Оглавление

[*Случайные события* 4](#_Toc516102191)

[Несовместные, достоверные, невозможные события. Пространство элементарных событий. 4](#_Toc516102192)

[Сумма, произведение событий. Противоположные события 4](#_Toc516102193)

[Классическое определение вероятности 4](#_Toc516102194)

[Геометрическая вероятность 4](#_Toc516102195)

[Теорема сложения: 4](#_Toc516102196)

[Условная вероятность. Теорема умножения. 4](#_Toc516102197)

[Независимое событие 5](#_Toc516102198)

[Теорема о полной вероятности: 5](#_Toc516102199)

[Теорема Байеса: 5](#_Toc516102200)

[Повторные независимые испытания 5](#_Toc516102201)

[*Случайные величины* 5](#_Toc516102202)

[Дискретная (прерывная) случайная величина 5](#_Toc516102203)

[Непрерывная случайная величина. Плотность распределения. 6](#_Toc516102204)

[Функция распределения 6](#_Toc516102205)

[Математическое ожидание 6](#_Toc516102206)

[Дисперсия (рассеяние) 6](#_Toc516102207)

[Мода, медиана, квантиль. 7](#_Toc516102208)

[Биномиальное распределение 7](#_Toc516102209)

[Равномерное распределение 7](#_Toc516102210)

[Нормальное распределение 7](#_Toc516102211)

[Функция Лапласа 7](#_Toc516102212)

[Вычисление вероятностей для нормального распределения с помощью функции Лапласа: 8](#_Toc516102213)

[*Системы случайных величин* 8](#_Toc516102214)

[Дискретная двумерная СВ. Матрица распределения 8](#_Toc516102215)

[Критерий независимости компонент ДССВ 8](#_Toc516102216)

[Непрерывная двумерная СВ. Плотность распределения и её свойства. 8](#_Toc516102217)

[Функция распределения 9](#_Toc516102218)

[Распределения компонент двумерной НСВ 9](#_Toc516102219)

[Критерий независимости компонент НССВ: 9](#_Toc516102220)

[Ковариация 9](#_Toc516102221)

[Независимость, некоррелированность ССВ: 10](#_Toc516102222)

[Свойства математического ожидания: 10](#_Toc516102223)

[Свойства дисперсии: 10](#_Toc516102224)

[Свойства коэффициента корреляции 10](#_Toc516102225)

[Функция регрессии 10](#_Toc516102226)

[*Предельные теоремы теории вероятности* 11](#_Toc516102227)

[Асимптотически нормальная СВ. Центральная предельная теорема. 11](#_Toc516102228)

[Предельные теоремы Муавра-Лапласа как следствия ЦПТ: 11](#_Toc516102229)

[Закон больших чисел (теорема Чебышева): 11](#_Toc516102230)

[Закон больших чисел (теорема Бернулли) 12](#_Toc516102231)

[*Оценки и методы оценивания* 12](#_Toc516102232)

[Оценки основных параметров генеральной совокупности по выборки 12](#_Toc516102233)

[Оценки основных параметров генеральной совокупности по вариационному ряду. 12](#_Toc516102234)

[Оценка плотности и функции распределения. 12](#_Toc516102235)

[Распределении, используемые в МатСтате. 13](#_Toc516102236)

[Точечная оценка параметра, несмещённость, состоятельность, сравнение оценок по эффективности. 13](#_Toc516102237)

[Понятие интервальной оценки. 13](#_Toc516102238)

[*Статистическая проверка гипотез* 14](#_Toc516102239)

[Постановка задачи: 14](#_Toc516102240)

[Ошибки первого и второго рода. Форма критической области при проверке простых гипотез о числовых характеристиках НГС. 14](#_Toc516102241)

[Критерии согласия. Критерий Пирсона. 14](#_Toc516102242)

[Критерий однородности. Критерий знаков. 14](#_Toc516102243)

[Критерии однородности. Критерий вилкоксона 14](#_Toc516102244)

[*Методы Монте-Карло* 15](#_Toc516102245)

[Какие задачи решаются методом статистических испытаний? 15](#_Toc516102246)

[На каких теоремах основан метод вычисления числовых характеристик? 15](#_Toc516102247)

[Общая схема метода, роль доверительной вероятности. 15](#_Toc516102248)

[Вычисление интегралов методом Монте-Карло. Простейший метод на примере определенного интеграла. 15](#_Toc516102249)

[Геометрический метод на примере определенного интеграла. 15](#_Toc516102250)

[Сравнение методов по точности и по трудоемкости 15](#_Toc516102251)

[Базовая случайная величина. Метод середин квадратов, метод вычетов. 16](#_Toc516102252)

[*Дисперсионный анализ* 16](#_Toc516102253)

[Постановка задачи однофакторного ДА (классическая схема). Требования к экспериментальным данным. 16](#_Toc516102254)

[Основное дисперсионное тождество. Проверка гипотезы о дисперсиях, обоснование эквивалентности ее первоначальной задаче. 16](#_Toc516102255)

[Когда применяется непараметрический дисперсионный анализ? Требования к экспериментальным данным, критерий Краскела-Уоллеса 16](#_Toc516102256)

[*Корреляционный анализ* 17](#_Toc516102257)

[Какие задачи решает корреляционный анализ? 17](#_Toc516102258)

[Выборочный коэффициент корреляции, (условия применения, свойства). 17](#_Toc516102259)

[Корреляционное отношение, его свойства. 18](#_Toc516102260)

[Коэффициент корреляции Спирмена (условия применения, свойства). 18](#_Toc516102261)

[*Регрессионный анализ* 18](#_Toc516102262)

[Основные задачи регрессионного анализа 18](#_Toc516102263)

[Паpная линейная pегpессия 18](#_Toc516102264)

[Метод наименьших квадpатов 19](#_Toc516102265)

[Проверка адекватности модели (три способа). 19](#_Toc516102266)

[Множественная регрессия 19](#_Toc516102267)

[Условия применимости МНК: 19](#_Toc516102268)

[Свойства оценок коэффициентов: 20](#_Toc516102269)

[*Временные ряды* 20](#_Toc516102270)

[Основные задачи при исследовани ВР. Агрегирование и сглаживание данных. 20](#_Toc516102271)

[Компоненты модели ВР: 20](#_Toc516102272)

[Гипотезы о существовании тренда, о наличии периодической составляющей и о случайности остатков ВP (критерии Фостера-Стюарта, поворотных точек, серий, статистика Дарбина-Уотсона). 21](#_Toc516102273)

[Гетероскедастичность данных. Тест Гольдфельда-Квандта. 21](#_Toc516102274)

[Взвешенный МНК 22](#_Toc516102275)

[Коррелированность объясняющих переменных и случайных членов. Метод инструментальных переменных 23](#_Toc516102276)

[Автокорреляция в остатках. Авторегрессионное преобразование. 23](#_Toc516102277)

[*Цепи Маркова* 23](#_Toc516102278)

[Матрица перехода системы. Орграф однородной ЦМ. Свойство стохастичности матрицы перехода. Классификация состояний ЦМ. 23](#_Toc516102279)

[Многошаговый переход в ЦМ. Уравнение Колмогоpова-Чепмена. 24](#_Toc516102280)

[Теорема о вероятности перехода в эргодическое состояние. 24](#_Toc516102281)

[Поглощающие ЦМ. Канонический вид матрицы перехода. Теорема о свойствах подматрицы Q. 25](#_Toc516102282)

[Фундаментальная матрица. Теорема. 25](#_Toc516102283)

[Теорема о вероятности поглощения в данном поглощающем состоянии: 25](#_Toc516102284)

[Эргодические ЦМ. Регулярность. Стационарный режим ЦМ. Предельные вероятности. 25](#_Toc516102285)

**ТЕОРИЯ**

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ**

# *Случайные события*

## Несовместные, достоверные, невозможные события. Пространство элементарных событий.

**Несовместные события –** событие, при появлении которого исключается появление другого события в одном и том же испытании. (# бросание монеты)

**Достоверные события –** события, вероятность появления которых равна единице.

**Невозможные события –** события, вероятность появления которых равна нулю.

**Пространство элементарных событий –** множество всех различных исходов некоторого испытания

## Сумма, произведение событий. Противоположные события

**Сумма двух событий А и Б –** событие, состоящее в появлении события А, или события Б, или обоих этих событий.

**Противоположные события –** два единственно возможных события, образующих полную группу. (Полная группа – события, сумма вероятностей которых равна единице).

**Произведение двух событий А и Б –** событие АБ, состоящее в совместном появлении (совмещении) этих событий.

## Классическое определение вероятности

Число, характеризующее степень возможности появления события. Вероятность события А – отношение числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу.

## Геометрическая вероятность

вероятности попадания точки в некоторую область (отрезок, часть плоскости и т.д.)

## Теорема сложения:

если события А и Б несовместные, то А + Б – событие, состоящее в появлении одного из этих событий, безразлично какого. Вероятность суммы несовместных событий равна сумме вероятностей этих событий *{ Р(А + Б) = Р(А) + Р(Б) }.*

## Условная вероятность. Теорема умножения.

**Условной вероятностью PA(B)** называют вероятность события В, вычисленную в предположении, что событие А уже наступило. *{ PA(B) = P(AB) / P(A) }*

**Теорема умножения:** вероятность совместного появления двух событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило. *{ P(AB) = P(A) PA(B) }*

## Независимое событие

Такое событие Б, на вероятность которого не влияет появление события А, т.е. если условная вероятность события Б равное его безусловной вероятности *{ PA(B) = P(B) }*

Если события независимые, то теорема умножения вероятностей принимает вид: *{ P(AB) = P(A)P(B) } –* **критерий независимости событий.**

## Теорема о полной вероятности:

Вероятность события А, которое может наступить лишь при условии появления одного из несовместных событий Б1, Б2, …, Бп, образующих полную группу, равна сумме произведений вероятностей каждого из этих событий на соответствующую условную вероятность события А: Р(А) = Р(В1)РВ1(А) + Р(Вп)РВп(А) + … + Р(Вп)РВп(А)

## Теорема Байеса:

теорема, позволяющая определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие. *{* P(A|B) = P(B|A) \* P(A) / P(B) *}*

## Повторные независимые испытания

производятся несколько испытаний, причём вероятность события А в каждом испытании не зависит от исходов других испытаний. Схема испытаний бернули: , где к – кол-во испытаний, где событие А выполняется, p – вероятность события А, q = 1 — p.

# *Случайные величины*

## Дискретная (прерывная) случайная величина

случайная величина, которая принимает отдельные, изолированные возможные значения с определёнными вероятностями. Число возможных значений может быть конечным или бесконечным.

При **табличном** задании закона **распределения** дискретной случайной величины первая строка таблицы содержит возможные значения, а вторая их вероятности.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | x1 | x2 | ... | xn |
| P | p1 | p2 | ... | pn |

## Непрерывная случайная величина. Плотность распределения.

**Непрерывная случайная величина –** случайная величина, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Число возможных значений бесконечно.

**Плотность распределения –** функция f(x) – первую производную от функции распределения F(x), т.е. предел отношения вероятности того, что значение случайной величины попадет в промежуток (x, x + Δx) к длине этого промежутка.

## Функция распределения

Функция F(x), определяющая вероятность того, что случайная величина Х в результате испытания примет значение, меньшее х, т.е.

**Свойства:**

1) Значения функции распределения принадлежат отрезку [0, 1]: 0 ≤ F(x) ≤ 1.

2) F(x) – неубывающая функция, т.е. F(x2) ≥ F(x1), если х2 >х1.

3) Если возможные значения случайной величины принадлежат интервалу (a, b), то 1) F(x) = 0 при х ≤ а; 2) F(x) = 1 при х ≥ b.

## Математическое ожидание

Сумма произведений всех её возможных значений на их вероятности

**Свойства:**

1) М.о. постоянной величины равно самой постоянной М(С) = С.

2) Постоянный множитель можно выносить за знак м.о.: М(СХ) = СМ(Х).

3) Математическое ожидание произведения двух независимых случайных величин равно произведению из м.о.: М(ХУ) = М(Х)М(У).

4) М.о. суммы двух случайных величин равно сумме м.о. слагаемых: М(Х + У) = М(Х) + М(У).

## Дисперсия (рассеяние)

м.о. квадрата отклонения случайной величины от её математического ожидания: .

**Свойства:**

1) Дисперсия постоянной величины С равна нулю: D(C) = 0.

2) Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат: D(CX) = C2D(X).

3) Дисперсия суммы двух независимых случайный величин равна сумме дисперсий этих величин: D(X + Y) = D(X) + D(Y)

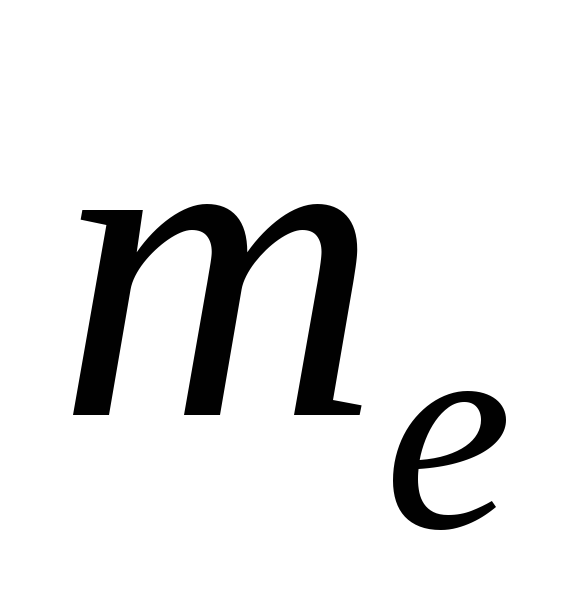
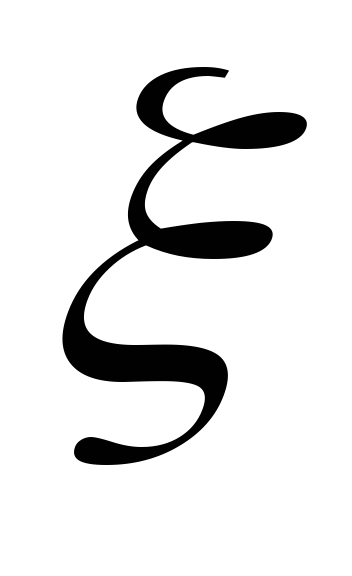
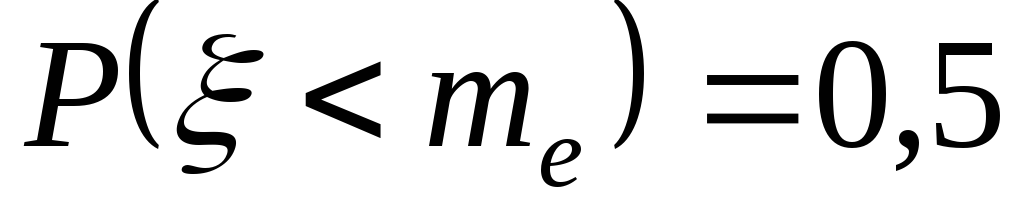
4) Дисперсия разности двух независимых случайных величин равна сумме этих дисперсий: D(X – Y) = D(X) + D(Y).

**Вычисление дисперсии:**

**Среднее квадратичное отклонение** – квадратный корень из дисперсии.

## Мода, медиана, квантиль.

**Модой** дискретной случайной величины ****называется ее значение****, имеющее наибольшую вероятность

***Медиана***  является значением случайной величины. Вероятность того, что случайная величина принимает значение меньше медианы, равна 0,5: Не все дискретные случайные величины имеют медиану.

Кванти́ль – число такое, что заданная случайная величина X  превышает его лишь с фиксированной вероятностью p.

## Биномиальное распределение

Распределение вероятностей, определяемой формулой Бернулли. Закон назван «биномиальным» потому, что правую часть равенства можно рассматривать как общий член разложения бинома Ньютона. { Bin(n, p) }.

Требования: 1) n – конечно

2) в каждом испытании два исхода: успех и неуспех

3) P(У) = p, Р(Н) = q, p + q = 1

M(X) = np; D(X) = npq

## Равномерное распределение

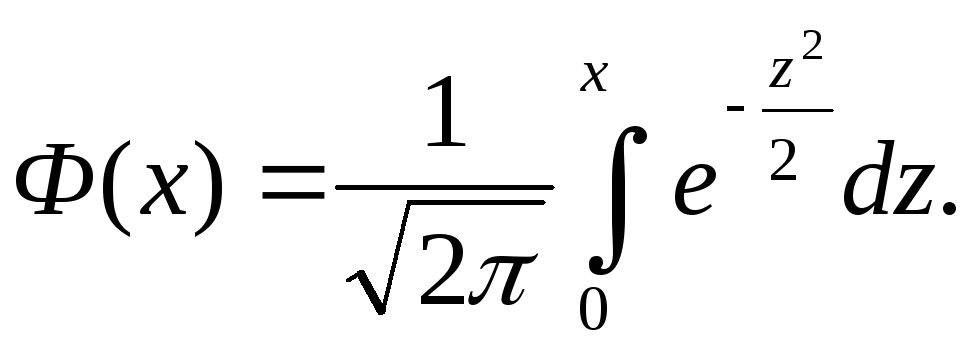
Распределение случайной величины, принимающая значения, принадлежащие интегралу [a, b], характеризующееся тем, что плотность на этом интервале постоянна.

## Нормальное распределение

Распределение НСВ, которое описывается плотностью , где а – математическое ожидание, сигма – среднее квадратическое отклонение нормального распределения.

## Функция Лапласа

**–** функция ошибок.

1) Функция Ф(х) нечетная: Ф(-х)= -Ф(х).

2) Функция Ф(х) монотонно возрастающая.

3) Ф(0)=0.

4) Ф(x > 5) = 0,5; Ф(x < -5)=-0,5.

## Вычисление вероятностей для нормального распределения с помощью функции Лапласа:

# X ~ N( 150, 10 )

P( x > 135 ) = P (135 < x < +∞ ) = Ф( (+∞ - 150) / 10) - Ф( (135 - 150) / 10) =

= Ф(+∞) - Ф( -1.5 ) = 0.5 + 0.4332 (по таблице).

# *Системы случайных величин*

## Дискретная двумерная СВ. Матрица распределения

**Дискретная двумерная СВ –** такая СВ, которая определяется двумя величинами (X, Y). X и Y – составляющие(компоненты).

**Матрица распределения –** закон распределения; матрица, первая строка которой перечисляет все возможные значения Х, а первый столбец все возможные значения Y. В клетке, стоящей на пересечении столбца xi и строкиyi указана вероятность p(xi,yi) того, что двумерная случайная величина примет значение (xi,yi). Сумма всех элементов таблице равна единице.

## Критерий независимости компонент ДССВ

Случайные величины X, Y называются **независимыми**, если независимыми являются события (X < x) и (Y < y) для любых вещественных x, y. В противном случае случайные величины (X, Y ) называются зависимыми.

Общее необходимое и достаточное условие независимости двух случайных величин: FXY (x, y) = FX(x) · FY(y) для любых вещественных x и y;

или P (AB) = P (A) P(B) для случая событий A = (X < x), B = (Y < y).

## Непрерывная двумерная СВ. Плотность распределения и её свойства.

Если существует неотрицательная функция fXY (x, y), называемая **двумерной плотностью вероятности**, такая, что вероятность попадания случайной величины (X, Y ) в область D равна двойному интегралу от плотности по области D:

**Плотность распределения –** вторая смешанная частная производная от функции распределения: .

**Свойства:**

1) Двумерная плотность вероятности неотрицательна.

2) Двойной несобственный интеграл с бесконечными пределами от двумерной плотности равен единице.

## Функция распределения

Функция F(x, y), определяющая для каждой пары чисел x, y вероятность того, что Х примет значение, меньшее х, и при этом У примет значение, меньше у. F(x, y) = P(X<x, Y<y)

**Свойства:**

1) Значение функции распределения удовлетворяют двойному неравенству 0 <= F(x, y) <= 1

2) F(x, y) есть неубывающая функция по каждому аргументу

3) Имеют место предельные отношения

3.1) F( -∞, y) = 0

3.1) F( -∞, -∞) = 0

3.1) F( x, -∞) = 0

3.1) F( ∞, ∞) = 1

4) При y = ∞ функция распределения становится функцией распределения компоненты Х и наоборот.

## Распределения компонент двумерной НСВ

**(безусл)**

**Условный закон распределения** компоненты Х при условии, что Y = yk называют совокупность возможных значений xi и соответствующих этим значениям условных вероятностей, определяемых равенством:

Аналогично для компоненты Y.

## Критерий независимости компонент НССВ:

Для того чтобы СВ X и Y были независимы, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения составляющих: F(x, y) = F1(x) F2(y)

## Ковариация

Корреляционный момент СВ X и Y **–** математическое ожидание произведения отклонений этих величин: .

**Вычислительная формула:**

Для дискретных случайных величин:

Для непрерывных случайных величин:

## Независимость, некоррелированность ССВ:

для того, чтобы X и Y были независимы, необходимо и достаточно выполнения одного из условий:

1. P(X = xi, Y = yi) = P(X = xi) P(Y = yi)
2. F(x, y) = Fx(x) Fy(y)
3. P(X = xi | Y = yk) = P(X = xi)
4. P(Y = yk | X = xi) = P(Y = yk)

**Некоррелированность ССВ:** СВ X и Y называются некоррелированными, если rxy = 0.

**Теорема 1:** если случайные величины X и Y независимы, то ковариация равна нулю.

## Свойства математического ожидания:

1. M(a + bX) = a + bM(X)
2. M(X + Y) = M(X) + M(Y)
3. M(XY) = mxmy + cov(X, Y)

## Свойства дисперсии:

1. D(a + bX) = b­2D(X)
2. D(X +- Y) = D(X) + D(Y) +- cov(X, Y)
3. Даже если X и Y независимы, то D(XY) != D(X)D(Y)

## Свойства коэффициента корреляции

1. Безразмерная величина.
2. К.к. независимых случайных величин равен нулю.

## Функция регрессии

Условное м.о. M(Y | x) есть функция от х: M(Y | x) = f(x), которую называют функцией регресии Y на X.

**Свойства:**

# *Предельные теоремы теории вероятности*

## Асимптотически нормальная СВ. Центральная предельная теорема.

**Асимптотически нормальная СВ –** СВ, распределение которой стремится к нормальному при увеличении размера выборки.

**Центральная предельная теорема –** теорема, утверждающая, что сумма большого количества независимых случайных величин имеет распределение близкое к нормальному.

## Предельные теоремы Муавра-Лапласа как следствия ЦПТ:

Пусть в каждом из https://www.matburo.ru/tv/tvbook/par_1_10.files/image002.gif независимых испытаний событие A может произойти с вероятностью https://www.matburo.ru/tv/tvbook/par_1_10.files/image004.gif, https://www.matburo.ru/tv/tvbook/par_1_10.files/image006.gif (условия [схемы Бернулли](https://www.matburo.ru/tvbook_sub.php?p=par17)). Обозначим как и раньше, через https://www.matburo.ru/tv/tvbook/par_1_10.files/image008.gif вероятность ровно https://www.matburo.ru/tv/tvbook/par_1_10.files/image010.gif появлений события А в https://www.matburo.ru/tv/tvbook/par_1_10.files/image002.gif испытаниях. кроме того, пусть https://www.matburo.ru/tv/tvbook/par_1_10.files/image012.gif– вероятность того, что число появлений события А находится между https://www.matburo.ru/tv/tvbook/par_1_10.files/image014.gif и https://www.matburo.ru/tv/tvbook/par_1_10.files/image016.gif.

Локальная теорема Лапласа.

Если n – велико, а р – отлично от 0 и 1, то

https://www.matburo.ru/tv/tvbook/par_1_10.files/image018.gif где https://www.matburo.ru/tv/tvbook/par_1_10.files/image020.gif - функция Гаусса (функция табулирована, таблицу можно скачать на странице [формул по теории вероятностей](https://www.matburo.ru/tv_spr.php)).

Интегральная теорема Лапласа.

Если n – велико, а р – отлично от 0 и 1, то P(n; k1, k2)https://www.matburo.ru/tv/tvbook/par_1_10.files/image022.gif где https://www.matburo.ru/tv/tvbook/par_1_10.files/image024.gif- функция Лапласа (функция табулирована).

Функции Гаусса и Лапласа обладают свойствами, которые необходимо знать при использовании таблиц значений этих функций:

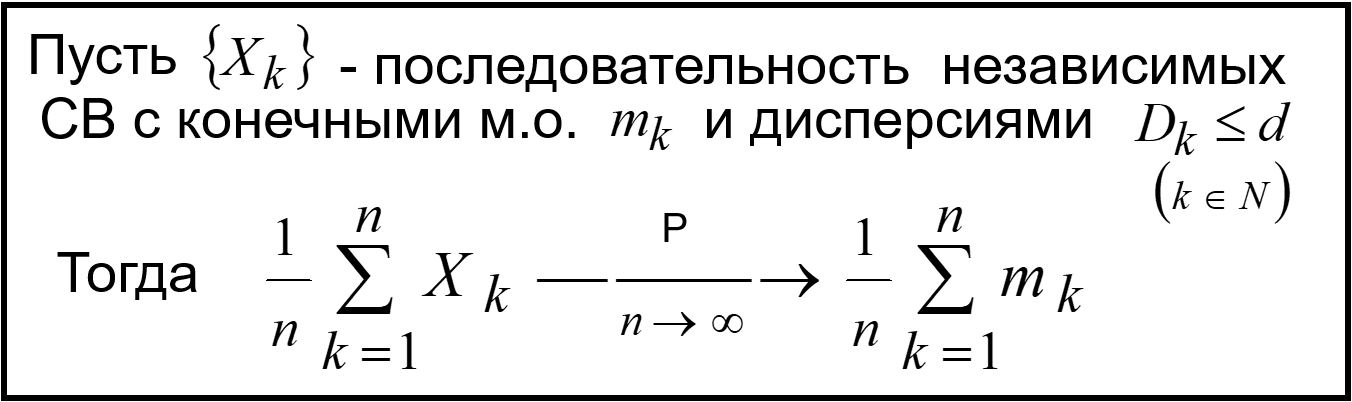
а) https://www.matburo.ru/tv/tvbook/par_1_10.files/image026.gif

б) при больших https://www.matburo.ru/tv/tvbook/par_1_10.files/image028.gif верно https://www.matburo.ru/tv/tvbook/par_1_10.files/image030.gif.

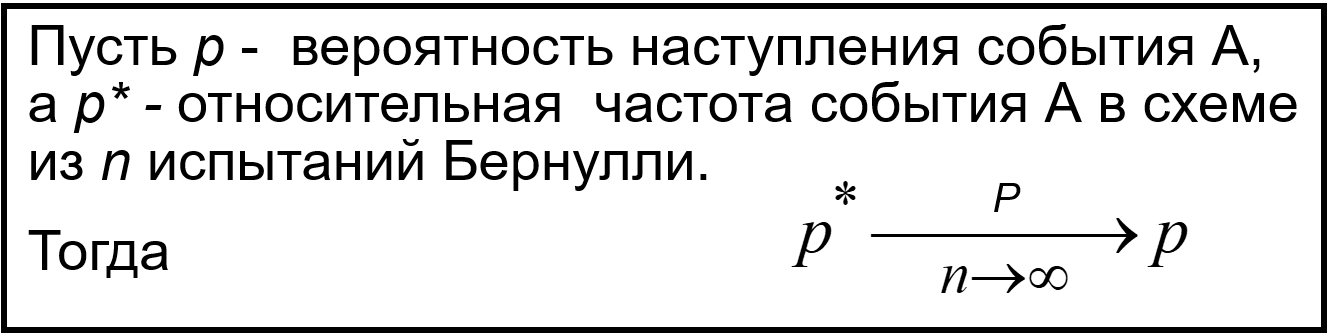
**Условия применения: ???**

## Закон больших чисел (теорема Чебышева):

Пусть СВ Х имеет конечные м.о. m и дисперсию D. Тогда для любого a > 0

**

## Закон больших чисел (теорема Бернулли)

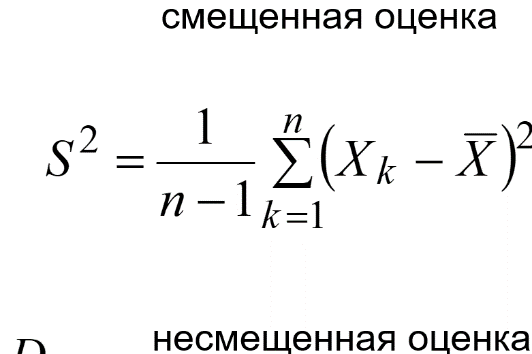
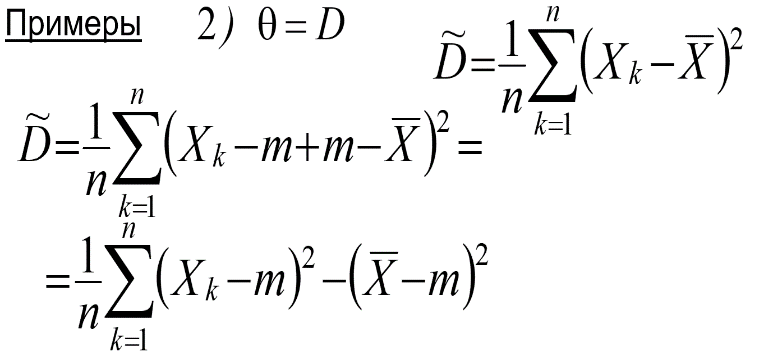
****

# *Оценки и методы оценивания*

## Оценки основных параметров генеральной совокупности по выборки

приближенное значение искомой характеристики (параметра), полученное по данным выборки.





## Оценки основных параметров генеральной совокупности по вариационному ряду.

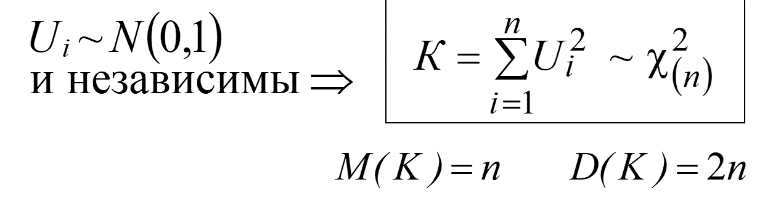
???

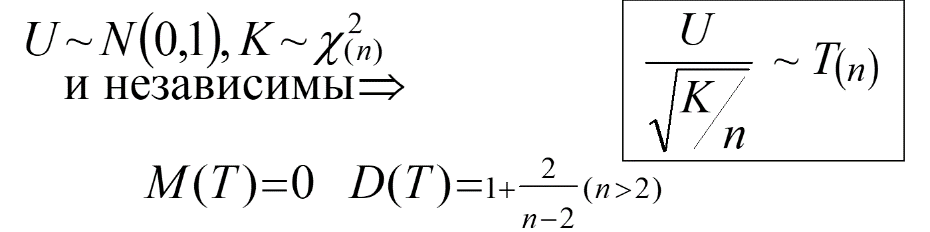
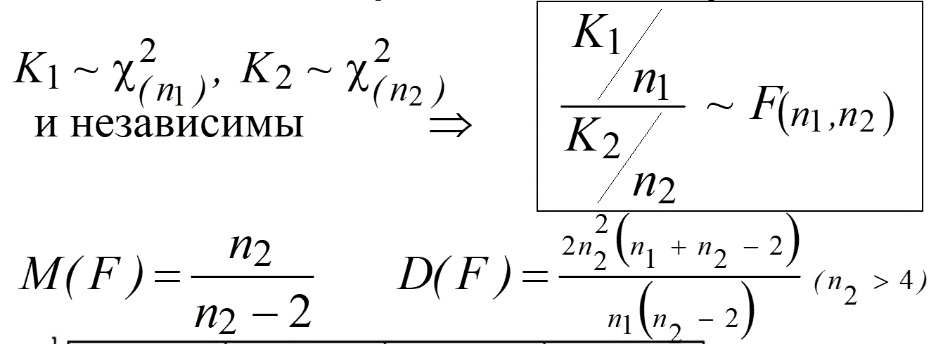
## Оценка плотности и функции распределения.

1. Непараметрические оценки: гистограмма, полигон, эмпирическая ф.

## Распределении, используемые в МатСтате.

1. Стандартное нормальное распределение U ~ N(0, 1); M(U) = 0; D(U) = 1 P(|X – mx| < e) = 2Ф(е/с.к.о.)
2. Распределение хи-хвадрат:



1. Распределение Стьюдента
2. Распределение Фишера 

## Точечная оценка параметра, несмещённость, состоятельность, сравнение оценок по эффективности.

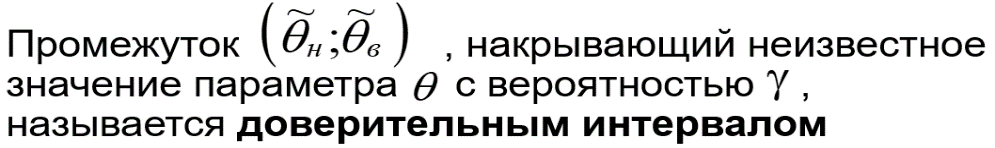
**Понятие точечной оценки параметра –** оценка, предположительно близкая к оцениваемому параметру.

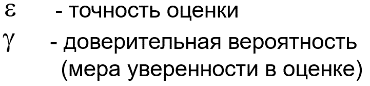
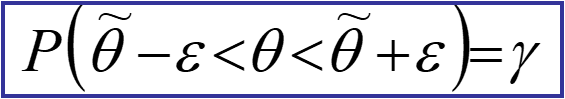
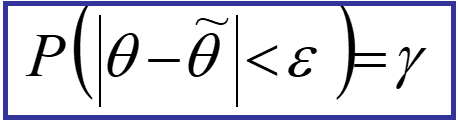
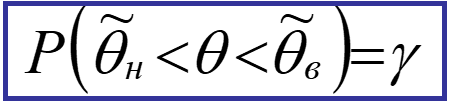
**Несмещенность –** статистическая оценка Θ\*, м.о. которой равно оцениваемому параметру Θ при любом объеме выборки, т.е. М(Θ\*) = Θ.

**Состоятельность –** статистическая оценка, которная при n стремящемся к бесконечности стремится по вероятности к оцениваемому параметру. Например, если дисперсия при n->oo стремится к нулю, то такая оценка оказывается и состоятельной.

**Сравнение оценок по эффективности:** несмещённая оценка ~Θ1 параметра Θ называется более эффективной оценкой, чем оценка ~Θ2, если её дисперсия меньше дисперсии ~Θ2.

## Понятие интервальной оценки.



****

**Точность и надежность оценки. Роль доверительной вероятности.**

# *Статистическая проверка гипотез*

## Постановка задачи:

Если закон распределения неизвестен, но имеет определённый вид, выдвигают гипотезу: генеральная совокупность распределена по закону А. Таким образом, идёт речь о виде предполагаемого распределения. Возможны и: о предполагаемой величине параметра, о равенстве параметров двух или нескольких распределений, о независимости выборок и т.д.

**Основной** (нулевой) называют выдвинутую гипотезу H0.

**Альтернативной** (конкурирующей) называют гипотезу H1, которая противоречит H0.

## Ошибки первого и второго рода. Форма критической области при проверке простых гипотез о числовых характеристиках НГС.

???

## Критерии согласия. Критерий Пирсона.

**Критерий согласия –** критерий проверки гипотезы о предполагаемом законе неизвестного распределения (проверка гипотез о соответствии эмпирического распределения и теоретическому распределению вероятностей). Имеется много критериев: «хи квадрат», Пирсона, Колмогорова, Смирнова и т.д.

**Критерий Пирсона --** [критерий](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9) для [проверки гипотезы](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B2%D0%B5%D1%80%D0%BA%D0%B0_%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D1%85_%D0%B3%D0%B8%D0%BF%D0%BE%D1%82%D0%B5%D0%B7) о принадлежности наблюдаемой выборки x1, x2, ..., xn объёмом n некоторому теоретическому [закону распределения](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) F(x, Θ).

## Критерий однородности. Критерий знаков.

**Критерии однородности --** это критерии проверки гипотез о том, что две (или более) выборки взяты из одного распределения вероятностей.

**Критерий знаков –** критерий, позволяющий проверить нулевую гипотезу, что выборка подчиняется биноминальному закону распределения с параметром p = ½. Критерий знаков можно использовать как непараметрический статистический критерий для проверки гипотеза равенства медианы заданному значению ( в частности нулю), а также отсутствия сдвига в двух связных выборках.

## Критерии однородности. Критерий вилкоксона

**Критерий Вилкоксона –** критерий для проверки однородности двух независимых выборок. Достоинство: он применим к случайным величинам, распределения которых неизвестны; требуется лишь, чтобы величины были непрерывными. Если однородны, то считается, что они извлечены из одной и той же генеральной совокупности, и имеют одинаковые, хоть и неизвестные непрерывные функции распределения. H0: F1(x) = F2(x).

# *Методы Монте-Карло*

## Какие задачи решаются методом статистических испытаний?

Этот метод указывает, как наиболее целесообразно выбрать случайную величину Х, как найти её возможные значения. Применяется в вычисления интегралов, в особенности многомерных; решении систем алгебраических уравнений высокого порядка; для исследования различного рода сложных систем (автоматического управления, экономических, биологических и т.д.)

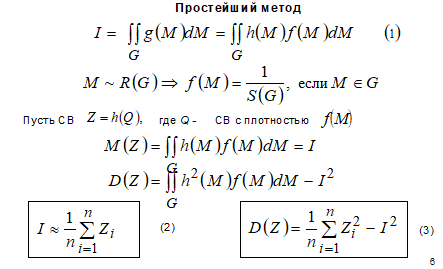
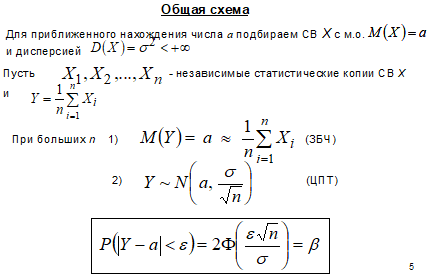
## На каких теоремах основан метод вычисления числовых характеристик?

???

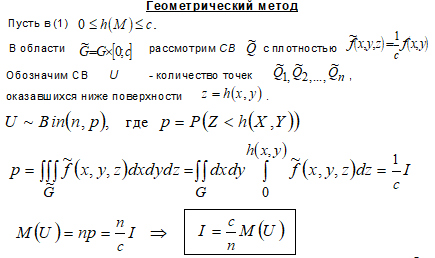
## Общая схема метода, роль доверительной вероятности.

???

## Вычисление интегралов методом Монте-Карло. Простейший метод на примере определенного интеграла.

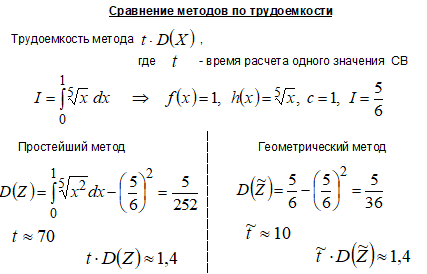
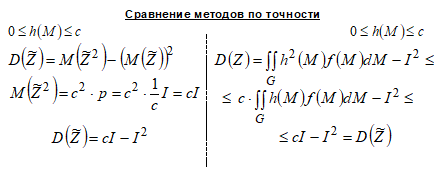


## Геометрический метод на примере определенного интеграла.

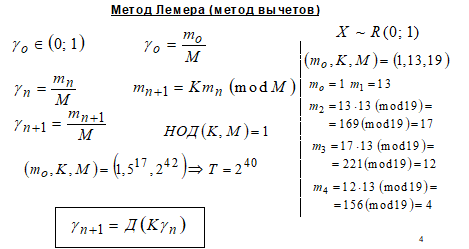
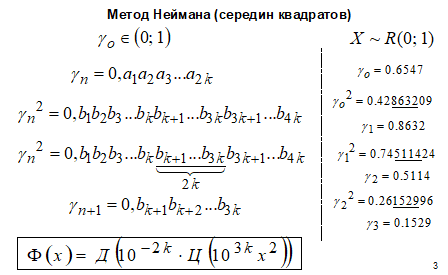




## Сравнение методов по точности и по трудоемкости



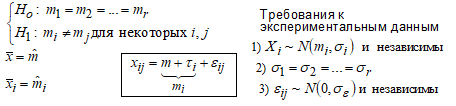
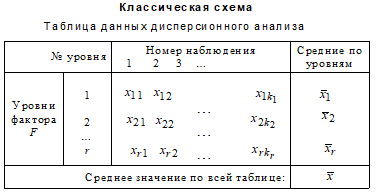
## Базовая случайная величина. Метод середин квадратов, метод вычетов.



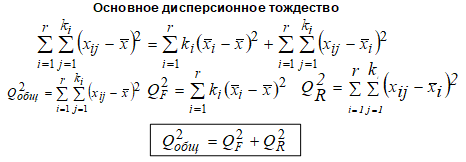
# *Дисперсионный анализ*

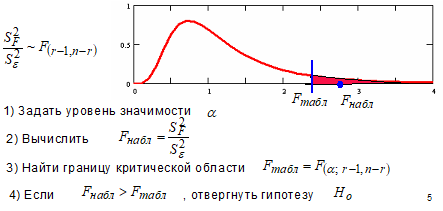
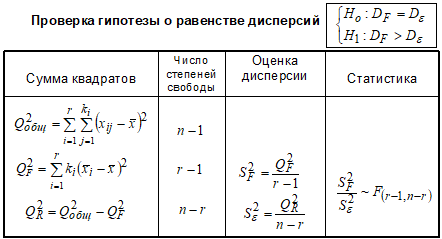
## Постановка задачи однофакторного ДА (классическая схема). Требования к экспериментальным данным.

Задача в выяснении, значимо или незначимо различаются выборочные средние (равенство всех м.о.). На практике: установить, оказывает ли существенно влияние некоторый качественный фактор F, который имеет р уровней на изучаемую величину Х.



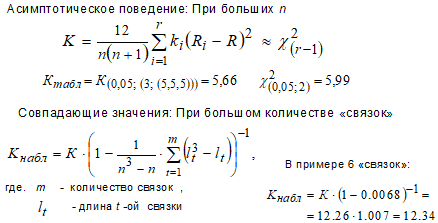
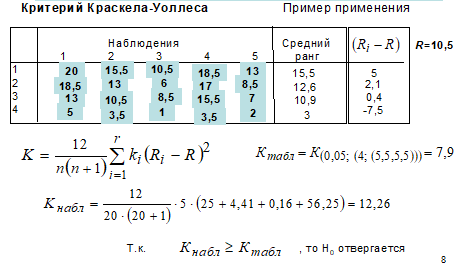
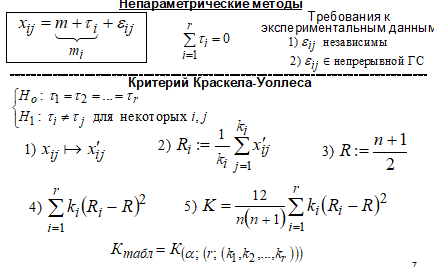
## Основное дисперсионное тождество. Проверка гипотезы о дисперсиях, обоснование эквивалентности ее первоначальной задаче.





## Когда применяется непараметрический дисперсионный анализ? Требования к экспериментальным данным, критерий Краскела-Уоллеса

Применяется когда ничего не известно о параметрах выборки (наверно).



# *Корреляционный анализ*

## Какие задачи решает корреляционный анализ?

1. Существует ли связь между факторами и откликами
2. Оценить силу связи
3. Выявить факторы, оказывающие наибольшее влияние на отклики

## Выборочный коэффициент корреляции, (условия применения, свойства).

Коэффициент корреляции Пирсона измеряет тесноту линейной связи между переменными X и Y:

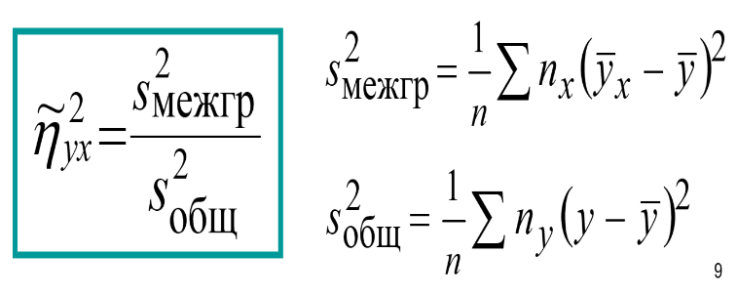
**Требования:**  2мерная нормальная генеральной совокупность, n< 50

Свойства:

1. Для любых переменных *X* и *Y* его абсолютная величина не превосходит единицы:
2. Значение коэффициента корреляции равно +1 или –1 тогда и только тогда, когда между переменными *X* и *Y* существует линейная функциональная связь
3. Если переменные *X* и *Y* независимы, то Еслито переменные *X* и *Y* называются **некоррелированными**. Некоррелированность переменных означает отсутствие между ними линейной стохастической зависимости, но не означает отсутствия связи вообще.
4. Точечной оценкой коэффициента корреляции является выборочный коэффициент корреляции который можно рассчитывать по формулам:

## Корреляционное отношение, его свойства.

Корреляционное отношение применяется в случае *нелинейной* зависимости между признаками и определяется через отношение *межгрупповой дисперсии* к *общей дисперсии*.



## Коэффициент корреляции Спирмена (условия применения, свойства).

Коэффициент корреляции Спирмена – мера линейной связи между случайными величинами. Корреляция Спирмена является ранговой, то есть для оценки силы связи используются не численные значения, а соответствующие им ранги.,где -- близость двух рангов, а xi’ yi’ –ранги, соответствующие их величинам.

**Свойства:**

1. Коэффициент Спирмена по абсолютной величине не превосходит единицы
2. принимает значения ±1 в случаях полной предсказуемости одной ранговой последовательности по другой
3. Проверка значимости коэффициента корреляции Спирмена проводится с помощью той же статистики, что и для коэффициента корреляции Пирсона

# *Регрессионный анализ*

## Основные задачи регрессионного анализа

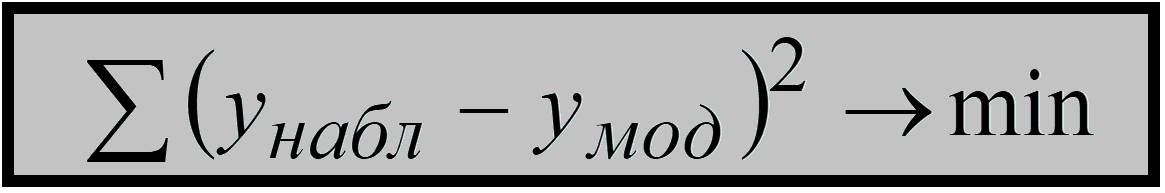
1. Подобрать форму связи (диаграмма рассеяния)
2. Оценить коэффициенты регрессии
3. Проверить качество модели (адекватность)
4. Осуществить прогноз по модельному уравнению

## Паpная линейная pегpессия

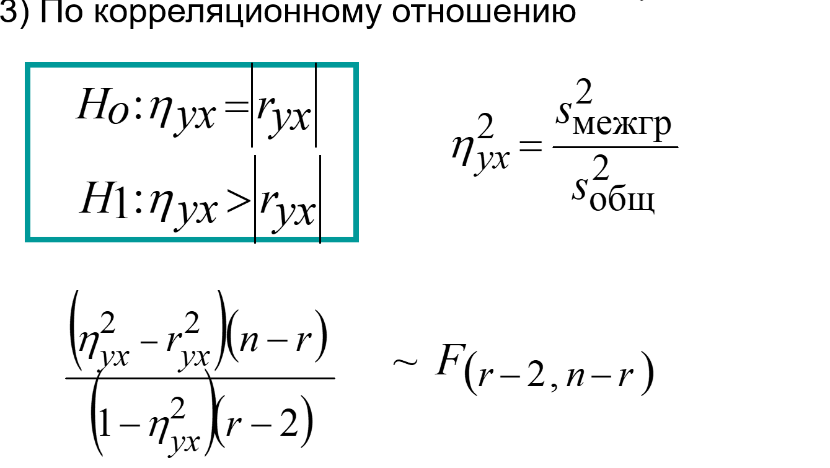
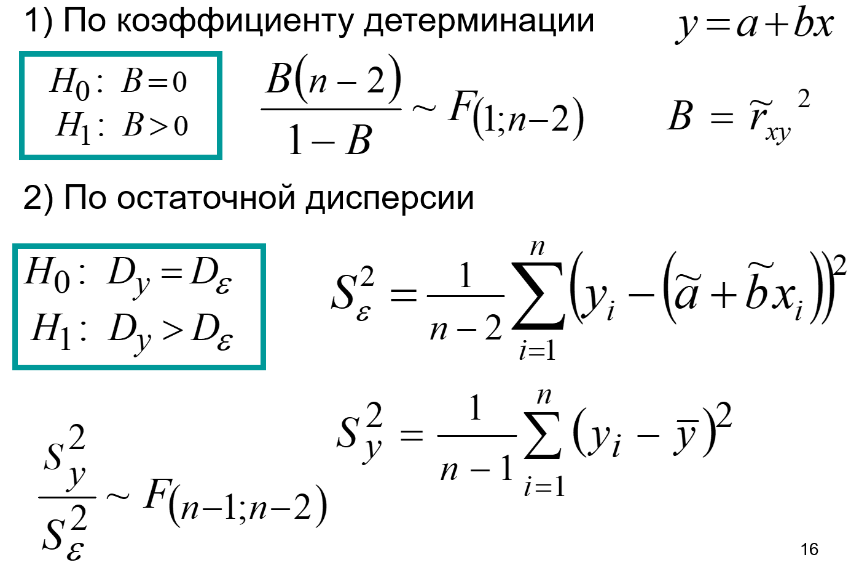
Представляет собой уравнение, описывающее связь между двумя переменными: зависимой переменной image021.png и независимой переменной image022.png. Иногда переменную image021.png называют результатом, а переменную image022.png – фактором: image021.png image023.png, при этом функция может быть как линейной, так и нелинейной.

## Метод наименьших квадpатов

Метод состоит в том, чтобы минимизировать отклонения теоретического значения отклика от наблюдаемого.



## Проверка адекватности модели (три способа).



## Множественная регрессия

Состоит в анализе связи между несколькими независимыми переменными X (называемые также регрессорами или предикторами) и зависимой переменной Y, которые связаны уравнением

## Условия применимости МНК:

1. Модель регрессии должна быть линейной по параметрам
2. х – не стохастическая переменная (заданная величина)
3. значения ошибки – случайные. Их изменение не образует определенной модели
4. число наблюдений должно быть больше числа оцениваемых параметров (5-6 раз)
5. значения переменной x не должны быть одинаковыми
6. изучаемая совокупность должна быть достаточно однородной
7. отсутствие взаимосвязи между фактором x и остатком
8. модель регрессии должна быть корректно специфицирована
9. в модели не должно наблюдаться тесной взаимосвязи между факторами

## Свойства оценок коэффициентов:

1. Несмещённость – м.о. коэффициента равно соответствующему истинному параметру регрессии M(bj) = bj
2. Эффективность – характеризуется наименьшей дисперсией D(bj) ->min
3. Состоятельность – при увеличении числа наблюдений увеличивается точность оценки.

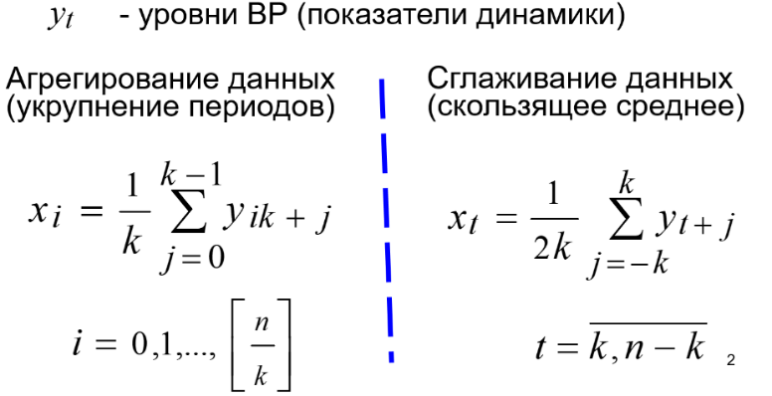
# *Временные ряды*

## Основные задачи при исследовани ВР. Агрегирование и сглаживание данных.

**Основные задачи при исследовании ВР:**

1. Предварительная обработка и анализ временных рядов
2. Моделирование временных рядов
3. Прогнозирование временных рядов

Агрегирование и сглаживание данных.



## Компоненты модели ВР:

Два вида компонент: *систематические* (результат воздействия постоянно действующих факторов. Они могут одновременно присутствовать во временном ряду) и *случайные* (случайный шум или ошибка, которая воздействует на временной ряд нерегулярно) составляющие.

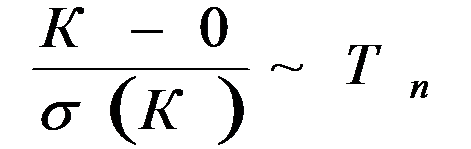
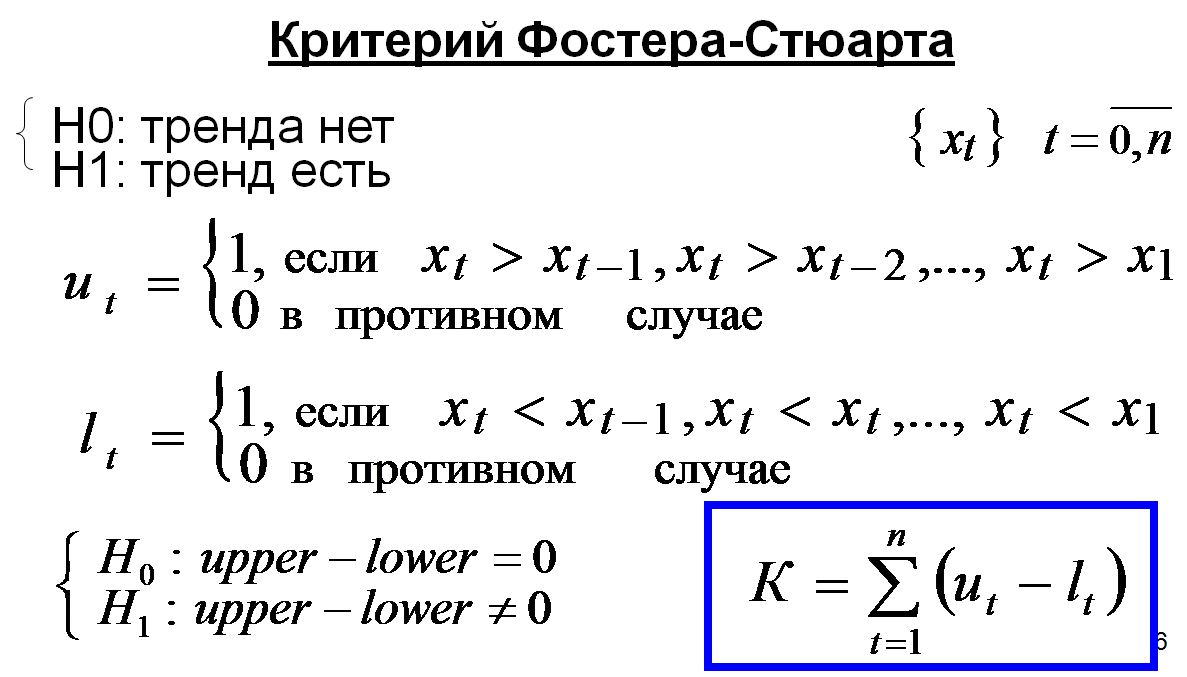
Систематические:

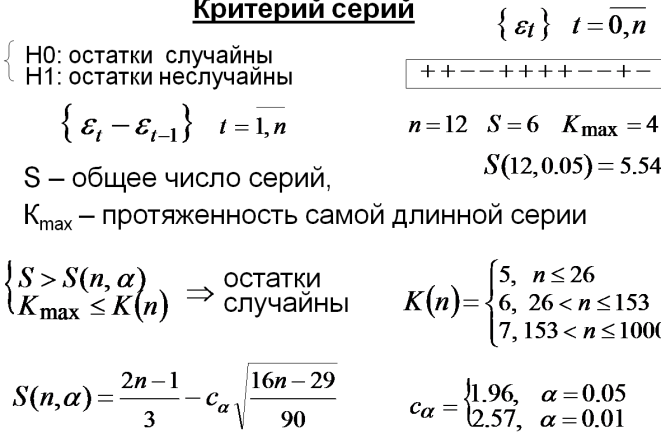
1. Тренд – линейная и нелинейная компонента, изменяющаяся во времени.
2. Сезонность – периодические колебания уровней временного ряда внутри года.
3. Цикличность – периодические колебания, выходящие за рамки одного года. Промежуток времени между двумя соседними вершинами или впадинами в масштабах года определяют, как длину цикла.

**Требования к критериям о компонентах ВР:**

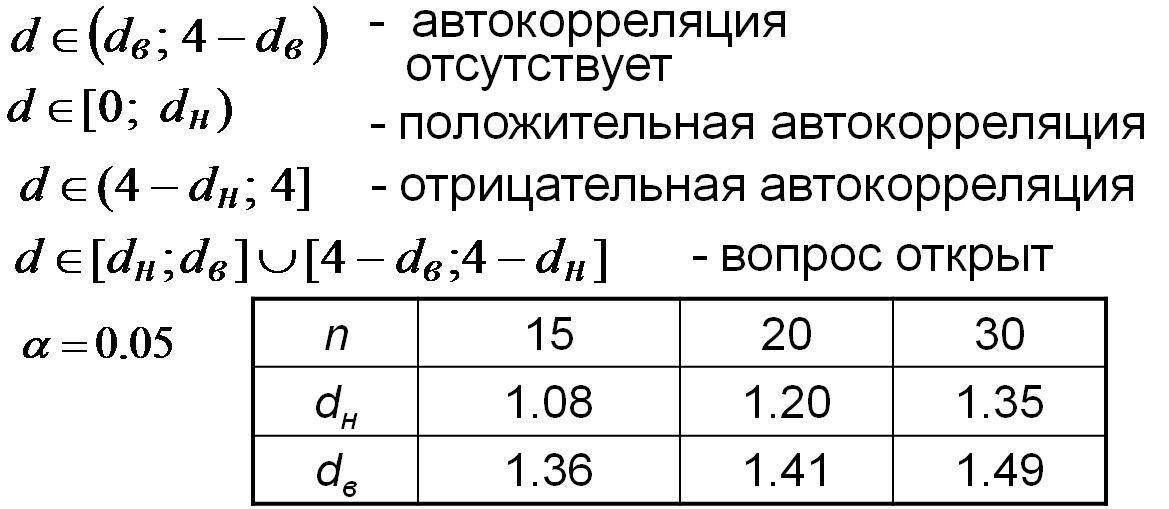
???

## Гипотезы о существовании тренда, о наличии периодической составляющей и о случайности остатков ВP (критерии Фостера-Стюарта, поворотных точек, серий, статистика Дарбина-Уотсона).





**---------------------------------------------------------------------------------------------------**



## Гетероскедастичность данных. Тест Гольдфельда-Квандта.

**Гетероскедастичность данных –** постоянство дисперсий случайных ошибок в различных наблюдениях.

**Тест Гольдфельда-Квандта** применяется, если случайные остатки предполагаются нормально распределенными случайными величинами и стандартное отклонение пропорционально значению переменной в этом наблюдении, т.е. , .

Процедура проверки:

1. Все наблюдения упорядочиваются по возрастанию фактора .
2. Упорядоченная совокупность разбивается на три группы размерностей соответственно. Причем должно быть больше чем число параметров модели.
3. Оцениваются отдельные регрессии для первой группы ( первых наблюдений) и для третьей группы ( последних наблюдений). Если предположение о пропорциональности дисперсий отклонений значениям фактора верно, то дисперсия регрессии по первой группы (рассчитываемая как ) будет существенно меньше дисперсии регрессии по третьей группе (рассчитываемой как ).
4. Формулируются:

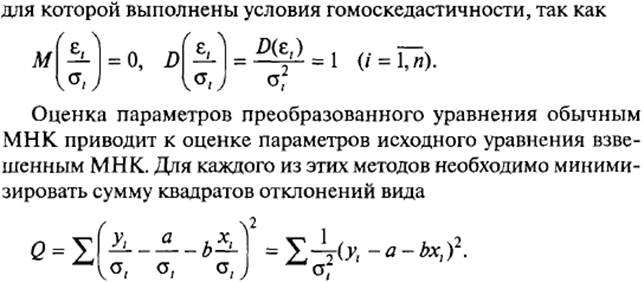
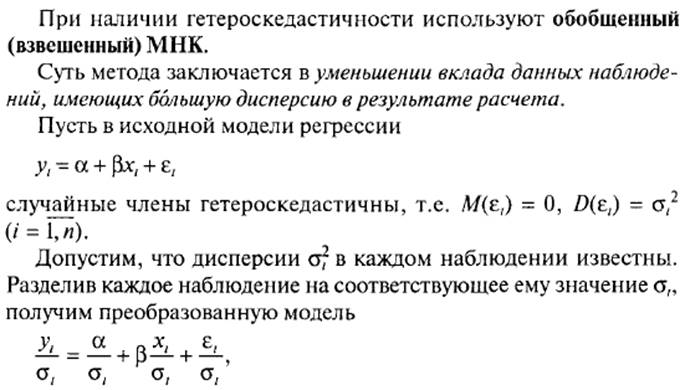
Основная гипотеза, предполагающая постоянство дисперсий случайных ошибок модели регрессии, т. е. присутствие в модели условия гомоскедастичности: .

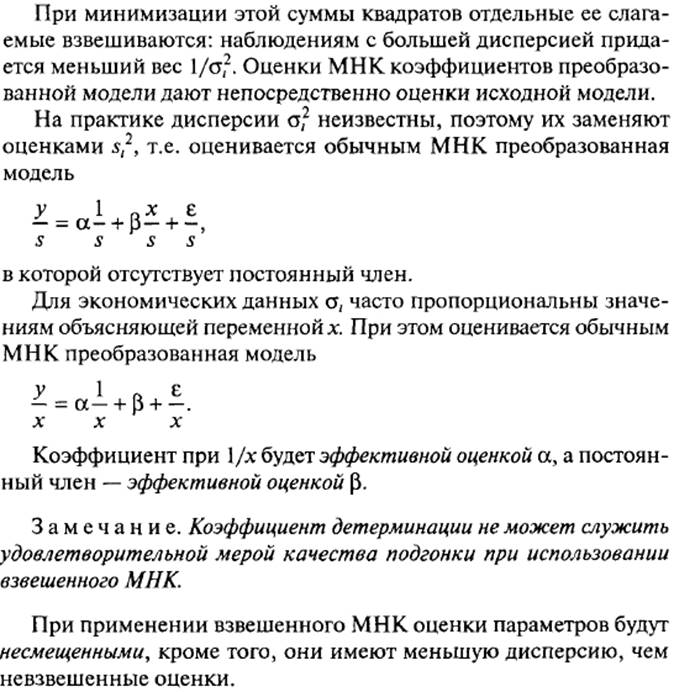
Альтернативная гипотеза, предполагающая непостоянство дисперсий случайных ошибок в различных наблюдениях, т. е. присутствие в модели условия гетероскедастичности:

1. Для сравнения соответствующих дисперсий вычисляется фактическое значение -критерия:

Здесь — число степеней свободы соответствующих выборочных дисперсий ( — количество объясняющих переменных в уравнении регрессии).

## Взвешенный МНК





## Коррелированность объясняющих переменных и случайных членов. Метод инструментальных переменных

???

## Автокорреляция в остатках. Авторегрессионное преобразование.

**Автокорреляция в остатках –** последовательная корреляция – кореляция между наблюдаемыми показателями, упорядоченными во времени или пространстве. Причины:

1) Автокорреляция может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

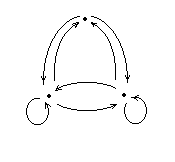
2) В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными.

**Авторегрессионное преобразование.** Сущность его заключается в построении модели по отклонениям значений временного ряда от выравненных по тренду значений.

# *Цепи Маркова*

## Матрица перехода системы. Орграф однородной ЦМ. Свойство стохастичности матрицы перехода. Классификация состояний ЦМ.

**Матрица перехода системы –** матрица, которая содержит все переходные вероятности этой системы.

**Орграф однородной ЦМ –** орграф состояний, вершины которого соответствуют состояниям ЦМ, а на дугах указаны вероятности перехода 

1

3

2

1/2

1/2

1/4 1/4

1/4

1/4

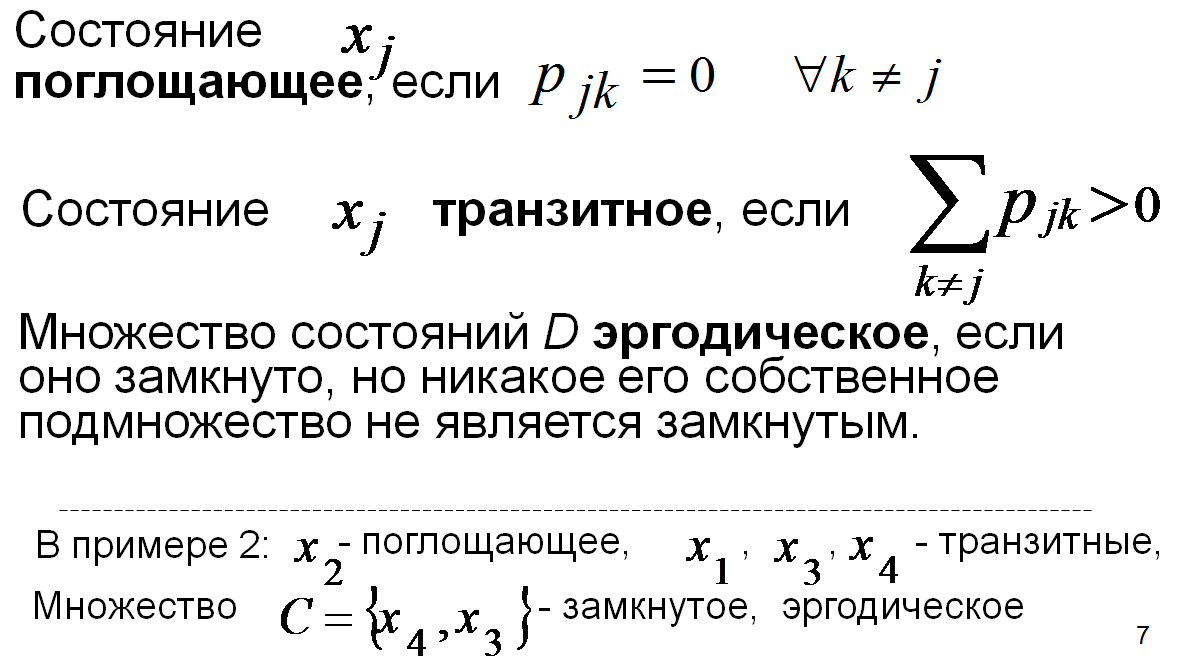
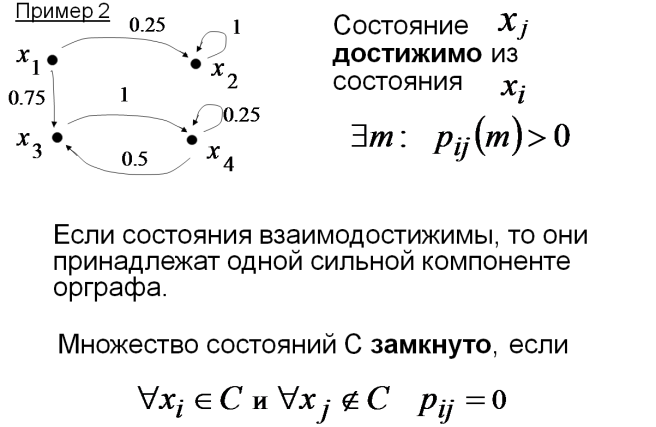
1/2

1/2

**Свойство стохастичности матрицы перехода:**

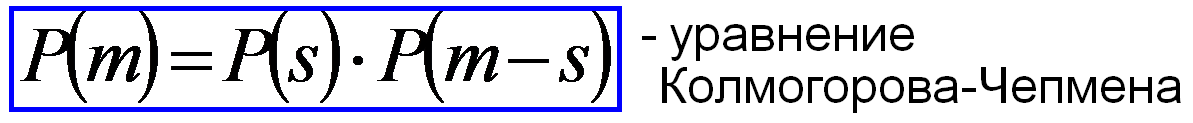
1. Каждый элемент находится в диапазоне [0; 1] => (0 <= pij <= 1)
2. Все элементы строки дают в сумме единицу.

**Классификация состояний ЦМ:**

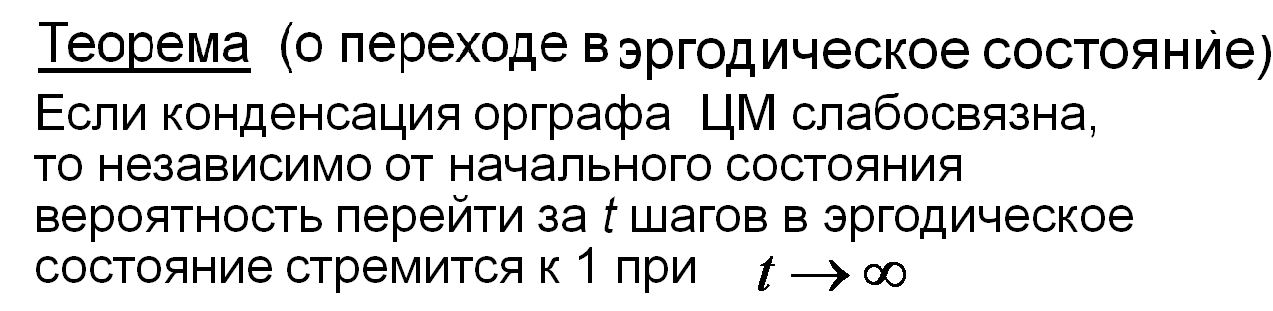
****

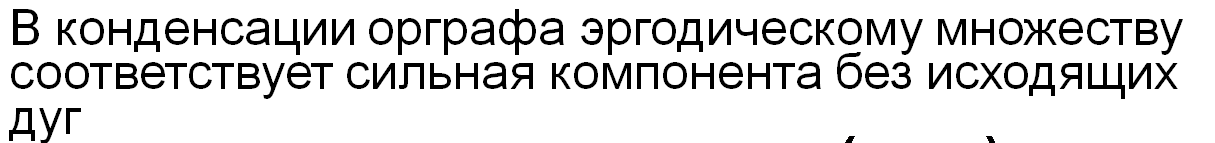
## Многошаговый переход в ЦМ. Уравнение Колмогоpова-Чепмена.

Матрица перехода за  шагов  равна -ой степени матрицы перехода  цепи Маркова. 

**,** где m – число шагов перехода xi в xj (из начального в конечное нужное), s – число шагов перейти в состояние xk (s < m)

## Теорема о вероятности перехода в эргодическое состояние.

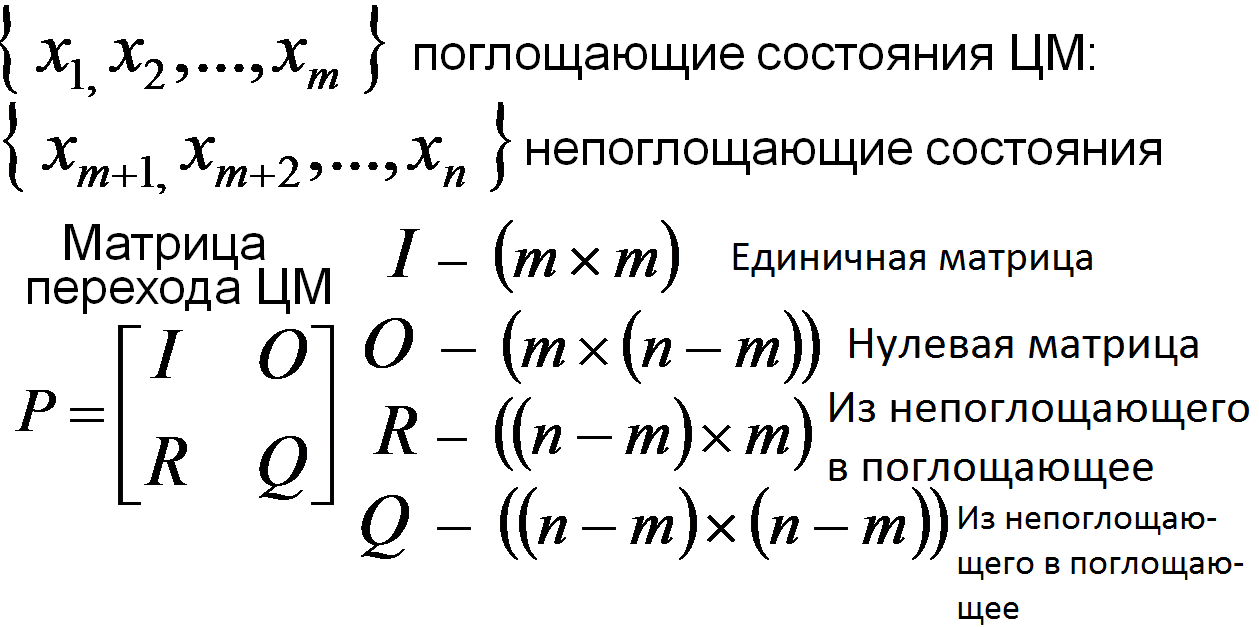
****

****

## Поглощающие ЦМ. Канонический вид матрицы перехода. Теорема о свойствах подматрицы Q.

**Поглощающие ЦМ –** если имеет хотя бы одно поглощающее состояние xp и для любого непоглощающего состояния xh

**Канонический вид матрицы перехода.**

****

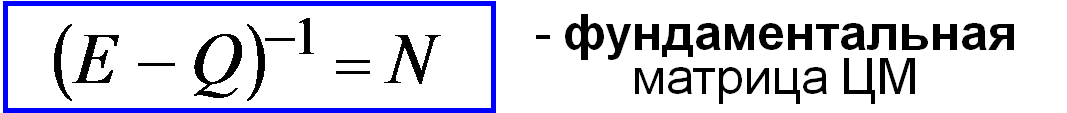
**Теорема о свойствах подматрицы Q:**

*Теорема:* Для поглощающей цепи с матрицей канонического вида справедливо:

а) ;

б) существует матрица, обратная к , и ее можно найти по формуле .

## Фундаментальная матрица. Теорема.

****

**Теорема об элементах фундаментальной матрицы:**

Элемент n­ij фундаментальной матрицы N равен среднему времени нахождения ЦМ в непоглощающем состоянии xj (начальном состоянии) до момента поглощения.

## Теорема о вероятности поглощения в данном поглощающем состоянии:

В поглощающей ЦМ с матрицей перехода канонического вида, вероятность перехода в заданное поглощающее состояние определяется матрицей B = NR, где bik – вероятность попасть в k-ое поглощающее состояние.

## Эргодические ЦМ. Регулярность. Стационарный режим ЦМ. Предельные вероятности.

**Эргодические ЦМ:** нет поглощающих состояний и из любого состояния можно попасть в любое другое. Орграф сильносвязен.

**Регулярность:** если существует такое k, что pij > 0 для любых i, j от 1 до n.

**Стационарный режим ЦМ:** при достаточно больших t вероятность нахождения в состоянии xi не зависит от того, каким было начальное состояние системы.

**Предельные вероятности:** вектор предельных вероятностей имеет вид . Нахождение вектора: . Данные вероятности показывают среднюю долю времени, которое ЦМ проводит в состоянии xj, j=(1,3), находясь в стационарном режиме.