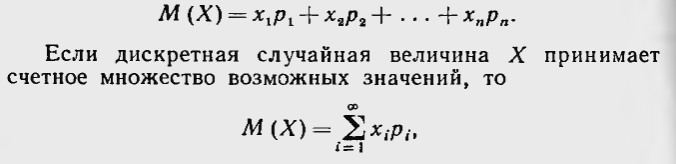
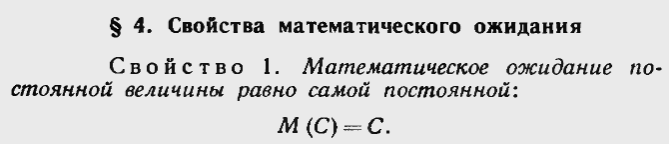
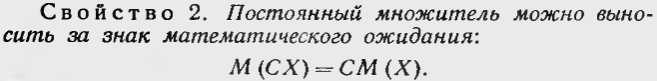
# Математическое ожидание

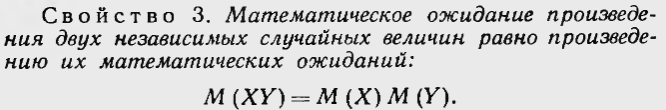
Мат. ожидание приближенно *равно среднему значению случайной величины*. Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности.

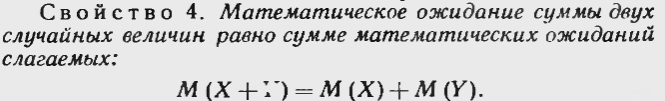


Вероятностный смысл полученного результата таков: математическое ожидание приближенно равно (тем точнее, чем больше число испытаний) среднему арифметическому наблюдаемых значений случайной величины.

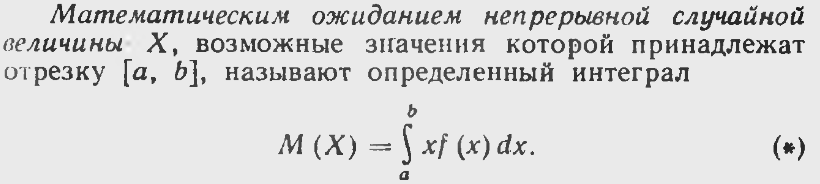




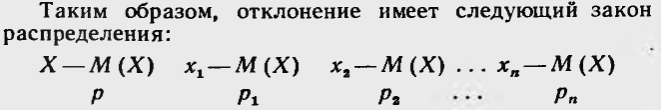


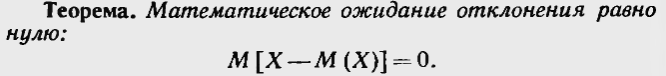


**Теорема**. Математическое ожидание М (X) числа появлений события А в п независимых испытаниях равно произведению числа испытаний на вероятность появления события в каждом испытании: М (X) = nр.



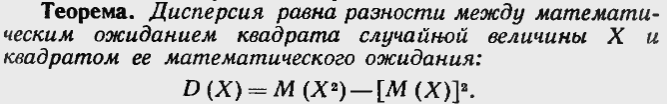
# Дисперсия

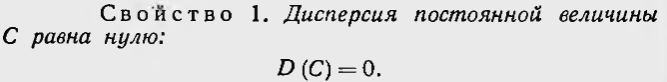


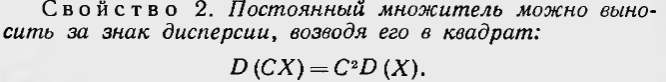


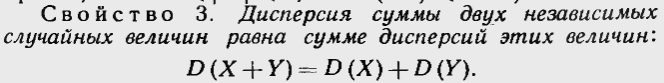
Дисперсией (рассеянием) дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

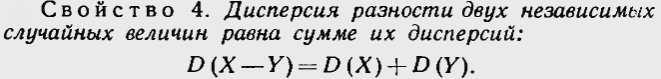




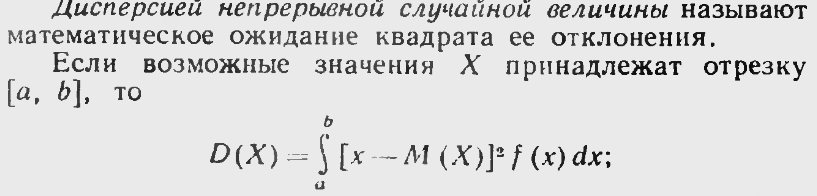






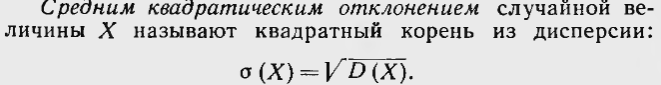


**Теорема**. Дисперсия числа появлений события А в n независимых испытаниях, в каждом из которых вероятность р появления события постоянна, равна произведению числа испытаний на вероятности появления и непоявления со бытия в одном испытании: D (X) = *npq*.

****

# Среднее квадратическое отклонение

Для оценки рассеяния возможных значений случайной величины вокруг ее среднего значения кроме дисперсии.

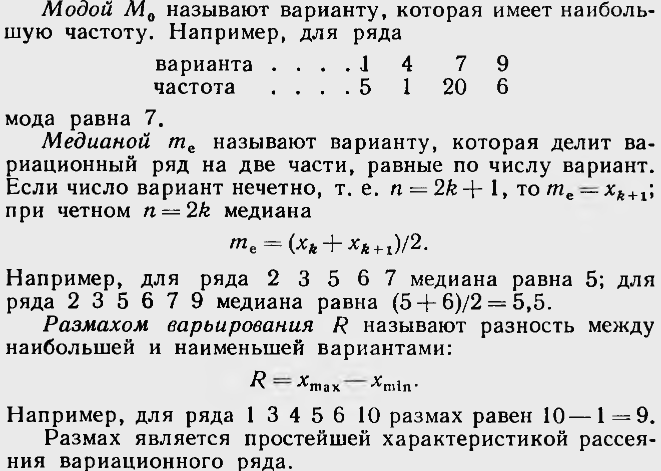


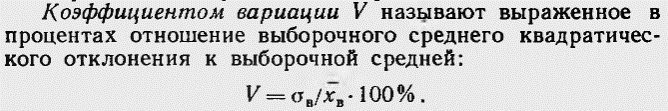
В тех случаях, когда желательно, чтобы **оценка рассеяния имела размерность случайной величины**, вычисляют среднее квадратическое отклонение, а не дисперсию. Например, если X выражается в линейных метрах, то sigma( Х ) будет выражаться также в линейных метрах, a D (X) — в квадратных метрах

**Теорема**. Среднее квадратическое отклонение суммы конечного числа взаимно независимых случайных величин равно квадратному корню из суммы квадратов средних квадратических отклонений этих величин:



# Мода, медиана, размах

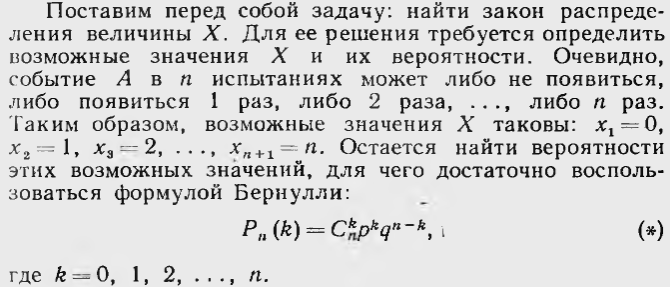




Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого

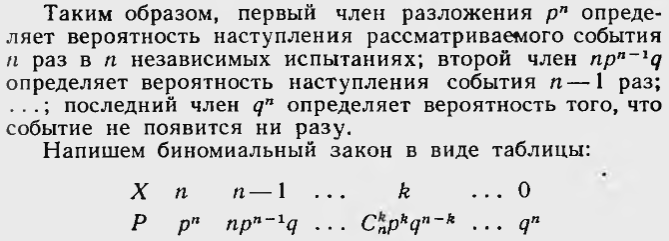
коэффициент вариации больше. Коэффициент вариации — безразмерная величина, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность, например если варианты одного ряда выражены в сантиметрах, а другого — в граммах.

# Биномиальное распределение



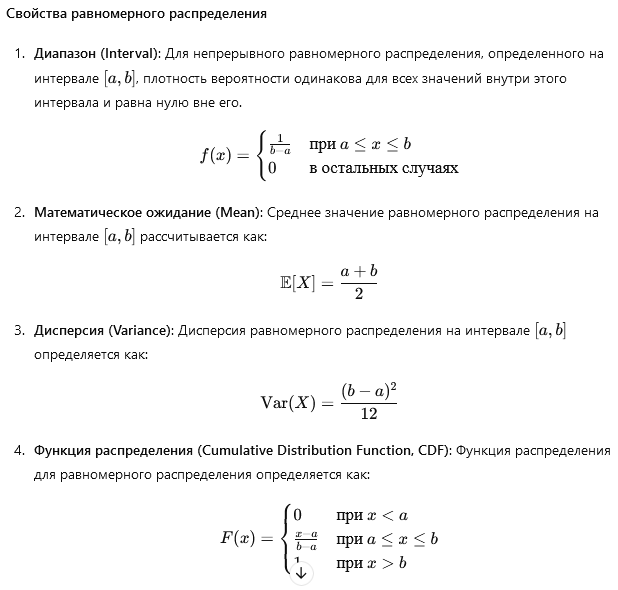
Биномиальным называют распределение вероятностей, определяемое формулой Бернулли. Закон назван «биномиальным» потому, что правую часть равенства (\*) можно рассматривать как общий член разложения бинома Ньютона:



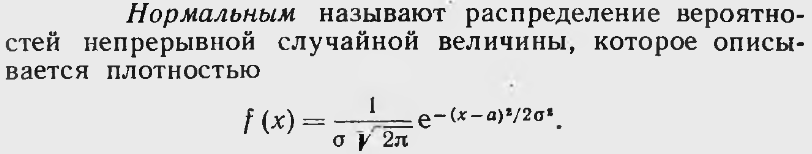


# Равномерное распределение

Распределение вероятностей называют равномерным, если на интервале, которому принадлежат все возможные значения случайной величины, плотность распределения сохраняет постоянное значение.

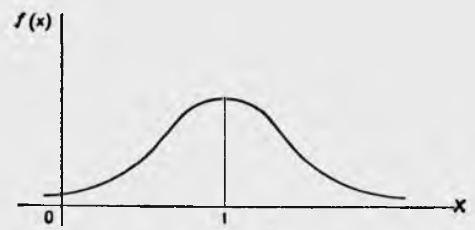


# Нормальное распределение



Мы видим, что нормальное распределение определяется двумя параметрами: ***а***и 𝞂. Достаточно знать эти параметры, чтобы задать нормальное распределение. Вероятностный смысл этих параметров таков: ***а*** есть математическое ожидание, 𝞂 — среднее квадратическое отклонение нормального распределения.

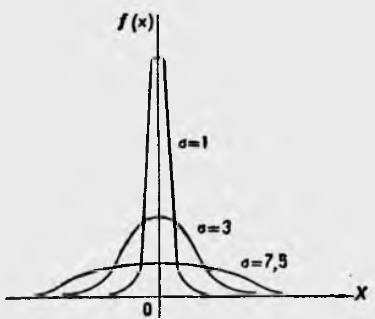
**График** плотности нормального распределения называют нормальной кривой (кривой Гаусса).



Влияние параметров нормального распределения на форму нормальной кривой

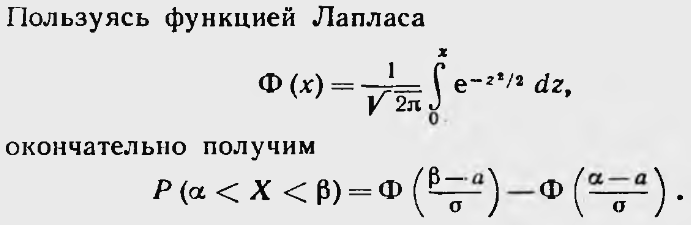
Изменение величины параметра а (математического ожидания) не изменяет формы нормальной кривой, а приводит лишь к ее сдвигу вдоль оси Ох: вправо, если а возрастает, и влево, если а убывает.

С возрастанием о максимальная ордината нормальной кривой убывает, а сама кривая становится более пологой, т. е. сжимается к оси Ох\ при убывании о нормальная кривая становится более «.островершинной-» и растягивается в положительном направлении оси Оу



При любых значениях параметров ***а*** и 𝞂 площадь, ограниченная нормальной кривой и осью х, остается равной единице.

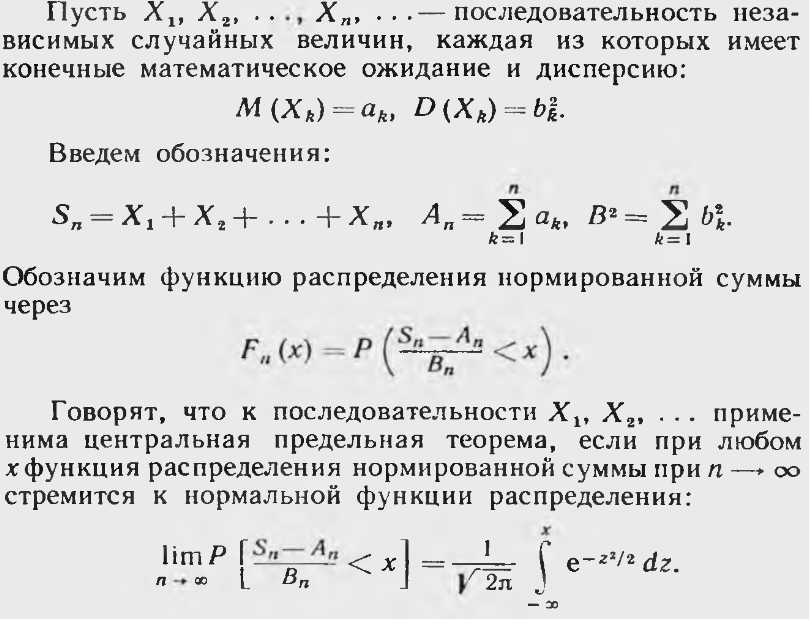
**Вероятность попадания в заданный интервал нормальной случайной величины**



# Центральная предельная теорема

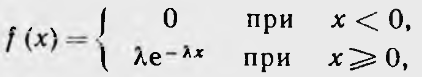
Известно, что нормально распределенные случайные величины широко распространены на практике. Чем это объясняется? Ответ на этот вопрос был дан выдающимся русским математиком А. М. Ляпуновым (**центральная предельная теорема**): если случайная величина X представляет собой *сумму* очень большого числа *взаимно независимых случайных величин*, *влияние* каждой из которых на всю сумму *ничтожно* *мало*(ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), то X имеет *распределение*, *близкое к нормальному*.

Условие применимости ЦПТ (Условие Ляпунова) Требовании, чтобы каждое слагаемое суммы оказывало на сумму ничтожное влияние.



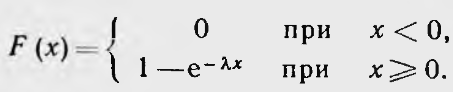
# Показательное распределение

Показательным (экспоненциальным) называют распределение вероятностей непрерывной случайной величины X, которое описывается **плотностью**:

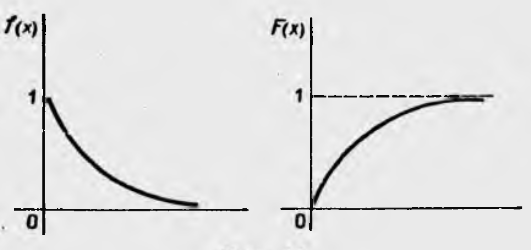


где λ — *постоянная положительная величина.*

**Функция распределения***:*



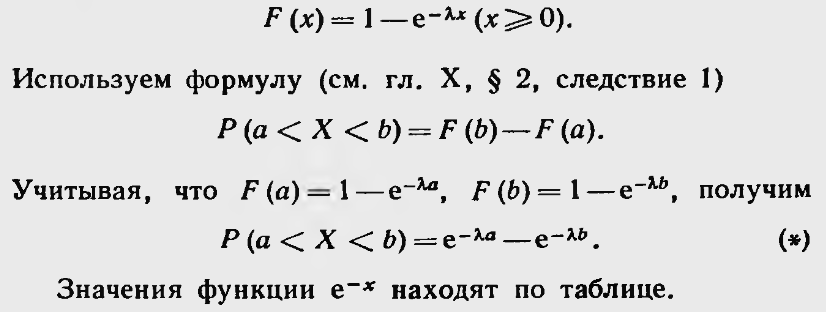
Показательное распределение определяется одним параметром λ.



**Преимуществo:** Обычно параметры неизвестны и приходится находить их оценки (приближенные значения); проще оценить один параметр, чем два или три и т. д.

Примером непрерывной случайной величины, распределенной по показательному закону, может служить время между появлениями двух последовательных событий простейшего потока

**Вероятность попадания в заданный интервал**



**Числовые характеристики**

*Математическое ожидание* показательного распределения равно обратной величине параметра λ: 

*Дисперсия*: D(X) = 2/ λ2

Среднее квадратическое отклонение: 

* математическое ожидание и среднее квадратическое отклонение показательного распределения равны между собой.

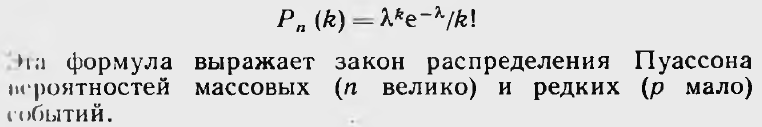
распределений и их свойства, какие значения принимают, параметры, какие параметрам можно задавать значения, функции распределения и плотности вероятности

# Распределение Пуассона

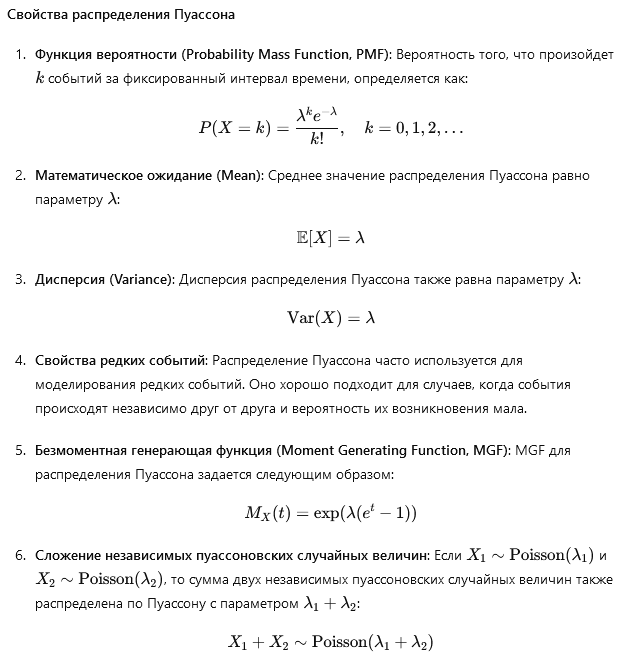
Пусть производится п независимых испытаний, в каждом из которых вероятность появления события А равна р. Для определения вероятности k появлений события в этих испытаниях используют формулу Бернулли.

Если же n велико, то пользуются асимптотической формулой Лапласа. Однако эта формула непригодна, если вероятность события мала (р <= 0,1). В этих случаях (n велико, р мало) прибегают к асимптотической формуле Пуассона.

Найти вероятность того, что при *очень большом числе испытаний*, в каждом из которых *вероятность события очень мала*, ***событие наступит ровно k*** *раз*. Сделаем важное допущение: произведение nр сохраняет постоянное значение, а именно nр = **λ (**математическое ожидание случайной величины**)**. Как будет следовать из дальнейшего, это означает, что ***среднее число появлений события* *в различных сериях испытаний***, т. е. при различных значениях n, остается неизменным.



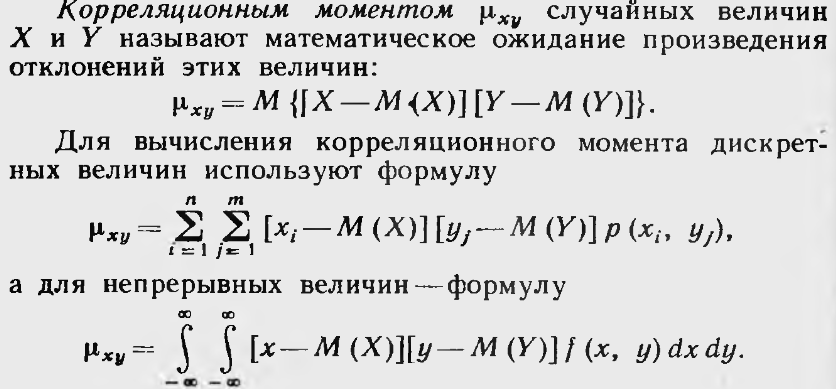
|  |  |
| --- | --- |
| Функция вероятности | Функция распределения |
| Функция вероятности | Функция распределения |



C увеличением λ λ распределение Пуассона стремится к распределению Гаусса(нормальному) со среднеквадратичным отклонением σ = λ   
и сдвигом λ λ.

# Коэффициент корреляции

Корреляционный момент

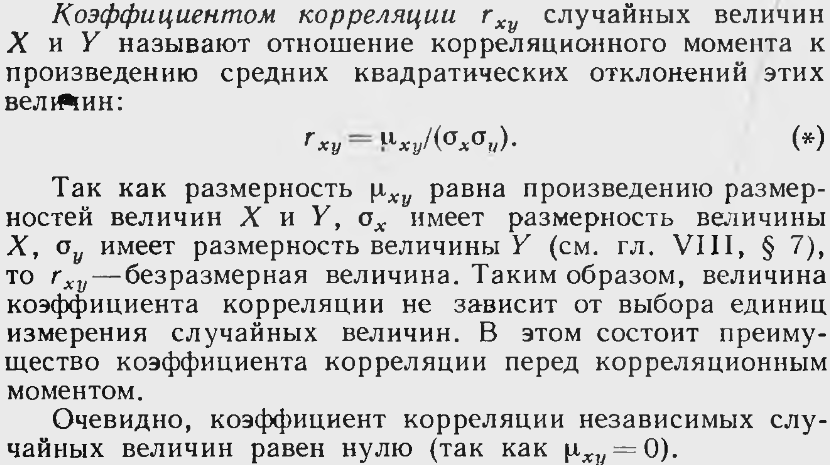


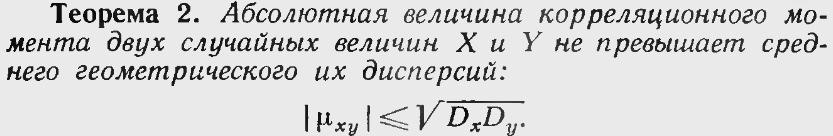
Корреляционный момент служит для характеристики связи между величинами X и Y. Корреляционный момент *равен нулю, если X и Y независимы*; следовательно, если корреляционный момент

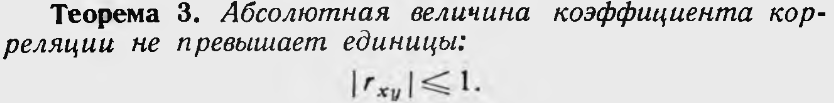
*не равен нулю, то X и Y — зависимые* случайные величины.

Величина корреляционного момента зависит от единиц измерения случайных величин.

Коэффициент корреляции

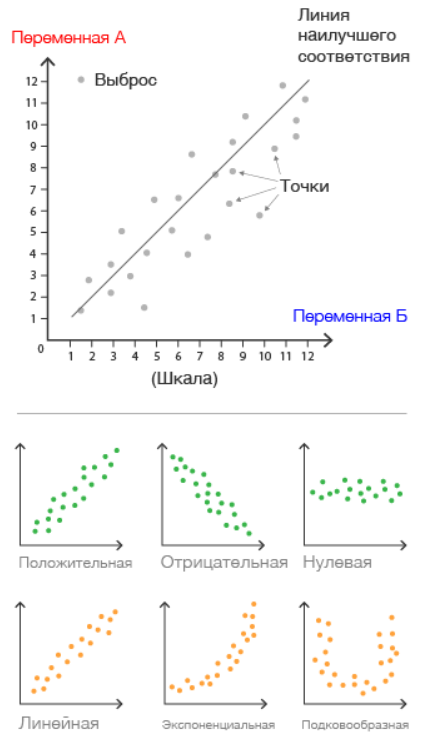




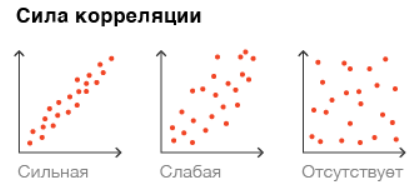
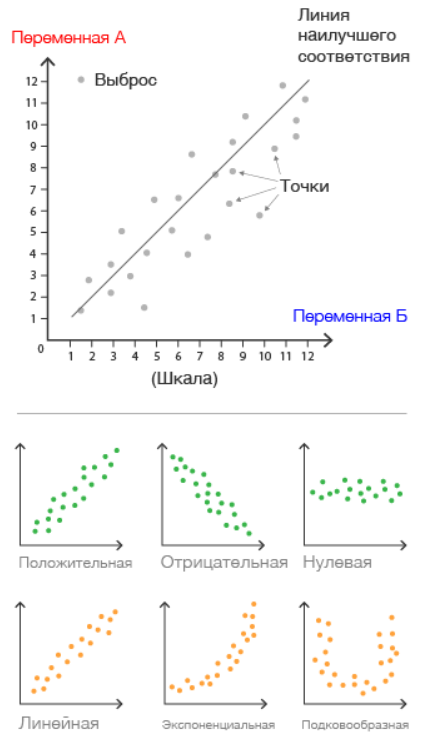


# Диаграмма рассеивания

На диаграммах рассеяния ряд точек, размещенных в декартовой системе координат, отображает значения по двум переменным. Присвоив каждой оси переменную, можно определить, существуют ли отношения или корреляция между этими двумя переменными.



Отображаемые на диаграммах рассеяния паттерны позволяют увидеть разные типы корреляции. Среди них: **положительная** (оба значения увеличиваются), **отрицательная** (одно значение увеличивается, в то время как второе уменьшается), **нулевая** (отсутствие корреляции), **линейная**, **экспоненциальная** и **подковообразная**. Сила корреляции определяется по тому, насколько близко расположены друг от друга точки на графике. Точки, которые значительно удалены от общего кластера точек, называются **выбросами**.



На графике могут использоваться линии или кривые, которые помогают при анализе и проводятся максимально близко ко всем точкам, чтобы продемонстрировать, как бы все эти точки могли потенциально выстроиться в одну линию. Этот элемент известен под названием **«линия наилучшего соответствия»** или **«линия общего направления»** и может использоваться для оценки через интерполяцию.

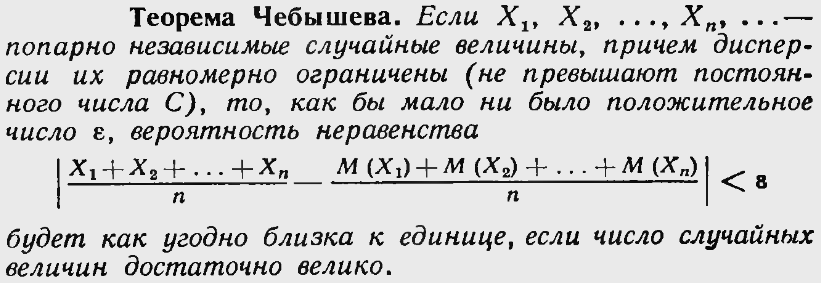
Диаграммы рассеяния идеально подходят в том случае, если у вас есть пара числовых данных, и вы хотите посмотреть, существует ли влияние одной переменной на другую. Однако не забывайте, что корреляция – это не причинная зависимость, поэтому на результаты может повлиять не принятая во внимание переменная.

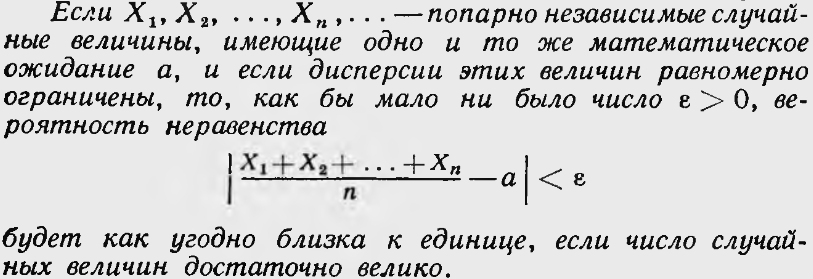
# Закон больших чисел

При некоторых сравнительно широких условиях суммарное поведение достаточно большого числа случайных величин почти утрачивает случайный характер и становится закономерным.

Неравенство Чебышева: Вероятность того, что отклонение случайной величины X от ее математического ожидания по абсолютной величине меньше положительного числа ε, не меньше, чем







Суть: хотя отдельные независимые случайные величины могут принимать значения, далекие от своих математических ожиданий, среднее арифметическое достаточно большого числа случайных величин с большой вероятностью принимает значения, близкие к определенному постоянному числу, а именно к числу (М + М (Х 2) + . . . + М (Хn))/n (или к числу *а* в частном случае).

Иными словами, отдельные случайные величины могут иметь значительный разброс, а их среднее арифметическое рассеянно мало.

*Среднее арифметическое достаточно большого числа независимых случайных величин (*дисперсии которых равномерно ограничены*) утрачивает характер случайной величины.*

**Теорема Бернулли**. Если в каждом из n независимых испытаний вероятность р появления события А постоянна, то, как угодно, близка к единице вероятность того, что отклонение относительной частоты от вероятности р по абсолютной величине будет сколь угодно малым, если

число испытаний достаточно велико. (n большое (P постоянно) -> отклонение маленькое)



# Простая случайная выборка

**Простая случайная выборка** — это вероятностный метод, используемый для отбора выборки для исследования. Основная характеристика простой случайной выборки состоит в том, что она дает *каждому элементу статистической совокупности одинаковую вероятность* *включения* в изучаемую выборку.

**Выборочной совокупностью** или просто выборкой называют совокупность случайно отобранных объектов. (Часть объектов)

**Генеральной совокупностью** называют совокупность объектов, из которых производится выборка. (Все объекты)

**Объемом** совокупности называют число объектов этой совокупности.

# Точность оценки

**Точечная оценка** – определяется ***одним числом***.

При *выборке малого объема* точечная оценка может значительно отличаться от оцениваемого параметра (*приводить к грубым ошибкам*). По этой причине при небольшом объеме выборки следует пользоваться интервальными оценками.

**Интервальная оценка** - определяется ***двумя числами*** — концами интервала. Интервальные оценки позволяют *установить точность и надежность* оценок.

**Точность оценки.** Пусть *β\* - статистическая оценка* (по данным выборки) *неизвестного* *β*. β\* тем точнее определяет параметр β, чем меньше абсолютная величина разности | β — β\* |.

Другим и словами, если δ > 0 и ***| β — β\*| < δ***, то ***чем меньше δ, тем оценка точнее.*** Таким образом, положительное ***число δ характеризует точность оценки***.

Однако статистические методы не позволяют категорически утверждать, что оценка β\* удовлетворяет неравенству *| β — β\*| < δ*; можно лишь говорить о вероятности *у*, с которой это неравенство осуществляется.

**Надежностью (доверительной вероятностью)** оценки *β* по *β*\* называют вероятность ***у***, с которой осуществляется неравенство ***| β — β\*| < δ***.

Обычно надежность оценки задается наперед, причем в качестве ***у*** берут

число, близкое к единице. Наиболее часто 0,95; 0,99 и 0,999.

**Доверительным** **интервалом** называют ***интервал (β \* — δ, β \* + δ),***

который покрывает неизвестный параметр с заданной надежностью ***y***.

# Ошибки 1 и 2 рода

**Статистическая гипотеза** – гипотеза ***о виде***неизвестногораспределения, ***или о параметрах*** известных ***распределений***.

Например:

1) генеральная совокупность распределена по закону Пуассона;

2) дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй— о параметрах двух известных распределений.

Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза.

Нулевой (*основной*) называют выдвинутую гипотезу Н0.

Конкурирующей (*альтернативной*) называют гипотезу Н1, которая противоречит нулевой.

*Простая* гипотеза, содержит только *одно предположение*.

*Сложная* гипотеза, состоит из конечного или бесконечного числа простых гипотез.

**Ошибка первого рода** - ***отвергнута правильная гипотеза***.

**Ошибка второго рода** - ***принята неправильная гипотеза***.

**Статистическим критерием** называют случайную величину ***К***, которая служит для проверки нулевой гипотезы.

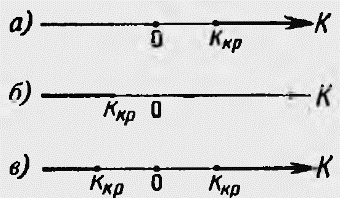
**Критической областью** называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

**Областью принятия гипотезы** называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез можно сформулировать так: если *наблюдаемое* *значение критерия принадлежит критической области — гипотезу отвергают*, если *наблюдаемое значение критерия принадлежит области принятия гипотезы — гипотезу принимают*.

Поскольку критерий К — одномерная случайная величина, все ее возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также являются интервалами и, следовательно, существуют точки, которые их разделяют.

*Критическими точками* (границами) kкр называют точки, отделяющие критическую область от области принятия гипотезы.



***Односторонняя критическая область:***

*Правостороння -* определяется неравенством К > kкр, где kкр — положительное число (а).

*Левосторонняя -* определяется неравенством K < kкр, где kкр — отрицательное число (б).

***Двусторонняя критическая область***, определяется неравенствами   
k1 < K < k2 , где k2 > k1 (в).

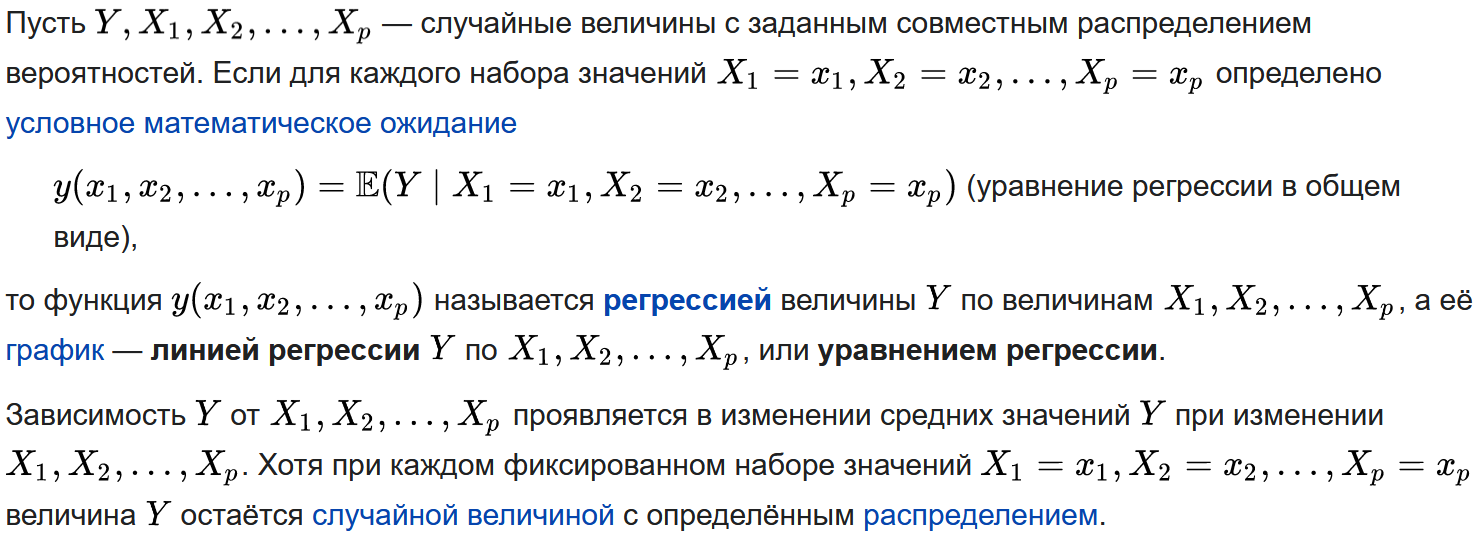
# Регрессионный анализ

***Регрессионный анализ* —** набор статистических методов *исследования влияния* одной или нескольких *независимых переменных* X1, X2, .. Xn *X 1 , X 2 , . . . , X p XX1на зависимую* переменную YYY.

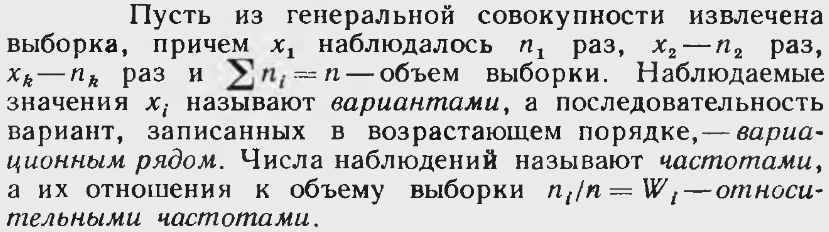
Наиболее распространённый вид регрессионного анализа — ***линейная регрессия***, когда находят *линейную функцию*, которая, согласно определённым математическим критериям, *наиболее соответствует данным*.

Цели:

* Предсказание значения зависимой переменной с помощью независимой.
* Определение вклада отдельных независимых переменных в вариацию зависимой.



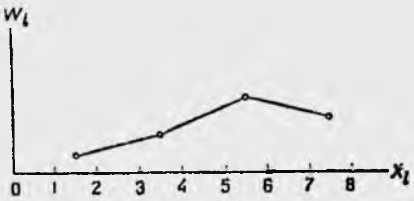
# Гистограмма



Для наглядности строят различные графики статистического распределения (полигон и гистограмму).

**Полигоном** **частот** называют ломаную, отрезки которой соединяют точки (х1; n1), (х2; n2) … (xk; nk). На горизонтальной оси откладывают варианты хi, а на вертикальной — частоты ni.

**Полигоном относительных частот** называют ломаную, отрезки которой соединяют точки (х1; W1), (х2; W2), … (xk; Wk).



В случае непрерывного признака целесообразно строить гистограмму. *Интервал* значений *разбивают на несколько* частичных *интервалов длиной h* и *находят* для каждого *сумму частот ni*.

**Гистограммой частот** называют ступенчатую фигуру, состоящую из прямоугольников, *основаниями* которых служат *частичные интервалы длиною h*, а *высоты* равны отношению ni/h (*плотность частоты*).

Площадь i-го частичного прямоугольника равна ni — сумме частот вариант i-гo интервала; следовательно, площадь гистограммы частот равна сумме всех частот, т. е. объему выборки.

