

СТРАТЕГИЯ ИССЛЕДОВАНИЯ¹



ВЛАДИМИР ПАНИОТТО,
генеральный директор компании
КМИС, профессор университета
«Киево- Могиланская академия»,



НАТАЛЬЯ ХАРЧЕНКО,
исполнительный директор КМИС
paniotto@kmis.kiev.ua
<http://www.kiis.com.ua>

ЭКСПЕРИМЕНТ СОЦИАЛЬНЫЙ – целенаправленное воздействие на социальную действительность для ее усовершенствования: во имя науки (менее тяжелый случай) или во имя народа (более тяжелый случай).

Е. Головаха. Психология и другие смешные науки. Альтернативный словарь. – М.: Смысл, 2012. – 142 с.

Немного терминологии

Под стратегией исследования мы будем понимать основные подходы к достижению цели исследования. Например, исследователям необходимо было выяснить, кто добрее – мужчины или женщины. Это можно сделать разными способами – (1) провести опрос, выясняя, что будут делать респонденты, если увидят заблудившегося ребенка; (2) уплатить ребенку, чтобы он делал вид,

что заблудился, и фиксировать, кто чаще к нему подходит – мужчины или женщины; (3) изучать документы благотворительных организаций и т.п. Выбор одного из этих подходов и есть стратегия исследования. Как синоним используют также понятие стратегический план исследования².

Исследования бывают разведывательными, описательными (или дескриптивными) и объяснительными. Планирование

¹ Статья представляет собой раздел книги авторов «Методы опроса», которая планируется к изданию в этом году.

² К сожалению, студенты часто путают стратегический план исследования с организационным планом и вместо описания подходов к проверке гипотез излагают последовательность действий для организации исследования (разработка выборки, печать анкет и т.п.).

каждого из типов исследования имеет свои проблемы, но наиболее сложным является планирование объяснительного исследования. Объяснительное исследование должно объяснить, почему разные люди или группы людей имеют различные значения тех или иных характеристик, что влияет на изучаемые свойства социальных общностей. К тому же разведывательные и описательные исследования в чистом виде используются очень редко, обычно разведывательные исследования используют качественные методы, а описательные являются частью объяснительных исследований. Например, кандидатам в президенты очень интересно знать величину их рейтинга и рейтинга конкурентов (описательное исследование), но этого им недостаточно, их интересует также, какие факторы влияют на рейтинг с тем, чтобы использовать эту информацию в своей предвыборной кампании. Например, им важно знать, в каких регионах проживает основная часть их избирателей, какое у них образование, уровень доходов и т.п., чтобы правильно построить рекламную кампанию. Т.е. нужны не только рейтинги, но и корреляции рейтингов с регионом проживания, образованием, уровнем дохода и другими характеристиками избирателей. Объяснительное исследование имеет целью выяснение корреляционных связей между теми или иными характеристиками объекта исследования, проверку гипотез о причинно-следственных связях.

Одним из основных способов проверки гипотез является эксперимент. Рассмотрим следующий пример, связанный с проверкой теории «козла отпущения». Эта теория объясняет расовые и этнические предубеждения тем, что при возникновении сложностей и неприятностей человек хочет вы-

местить зло на ком-то, кто более доступен. «Поскольку закономерные объекты нашей агрессии обычно недоступны (учителя, начальство, управляющий банком и т. д.), мы перемещаем всю порожденную фрустрацией агрессию на удобную альтернативу. Это может быть человек, группа людей, животное или даже неодушевленный объект. В истории можно найти немало примеров, когда во времена экономического спада агрессия в форме активной дискриминации направлялась на тех, кто рассматривался в качестве «законной» мишени. На волне антисемитизма в Германии 1930-х годов евреев обвиняли в упадке немецкой экономики. На юге Америки негры когда-то служили «козлами отпущения» в неурожайные годы. Чем хуже был урожай хлопка, тем больше негров подвергалось линчеванию»³. Проверка этой теории осуществлялась экспериментально. «Социальные психологи Миллер и Бугельский провели, например, следующий опыт. Группу подростков, в которой было несколько японцев и мексиканцев, вывезли в летний лагерь. Затем руководство лагеря сознательно создало ряд трудностей, вызвавших у ребят состояние фрустрации (напряженности). Японцы и мексиканцы не имели к этим трудностям никакого отношения, тем не менее враждебность против них выросла, товарищи вымещали на них свое раздражение.»⁴

Есть и другой подход. В нашем примере вместо реального эксперимента можно провести исследование, репрезентативное для данной страны, выделить группу людей, которые испытывают сложности, группу людей, которые не испытывают сложностей, и сравнить их уровень ксенофобии. Если у тех, у кого есть неприятности в жизни, уровень ксенофобии выше, чем у благопо-

³ см. Теория «козла отпущения», козел отпущения. Психология. А – Я: Словарь-справочник / М. Кордуэлл. - Пер. с англ. К. Ткаченко. - М.: ФАИР-ПРЕСС, 2000. - 448 с.

⁴ И.С. Кон. Психология предрассудка. «Новый мир», М, 1966, № 9

лучных людей, гипотеза подтвердилась. В опросе, проведенном нами в феврале 2011, мы задавали респондентам вопрос об их уровне счастья и задавали также ряд вопросов об отношении к различным этническим группам (была использована шкала Богардуса). На основании вопросов об отношении к этническим группам был рассчитан индекс ксенофобии⁵, который изменяется от 1 до 7 (1 – минимальный уровень ксенофобии, 7 – максимальный). У счастливых людей значение индекса было равно 4.28, а у несчастливых – 4.09 (разница небольшая, но статистически значимая). Так что гипотеза получила некоторое эмпирическое подтверждение. Такой подход называется квази-экспериментальной или аналитико-экспериментальной стратегией. Экспериментатор не оказывает непосредственного воздействия на эксперимент, а пользуется уже существующими группами для изучения интересующих его процессов.

Именно этот подход доминирует в социологии. Собственно, зачем ставить эксперименты, когда сама жизнь уже поставила миллионы экспериментов, требуется только отобрать нужные группы⁶. Поэтому квази-экспериментальный подход иногда называют также *ex post facto* эксперимент, т.е. эксперимент, использующий уже произошедшие события.

Здесь и далее мы будем говорить об эксперименте в широком смысле, имея в виду любую объяснительную стратегию - и ре-

альный эксперимент и квази-эксперимент. В большинстве случаев все сказанное о реальном эксперименте касается и квази-эксперимента.

Как проверяется наличие причинно-следственных связей (проверка каузальных гипотез).

Как уже было сказано, главной целью объяснительного исследования является выяснение связей между теми или иными характеристиками объекта исследования, проверка гипотез о причинно-следственных связях. Часто причинно-следственные связи называются каузальными (от английского *causal* – причинный, не путать с *casual*⁷).

Эксперименты и квази-эксперименты являются средством проверки каузальной гипотез. Существуют три необходимые условия для утверждения с определенной долей уверенности, что отношения между двумя переменными является каузальным и что событие или признак А является причиной события (или признака) В.

1) Первое условие относится к временной последовательности и формулируется следующим образом: причина должна предшествовать следствию, А должно произойти раньше В.

2) Второе, необходимое условие для обоснованного вывода о причинной зависимости В от А заключается в том, что события

⁵ Методология построения индекса изложена в статье В.Паниотто. Динаміка ксенофобії і антисемітизму в Україні (1994-2007) - – Соціологія: теорія, методи, маркетинг. 2008, N.1, с. 197-214

⁶ Заметим, что квази-эксперимент - это более гуманный метод проведения исследований, чем реальный эксперимент (см.также определение Е. Головаха в эпиграфе). И хотя Е.Головаха издевается также и над квази-экспериментом («Квази-эксперимент – наиболее популярный метод социологического исследования, позволяющий получать квазисоциологическое знание о квазисоциальных явлениях.» Е.Головаха. Психология и другие смешные науки. Альтернативный словарь. – М.: Смысл, 2012. – 142 с.), мы намерены показать, что при правильной организации он является наиболее адекватным методом проверки социологических гипотез

⁷ Один из авторов получил приглашение на прием в посольство Великобритании с указанием на *dress code -business casual* (стиль одежды, допускающий некоторые отклонения от строгой деловой одежды) и долго думал над этой фразой, не понимая, при чем здесь причинность. Потом уже присмотрелся внимательно к слову *casual* и понял свою ошибку

А и В должны быть статистически связаны друг с другом. То есть, А и В встречаются вместе намного чаще, чем это возможно за счет случайных совпадений или изменение А приводит и к изменениям В. Если возможная причина и следствие статистически не связаны друг с другом, одно не может быть причиной другого.

3) Третье, необходимое условие каузального вывода состоит в том, что не должно быть правдоподобного альтернативного объяснения для появления В кроме появления А.

Например, если все студенты, который прочли нашу книгу, сдали успешно экзамен по методологии проведения опросов населения, то можно предположить, что чтение книги – причина хорошей оценки на экза-

менах. Но если эти студенты также посещали лекции, семинары, читали другие книги, то нет уверенности, что хорошая оценка на экзаменах – следствие чтения именно нашей книги, есть и альтернативные объяснения (посещение лекций, например).

Третье условие труднее всего соблюсти (особенно это относится к квазиэкспериментам),

так как кажущееся наличие связи между А и В может быть обусловлена действием третьим событием (третьей переменной) С, которая и вызывает изменения В. И доказать, что таких признаков С не существует, очень трудно, мы только можем выдвигать разные гипотезы об альтернативах, проверять их и показывать, что они не влияют на А.

Таблица 1
Политические ориентации владельцев кошек и собак (% взрослого населения Украины, готовых голосовать за те или иные партии)

Партии	Кого из этих видов домашних животных Вы держите дома?		
	собак	кошек	Разность
Партия регионов Украины (В. Янукович)	28.6	34.5	-5.9*
Блок Юлии Тимошенко	15.9	12.9	2.9
Блок Ющенко «Наша Украина»	12.3	12.9	-0.6
Народный блок Владимира Литвина	3.1	3.3	-0.2
Коммунистическая партия Украины (П.Симоненко)	1.8	5.1	-3.3*
Социалистическая партия Украины (Мороз)	1.8	3.9	-2.1
Гражданская партия «ПОРА» (В.Каськів)	0.9	1.2	-0.3
Социал-демократическая партия Украины (В.Медведчук)	0.4	0.9	-0.5
Украинский народный блок Костенко-Плюща	0	0.6	-0.6
Трудно сказать	35.2	24.7	10.5
Итого	100	100	0

*) значимо на уровне 5%

Пример.

Рассмотрим результаты опроса, проведенного Киевским международным институтом социологии в декабре 2005 года⁸ (таблица1)

Мы видим, что в ХХХХ году политические ориентации владельцев кошек и владельцев собак отличались, в частности владельцы кошек более охотно были готовы голосовать за партию регионов и за коммунистов, чем владельцы собак. Неужели владение кошками повышало уровень симпатии к партии Януковича и к коммунистам? И с чем это может быть связано? Не можем же мы предположить, что кошки агитировали своих хозяев?

Конечно же нет, дело тут в другом. В городах легче держать кошек, чем собак, кошек не нужно выгуливать и они обычно меньше, чем собаки и лучше переносят ограниченные пространства городских квартир (в селе собак держат на улице и нет таких сложностей, как в городе). Поэтому среди владельцев кошек больше городских жителей, чем среди владельцев собак (см.таблицу2.).

Таким образом, реально на различия в ориентациях (признак А) в данном случае влияет не владение тем или иным животным (признак В), а третий признак (С) – место жительства (город или село). Если анализировать ситуацию более глубоко, то обнаружится, что на Востоке страны, где партия

регионов и коммунисты были наиболее популярны, проживает намного больше городского населения, чем в других регионах, так что скорее всего на политические ориентации влияет даже не тип проживания (город или село), а регион проживания респондентов.

Планирование эксперимента

Как уже указывалось, для определения того, влияет ли переменная А на переменную В стратегия исследования предполагает проведение эксперимента или квази-эксперимента.

Планирование эксперимента предполагает последовательность действий:

1) определение объектов, которые будут использованы как экспериментальная и контрольная группа.

Пусть, например, мы хотим проверить гипотезу о том, что просвещение улучшает нравы или, говоря современным языком и сужая задачу, что повышение уровня образования (переменная В) приводит к повышению уровня доброты (переменная А). Одним из способов проверки этой гипотезы может быть проведение эксперимента, в котором экспериментальной группой будет группа лиц с высоким уровнем образования, а контрольной группой – с низким уровнем. Если в экспериментальной группе

Таблица 2**Процент жителей города и села среди владельцев собак и кошек**

Держите ли Вы дома	Проживают в городе	... в селе	
Собак	43.4	56.6	100
Кошек	51.4	48.6	100

⁸ Данные репрезентативны для населения Украины в возрасте 18 лет и старше, проживающих во всех областях Украины и в Крыму, методом личного интервью опрошено 2028 респондентов, ошибка выборки с учетом дизайн-эффекта составляет 3.3%

средний уровень доброты будет выше, чем в контрольной группе – гипотеза подтвердилась.

2) выделения переменных, которые могут повлиять на эксперимент

Если использовать наш пример, то кроме образования на формирование доброты могут иметь влияние, например, пол, возраст, место проживания (город или село) и происхождение (тип семьи)

3) выбор методов уравнивания неэкспериментальных переменных.

Основная и контрольная группа должны быть одинаковыми по распределению всех признаков, которые могут повлиять на экспериментальную переменную А (в нашем примере – это доброта). До 1900-1910 гг главным методом выравнивания экспериментальной и контрольной группы был точечный подбор⁹. Например, в экспериментальную группу включены сто человек с высшим образованием. Тогда в контрольную группу подбирают людей, имеющих более низкое образование (образование – это экспериментальный фактор, по которому должны быть различия) и похожих по другим признакам. Скажем, первый человек в экспериментальной группе это женщина с высшим образованием 30 лет, проживающая в городе, из семьи интеллигенции. Тогда в контрольную группу тоже включают городскую женщину 30 лет из семьи интеллигенции, но эта женщина без высшего образования. И так к каждому члену экспериментальной группы подбирают «двойника» из контрольной группы. Однако, с начала 20 века доминирует подход «рандомизации» экспериментальной и контрольной групп, обе группы представляют собой стохастическую выборку из целевой генсовкупности с равной вероятностью попадания каждого элемента в экспериментальную группу (то

же самое касается и контрольной группы). В данном примере экспериментальная группа – это стохастическая выборка 100 респондентов из всех лиц с высшим образованием, а контрольная – такая же выборка из генсовкупности тех, кто не имеет высшего образования.

Наконец, еще два очевидных пункта плана действий:

3) выбор логики организации и условий проведения эксперимента и

4) выбор индикаторов и средств контроля результатов экспериментов, т.е. концептуализация и операционализация основных понятий и выбор критериев оценки значимости различий результатов экспериментальной и контрольной групп.

Валидность эксперимента: внутренняя и внешняя

Основной вопрос – действительно ли наша экспериментальная стратегия исследования показала, что переменная В является причиной переменной А и нет альтернативных переменных, которые могут повлиять на А. Это свойство называют внутренней валидностью эксперимента. Эксперимент обладает внутренней валидностью, если он показывает, что экспериментальное действие (экспериментальная переменная) действительно повлияла на те изменения, которые имели место в эксперименте.

Мы не сомневаемся, что проницательный читатель уже догадался, что раз есть внутренняя валидность, то есть и внешняя. Внешняя валидность эксперимента показывает на какие популяции, ситуации, другие переменные могут быть распространены результаты эксперимента. Если вспомнить пример экспериментов Миллера и Бугельского, которые вывели подростков на

⁹ Налимов В.В. Теория эксперимента. - М.: Наука, 1971. – 208 с

остров и показали, что создание трудностей повышало уровень ксенофобии, то возникают вопросы - а будет ли тот же результат, если это будут не подростки, а взрослые; а проявятся ли те же закономерности не на острове, а в обычной жизни; а распространятся ли результаты эксперимента на другие проявления враждебности. Первый из этих вопросов – на какие группы, на какую генсовокупность можно распространить результаты – это вопрос репрезентативности, так что внешняя валидность – это что-то вроде расширенной репрезентативности (возможности распространить результаты не только на другие группы людей, но и на другие ситуации и переменные).

Чтобы облегчить поиск и проверку возможных альтернативных признаков, которые могут повлиять на переменную А, кроме В, ниже приведены некоторые типичные ситуации таких признаков – факторов, угрожающих внутренней валидности экспериментального плана (здесь и далее мы следуем работам Дональда Кэмпбелла¹⁰)

Факторы, угрожающие внутренней валидности

1. Фон (history) - события, которые повлияли на изменения между первым и вторым измерением кроме экспериментального влияния. Имеется в виду эксперимент, в котором была зафиксирована начальная ситуация, потом произошло экспериментальное воздействие и проведено второе измерение для оценки влияния этого воздействия (как в приведенном выше примере – измерили уровень ксенофобии учащихся, вывезенных на остров, затем создали им трудности, за-

тем опять измерили уровень ксенофобии).

Пример 1.

В 1985 году отдел методологии и методов социологии Института философии Академии наук УССР провел опрос населения г. Киева, в ходе которого среди прочих задач проводилось также измерение удовлетворенности жизнью и ее отдельными аспектами. Кроме того, определялись жизненные ценности населения (использовалась методика В.Хмелько¹¹, жизненные ценности измеряются как коэффициент парной регрессии, показывающий, какой вклад удовлетворенность той или иной стороной жизни – семьей, работой, уровнем обеспеченности и т.п. – вносит в удовлетворенность жизнью в целом; чем выше вклад в удовлетворенность жизнью – тем выше ценность этого аспекта жизни).

В 1989 году был проведен второй опрос для измерения того, как повлияла на исследуемые показатели политика гласности и перестройки, начатая в 1985 г. руководством Советского Союза. Результаты показали, что индекс удовлетворенностью жизнью, который измерялся по шкале от -100 до +100, упал за 4 года с +36 до +1, а ценность семьи снизилась. Исследователи интерпретировали эти изменения как следствия гласности и перестройки. Действительно, за эти годы не произошло таких резких изменений в экономике страны, как после коллапса Советского Союза в 1991 и разрыва экономических связей. Но удовлетворенность жизнью зависит от отношения обеспеченности основных потребностей к уровню притязаний. Уровень обеспеченности за эти годы существенно не изменился, а вот уровень притязаний существенно возрос. В силу по-

¹⁰ Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. - М.: Прогресс, 1989. – 392с.

¹¹ Хмелько В.Е. Социальная направленность личности: некоторые вопросы теории и методики социологических исследований. — К.: Политиздат Украины, 1988

литики гласности средства массовой информации показали истинное положение вещей в стране и за рубежом, открылись границы, участились зарубежные поездки, население осознало, в какой глубокой яме (если не высказываться более резко) находится страна по сравнению с основными капиталистическими странами. И уровень притязаний существенно возрос, следовательно отношение уровня обеспеченности к уровню притязаний (удовлетворенность жизнью) упало. И хотя эта логика выглядит убедительно, но полной уверенности в правильности выводов нет. Дело в том, что в 1986г. произошла Чернобыльская авария, которая оказалась шоком для населения Киева. Возможно, что на падение удовлетворенности жизнью оказала влияние Чернобыльская авария, а не гласность и перестройка. Кроме того, исследования катастроф показывают, что часто в результате падает ценность семьи. И это тоже несколько подтачивает уверенность в валидности выводов, сделанных исследователями. Вот если бы в выборку входили еще какие-либо территории, которые были бы намного дальше от Чернобыльской зоны, например, Харьков, тогда возможно было отделить влияние гласности от влияния аварии (гласность влияла на всю территорию страны, а авария на зоны риска, различие характеристик населения Харькова и населения Киева показало бы влияние аварии).

Пример 2.

20 февраля 2012 года компания «Социальный мониторинг» (СМ) закончила опрос репрезентативный для избирателей Киева, а через неделю, 27 февраля аналогичный опрос закончил Киевский международный институт социологии (КМИС), результаты приведены в таблице 3

В обоих опросах Виталий Кличко занял первое место, Александр Попов – второе, остальные кандидаты (А. Гриценко, Н. Катеринчук и другие) получили меньше 5%, мы не приводили их рейтинге в таблице. Все кандидаты вели агитационные кампании, поэтому в командах кандидатов возник вопрос – что же такого сделала команда Кличко, что за неделю рейтинг Виталия Кличка вырос почти на 13%. Такой же вопрос возник и в команде Кличко, так как за эту неделю они не предпринимали каких-либо экстраординарных усилий в своей рекламной компании в Киеве. Итак, мы имеем контрольную группу – выборку, опрошенную «Социальным мониторингом», и характеризующую избирателей Киева, имеем экспериментальное воздействие – агитационную кампанию и имеем экспериментальную группу – выборку, опрошенную КМИСом и тоже характеризующую избирателей Киева. Валиден ли этот квази-эксперимент, можно ли сказать, что рейтинг В. Кличко вырос в результате агитационной кампании?

Таблица 3

Ответы на вопрос «Если бы в ближайшее воскресенье проходили выборы мэра Киева, приняли бы Вы участие в голосовании и если да, то каким был бы Ваш выбор?»

Варианты ответов	СМ	КМИС
Кличко Виталий	26.10%	39.00%
Попов Александр	23.60%	23.60%
Другие ответы (другие кандидаты, трудно сказать)	50.30%	37.40%
Итого	100.00%	100.00%

Анализ показал следующее. Социологические исследования не проводятся мгновенно, опрос СМ проходил с 10 по 20 февраля, а опрос КМИСа с 15 по 27 февраля. А 18 февраля поздно вечером в Мюнхене проходил бой Виталия Кличко с Дерекком Чисорой, который В. Кличко очень уверенно выиграл. В последующие дни этот бой активно обсуждали, демонстрировали фрагменты и о победе В. Кличко сообщали во всех новостях. Из 11 дней опроса СМ 9 дней респондентов опрашивали ДО боя, т.е. не менее 80% опрошенных отвечали, не зная о победе В.Кличко. Что касается нашего опроса (опроса КМИСа), то тут мы можем сказать точно, до 19 февраля было опрошено лишь 18% респондентов, т.е. 82% опрошены ПОСЛЕ боя В. Кличко. Поэтому валидность предположения о влиянии агитации является весьма сомнительной, мы полагаем, что главной причиной повышения рейтинга В. Кличко является медиа-эффект повышенного позитивного внимания к В. Кличко после боя (тем более, что через некоторое время рейтинг В. Кличко снизился).

2. Естественное развитие, взросление (maturation) - изменения со временем которые не связаны с конкретными событиями (взросления, рост жажды, голода, усталости и т.д.) В одном из методов определения эффективности рекламных роликов группе лиц, отправившихся в прогулочный круиз на катере, показывают рекламу напитков и фиксируют сколько напитков было куплено и выпито до, а сколько после показа рекламы. При этом возникает вопрос – действительно ли рост потребления напитков связан с рекламой или же просто через некоторое время у людей возникает жажда в силу естественных процессов.

Еще один пример – произошла реформа школьного образования и проводится

влияние новой системы на подростков – тестируется группа подростков 10-14 лет, и через два года тестирование повторяется. Можно ли сказать, что произошедшие изменения связаны с новой системой образования? Нет, такой уверенности нет, группа школьников повзрослела, теперь им 12-16 лет, в этом возрасте за два года происходят существенные изменения (например, растет IQ, этот рост прекращается только после 21 года).

3. Эффект тестирования, замера (testing)

Наиболее известные эксперименты, столкнувшиеся с эффектом тестирования – знаменитые Хотторнские эксперименты¹². Беспрецедентные по сложности и длительности эксперименты длились с 1924 по 1932г., исследователи по заказу компании Вестерн Электрик на Хотторнском заводе, расположенном в небольшом городе Хотторн (Hawthorne), пытались найти факторы, повышающие производительность труда сборщиц реле. К удивлению исследователей производительность труда росла и в основной группе, где проводили всякие улучшения условий труда и в контрольной группе, где не было никаких изменений. В итоге было обнаружено явление, названное Хотторнским эффектом – сам факт проведения исследования, внимание компании к работницам оказал большее влияние на результаты, чем экспериментальные факторы

4. Нестабильность, низкая надежность инструмента измерения результатов эксперимента.

Это наиболее очевидный фактор, который может поставить под угрозу валидность эксперимента. Если инструмент ненадежен, и мы при его использовании получаем каждый раз разные результаты даже когда ничего не происходит, то откуда мы узнаем, что на этот раз изменения связаны с действием

¹² Е.Гіденс. Соціологія. – Київ: Основи, 1999

экспериментальных факторов, а не случайных отклонений из-за ненадежного инструмента исследования.

5. Статистическая регрессия – изменения, которые происходят, когда группы отбираются на базе крайних оценок.

Этот фактор является самым неочевидным и, нужно сказать, что когда мы сами в свое время впервые прочли об этом факторе, то были несколько удивлены. Проще всего объяснить это на следующем примере. Пусть тестируется некоторая методика обучения, мы проведем начальный тест и отберем студентов, которые получили минимальные оценки. Например, тест имеет значения от 1 до 100 и мы отобрали студентов, имеющих значение от 1 до 10 (среднее, скажем, 3.8). Далее мы хотим провести обучение и посмотреть, увеличатся ли оценки студентов, вырастет ли среднее значение оценок. Результаты теста подвержены некоторым случайным колебаниям, если бы студенты заполняли несколько тестов подряд, то один раз их результаты были бы чуть лучше, другой раз чуть хуже при том же уровне знаний (усталось, голод, с кем-то поругались, неприятный звонок, что-то потеряли или наоборот, произошло что-то вдохновляющее – разные факторы могут повлиять на выполнение теста). Коэффициент корреляции между двумя тестами практически никогда не бывает равным единице. Из этого факта следует статистический вывод – студенты, имеющие наиболее низкие показатели в следующий раз в среднем покажут результат (т.е. выше, чем 3.8) лучше даже без всякого обучения, часть из них случайно покажет чуть лучше результаты, а другие не смогут показать чуть хуже, так как их оценки уже минимальны. С другой стороны, если мы отберем группу студентов с наилучшими оценками, а потом повторим тест, то эта группа в следующий раз по тем же причинам покажет результат, который хуже, чем в первый раз.

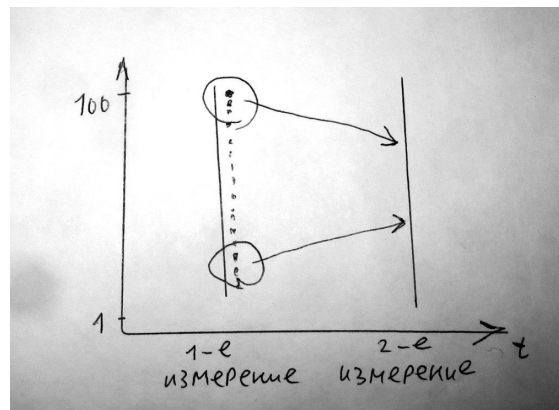


Рис. 1. Статистическая регрессия к среднему при отборе крайних групп

6. Отбор (selection) - неэквивалентность отобранных групп по составу

Если мы сравниваем экспериментальную и контрольную группу, то они должны быть одинаковы по всем характеристикам, кроме экспериментальной переменной. Однако иногда это сложно обеспечить. В приведенном нами ранее примере сравнения политических ориентаций владельцев кошек и собак, эти две группы были неэквивалентными по доле лиц, живущих в селе и в городе.

Когда мы выровняли эти группы по этому признаку, сравнили владельцев кошек и собак, живущих в крупном городе, разница в политических ориентациях этих двух групп исчезла.

7. Отсев (attrition, experimental mortality) - неравномерность выбывание респондентов из групп, которые сравниваются

Речь идет о ситуациях, когда сначала контрольная и экспериментальные группы были одинаковыми, но потом из-за выбывания членов группы их состав стал отличаться. Или мы провели измерения в какой-то группе, потом произошло не только экспериментальное воздействие, но и изменение состава группы, поэтому после второго измерения возникает проблема отделения эф-

факта экспериментального воздействия от эффекта отличающихся структур групп.

Типичным примером являются так называемый лонгитюдные исследования, т.е. исследования одних и тех же респондентов в течение нескольких лет – возраст респондентов увеличивается и мы не знаем, какие изменения связаны с возрастом, а какие – с социальными изменениями.

Еще один пример – исследование доходов и расходов населения Украины. В советской Украине и в первые годы независимости Украины правительство опиралось на данные панели дневников расходов и доходов, которые вела репрезентативная выборка респондентов в течение года, получая за это небольшую оплату. При этом был большой отсев и состав группы менялся – менее обеспеченные и более организованные респонденты дольше оставались в панели и продолжали работать, но полученные таким образом изменения в бюджетах не отражали реальной ситуации в стране. Поэтому Госкомстат в 1999 перешел на другую методологию.

Еще одна проблема, с которой пришлось столкнуться социологам и не только социологам – аннексия Крыма Россией в 2014 году и сложности проведения опросов в Крыму после этого. И если даже мы считаем, что Крым – часть Украины, то результаты по Крыму не интересуют многих наших заказчиков, так как жители Крыма практически не участвуют в выборах в Украине. Если мы изучаем динамику процессов в Украине, то сравнение результатов по Украине в целом с предыдущими годами сталкивается с тем, что часть изменений связана с отсутствием Крыма в нынешних опросах. Например, задавая вопрос «Как бы Вы проголосовали, если бы проводился референдум о присоединении к ЕС?» мы получили намного больший процент проголосовавших бы за ЕС, чем в 2013 году, но частично из-за того, что Крым не опрашивается (и какой смысл проводить опрос в Крыму в этом

случае, если при проведении референдума в Украине жители Крыма в нем не участвовали бы?). Де факто мы имеем несовпадение состава сравниваемых генеральных совокупностей до и после 2014 года.

Факторы, угрожающие внешней валидности

Факторы, угрожающие внешней валидности, для социологов не очень опасны. Если психологи часто исследуют лишь несколько небольших групп людей, полагая, что психологические характеристики не зависят от социального статуса или региона, то социологи всегда озабочены проблемами репрезентативности и проводят массовые опросы, используя не реальные эксперименты, а квази-экспериментальные планы. Проблема обобщения результатов исследования и распространения выводов на генеральную совокупность для нас является стандартной. Поэтому рассмотрим лишь две таких угрозы.

1. Реактивный эффект или эффект взаимодействия тестирования и экспериментального эффекта – повышение чувствительности к экспериментальному воздействию за счет предварительного тестирования

Пример. Тест Айзенка на коэффициент интеллекта (IQ) предназначен для измерения скорости протекания умственных процессов, и по замыслу Айзенка измеряет врожденные характеристики человека. Оказалось, однако, что существует определенный элемент обучения: люди, которые заполняют тест второй и третий раз делают это лучше, они совершенствуют стратегию заполнения теста, не застревают на сложных для них задачах, а сначала делают то, что им проще и т.п. Поэтому считается, что первые несколько измерений являются подготовкой и истинное измерение коэффициента интеллекта начинается с 3-го или 4-го измерения.

2. Взаимодействие отбора и экспериментального воздействия, когда именно сам характер экспериментального воздействия приводит к тому, что состав экспериментальной группы становится неэквивалентным составу контрольной группы. Чаще всего это происходит за счет самоотборов.

Например, в опросах некоторых телеканалов и радиостанций задается вопрос с двумя альтернативами, скажем, «да» и «нет» и указываются два телефона – звонок по одному из них означает «да», звонок по другому – «нет». Были ли Вы в филармонии в этом году? Умеете ли Вы играть на скрипке? Знаете ли Вы испанский язык? Во всех этих вопросах у людей, которые были, умеют, знают выше заинтересованность в теме и больше стимулов позвонить, чем у тех, кто не был, не умеет, не знает. В итоге процент посещающих филармонию или играющих на скрипке среди зрителей этого телеканала оказывается намного выше, чем в действительности. А так как неизвестно, насколько выше – на 10%, вдвое выше, в десять раз выше – то результаты таких опросов совершенно бесполезны.

Еще один типичный пример – программы технической помощи различным организациям, предусматривающие обучение персонала или реформы в этой организации или тренинги для тех или иных специалистов. Когда потом сравнивают результат тех, кому оказана помощь или кто прошел тренинг, с теми, кто не прошел – результаты впечатляют. Однако, во многих случаях в таких программах происходит самоотбор, на участие в программах соглашаются организации, руководство которых более прогрессивно и больше заинтересовано в результате, поэтому в этой организации с самого начала могла быть лучше ситуация, чем у тех, кто отказался. И то же самое с тренингами, на которые самоотбираются наиболее активные и способные люди.

Примеры типичных планов экспериментов

Мы рассмотрим пять планов, но первые три плана, хотя и часто используются, имеют очень много недостатков (в одной из работ Д. Кембелл назвал их доэкспериментальными), лишь 4-й и 5-й планы реально являются экспериментальными и рекомендуемыми для использования.

Но сначала введем некоторые обозначения.

X - экспериментальная переменная (действие, влияние), которая подлежит измерению.

O – измерение (от observation – наблюдение, которое иногда используется как синоним измерения)

R – означает, что группа набрана с помощью случайного репрезентативного отбора (от random - случайный)

1. Исследования одиночного случая

X O - группа подверглась экспериментальному воздействию X после чего произошло измерение параметров, интересующих исследователя.

Например, группу записавшихся на тренинг обучают чему-то, а потом оценивают результаты или дают лекарства группе больных и оценивают, насколько это им помогло (то есть сравнение осуществляется с предположением как было бы без влияния или с каким-либо идеалом). Этот план подвержен практически всем описанным выше влияниям, он не защищен от угроз внутренней и внешней валидности и практически бесполезен. Очень многие мошенничества или распространение бесполезных товаров основано на такой схеме. Например, людям предлагаются какие-то биодобавки и говорят, что их постоянное применение в течение 8-10 месяцев приводит к значительному улучшению здоровья. В доказательство приводят свидетельства тех людей, кто стал

энтузиастом их применения, так как его состояние улучшилось. Части людей, которые принимают эти добавки, стало действительно лучше, части — хуже и они перестают принимать их, а у части ничего не изменилось, но им кажется, что стало лучше. Первая и третья группа продолжают сами их по-

купать и рекомендовать другим¹³.

При использовании этой схемы эксперимента логика приписывания результата именно рассматриваемому экспериментальному воздействию точно такая же, в статье «Огурцы Вас погубят» из сборника «Физики продолжают шутить».

*Огурцы вас погубят! — Физики продолжают шутить.
— М.: Мир, 1968*

Каждый съеденный огурец приближает вас к смерти. Удивительно, как думающие люди до сих пор не распознали смертоносности этого растительного продукта и даже прибегают к его названию для сравнения в положительном смысле («как огурчик!»). И, несмотря ни на что, производство консервированных огурцов растет. С огурцами связаны все главные телесные недуги и все вообще людские несчастья.

Практически все люди, страдающие хроническими заболеваниями, ели огурцы. Эффект явно кумулятивен.

99.9% всех людей, умерших от рака, при жизни ели огурцы.

100% всех солдат ели огурцы.

99.7% всех лиц, ставших жертвами автомобильных и авиационных катастроф, употребляли огурцы в пищу в течение двух недель, предшествовавших фатальному несчастному случаю.

93.1% всех малолетних преступников происходят из семей, где огурцы потребляли постоянно. Есть данные и о том, что вредное действие огурцов сказывается очень долго: среди людей, родившихся в 1839 г. и питавшихся впоследствии огурцами, смертность равна 100%. Все лица рождения 1869-1879 гг. имеют дряблую морщинистую кожу, потеряли почти все зубы, практически ослепли (если болезни, вызванные потреблением огурцов, не свели их уже давно в могилу). Еще более убедителен результат, полученный известным коллективом ученых-медиков: морские свинки, которым принудительно скармливали по 20 фунтов огурцов в день в течение месяца, потеряли всякий аппетит!

Единственный способ избежать вредного действия огурцов — изменить диету. Ешьте, например, суп из болотных орхидей. От него, насколько нам известно, еще никто не умирал.

Напечатано в журнале «The Journal of Irreproducible Results».

¹³ Мы не против биодобавок, мы в этом не разбираемся, может быть, они необыкновенно полезны. Мы против того способа доказательства их полезности, которые используют некоторые распространители, и против их бездумного применения без попыток оценить их полезность именно для Вас

2. План с предыдущим и конечным тестирования в одной группе

O1 X O2 - отличается от первого плана начальным тестированием O1, затем, как и в первом плане осуществляется экспериментальной воздействию и затем опять измерение O2

Этот план несколько лучше, чем первый. Если бы, начиная прием биодобавок, покупатели зафиксировали бы свое состояние здоровья, сходили бы к независимым от продавцов врачам, сделали бы различные анализы, то связать состояние здоровья с приемом биодобавок было бы проще. Однако, этот план тоже подвержен действию почти всех угроз валидности, которые были рассмотрены ранее.

Во-первых, может повлиять фон. Д. Кемпбелл приводит в качестве примера исследование Колльера 1940 года, который исследовал влияние изучения материалов нацистской пропаганды на установки студентов по схеме O1 X O2. В период между первым и вторым замером пала Франция, что, скорее всего, оказало большее влияние на студентов, чем изучение предложенных Колльером материалов. Приведенный нами ранее пример двух опросов в Киеве тоже был проведен по этой схеме и тоже, кроме влияния экспериментальной переменной (гласность), было альтернативное событие – Чернобыльская авария. Во-вторых, влияет и естественное развитие, между первым и вторым измерением испытуемые в зависимости от времени между первым и вторым измерениями могут проголодаться, устать, заскучать, стать старше и т.п. Далее аналогично может влиять и начальное тестирование и нестабильность инструмента измерения и статистическая регрессия и отсев.

3. Группа, на которую влияют, сравнивается с той, на которую не влияют

Первая группа	X O1
Вторая группа	O2

Например, сравниваются те, кто смотрел определенную телевизионную передачу, скажем, рекламу шоколадных батончиков, с теми, кто не смотрел. X здесь – трансляция рекламы, а измерения O1 и O2 – социологический опрос, измеряющий желание респондентов купить этот батончик. Или сравниваются те, кто занимался на курсах скоростного чтения текстов, с теми, кто не занимался. Этот план уже валиден по отношению к влиянию фона, фон влияет и на первую и на вторую группу, поэтому разница измерений O1 и O2 не может быть обусловлена влиянием фона, фон тут не является альтернативным объяснением полученных результатов. Поскольку нет первого замера, то мы тут не говорим о естественном развитии и начальном тестировании.

Однако отбор и отсев – это те факторы, которые могут повлиять на валидность этого экспериментального плана. «В педагогических исследованиях с этой проблемой чаще всего сталкиваются в работах, цель которых – установить влияние обучения в колледже путем сравнения характеристик новичков, не подвергшихся влиянию X, и выпускников, подвергшихся такому влиянию. Если в подобных исследованиях устанавливается, что, к примеру, вновь поступившие девушки более красивы, чем выпускницы, то мы воздерживаемся от вывода, что наш суровый способ обучения пагубно влияет на красоту, а скорее осмеливаемся утверждать, что красивые девушки чаще выбывают из колледжа до завершения обучения в связи с замуже-

ством. ... (Конечно, если бы мы сравнивали тех же самых девушек, когда они поступали, и когда они окончили колледж, такая проблема не возникла бы, и мы имели бы дело с планом 2).¹⁴

Мы не уверены, что красивые девушки чаще бросают учебу и навряд ли им легче выйти замуж (это зависит скорее от характера, чем внешности, и от многих других характеристик, включая интеллект, чувство юмора, запах и т.п., к тому же, как показывают теории социального обмена, у красивых девушек выше уровень притязаний). Но этот пример показывает, что 3-й экспериментальный план не намного лучше второго.

4. План с предыдущим и конечным измерением и контрольной группой

R O1 X O2 - экспериментальная группа, набранная с помощью стохастического отбора отбор, начальное измерение O1, экспериментальное воздействие X, измерение O2

R O3 _ O4 - контрольная группа, набранная с помощью стохастического отбора отбор, начальное измерение O1, экспериментальное воздействие отсутствует, измерение O2

Этот план с двумя эквивалентными группами — основной (экспериментальной) и контрольной — является научно обоснованным и позволяет во многом контролировать перечисленные угрозы валидности. Наличие контрольной группы позволяет контролировать фон (так же, как в плане 3), естественное развитие и начальное тестирование одинаково влияет на обе группы и не может помешать вычленению влияния

X, эффект регрессии контролируется при сравнении средних для всего массива. Влияние нестабильности инструмента можно оценить, сравнивая значимость различий для двух случайных выборок. Влияние отбора и отсева устраняется стохастической эквивалентностью двух групп, но проблема взаимодействия отбора с X сохраняется¹⁵

Один из наиболее известных типов применения этого плана — использование в медицине в качестве контрольной группы так называемую группу плацебо. Плацебо (лат. *placere* — нравиться) означает препарат, у которого нет лечебных свойств («фальшивые лекарства»). Термин «эффект плацебо» введен в научный оборот американским врачом Генри Бичером в 1955 г., который выяснил, что примерно треть пациентов выздоравливают от таблеток-«пустышек», которые не содержат действующих веществ. Чтобы определить реальный эффект лекарства и отделить его от эффекта плацебо кроме экспериментальной группы, получающей лекарство, формируется эквивалентная группа, получающая плацебо.

Есть анекдот о плацебо, который показывает, что существуют и другие неконтролируемые факторы.

«Больница.

Пациент приходит к лечащему врачу:- Доктор, почему Вы изменили мне назначение, даете какие-то другие таблетки?

Доктор: Почему Вы так думаете? Эти таблетки чем-то отличаются?

Пациент: На вид они такие же. Но когда я бросал в унитаз те, что Вы мне раньше давали - они тонули, а эти плавают.»

¹⁴ Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. - М.: Прогресс, 1989, с.62.

¹⁵ Настоящие исследователи хорошо понимают важность контрольной группы. В качестве примера приведем анекдот. Жена статистика родила близнецов. Он был в восторге и тут же позвонил священнику договориться о крещении. «Поздравляю, - сказал священник, - принесите их в церковь в воскресенье, и мы будем крестить их». «Нет, - ответил статистик, - Крестить будем одного, второго оставим в качестве контрольной группы.» Кстати, «близнецовый метод» - сравнение близнецов, выросших вместе и выросших отдельно — является одним из важных способов отделения врожденных характеристик человека от формируемых путем социализации.

5. План Соломона для 4-х групп.

R O1 X O2 - первая экспериментальная группа с начальным и конечным измерением

R O3 _ O4 - первая контрольная группа с начальным и конечным измерением

R _ X O5 - вторая экспериментальная группа только с конечным измерением

R _ _ O6 - вторая контрольная группа только с конечным измерением

Этот план назван не в честь библейского Царя Соломона, а по имени разработавшего его американского ученого Р.Соломона.

Этот план является усовершенствованием плана 4 для 2-х групп и позволяет отделить экспериментальный эффект от эффекта первого измерения:

Среднее для столбцов - главный эффект X равен $(O2+O5)/2$

Среднее по срокам - эффект первого измерения равен $(O4+O6)/2$

Некоторые специальные типы исследований и исследовательских стратегий.

Некоторые типы исследований имеют специфический дизайн и требуют специальных исследовательских стратегий. Мы перечислим некоторые из них, но не будем рассматривать, так как это требует более глубокого рассмотрения и не может быть осуществлено в рамках этой статьи.

Обычные разовые одномоментные репрезентативные исследования называют кросс-секционными (cross-sectional survey) - это исследования, предполагающие сбор данных в один определенный момент вре-

мени. Это как бы одна засечка на шкале времени (cross-sectional – это поперечное сечение).

В отличие от этого лонгитюдные исследования (longitudinal survey) предполагают повторяющиеся время от времени исследования со сбором информации по одним и тем же переменным. Можно выделить три типа лонгитюдных исследований – трендовые, когортные и панельные¹⁶.

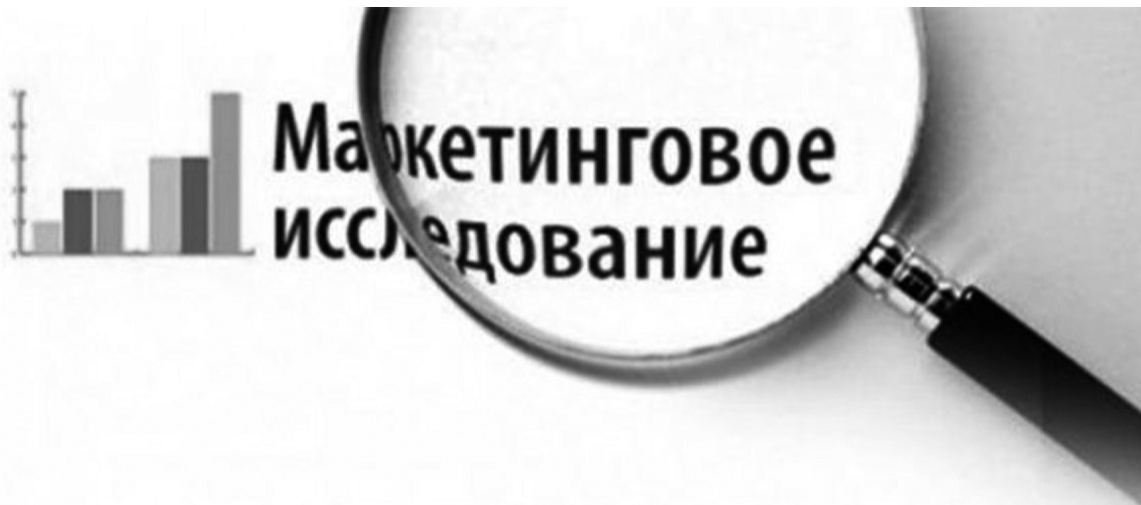
Трендовые или повторные исследования – это повторяющиеся исследования по репрезентативным для данной генсовокупности выборкам по той же системе показателей (признаков), просто серия кросс-секционных исследований¹⁷.

Когортные исследования - это когда мы с некоторым временным промежутком исследуем одну и ту же категорию людей. Чаще всего, это возрастные когорты. Или, скажем, выпускники школ данного года (и мы можем раз в пять лет проводить опросы и сравнивать выпускников школ 2005, 2010 и 2015 годов).

Панельные исследования – когда мы время от времени опрашиваем одних и тех же людей (панель). Отличие от когортных исследований состоит в том, что в когортных исследованиях мы изучаем одну и ту же категорию людей, а не одних и тех же людей. И в 2005 и в 2010 и в 2015 это опрос 17-18-летних школьников, мы получим три волны такого опроса и сможем сказать, как изменились выпускники школы за 10 лет, например, как изменилось восприятие престижа профессий, какие профессии стали более, а какие менее престижными. Если же мы в 2005 году

¹⁶ Терминология может несколько отличаться (иногда, например, лонгитюдные исследования считают синонимом панельных исследований), мы придерживаемся терминологии учебника Эрла Бабби, который уже больше двадцати лет используется как один из основных учебников на ф-те социологии Национального университета «Киево-Могилянская академия». Earl R. Babbie The Practice of Social Research. 8-th edition. – Wadsworth Publishing Company, 1998, pp.101-105

¹⁷ Как, например, национальные переписи или мониторинговые опросы КМИСа для оценки динамики уровня ксенофобии в Украине, уровня бедности, поддержки независимости Украины (см. <http://kiis.com.ua/?lang=ukr&cat=reports&id=529&page=5>).



начали панельный опрос выпускников, то мы должны найти тех же самых людей и в 2010 и в 2015 и опросить их, им будет уже не 18, а 23 года, а потом 28 лет. Мы не сможем сказать по этим данным, как меняется престиж профессий у выпускников школ и на какие специальности можно ожидать высокие конкурсы, по панельному опросу мы сможем сказать, как с возрастом изменяется восприятие мира тех, кто закончил школу в 2005 году, пригодились ли им школьные знания, что было наиболее и что наименее полезным для их дальнейшей жизни.

Из всех этих типов исследований панельные исследования являются наиболее сложными, возникает множество проблем¹⁸. Во-первых, поиск (tracing) респондентов, которые за 5-10-15 лет переезжают в другое место, во-вторых, отсев участников панели (panel attrition, panel mortality), часть из них не удается найти, другие отказываются от-

вечать, социологи надоедают им за эти годы (особенно, если опросы проводятся ежегодно или чаще), в-третьих, панель стареет и необходимо ее «ремонтировать». Все это требует специальной стратегии исследования и специального методического дизайна¹⁹.

Наконец, еще один тип исследований – это оценочные исследования (evaluation research, social assessment). Эти исследования представляют собой применение процедур социального исследования для оценки внедрения и полезности программ социального вмешательства (интервенции). Иными словами оценочные исследования это использование исследовательской методологии для оценки и улучшения планирования, мониторинга, эффективности, программ в области здравоохранения, образования, благосостояния и других программ обслуживания человека²⁰.

¹⁸ Methodology of Longitudinal Surveys./Ed.by Peter Lynn. UK: John Wiley & Sons, 2009/ - 396p. Джесси Рассел «Лонгитюдное исследование». – М.: Издательство «VSD», 2013.

¹⁹ Не можем отказать себе в удовольствии привести определение панели Е. Головахи, использующего двусмысленность этого термина. « ПАНЕЛЬ – группа респондентов, периодически удовлетворяющая исследовательские потребности социологов». Е. Головаха. Психология и другие смешные науки. Альтернативный словарь. – М.: Смысл, 2012. – 142 с.

²⁰ Оценка программ: методология и практика./ Под ред. А.И. Кузьмина, Е.О. Салливан, Н.А. Кошелевой. – М.: Издательство «Престо-РК», 2009. – 396с. А здесь (<http://www.cehd.umn.edu/OLPD/MESI/resources/default.html>) Вы найдете многочисленные онлайн источники информации об этом типе исследования