

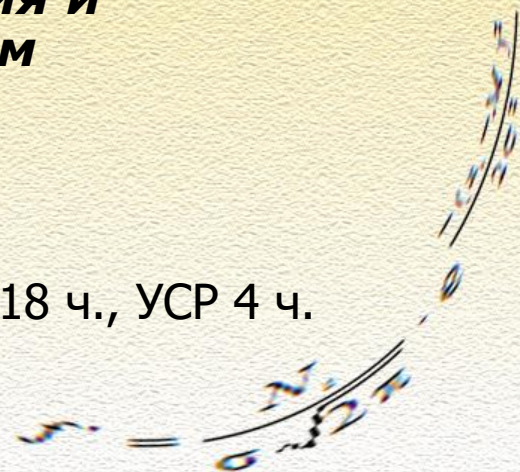


МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ

Карпиченко Александр Александрович

*доцент кафедры почвоведения и
геоинформационных систем*

Всего 40 ч., в т.ч. лекций 18 ч., лабораторных 18 ч., УСР 4 ч.



Литература

- elib.bsu.by/handle/123456789/17887
- geo.bsu.by
- Математические методы в географии: учебно-методическое пособие / Н.К. Чертко, А.А. Карпиченко. — Минск: БГУ, 2009.

Литература

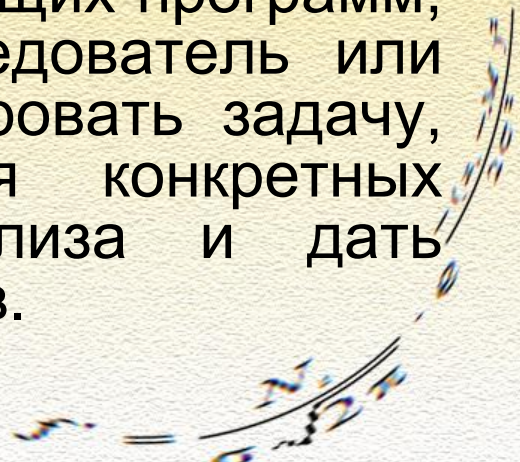
Математические методы в географии: электронный учебно-методический комплекс для специальностей: 1-31 02 01 «География (по направлениям)», направление специальности: 1-31 02 01-02 «География (научно-педагогическая деятельность)»; 1-31 02 03 «Космоаэрокартография»; 1-33 01 02 «Геоэкология» / А. А. Карпиченко, Н. К. Чертко; БГУ, фак. географии и геоинформатики, каф. почвоведения и геоинформационных систем. – Минск: БГУ, 2019. – 61 с. №003026032020, Деп. в БГУ 26.03.2020. Режим доступа: <http://elib.bsu.by/handle/123456789/241510>



Введение

Географические исследования и практические задачи базируются на большом объеме количественной информации, которую необходимо **объективно** оценить и провести группировку или классификацию, доказать зависимость или провести моделирование, выявить оптимальные условия развития или установить пространственные закономерности развития объектов или явлений, дать прогноз их эволюции.

Эти вопросы успешно решаются с помощью математических методов и соответствующих программ, разработанных для компьютера. Исследователь или практик должен лишь четко сформулировать задачу, выбрать наиболее подходящий для конкретных условий математический метод анализа и дать объективную интерпретацию результатов.



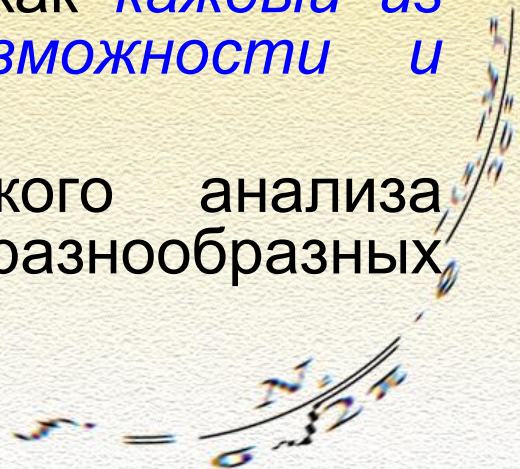


Введение

Математика позволяет решать задачи *частные* и *общие*. Например, расход воды в реке рассчитывается на основе специальной частной формулы, а загрязнение воды в реке под воздействием предприятия оценивается с применением *факторного анализа* – общего для решения многих специальных географических задач.

Механический подход при использовании математики недопустим. В конкретной ситуации надо выбрать надежный математический прием, так как *каждый из методов анализа имеет свои возможности и ограниченную область применения*.

Большинство методов статистического анализа универсальны и могут применяться в разнообразных отраслях деятельности человека.



Введение

Значительное влияние на развитие математических методов оказали открытый **Яковом Бернулли** (1654–1705) *закон больших чисел* и *теория вероятности*, основы которой разработал французский математик и астроном **Пьер Симон Лаплас** (1749–1827).

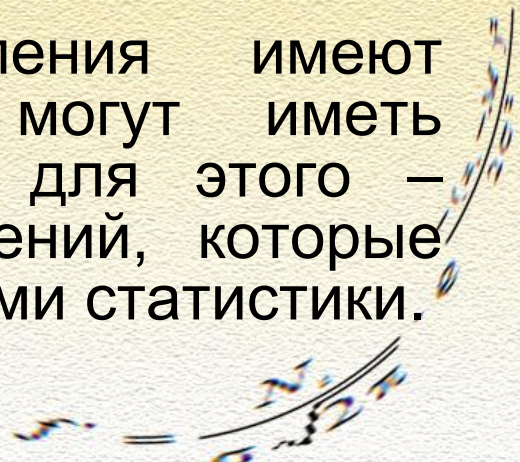




Введение

На основе *теории вероятности*, которая позволяет выявить определенные тенденции в кажущемся хаосе случайных явлений, появилась **математическая статистика**. Она позволяет дать количественную оценку вероятностей различных явлений, которые не имеют постоянных, всегда одних и тех же исходов. Большинству природных и экономических явлений свойственна **вариабельность** (изменение в определенных пределах). Например, температура воздуха меняется ежечасно, ежедневно, ежемесячно.

Однако многие хаотические явления имеют упорядоченную структуру, поэтому могут иметь конкретную оценку. Главное условие для этого — статистическая устойчивость этих явлений, которые можно описать математическими методами статистики.

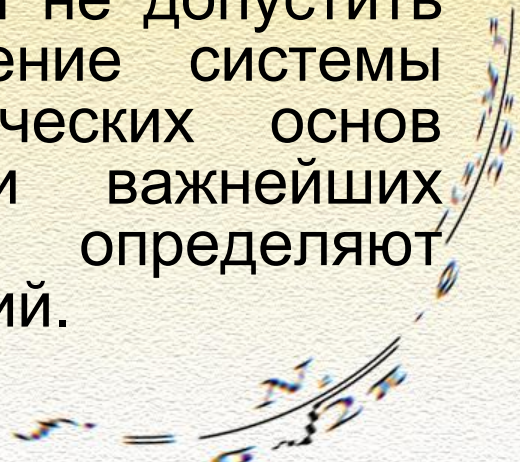




Введение

Следует иметь в виду, чрезмерное увеличение объема исходной информации ведет к увеличению «информационного шума» (роста числа помех). Достигнув известного предела «шум» подавляет исходную информацию.

Чем сложнее система, тем больше рассматриваемых взаимосвязанных переменных, тем труднее установить множество отношений. *Количественные методы анализа помогают выбрать ведущие факторы, причины, признаки.* В таких случаях важно понимание смысла математических методов, чтобы не допустить ошибочных выводов. Начинать изучение системы необходимо с усвоения методологических основ организации самих исследований и важнейших элементов системологии, которые определяют последовательность дальнейших действий.



Введение

Современные географические методы исследования (сравнительно-географический, системный и другие) необходимо использовать в сочетании с математическим обоснованием результатов. Математические методы позволяют широко использовать **системный анализ**, как наиболее совершенный.

Любой географический объект исследования может быть представлен как **система** – определенный объект, состоящий из множества частей, которые взаимосвязаны не только между собой, но и с соседними объектами-системами. Установить целостность и структуру, иерархичность, величину и направленность связей в системе, их характер позволяют математические методы путем создания формализованных систем.

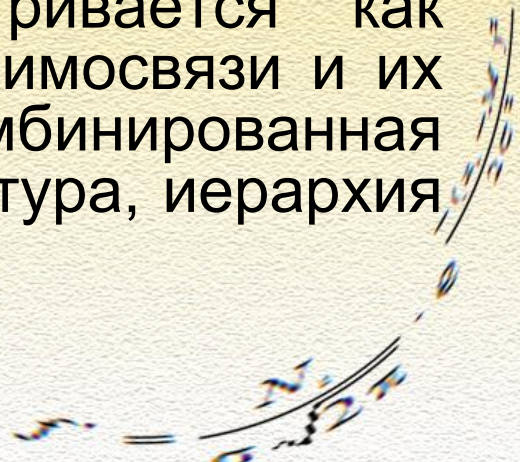


Введение

Системный подход основан на исследовании объектов как систем, создает единую теоретическую модель.

Системный анализ представляет собой совокупность методологических средств, позволяющих обосновать проблемы научно-практического характера.

Успешное использование системного анализа возможно при реализации следующих важнейших принципов, опирающихся на математические методы: выявляется и формулируется конечная цель исследования; система-объект рассматривается как единое целое, в ней выявляются все взаимосвязи и их результаты; строится обобщенная комбинированная модель (модели), где отображаются структура, иерархия и взаимосвязи.



Введение

Выделяются две группы систем: *материальные* и *абстрактные*. Традиционные методы географии изучают материальные системы. *Социальные системы* через техногенез могут оказывать воздействие на природные. По развитию выделяют системы *статичные* (предприятия) и *динамичные* (ландшафт).

По характеру взаимодействия системы делятся на *закрытые* (в них не поступает и из них не выводится вещество, происходит лишь обмен энергией) и *открытые* (постоянно происходит ввод и вывод вещества и обмен энергией). В открытой системе, например, ландшафте постоянно протекающие процессы и явления создают подвижное равновесие, т.е. некоторую стабильность в определенных условиях среды и общества.



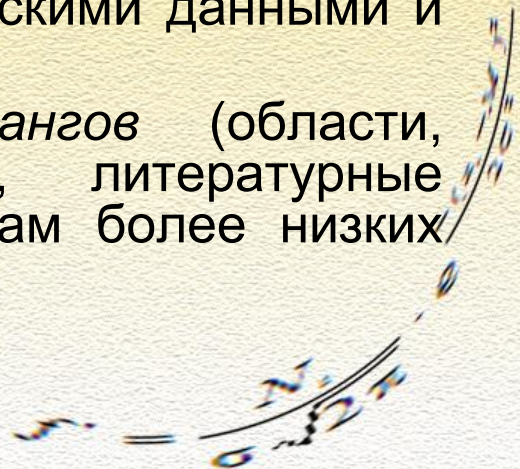
1. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Источником материала для статистической обработки могут быть **собственные экспериментальные исследования, статистическая информация, аналитические данные других исследователей, фондовые материалы, литературные источники, географические карты, аэрофотоснимки.**

При изучении территориальных комплексов *низших рангов* (фаций, урочищ), промышленных предприятий, объектов сельскохозяйственного назначения наиболее ценными для статистической обработки являются материалы собственных исследований.

При изучении объектов *среднего ранга* возрастает роль отраслевых и специальных карт вместе с авторскими данными и литературными источниками.

Для исследования объектов *высоких рангов* (области, провинции, регионы) используются карты, литературные источники, обобщающие материалы по объектам более низких рангов.

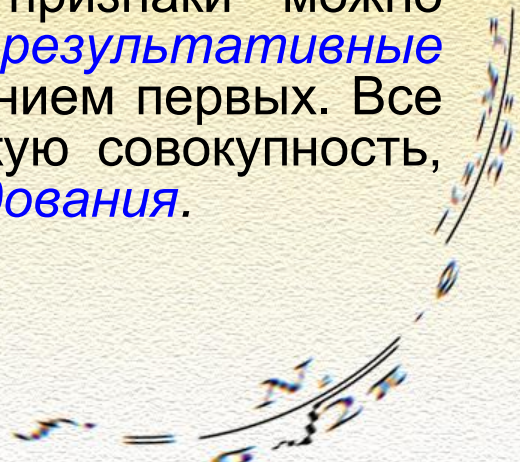




1.1. Генеральная совокупность и выборка

Первичным элементом в статистике является *единица наблюдения* (варианта, дата): 3 4 3 4 3 3 3 3. Их ряд образуют *статистическую совокупность*, которая характеризует *объект исследования*.

Большинство единиц наблюдения имеют *вероятностный, случайный характер*. По виду исследуемые *признаки* могут быть *качественными* и *количественными*. Количественные признаки имеют числовое выражение, качественные – словесное (образование начальное, среднее, высшее). Качественным признакам при статистической обработке присваивают балл или ранг соответственно их смыслу (начальное образование – 1 балл, среднее – 2, высшее – 3). Исследуемые признаки можно подразделить на *факторные* (факториальные) и *результативные* (результатирующие); вторые изменяются под влиянием первых. Все единицы наблюдения, входящие в статистическую совокупность, объединены *единством места и времени исследования*.





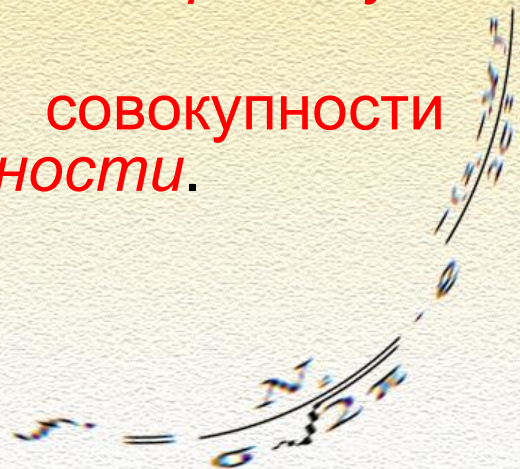
1.1. Генеральная совокупность и выборка

Чрезмерное увеличение объема любой исходной информации ведет к увеличению «информационного шума» (погрешностей), который подавляет искомую исследователем информацию. Это отражается на *вариабельности* (изменчивости, случайности) процессов и явлений.

По *времени* наблюдение может быть *текущим* (непрерывным) и *единовременным* (в один и тот же момент времени в разных точках – метеонаблюдения на постах). По *охвату* исследование может быть *сплошное* и *не сплошное*. Эта особенность определяет ход и методику статистического анализа.

Сплошное статистическое исследование (перепись всего населения республики) образует генеральную совокупность.

Общее число членов генеральной совокупности называют объемом генеральной совокупности.

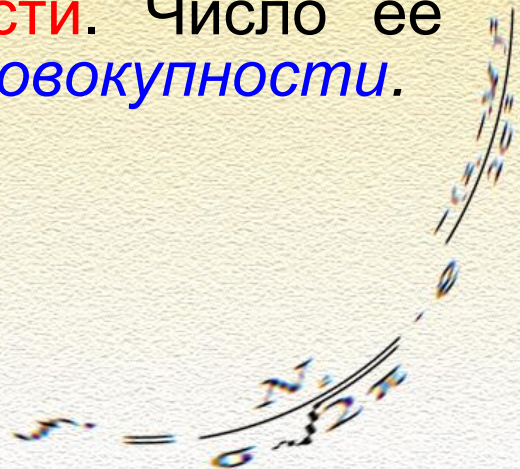




1.1. Генеральная совокупность и выборка

Из-за больших размеров генеральной совокупности (перепись населения) или из-за отсутствия определенных границ этой совокупности (Европа) оно проводится редко.

На исследование генеральной совокупности затрачивается много средств и времени, поэтому ограничиваются методом *выборочного исследования* (не сплошного) из *генеральной совокупности*. **Выборка образует совокупность наблюдений, полученных с целью объективной характеристики и получения информации о генеральной совокупности.** Число ее членов называют *объемом выборочной совокупности*.





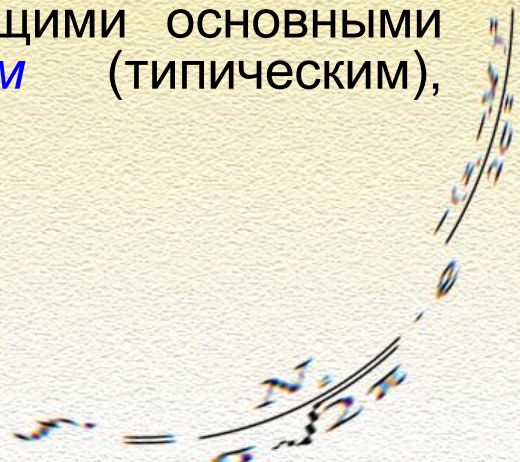
1.1. Генеральная совокупность и выборка

Выборочное исследование можно проводить такими методами, как *монографический, основного массива и выборочным*.

Монографический метод используется для описания объекта с какими-либо особенностями (зонирование города с развитой машиностроительной промышленностью). Выводы могут быть распространены только на группу аналогичных объектов.

Метод основного массива дает представление о конкретном объекте, поэтому переносить полученные закономерности на другие объекты нельзя (бассейн р. Неман). Наиболее распространен метод выборочного исследования из генеральной совокупности.

Выборка может быть представлена следующими основными типами отбора: *случайным, направленным* (типическим), *смешанным*.

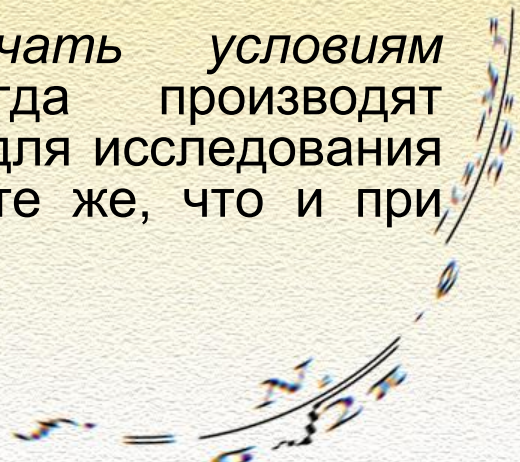




1.1. Генеральная совокупность и выборка

При **случайном отборе** все объекты имеют одинаковую возможность попасть в выборку. В его основе лежит перемешивание. Для этого можно использовать таблицу случайных чисел (прил. 2 в «Математические методы в географии»). Начав с любой четырехзначной колонки и, двигаясь по столбцу сверху вниз или снизу вверх, выписывают первые или последние однозначные цифры для объема выборки до 9, двухзначные – для объема выборки от 10 до 99, трехзначные – для объема выборки от 100 до 999 и т. д. По второму варианту из объектов в списке (алфавитном или ином) в выборку включают каждый третий или пятый, или десятый (механическая выборка) и т. д.

*Случайная выборка может не отвечать условиям исследования из-за неоднородности. Тогда производят **целенаправленный (когортный) отбор**, выбирая для исследования типичные объекты. Правила отбора остаются те же, что и при случайном отборе.*

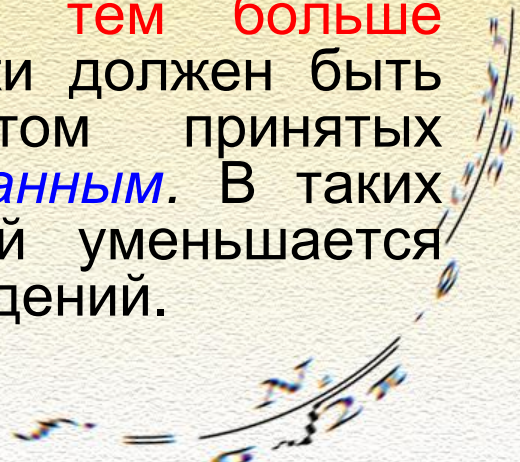




1.1. Генеральная совокупность и выборка

Смешанный отбор производят в тех случаях, когда необходимо дать характеристику неоднородного объекта. Например, холмисто-моренный ландшафт делят фации с однородными условиями, в каждой из которых производят случайный отбор. Полученные результаты объединяют в одну выборку.

Соблюдения правил составления выборки дают возможность *наиболее полно и точно, т. е. репрезентативно*, характеризовать генеральную совокупность. Величина *ошибки репрезентативности* зависит от изменчивости изучаемого признака. **Чем больше разброс значений изучаемого признака, тем больше статистическая ошибка.** Отбор для выборки должен быть также *научно обоснованным* с учетом принятых методических правил, т. е. *рентдомизированным*. В таких случаях при меньшем числе наблюдений уменьшается вероятность систематических ошибок наблюдений.



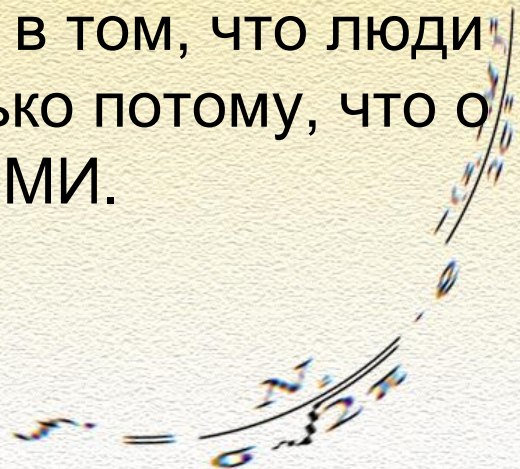


1.1. Генеральная совокупность и выборка

Систематическая ошибка отбора — статистическое понятие, показывающее, что выводы, сделанные применительно к какой-либо группе, могут оказаться неточными вследствие неправильного отбора в эту группу.

В общем случае часто невозможно выделить систематическую ошибку выборки только на основе статистических методов.

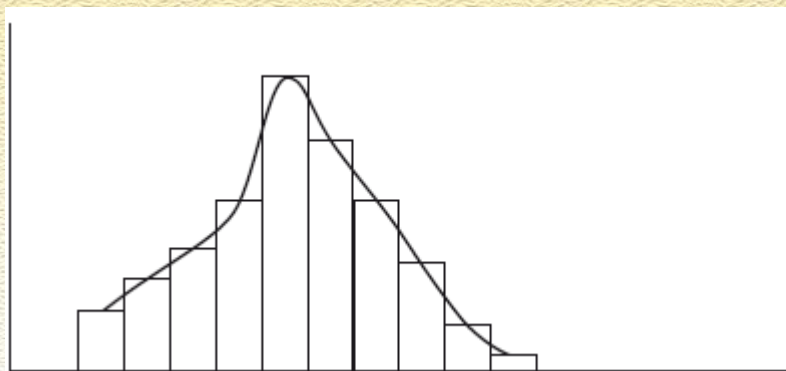
Аналогом ошибки выжившего является парадокс доступности информации. Он заключается в том, что люди могут считать исход более вероятным только потому, что о подобных исходах больше сообщается в СМИ.



1.1. Генеральная совокупность и выборка

На втором этапе статистического исследования проводят *сводку* и *группировку* данных. Варианты группировок следующие: разделение анализируемой статистической совокупности на группы по тем или иным признакам; объединение мелких однородных групп в более крупные; комплексная группировка на основе многих учетных признаков, даже если они разнородные.

Типологическая группировка выделяет в совокупности качественно однородные в существенном отношении группы. Группировка по своей сути представляет собой процесс классификации. В государственной статистике используют *классификаторы* – специальные справочники, инструкции, указания.

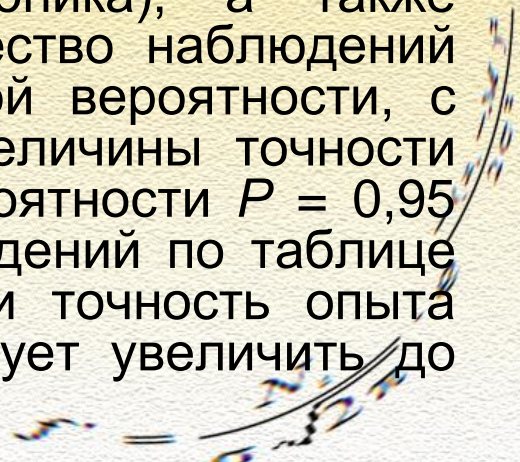




1.1. Генеральная совокупность и выборка

Самым сложным является определение **объема наблюдений в исследованиях**, который необходим для получения надежного представления о характере изменчивости признака в генеральной совокупности. *Если объект исследуется впервые, то определить объем выборки трудно.* В большинстве случаев достаточно точные результаты получают при объеме выборки около 100. Оптимальный объем выборки обычно пропорционален степени изменчивости признака. Если признак сильно изменяется, то количество измерений следует увеличить.

Предложены также другие способы определения величины выборочной совокупности при исследованиях: **по таблице достаточно больших чисел** (прил. 1 учебника), а также **расчетным способом**. В обоих случаях количество наблюдений определяется исходя из величины допустимой вероятности, с какой предполагается делать заключения, и величины точности опыта. Например, при допустимом уровне вероятности $P = 0,95$ (95 %) и точности опыта $p = 5 \%$ число наблюдений по таблице достаточно больших чисел составит 384. Если точность опыта увеличить до 1 %, то число наблюдений следует увеличить до 9603.



1.1. Генеральная совокупность и выборка

Чаще всего **ориентировочный объем** (N) выборочной совокупности рассчитывают по формулам, в которых вероятность заменяют степенью варьирования:

$$N = \sigma^2 / m_M^2,$$

где σ^2 – дисперсия; m_M – ошибка среднего арифметического.

Допустим, варьирование признака (колебание температуры) составляет 7°C , тогда число наблюдений выборочной совокупности с ошибкой среднего арифметического $m = \pm 0,5^\circ\text{C}$ составит: $N = \sigma^2 / m_M^2 = 7^2 / 0,5^2 = 196$.

Объем выборочной совокупности можно также определить по ожидаемому коэффициенту вариации (V) и точности опыта (p) с учетом поправочного коэффициента (1,96) для уровня вероятности 0,95 и 0,99:

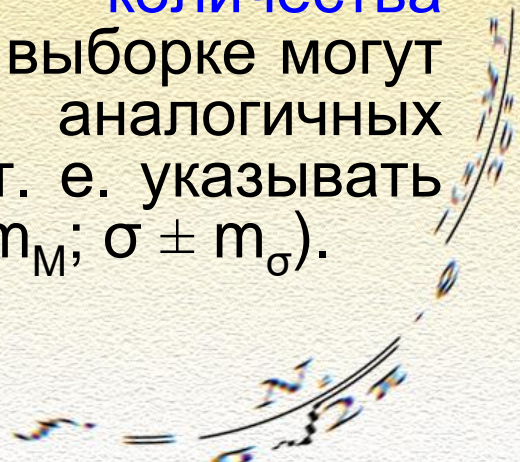
$$N = (1,96 \cdot V)^2 / p^2.$$



1.1. Генеральная совокупность и выборка

Например: Для расчета коэффициента увлажнения в зависимости от количества выпадающих осадков и испарения с ожидаемой точностью опыта 3 % и коэффициента вариации 30 % потребуется следующий объем выборочной совокупности $N = (1,96 \cdot 30)^2 / 3^2 = 384$.

Определение объема выборочной совокупности необходимо для получения достоверной информации о генеральной совокупности путем расчета минимального, но объективного количества наблюдений. Полученные параметры по выборке могут служить приблизительными оценками аналогичных параметров генеральной совокупности, т. е. указывать пределы, в которых они заключены ($M \pm m_M; \sigma \pm m_\sigma$).



1.2. Обработка вариационного ряда

Варианты в статистической совокупности подвергаются обработке. Для этого составляется *вариационный ряд*, т. е. варианты располагают по возрастающим или убывающим величинам. Варианты в выборке, относящиеся к одному и тому же признаку, практически не совпадают между собой, или *варьируют*. Те варианты, которые резко отличаются от вариантов статистической совокупности и вызывают сомнение у исследователя определяются как *артефакт*. Они располагаются в начале или в конце вариационного ряда. Артефакт исключается из статистической совокупности и не подлежит обработке. Например, в приведенных вариационных рядах: 2, 9, 11, 12, 13, 15 и 25, 27, 29, 32, 55 почти все соседние показатели весьма близки по значению. Вызывают сомнение варианты 2 в первом ряду и 55 во втором. Их можно принять за артефакт и исключить (выбраковать) из обработки. Выбраковка должна быть статистически доказана.

1.2. Обработка вариационного ряда

Существующие критерии выбраковки основываются, как правило, на допущении, что выборка распределяется по нормальному или близкому к нему закону. В качестве критерия выбраковки может быть использован критерий τ (прил. 3). Если критерий τ вычисленный (фактический) больше или равен критерию τ табличному ($\tau_{\text{ф}} \geq \tau_{\text{т}}$) при объеме выборки N и уровне значимости α (0,05 или 0,01), то соответствующие значения вариантов выборки (x) допустимо отбросить как артефакт. Значения τ для вызывающей сомнение величины вычисляются по следующим формулам:

$$\tau_1 = (x_2 - x_1) / (x_{n-1} - x_1)$$

для наименьшего значения переменной величины в вариационном ряду (x_1);

$$\tau_n = (x_n - x_{n-1}) / (x_n - x_2)$$

для максимального значения переменной в вариационном ряду.

1.2. Обработка вариационного ряда

При составлении вариационного ряда по урожайности сельскохозяйственных культур в разрезе хозяйств одного из районов получен следующий ряд значений: 10,8; 12,5; 12,9; 13,2; 20,2 (ц/га). Вызывает сомнение максимальное значение в выборке варианты 20,2. Следует доказать, можно ли ее отнести к артефакту. Подставляем необходимые данные в формулу:

$$\tau_5 = (x_5 - x_4) / (x_5 - x_2) = (20,2 - 13,2) / (20,2 - 12,5) = 0,958.$$

Вычисленное значение критерия ($\tau_5 = 0,958$) сравнивают с табличным значением (τ_T), учитывая объем выборки ($N = 5$). В прил. 3 критическое значение критерия артефакта для $N = 5$ и уровня значимости $\alpha 0,05$ и $0,01$ соответственно будут равны $0,807$ и $0,916$, что меньше расчетного значения ($\tau_5 = 0,958$). Поэтому варианту $20,2$ признают артефактом и исключают из статистической обработки как сомнительную.