МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ

Карпиченко Александр Александрович

доцент кафедры почвоведения и земельных информационных систем

Литература

- Математические методы в географии: учебно-методическое пособие / Н. К. Чертко, А. А. Карпиченко. – Минск: БГУ, 2009.

Электронная библиотека БГУ

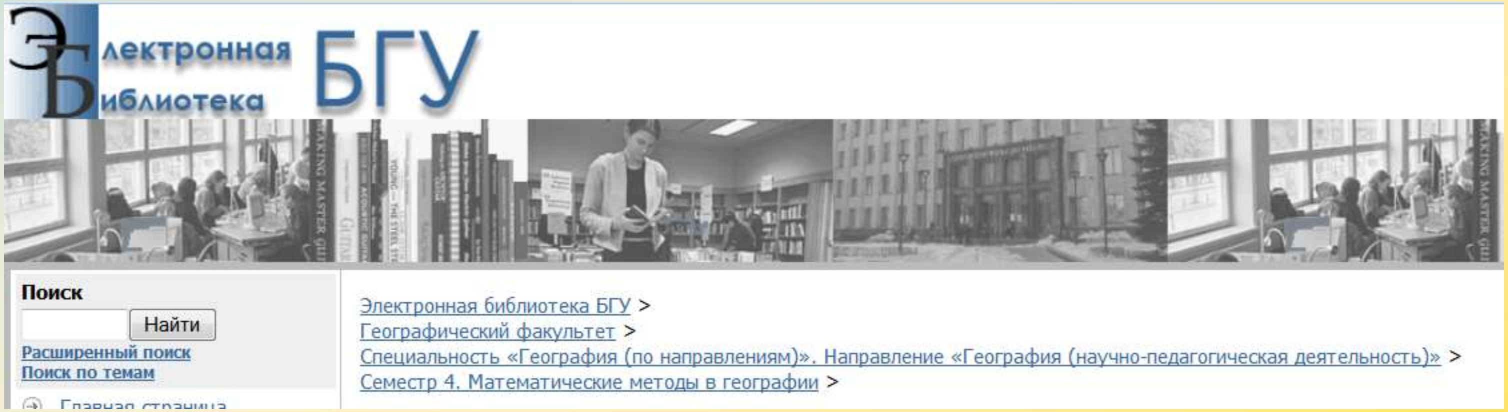
elib.bsu.by

Географический факультет

Специальность «География (по направлениям)». Направление «География (научно-педагогическая деятельность)»

Семестр 4. Математические методы в географии

Заглавия



Электронная библиотека БГУ

elib.bsu.by

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* (результирующие); вторые изменяются под влиянием первых. Все единицы наблюдения, входящие в статистическую совокупность, объединены *единством места и времени исследования.*

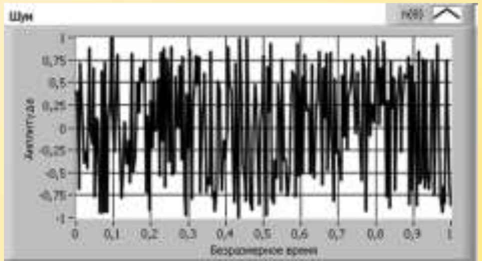
1.1. Генеральная совокупность и выборка

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*.

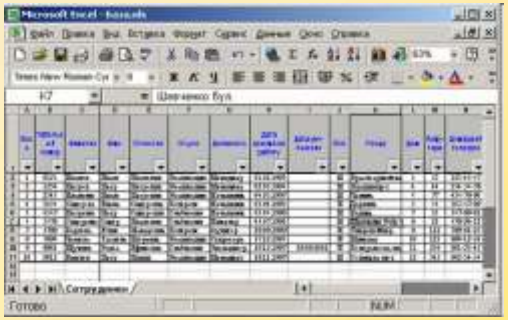


Карпиченко А.А.

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.*



Карпиченко А.А.

1.1. Генеральная совокупность и выборка

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

$$N = \sigma^2 / m^2_M ,$$

где σ – среднее квадратическое отклонение; m_M – ошибка среднего арифметического.

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

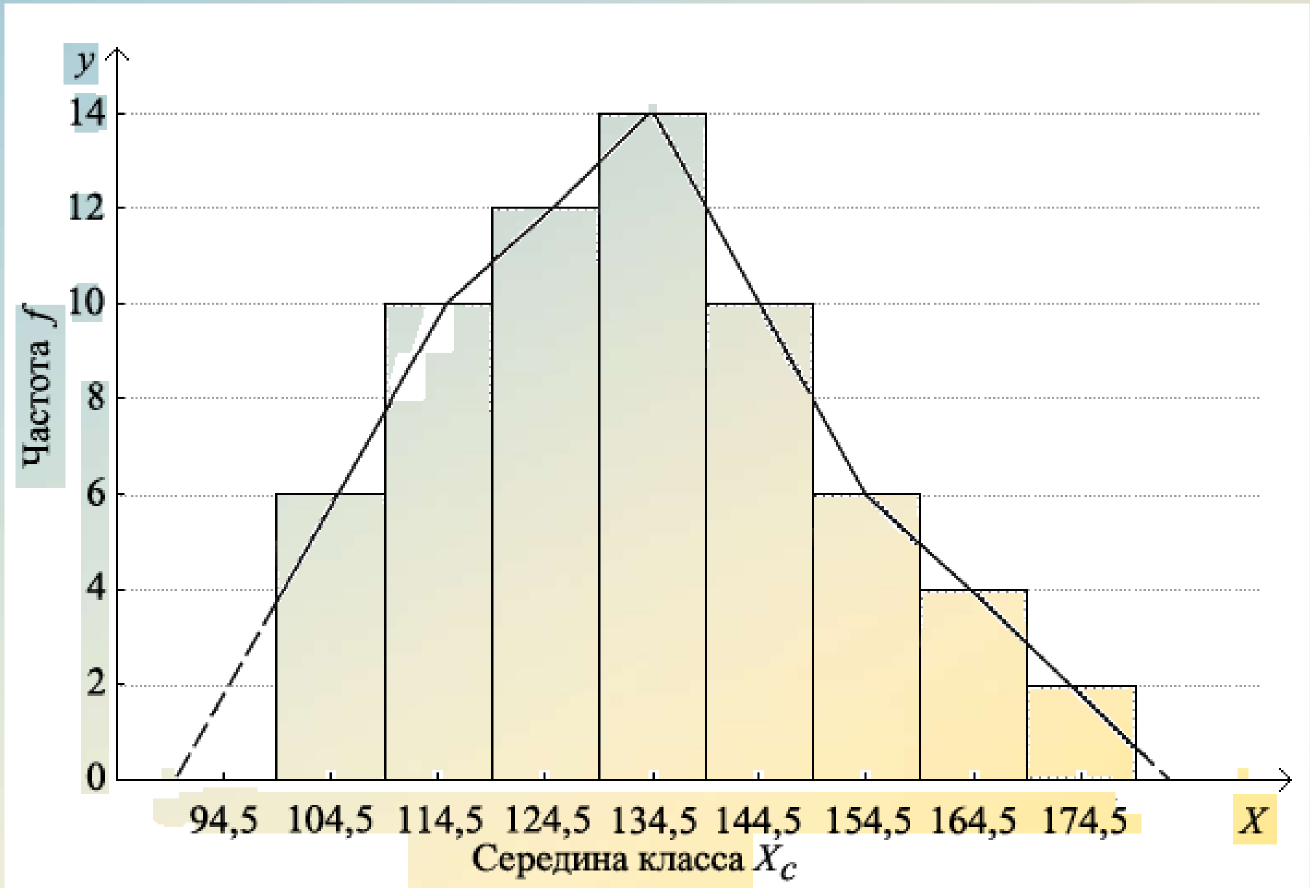
Объем выборочной совокупности можно также определить по ожидаемому коэффициенту вариации (V) и точности опыта (p) с учетом поправочного коэффициента (1,96) для уровня вероятности 0,95 и 0,99:

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

1.2. 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.

1.2. Обработка вариационного ряда



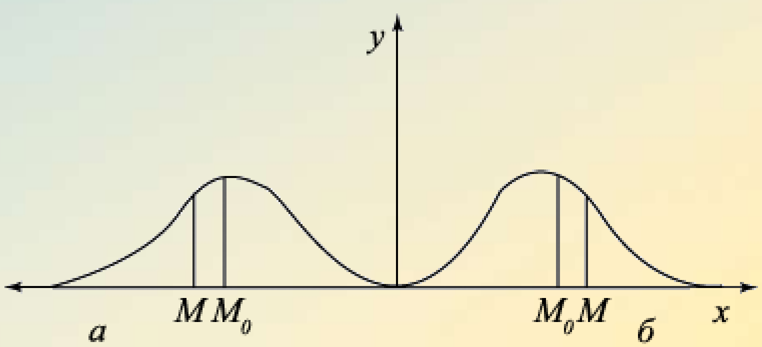
Способы графического представления вариационного ряда: кривая распределения и гистограмма

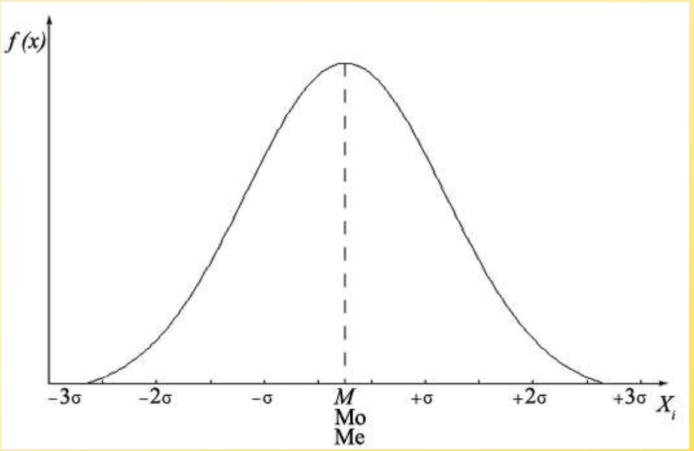
Карпиченко А.А.

1.2. Обработка вариационного ряда

Показатели асимметрии и эксцесса. Распределение частот в изучаемом объекте не всегда подчиняется закону нормального распределения. Это особенно четко проявляется при выражении вариационного ряда в виде графика. Распределение частот может быть представлено *асимметричной*, *островершинной* или *туповершинной* кривой.

Асимметрия кривой распределения обусловлена неравномерным размещением вариант по обе стороны от модального значения признака. Если число вариант больше справа от моды, распределение имеет положительную асимметрию, если слева – отрицательную.





а – отрицательная асимметрия, *б* – положительная асимметрия

1.2. Обработка вариационного ряда

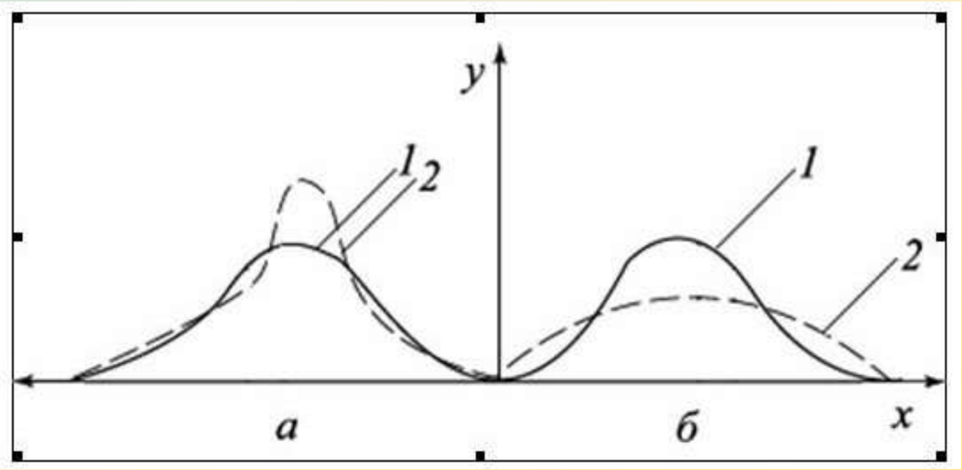
При получении асимметричной кривой следует проверить асимметричность распределения. Если асимметричность не будет доказана по критерию Стьюдента, то рассматриваемое распределение относят к симметричному. Для проверки асимметричности распределения вычисляют коэффициент асимметрии, его ошибку, затем на основании показателя достоверности устанавливают вид кривой распределения. Коэффициент асимметрии находят:

$$K_{as} = (M - \mathrm{Mo}) / \sigma, \text{ или } K_{as} = (M - \mathrm{Me}) / \sigma.$$

1.2. Обработка вариационного ряда

Эксцесс кривой распределения *(Е)* имеет место в тех случаях, когда большинство вариантов совокупности сосредоточено около среднего арифметического. Тогда эмпирическая кривая распределения отклоняется от нормальной теоретической кривой у ее вершины и количественно выражается показателем эксцесса.

Положительный эксцесс представлен островершинной кривой (эксцессивной, или лептокуртичной) (рис. *а*), отрицательный – плосковершинной (депрессивной, или платикуртичной) (рис. *б*). При сильном отрицательном эксцессе кривая может приобрести вид двухвершинной:



1 – теоретическая линия распределения, *2* – эмпирическая линия распределения

Карпиченко А.А.

1.3. Показатели описательной статистики

Статистические показатели распределения

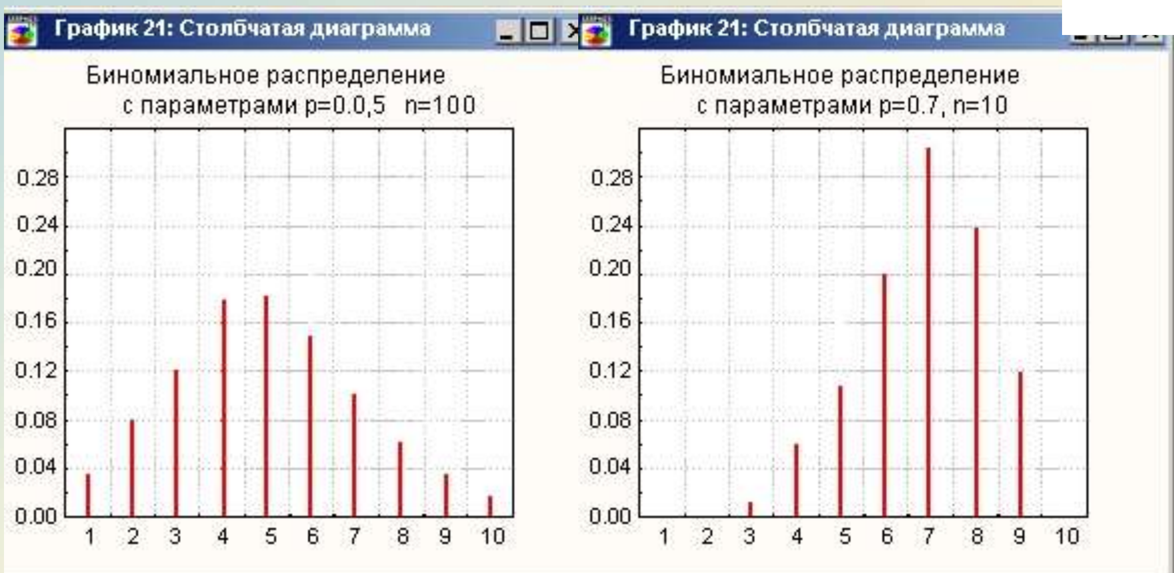
Показатели	Назначение показателей	Показатели распределения
Центра распределения (средние величины)	Описывают положение середины распределения	*Структурные (непараметрические) средние:* мода (Мо) и медиана (Ме). *Степенные средние:* среднее арифметическое (M), среднее гармоническое ($M_{гар}$), среднее геометрическое ($M_г$), среднее квадратическое ($M_{кв}$), среднее кубическое ($M_{куб}$), среднее взвешенное ($M_{взв}$)
Рассеивания вариант	Описывают степень разброса (вариабельности, изменчивости) вариант	Лимит (lim) Размах варьирования (*ampl*) Среднеквадратическое отклонение (σ) Дисперсия (σ^2) Коэффициент вариации (V) Квантили ($V_{0,25;\,0,5;\,0,75}$)
Формы распределения	Описывают симметрию и островершинность распределения вариант около центра	Коэффициент ассиметрии (As) Эксцесс (E) Гистограмма Полигон распределения

Карпиченко А.А.

1.4. Теоретические функции распределения

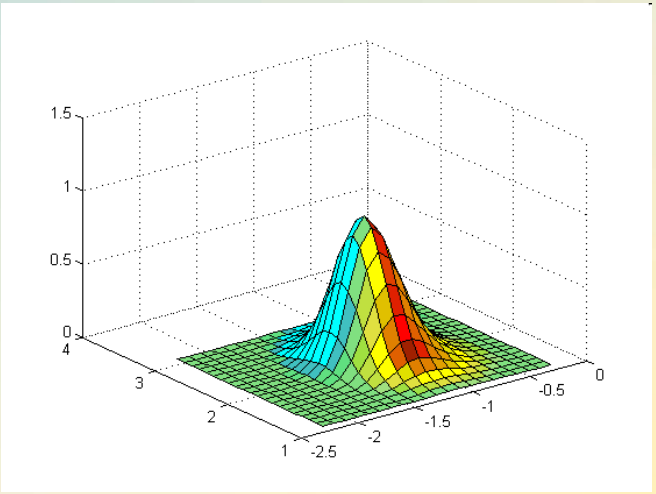
Биномиальное распределение (распределение Бернулли) возникает, когда оценивается сколько раз происходит событие в серии определенного числа независимых, выполняемых в одинаковых условиях наблюдений. Разброс вариант – следствие влияния ряда независимых и случайно сочетающихся факторов (есть событие или его нет). Характерно для альтернативного типа изменчивости признака.

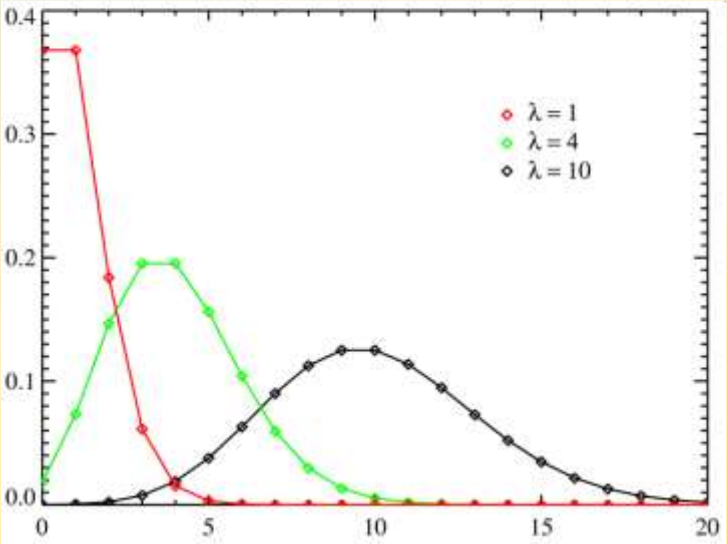
$$F_Y(y) \equiv \mathbb{P}(Y \leqslant y) = \sum_{k=0}^{\lfloor y \rfloor} \binom{n}{k} p^k q^{n-k}, \; y \in \mathbb{R}$$



1.4. Теоретические функции распределения

Распределение Пуассона рассматривается как предельный случай биномиального распределения и используется для характеристики редких событий. Отличительная особенность распределения Пуассона – величина дисперсии близка к величине среднего арифметического, например, длительное наводнение. Это проявляется в ситуациях, когда в определенный отрезок времени или на определенном пространстве происходит случайное число каких-либо событий, например, длительно повторяющиеся ураганы в течение одного летнего периода. На графике это распределение представляется в виде резко выраженной асимметрии.





1.4. Теоретические функции распределения

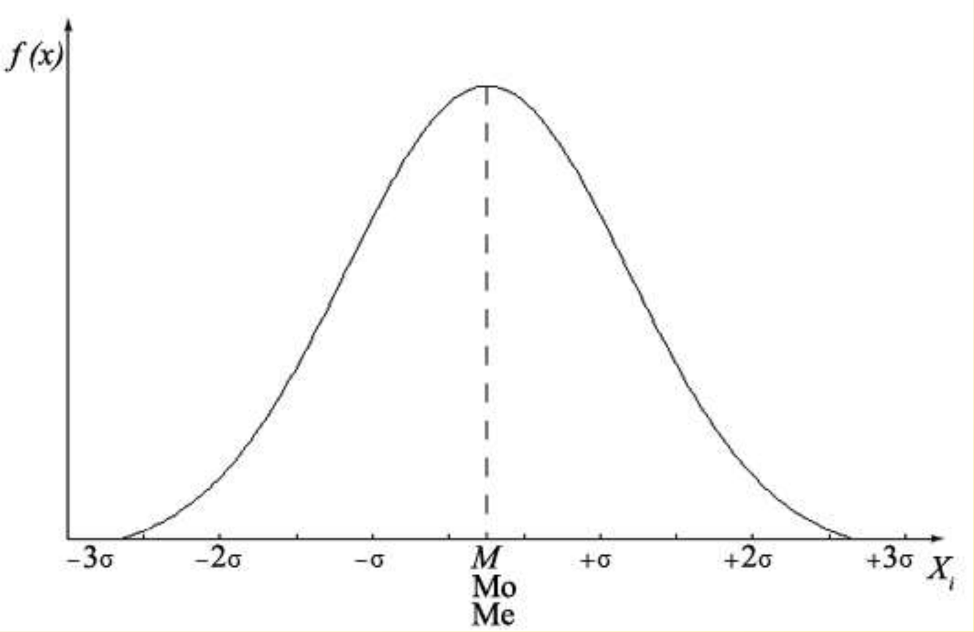
Нормальное распределение (распределение Гаусса) используется для приближенного описания явлений, которые носят вероятностный, случайный характер. Приоритет в открытии этого закона принадлежит Де Муавру (1733), но его связывают с именем Гаусса, исследовавшего его в начале 19 в.

Распределение Гаусса имеет место среди природных и экономических явлений. В системе признак варьирует под влиянием большого количества взаимно независимых факторов, каждый из которых мало влияет на его общую вариабельность. Причем одни факторы приводят к возрастанию величины признака, другие – к уменьшению. Встречаемость вариантов, занимающих середину совокупности, максимальна. Такое распределение считается нормой для случайных величин, поэтому оно получило название нормального.

1.4. Теоретические функции распределения

Графически нормальное распределение выражается плавной симметричной куполообразной кривой с приближающимися к оси абсцисс ветвями (кривая плотности нормального распределения).

Кривая показывает, что большие отклонения от средней встречаются реже, чем малые. С уменьшением среднего квадратического отклонения (σ) кривая нормального распределения становится все более островершинной. Площадь, заключенная под кривой нормального, всегда принимается равной единице.



Карпиченко А.А.

1.4. Теоретические функции распределения

При нормальном распределении среднее, мода и медиана совпадают. Кривая плотности не пересекает оси абсцисс, что подтверждает вероятность существования неограниченно больших отклонений. Уравнение нормального распределения можно записать в нескольких модификациях. Наиболее часто используется следующая формула, где *f* `– искомая ордината кривой (теоретическая частота); в степень числа *e* входит величина $(x_i - M_x)/\sigma$, получившая название нормированного отклонения.

$$f' = \frac{N_i}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x_i - M)^2}{2\sigma^2}}, \qquad (1.17)$$

1.4. Теоретические функции распределения

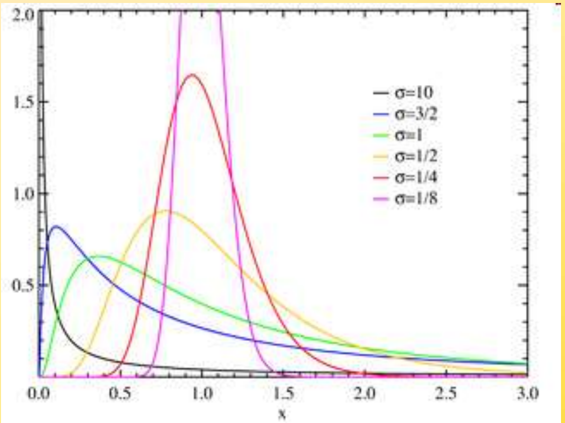
При нормальном распределении около 68,3 % всех вариант отклоняется от среднего значения не более, чем на величину среднего квадратического отклонения (±σ). Соответственно в пределах от –2σ до +2σ находится 95,5 % вариант, в пределах от –3σ до +3σ – 99,7 %.

Отклонение вариант от нормального закона распределения может указывать на влияние какого-либо другого фактора на статистическую совокупность.

1.4. Теоретические функции распределения

Логнормальное распределение. Некоторые распределения при изучении географических объектов имеют выраженную асимметрию (например распределение микроэлементов в почвах), поэтому представляет практический интерес преобразование асимметричного распределения в симметричное (нормальное). Иногда это возможно, если каждую варианту выборки выразить в виде логарифма ($\lg x_i$). В тех случаях, когда логарифм случайной величины (x_i) подчиняется нормальному распределению, а сами значения случайных величин распределены асимметрично, распределение случайной величины принято называть логарифмически нормальным, или логнормальным. Уравнение логнормального распределения имеет вид:

$$f' = \frac{1}{\sigma_{x_i}} \cdot \frac{e^{\frac{-(\lg x_i - M)^2}{2\sigma^2}}}{\sqrt{2\pi}}.$$



1.5. Статистические критерии различия

Различие между двумя выборками устанавливается с помощью ряда критериев: ***t* – распределение Стьюдента, наименьшего существенного различия (НСР), *F* – распределения Фишера, критерия соответствия (χ^2).**

Каждый из критериев применяется при определенных условиях, которые задаются целью исследования. Несоблюдение указанных условий может привести к ошибочным выводам.

1.5. Статистические критерии различия

Критерий Стьюдента. Используется для оценки сходства или различия между выборочными совокупностями по разности величин их средних арифметических ($d = M_{\text{большая}} - M_{\text{меньшая}}$) и ее отношения к ошибке этой разности (m_d) при условии распределения вариант в группах по закону нормального распределения и подтверждается равенство разброса вариант в выборке (близкие дисперсии сравниваемых выборок). Не допускается применения критерия в случае балльного характера сравниваемых числовых признаков.

1.5. Статистические критерии различия

Наименьшая существенная разность (НСР). Используется в дисперсионном анализе. Показывает то минимальное различие между средними, начиная с которого при выбранном уровне вероятности средние сравниваемые показатели существенно отличаются друг от друга. Величина критерия выражается в тех же единицах, что и сравниваемые средние выборочных совокупностей и определяется по формуле:

$$\text{НСР} = t_{\text{табл}} \cdot m_d ,$$

где m_d – ошибка разницы средних; $t_{\text{табл}}$ – табличное значение критерия Стьюдента при уровне вероятности 0,95 или 0,99 и степени свободы, определяемой экспериментом.

Если разность между сравниваемыми средними в условиях эксперимента **больше или равна величине НСР** при P 0,95 или 0,99, то **различие существенно**.

1.5. Статистические критерии различия

Критерий Фишера. В выборочных совокупностях дисперсии могут существенно отличаться друг от друга. В таких случаях установление различий между выборочными совокупностями проводится по критерию Фишера (F – *положительное асимметричное распределение*). Расчет производится по формуле:

$$F = \sigma^2_{\text{большая}} / \sigma^2_{\text{меньшая}} \qquad (1.25)$$

Если величина расчетного критерия Фишера ($F_{\text{ф}}$) не превышает величины приведенного в таблице ($F_{\text{т}}$) (прил. 5 учебника), то различие между сравниваемыми дисперсиями считается **недостоверным**. При $F_{\text{ф}} > F_{\text{т}}$ эти дисперсии достоверно различны, как и сравниваемые по ним генеральные совокупности. Степень свободы рассчитывается для сравниваемых выборок отдельно по формуле $\nu = N - 1$.

1.5. Статистические критерии различия

Критерий Пирсона (хи-квадрат, χ^2). Для оценки соответствия или расхождения полученных эмпирических данных и теоретических (расчетных, прогнозных) распределений применяются статистические *критерии согласия*. Среди них наибольшее распространение получил непараметрический критерий К. Пирсона – хи-квадрат. Его можно использовать с различными формами распределения совокупностей. Как и любой другой статистический критерий, он не доказывает справедливость нулевой гипотезы, а лишь устанавливает с определенной вероятностью ее согласие или несогласие с экспериментальными данными. Критерий применяется при условии наличия не менее 5 наблюдений или частот в каждой группе, классе или совокупности. Малые частоты объединяют. Вычисление проводят по формуле:

$$\chi^2 = \sum [(\varphi - \varphi')^2 / \sum \varphi'],$$

где φ, φ'– наблюдения или частоты в опыте соответственно эмпирически или теоретически ожидаемые.

1.5. Статистические критерии различия

Значения χ^2 могут быть только положительными и возрастать от нуля до бесконечности. Если вычисленный критерий хи-квадрат больше табличного (теоретического) значения, нулевая гипотеза, которая предполагает соответствие эмпирического и теоретического распределений, отвергается, при $\chi^2_{выч} < \chi^2_{табл}$ нулевая гипотеза принимается.

Критерий Пирсона тем меньше, чем меньше различаются эмпирические и теоретические частоты. Он не позволяет обнаружить различия, которые скрадывает группировка (объединение малых частот в одну группу). Его удобно использовать, так как не требуется вычислений средних дисперсий.