



Исследования развития и бытия личности

Андрей Хвостов

ПРОБЛЕМЫ СТАТИСТИЧЕСКОГО АНАЛИЗА В ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Современная психология и социология широко используют самые разнообразные статистические методы. Они позволяют наглядно описать некоторое явление или процесс, выявить закономерности, сделать выводы или прогноз. Как пишет Е.В. Сидоренко [1]: «Стало принято использовать математические методы, как принято жениться молодому человеку, если он хочет сделать дипломатическую или политическую карьеру...» При этом «мода» иногда доходит до того, что при планировании эксперимента предлагается строить гипотезу, исходя из расчета определенных статистических процедур получения результатов, их оценивания и анализа [2], а статистическая верификация выводов считается обязательной.

Что важнее — качественный или количественный анализ?

С другой стороны, большинство исследователей полагают, что основным методологическим принципом обработки полученных в эксперименте данных является неперенный сплав количественного и качественного анализов, и сама по себе количественная оценка ничего не значит без качественного акцента, без содержательной характеристики [3]. Например, в таком подходе иногда достаточно описать признак и убедиться в

Основные задачи
статистического
анализа

его представленности, поскольку ему невозможно приписать ранг или числовое выражение. Центр тяжести доказательства перемещается в подобных ситуациях в другую плоскость: нужно установить не уровень выраженности признака, не степень его отличия от другого уровня, а прежде всего подметить, обнаружить данный признак, как таковой. Если удастся это сделать, то никаких других показателей для диагноза не требуется, специалист понимает, что наличие данного признака выделяет человека или группу из среды других людей, которые данным признаком не обладают [4]. Например, если китайцы, в отличие от русских, вообще не демонстрируют агрессию на депривацию отдельного звена самосознания, то статистической обработки не предполагается; вполне обоснованно говорить о некоторых качественных отличиях двух культур.

Тем не менее цель современной психологической диагностики в ее наиболее распространенных формах заключается в том, чтобы фиксировать и описывать в упорядоченном виде психологические различия как между людьми, так и между группами людей. Статистические приемы при этом привлекаются для того, чтобы доказать, что сравниваемый признак (или признаки) свидетельствует об отличиях диагностируемых — лица или группы — от тех, с кем сравнивают, и что это отличие значимо. Статистический анализ часто применяется для выявления связи между отдельными психическими процессами, условиями и т.д. Все более распространенными становятся методы многомерного анализа — факторный и кластерный, которые предполагают качественную интерпретацию, позволяют описать структуру феномена или выявить качественно отличные типы.

Притом что каждая из обозначенных целей (из которых перечислены только некоторые) предполагает целый спектр математических моделей, каждая из которых адекватна в определенных условиях. Трудности в эксперименте начинаются еще раньше, до выбора статистического метода.

Проблема 1: необходимая и достаточная
выборка

Начнем с того, какую выборку можно считать репрезентативной? Уже здесь мнения расходятся. Так, П. Клайн полагает, что в каждой подгруппе должно быть как минимум 300 человек [5]. Подгруппа образуется такими стратификационными переменными, как пол, возраст, образование, социальное положение и т.д. Например, одной отдельной подгруппой будут мужчины от 18 до 25 лет с высшим образованием. Э. Ноэль [6] предполагает, что необходимым минимумом является выборка из 200 человек. Однако такие количества указываются для расчета строгих тестовых норм или репрезентативных социологических выборок. Предполагается, что такие выборки, как тридцать бизнесменов, двадцать

два инженера из технических лабораторий — относительно бесполезны, поскольку не могут использоваться как нормативная группа. С другой стороны, маленькая выборка имеет «право на существование». Во-первых, для рабочих целей принято иметь не менее тридцати измерений исследуемого признака (в одной системе условий) [7], но не уточняется, какой анализ предполагается для воспроизводимых результатов. И.И. Елисеева и М.М. Юзбашев утверждают, что при 30 наблюдениях погрешность становится значительной, несоответствие результатов наблюдается уже при значениях менее 100 [8]. П. Клайн полагает, что для воспроизводимых результатов факторного анализа необходимо 100 испытуемых [9], хотя для максимального уменьшения стандартных погрешностей требуется примерно 200 испытуемых. «Магическая» цифра в 100 наблюдений приводится Дж. Гласс и Дж. Стенли [10], которые считают выборку достаточной для получения выборочных распределений, близких к нормальному в большинстве случаев, и другими авторами также в связи с тем, что она является основанием для применения несколько различных законов распределения [11]. Таким образом, хотя мнения расходятся, но друг другу не противоречат. Здесь стоит обратиться к мнению Р. Готтсданкер, который принимает во внимание статистические ошибки совершенно иного рода.

- Р. Готтсданкер полагает, что по мере значительного увеличения числа испытуемых с целью увеличения силы проверки на значимость происходит любопытная вещь [12]. Если значимость не будет обнаружена на 20 испытуемых, она появится для 200, или 2000, или 2 000 000. В этом нет никакой мистики. Возможно, что буквы распознаются лучше, чем числа, только потому, что в каждой тысяче находится несколько испытуемых, которые реагируют отрицательно на числа в силу неудачного опыта в школьной математике. Соответственно, на большой выборке будет показано, что числа распознаются хуже букв, хотя реально это справедливо в отношении только тех нескольких испытуемых из тысячи. Такие дополнительные факторы обнаруживают всегда свое присутствие при анализе больших массивов данных. Поэтому с особенной осторожностью следует подходить к случаям, которые требуют для выявления действия независимой переменной слишком большого количества данных. Гораздо большее впечатление производят статистически значимые различия, полученные на относительно небольшом числе испытуемых или проб.

Так сколько все же испытуемых необходимо и достаточно? Вопрос необходимо дифференцировать. Прежде всего, ограничимся психологическими исследованиями, не рассматривая далее социоло-

гию. Далее, достаточных для воспроизводимых выводов? Для интерпретируемых? Или для достоверных (статистически значимых)? Для воспроизводимых — чем больше, тем лучше. Для интерпретируемых достаточно минимума, при этом, как мы видели, увеличение выборки даже нежелательно. Что касается значимости, то она, если верить Р. Готтсданкер, зависит не столько от самого психологического феномена, сколько от факторов, которые к исследуемому явлению отношения не имеют. Таким образом, в условиях, когда точную цифру оптимальной выборки назвать нельзя, можно лишь рекомендовать, принимая во внимание «средневзвешенное» мнение ряда авторов. Для рабочих целей, экспериментов, носящих пилотажный характер, не претендующих на репрезентативность, можно установить минимальный объем подгруппы в 30 человек, который обозначается как малая выборка, оптимальный — около 100 человек, выборка с большим объемом, безусловно, считается большой [13].

Дальнейшее увеличение числа испытуемых в подгруппе нецелесообразно (если речь идет о пилотажных исследованиях), при выборке более 200 человек в подгруппе действительно выявляется большое количество «новых» закономерностей, которые статистически значимы и требуют интерпретации. Но реально они отражают очень небольшие в абсолютных величинах связи (между переменными) или отличия (от других подгрупп), которые субъективно самим исследователем не воспринимаются и трактовке поддаются с большим трудом, если поддаются вообще. Например, если утрировать, можно найти на 1000 испытуемых связь между темнотой волос и экстраверсией; она может быть очень небольшой (линейная корреляция 0,081 в абсолютных значениях), но статистически достоверной (на 1% уровне значимости). Тогда как равной по достоверности мерой связи для 30 испытуемых будет показатель 0,463 (вместо 0,081). Как говорится, почувствуйте разницу.

Проблема 2: закономерности реальные и случайные

Как можно снизить вероятность ошибки (она называется ошибкой 1-го рода), при которой мы принимаем гипотезу о наличии связи, различиях и т.д. Прежде всего — реальная воспроизводимость результатов. В некоторых случаях связи (закономерности в целом) неустойчивы, поскольку неустойчивы порождающие их условия. Например, закономерности, выявленные в детском возрасте, совершенно нельзя переносить на старшие возрастные группы. Более того, могут «отменяться» закономерности, вызванные спорадическими социальными условиями. В частности, закономерности могут возникнуть в результате перекрещивающегося действия нескольких законов, и не всегда очевидно, с действиями каких законов мы имеем дело [14].

Однако закономерность на данном этапе социогенеза и филогенеза может быть реальной, хотя и достоверно значимой только для определенных условий. Отделить случайность от закономерности здесь уже довольно сложно, поскольку определить действие отдельных факторов иногда крайне сложно. Первое, что можно предложить в таком случае, — исходить из теоретических предположений, заранее производить качественный отбор тех характеристик явления, который будет подвергнут статистической обработке. Конечно, это придает исследованию «ограниченный» характер и могут упускаться реальные закономерности, носящие латентный характер.

Если же принимать во внимание все закономерности, выявленные в ходе эксперимента, можно обращать внимание только на те закономерности, которые опять же соответствуют ряду критериев, не имеющих статистического характера. Это снова степень соответствия результатов теоретическим представлениям. В частности, это может быть степень соответствия отдельной закономерности общей структуре явления. Обобщенно — полученные данные должны относительно хорошо интерпретироваться [15]. Конечно, сразу возникает вопрос о субъективности интерпретации. Е.В. Сидоренко на это справедливо отвечает, что субъективна любая интерпретация, и также подчеркивает, что критерием адекватности статистической обработки является осмысленность результатов. Тем не менее защитой от субъективности может быть опрос экспертов. Например, автор иногда прибегал к такому приему: просил сотрудников (лаборатория социально-психологических проблем сознания ПИ РАО) обозначить (назвать) выделенные факторы или кластеры дескрипторов. В дальнейшем, иногда даже не соглашаясь с большинством экспертов, высказанная ими точка зрения часто помогает посмотреть на явление с новой точки зрения. Экспертами совершенно не обязательно должны быть психологи. Например, при сравнении двух культур лучшим экспертом будет человек, хорошо знакомый с обоими. Именно он сможет подтвердить, обоснованы ли выводы об отличиях или сходстве.

Проблема 3: выбор
статистического
метода

Следующий вопрос, который требует рассмотрения в связи с проблемой некоторого противоречия воспроизводимости и интерпретируемости, — проблема мощности статистического критерия. Собственно, в общем виде мощность описывается как способность показать отличия (связи), а не допустить ошибку, состоящую в принятии гипотезы об их отсутствии. Здесь перед исследователем встает несколько проблем. Первая — выбор статистического критерия. И здесь мнения расходятся в большей степени, чем при определении выборки.

Как полагает Б.А. Сосновский, практика статистической обработки требует следующего: в случае обнаружения недостоверности различий (там, где они предполагаются) с помощью какого-либо простейшего критерия нужно прибегнуть к другому критерию — критерию более мощному (чувствительному). Что получается при таком подходе? Как показывает сам А.Б. Сосновский, серийный критерий при сравнении двух выборок может не показывать различия, но оно будет значимым на пятипроцентном уровне при применении критерия числа инверсий (Вилкоксона — Манна — Уитни) [16]. Но если пойти далее, то на тех же условных выборках можно, применяя t-критерий Стьюдента, получить различие не только на пятипроцентном уровне значимости, но и на однопроцентном уровне. Для этого вполне достаточно предположить (и проверить) гипотезу о том, что выборки связаны или что они находятся в одном так называемом «хвосте» распределения.

Е.В. Сидоренко полагает, что выбор статистической модели следует прежде всего определять, исходя из гипотезы (модель должна подходить для доказательства научных предположений) [17]. Если правильно понимать, то если нужны отличия (связи) — использовать мощный критерий, не нужны — не столь мощные. Далее обращать внимание на ограничения критерия. Однако, как указывает Ю.Н. Толстова, для некоторых методов, показавших свою эффективность при решении практических задач, отсутствуют строгие доказательства корректности их применения [18]. Более того, в одном из подходов выбор скорее предполагает не столько использование корректного метода, сколько анализ того, в какой мере могут быть нарушены условия (ограничения) применимости известных методов, чтобы результаты «не слишком» искажали реальность. Обобщенно — применять заведомо корректные методы, максимально уменьшая возможные искажения.

Например, для измерения связи даже между двумя номинальными признаками могут служить более сотни известных коэффициентов соответствующего плана [19]. Хотя самым мощным и теоретически обоснованным непараметрическим является критерий χ^2 Пирсона [20], он имеет некоторые ограничения и должен в соответствующих условиях заменяться другими. Но, учитывая его преимущества, иногда рекомендуется упорядочивать исходные данные «под критерий» — соблюдать минимальную частоту отдельного признака, общий объем распределений и т.д.

Хотя субъективность при этом остается. Например, возвращаясь к рассмотренному выше примеру возможности предположить (и проверить) гипотезу о том, что выборки связаны или что они находятся в одном так называемом «хвосте» распределения. Применяя t-критерий Стьюдента, можно получить различие не только

на пятипроцентном уровне значимости, но и на однопроцентном уровне. Как считают Дж. Гласс и Дж. Стенли, всегда есть возможность «объединить в пары» данные из двух различных выборок [21]. Пары можно описать наименованиями типа «брат — сестра», «до — после» и т.д. При этом применяется модель t-критерий Стьюдента, показывающая гораздо более достоверно значимые отличия, чем при использовании модели для независимых выборок. То же относится к «хвостам распределения», хотя для этой гипотезы нужны весьма веские основания. Допустим, что мы исследовали одних и тех же испытуемых до и после некоторого события; тогда выборки связаны. Однако возможно ли корректно связать (и обосновать как зависимые) такие выборки, как «брат и сестра», — проблематично. Так же можно согласиться с правомерностью исследований эффективности рекламы (до и после) с использованием модели зависимых выборок, но сомнительно, когда речь идет о сравнении курсов двух акций [22]. Если подойти к проблеме «по Б.А. Сосновскому» — надо использовать более мощный критерий и искать различия; если «по Е.В. Сидоренко» — тот, который больше соответствует гипотезе.

Проблема 4: выбор
статистического пакета
(программы)

Наконец, еще одна проблема может возникнуть уже не из-за методологических, сколько чисто математических несоответствий моделей, заложенных в разных программах. В настоящее время широко используются такие статистические пакеты программ, как Microsoft Excel, SPSS, Minitab Statistical Software и др. Однако притом что в большинстве случаев результаты обработки данных, полученные в разных программах, совпадают, возможны и большие «разногласия». Так, широко используемый критерий χ^2 (Пирсона) вычисляется (скорее, используется) по-разному. По сути, он предназначен для сравнения частот распределений: двух эмпирических или эмпирического и теоретического. Однако в программах Microsoft Excel и SPSS возможно соотнесение только эмпирических и теоретических (ожидаемых) распределений, в программе Minitab Statistical Software — только двух эмпирических. Не вдаваясь в тонкости, используя программы Microsoft Excel и SPSS, нельзя прямо установить, отличаются ли частоты, например, реакций на депривацию у мужчин и женщин, русских и китайцев. Но это позволяет сделать программа Minitab Statistical Software, которая имеет свои ограничения. Однако проблема не столько в возможностях, сколько в том, что эти возможности практически не «озвучиваются». Неопытному пользователю может быть совершенно непонятно, что он высчитывает.

К сожалению, если в таких «классических» методах анализа, как χ^2 или Т-Тест, еще можно разобраться с помощью учебников, то в ряде случаев дело обстоит сложнее. Например, популярные многомерные методы анализа данных (типа факторного или кластер-

ного анализа), если рассматриваются, то в большинстве случаев на математическом языке, иногда — на уровне «невнятных» абстракций. Так, в пакете SPSS-10 предусматривается 8 способов измерения расстояний между объектами и 7 — кластеризации (расстояния между кластерами). Возможно много комбинаций, каждый приведет к своему результату. Что выбрать?

А.М. Дубров с соавторами предлагают: когда формализовать цель задачи не удастся, критерием качества классификации считать возможность содержательной интерпретации найденных групп [23]. Весьма субъективный критерий и возможностей для «махинаций» очень много. Также рекомендуется сравнение результатов, проводимых разными процедурами, но конкретно алгоритм не описан. Несколько более объективные критерии в выборе расстояний предлагают И.И. Елиссеева и М.М. Юзбашев [24]. В.П. Боровиков хотя и рекомендует «вжиться» в стоящую задачу для выбора метода (поскольку однозначного ответа не существует), но все же дает вполне понятные описания преимуществ и недостатков отдельных алгоритмов [25]; меры связи между членами кластера, величина самого кластера и т.д.

Резюме

В качестве решения тех проблем, которые были поставлены в данной статье, можно предложить следующее. Использовать «необходимо достаточные» выборки (минимум 30, желательно от 50, оптимально около 100 человек). Применять проверенные методы статистической обработки (они достаточно мощные, чтобы выявить закономерности), а не идти от «слабых» к «сильным»; лучше выявить «лишнюю» связь, чем ее пропустить. Если полученную закономерность не удастся интерпретировать, стоит проверить ее на воспроизводимость. Интерпретировать только то, что интерпретируется, и при этом можно обратиться к мало используемой в последнее время процедуре — экспертной оценке; попросту говоря, спросить у специалиста, похоже ли это на реальность и что результат может означать.

Автор понимает, что в настоящей статье демонстрируется только «вершина айсберга»; проблем статистического анализа гораздо больше. Однако, как представляется, все же удалось показать общие пути решения некоторых основных проблем и в ряде случаев дать конкретные рекомендации.

-
1. *Сидоренко Е.В.* Методы математической обработки в психологии. СПб., 2002. С. 5.
 2. *Ермолаев-Томин О.Ю.* Прикладная статистика для психологов: Программа спецкурса. М., 1998. С. 3.
 3. *Гуревич К.М.* Проблемы современной психологической диагнос-

- тики // Психологическая диагностика: Проблемы и исследования / Под ред. К.М. Гуревича. М., 1981. С. 14.
4. Там же. С. 19.
 5. *Клайн П.* Справочное руководство по конструированию тестов. Киев, 1994. С. 215.
 6. *Ноэль Э.* Массовые опросы. М., 1978.
 7. *Сосновский Б.А.* Лабораторный практикум по общей психологии. М., 1979. С. 18.
 8. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики: Учебник. М., 2000. С. 190.
 9. *Клайн П.* Справочное руководство по конструированию тестов. Киев, 1994. С. 251.
 10. *Гласс Дж., Стэтли Дж.* Статистические методы в педагогике и психологии. М., 1976. С. 222.
 11. *Дубров А.М., Мхитарян В.С., Трошин Л.И.* Многомерные статистические методы: Учебник. М., 1998. С. 49.
 12. *Готтсданкер Р.* Основы психологического эксперимента. М., 1982. С. 253.
 13. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики: Учебник. М., 2000. С. 190.
 14. *Маслов П.П.* Статистика в социологии. М., 1971. С. 53.
 15. *Мельников В.М., Ямпольский Л.Т.* Введение в экспериментальную психологию личности: Учебное пособие. М., 1985. С. 107.
 16. *Сосновский Б.А.* Лабораторный практикум по общей психологии. М., 1979. С. 137–138.
 17. *Сидоренко Е.В.* Методы математической обработки в психологии. СПб., 2002. С. 35.
 18. *Толстова Ю.Н.* Анализ социологических данных. М., 2000. С. 87.
 19. Там же. С. 95.
 20. *Суходольский Г.В.* Основы математической статистики для психологов. Л., 1972. С. 294.
 21. *Гласс Дж., Стэтли Дж.* Статистические методы в педагогике и психологии. М., 1976. С. 270.
 22. *Боровиков В.П.* Программа STATISTICA для студентов и инженеров. М., 2001. С. 109–117.
 23. *Дубров А.М., Мхитарян В.С., Трошин Л.И.* Многомерные статистические методы: Учебник. М., 1998. С. 242.
 24. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики: Учебник. М., 2000. С. 133–155.
 25. *Боровиков В.П.* Программа STATISTICA для студентов и инженеров. М., 2001. С. 186–187.

