7.5. Задачи, решаемые при применении выборочного метода

При применении выборочного наблюдения возникают три основные задачи:

• определение объема выборки, необходимого для получения требуемой точности результатов с заданной вероятностью;

• определение возможного предела ошибки репрезентативности, гарантированного с заданной вероятностью, и сравнение его с величиной допустимой погрешности.

• определение вероятности того, что Ошибка выборки не превысит допустимой погрешности.

Все эти задачи решаются на основе теоремы Чебышева, согласно которой Р {[ х - ? | < ε } ≥ 1 - h, когда п - достаточно большое число; ε и h — сколь угодно малые положительные числа. Это соотношение, как было показано в п. 7.3, может быть выражено через формулу предельной ошибки выборки ?x = tsx или ?p = ts. Решение указанных задач зависит от того, какие величины в формуле предельной ошибки заданы, а какие нужно найти.

Объем выборки рассчитывается на стадии проектирования выборочного обследования. Так как



то

формула 1



где ? - допустимая погрешность,, которая задается исследователем исходя из требуемой точности результатов проектируемой выборки;

t - табличная величина, соответствующая заданной доверительной вероятности F(t), с которой будут гарантированы оценки генеральной совокупности по данным выборочного обследования;

?2 — генеральная дисперсия.

Последняя величина, как правило, неизвестна. Используются какие-либо ее оценки: результаты прошлых обследований той же совокупности, если ее структура и условия развития достаточно стабильны, или же зная примерную величину средней, находят дисперсию из соотношения



если известны xmax и хmin, то можно определить среднее квадратическое отклонение в соответствии с правилом «трех сигм»



так как в нормальном распределении в размахе вариации «укладывается» 6?(±3?). Если распределение заведомо асимметричное, то

.



Для относительной величины принимают максимальную величину дисперсии ?2max = 0,5•0,5 = 0,25.

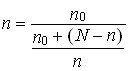
При расчете п не следует гнаться за большими значениями t и малыми значениями ?, так как это приведет к увеличению объема выборки, а следовательно, к увеличению затрат средств, труда и времени, вовсе не являющемуся необходимым.

Формула 1 не учитывает бесповторности отбора и дает максимальную величину выборки, которую можно скорректировать «на бесповторность». Так как



то на основе 1 формулы получаем выражение скорректированного объема выборки (п):

формула 2



где



При больших размерах генеральной совокупности скорректированный Объем выборки незначительно отличается от n0.

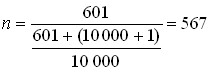
Например, для изучения структуры и стоимости покупок в универмаге из 10 000 покупателей следует отобрать определенное число человек, которое бы обеспечивало с вероятностью 0,95 определение средней стоимости покупок с точностью не менее 2 тыс. руб. Дисперсию примем по прошлому обследованию равной 625.

человек,



тогда скорректированная численность

человек (? 570 человек).



При проектировании районированной выборки рассчитанный объем выборки распределяют пропорционально численности районов (пропорциональный отбор):

формула 3



где пi — объем выборки для i-го района;

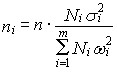
Ni - объем i-го района в генеральной совокупности;

п - общий объем выборки;

N - общий объем генеральной совокупности.

При различиях в однородности выделенных районов лучшие результаты дает распределение запланированного объема выборки между районами не только с учетом их объема, но и с учетом дисперсии признака (оптимальный отбор). В этом случае объем выборки в i-м районе определяется как

формула 3



где ?2i - дисперсия признака х в i-м районе.

При любом виде проектируемой выборки расчет объема выборки начинают по формуле повторного отбора (7.20). Если в результате расчета п доля отбора превысит 5%, проводят второй вариант расчета по формуле бесповторного отбора, либо по формуле 2, либо как



Если доля отбора меньше 5%, к формуле бесповторного отбора не переходят, так как это не скажется существенно на величине п.

Выборка должна быть такой, чтобы выборочные показатели по всем основным характеристикам были репрезентативны. Поэтому численность выборки рассчитывают многократно исходя из допустимых ошибок разных показателей, значения которых в генеральной совокупности известны.

Например, при выборочном учете детей школьного возраста требуется определить число семей, которые надо обследовать. При этом надо учесть: а) число детей в возрасте 6-7 лет, б) число детей в возрасте 6-15 лет; в) число детей в возрасте 16-17 лет;

г) среднедушевой доход (например, для решения вопроса о строительстве базы отдыха).

Так как репрезентируемые признаки могут иметь разную размерность, то допустимая погрешность для каждого их них задается в виде относительной величины (? : х?) (например, планируется, что в определении среднего размера семьи ошибка должна быть не больше 2%, в определении дохода - не больше 3% и т.д.). В этом случае вместо дисперсии в формуле (7.20) берется квадрат коэффициента вариации.

Вычислив значение п, на основе каждой из характеристик получаем разные объемы выборки: 1200; 300; 700; 100. Обследовать необходимо 1200 семей, т.е. из рассчитанных численностей берется максимальная. При резких различиях необходимых объемов выборки для разных вопросов программы проводится многофазный отбор. В рассмотренном примере среднедушевой доход достаточно учитывать в одной из каждых 12 семей, попавших в выборку.

Многофазный отбор, как правило, довольно сложно организовать, может быть нарушен принцип случайности отбора. Поэтому для обеспечения репрезентативности оказывается выгоднее затратить больше средств на учет большего числа единиц совокупности. Многофазный отбор целесообразно применять, если соотношение между рассчитанными объемами выборки по крайней мере 1:6.

Поскольку расчет необходимой численности выборки основан не на точных, а на предположительных данных о колеблемости в совокупности, следует соблюдать следующие рекомендации: абсолютную величину п округлять только вверх; долю отбора округлять только вниз, т.е. из предосторожности планировать несколько больший объем выборки, чем показывают расчеты.

Объем многоступенчатой выборки рекомендуется увеличить не менее чем на 10% от рассчитанной численности, поскольку, как было показано в предыдущем параграфе, многоступенчатость отбора увеличивает ошибку выборки.

После проведения выборки рассчитывают возможные ошибки . выборочных показателей (ошибки репрезентативности), которые используются для оценки результатов выборки и для получения характеристик генеральной совокупности.

Пример. На электроламповом заводе взято для проверки 100 ламп. Средняя продолжительность их горения оказалась 1420 ч со средним квадратическим отклонением 61,03 ч. Поскольку приемщика продукции интересует качество всей партии (50 тыс. электроламп), оценивают точность полученной средней. Средняя возможная ошибка вычисленной выборочной средней

ч.



С вероятностью 0,954 предел возможной ошибки

?х = 2•6,1 = ± 12,2 ч.

С вероятностью 0,954 можно утверждать, что средняя продолжительность горения 1 электролампы во всей партии будет находиться в пределах от 1408 до 1432 ч; 46 электроламп из 1000 могут иметь срок горения, выходящий за эти пределы.

Приемщика продукции интересуют отклонения от вычисленных пределов только в сторону сокращения продолжительности горения. Меньше чем 1408 ч могут гореть 23 лампы из 1000. На основании этого приемщик продукции решает вопрос о годности всей партии электроламп.

Решение вопроса может быть уточнено: определим, у какой доли ламп срок службы окажется меньше установленного лимита. Для потребителя продукции таким лимитом являются 1410 ч, продукция с меньшим сроком горения неприемлема.

При контрольной проверке 100 ламп 100 ламп горели менее 1410 ч, их удельный вес р = 0,1, или 10%. Средняя возможная ошибка этой доли

, или ± 3%



С вероятностью 0,954 предел ошибки доли Д^ = 2 • 0,03 = ± 0,06, или ±6%. Следовательно, во всей партии можно ожидать от 4 до 16% некачественных электроламп.

Чаще всего делают заключение об удовлетворительности выборки, сопоставляя получившиеся пределы ошибок выборочных показателей с величинами допустимых погрешностей. Может получиться, что предел ошибки, рассчитанный с заданной вероятностью, окажется выше допустимого размера погрешности. В этих случаях определяют вероятность того, что ошибка выборки не превзойдет допускаемую погрешность. Решение этой задачи и заключается в отыскании Fft) на основе формулы предела ошибки выборки:



где ? - допустимый размер погрешности оцениваемого показателя;

s2 - дисперсия показателя, рассчитанная по данным выборочного наблюдения;

n — объем проведенной выборки.

ставится условие, что минимальный срок горения электроламп 1410 ч, то, учитывая среднюю продолжительность горения по выборке (х= 1420 ч), допустимая погрешность равна 10 ч: 1410 - 1420 = - 10 ч.

Как было установлено выше, с вероятностью 0,954 предел возможной ошибки выборочной средней составил 12,2 ч, что превосходит допустимую погрешность. Является ли это основанием для браковки всей партии? Для ответа на этот вопрос определяют вероятность риска при приемке продукции:

, отсюда t= 1,64.



Соответствующая доверительная вероятность 0,899 (см. приложение, табл. 1). Вероятность того, что средний срок горения лампы меньше 1410 ч, равна:



Следовательно, из 100 ламп 5 могут гореть менее 1410 ч - риск появления некачественной продукции достаточно высок.

Аналогично можно определить вероятность того, что предел ошибки доли не превысит допускаемую погрешность доли.

Оценки надежности выборочных показателей, как показано на примере, позволяют принять обоснованные решения в отношении генеральной совокупности.