходит к локальному максимуму логистической вероятности математического ожидания наблюдаемых значений. Как и в регрессионном анализе, использование метода максимизации правдоподобия приводит к снижению дисперсии и риску смоделировать шум, однако в настоящее время он получил наиболее широкое применение благодаря тому, что достаточно давно получил реализацию в популярных статистических пакетах (в SPSS, начиная с 8-ой версии).

Метод **подстановки значений с помощью факторного анализа** был разработан сравнительно недавно, в самом конце XX века. Первоначально Майклом Веделем и Вагнером Камакурой[1] была предложена модификация алгоритма факторного анализа при наличии большого числа пропущенных данных. Но оказалось, что, используя алгоритм максимизации правдоподобия, можно решить и обратную задачу — подстановку пропущенных значений исходных переменных на базе полученных латентных переменных, т. е. факторов. Этот метод основан на предположении, что все факторы можно вычислить на основе известных, наблюдаемых значений переменных. Метод работает с различными типами данных, что является большим преимуществом. Разработчики метода особенно подчеркивают, что предложенная ими модель подстановки оправдана даже в случае пропуска 50 % данных и дает более достоверные результаты, чем анализ полных наблюдений, метод, использующий доступные данные, и подстановка среднего по выборке.

Метод множественного восстановления данных

Среди методов восстановления пропущенных данных наилучшим на данный момент является метод множественного восстановления (multiple imputations), предложенный в конце 1970-х Дональдом Рубином[2]. От предыдущих этот метод отличает сложность и ресурсоёмкость, что ставит перед практикующими исследователями серьезные методологические трудности, препятствующие его грамотной практической реализации. А публикации на русском языке, посвященные данному методу, можно пересчитать по пальцам одной руки[3,1]. В первую очередь

[1] Kamakura W.A., Wedel M. Factor Analysis and Missing Data // Journal of Marketing Research. 2000. Vol. 37. No. 4: Nov. P. 490–498.

[2] Rubin D.B. Inference and Missing Data // Biometrica. 1976. Vol. 63. P. 581–592.

[3] Косьяненко А.В. Опыт восстановления пропущенной рыночной информации на основе Байесовского подхода: Препринт WP16/2007/02. М.: ГУ–ВШЭ, 2007.

это вызвано тем, что до недавнего времени в статистическом пакете SPSS, который *de facto* является в России стандартом в области анализа данных в социальных науках, этот метод отсутствовал. Другая причина, сдерживающая применение метода, состоит в том, что в его основу положены редко встречающиеся в практике прикладных социальных исследований байесовский подход и алгоритм Монте-Карло с применением марковских цепей для статистического моделирования. Поэтому закономерен вопрос: стоит ли в данном случае полученный результат затраченных усилий.

Суть метода множественного восстановления данных заключается в том, чтобы одновременно генерировать несколько значений пропущенной величины вместо того, чтобы заменять пропущенную информацию одним значением. Не ограничиваясь однократным вычислением ожидаемого (среднего) значения для пропущенной информации, можно случайным образом подставить значения, вычисленные на основе предсказанного распределения переменной. Для практического использования всех восстановленных таким образом данных генерируются наборы баз данных с различными вариантами подстановок пропущенных данных.

Метод многократной подстановки включает несколько этапов. На первом этапе необходимо выбрать модель, с помощью которой будут предсказываться пропущенные значения. На втором этапе производится несколько вариантов подстановки значений, при этом для каждого варианта генерируется своя база данных с набором восстановленных данных. Завершается алгоритм анализом всех полученных баз данных с подставленными значениями и объединением оценок интересующих нас параметров в общие окончательные показатели (см. Рис.) с помощью алгоритма Рубина[2].

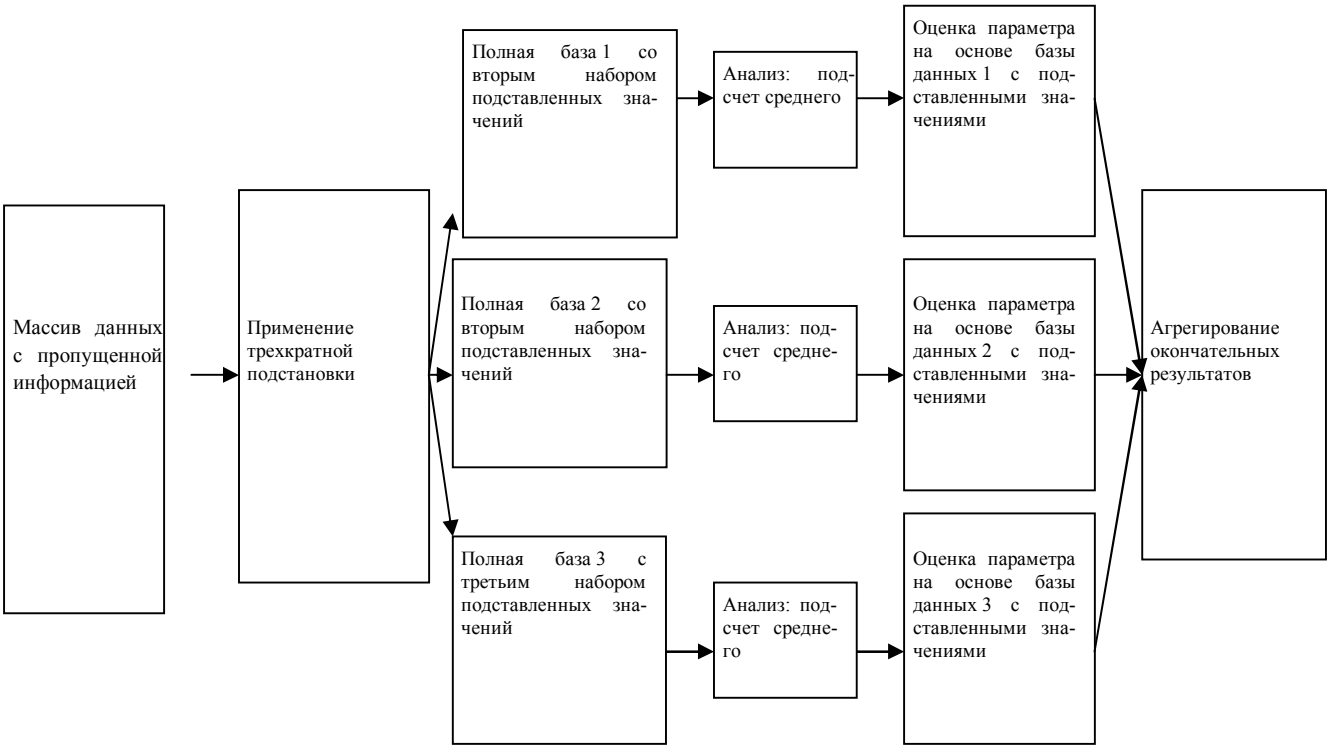


Рис. Схема анализа данных с использованием метода множественного восстановления данных

Каким образом происходит агрегирование данных и оценка их статистических параметров[3]? Допустим, мы создали m наборов данных с восстановленными пропусками. Для каждого набора мы должны вычислить оценки средних и стандартные ошибки (оценки дисперсий). Обозна-

[1] Зангиева И.К. Решение проблемы неполноты данных массовых опросов // Российская социология завтрашнего дня / Под. ред. Г.В. Иванченко, И.С. Чирикова; факультет социологии ГУ–ВШЭ. М., 2008. С. 84–95.

[2] Rubin D.B. Multiple Imputation for Nonresponse in Surveys. New York: J. Wiley & Sons, 1987.

[3] Schafer J.L. Multiple imputation: A primer // Statistical Methods in Medical Research. 1999. Vol. 8: No. 1. P. 3–15.

чим оценку интересующего нас параметра (например, коэффициента в уравнении регрессии), как $\hat{Q}_j$, полученное из j-ого набора данных (j=1,2,...,m), а U_j — стандартная ошибка параметра $\hat{Q}_j$. Тогда агрегированная оценка среднего будет равна:

$$\overline{Q} = \frac{1}{m}\sum_{j=1}^{m}\hat{Q}_j.$$

Для оценки агрегированной стандартной ошибки сначала вычисляется внутригрупповая дисперсия:

$$\overline{U} = \frac{1}{m}\sum_{j=1}^{m}U_j.$$

и межгрупповая дисперсия:

$$B = \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{Q}_j - \overline{Q}\right)^2.$$

Тогда общая дисперсия будет равна:

$$T = \overline{U} + \left(1+\frac{1}{m}\right)B.$$

Доверительные интервалы вычисляются стандартно, но с учетом того, что используется значение статистики Стьюдента с df степенями свободы, вычисляемой по формуле:

$$df = (m-1)\left(1+\frac{m\,\overline{U}}{(m+1)B}\right)^2.$$

Проверка нулевой гипотезы Q=0 выполняется сравнением соотношения $t = \overline{Q}/\sqrt{T}$ с тем же распределением Стьюдента.

Как было доказано разработчиками метода, в большинстве случаев достаточно небольшого числа подстановок — от трех до десяти. Эффективность оценки, основанной на m-ом количестве подстановок, можно вычислить по формуле[1]:

$$\varepsilon = \left(1+\frac{\gamma}{m}\right)^{-1},$$

где γ — доля пропущенной информации в массиве данных. Наглядное представление об эффективности даёт следующая таблица.

Таблица

Эффективность оценки пропущенной информации
методом множественного восстановления данных

	Эффективность оценки (%)			
m	γ =0.1	γ =0.3	γ =0.5	γ =0.7
3	97	91	86	81
5	98	94	91	88
10	99	97	95	93

Для эффективного вычисления среднего, дисперсии и ковариации (или корреляции), используется алгоритм максимизации правдоподобия (EM), позволяющий учесть все значения массива данных, в том числе и частично пропущенные. Для генерации значений для пропущенной информации обычно используются процедуры, имеющие в своей основе цепи Маркова и метод Монте-Карло.

Метод работает следующим образом. Приращение данных начинается со случайного приписывания значений пропущенной информации, после чего новые оценки параметра извлекаются из предыдущего распределения Байеса, основанного на приписанных данных. Алгоритм под-

[1] Schafer J.L. Analysis of Incomplete Multivariate Data. London: Chapman & Hall, 1997.

становки данных состоит из двух шагов. На первом шаге новые значения для пропущенной информации приписываются исходя из условного распределения наблюдаемых данных и последних приписанных значений параметров. На втором шаге новые значения для параметров симулируются путем извлечения их из последующих распределений Байеса наблюдаемых и последних приписываемых значений для пропущенной информации. Начальные значения параметров для первых подстановок можно получить с помощью алгоритма максимизации правдоподобия. Многократные подстановки производятся путем применения цепей Маркова и метода Монте-Карло. Проведение этой процедуры для массива данных с большим числом переменных может занять много времени.

Метод множественного восстановления данных довольно сложен как в техническом, так и в содержательном плане, поэтому вполне закономерным является вопрос о тех преимуществах, которые этот метод имеет по сравнению с другими. Был выполнен вычислительный эксперимент[1] на 1000 тестовых массивов данных объемом порядка 300 наблюдений в каждом и сгенерированными пропущенными данными. Во всех массивах даны был одинаковый процент пропущенной информации, случайно распределенной по одним и тем же законам распределения. Массивы данных обрабатывались различными способами борьбы с пропущенными значениями: подстановка среднего, подстановка на основе регрессионного анализа, модель максимизации правдоподобия (EM) и модель множественных подстановок. Сравнение метода множественного восстановления данных с другими показало, что он позволяет избежать систематических ошибок в оценке коэффициентов регрессионных моделей, а также превосходит другие методы при большом проценте пропущенных данных. Можно считать, что в настоящее время модель многократных подстановок, с учетом возможности его практического применения (начиная с 17-ой версии SPSS) и качества получаемых результатов, является наилучшим из имеющихся. В отличие от других рассмотренных нами методов, он учитывает неточность подстановок, что приводит к получению более близких к реальности результатов анализа.

Применение метода в практике исследований не должно ограничиваться только вынужденным восстановлением пропущенных данных *post hoc*[2]. При использовании современных компьютерных методов сбора информации появилась необходимость сокращать время интервью для увеличения доли полностью заполненных анкет. В таких случаях приходится сознательно закладывать в дизайн исследования пропуски в данных, например, случайным образом исключая отдельные вопросы или блоки вопросов для разных респондентов, а затем для способов восстановления пропущенной информации использовать метод множественного восстановления данных.

Заключение

Пропущенная информация в массивах данных социологических и маркетинговых исследованиях является скорее правилом, чем исключением. Поэтому обязательным этапом исследования является процедура редактирования данных. Хотя часть информации может быть восстановлена на основе заполненных анкет, большая часть пропусков либо оставляется «как есть», либо заполняется исследователями на основании тех или иных соображений. В ряде случаев исследователь вставляет на место пропущенных данных какие-то осмысленные, на его взгляд, цифры и коды. В других случаях исследователи, стараясь избежать субъективного подхода, прибегают к различным методам восстановления пропущенной информации.

В работе были рассмотрены различные типы пропусков в данных и методы их восстановления. Для широкого применения рекомендован метод множественного восстановления данных, предложенный в конце 1970-х Дональдом Рубином. Метод достаточно сложен, как с точки зрения используемого математического аппарата, так и содержательной интерпретации результатов, что сдерживает его широкое применение в исследованиях, несмотря на его явные преимущества над другими методами восстановления данных. Например, он позволяет избежать систематических

[1] Ramaswamy V., Raghunathan T.E., Cohen S.H., Ozcan K. A Multiple Imputation Approach of Missing Data in Marketing Research: Unpublished working paper / University of Michigan, Department of Business Administration. 2001.

[2] Rubin D.B. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations // Journal of Business & Economic Statistics. 1986. Vol. 4. No. 1.

ошибок в оценке коэффициентов регрессионных моделей, а также превосходит другие методы при большом проценте пропущенных данных. Удобная реализация данного метода в последних версиях широко распространенного статистического пакета SPSS позволяет рекомендовать его как наилучший на текущий момент метод восстановления данных.

Использование кластерного анализа в парадигме Data Mining для изучения структуры рынка труда

Мальцева Анна Васильевна
Алтайский государственный университет

Рынок труда является неизменно актуальной проблематикой научных изысканий различного дисциплинарного уровня. Эта тема представляется в целом безграничной ввиду постоянного развития объекта, что требует поиска все новых теоретических, методных и технологических решений анализа его состояний. Неизменной при любых трансформациях рынка труда, вызванных политическими или экономическими причинами, остается сущностная его характеристика — соотношение предложения и спроса рабочей силы. Такая постановка вопроса довольно очевидна и все же именно она может стать отправной точкой определения структуры рынка труда с несколько непривычной стороны: со стороны описывающих рынок труда данных, содержащихся в базах данных государственных служб занятости. Подобные хранилища эмпирического материала непривычны для социолога, но необходимы при анализе структуры рынка труда, поскольку содержат в себе всестороннее описание объекта исследования (исключением является разве что вероисповедание обращающихся граждан или наличия у них родственников за рубежом). *Благодаря такому подходу рынок труда как систему можно описать на макроуровне с помощью современных информационных технологий. Когда макроуровень будет взят в размерности, позволяющей выделять надындивидуальные структурные образования, масштаб которых ограничен системой рынка труда.*

Использование сведений из баз данных развивается в рамках информационного подхода к анализу данных, традиционно в бизнесе для описания бизнес-процессов[1]. Данный подход позволяет изучать объект и содержащиеся в нем закономерности в его естественном состоянии. Это, кроме всего прочего, позволяет еще избежать директивности в определении базовой модели анализа объекта исследования, поскольку она в дальнейшем «подстраивается» под данные и используется для уточнения выявленных закономерностей. Информационный подход является одним из методологических оснований парадигмы анализа данных Data Mining или «глубокого анализа данных». В литературе данная парадигма еще называется методикой извлечения знаний, к которым также относят Knowledge Discovery in Databases (KDD, поиск знаний в базах данных). Если рассматривать методический вопрос с учетом последующих технологий реализации аналитических решений, то важно уточнить иерархию обозначенных выше терминов: «Data Mining является ядром методики» Knowledge Discovery in Databases[2]. Процесс извлечения знаний из баз данных (KDD) — это процедура «получения из данных знаний в виде зависимостей, правил, моделей»[3], включает в себя этапы выборки данных, очистки и трансформации, моделирования и интерпретации полученных результатов. При использовании парадигмы Data Mining возможно обнаружение «в сырых данных ранее неизвестных, нетривиальных»[4] зависимостей или знаний. «Знания должны описывать новые связи между свойствами, предсказывать

[1] Колин К.К. Информационный подход как фундаментальный метод научного познания // Межотраслевая информационная служба / М.: ВИМИ, 1998. Вып. 1 (102). С. 3–17; Мальцева А.В. О результатах применения частной методики сегментации рынка труда // Вестник Самарского государственного университета. 2010. №7; Мальцева А.В. Социологический анализ рынка труда: поиск новых источников эмпирических данных // Сборник материалов VI Международной научно-практической конференции «Наука и современность – 2010» / Под общ. ред. С.С. Чернова. Новосибирск: Изд-во НГТУ, 2010. Ч. 2. С. 36–39; Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2010.

[2] Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2010. С. 40–43.

[3] Паклин Н.Б., Орешков В.И., 2010. С. 41.

[4] Паклин Н.Б., Орешков В.И., 2010. С. 42.