

Генеральная совокупность. Выборки. Типы выборок.

Доц. каф.ПИ Платонова И.Ю.

Генеральная совокупность

Совокупность всех возможных элементов объекта исследования, которая подлежит изучению в рамках конкретного исследования, и на которую будут распространяться выводы проведенного исследования, называется **генеральной совокупностью**.

Суммарная численность объектов наблюдения (число учащихся, число образовательных учреждений, число выпускников и т.д.), обладающих определенным набором признаков (пол, возраст, численность, успеваемость и т.д.), ограниченная в пространстве и времени.

Выборочная совокупность

Процесс формирования выборочной совокупности называется выборкой. Важнейшим вопросом теории выборки является обеспечение репрезентативности выборочной совокупности.

Репрезентативность - свойство выборочной совокупности воспроизводить в себе основные параметры и значимые элементы структуры генеральной совокупности.

Репрезентативность выборки - показатель объективности полученной информации и правомочности ее распространения по части на целое. Выборка считается репрезентативной, если отклонение выборочной совокупности от генеральной не превышает 5%. Если отклонение превышает 5%, это считается ошибкой выборки.

Выборка (Выборочная совокупность)

- Часть объектов из генеральной совокупности, отобранных для изучения, с тем чтобы сделать заключение обо всей генеральной совокупности.
- Для того чтобы заключение, полученное путем изучения выборки, можно было распространить на всю генеральную совокупность, выборка должна обладать свойством репрезентативности.

Выборка (Выборочная совокупность)

- Когда для каждого объекта в выборке измерено значение одной переменной, популяция и выборка называются **одномерными**. Если же для каждого объекта регистрируются значения двух или нескольких переменных, такие данные называются **многомерными**.
- Одной из основных задач статистического анализа является получение по имеющейся выборке достоверных сведений о интересующих исследователя характеристиках генеральной совокупности. Поэтому важным требованием к выборке является ее **репрезентативность**, то есть правильная представимость в ней пропорций генеральной совокупности. Достижению репрезентативности может способствовать такая организация эксперимента, при которой элементы выборки извлекаются из генеральной совокупности случайным образом.

-

Репрезентативность выборки

- Свойство выборки корректно отражать генеральную совокупность.
- Одна и та же выборка может быть репрезентативной и нерепрезентативной для разных генеральных совокупностей.

Пример:

- Важно понимать, что репрезентативность выборки и ошибка выборки – разные явления. Репрезентативность, в отличие от ошибки никак не зависит от размера выборки.

Ошибка выборки

- Отклонение результатов, полученных с помощью выборочного наблюдения от истинных данных генеральной совокупности.
Ошибка выборки бывает двух видов – **статистическая и систематическая**.
- **Статистическая ошибка** зависит от размера выборки. Чем больше размер выборки, тем она ниже.
Пример:
Для простой случайной выборки размером 400 единиц максимальная статистическая ошибка (с 95% доверительной вероятностью) составляет 5%, для выборки в 600 единиц – 4%, для выборки в 1100 единиц – 3%. Обычно, когда говорят об ошибке выборки, подразумевают именно статистическую ошибку.

Ошибка выборки

Систематическая ошибка зависит от различных факторов, оказывающих постоянное воздействие на исследование и смещающих результаты исследования в определенную сторону.

Пример:

- Использование любых вероятностных выборок занижает долю людей с высоким доходом, ведущих активный образ жизни. Происходит это в силу того, что таких людей гораздо сложнее застать в каком-либо определенном месте (например, дома).
- Проблема респондентов, отказывающихся отвечать на вопросы анкеты (доля «отказников», для разных опросов, колеблется от 50% до 80%)

В некоторых случаях, когда известны истинные распределения, систематическую ошибку можно нивелировать введением квот или перевзвешиванием данных, но в большинстве реальных исследований даже оценить ее бывает достаточно проблематично.

Типы выборок

- **Выборки делятся на два типа:**
 - вероятностные
 - невероятностные
- **Фундаментальное различие между выборками состоит в их принадлежности к вероятностным или невероятностным выборкам. Вероятностную выборку еще часто называют случайной выборкой, и только в отношении случайных, или вероятностных, выборок можно быть уверенными, что они лишены тенденциозности. В соответствии с определением все члены генеральной совокупности случайной выборки имеют равные шансы быть в ней представленными, и самый очевидный пример случайной выборки — это обычная лотерея. Все шары или числа, остающиеся в розыгрыше, сохраняют равные шансы быть выбранными в следующий раз. Ясно, что никакая тенденция не влияет на выбор чисел в лотерее.**

Вероятностные выборки

- **1. Случайная выборка (простой случайный отбор)**
Такая выборка предполагает однородность генеральной совокупности, одинаковую вероятность доступности всех элементов, наличие полного списка всех элементов. При отборе элементов, как правило, используется таблица случайных чисел.
- **2. Механическая (систематическая) выборка**
Разновидность случайной выборки, упорядоченная по какому-либо признаку (алфавитный порядок, номер телефона, дата рождения и т.д.). Первый элемент отбирается случайно, затем, с шагом 'n' отбирается каждый 'k'-ый элемент. Размер генеральной совокупности, при этом – $N=n*k$

Вероятностные выборки

3. **Стратифицированная** (районированная)

Применяется в случае неоднородности генеральной совокупности. Генеральная совокупность разбивается на группы (страты). В каждой страте отбор осуществляется случайным или механическим образом.

4. **Серийная** (гнездовая или кластерная) выборка

При серийной выборке единицами отбора выступают не сами объекты, а группы (кластеры или гнёзда). Группы отбираются случайным образом. Объекты внутри групп обследуются сплошным образом.

Невероятностная выборка

- Отбор в такой выборке осуществляется не по принципам случайности, а по субъективным критериям – доступности, типичности, равного представительства и т.д..

Квотная выборка

Изначально выделяется некоторое количество групп объектов (например, мужчины в возрасте 20-30 лет, 31-45 лет и 46-60 лет; лица с доходом до 13 тысяч , с доходом от 13 до 30 тысяч и с доходом свыше 30 тысяч) Для каждой группы задается количество объектов, которые должны быть обследованы. Количество объектов, которые должны попасть в каждую из групп, задается, чаще всего, либо пропорционально заранее известной доле группы в генеральной совокупности, либо одинаковым для каждой группы. Внутри групп объекты отбираются произвольно. Квотные выборки используются в маркетинговых исследованиях достаточно часто.

Невероятностная выборка

Метод снежного кома

Выборка строится следующим образом. У каждого респондента, начиная с первого, просят контакты его друзей, коллег, знакомых, которые подходили бы под условия отбора и могли бы принять участие в исследовании. Таким образом, за исключением первого шага, выборка формируется с участием самих объектов исследования. Метод часто применяется, когда необходимо найти и опросить труднодоступные группы респондентов (например, респондентов, имеющих высокий доход, респондентов, принадлежащих к одной профессиональной группе, респондентов, имеющих какие-либо схожие хобби/увлечения и т.д.)

Невероятные выборки

Стихийная выборка

Опрашиваются наиболее доступные респонденты. Типичные примеры стихийных выборок – опросы в газетах/журналах, анкеты, отданные респондентам на самозаполнение, большинство интернет-опросов. Размер и состав стихийных выборок заранее не известен, и определяется только одним параметром – активностью респондентов.

Выборка типичных случаев

Отбираются единицы генеральной совокупности, обладающие средним (типичным) значением признака. При этом возникает проблема выбора признака и определения его типичного значения.

Основные способы организации выборочного наблюдения

Существуют пять основных способов организации выборочного наблюдения:

1. **простой случайный отбор**, при котором объектов случайно извлекаются из генеральной совокупности объектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность.

Такие выборки называются **собственно-случайными**

2. **простой отбор с помощью регулярной процедуры** осуществляется с помощью механической составляющей (например, даты, дня недели, номера квартиры, буквы алфавита и др.) и полученные таким способом выборки называются **механическими**

Основные способы организации выборочного наблюдения

Стратифицированный отбор

заключается в том, что генеральная совокупность объема подразделяется на подсовокупности или слои (страты) объема так что страты представляют собой однородные объекты с точки зрения статистических характеристик (например, население делится на страты по возрастным группам или социальной принадлежности; предприятия — по отраслям). В этом случае выборки называются стратифицированными (иначе, расслоенными, типическими, районированными);

Основные способы организации выборочного наблюдения

- 4. **методы серийного отбора** используются для формирования **серийных или гнездовых выборок**. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или население при территориально-административном делении страны). Отбор серий можно осуществить собственно-случайным или механическим способом. При этом проводится сплошное обследование определенной партии товара, или целой территориальной единицы (жилого дома или квартала);
- 5. **комбинированный (ступенчатый) отбор** может сочетать в себе сразу несколько способов отбора (например, стратифицированный и случайный или случайный и механический); такая выборка называется комбинированной.

№ п/п	Тип выборки	Краткое описание	Достоинства	Недостатки
1	Простая случайная	Из однородной совокупности, все элементы которой известны и могут быть пронумерованы, отбор единиц выборки осуществляется с помощью таблиц случайных чисел.	Предварительная информация о генеральной совокупности очень проста – перечень или опись ее элементов. Легко классифицируются и вычисляются ошибки.	Необходима основа выборки – перечень всех элементов генеральной совокупности, что делает выборку малоприменимой для очень больших генеральных совокупностей. Требуется качественно однородных совокупностей, что в социальной сфере бывает редко.
2	Систематическая	Необходим полный список единиц генеральной совокупности. В выборку отбирается по одному объекту через интервал, равный шагу отбора - отношению объема генеральной совокупности к объему выборки	Позволяет с помощью простой техники отбора при небольшом объеме выборки охватить сравнительно большие генеральные совокупности	Существует опасность совпадения интервала отбора со скрытой периодичностью распределения признака в генеральной совокупности, что может привести к смещениям. Неопределенность выбора первоначального объекта (начальной точки отбора).
3	Гнездовая	Выборочные единицы отбираются с помощью одного из способов случайного отбора. Единицы отбора представляют собой статистические группы (гнезда), которые целиком или выборочно подвергаются обследованию.	Уменьшаются затраты (по сравнению с простой случайной выборкой равного объема) на организацию процедур отбора, например, при пространственно разбросанной генеральной совокупности. Относительно простое составление основы выборки для выбранных гнезд, чем полный перечень элементов, соответствующих всем гнездам в генеральной совокупности.	При формировании искусственных гнезд создается трудность отнесения каждого отдельного элемента генеральной совокупности только к одному гнезду и обеспечения приблизительно одинаковых размеров гнезд.
4	Стратифицированная (районированная) выборка, расслоенный отбор) А. Пропорциональное размещение	Исследуемая совокупность предварительно разделяется на страты (слои) в соответствии с генеральным распределением известных и значимых для исследования признаков; из каждой страты извлекается выборка Объем выборки из страты пропорционален размеру страты в генеральной совокупности	При удачном подразделении совокупности на однородные группы расслоенный отбор дает выигрыш в точности по сравнению с простым случайным.	Если классификация страты не совпадает с распределением изучаемых признаков, возможно смещение выборки по этим признакам.