

ВАЛИДНОСТЬ - 1. Показатель качества метода, его способность давать результаты, адекватно отражающие изучаемое явление, т.е. именно те результаты, для получения которых он предназначен (В. метода). 2. Мера соответствия теории эмпирическим данным, возможность делать разумно точные предсказания на основании теории (В. теории). 3. Мера соответствия результатов изучаемой реальности, точнее, представлениям о реальности (В. результатов). В данной статье В. обсуждается в первом смысле. В позитивистской науке различают В. измерения, внутреннюю и внешнюю В. экспериментальных процедур и В. статистического вывода.

В. измерительных процедур

Под измерением понимают процесс связывания теоретического понятия с одной или несколькими латентными переменными, а этих последних - с наблюдаемыми переменными. В классической теории результат измерения включает два не коррелирующих компонента: истинный и ошибочный.

В отношении ошибок измерения делаются предположения, при выполнении которых поведение ошибок становится известным. В. измерительной процедуры - мера ее соответствия измеряемому конструкту. Чаще других оценивают критериальную (часто называемую эмпирической), конструктивную, конвергентную и дискриминантную В. В качестве вспомогательных методов установления В. используют реферирование литературы по вопросу, экспертные оценки.

Критериальная В. измерительной процедуры оценивается по тесноте статистической связи (корреляции) между измеренными результатами и внешним критерием. Так, критериальная В. шкалы лидерства означает, что полученный показатель коррелирует с некоторым независимо полученным показателем, например, социометрическими данными. Указание на критериальную В. предполагает не только сообщение о коэффициенте корреляции между показателями теста и внешним показателем, но также о всех обстоятельствах исследования: как и в какой ситуации был измерен критерий, на какой выборке проводилось исследование и др. Критерии могут быть текущими (измеряемыми одновременно с применением валидизируемой процедуры) и прогностическими. Прогностическая В. высоко желательна для практически ориентированных методов измерения. Недостатками подхода является трудности с подбором критерия и риск измерения невалидного критерия, что существенно снижает полученный коэффициент корреляции.

Как оцениваемое поведение, так и критерий могут меняться во времени, а выборки могут быть смещенными. Если бы исследователю удалось найти совершенный критерий, создание измерительной процедуры стало бы избыточным. В этой связи понятен

парадокс, сформулированный Дж.Келли: В. метода - это его способность предоставлять и без того известную информацию. Зависимость критериальной В. от свойств критерия, невозможность нахождения удовлетворительного объективного показателя во многих областях знания существенно снижают ценность критериальной В. как показателя качества измерительной процедуры.

Конструктная В. устанавливается по статистической связи между показателями данной процедуры и других методик, измеряющих родственный конструкт. Гипотезы о статистической связи формулируются до их проверки, на основании содержательной теории.

Конструктная валидизация - длительный процесс, и ни одна эмпирическая корреляция не может гарантировать достоверности измерения. Теоретически постулируется связь между конструктами, оценивается корреляция между индикаторами этих конструктов, на основании полученных данных пересматриваются теоретически ожидаемые связи между конструктами или между конструктами и индикаторами. Подбирается новый конструкт (или новый индикатор, или новая теория о связи между конструктами), и процесс повторяется снова.

Оба конструкта могут оказаться неродственными. Так бывает, когда измерительные методы, выступающие под одним и тем же названием, строятся на разных теоретических основаниях.

Другие проблемы связаны с ситуациями, когда измеренный показатель, привлекаемый для валидизации процедуры, измеряется недостаточно валидно; один и тот же индикатор одновременно измеряет разные конструкты; ошибки измерения коррелируют между собой. Если исследователь уверен/а в том, что структурная схема корректна (отсутствуют связи между индикаторами и посторонними конструктами, а корреляция между ошибками измерения равна нулю), он/а может найти корреляцию между конструктами, скорректированную на ненадежность индикаторов.

Для оценки конструктной В. измерительной процедуры часто используют факторный анализ индикаторов. Под этим именем понимается широкий набор методов снижения размерности данных, когда из множества измеренных переменных извлекается небольшое количество латентных переменных (факторов). Количество и теоретическая интерпретируемость факторов трактуется как мера В. измерительного метода, а факторные нагрузки измеренных переменных - как мера В. индикаторов. Факторная нагрузка - это стандартизованная мера структурной связи между общим фактором (латентной переменной) и индикатором. Ее часто определяют как корреляцию между фактором и переменной. Широко известные методы факторного анализа (например,

анализ главных осей) является эксплораторной процедурой, т.е. не позволяет проверять статистических гипотез, а результат анализа существенно определяется техническими решениями, который принимает исследователь.

В этом смысле возможности метода как средства валидизации ограничены. Чтобы ослабить произвол при проведении факторного анализа, используют кросс-валидизацию: выборка случайно расщепляется пополам, факторы извлекаются на одной половине выборки, а оправданность и устойчивость факторного решения проверяется на другой половине.

Понятия конвергентной и дискриминантной В. введены Д. Кембеллом и Д. Фиске как два взаимосвязанных показателя достоверности метода. Конвергентная В. требует "сходимости" (высокой корреляции) родственных конструкторов, дискриминантная В. - отсутствия корреляций между измерениями неродственных конструкторов. Более строго эта концепция реализована в подходе, известном под именем "Много черт - много методов", или МЧММ (MTMM - many traits, many methods).

Суть его в том, что несколько свойств измеряются несколькими методами. Между показателями одного и того же свойства, измеренного с помощью разных методов, должны быть высокие корреляции, между показателями разных свойств, измеренными с помощью одного метода, - низкие. Третьим условием является превышение первого набора коэффициентов корреляции над вторым.

Есть и более строгие статистические методы анализа корреляционной матрицы МЧММ. В последние годы приобрели популярность структурные уравнения, в частности, конфирматорный факторный анализ, который в соответствии с содержательной моделью раскладывает ковариации измерений на компоненты, связанные с влиянием методов, свойств, и ошибки. Полученные структурные коэффициенты интерпретируются как показатели В.

В. экспериментального метода

Круг понятий, связанных с оценкой В. эксперимента (и результатов эксперимента), был разработан методологами-постпозитивистами в 1960-х (Д. Кэмпбелл, Т. Кук, Дж. Стэнли и др.).

В эпистемологической традиции, восходящей к Дж.С.Миллю и связывающей причинность с манипуляцией, истинный эксперимент со случайным распределением

испытуемых по условиям рассматривается как единственное средство убедительной проверки каузальных гипотез.

С этой точки зрения причинные утверждения в неэкспериментальных науках (социологии) нелегитимны. Предшествование во времени предполагаемой причины (независимой переменной) следствию (зависимой переменной) как одно из условий каузальной связи в эксперименте обеспечивается манипуляцией и измерением ее последствий. Сложнее выполнить другое требование к каузальному выводу - отсутствие правдоподобных альтернативных объяснений. Под внутренней В. понимается уверенность в том, что именно независимая переменная X служит основной причиной систематической изменчивости значений зависимой переменной Y, т.е. отсутствуют другие переменные, опосредующие связь между X и Y. Мы не в состоянии гарантировать В. эксперимента, но с учетом особенностей используемой исследовательской процедуры можем идентифицировать и устранить основные причины невалидности.

Кэмпбелл и Стэнли предложили классификацию причин снижения внутренней В. История (фон) - это события, которые произошли между X и Y и могли повлиять на их ковариацию. Естественные изменения - перемены во внутреннем состоянии испытуемых, не связанные с влиянием X, такие как усталость, взросление, научение. Эффект инструмента связан с нежелательными изменениями измерительных процедур (технические поломки, усталость наблюдателей). Эффект статистической регрессии (соскальзывание к среднему) наблюдается тогда, когда экспериментальные и/или контрольные группы отбирались на основании крайних значений релевантных показателей; по причинам статистического свойства к моменту второго измерения максимальные значения показателей понизятся, а минимальные повысятся.

Отбор групп как причина невалидности связан с неслучайным (неэквивалентным) отбором испытуемых или других единиц исследования. Отсев ("экспериментальная смертность") означает неравномерное и неслучайное выбывание участников исследования. Взаимодействие отбора с другими факторами описывает ситуацию, когда неслучайно отобранные испытуемые имеют разную личную историю, склонность к усталости, отсев и др. Внутренняя В. оценивается качественно по степени отклонения исследовательской процедуры от некоторого идеального случая. Это касается также неэкспериментальных планов исследования, в которых к приведенному перечню причин снижения валидности добавляются новые угрозы.

Внешняя В. характеризует степень обобщаемости полученных результатов на иные генеральные совокупности и контексты.

Успешное воспроизведение исследования служит одним из важных условий оценки достоверности теорий и помогает уточнить диапазон пригодности последних. Одной из стратегий достижения внешней В. также является идентификация и устранение причин, ответственных за случайное получение результата. Это, в частности, реактивный эффект (влияние процедуры начального измерения на поведение испытуемых) и взаимная интерференция экспериментальных воздействий. Другие стратегии основаны на использовании плана рандомизированного эксперимента с единственным измерением после исследования (что снимает реактивный эффект); снижение риска влияния экспериментатора (напр., минимизация контактов экспериментатора с испытуемым, увеличение количества исследователей или двойной слепой метод, когда исследователь до окончания эксперимента не знает, кто из испытуемых каким воздействиям подвергается); повышение экологической В. (реалистичности) исследования. Важную методологическую роль в обеспечении В. научного исследования играет воспроизведение с расширением, когда каждая последующая репликация эксперимента предусматривает небольшие процедурные модификации.

В. статистического вывода

Под В. статистического вывода понимается выполнение комплекса условий, уменьшающих вероятность неверного статистического решения о нулевой гипотезе (о равенстве параметра некоторому значению или соответствии статистической модели полученным данным) или величине эффекта. Основным методом принятия статистических решений в последние десятилетия стала проверка статистической значимости, т.е. о равенстве параметра некоторому значению. Для этого используются критерии z , t , F , χ^2 и др. Эмпирически полученное значение критерия сравнивается с критическим, и на основании этого сравнения нуль-гипотеза отвергается или не отвергается. Т.о., решение носит бинарный характер (да-нет). В этом процессе можно совершить одну из трех ошибок: неверно отвергнуть истинную нуль-гипотезу (ошибка первого рода, α), неверно принять истинную альтернативную гипотезу (ошибка второго рода, β) и неверно сформулировать статистические гипотезы, т.е. неправильно перевести исследовательский вопрос на язык статистики.

Угрозы В. статистического вывода многочисленны и имеют разную природу. Использование невалидных индикаторов (см. раздел данной статьи

Валидность измерительных процедур) не дает возможности уверенно интерпретировать полученный результат. Использование ненадежно измеренных переменных приводит к существенному занижению полученных статистик.

Большой класс угроз В. статистического вывода связан с неверным формулированием модели.

В частности, более известные линейные модели выбираются в тех случаях, когда лежащие в их основе предположения нарушаются: связь между переменными носит нелинейный характер, присутствуют экстремальные значения, дисперсии зависимой переменной для разных уровней независимой переменной неравны, переменные измерены более грубо, чем того требует модель, отсутствующие значения переменных распределены неслучайным образом. В стандартных пакетах статистического анализа имеются методы диагностики нарушений предположений модели.

Более сложные концептуальные ошибки совершаются в связи с неверным определением переменных как независимых или коварирующих, неправильным выбором моделей для зависимых (повторных) или независимых измерений, фиксированных или случайных эффектов. При сравнении нескольких средних или оценке значимости нескольких корреляций исследователи не всегда учитывают инфляцию ошибки первого рода: так, для j переменных мы получаем $k = j(j - 1)/2$ корреляций, и вероятность случайно получить хотя бы один значимый коэффициент корреляции, а общ., равна $1 - (1 - \alpha)^k$. Использование слишком большого числа переменных при небольшом числе наблюдений (респондентов) создает "сверхпригодность" модели, когда модель идеально описывает выборочные данные, но неадекватна для описания генеральной совокупности.

Многие простые ошибки статистического вывода обусловлены причудливым и механическим объединением двух разных подходов к статистическому решению - Фишера и Неймана-Пирсона.

Последний подход подчеркивает важность мощности критерия для оценки значимости. Так, результат может оказаться статистически незначимым по причинам, связанным с недостаточной мощностью метода: слабость экспериментального воздействия, маленькая или гетерогенная выборка, ненадежное измерение переменных, чрезмерно малое значение ошибки первого рода. Статистическая значимость - это вероятность того, что полученный результат случаен на выборке данного объема, однако эта величина ничего не говорит о величине эффекта. Современные руководства рекомендуют сообщать в публикациях не только показатель значимости (p), но значения эффекта (типичные показатели - r Пирсона, t Стьюдента, d Коэна, g Хиджеса, η^2 для дисперсионного анализа и др.). Другими альтернативами бинарным решениям о статистической значимости являются доверительные интервалы, байесовы статистики и, в более широком смысле, воспроизведение с расширением и метаанализ.

В. в качественных исследованиях

В. в качественных исследованиях не может быть оценена с помощью описанных выше классических методов. Ряд авторов показывают примитивность имеющихся подходов к оценке В. и "критериологии" в целом (Т.Швандт, Дж

.Смит). В то же время отказ от идеи достоверности/аутентичности в отличие от дискредитированной объективности выглядит слишком радикальным даже для постмодернистских исследователей. Поэтому другие авторы (Э.Губа, И

.Линкольн, С.Квале, С.Мишлер) пытаются переформулировать традиционные критерии валидности, сделав их менее позитивистскими.

Обсуждаются новые процедуры обоснования достоверности: коммуникативная валидизация (участие испытуемых в обсуждении результатов; привлечение к обсуждению коллег), процедурная валидизация (точность, тщательность и полнота полевых заметок или получаемых данных; полное, сбалансированное и прозрачное описание результатов, открытое для иных интерпретаций; чувствительность к обратной связи со стороны коллег; перепроверка выводов на других фрагментах полученного материала) и т.д. Среди прочих подходов к обоснованию достоверности результатов следует упомянуть установление степени правдоподобия (plausibility) как оценки полученного знания с позиции уже имеющихся знаний; доверие (credibility) как оценка и коллективного осмысления результатов с учетом природы феномена, и обстоятельств его наблюдения; укорененность в данных, включенность в контекст исследовательской программы (dependability), которая основана на тщательном изучении и оценке процедурных аспектов; чувствительность как способность исследователя увидеть социальную проблему и способствовать ее решению; онтологическая и образовательная аутентичность - способность повысить сознательность участников исследования (в первом случае) и их окружения (во втором); каталитическая аутентичность как влияние на социальные программы, способствующие улучшению качества жизни изучаемой популяции.

Постмодернистскими авторами обсуждаются принципиально новые понятия и принципы обоснования знания: ироническая (Ж.Бодрийар), неопрагматическая (Ж.-Ф.Лиотар), ризоматическая (Ж.Деррида) валидизация. Н

.Дензин обосновал метод триангуляции как радикальной альтернативы традиционным подходам к валидизации. Триангуляция есть сочетанное и взаимообогащающее использование разных методов, методологий, данных, теорий и/или исследователей. Множественность подходов и приемов призвана ослабить эпистемологические ограничения, перейти установленные границы, выявить новые стороны феномена. Следует отметить, что идея триангуляции [истины] была порождена в постпозитивизме (Д.Кэмпбелл), где она понималась как некоторая процедура, набор полезных трюков: в дополнение к опросу "обычных" людей опросить экспертов; случайным образом расщепить выборку испытуемых пополам и провести анализ данных отдельно; исключить из анализа одну переменную и посмотреть, как изменится модель; валидизировать конструкт с помощью подхода "много черт - много методов" и т.д.

Чтобы избежать нежелательных ассоциаций с позитивизмом, постмодернистская исследовательница-этнограф Л.Ричардсон в радикальном духе отказывается от идеи триангуляции и предлагает иную привлекательную метафору - кристалла, а не треугольника, кристаллизации, а не триангуляции. Кристалл одновременно создает, преломляет и искажает, он многогранен, и ни одна из его граней не более достоверна, чем любая другая. Как и знание, кристалл растет и разрушается.

Метафорический подход к проблеме обоснования знания и оценки его качества неслучаен. Он призван разрушить критериальные подходы и представления о стандартах. Социальное исследование обладает ценностью лишь тогда, когда дает равные возможности высказаться представителям разных групп, создает равновероятные версии и подавляет доминирующие, "правильные" интерпретации.

С.В. Сивуха