1 Случайный процесс.+

2 Классификация случайных процессов.+

3 Граф состояний.+

4 Классификация состояния цепи.+

5 Марковский процесс.+

6 Цепь Маркова и цепь Маркова с непрерывным временем.+

7 Способ задания цепи Маркова.+

8 Однородные цепи Маркова.+

9 Переходные вероятности.+

10 Стационарное распределение для цепей Маркова.+

11 Условия существования стационарного режима.+

12 Правило поиска предельных вероятностей.+

13 Формула нахождения вероятности перехода за m шагов.+

14 Теорема о предельных вероятностях.+

15 Выборка.+

16 Эмпирическая функция распределения.+

17 Гистограмма.+

18 Вариационный ряд.+

19 Стат. ряд частот.+

20 Стат. ряд частностей.+

21 Полигон.+

22 Интервальный вариационный ряд.+

23 Размах выборки.+

24 Выборочное среднее и его свойства.+

25 Выборочная дисперсия и ее свойства.+

26 Исправленная выборочная дисперсия и ее свойства.+

27 Выборочное среднеквадратичное отклонение.+

28 Эмпирические центральные моменты.+

29 Эмпирические начальные моменты.+

30 Точечная оценка.+

31 Несмещенная стат. оценка.+

32 Состоятельная стат. оценка.+

33 Эффективная стат. оценка.+

34 Асимптотически эффективная стат. оценка.+

35 Свойства выборочного среднего и исправленной выборочной дисперсии для нормального распределения генеральной совокупности.+

36 Выборочная медиана.+

37 Выборочный коэф. асимметрии+

38 Выборочный коэф. эксцесса.+

39 Распределение хи-квадрат с n степенями свободы.+

40 Распределение Стьюдента с n степенями свободы.+

41 Распределение Фишера с n1 и n2 степенями свободы+

42 Теорема Фишера.+

43 Теорема о распределении случ. велич.+

44 Теорема о распределении случ. велич.+

45 Интервальные оценки, доверительный интервал.+

46 Метод Неймана построения доверительных интервалов.+

47 Доверительный интервал для мат. ожидания при известной дис­персии нормально распределенной генеральной совокупности.+

48 Доверительный интервал для мат. ожидания при неизвестной дисперсии нормального распределения генеральной совокупности+

49 Построение доверительного интервала для мат. ожидания в случае большой выборки+

50 Построение доверительного интервала для среднеквадратичного отклонения и дисперсии нормального распределения генеральной совокупности.+

51 Статистическая гипотеза. Простая гипотеза. Основная и альтернативная гипотезы+

52 Критерий согласия+

53 Мера отклонения и различные критерии+

54 Критерий "хи" квадрат в случае простой гипотезы+

55 Критерий "хи" квадрат в случае, когда по выборке оценива­ются параметры+

56 Критерий Колмогорова+

57 Критерии однородности и независимости+

58 Критерий Колмогорова-Смирнова+

59 Критерий Вилкоксона-Манна-Уитни+

60 Статическая оценка коэффициента корреляции Пирсона

61 Коэффициент корреляции Спирмана

62 Проверка гипотезы о независимости признаков на основании критерия Спирмана

63 Коэффициент корреляции Кендалла

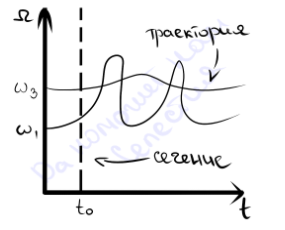
64 Проверка гипотезы о независимости признаков на основании критерия Кендалла

65 Задача регрессионного анализа

66 Статическая оценка параметров линейной зависимости между случайными величинами+

**1 Случайный процесс**

Случайный процесс - семейство случайных величин, индексированных параметром t, который зачастую играет роль времени или координат.

Случайный процесс (СП) это некоторый процесс или явление, поведение которого в течение времени и результат заранее предсказывать невозможно. Примеры случайных процессов: динамика изменения курса валют или акций, выручка или прибыль организации с течением времени, объемы продаж товара и т.д.

ვ(w, t0) - сечение случайного процесса

При фиксированном w, будет траекторией

**2 Классификация случайных процессов**

Пусть ვ(t) - случайная величина, t ∈ Ө

X = {ვ(t), t ∈ Ө} - мн-во значений СП

1. Ө - является счетным числом ⇒ ვ(t) - СП с дискретным временем
2. Ө = [a, b] ⇒ ვ(t) - СП с непрерывным временем
3. X - дискретн, Ө = [a, b] ⇒ СП с дискретным множеством состояний
4. X- непр, Ө - непр ⇒ СП с непрерывным множеством состояний

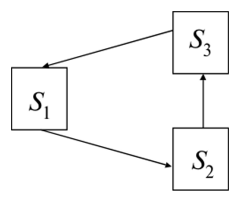
Примеры:

1. подбрасывание монетки
2. каждый день в 8 утра измеряли температуру воздуха
3. пусть есть устройство из узлов. число отказавших узлов - ვ(t)
4. величина напряжения в электросети

**3 Граф состояний**

Пусть дана цепь Маркова. По этой цепи построим модель (граф состояний).

* вершины обозначают состояния
* ребра показывают переходы между двумя состояниями



Граф состояний геометрически изображает возможные состояния системы и ее возможные переходы из состояния в состояние.

**4 Классификация состояния цепи**

i, j, Pij(k) > 0 ⇒ состояние j достижимо из состояния j

Определение:

Мн-во состояний цепи - замкнутый класс, если любые 2 состояния явл. достижимыми и для ∀ i ∈ C, ∀ j ∉ C, Pij = 0

Для любой цепи Маркова существует хотя бы один замкнутый класс

Определение:

Источник - состояние, в которое нельзя войти, но можно выйти из

Поглощающее - состояние, в которое можно войти, но нельзя выйти из

Изолированное - нельзя войти и выйти

Транзитное - можно войти и выйти

Эргодическое подмн. сост. - из любого состояния мн. можно перейти в любое другое подмн.

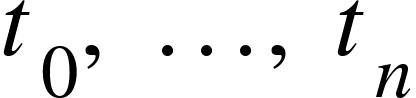
Соседние состояния - состояния, которые достижимы за один временной шаг

**5 Марковский процесс**

Марковский процесс – это случайный процесс в системе S с дискретными состояниями s1,s2,s3…sn, если для любого момента времени t0, вероятность каждого из состояний системы в будущем (при t>t0) зависит только от ее состояния в настоящем (t=t0) и не зависит от ее поведения в прошлом (t<t0).

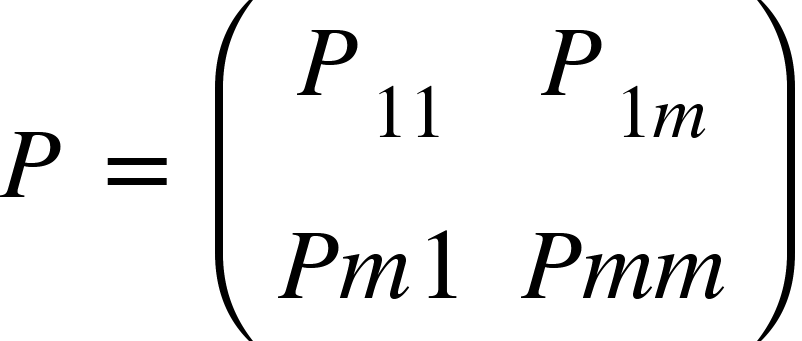
[Марковская цепь](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%80%D0%BA%D0%BE%D0%B2%D1%81%D0%BA%D0%B0%D1%8F_%D1%86%D0%B5%D0%BF%D1%8C) — частный случай марковского процесса, когда пространство его состояний дискретно.

С дискретным временем:

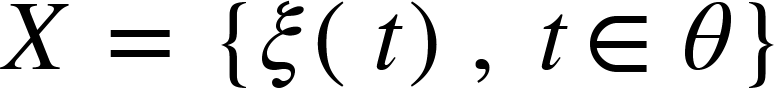
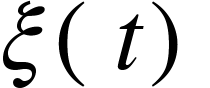
Случайный процесс, протекающий в некоторой системе S, переходы состояний происходят в заранее известные моменты времени  (это шаги или этапы процесса). В промежутки времени между смежными шагами состояние системы не изменяется.

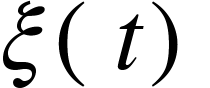
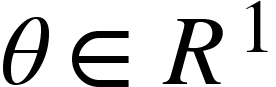
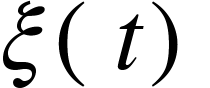
С непрерывным временем:

Случайный процесс, протекающий в некоторой системе S, переходы состояний возможны в любые, заранее неизвестные, случайные моменты времени.

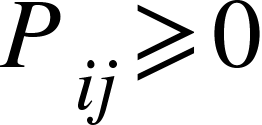
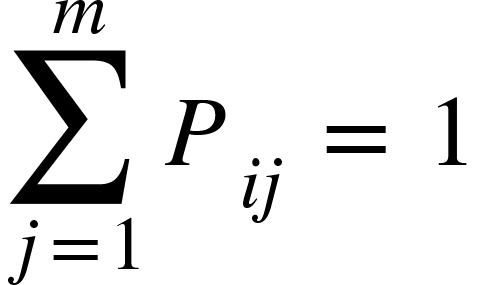
 - стохастическая матрица. (Каждая задает Марковский процесс)

**6+8 Цепь Маркова и цепь Маркова с непрерывным временем, однородные цепи Маркова**

Пусть есть Марковский случайный процесс .  - цепь Маркова, если <math xmlns="http://www.w3.org/1998/Math/MathML"><mi>&#x3B8;</mi></math> - счетное множество, при этом часть требуют, чтобы множество состояний Х было конечным или счетным.

- цепь Маркова с непрерывным временем, если , Х не более чем счетно. Т. е. цепью Маркова с непрерывным временем называется цепь, изменение состояний которой возможно в любые случайные моменты времени.

Свойства однор. цепи:

1. 
2.  - матрица переходных вероятностей является [стохастической](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%BE%D1%85%D0%B0%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0)

Цепь Маркова называется однородной, если матрица переходных вероятностей не зависит от номера шага (n), то есть



В противном случае цепь Маркова называется неоднородной.

Цепь Маркова называется конечной однородной, если она однородна, а множество состояний конечно.

Задается матрицей переходных состояний и вектором начальных состояний.

**7 Способ задания цепи Маркова**

1. Матрица перехода.
2. Граф вероятностей переходов.

Задается матрицей переходных состояний и вектором начальных состояний.

Для описания цепи Маркова удобно использовать граф вероятностей переходов, вершины которого обозначают возможные состояния системы, стрелки от одной вершины к другой указывают возможные переходы между состояниями, а число над стрелкой задаёт вероятность такого перехода.

**9 Переходные вероятности**

*Переходной вероятностью рij* называют условную вероятность того, что из состояния *i* в итоге следующего испытания система перейдет в состояние *j*.

*Матрицей перехода системы* называют матрицу, ко­торая содержит все переходные вероятности этой сис­темы

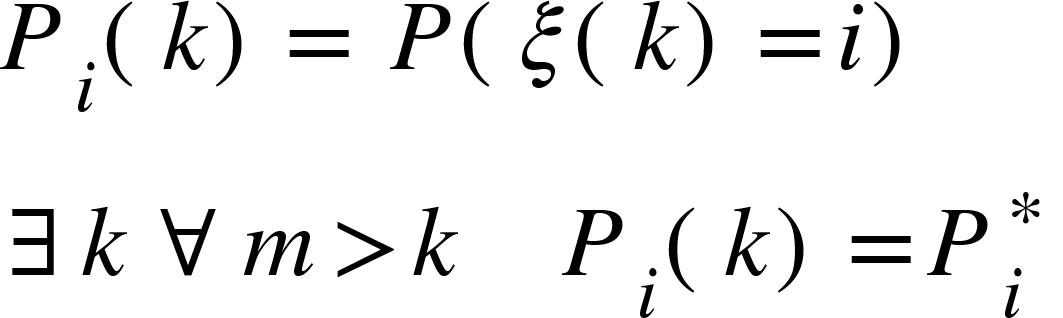
Так как в каждой строке матрицы помещены вероят­ности событий (перехода из одного и того же состояния *i* в любое возможное состояние *j*). которые образуют полную группу, то сумма вероятностей этих событий равна единице.

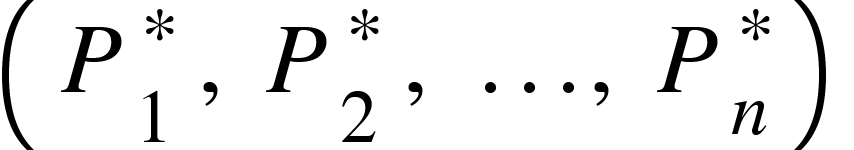
**10 Стационарное распределение для цепей Маркова**

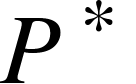
Стационарное распределение цепи Маркова — это такое распределение вероятности, которое не меняется с течением времени.

Пусть — однородная [цепь Маркова](https://ru.wikipedia.org/wiki/%D0%A6%D0%B5%D0%BF%D1%8C_%D0%9C%D0%B0%D1%80%D0%BA%D0%BE%D0%B2%D0%B0) с дискретным временем, [счётным](https://ru.wikipedia.org/wiki/%D0%A1%D1%87%D1%91%D1%82%D0%BD%D0%BE%D0%B5_%D0%BC%D0%BD%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2%D0%BE) пространством состояний  и матрицей переходных вероятностей  Тогда [дискретное распределение](https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5) называется стационарным (инвариантным), если 

Например, процесс изменения напряжения в сети питания технического устройства, пройдя сразу после включения через ряд колебаний, по прошествии времени устанавливается.

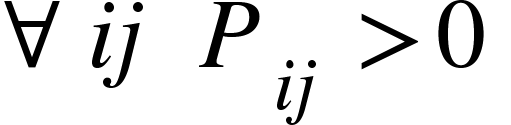


 - стационарный режим (состояние).

Смысл - доля времени, когда система находится в данных состояниях.

**11 Условия существования стационарного режима**

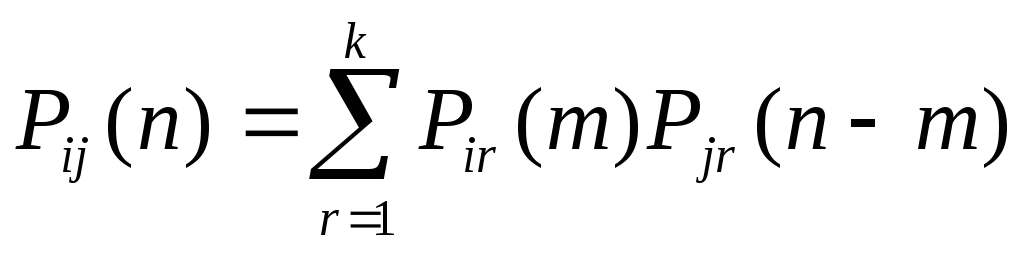
1. Множество состояний должно быть эргодическим.
2. Цепь должна быть однородной.
3. Цепь должна быть хорошо перемешанна (не будет явных циклов).

Если для , тогда существуют стационарные вероятности.

**12 Правило поиска предельных вероятностей**

Для всех узлов высчитываем суммарную интенсивность всех потоков, а также сумму произведений вероятностей всех состояний, получаем уравнение. Проделываем для всех узлов. В итоге получаем систему уравнений. Исключается одно из уравнений системы и вместо него добавляем условие нормировки (P1(t) + … + Pn(t) = 1). Приравниваем производную к 0, решаем систему уравнений.

**13 Формула нахождения вероятности перехода за m шагов.**

Введем промежуточное (между *i* и *j*) состояние *r*. Другими словами, будем считать, что из первоначального состояния *i* за *m* шагов система перейдет в промежуточное состояние *r* с вероятностью pij(n-m), после чего, за оставшиеся n-m шагов из промежуточного состояния r она перейдет в конечное состояние j с вероятностью pij(n-m). По формуле полной вероятности получаем:

**14 Теорема о предельных вероятностях.**

Если для любого i,j - Pij > 0, все элементы > 0, тогда ∃ пред. вероятности (среднее относительное время пребывания системы в этом состоянии).

Следствие: Пусть существует такое k, где все элементы больше p^k, тогда существует предельная вероятность.

**15 Выборка.**

Выборка - независимые наблюдения над одной и той же случайной величиной.

Способы представления выборки:

1. вариационный ряд частот
2. полигон частот
3. гистограмма частот
4. интервальный вариационный ряд

**16 Эмпирическая функция распределения.**

Функцией распределения выборки называют функцию, которая определяет для каждого значения x частоту события, таких что X < x и предназначена для оценки теоретический функции распределения генеральной совокупности в математической статистике., nx – количество наблюдений, n – общей объем выборки.

**17 Гистограмма.**

Гистограмма – это метод графического представления распределений непрерывно варьирующих признаков и состоит из примыкающих друг к другу прямоугольников, где:

* Основание – это соответстующий интервал
* Высота – частота соответствующего интервала

, где n - частота i-го интервала группировки, h - ширина i-го интервала группировки.

**18 Вариационный ряд.**

Вариационный ряд - упорядоченная по возрастанию/убыванию выборка

Крайние члены называются экстремальными значениями вариационного ряда.

**19 Стат. ряд частот**

Вариационным (статистическим) рядом называется таблица, первая строка которой содержит в порядке возрастания элементы ', а вторая - их частоты (относительные частоты).

Частоты - это сколько значений попали в интервал.

**20 Стат. ряд частностей**

Частности – это частоты, выраженные в виде относительных величин.

Сумма частостей равна единице или 100 %. Замена частот частостями позволяет сопоставлять вариационные ряды с разным числом наблюдений.

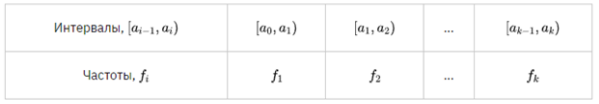
**21 Полигон**

Элементы (в порядке возрастания) - xi, частоты - mi, относительные частоты - fi.

**Полигоном частот** (относительных частот) выборки называется ломаная с вершинами в точках (xi, mi) ( (xi, fi) ).

**22 Интервальный вариационный ряд**

Интервальный вариационный ряд - ряд распределения, в котором однородные группы составлены по признаку, меняющемуся непрерывно или принимающему слишком много значений

****

**23 Размах выборки**

Разность между наибольшим и наименьшим значением выборки называют размахом вариации (выборки). R = xmax - xmin

**24 Выборочное среднее и его свойства.**

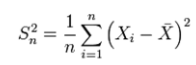
Выборочное среднее - приближение теоретического среднего распределения, основанное на выборке из него.

Является статистической оценкой для матожидания

Свойства:

1. Несмещенность
2. Состоятельность
3. Выборочное среднее из нормальной выборки — [эффективная оценка](https://ru.wikipedia.org/wiki/%D0%AD%D1%84%D1%84%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) её среднего
4. Асимпотически нормальная оценка

**25+26 Выборочная дисперсия и исправленная выборочная дисперсия, свойства.**

Выборочная дисперсия — это оценка теоретической дисперсии распределения, рассчитанная на основе данных выборки. Виды выборочных дисперсий:

1. смещенная или неисправленная
2. несмещенная или исправленная

X c чертой - выборочное среднее, n - объем выборки, у исправленной - вместо 1/n ⇒ 1/n-1

Свойства:

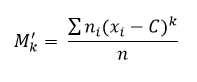
1. Смещенность и несмещенность
2. Состоятельность (для обоих типов)

**27 Выборочное среднеквадратичное отклонение.**

Выборочное среднеквадратическое отклонение - характеристика распределения случайной величины, показывающая среднюю степень разброса значений величины относительно мат. ожидания.Определяется как корень из выборочной дисперсии. Может быть (не-) смещенной при использовании в подсчетах соответственно (не-) и смещенной дисперсии.

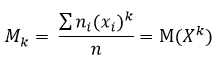
**28-29 Эмпирические начальные и центральные моменты.**

Обычным эмпирическим моментом порядка k называют среднее значение   
*k-x* степеней разностей *х* - *С:*



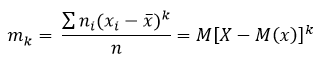
Где хi — наблюдаемый вариант, *ni —* частота варианта, *n* = объем выборки, *С —* произвольное постоянное число.

**Начальным эмпирическим моментом** порядка k называют обычный момент порядка k при С = 0, или Математическим ожиданием от величины в k-й степени



(начальный эмпирический момент первого порядка равен выборочной средней)

**Центральным эмпирическим моментом** порядка k называют обычный момент порядка k при С = (выборочное среднее), или математическим ожиданием величины: (X – M(X))^k



(центральный эмпирический момент второго порядка равен выборочной дисперсии)

# 30-34 Точечная оценка: несмещенная, cостоятельная, эффективная, aсимптотически эффективная стат. оценки.

Пускай выборка x1…xN зависит от параметра *ፀ* ∈ Θ. Тогда статистику,

, принимающую значения в Θ, называют точечной оценкой параметра *ፀ.*

1) Оценка называется [несмещённой](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0), если её математическое ожидание равно оцениваемому параметру генеральной совокупности

2) Оценка называется [состоятельной](https://ru.wikipedia.org/wiki/%D0%A1%D0%BE%D1%81%D1%82%D0%BE%D1%8F%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0), если она с увеличением объема выборки n стремится [по вероятности](https://ru.wikipedia.org/wiki/%D0%A1%D1%85%D0%BE%D0%B4%D0%B8%D0%BC%D0%BE%D1%81%D1%82%D1%8C_%D0%BF%D0%BE_%D0%BC%D0%B5%D1%80%D0%B5) к параметру генеральной совокупности

3) Оценка называется [эффективной](https://ru.wikipedia.org/wiki/%D0%AD%D1%84%D1%84%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0), если она обладает минимальной дисперсией среди всех возможных несмещенных точечных оценок

4) Асимптотически эффективная оценка — оценка, [распределение](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) которой стремится к [нормальному распределению](https://ru.wikipedia.org/wiki/%D0%9D%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5) при увеличении размера выборки. Асимптотически эффективная оценка является асимптотически эффективной, если асимптотическая ковариационная матрица минимальна в данном классе оценок.

**35 Свойства выборочного среднего и исправленной выборочной дисперсии для нормального распределения генеральной совокупности.**

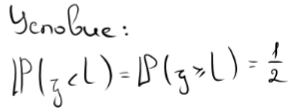
Выборочное среднее:

1. Несмещенное
2. Состоятельное
3. В случае норм. распр. ген. совокупности - эффективное

Исправл. выб. D:

1. Несмещенная
2. В случае норм. распр. ген. совокупности - асимпт. эффективная

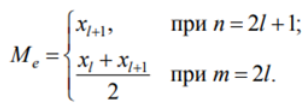
**36 Выборочная медиана.**



Медиана  вариационного ряда\* – это значение, которая делит его на две равные части (по количеству вариант).

\* не важно, [дискретного](http://www.mathprofi.ru/diskretnyi_variacionnyi_ryad.html) или [интервального](http://www.mathprofi.ru/intervalnyi_variacionnyi_ryad.html), генеральной совокупности или выборочной.

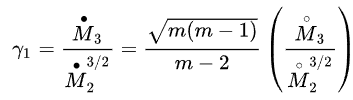
Формула зависит от четности:



**37 Выборочный коэф. асимметрии**

Пусть задана [случайная выборка](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0)  наблюдений 

Выборочный коэффициент асимметрии определяется формулой:



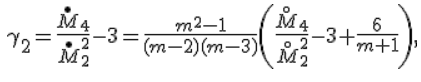
Где:

*  - выборочный центральный момент k-го порядка;
* - [несмещённая оценка](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) центрального момента второго порядка;
*  - [несмещённая оценка](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) центрального момента третьего порядка.

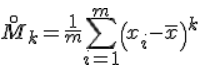
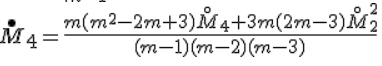
**38 Выборочный коэф. эксцесса.**

Пусть задана выборка  наблюдений 

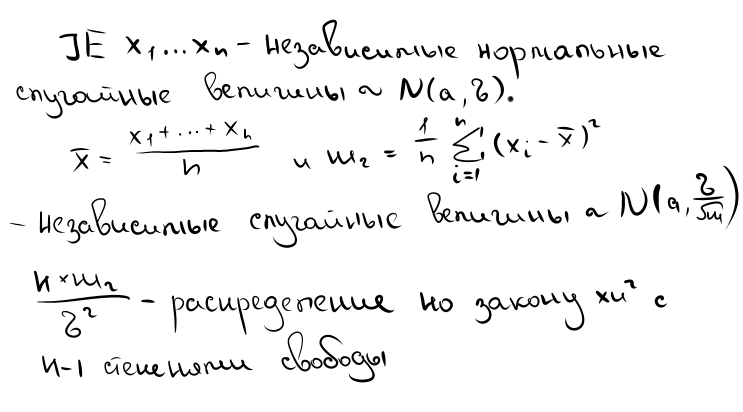
Выборочный коэффициент эксцесса ([несмещенная оценка](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0)) определяется формулой:



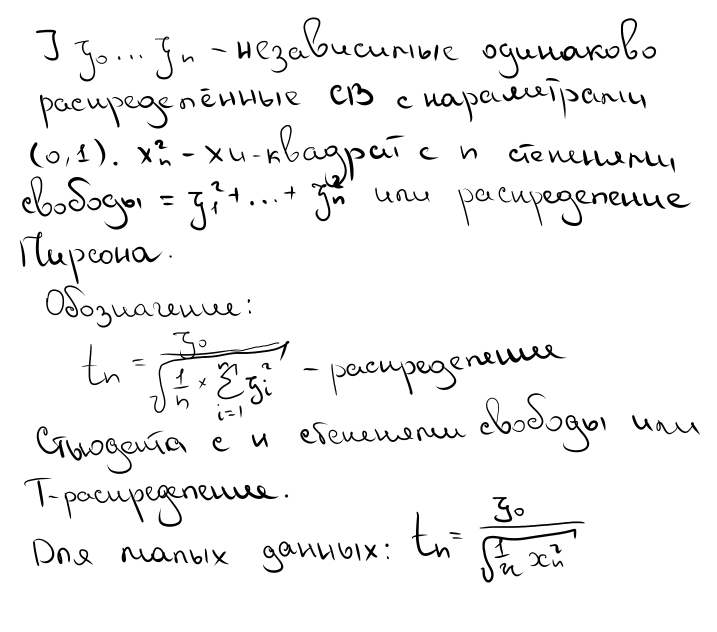
Где:

*  - выборочный центральный момент k-го порядка;
*  - [несмещённая оценка](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) выборочного центрального момента второго порядка;
*  - [несмещённая оценка](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) выборочного центрального момента четвёртого порядка.

# 39 Распределение хи-квадрат с n степенями свободы.

****

**40 Распределение Стьюдента с n степенями свободы.**



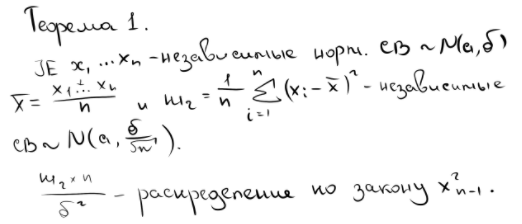
# 41 Распределение Фишера с n1 и n2 степенями свободы.

Пусть xn12 и xn22 - независимые случайные величины, имеющие хи-квадрат распределение с n1 и n2 степенями свободы)

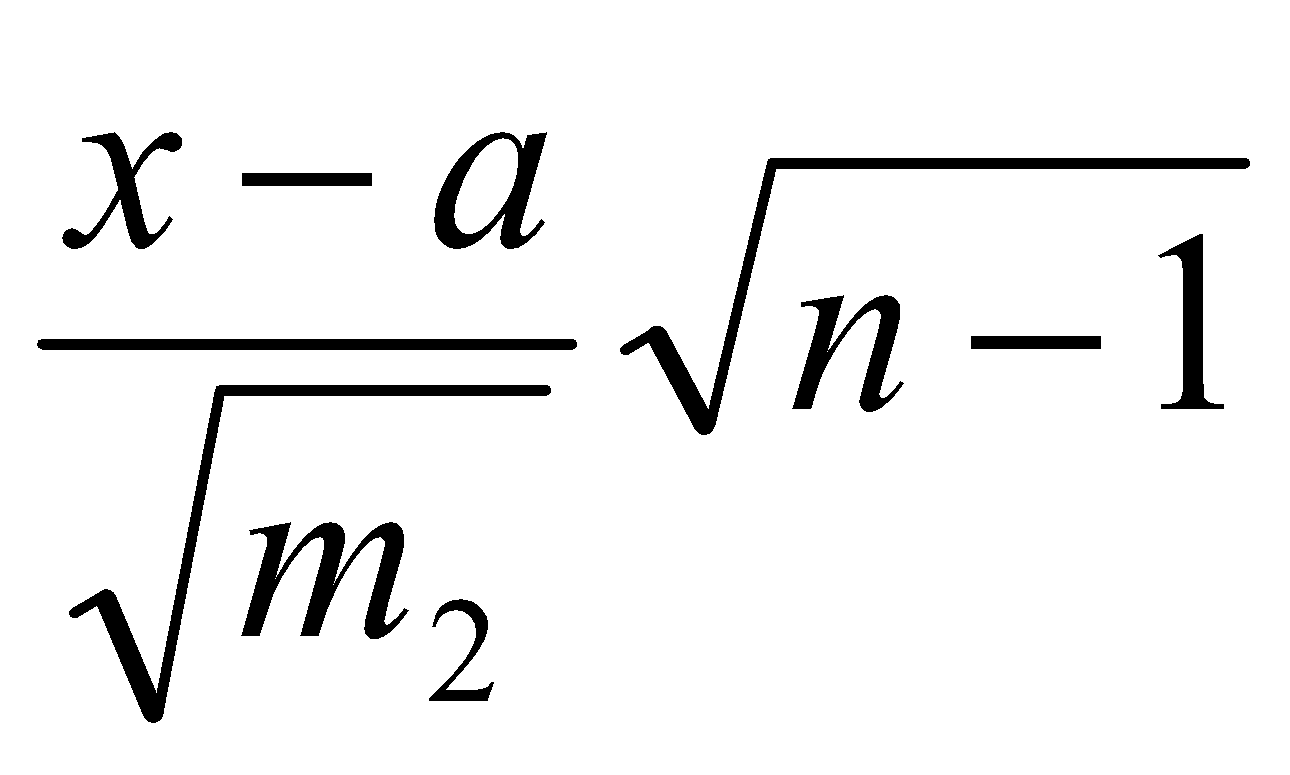
Распределение Фишера с n1 и n2 степенями свободы будет:

Fn,m = (xn12/n1)/(xn22/n2)

# 42 Теорема Фишера

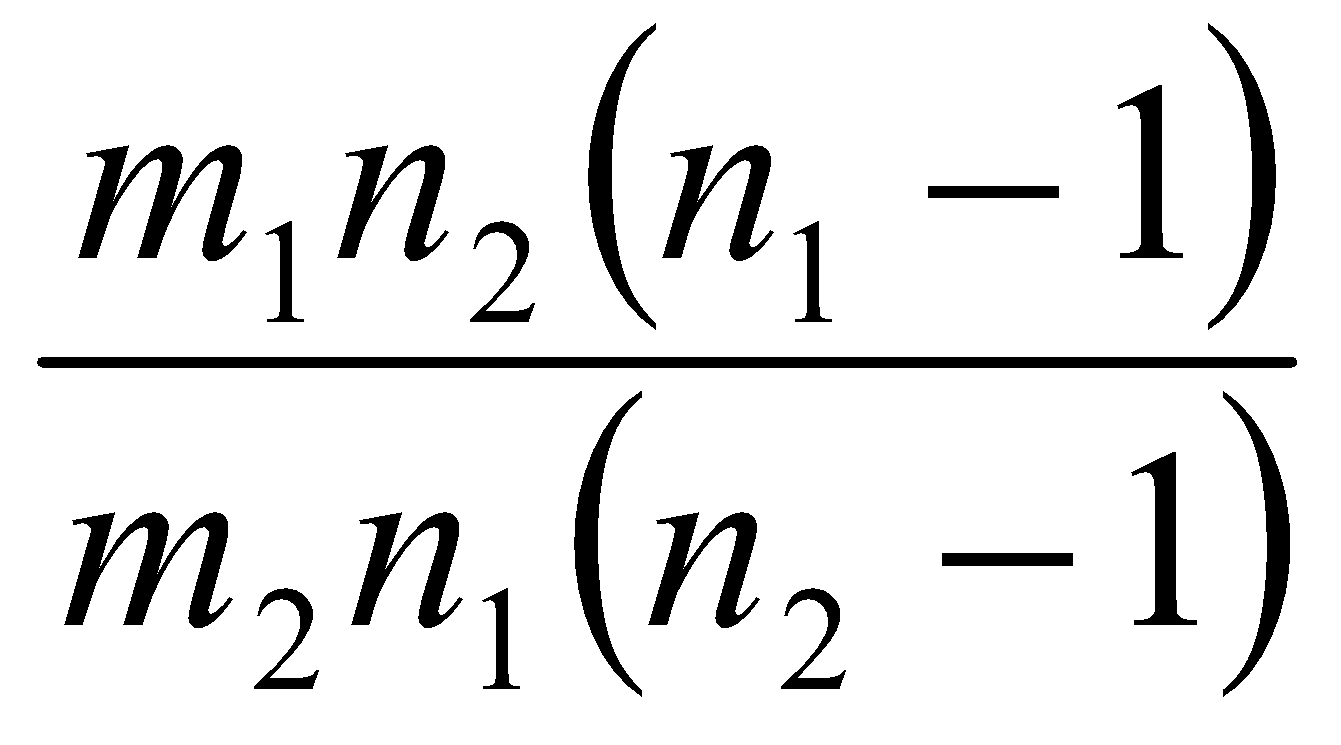
****

# 43 Теорема о распределении случ. велич.

 - имеет распределение стьюдента с n-1 степенью свободы.

a - матожидание, m2 - дисперсия, n - размер выборки

# 44 Теорема о распределении случ. велич.

 - имеет распределение Фишера с , степенями свободы.

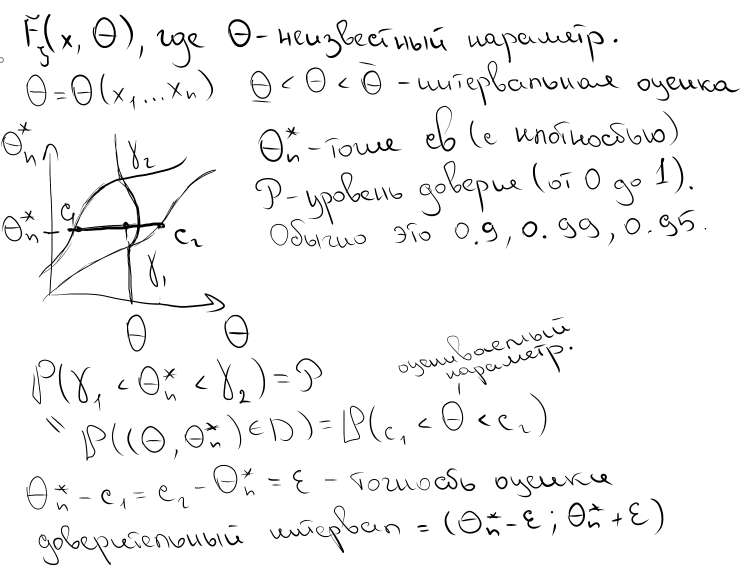
m1 и m2 - дисперсии, n1 и n2 - размеры выборок

# 45 Интервальные оценки, доверительный интервал

Рассмотрим выборку X из ген. совокупности . Пусть для нее задана статистическая параметрическая модель

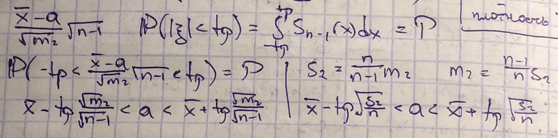
Введем такие статистики 1(X) и 2(X), чтобы с вероятностью выполнялось равенство: P{1(X) 2(X)} = . В этом случае говорят об интервальной оценке для параметра . Интервал (1(X); 2(X)) - называют доверительным интервалом, а коэффициент - уровнем доверия.

**46 Метод Неймана построения доверительных интервалов**

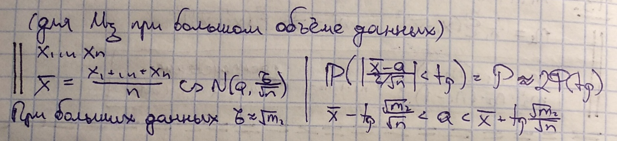


# 47 Доверительный интервал для мат. ожидания при известной дис­персии нормально распределенной генеральной совокупности.

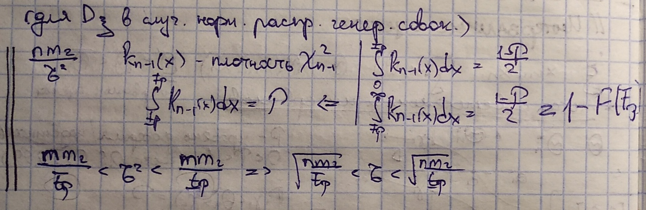
# 48 Доверительный интервал для мат. ожидания при неизвестной дисперсии нормального распределения генеральной совокупности

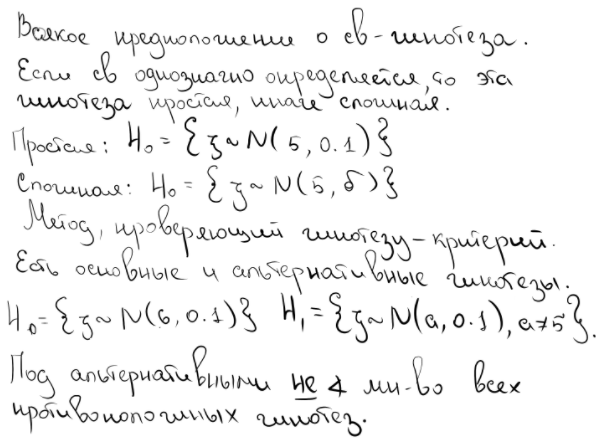


# 49 Построение доверительного интервала для мат. ожидания в случае большой выборки

****

# 50 Построение доверительного интервала для среднеквадратичного отклонения и дисперсии нормального распределения генеральной совокупности.

****

**51 Статическая гипотеза. Простая гипотеза. Основная и альтернативная гипотезы.** 

# 52 Критерий согласия

Критерий согласия - проверка гипотезы о согласии имеющихся статистических данных с выбранным теоретическим распределением. В качестве конкурирующей гипотезы могут выступать различные распределения.

Для проверки можно использовать критерий хи-квадрат Пирсона.

# 53 Мера отклонения и различные критерии

Мера отклонения - количественная разница между ожидаемым значением для критерия и действительным для выборки.

Например, в критерии Колмогорова мера отклонения (Dn) - расстояние между теор. и эмп. функциями распр.

# 54+55 Критерий хи квадрат в случае простой гипотезы, и когда по выборке оцениваются параметры

Различие между простой и сложной гипотезой - степень свободы, в случае параметров степень свободы будет меньше.

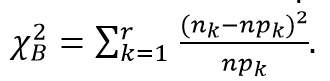
Простая гипотеза - гипотеза о согласии, гипотеза с оцениванием параметров - когда мы проверяем согласие с некоторым семейством распределений

1. Найти число параметров, обозначить их как *l*.

2. Если X-дискретная случайная величина, то определить частоты nk, k=1…r. Если X - непрерывная случайная величина, то разбить область значений на r интервалов Δ1…Δr и определить число элементов выборки nk принадлежащие каждому интервалу.

3. Если X - дискретная случайная величина, используя закон распр. F(x), вычислить вероятности pK с которыми случайная величина X принимает каждое из значений. В случае, если X - непрерывная случайная величина, определяется вероятность попадания в каждый интервал.

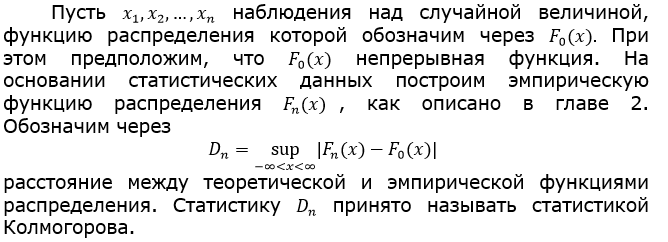
4. Вычислить статистику критерия



5. Сравнить полученное значение с критическим, определить противоречат ли опытные данные гипотезе.

(Замечание, все интервалы должны иметь длину не менее 9, npk 9)

# 56 Критерий Колмогорова



Критерий Колмогорова - проверка гипотезы о принадлежности выборки некоторому закону распределения.

Строится эмпирическая функция распределения выборки F(x), теоритическая функция распределения выборки Fз(x).

Находятся значения Dm,n+ и Dm,n- (для всех элементов выборки):

Dm,n+ = значение Fз(x) - значение F(x) (на правом конце соотв. отрезка)

Dm,n- = значение F(x) (на левом конце соотв. отрезка) - значение Fз(x)

Затем находится значение D: D=max(Dm,n+, Dm,n-)

Находим статистику: S =

По таблице находим критическое значение с необоходимым уровнем доверия. Если найденное значение не превышает критическое, значит данные не противоречат гипотезе о принадлежности выборки некоторому закону распределения

# 57 Критерии однородности и независимости

Критерий однородности - выборки распределены по одному закону

Гипотеза - H0 = {F1(x) = F2(y)}, где x(x1, … xn) и y(y1, … yn) - серии наблюдений над случайными величинами, а F1(x) и F2(y) - их функции распределения.

Критерий независимости - выборки принадлежат к одной и той же серии наблюдений

Гипотеза - H0 = {F(x, y) = F1(x)F2(y)}, где F1(x), F2(y) - одномерные функции распределения, а (x, y) - двумерная случайная величина.

# 58 Критерий Колмогорова-Смирнова

Строится эмпирическая функция распределения первой выборки F(x), для второй выборки строим функцию G\*(y) (эмп. функцию распр. второй выборки относительно эмп. функции распр. первой выборки)

Находятся значения Dm,n+ и Dm,n- (для всех элементов 2 выборки):

Dm,n+ = G(y) - значение F(x) (на правом конце соотв. отрезка)

Dm,n- = значение F(x) (на левом конце соотв. отрезка) - G(y)

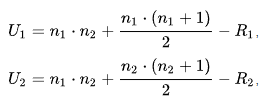
Затем находится значение D: D=max(Dm,n+, Dm,n-)

Находим статистику: S =

По таблице находим критическое значение с необоходимым уровнем доверия. Если найденное значение не превышает критическое, значит опытные данные не противоречат гипотезе однородности.

**59 Критерий Вилкоксона-Манна-Уитни**

Для применения U-критерия Манна — Уитни нужно произвести следующие операции.

1. Составить единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг (при наличии повторяющихся элементов в выборке использовать средний ранг). Общее количество рангов получится равным N = n1 + n2, где n1 — количество элементов в первой выборке, а n2 — количество элементов во второй выборке.
2. Разделить единый ранжированный ряд на два, состоящих соответственно из единиц первой и второй выборок. Подсчитать отдельно сумму рангов, пришедшихся на долю элементов первой выборки R1, и отдельно — на долю элементов второй выборки R2, затем вычислить:
3. 
4. Определить значение U-статистики Манна-Уитни по формуле U = min{U1, U2}
5. Если значение U табличного, то принимается альтернативная гипотеза, иначе принимается нулевая гипотеза (не отличаются выборки).

**60 Статическая оценка коэффициента корреляции Пирсона**

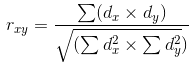
Оценка статистической значимости коэффициента корреляции осуществляется при помощи t-критерия, рассчитываемого по следующей формуле:



Значение tr сравнивается с крит. значением при определенном уровне значимости и числе степеней свободы n-2.

Если tr > tкрит, то делается вывод о статистической значимости выявленной корреляционной связи.

Расчет коэффициента корреляции Пирсона производится по следующей формуле:

, где (Величины отклонений) , X, Y - значения , M - среднее

**61 Коэффициент корреляции Спирмана**

Заданы две выборки 

Коэффициент корреляции Спирмена вычисляется по формуле:



где ,

Коэффициент \rho принимает значения из отрезка [-1;\;1]. Равенство \rho=1 указывает на строгую прямую линейную зависимость, \rho=-1 на обратную.

**62 Проверка гипотезы о независимости признаков на основании критерия Спирмана**

: связи между признаками нет, они независимы

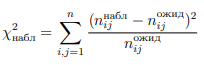
: связь между признаками есть, они не независимы

Чтобы сравнить набл и крит значение статистики критерия, необходимо определить ожидаемые частоты.

Общая формула расчета выглядит так: , где i и j – номер строки и столбца, в которых находится интересующее число n.

Наблюдаемые частоты мы берем из таблицы.

Статистика критерия имеет распределение хи-квадрат (x^2). Наблюдаемое значение статистики считается следующим образом:



Критическое значение: , где r – число строк в таблице сопряженности, c – число столбцов в таблице.

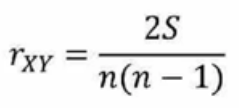
Если наблюдаемое значение меньше критического, то делаем вывод о том, что на уровне значимости 5% нет оснований отвергнуть нулевую гипотезу о независимости признаков.

**63 Коэффициент корреляции Кендалла**

Коэффициент корреляции Кенделла - мера линейной связи между случайными величинами.

Заданы две выборки x = (x_1,\ldots,x_n),\; y = (y_1,\ldots,y_n)

Коэффициент корреляции Кенделла вычисляется по формуле:

, где S = P - Q

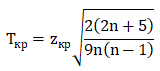
* P - суммарное число наблюдений, следующих за текущими наблюдениями с большим значением рангов Y
* Q - суммарное число наблюдений, следующих за текущими наблюдениями с меньшим значением рангов Y (равно не учит.)

Коэффициент \tau принимает значения из отрезка [-1;\;1]. Равенство \tau=1 указывает на строгую прямую линейную зависимость, \tau=-1 на обратную.

**64 Проверка гипотезы о независимости признаков на основании критерия Кендалла**

: связи между признаками нет, они независимы (r = 0)

: связь между признаками есть, они не независимы (r != 0)

1. Необходимо вычислить коэффициент ранговой корреляции Кендалла (формула прошлого вопроса)
2. Необходимо вычислить крит. точку: ,

где  – объем выборки;  – критическая точка двусторонней критической области, которую находят [по таблице функции Лапласа](https://100task.ru/sample/119.aspx)

Если  – нет оснований отвергнуть нулевую гипотезу. Ранговая корреляционная связь между признаками незначимая. Если  – нулевую гипотезу отвергают. Между признаками существует значимая ранговая корреляционная связь.

**65 Задача регрессионного анализа**

Регрессионный анализ - инструмент для количественного определения значения одной переменной на основании другой.

Цели:

1. Определение степени детерминированности вариации критериальной (зависимой) переменной предикторами (независимыми переменными)
2. Предсказание значения зависимой переменной с помощью независимой(-ых)
3. Определение вклада отдельных независимых переменных в вариацию зависимой

Простейшая модель регрессии – линейная регрессия. Линейная регрессия для x и y : 

Задача регрессионного анализа сводится к поиску коэффициентов a и b:

, где r - коэф кор Пирсона, а - стандартные отклонения для X и Y.

, так как

Зависимая переменная - пер. в регрессии, которую нельзя изменить.

Независимая переменная - которую можно изменить.

Коэффициенты регрессии - коэф, которые рассчитываются в результате регрессионного анализа.

# 66 Статическая оценка параметров линейной зависимости между случайными величинами

Метод наименьших квадратов:

Пусть имеются n наблюдений признаков x и y, известен вид уравнения регрессии f(xi)

Находятся такие значения параметров a и b, которые смогут минимизировать сумму квадратов отклонений фактических знач. yi от теор. знач. f(xi) для всех наблюдений i=1:n

⇒

Если случайные величины X и Y связаны линейной корр. зависимостью, то уравнение регресси Y на X имеет вид: y = B0 + B1x, уравнение X на Y: x = B’0+B’1x

Коэф. уравнений находятся методом наименьших квадратов, имеют вид:

B0 = my - B1mx, B1 = pxyy/x, B’1 = pxyx/y, B’0 = mx - B’1my,

pxy - коэф. корреляции, m - мат ожидание, 2 - дисперсия случайных величин

Предварительное представление о зависимости между *Х* и *Y* можно получить, отмечая элементы выборки (X, Y) в виде точек на плоскости с выбранной системой координат. Такое представление выборки системы двух случайных величин называется диаграммой рассеивания.