

Лекция 3.2. Статистическое наблюдение (Выборка)

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ – вид статистического наблюдения, организованный на основе выборки, при котором отбирается часть единиц изучаемой совокупности, по определённым правилам, из общей совокупности единиц. Совокупность единиц, из которых осуществляется отбор, называется ген. совокупностью. Отбор из генеральной совокупности проводится т.о., чтобы можно было получить достаточно точное представление об основных параметрах совокупности в целом.

Главное требование, которому должна отвечать выборка, – это требование репрезентативности или, иначе говоря, представительности. При рассмотрении любой схемы извлечения выборки должны быть учтены два фактора: использование вероятностной процедуры и степень объективности в действиях специалиста, формирующего выборку.

Наиболее важный принцип в применении выборочного наблюдения обеспечение случайности отбора, т.е. обеспечение равной возможности всем единицам, входящим в состав генеральной совокупности, быть избранными. Теоретические основы выборочного наблюдения обоснованы в теоремах Чебышева и Ляпунова.

При выборочном наблюдении в первую очередь решается 2 задачи: Определение ошибок выборки (средней и предельной); определение численности выборочной совокупности. Для каждого способа выборки свои алгоритмы расчета решения этих задач (см. таблицы).

В таблице 1 представлены обозначения, используемые в расчетах:

N	объём генеральной совокупности;
n	объём выборочной совокупности;
\bar{x}	генеральная средняя;
\tilde{x}	выборочная средняя;
P	генеральная доля;
w	выборочная доля;
σ^2	генеральная дисперсия
S^2	выборочная дисперсия
σ	среднее квадратическое отклонение в генеральной совокупности;
S	среднее квадратическое отклонение в выборочной совокупности;
Δ	предельная ошибка;
μ	величина средней квадратической ошибки.

Выборочное наблюдение различают по способу организации выборки:

1. Собственно – случайный отбор или простая случайная выборка – это отбор единиц из генеральной совокупности без всякого расчленения на части и группы. Основа выборки при этом может быть случайной, ранжированной, несистематизированной. При этом способе полностью соблюдается принцип случайности, для чего используют метод жеребьевки или табл. случайных чисел. Возможен повторный (возвратный) и бесповторный (безвозвратный) отбор единиц.

Табл. 2

Формулы ошибок простой случайной выборки

	Способ отбора единиц	
	повторный	бесповторный
Средняя ошибка		
для средней	$\mu_x = \sqrt{\frac{S^2}{n}}$	$\mu_x = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$
для доли	$\mu_p = \sqrt{\frac{w(1-w)}{n}}$	$\mu_p = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Предельная ошибка		
для средней	$\Delta_x = t \sqrt{\frac{S^2}{n}}$	$\Delta_x = t \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$
для доли	$\Delta_p = t \sqrt{\frac{w(1-w)}{n}}$	$\Delta_p = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$

Табл. 3

Формулы для определения численности случайной выборки

	Способ отбора единиц	
	повторный	бесповторный
для средней	$n = \frac{t^2 S^2}{\Delta_x^2}$	$n = \frac{t^2 N S^2}{\Delta_x^2 N + t^2 S^2}$
для доли	$n = \frac{t^2 w(1-w)}{\Delta_p^2}$	$n = \frac{t^2 N w(1-w)}{\Delta_p^2 N + t^2 w(1-w)}$

2. Механический отбор, состоящий в отборе единиц из упорядоченного ряда единиц ген. совокупности, с равным шагом. При организации механического отбора из ген. совокупности возникают две задачи: определения «шага отсчёта» (расстояния между отбираемыми единицами), выбора единицы, с которой надо начинать отсчёт.

«Шаг отсчета» определяется путем деления численности генеральной совокупности на численность выборочной совокупности.

Так как в генеральной совокупности единицы располагаются случайным образом, то механический отбор можно рассматривать как разновидность случайного бесповторного отбора. Поэтому для ошибки

механической выборки применяются формулы случайной бесповторной выборки.

3. Типический отбор, или отбор с аллокациями (расслоенная, стратифицированная или районированная выборка) – отбор единиц из типических групп, на которые разделена ген. совокупность. Расслоенный случайный отбор предусматривает предварительное деление совокупности, содержащей N единиц, на гомогенные группы и проведение простого случайного отбора в каждом слое. Расслоение можно рассматривать как процедуру извлечения выборок, в которой на простой случайный отбор наложены некоторые ограничения или условия. При выполнении определённых условий и наложении правильных ограничений можно получить значительный выигрыш в точности и, как правило, с малыми дополнительными затратами либо вовсе без них. При расслоенном случайном отборе управление обследованием значительно упрощено. Однако сама процедура предполагает знание объёмов слоев, общего числа единиц в выборке, а также определение долей отбора в каждом слое. Расслоение может дать выигрыш в точности при оценивании характеристик всей совокупности. Если каждый слой однороден в том смысле, что результаты измерений в нем мало изменяются от единицы к единице, то можно получить точную оценку среднего значения для любого слоя по небольшой выборке в этом слое. Затем эти оценки можно объединить в одну точную оценку для всей совокупности.

Типическая выборка равнозначна ряду случайных выборок из меньших совокупностей (типических групп или слоев), если в каждой из них единицы (элементы) отбираются в случайном порядке. Основное назначение типической выборки заключается в том, чтобы за счёт расслоения совокупности получить более высокую точность результатов выборки по сравнению с простым случайным отбором при том же объёме выборочной совокупности или такую же точность при меньшем объёме выборки. Чем эффективнее процедура расслоения, тем точнее результаты. Для расслоения, как правило, используется метод группировок. Эту выборку применяют в целях равномерного представления различных типов или различных регионов, и в этом случае её называют районированной.

Формулы для расчета ошибок типической выборки приведены в табл.4.

Табл. 4

Формулы ошибок типической выборки

	Способ отбора единиц	
	повторный	бесповторный
Для средней		
при пропорциональном размещении единиц	$\mu_x = \sqrt{\frac{S^2}{n}}$	$\mu_x = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$
при оптимальном размещении единиц	$\mu_p = \frac{1}{N} \sqrt{\sum \frac{S_i^2 N_i^2}{n_i}}$	$\mu_p = \frac{1}{N} \sqrt{\sum \frac{S_i^2 N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)}$

Для доли		
при пропорциональном размещении единиц	$\mu_p = \sqrt{\frac{w(1-w)}{n}}$	$\mu_p = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
при оптимальном размещении единиц	$\mu_p = \frac{1}{N} \sqrt{\sum \frac{w_i(1-w_i)N_i^2}{n_i}}$	$\mu_p = \sqrt{\sum \frac{w_i(1-w_i)N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)}$

Она имеет широкое применение при изучении неоднородных совокупностей, когда неприменимы ни случайная, ни механическая В вследствие очень высокой вариации системы изучаемых признаков, асимметричности распределения их значений. Может осуществляться равномерное, пропорциональное и оптимальное размещение аллокаций. Оптимальное размещение минимизирует среднюю ошибку выборочного наблюдения. Впервые задачу эффективной организации выборочного наблюдения на основе типической выборки решил в 1920 А.А. Чупров и независимо от него в 1934 Е. Нейман. В статистике такое размещение называется «неймановым».

4. Серийный отбор – выборка не единиц, а некоторых однородных групп (серий, гнезд) изучаемой совокупности. В пределах попавших в выборку серий единицы совокупности подвергаются сплошному наблюдению.

Стандартная ошибка пр равновеликих сериях определяется по формулам в табл.5.

Табл. 5

Формулы ошибок серийной выборки

	Способ отбора единиц	
	повторный	бесповторный
для средней	$\mu_x = \sqrt{\frac{\delta_x^2}{m}}$	$\mu_x = \sqrt{\frac{\delta_x^2}{m} \left(1 - \frac{m}{M}\right)}$
для доли	$\mu_p = \sqrt{\frac{\delta_w^2}{m}}$	$\mu_p = \sqrt{\frac{\delta_w^2}{m} \left(1 - \frac{m}{M}\right)}$

5. Комбинированный отбор – использование одновременно сплошного наблюдения и выборочного наблюдения (часто нескольких способов выборочного наблюдения, например, серийного и собственно-случайного; механического и гнездового).

Средняя ошибка комбинированной выборки определяется по формулам табл.6.

Табл. 6

Формулы средней ошибки комбинированной выборки

	Способ отбора единиц	
	повторный	бесповторный
средняя ошибка	$\mu_x = \sqrt{\frac{S^2}{n} + \frac{\delta^2}{m}}$	$\mu_x = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right) + \frac{\delta^2}{m} \left(1 - \frac{m}{M}\right)}$

6. Многоступенчатый отбор предполагает отбор из ген. совокупности сначала укрупнённых групп, затем меньших по объёму, и до тех пор, пока не будут отобраны те группы (серии) или отдельные единицы, которые будут подвергнуты наблюдению. На каждой ступени отбор может вестись случайным способом. Многофазный отбор осуществляется т.о., что одни сведения собираются от всех единиц отбора, а затем производится подвыборка, единицы которой обследуются по более широкой программе. В частном случае на первой фазе может быть произведено не выборочного наблюдения, а сплошное наблюдение. Особенность многофазной выборки заключается в том, что подвыборки извлекаются из выборок предыдущих фаз, и, следовательно, на всех фазах используются одни и те же единицы отбора. Многоступенчатая выборка имеет широкое применение в переписях населения. При использовании многоступенчатой выборки характерно сочетание выборочного наблюдения одной и той же совокупности по различным программам наблюдения, отличающимся друг от друга шириной обследования (т.е. количеством вопросов, по которым опрашивается население на каждой из фаз сформированных подвыборок). Данные, полученные на первой фазе, могут быть использованы в качестве дополнительной информации для повышения точности оценок статистических характеристик по изучаемым признакам на второй фазе. Следует также отметить, что информацией на первой фазе можно воспользоваться для расслоения единиц первой фазы и произвести формирование выборки второй фазы с единой или переменной долей отбора. Расчёт ошибки многофазной выборки производится для каждой фазы отдельно: взаимопроникающая выборки – форма выборочного наблюдения, при котором из генеральной совокупности одним и тем же способом формируются две (или более) выборочные совокупности. При этом происходит взаимное уточнение результатов обследования.

По степени охвата единиц изучаемой совокупности различают большие и малые выборки выборочное наблюдение относительно крупного размера, предназначенное для проведения многоцелевых обследований единого объекта наблюдения, в зарубежной статистике называют гл. выборкой. выборочное наблюдение, при котором наблюдением охвачено менее 25 – 30 единиц, принято называть малыми выборками. Оно обычно применяется в том случае, если невозможно или нецелесообразно использовать большую выборку (например, исследование продукции на прочность, связанное с её порчей). Выводы, сделанные на основе малой совокупности, справедливы лишь при нормальном распределении значений изучаемого признака в ген. совокупности.

Преимущество выборочного наблюдения в сравнении с другими видами сплошного наблюдения состоит в том, что по результатам этого наблюдения можно оценить искомые параметры генеральной совокупности, однако между ними существует некоторое расхождение, называемое

ошибкой выборки. Общая её величина складывается из ошибок регистрации и ошибок репрезентативности.

Ошибки репрезентативности обусловлены вариацией изучаемого признака в совокупности и недостаточно полным или адекватным представлением в выборке различных категорий единиц генеральной совокупности. Величина ошибки характеризует степень надёжности результатов выборочного наблюдения, и её знание необходимо при оценке параметров генеральной совокупности. Величина ошибки зависит от способа формирования выборки; от объёма выборочного наблюдения и от степени вариации признака в генеральной совокупности. Для каждого способа выборки величина ошибки определяется по соответствующим формулам выборочного наблюдения, содержащее «грубые ошибки» называют засоренной выборкой. Для учёта «грубых ошибок» используются методы робастного оценивания.

На заключительном этапе выборочного наблюдения осуществляется распространение полученных результатов на ген. совокупность. При этом проверяется, насколько адекватно представлена генеральная совокупность в выборочном наблюдении, и какова степень соответствия фактически полученной ошибки выборки её запланированному уровню. Общее значение изучаемого показателя для совокупности в целом определяется двумя способами: методом прямого счёта и методом коэффициентов. Прямой счёт состоит в том, что искомая характеристика находится по данным выборочного наблюдения и, с учётом доверительного интервала (предельной ошибки выборки), при заданном уровне значимости, распространяется на генеральную совокупность. Метод коэффициентов применяется обычно для внесения поправок в данные сплошного наблюдения, в котором обнаружены ошибки регистрации. Для этого проводится тщательное повторное наблюдение на основе выборочного наблюдения. Поправочные коэффициенты рассчитываются как частное от деления соответствующих данных контролирующего выборочного наблюдения и сплошного наблюдения. Умножением общего итога сплошного наблюдения на поправочный коэффициент получают данные, которые затем принимают за окончательные данные сплошного наблюдения. Они обычно называются итогами с поправкой на недоучёт.

Статистическое наблюдение, организованное на основе использования выборочного наблюдения, принято называть микропереписью, или микроценозом.

КОНТРОЛЬ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ – один из этапов статистического исследования, состоящий в сравнении полученных значений признаков с некоторыми эталонными значениями. При контроле статистической информации производится проверка правильности заполнения реквизитов документа и содержащихся в нем сведений, с целью выявления и устранения возможных ошибок. Контроль достоверности

информации осуществляется на всех этапах сбора и обработки статистических данных. Особое значение имеет входной контроль, т.е. проверка достоверности данных, получаемых после этапов наблюдения и подготовки данных к электронной обработке, поскольку ошибки, допущенные на этих этапах, искажают результаты последующих расчётов и, как правило, не могут быть впоследствии устранены. Контроль достоверности статистических данных предполагает, прежде всего, проверку полноты собранных данных. Она состоит в выяснении того, от всех ли источников поступили сведения, все ли показатели и ответы имеются в заполненных формулярах и формах отчётности. С проверкой полноты тесно связана работа по получению не поступивших во время данных. Целый ряд статистических работ, и, прежде всего, сводка, не могут выполняться по неполным данным. Контроль процесса сбора первичных данных направлен на обеспечение высокого уровня собираемости данных и заполнения формуляров. В процессе наблюдения всегда могут возникнуть ошибки, поэтому весь собранный материал должен быть подвергнут контролю в целях обнаружения ошибок и последующего их устранения или редактирования первичных данных. На качество материалов, полученных в ходе статистического наблюдения, влияет характер контроля формуляров и другой учетной документации, осуществляемого в процессе их приемки. Различают контроль формы и контроль содержания. Контроль формы включает обнаружения пропусков, отсутствие ответов. Содержательный контроль осуществляется в двух формах: и арифметический (счётный) контроль и логический контроль.

Арифметический контроль применяют к количественным ответам, полученным как сумма, разность, произведение или часть других показателей, он заключается в проверке правильности вычислений и основан на жёсткой связи между признаками, которая может быть проверена арифметическими действиями: сложением, вычитанием, умножением, делением. Связь такого рода часто отражается в заголовках граф отчётности и в подсказах к графам и строкам статистических форм. Арифметический контроль используется для проверки итоговых сумм. Если представленное число слагаемых не является полным, то сумма слагаемых должна быть меньше либо равна общему итогу, но не может превышать его. Там, где имеет место расчёт взаимосвязанных показателей, контроль возможен в результате установления некоторой их количественной связи, а также балансовым методом.

Логический контроль осуществляют путем проверки содержательной связи между значениями признаков по каждой единице наблюдения, то есть определяется соответствие ответа поставленному вопросу или соответствие между ответами на разные вопросы программы. В тех случаях, когда нарушается логическая связь признаков, необходимо установить, содержит ли формуляр действительно допущенную ошибку регистрации либо наблюдаемое сочетание признаков предопределено неординарностью

единицы наблюдения или ситуации. Логическую неточность в ответах устанавливают разными способами, а именно путем сопоставления фактических показателей с ожидаемыми значениями; с ранее полученными значениями; с показателями единиц совокупности, сходных по типу и условиям функционирования, с показателями, полученными из нескольких других источников и т.п. Логический контроль, в отличие от арифметического, непосредственно не устанавливает ошибку наблюдения, а лишь сигнализирует о её возможности и требует дополнительной связи с единицей наблюдения. Обычно для проверки поступающего материала наблюдения составляется схема контроля, в которую включаются все увязки между вопросами программы наблюдения: как арифметические, так и логические. При контроле результатов статистического наблюдения привлекают материалы из разных источников. Все ошибки, обнаруженные в ходе контроля, по возможности должны быть исправлены. Для этого порой проводят контрольные опросы, запросы почтой, по телефону, факсу и пр. Так, при проведении специальных статистических наблюдений важным средством проверки достоверности являются контрольные обходы и повторные опросы. Контрольные обходы применяются при переписях и других обследованиях, проводимых способом непосредственного наблюдения.

В практической статистике применяется автоматизированная *система редактирования* собираемых данных. Разработка такой системы ставит своими целями эффективное распределение ресурсов на редактирование данных для каждого проводимого наблюдения; ориентированность на применение универсальных программно-технологических средств; использование отлаженных на практике («лучших» в данной области статистики) алгоритмов автоматического редактирования; сохранение на относительно низком уровне затрат на обучение персонала, привлекаемого к данной работе и снижение количества повторных обращений к респондентам.

По используемым методам редактирование может быть классифицировано на два типа:

- Первый тип - статистическое редактирование данных, ошибочных с той или иной степенью вероятности; редактирование по результатам логического и арифметического контроля. Первый тип связан с широким классом стохастических методов замещения ошибочных значений и заполнения пропусков данных.
- Второй тип представляет собой процесс контроля соответствия значения признаков и показателей заранее определенным правилам или соотношениям. Значения данных, которые не проходят логический или арифметический контроль, являются ошибочными и должны быть исправлены или исключены из обследования. Коррекция основана на применении детерминированных методов, т.е. с помощью расчётных

формул, аргументами в которых выступают признанные верными значения признаков и показателей этой же единицы текущего периода наблюдения или любого предшествующего, либо данные других наблюдений или административных источников.

При редактировании обеспечивается непротиворечивость первичных данных наблюдения во временном разрезе. В случае, когда результаты наблюдения должны быть согласованными с итогами других наблюдений, этап редактирования должен обеспечивать сопоставимость этих данных на микроуровне. Обязательное требование к процедурам редактирования – накопление информации о внесённых с целью исправления изменениях в пообъектных данных, для обеспечения, в случае необходимости, возможности возврата к исходным значениям признаков. При проведении наблюдения целесообразно выполнение целевого редактирования, направленного на исправление ошибок, в наибольшей степени влияющих на сводные итоги. Для сплошных наблюдений в статистике предприятий основные усилия по контролю и редактированию направляются на исправление ошибочных данных крупных, более сложных по своей структуре предприятий, поскольку вклад в значение показателей более мелких единиц менее значимы.

При редактировании и заполнении пропусков данных выборочных наблюдений учитываются выборочные веса единиц, присвоенных им на стадии планирования выборки. Для обследований с большим числом разрезов разработки и, соответственно, со значительным количеством выходных таблиц, из общего списка таблиц выбирают только ключевые, по возможности, с различными видами взаимосвязей показателей. Они контролируются на непротиворечивость.

Для содержательного статистического контроля успешно используется графический метод представления данных. Он помогает эффективно идентифицировать нетипичные и экстремальные значения показателей.

Для получения достоверных статистических данных решающее значение имеет организация наблюдения и контроля, которая должна предусматривать комплекс эффективных мер для предупреждения и выявления ошибок.

Контроль сбора данных осуществляется путём проведения мониторинга для отслеживания доли полученных ответов от респондентов. Мониторинг способствует эффективному использованию ресурсного обеспечения на стадиях сбора и обработки данных. При его проведении, помимо текущих ответов респондентов, поступивших на определенный момент, автоматически пересчитываются основные показатели наблюдения в целях определения стабилизации значений показателей с поступлением новых отчётов. Рассчитывается соотношение текущих значений показателей с итоговыми значениями за предшествующий период наблюдения. Результаты собираемости используются для определения приоритетных

групп респондентов для будущих наблюдений с целью разработки мероприятий по сбору информации. Для получения качественной информации и во избежание массовых ошибок в некоторых, особо крупных обследованиях проводится разъяснительная работа в средствах массовой информации о цели и задачах наблюдения, напр. перепись нас. или с.-х. перепись. В ряде случаев считается целесообразным стимулировать респондентов путём предоставления им результатов обследований. Такая практика широко распространена в официальной статистике США.

В широком смысле контроль статистической информации (не только контроль достоверности информации, и контроль соответствия деятельности территориальных органов и подведомственных организаций Федеральному Закону о статистике) является одной из важнейших функций Федеральной службы гос. статистики (Росстат). Росстат и его региональные органы, для повышения достоверности данных первичного статистического учёта, периодически контролируют всю документацию. Наряду с проверкой документов первичного учёта статистические органы уделяют внимание контролю качества статистической отчётности путём изучения её полноты, правильности заполнения форм, документальной обоснованности данными первичного, оперативно-технического и бухгалтерского учёта.

Литература:

Теория статистики. – Учебник/ под редакцией проф. Г.Л. Громыко.- 2-е изд., перераб. и доп. - М.: ИНФРА–М, 2005 г.- 476 с., гл. 1, стр. 28-37; стр. 149-193.

Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник/Под ред чл.-корр. РАН И.И. Елисеевой, 4-е изд., перераб. и доп. – М.: Финансы и статистика , 1998, с. 7-12, с.29-38.

Ефимова М.Р., Петрова Е.В., Румянцева В.И. Общая теория статистики: Учебник. – 2-е изд, испр. И доп. – М.: ИНФРА – М, 2005, с. 27-50.

Образцова О.И. Курс лекций по теории статистического наблюдения

Организация государственной статистики в Российской Федерации/ Госкомстат России. – М., 2004, с. 248-307.

Поповская Е.В. Сплошное обследование малых предприятий: проблемы организации и основные этапы подготовки к его проведению. Вопросы статистики, 2000, №9, стр 3-5.

Статистика: учебник,/ Под ред. В.С. Мхитаряна. – М.: Экономист, 2005.с. 28-43.

Статистический словарь, издание второе, переработанное и дополненное. – М.; Финансы и статистика, 1989, с. 280 – 281, 309, 455.

Суринов А.Е. Официальная статистика в России: проблемы реформирования. – ИИЦ « Статистика России», 2002

Чернышева Т. Выборочное наблюдение в социально-экономических исследованиях и основные этапы их подготовки. Вопросы статистики. 1996, №11, стр. 12-18.