

Тема: ВЫБОРОЧНАЯ СОВОКУПНОСТЬ И СТАТИСТИЧЕСКИЕ ОШИБКИ

Как вам уже известно, объектом исследования в статистике является статистическая совокупность, т.е. группа, состоящая из большого числа относительно похожих единиц (жителей какой-либо территории, больных каким-либо заболеванием, случаев впервые возникшего заболевания, случаев выхода на инвалидность, случаев неудовлетворенности медицинской помощью и пр.). Сначала исследователь изучает (описывает, характеризует) каждую единицу совокупности по тем учитываемым признакам, которые, как он полагает, позволят ему наилучшим образом осветить изучаемую проблему. Затем исследователь, воспользовавшись методами статистики, обобщает и анализирует полученные сведения в целом по совокупности.

Генеральная и выборочная совокупности

Изучаемая статистическая совокупность может рассматриваться как *генеральная* или как *выборочная*. Это обстоятельство определяет подходы к анализу полученных данных. Дадим определения этим понятиям.

Генеральная совокупность – это совокупность всех без исключения единиц изучаемого объекта, всех единиц, которые соответствуют цели исследования.

Выборочная совокупность (выборка) – это часть единиц генеральной совокупности, отобранная специальным методом и предназначенная для характеристики всей генеральной совокупности.

Пример.

Цель исследования: изучить здоровье современных студентов ИвГМА.

Тогда:

- *Генеральной совокупностью* будут являться все до единого студенты ИвГМА, обучающиеся в данном году. Объем совокупности составит, предположим, две тысячи человек ($N=2000$).
- *Выборочной совокупностью* могут являться, предположим, 180 студентов, обучающихся ИвГМА в данном году ($n=180$).

Обратите внимание, что одна и та же статистическая совокупность может рассматриваться или как *генеральная*, или как *выборочная*. Это зависит от цели исследования.

Пример.

Изучаемая совокупность – все студенты ИвГМА, обучающиеся в данном году. Эта совокупность будет являться:

- *Генеральной*, если цель исследования – изучить здоровье студентов ИвГМА, обучающихся в данном году.
- *Выборочной*, если цель исследования – изучить здоровье современных студентов медицинских вузов России.

Выборочная совокупность изучается тогда, когда изучить генеральную совокупность невозможно или нерационально, т.е. неразумно (от лат. ratio – разум). Это обычно бывает в следующих случаях:

1) *Генеральная совокупность очень велика, поэтому изучить ее всю целиком нерационально (эта работа потребует очень много времени и средств).*

Пример:

Изучается здоровье очень большой группы населения, предположим, здоровье женщин репродуктивного возраста (от 15 до 49 лет), проживающих в Ивановской области. Численность женщин этого возраста в Ивановской области (т.е. объем генеральной совокупности) в настоящее время составляет около 260 тысяч. Чтобы провести медицинский осмотр и(или) проанкетировать такое огромное число женщин потребуются

десяток лет, сотни специалистов-исследователей и миллионы рублей. Поэтому формируют выборку, например, из нескольких сотен женщин (предположим, $n=600$), изучают их здоровье, а затем распространяют, «переносят» сделанные выводы на всех женщин репродуктивного возраста, проживающих сейчас в Ивановской области (т.е. всю генеральную совокупность).

2) *Генеральная совокупность не имеет границ (прежде всего, временных).*

Пример:

Изучается эффективность нового лекарственного средства, предположительно обладающего гипотензивным действием. Чтобы доказать, что это средство надежно и стабильно снижает артериальное давление, необходимо его опробовать на **всех (!!!)** больных артериальной гипертензией – живущих сейчас во всех странах, а также всех, кто будет жить на Земле в будущем. Именно такая группа будет являться в данном случае генеральной совокупностью. Однако выполнить такое исследование совершенно невозможно. Поэтому формируют выборку, состоящую из нескольких сотен современных больных артериальной гипертензией (предположим, $n=400$), апробируют лекарственное средство на этих пациентах, а затем сделанные выводы относительно его эффективности и безопасности распространяют на всю генеральную совокупность больных артериальной гипертензией.

3) *Изучение всей генеральной совокупности приведет к ее уничтожению.*

Пример:

Необходимо оценить качество поступившей в аптечную сеть с завода-изготовителя партии лекарственного препарата. Вся партия – это генеральная совокупность (предположим, $N=2000$). Препарат выпускается в форме раствора в герметично закупоренных флаконах. Чтобы оценить концентрацию различных веществ в растворе, необходимо выполнить ряд химических опытов. Для этого требуется вскрыть герметичный флакон и использовать больше половины его содержимого. Понятно, что после этого исследования применять данный флакон с лекарственным препаратом уже нельзя. Если будут вскрыты и проверены все флаконы этой партии (т.е. изучена вся генеральная совокупность), то будут получены совершенно точные сведения о качестве изучаемой партии лекарства, но при этом вся она будет полностью уничтожена. Поэтому формируют выборку из нескольких флаконов, вскрывают и проверяют их, а сделанные по выборке выводы о качестве лекарственного препарата распространяют на всю партию лекарства (т.е. всю генеральную совокупность).

Возникает вопрос: а можно ли доверять выборочным данным (т.е. данным, которые получены при изучении выборки)? Можно ли утверждать, что, если лекарственное средство эффективно для 75 из 100 испытуемых, то и вся генеральная совокупность больных этим заболеванием будет характеризоваться такими же свойствами? То есть у троих из каждых четверых пациентов это средство будет давать ожидаемый эффект? Да, это утверждать можно. Правда, не со 100%-ной уверенностью. Как точно измерять уверенность при статистических утверждениях, вы узнаете на учебных занятиях в следующем семестре. Пока же разберемся, почему можно доверять результатам выборочного исследования.

Возможность изучать генеральную совокупность по выборке опирается на **закон больших чисел**. Этот закон является одним из основных в теории вероятностей. Простейшую схему его проявления впервые описал швейцарский математик Якоб Бернулли (1655-1705). Если применить этот математический закон к статистическому исследованию, то он будет выглядеть так:

1. При увеличении числа наблюдений в выборке результаты выборочного исследования все больше и больше становятся похожими на истинные характеристики генеральной совокупности.

2. При достижении определенного числа наблюдений в выборке результаты выборочного исследования станут максимально близкими к истинным характеристикам генеральной совокупности.

Другими словами, чем больше объем выборки (чем больше в ней единиц наблюдения), тем точнее исследование, тем больше можно доверять выводам, которые сделал исследователь при изучении этой выборки. Однако, начиная с некоторого объема наблюдения, точность исследования почти перестает изменяться. Увеличение числа единиц наблюдения не дает существенного повышения точности – исследователь просто будет делать лишнюю работу. Он мог бы получить те же самые сведения, изучив меньшее число единиц наблюдения.

Пример.

Изучается распространенность табакокурения среди студентов России. Генеральная совокупность очень велика, поэтому изучать ее всю целиком не стоит. Проводится выборочное исследование. Были опрошены 10 случайно выбранных студентов разных вузов. Они ответили, что не курят (предположим, ответили искренне). Можно ли на основании этих данных делать вывод, что курящих студентов в российских вузах нет? Конечно же, нельзя. Выборка слишком мала, она не способна охарактеризовать всю генеральную совокупность. В выборку случайно попали только некурящие студенты. Но если увеличивать объем выборки, то повышается вероятность того, что в ней окажутся и курящие студенты тоже. Причем, чем больше будет становиться объем выборки, тем ближе к истинному будет в ней соотношение курящих и некурящих. При достижении некоторого объема выборки это соотношение будет крайне мало отличаться от истинного.

Итак, малая выборка не может точно охарактеризовать генеральную совокупность. А всегда ли способна это сделать большая выборка? Нет, не всегда. Здесь нам нужно познакомиться с понятием репрезентативности.

Репрезентативность и ее виды

Выборочная совокупность должна быть *репрезентативной* (от англ. representative – представительный), т.е. должна правильно представлять генеральную совокупность, достаточно точно отражать ее свойства.

Выборка должна быть репрезентативной и в плане количества, и в плане качества.

- **Количественная репрезентативность** обеспечивается *достаточным количеством единиц наблюдения* в выборке, таким количеством, которое гарантирует получение надежных, доказательных, статистически достоверных результатов. Закон больших чисел говорит именно об этом.
- **Качественная репрезентативность** обеспечивается *одинаковой структурой* выборочной и генеральной совокупностей (по полу, возрасту, расовой принадлежности, национальности, уровню образования и пр. факторным признакам, которые могут оказать влияние на полученные результаты).

Чтобы понятнее была суть качественной репрезентативности, расскажем следующую историю. Одним из наиболее ярких примеров фиаско в истории применения статистики является заключение, сделанное при обработке результатов опроса, проведенного в США в 1936 году журналом «Literary Digest». Редакцией этого журнала было разослано 10 миллионов (!!!) опросных листов, в которых респондентам было предложено ответить на вопрос, за кого они будут голосовать на предстоящих президентских выборах – за республиканца А.Лэндона или демократа Ф.Рузвельта. Было возвращено более 2 миллионов (!!!) заполненных опросных листов. Это была огромная выборка из генеральной совокупности американских избирателей. Количественная репрезентативность такой выборки не может вызывать никаких сомнений. На основании полученных сведений с большой долей уверенности был сделан вывод, что выборы

несомненно выиграет республиканец Альфред М. Лэндон. Настал день выборов. Теперь свое мнение высказали все избиратели (т.е. была получена характеристика всей генеральной совокупности). С большим отрывом, набрав более 60% голосов избирателей, победил Франклин Делано Рузвельт.

Причиной ложного статистического прогноза стало то, что выборка потенциальных избирателей, хотя и была репрезентативной в количественном отношении, не обладала *качественной репрезентативностью*. Работники «Literary Digest» плохо продумали технологию отбора единиц наблюдения для своего выборочного исследования. Опросные листы были разосланы людям:

- по адресам, полученным из телефонных книг,
 - по адресам из полицейской базы данных о владельцах автомобилей.
- Иметь телефон или автомобиль в 30-е годы XX века могли себе позволить только достаточно обеспеченные американцы. В выборке практически не было представителей огромной массы малообеспеченных слоев населения, которые как раз и собирались голосовать за Ф.Рузвельта.

Отсутствие репрезентативности выборки приводит к неверным результатам расчетов и некорректным выводам, т.е. к появлению статистических ошибок. Для статистики ошибки являются очень серьезной проблемой, и многие вдающиеся люди, не скрывая иронии, это отмечали. Например, великий американский писатель, журналист и общественный деятель Марк Твен (1835-1910) сказал следующее: «*Существует три вида лжи – ложь, наглая ложь и статистика.*» Британскому премьер-министру Бенджамину Дизраэли (1804-1881) приписывают похожее высказывание: «*Имеются четыре вида лжи – ложь, наглая ложь, предвыборные обещания и статистика.*» Разберемся, какие же бывают статистические ошибки, и как можно их избежать, или хотя бы уменьшить.

Статистические ошибки

Существует два принципиально разных вида статистических ошибок:

- а) случайная ошибка,
- б) систематическая ошибка (смещение).

а) Случайная ошибка

Случайная ошибка – это отклонение результата выборочного исследования от истинных характеристик генеральной совокупности вследствие разнообразия (вариабельности) изучаемого признака.

Если бы все единицы генеральной совокупности имели одно и то же значение изучаемого признака, то никакой случайной ошибки в статистических исследованиях не было бы.

- Если бы изучаемое лекарственное средство было одинаково эффективным для всех больных данным заболеванием, то какую бы выборку этих больных мы ни рассматривали, результаты оценки эффективности лекарства были бы всегда одинаковы – 100%.
- Если бы все студенты российских вузов были курильщиками, то какую бы выборку студентов мы ни изучали, в нее попадали бы только курящие студенты.
- Если бы все 10 летние мальчики имели рост 140 см, то в любой выборке 10-летних мальчиков средний рост был бы 140 см.
- Если бы все дети рождались с массой тела 3200 г., то в любой выборке новорожденных средняя масса тела была бы 3200 г.

Перечень примеров может быть бесконечен.

Но мир устроен иначе. Английский статистик Фрэнк Йейтс в 1937 году писал: «Большинству биологических объектов свойственна вариабельность и прелесть простоты и воспроизводимости физических или химических экспериментов утрачивается. А значит – на передний план начинают выдвигаться статистические проблемы.»

Исследователь знает, что у единиц генеральной совокупности изучаемые им учитываемые признаки имеют разные значения. А вот какие именно единицы оказались в его выборке, часто или редко встречающиеся в генеральной совокупности, исследователю неизвестно.

- Эффективность лекарств для разных людей различна. А какие люди, чувствительные или малочувствительные к этому лекарству, случайно попали в выборку, исследователь не знает.
- Исследователь, изучающий распространенность табакокурения среди студентов, не может знать, кто преимущественно попал в его выборку – студенты из числа курящих или же из числа некурящих.
- Исследователь, изучающий физическое развитие 10-летних мальчиков, понимает, что в его выборку могут случайно попасть преимущественно дети из числа невысоких или же преимущественно высокие дети.
- Исследователь, изучающий акушерские проблемы конкретной соматической патологии и замеряющий массу тела детей, которые родились у женщин с данной патологией, предполагает, что в его выборке могут случайно оказаться как преимущественно дети с малой массой, так и преимущественно с высокой массой.

Исследователь стремится изучить, охарактеризовать генеральную совокупность. Но сделать это невозможно. Тогда исследователь изучает выборку. Затем, пользуясь статистическими методами, он описывает выборочные данные, т.е. получает выборочные оценки (частоты, доли, средние и пр.). Он предполагает, что истинные характеристики генеральной совокупности несколько отличаются от полученных им выборочных оценок. Насколько отличаются? Это как раз и показывает случайная ошибка.

Кратко охарактеризуем случайную ошибку:

- ✓ Случайная ошибка – это неизбежная ошибка любого выборочного исследования, связанная со случайностью отбора единиц из генеральной совокупности. Часть никогда не сможет абсолютно точно охарактеризовать целое, какой бы большой она ни была. При сплошном исследовании (т.е. изучении всей генеральной совокупности) случайной ошибки не будет, т.к. исследователь узнает истинные характеристики изучаемого объекта.
- ✓ Случайная ошибка связана с количественной репрезентативностью. Величину случайной ошибки можно уменьшить, увеличив объем выборки (вспомним закон больших чисел).
- ✓ Результаты выборочного исследования могут оказаться как завышенными, так и заниженными по сравнению с истинными характеристиками генеральной совокупности.
- ✓ Величину случайной ошибки можно рассчитать по статистическим формулам, опираясь на оценки варибельности признака и объем наблюдения (с этими формулами мы будем знакомиться в следующем семестре).
- ✓ Выборочные оценки (т.е. результаты, полученные по выборке) часто представляют с указанием рассчитанной величины случайной ошибки. Поскольку выборочные оценки могут быть как завышены, так и занижены, случайную ошибку откладывают как в меньшую, так и в большую сторону, т.е. ставят знак «±» («плюс-минус»). Например:
 - «При использовании лекарственного средства «А» отметили улучшение самочувствия $78,3 \pm 2,5$ из 100 пациентов»
 - «Распространенность табакокурения среди студентов российских вузов составляет $43,4 \pm 5,8$ на 100 опрошенных»
 - «Средний рост 10-летних мальчиков в городе М. составляет $138,1 \pm 1,7$ см»
 - «Средняя масса тела детей, родившихся у женщин с данной соматической патологией, составляет 4145 ± 68 граммов»

б) Систематическая ошибка (смещение)

Систематическая ошибка (смещение) принципиально отличается от случайной ошибки.

Систематическая ошибка (смещение) – это постоянное, неслучайное, одностороннее отклонение результатов исследований от истинных характеристик генеральной совокупности.

Сначала рассмотрим, чем систематическая ошибка (смещение) отличается от случайной ошибки. Итак:

- ✓ Систематическая ошибка (смещение) может появиться как при выборочном исследовании, так и при сплошном. При хорошо спланированном и выполненном исследовании этой ошибки может и не быть.
- ✓ Систематическая ошибка (смещение) не связана с количественной репрезентативностью, ее величина почти не зависит от числа единиц наблюдения. В хорошо спланированном исследовании она может отсутствовать в малой выборке. В плохо спланированном исследовании она может быть огромной при изучении всей генеральной совокупности.
- ✓ Систематическая ошибка (смещение) часто бывает связана с качественной репрезентативностью выборки.
- ✓ Систематическая ошибка (смещение) искажает истинные характеристики генеральной совокупности только в одну сторону (либо в большую, либо в меньшую). Например, выборки, извлеченные из одной генеральной совокупности, снова и снова будут давать только завышенные оценки (либо только заниженные).
- ✓ Величину систематической ошибки (смещения) в большинстве случаев определить (рассчитать) нельзя. Можно только предположительно судить о направлении смещения. Часто исследователю вообще остается неизвестным, являются ли полученные оценки смещенными или же нет.
- ✓ Систематическая ошибка (смещение) обычно появляется в плохо спланированном и неаккуратно выполненном исследовании. Она может возникнуть на любом этапе работы – и при сборе материала, и при его разработке, и при его анализе, и при формулировании выводов. Исследователь должен стремиться так организовать и выполнить свою работу, чтобы вероятность систематических ошибок (смещений) была минимальной.

Теперь разберемся в причинах появления систематических ошибок (смещений). Рассмотрим следующий пример.

Проведено экспериментальное исследование с целью изучить эффективность нового лекарственного средства «А». Испытуемых разделили на две группы:

- первая группа в течение месяца принимала новое лекарственное средство «А»,
- вторая группа также в течение месяца принимала лекарственное средство «В» (в настоящее время именно оно обычно используется для лечения изучаемой патологии).

Выявлено, что лекарство «А» более эффективно, чем лекарство «В». Это может быть истиной. Но может оказаться и ложным выводом. На самом деле эти лекарства по своей эффективности могут быть примерно одинаковы, либо новое лекарство «А» даже менее эффективно, чем лекарство «В».

Посмотрим, какие систематические ошибки могли «обмануть» исследователя, привести его к ложному заключению:

- 1) Исследователь неправильно сформировал сравниваемые группы пациентов. Пациенты, получавшие новое лекарство «А», имели более благоприятную форму заболевания (менее тяжелую). Нарушенные процессы в их организме под влиянием лечения восстановились быстрее и надежнее.

- 2) Лекарство «А» приятнее на вкус, поэтому пациенты, получавшие его, строже соблюдали схему лечения, не пропускали время приема.
- 3) Лекарство «А» – это новое лекарственное средство, а лекарство «В» – старое. Многие люди больше доверяют всему новому. Поэтому и сами пациенты, и врач-исследователь, наблюдавший за изменением их здоровья, были склонны думать, что новое лекарственное средство непременно действует лучше.
- 4) Фирма-производитель лекарства «А» очень заинтересована в том, чтобы была научно доказана высокая эффективность разработанного нового препарата. Это позволило бы получать большую прибыль от его продажи. Поэтому фирма высоко оплачивала труд врачей-исследователей, дарила им ценные подарки, приглашала на праздничные мероприятия, помогала публиковать научные статьи и пр. В обмен на полученное материальное вознаграждение врачи-исследователи на этапе сбора или этапе обработки данных умышленно «откорректировали» собранные сведения, чтобы доказать эффективность и безопасность лекарства «А».
- 5) В процессе сбора первичного статистического материала или в процессе его обработки были допущены ошибки, например, были перепутаны учетные документы. В результате статистические карты пациентов, получавших лекарство «В», были обработаны как документация пациентов, получавших лекарство «А».
- 6) Исследователь допустил ошибки в ходе анализа собранных сведений – неверно сгруппировал данные, выбрал для расчета неверную формулу, неправильно интерпретировал результаты расчетов.

Перечень примеров можно продолжать и продолжать.

Рассмотрим самые частые *причины появления систематических ошибок (смещений)*.

1. Неправильный отбор единиц наблюдения

В этом случае выборка по своей структуре не похожа на генеральную совокупность (т.е. отсутствует качественная репрезентативность выборки) (как в примере с социологическим исследованием «Literary Digest»).

Дизайн исследования может предусматривать изучение не одной выборки, а нескольких выборок в сравнении. Например, формируется основная группа, принимающая лекарственный препарат, и контрольная группа, принимающая плацебо. При неправильном отборе сравниваемые группы могут существенно различаться по некоторым важным факторным признакам, влияющим на результат (пол, возраст, степень тяжести заболевания, сопутствующие заболевания и пр.).

Такая ошибка возникает при формировании исследуемых совокупностей. О ней следует помнить на подготовительном этапе в ходе планирования исследования.

2. Особенности измерения признака

Разные методы измерения одного и того же признака могут дать разные результаты. Например, при измерении диастолического артериального давления с помощью сфигмоманометра результаты будут примерно на 10 мм.рт.ст. выше, чем при измерении с помощью внутриаrтериальной канюли. При измерении систолического артериального давления результаты будут выше, если испытуемый во время обследования говорит.

О возможности такого смещения следует помнить на втором этапе исследования при сборе статистического материала.

3. Воздействие вмешивающихся факторов

Два или более явления могут быть взаимосвязаны (как говорят «ходят вместе»). Размер одного явления изменяется, и размер другого явления тоже изменяется. Но это совершенно не значит, что одно из явлений – это причина другого. Возможно, есть какой-

то третий фактор, который оказывает влияние на оба изучаемых явления, вызывая их изменения. Об этой ошибке следует помнить в ходе анализа материала.

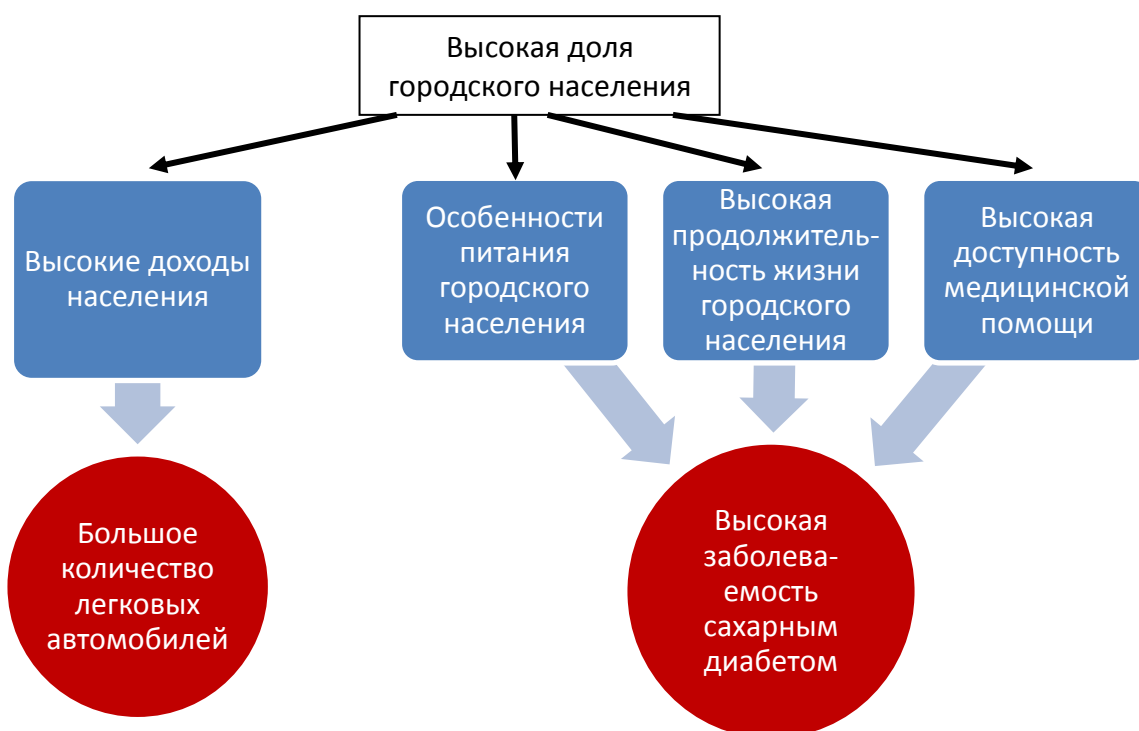
Пример:

Распространенность сахарного диабета в РФ существенно различается в разных регионах. При анализе данных была выявлена достоверная сильная прямая связь между уровнем заболеваемости населения сахарным диабетом и количеством легковых автомобилей в расчете на душу населения в этом регионе. Идея о том, что выхлопные газы именно легковых автомобилей вызывают развитие сахарного диабета, конечно, интересная, но вряд ли соответствует истине.

Исследователи предположили, что выявленная «связь» количества легковых автомобилей и заболеваемости сахарным диабетом, скорее всего, является лишь «наводкой» от других факторов, в частности, от доли городского населения в регионе (рисунок 1).

Рисунок 1

**Пример систематической ошибки,
вызванной воздействием вмешивающегося фактора**



Если доля городского населения в регионе выше, тогда:

С одной стороны:

- Средние доходы населения выше, а значит – количество легковых автомобилей в расчете на душу населения тоже будет выше.

С другой стороны:

- Особенности питания городского населения провоцируют заболеваемость сахарным диабетом.
- Доступность медицинской помощи городскому населению выше, следовательно, больше будет выявляться случаев сахарного диабета. (У городского жителя диабет обнаружат и отметят в официальной статистической документации, а селянин всю жизнь может прожить с нераспознанным диабетом)
- Средняя продолжительность жизни городского населения больше, следовательно, распространенность сахарного диабета будет выше. (В городе человек с сахарным диабетом проживет, предположим, до 70 лет, и его болезнь каждый год будут отмечать

в учетных и отчетных статистических документах, а селянин проживет меньше, случай его заболевания исчезнет из популяции и учитываться уже не будет).

4. Умышленное искажение данных испытуемым или исследователем

Такое смещение может появиться при изучении распространенности поведения, осуждаемого обществом (табакокурение, злоупотребление алкоголем, употребление наркотиков, число искусственных аборт и пр.). Полученные данные, скорее всего, окажутся заниженными, т.к. многие испытуемые не скажут правду, не признаются в своем неблагоприятном поведении.

А вот при изучении документации, которая отражает качество работы медперсонала, полученные данные, скорее всего, окажутся завышенными, особенно если учитываемые показатели качества влияют на оплату труда. Ведь каждому работнику хочется выглядеть в глазах начальства лучше и зарабатывать больше.

О таком смещении следует помнить на этапе планирования исследования и этапе сбора материала. К сожалению, избавиться от него в большинстве случаев невозможно. Можно лишь предположить направление смещения (в большую или меньшую сторону по отношению к истинным характеристикам генеральной совокупности), но узнать точный размер смещения нельзя.

Возможность появления систематической ошибки еще не означает, что она действительно присутствует в конкретном исследовании. Может оказаться, что исследователь настолько хорошо все продумал, учел и предусмотрел, что результат исследования будет *несмещенным*, правильным.

В плохо продуманном исследовании могут присутствовать систематические ошибки различных видов (обусловленные и неправильным отбором, и некачественными измерениями, и неверным анализом).

Систематическая и случайная ошибки не исключают друг друга и могут обе присутствовать в результатах исследования.

Выборочные исследования выполняются очень часто. Рассмотрим методы, позволяющие сформировать выборку, отобрать из генеральной совокупности единицы для изучения.

Методы формирования выборки

Методы формирования выборки бывают:

1. Не требующие предварительного деления генеральной совокупности на части (случайный, механический, серийный).
2. Требующие предварительного деления генеральной совокупности на части (типологический).

1) Случайный отбор

В основе случайного отбора – перемешивание. Единицы наблюдения отбираются наугад. Все единицы генеральной совокупности имеют равные возможности попасть в выборку.

2) Механический отбор

Единицы наблюдения отбираются по какому-либо условному признаку, ничего не значащему для цели исследования и никак не влияющему на результат. Например:

- каждый пятый случай лечения (если нужно отобрать 20% генеральной совокупности);
- пациенты с фамилией на букву «М» (вряд ли фамилия может существенно повлиять на здоровье);
- нечетные даты поступления в стационар и пр.

Однако, при использовании этого метода следует быть осторожным. Он не всегда позволяет сформировать репрезентативную в качественном отношении выборку. Например, в соответствии с какими-либо договоренностями по нечетным датам в данный

стационар госпитализируют пациентов из района К., особенности экологии в котором способствуют более тяжелому течению изучаемого заболевания.

3) Серийный (гнездный) отбор

При этом единицы наблюдения отбираются не по одной, а группами. Эти группы по другому называются «серии», «гнезда». Например:

- На территории района тридцать примерно одинаковых населенных пунктов. Для изучения отберем жителей населенных пунктов «А», «В» и «С».
- Изучается здоровье детей дошкольного возраста. Для изучения можно отобрать детей, посещающих детские сады №2, №12 и №22.

Внутри каждой группы (серии, гнезда) изучаются **все единицы**, соответствующие критериям включения в исследование, без какого-либо дополнительного отбора.

Отбор самих серий осуществляется с помощью случайного или механического метода.

4) Типологический (типический) отбор

Генеральная совокупность предварительно разбивается на типы (по полу, возрасту, уровню образования и пр.) с последующим отбором единиц наблюдения из каждой типологической группы пропорционально численности этих групп.

Например, известно, что в генеральной совокупности численностью 10 000 человек 20% составляют мужчины и 80% - женщины. Предполагается, что фактор пола имеет существенное влияние на изучаемое явление (например, у мужчин изучаемая патология развивается чаще). При типологическом отборе структура выборки по полу должна быть такой же, как и структура генеральной совокупности, т.е. в ней так же 20% должны составлять мужчины и 80% – женщины. Для изучения необходимо сформировать выборку объемом 500 человек. При типологическом отборе в ней должно быть 100 мужчин (20% от 500) и 400 женщин (80% от 500).

5) Многоступенчатый отбор

На каждой ступени исследования формируется выборка, являющаяся частью совокупности предыдущей ступени. Набор учитываемых признаков для каждой ступени свой. Например:

- Изучаются все работники предприятия ($n_1=6000$)
- Изучаются женщины, работающие на предприятии ($n_2=3500$)
- Изучаются женщины, работающие на предприятии и имеющие детей ($n_3=2700$)
- Изучаются женщины, работающие на предприятии и имеющие детей с врожденными заболеваниями ($n_4=120$) (рисунок 2)

Рисунок 2

Пример многоступенчатого отбора



б) Парно-сопряженный отбор (метод «копи-пара»)

Этот вид отбора применяется, если необходимо сформировать несколько групп испытуемых. Единицы наблюдения подбираются парами: одна в основной группе, одна в контрольной группе. Эти парные единицы являются копиями друг для друга, т.е. имеют одинаковые значения самых важных для данного исследования факторных признаков (пол, возраст, расовая принадлежность, уровень образования, форма заболевания и пр.). Сравнимые группы можно уравновесить максимум по 5-7 признакам (рисунок 3).

Рисунок 3

Пример парно-сопряженного отбора



В случае парно-сопряженного отбора структура сравниваемых групп будет одинаковой. Если по итогам исследования будет выявлено различие групп, то можно с большей уверенностью утверждать, что различия эти не случайны, что они обусловлены воздействием изучаемого медицинского вмешательства.