Тема лекции 7. **Законы распределения - I.**

Характерные черты варьирования. В любом более или менее симметричном вариационном ряду заметна одна характерная особенность - накапливание вариант в центральных классах и постепенное убывание их численности по мере удаления от центра ряда. Эта особенность варьирования количественных признаков встречается довольно часто. Так, например, среди взрослого населения чаще встречаются люди среднего роста, а индивиды очень высокого или очень низкого роста - значительно реже. Однако в массе одновозрастных индивидов людей выше и ниже среднего роста оказывается примерно одинаковое количество.

Пусть большая масса людей одного пола и возраста разделена на отдельные группы так, чтобы в каждую группу вошли индивиды приблизительно одинакового роста. Пусть затем от каждой группы выделено по одному представителю, которые построены в одну шеренгу по ранжиру - от самого низкорослого до самого высокорослого. Если в затылок им поставить членов группы, получается «живая диаграмма» распределения, более или менее симметричная. Отмеченная черта варьирования обнаруживается не только в распределении людей по росту, но и по многим другим признакам, в частности по размерам обуви.

Впервые на эту закономерность варьирования обратил внимание А. Кетле (1835), исследовавший распределение нескольких тысяч американских солдат по росту (длине тела). «Человеческий рост, - писал он,- изменяющийся, по-видимому, самым случайным образом, тем не менее подчиняется самым точным законам; и эта особенность свойственна не только росту; она проявляется также и в весе, силе, быстроте передвижений человека, во всех его физических... и нравственных способностях. Этот великий принцип... разнообразящий проявление человеческих способностей... кажется нам одним из самых удивительных законов мира».

Описанная закономерность относится не только к человеку. Выше рассмотрены варьирование глазков в клубнях картофеля, распределение кальция в сыворотке крови обезьян и т.д. Подобных примеров можно привести много. Особенно примечательно, что не только распределение живых существ и продуктов их жизнедеятельности, но и случайные ошибки измерений подчиняются этой закономерности. «Не удивительно ли, - писал А. Кетле, - что случайные ошибки располагаются в таком совершенном порядке, и наши бессознательные промахи проявляются с такой симметрией, которая, кажется, могла бы быть результатом тщательно обдуманных расчетов».

Таким образом, прослеживается широко распространенная в природе закономерность: в массе относительно однородных единиц, составляющих статистическую совокупность, большинство членов оказывается среднего или близкого к нему размера, и чем дальше они отстоят от среднего уровня варьирующего признака, тем реже встречаются в данной совокупности. И это независимо от формы распределения, что указывает на определенную связь между числовыми значениями варьирующих признаков и частотой их встречаемости в данной совокупности.

Наглядным выражением этой связи и служат *вариационный ряд* и его линейный график - *вариационная кривая*. Эту закономерность можно воссоздать априори в виде математической модели, не опасаясь впасть в противоречие с фактами.

Случайные события. Подброшенный камень падает вниз, брошенный в воду - тонет. По длине одной из сторон куба можно точно определить его объем. На языке теории вероятностей всякий результат, или исход, однократного испытания называется *событием*. Под испытанием, которое может повторяться бесконечно большое число раз, подразумевают комплекс условий, необходимых для того, чтобы тот или иной исход мог осуществиться. У каждого из отмеченных здесь событий есть только один исход, заранее предсказуемый. Такие события называются *достоверными*. Если же при осуществлении комплекса условий события заведомо произойти не могут, они называются *невозможными*.

Существуют события, исход которых заранее непредсказуем. Если подбросить монету, то заранее нельзя сказать, как она упадет - вверх гербом или решкой. Здесь исход испытания зависит от случая. События, исход которых при осуществлении комплекса условий точно предсказать нельзя, называют *случайными.*

Случайные события (обозначаются начальными прописными буквами латинского алфавита *A, B, C,...)* называются *несовместимыми* или *несовместными,* если в серии испытаний всякий раз возможно осуществление только одного из них. Например, при метании монеты она может упасть вверх гербом или решкой. Здесь два разновозможных и несовместных исхода. События, которые в данных условиях могут произойти одновременно, называются *совместными.*

Вероятность события и ее свойства. Случайное событие можно предсказать лишь с некоторой уверенностью или вероятностью, которую данное событие имеет. При этом *вероятность* рассматривают как числовую меру объективной возможности осуществления события A при единичном испытании и обозначают символом $P(A)$. Согласно классическому определению (лек.1, формула 1), вероятность события A выражается отношением числа благоприятствующих осуществлению этого события исходов m к числу всех равновозможных и несовместных исходов n.

Из «классического» определения вероятности следует, что она представляет *число, заключенное между нулем и единицей*, т. е. выражается в долях единицы (может быть выражена в процентах от общего числа испытаний). Очевидно, вероятность достоверного события *А* равна единице, а вероятность невозможного события *Л* равна нулю. Считают, что события, имеющие очень малую вероятность, в единичных испытаниях не произойдут, т. е. такие события рассматривают как *практически невозможные.* Если же вероятность события достаточно велика, его принято считать *практически достоверным.* Этот принцип практической уверенности в прогнозировании исходов случайных событий позволяет использовать теорию вероятностей в практических целях.

Закон больших чисел. Многочисленные опыты и наблюдения показали, что частоты ожидаемых случайных событий приближаются к их вероятности по мере увеличения числа испытаний *n*. Так, если одну и ту же монету подбрасывать большое число раз, то невозможно ожидать, чтобы во всех без исключения случаях выпадал только герб или только решка. Ясно, что в каком-то числе случаев выпадет герб, а в других случаях - решка. Примечательно, что чем больше число испытаний, тем ближе к единице оказывается отношение выпавших гербов и решек, а частота каждого события становится ближе к его вероятности.

Закон больших чисел, теоретическое обоснование которому было дано Я. Бернулли (1713). Этот закон утверждает, что частость m/n события A будет сколь угодно близкой к его вероятности p, если число испытаний неограниченно возрастает. Как было показано выше, частость события и его вероятность не совпадают. Разница между ними уменьшается при увеличении числа испытаний. Закон больших чисел имеет объективный характер: их действие не зависит от сознания и воли людей.

А. Кетле и другие статистики собрали большой материал, убедительно свидетельствующий о наличии внутренней связи между случайностью и закономерностью, существующей в сфере массовых явлений.

Типы вариационных рядов. Вариационные ряды подразделяют на эмпирические и теоретические. Эмпирическими рядами будут такие, которые составлены на основании конкретных данных, полученных из выборочной совокупности.

Число членов в эмпирических рядах может быть малым или большим, но не достигающим бесконечно большого числа. В эмпирических вариационных рядах распределение членов выборки по классам выражается частотами.

Теоретические вариационные ряды отражают закономерности распределения членов совокупности по классам варьирующего признака при бесконечно большом числе наблюдений (т.е. $n \to \infty$).

Теоретический вариационный ряд служит пределом, к которому стремится эмпирический ряд при увеличении объема выборки до бесконечности. В теоретических рядах встречаемость членов совокупности, отклоняющихся на определенную величину от средней арифметической ряда, выражается вероятностью (P), а не частотой, которая применима к эмпирическому ряду. Следовательно, вероятность служит мерой возможности появления какого-либо события. Вариационные ряды любого типа могут быть изображены графически.

Случайные величины. Как было показано выше, варьирующие признаки в математике рассматривают как переменные случайные величины, способные в одних и тех же условиях испытания принимать различные числовые значения, которые заранее невозможно предсказать. Случайные величины делят на дискретные и непрерывные. Случайная величина называется *дискретной*, если она может принимать только определенные фиксированные значения, которые обычно выражаются целыми числами.

Если же случайная величина способна принимать любые числовые значения, она называется *непрерывной*. Очевидно, что счетные признаки относятся к дискретным случайным величинам, тогда как признаки мерные, варьирующие непрерывно, являются величинами непрерывными.

Случайная величина x в серии независимых повторных испытаний может принимать самые различные значения, но в каждом отдельном испытании она принимает единственное из возможных значений x_i.

Закон распределения случайных величин. Функция *f{x),* связывающая значения x_i переменной случайной величины x с их вероятностями p_i, называется *законом распределения этой величины.* Закон распределения случайной величины можно задать таблично, выразить графически в виде кривой вероятности и описать соответствующей формулой.

Закон распределения дискретной случайной величины может, например, выражаться в виде биномиальной кривой и описываться формулой Бернулли, которая позволяет находить вероятные значения этой величины в серии независимых испытаний.

В отношении же непрерывной случайной величины речь может идти лишь о тех значениях, которые она способна принять с той или иной вероятностью в интервале от и до. Этот интервал может быть каким угодно: и большим, и малым.

Выдающиеся математики - А. Муавр (1733), И. Г. Ламберт (1765), П. Лаплас (1795) и К. Гаусс (1821) - установили, что очень часто вероятность P любого значения x_i непрерывно распределяющейся случайной величины x находится в интервале от x до $x+dx$ и выражается формулой

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}dx \qquad (30)$$

где dx - малая величина, определяющая ширину интервалами е - математические константы (π - отношение длины окружности к ее диаметру, равное 3,1416; е = 2,7183 - основание натуральных логарифмов);

σ — стандартное отклонение, характеризующее степень рассеяния значений x_i случайной величины X вокруг средней μ, называемой *математическим ожиданием.*

В показатель степени числа е входит *нормированное отклонение* $t=(x_i-\mu)/\sigma$ - величина, играющая важную роль в исследовании свойств нормального распределения, описываемого вышеприведенной формулой.

Как видно из этой формулы, закон нормального распределения (нормальный закон) выражает функциональную зависимость между вероятностью $P(X)$ и нормированным отклонением t. Он утверждает, что вероятность отклонения любой варианты x_i от центра распределения μ, где $x_i- \mu=0$, определяется функцией нормированного отклонения t. Графически эта функция выражается в виде кривой вероятности, называемой *нормальной кривой*. Форма и положение этой кривой определяются только двумя параметрами: μ и σ. При изменении величины μ форма нормальной кривой не меняется, лишь график ее смещается вправо или влево.

Изменение же величины σ влечет за собой изменение только ширины кривой: при уменьшении σ кривая делается более узкой за счет меньшего рассеяния вариант вокруг средней, а при увеличении σ кривая расширяется. Во всех случаях, однако, нормальная кривая остается строго симметричной относительно центра распределения, сохраняя правильную колоколообразную форму (рис.7.1, 7.2).

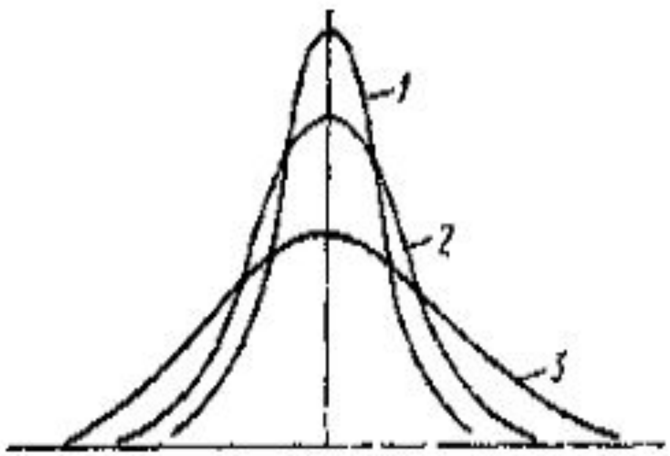


Рисунок 7.1 Нормальные кривые (1,2,3 при разных значениях параметра σ)

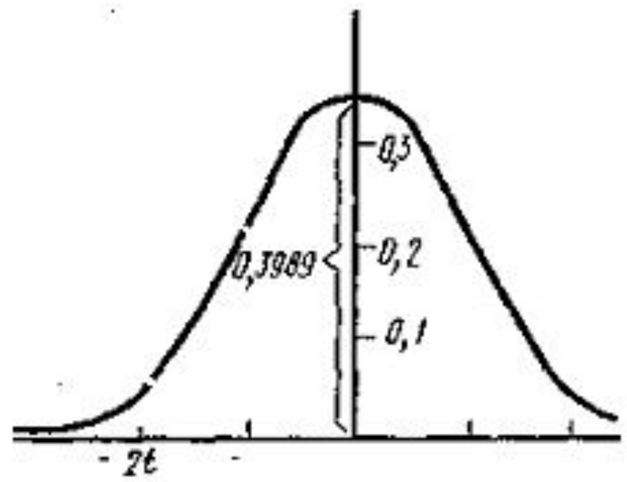


Рисунок 7.2 Стандартизованная форма нормальной кривой (при σ=1)

Нормальная кривая с параметрами μ = 0 и σ=1 называется *нормальной* или *стандартизованной кривой*. Она описывается формулой. Стандартная кривая имеет площадь, равную единице.

Типы распределения. Существуют следующие типы вариационных рядов и соответственно типы вариационных кривых: нормальные, биноминальные; асимметричные; эксцессивные; пуассоновы; трансгрессивные.

Нормальное распределение. В большинстве распределений, с которыми приходится встречаться биологу, экологу, медицинскому работнику, проявляется определенная закономерность: крайние значения - наименьшее и наибольшее - появляются редко; чем ближе значение признака к средней арифметической, тем оно чаще встречается; в центре распределения имеются такие значения, которые встречаются наиболее часто и образуют в вариационном ряду модальный класс.

Такое распределение значений признака так часто встречается в самых различных областях науки и практики, что первоначально принималось за норму всякого массового случайного проявления признаков и в соответствии с этим получило особое название: нормальное распределение.

В настоящее время *нормальным* называется такое распределение, которое следует закону, выраженному определенной математической формулой, предложенной Гауссом и независимо от него Лапласом.

Нормальное распределение имеет большое значение для решения конкретных задач, кроме этого, оно играет существенную роль в теории вариационной статистики. Закономерности нормального распределения выражены уравнением Гаусса – Лапласа:

$$y_x = \frac{n}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\overline{X})^2}{2 \cdot \sigma^2}} \qquad (31)$$

где y_x – искомая ордината нормальной кривой, показывающая теоретическую частоту для конкретной величины данного варьирующего признака;

$\overline{X}$ - средняя арифметическая данного вариационного ряда;

n – объем выборки;

σ – среднее квадратическое отклонение;

e – основание натуральных логарифмов, равное 2,71828;

π – постоянное число пи, равное 3,1416; *x* – варианта.

Пользуясь уравнением нормального распределения, можно определить теоретические частоты для любого эмпирического ряда, так как величина у является функцией переменной величины х.

В большинстве работ при характеристике фактических распределений и определении достоверности выборочных величин предполагается, что имеется нормальное распределение значений признака. Поэтому необходимо бывает проверить это допущение, т. е. установить, достаточно ли изучаемое распределение соответствует нормальному, можно ли считать, что отступления фактических частот от таких, которые точно следуют закону нормального распределения, произошли от случайных причин.

Для этой цели необходимо:

1. Рассчитать теоретический нормальный ряд на основе показателей изучаемого эмпирического ряда, использовав свойства первой функции нормированного отклонения.

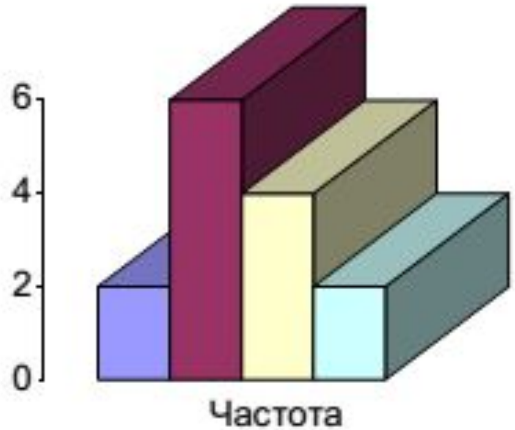
2. Установить достоверность расхождения между эмпирическим и теоретическим рядами, применив один из критериев расхождения распределений.

Свойства нормального распределения:

1. Нормальное распределение имеет строгую закономерность между величиной ординат (y), отрезками оси абсцисс (x), выраженными через нормированное отклонение (t), и площадью кривой, ограниченной её ветвями и осью абсцисс.

2. Нормированное отклонение (t) имеет три функции, позволяющие определять ряд показателей нормального распределения. Первую функцию f(t) используют для установления теоретических частот ряда. Вторая функция φ(t) позволяет определять долю площадей или несколько долей площади нормальной кривой, что соответствует числу (или %) членов совокупности, входящих в границу площади, отсеченной ординатами. Третья функция F(t) предназначена для определения величины частных средних арифметических той или иной отсеченной доли площади нормальной кривой.

Таблица 7.1
Нормальное распределение

График	Общая характеристика
	Нормальная кривая имеет колокообразную форму, симметрична по отношению к перпендикуляру. Закономерность распределения такова, что 99,7% всех членов совокупности находятся в границах $\overline{X} \pm 3\,\sigma$. Характеризуется с помощью $\overline{X}$ и σ. Чем меньше σ, тем выше кривая и тем меньше ее основание и, наоборот.

3. Закономерности нормального распределения и три функции нормированного отклонения дают возможность решать многие вопросы планирования и прогнозирования основных показателей изменчивости и среднего арифметического значения варьирующего признака в различных популяциях.

Литература:

Основная – 1 [24-66]; 2 [т.1-82-99]; 3 [163-183].

Дополнительная – 2 [34-80]; 3 [28-46]; 5 [142-184].

Контрольные вопросы:

1. Описать характерные черты варьирования.
2. Случайные события и случайные величины.
3. Вероятность события и её свойства.
4. Закон больших чисел.
5. О чем свидетельствует закон распределения случайных величин?
6. Нормальное распределение: основная характеристика и особенность.