

ТЕСТЫ ШКОЛЬНЫХ ДОСТИЖЕНИЙ: НОРМИРОВАНИЕ ТЕСТОВ

// Школьные технологии. – 1999. - № 3. –
С. 179 – 198.

Уривки

ВЫБОРКА

ВАРИАНТЫ РАСЧЕТА ОБЪЕМОВ ВЫБОРКИ

Корректное построение выборки — основа для стандартизации: все зависит от выборки заданий для теста и выборки репрезентативной выборочной совокупности. При формировании выборки следует учитывать две важные переменные: объем и репрезентативность. Выборка должна точно отражать категорию лиц, для которых предназначен тест (конечно, может быть несколько таких категорий и, следовательно, и несколько выборок или же одна большая выборка), а также быть достаточной большой для обеспечения столь малой стандартной погрешности нормативных данных, чтобы ею можно было пренебречь. Ниже приведены расчеты объемов выборки для всех участников педагогического процесса, а не только учеников.

Объем выборки — число единиц наблюдения, включаемых в выборочную совокупность.

Что определяет объем выборки, и от чего он зависит? В общем виде объем выборки определяется следующими четырьмя обстоятельствами:

1. Задачи и условия проведения исследования,

1.1. Получение экспресс-информации, фундаментальное исследование, необходимость получения прогноза по результатам исследования требуют разной степени точности и определяют точность и доверительную вероятность результатов проводимого исследования.

1.2. Объем выборки зависит от разрешающей величины шкалы изучения (дихотомическая, процентная, семантический дифференциал) и требует учета при выборе способа расчета объема выборки.

1.3. От условий проведения зависит способ определения объема выборки, при этом возможны три случая:

— об измеряемых величинах есть статистическая информация по предыдущим исследованиям;

— будет проведено пробное исследование;

— о характеристиках статистических величин ничего не известно.

В последнем случае возможно получение оценок, необходимых статистических данных. Дисперсия может быть оценена по правилу “шести сигм”:

$$\delta = \frac{x_{\max} - x_{\min}}{x_{\min}} = \frac{R}{6}$$

где x_{\max} — максимальное значение шкалы измерения,

x_{\min} — минимальное значение шкалы.

Остальные необходимые статистические данные могут быть либо заданы, либо оценены как максимально теоретически возможные. Таким образом, задача расчета объема выборки в этом случае может быть сведена к задаче с известными статистическими данными.

2. Степень однородности генеральной совокупности. Объем выборки определяется разбросом измеряемого показателя в генеральной совокупности, выражаемым дисперсией или стандартным отклонением.

3. Вероятность, с которой гарантируется достоверность результата.

Определяется условиями исследования. Обычно в практике исследований ее величина колеблется от 85% до 99%. Наиболее часто используется вероятность 95% (0,95).

4. Точность результатов: определяемая предельной ошибкой репрезентативности.

Предельная ошибка репрезентативности задается обычно: в пределах от 0,01 до 0,10 с наиболее частым употреблением: 5% (0,05).

Определяющим фактором является первое обстоятельство (задачи исследования), но объем выборки зависит также от остальных трех обстоятельств (степень однородности генеральной совокупности, вероятность, с которой гарантируются результаты, предельная ошибка репрезентативности).

Дадим обозначения нескольким известным статистическим понятиям, без которых дальнейшее объяснение затруднено;

P — вероятность,

n — объем выборки,

Δ — предельная ошибка репрезентативности,

σ^2 — дисперсия,

σ — стандартное отклонение,

N — объем генеральной совокупности,

t — табулированная константа, смысл которой определим несколько позже (см. расчеты по формулам),

$\alpha = 1 - P$ — уровень значимости.

ОБЩИЙ ПОДХОД К РАСЧЕТУ ОБЪЕМА ВЫБОРКИ

Основная величина, от которой зависит величина объема выборки — дисперсия и ее поведение (изменение). При выборочном методе исследования дисперсия изучаемого признака складывается из двух составляющих — дисперсии генеральной совокупности (собственно мера рассеяния признака как такового) и дисперсии, вызванной неоднородностью генеральной совокупности. При малых объемах выборки обе составляющие вносят в нее свой вклад, при этом вторая составляющая может как увеличивать, так и уменьшать дисперсию. С увеличением выборки, составляющая, вызванная неоднородностью генеральной совокупности уменьшается, значение дисперсии при этом перестает изменяться, при этом дальнейшее увеличение объема выборки нецелесообразно, поскольку ошибка выборки становится пренебрежительно мала.

Графически это может быть выражено следующим образом:

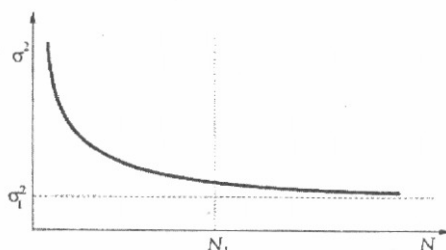


Рис. 1. График зависимости дисперсии от объема выборки

N_1 — оптимальный объем выборки.

СПОСОБЫ РАСЧЕТА ОБЪЕМА ВЫБОРКИ

ЗАМЕЧАНИЕ. Объем выборки, рассчитываемый в данной части, относится к:

1) чисто случайной выборке (безотносительно к способу построения выборки — по таблицам случайных чисел, периодическому и пр.);

2) бесповторному случаю — один и тот же испытуемый в одном исследовании дважды не тестируется.

Это обусловлено тремя обстоятельствами:

а) объем выборки для чисто случайного способа отбора не является максимальным, однако в практике применяется только этот способ;

б) объем выборки для доли и среднего рассчитывается одинаковым способом, считая признаки альтернативными;

в) расчеты более сложных выборок уменьшают объем выборки и их проведение подробно описано в литературе.

Степень близости выборочной средней к генеральной средней при постоянной дисперсии при повторном отборе зависит только от объема выборки, а не от удельного веса выборочной совокупности в генеральной. Объем выборочной совокупности не зависит от объема генеральной

Объем выборочной совокупности рассчитывают четырьмя основными способами.

1. По изменению дисперсии.
2. По таблицам достаточно больших чисел и номограммам больших чисел.
3. Эмпирически по среднему квадратичному отклонению.
4. По формулам математической статистики.

Приведем варианты расчета объемов выборки всеми четырьмя способами из одинаковой генеральной совокупности.

1. По изменению дисперсии

Расчет объема выборки по изменению дисперсии вытекает из общего подхода к расчету объема выборки. Применяется, когда известны значения статистик, характеризующих различные объемы совокупностей.

Для примера изучения образовательной системы Санкт-Петербурга взяты следующие совокупности: учителя, родители, учащиеся. Расчеты проведены

для двух вопросов в каждой группе — один носит оценочный характер по отношению к образовательному заведению, второй — самооценка.

В таблице представлены простые статистики для разных выборок совокупностей для наиболее употребительной пятиразрядной шкалы.

Курсивом в таблицах выделены значения дисперсии, статистически равнозначные, и обозначен момент, с которого дальнейшее увеличение выборки не уменьшает дисперсию. Таким образом, для данных вопросов объемы выборки могут быть определены в 150—350 единиц.

Таблица 1.

категория	объем выборки	дисперсия вопрос 1	дисперсия вопрос 2
Учителя	100	0,35	0,54
	150	0,51	0,55
	200	0,46	0,67
	300	0,43	0,63
	400	0,44	0,62
	500	0,44	0,62
	600	0,43	0,62

Таблица 2.

категория	объем выборки	дисперсия вопрос 1	дисперсия вопрос 2
Ученики	50	1,37	1,86
	100	1,25	1,81
	200	1,03	1,89
	400	1,01	1,95
	500	1,00	1,93
	700	1,00	1,93

Таблица 3.

P	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02
0,85	51	63	80	105	143	207	323	575	1295
0,90	67	83	10*	138	187	270	422	751	1660
0,95	96	118	150	195	266	384	600	1067	2400
0,99	165	204	259	338	400	633	1036	1843	4146
0,997	220	271	344	449	61	880	1376	2446	5504
0,999	270	334	422	552	751	1082	1691	3007	6767

Достаточно большие числа наблюдений при допустимой ошибке (α) и доверительных вероятностях (P), рассчитанные без учета величины N — объем генеральной совокупности.

2. РАСЧЕТ ОБЪЕМА ВЫБОРКИ ПО ТАБЛИЦАМ ДОСТАТОЧНО БОЛЬШИХ ЧИСЕЛ И НОМОГРАММ ДОСТАТОЧНО БОЛЬШИХ ЧИСЕЛ

Расчеты объема выборки по номограммам и таблицам дают примерно одинаковые результаты, но применяются при разных условиях. Таблицы достаточно больших чисел используют, если ничего нельзя сказать ни о средних показателях, ни о генеральной совокупности. При этом используются следующие вполне корректные (не уменьшающие объема выборки) допущения;

— объем генеральной совокупности не влияет на объем выборки;

— численность выборки зависит от P и величины вероятности, с которой делается заключение о достоверности выводов предельной ошибки репрезентативности и вероятности появления события.

Поскольку таблицы достаточно больших чисел построены с расчетом максимального значения признаков, то полученное число окажется несколько больше необходимого. Одним из видов таблиц достаточно больших чисел является следующая таблица 3.

Чтобы найти по таблице достаточно большую численность выборки для обеспечения точности оценки $\alpha = 0,05$, ее надежности (доверительной вероятности) $P = 0,95$, на пересечении находим число 384, которое и является исходным достаточно большим числом.

Для случая изучения образовательных систем, наиболее употребителен максимальный объем выборки в 400—600 единиц.

Возможная ошибка тем меньше, чем больше численность выборки. Причем, чтобы уменьшить ошибку вдвое, численность выборки приходится увеличить в 4 раза. В таблице выделены значения выборок наиболее употребительных в практике социально-педагогических исследований на уровне города. По этой же таблице можно провести обратную работу — определение предельной допустимой ошибки и вероятности при известной выборке.

Если рассеяние признака известно на основании предыдущих исследований, то объем выборочной совокупности может быть определен по номограммам больших чисел, при этом объемы выборки при тех же значениях P и α меньше, чем определенные по таблицам.

3. РАСЧЕТ ОБЪЕМА ВЫБОРКИ ЭМПИРИЧЕСКИМ СПОСОБОМ

Определение объема выборки по процентному выражению среднего квадратического отклонения от величины измеряемого параметра. Согласно эмпирическому закону объем выборки увеличивают до тех пор, пока среднее квадратическое отклонение, выраженное в процентах от величины средней, не станет меньше или равным 50 %.

Проведем расчет для вопроса 1 родителей.

выборка	отношение, %%
150	52

300	49
400	46
500	44
700	45

Как видно из расчета, устойчивое значение отношения менее 50% соответствует выборке в 300 — 400 единиц. Аналогичные расчеты проведенные для учеников и учителей дают значения 200 — 300 и 300 — 400 соответственно.

4. РАСЧЕТЫ ПО СТАТИСТИЧЕСКИМ ФОРМУЛАМ

Напомним еще раз условия, для которых проведены настоящие расчеты (см. замечание в начале этого раздела).

Различные источники дают одну формулу для расчета объема выборки для нашего случая:

$$n = \frac{t^2 \cdot \delta^2 \cdot N}{N \cdot \Delta^2 + t^2 \cdot \delta^2}$$

где n — объем выборки.

Некоторые сложности появляются с определением величины t .

В разных работах эта величина определяется по-разному. Она определяется как нормированное отношение или стандартизированная разность и заданы ее значения:

$t = 2$ при $P = 0,95$ и $t = 3$ при $P = 0,99$, где P — вероятность (надежность заключения).

Эта величина может не вычисляться, а устанавливаться равной 4.

Она определяется как критическая точка стандартизированного нормального распределения и составляет:

$t = 1,6449$, при $\alpha = 0,05$,

и $t = 2,3263$, при $\alpha = 0,01$, где α — уровень значимости.

Эта величина определяется также как величина, связывающая среднюю (μ) и предельную ошибку выборки (Δ)

$$t = \frac{\Delta}{\mu}$$

Отсюда становится ясным статистический смысл этой величины — мера риска допущения ошибки связанной с выборкой — предельная ошибка выборки равна t -кратному числу предельных ошибок выборки. Табличные значения этой величины следующие:

$t = 1,96$, при $\alpha = 0,05$, и

$t = 2,58$, при $\alpha = 0,01$.

В этой же работе определены источники этих значений, вытекающие из значений функции Лапласа для нормального распределения. Кроме того, в этой работе для выяснения этой величины рекомендуется использовать эмпирическую формулу:

$$t = 3 + \frac{6}{n - 4}$$

Подробное рассмотрение этой величины вызвано ее значительным влиянием на получаемый в результате расчетов объем выборки, Разброс итоговых значений количества единиц наблюдения весьма значителен при констатировании остальных составляющих, так для случая $N = 29962$ — количества учителей в Санкт-Петербурге, при равенстве прочих переменных получаем следующие значения:

При $t = 2$, $n = 1198$, а

при $t = 3$, $n = 1752$, таким образом, разброс составляет более 600 единиц, что для практики исследований весьма существенно.

Расчет объема выборки по формулам для различных совокупностей образовательной системы города дает следующие результаты:

1. Для учителей города, при $N = 29962$, $n = 286$;
2. Для учащихся в школах города $N = 540481$, $n = 298$;
3. Для образовательных учреждений типа «школа М662,л—7В»;
4. Для классов $N = 2018$, $n = 184$.

НЕКОТОРЫЕ ВЫВОДЫ

Рассчитанные четырьмя различными способами объемы выборок дают примерно одинаковые результаты при сохранении показателей качества.

Достаточными объемами выборки можно признать значения в пределах от 150 до 600 единиц — это максимальное и минимальное значения для выборок с нашими граничными условиями. При этом необходимо заметить, что построение более сложных выборок приводит, как правило, к некоторому уменьшению количества измеряемых объектов.

В заключение необходимо сказать о стойком убеждении среди педагогов и социологов в необходимости 5 или 10 % выборки для получения надежных результатов. Анализ литературных источников, посвященных расчетам объемов выборок, ни теоретически, ни эмпирически не подтверждает такие объемы.

ОЦЕНОЧНЫЙ ПОДХОД К РАСЧЕТУ ОБЪЕМА ВЫБОРКИ С УЧЕТОМ СТРАТИФИКАЦИИ

Как мы показали, для простого уменьшения значения стандартной погрешности вполне достаточна будет выборка из 300—500 испытуемых. Однако репрезентативность выборки не прямо зависит от ее объема, до сих пор мы рассматривали выборки без выделения страт. Для стратифицированных выборок существуют отдельные формулы расчета объемов. Однако для практической работы более полезно будет рассмотреть вопрос оценки, а не расчета объема выборочной совокупности.

При получении нормы для общей популяции, например, детей школьного возраста, необходима выборка больше, чем выборка из столь ограниченной популяции, как укротители львов или факиры. До начала исследования нельзя сделать никакого утверждения относительно объема выборки безотносительно той популяции (категории лиц), из которой она подбирается. Здесь проясняется тот момент, что репрезентативность выборки является более важной, чем ее размер. Маленькая, но репрезентативная нормативная выборка будет предпочтительнее, чем большая, но непропорционально представленная.

Получение репрезентативной нормативной выборки

Ясно, что наиболее неоднородной популяцией является генеральная популяция (все население), а все остальные являются ее подмножествами. По этой причине получить выборку заданного определенного качества из генеральной популяции наиболее трудная задача. Однако в практике тестов школьных достижений этого и не нужно.

Правила для общих норм

Принципиальным для оценки объема выборки является то, как будут использованы полученные данные. Если на их основе будут получаться нормы, то проведение расчетов объемов должно быть максимально строгим, если же предполагается критериальное оценивание, то оно требует меньшей точности и, соответственно, меньших затрат.

Для получения норм для генеральной популяции можно сформулировать несколько общих правил:

1. Выборка должна быть стратифицирована. То есть в выборке должны быть представители разных групп, реально существующих в генеральной совокупности, желательно в такой же пропорции.

Кроме этого, страты должны быть выделены таким образом, чтобы дисперсия их ответов была максимальной. Обычно в психолого-социологических исследованиях адекватной

признается стратификация на четыре уровня. Важными стратификационными переменными являются социальное положение, возраст и пол. для тестов школьных достижений — возраст и пол.

2. В каждой подгруппе должно быть достаточное количество испытуемых, чтобы сформировать адекватную выборку. Выборка для всего населения означает, что с учетом всех возможных классификаций, необходимо очень большое количество испытуемых, например: 4 (социальное положение) \times 2 (пол) \times 5 (возраст) дает 40 категорий по 300 (среднее значение предварительных расчетов) испытуемых в каждой, что дает общее количество 12 000 испытуемых. Следовательно, обеспечение адекватности норм для генеральной совокупности требует огромных затрат ресурсов. Нормы, полученные в менее масштабных исследованиях, также полезны, но должны использоваться с осторожностью.

По приведенной выше причине, а именно из-за необходимости привлечения огромных ресурсов, многие разработчики психологических тестов указывают более специфические групповые нормы, специально соответствующие характеру и практическому применению теста.

Формирование выборок для специфических групп

Для получения адекватных норм для специфических групп необходимо стратифицировать выборку по основной переменной, влияющей на эти группы. На примере будет показано, как это делается в целом, хотя для выбора стратифицирующих переменных необходимы значительные исследования конкретной группы.

Учащиеся одной параллели

Чтобы получить соответствующую выборку для учащихся одной параллели, необходимо обеспечить ее адекватность. Так, основанием для формирования выборки может быть подбор всех учащихся из шестых классов. Для школ некоторыми основными стратифицирующими переменными являются:

1. Тип образовательного учреждения или реализуемой образовательной программы — школа, гимназия, лицей, авторская школа. Как показывает опыт, это разделение на категории можно отбросить, поскольку наиболее существенным для стратификации является тип образовательной программы, выделенный по следующим двум основаниям:

а) по уровню преподавания: коррекционная (компенсирующего обучения), базового уровня, повышенного уровня;

б) по профилю: физико-математический, естественнонаучный, гуманитарный, эстетический, военно-спортивный, профессионального обучения.

Таким образом, основанием отбора, если речь идет об учебных заведениях общего образования, должна быть образовательная программа, а единицей отбора — тот или иной класс, в котором данная программа реализуется.

2. Пол — мальчики, девочки, смешанная группа.

3. Географическое размещение — центр города, спальный район, поселок, сельская местность,

4. Статус школы в системе образования — начальная, неполного среднего, полного среднего, начального профессионального образования.

В зависимости от целей тестирования количество стратификационных групп может быть уменьшено, например, если тестирование рассчитано только на городскую полную среднюю школу для детей обучающихся в гимназических и обычных классах, то по позиции “статус” стратификацию можно не учитывать, а по позиции “уровень преподавания” — уменьшить до двух.

Таким образом для расчета норм по параллели классов может потребоваться 2 (уровень преподавания) \times 2 (пол) \times 2 (район города) \times 300, получим 2400.

Будет трудно отобрать выборку, точно сбалансированную по всем переменным (поскольку на окраине города достаточно мало школ реализующих профильную программу

обучения на повышенном уровне), но тем не менее выборка, отражающая пропорции каждой из школ в общей популяции, даже если было по две школы на категорию, должна была бы дать адекватные нормы.

П.Клайн предлагает следующие правила формирования выборок для специальных групп.

1) найдите наиболее важные переменные, релевантные для данных групп, и используйте их как основание для стратифицирования выборки;

2) подберите настолько большую выборку, насколько возможно, как минимум из 300 испытуемых;

3) помните, что маленькая выборка — это лучше, чем вообще ничего. Если использовались небольшие выборки, акцентируйте внимание пользователей теста на том, чтобы они не использовали нормы, а если будут делать это, с предельной осторожностью.

Уже написано достаточно, чтобы стало ясно, что при установлении норм нет никаких теоретических проблем. Наоборот, обычной трудностью является отсутствие ресурсов: времени, денег, испытуемых и помощников для проведения тестирования и обработки тестов. Однако, если тест предназначен для практических целей: отбора детей для продолжения образования, профотбора и профориентации, то естественно, что его нормы должны удовлетворять высоким стандартам, описанным здесь. Существенно важно использование стратифицированных и больших выборок.