ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Теория вероятностей и математическая статистика

Направление подготовки (специальность): Экономика
Профиль образовательной программы: Экономика предприятий и организаций
Форма обучения: очная

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция № 1 (4 часа).

Тема: «Случайные события. Вероятность события» (интерактивная форма)

1.1.1 Вопросы лекции:

1. Основные понятия теории вероятностей
2. Классическое определение вероятности
3. Элементы комбинаторики
4. Геометрическое определение вероятности

5 Статистическое определение вероятности

1.1.2 Краткое содержание вопросов:

1. Основные понятия теории вероятностей

Теория вероятностей является основой математической статистики. ***Математическая статистика*** *– раздел математики, изучающий математические методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления статистических закономерностей.*

Теория вероятностей *– это математическая наука, изучающая закономерности случайных явлений (событий величин, функций, процессов и др.)*

Предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.

Одним из основных понятий теории вероятностей является понятие случайного события.

Случайное событие *– это любой факт, который может либо произойти, либо не произойти при выполнении некоторого комплекса условий.*

Обычно случайные события обозначаются заглавными латинскими буквами: A, B, C.

Геометрически случайные события удобно изображать с помощью диаграмм Эйлера-Венна.

Виды событий: ***достоверные*** (Ω); ***невозможные*** (Ø); - А является частным случаем В; ***равносильные; несовместные*** (***несовместимые***); ***совместные*** (***совместимые; равновозможные.***

Операции над событиями

Суммой событий А и В называют событие **А+В**, состоящее в наступлении хотя бы одного из этих событий А или В.

Произведением событий А и В называют событие А·В, состоящее в одновременном наступлении этих событий А и В.

Событие $\overline{A}$ называется противоположным событием (дополнением) события А, если непоявление одного события влечет появление другого. Сумма противоположных событий есть событие достоверное, а произведение невозможное:

События $\{H_i\}_{i=1}^{n}$ образуют полную группу попарно несовместимых событий, если любые два из них несовместны и хотя бы одно непременно должно произойти в результате испытания:

$\sum_{i=1}^{n} H_i = \Omega$ - полнота,

$H_i \cdot H_j = \varnothing,\ \forall i \neq j$ - попарная несовместимость.

2. Классическое определение вероятности

Вероятность события — это численная мера объективной возможности его появления.

В соответствии с классическим определением:

Вероятность P(A) события A равняется отношению числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу*:*

$$P(A) = \frac{m}{n}. \qquad (*)$$

При этом полагают, что:

- испытание содержит конечное число исходов;
- все исходы испытания равновозможны и несовместимы.

Такого рода опыт называют «схемой случаев» (или «схемой урн», так как любую подобную вероятностную задачу можно свести к задаче с урнами с разноцветными шарами).

Из классического определения вероятности вытекают следующие очевидные ***свойства вероятности события***.

3. Элементы комбинаторики

Комбинаторика — это раздел математики, изучающий методы решения задач на подсчет числа различных комбинаций.

В комбинаторике есть два важных правила, часто применяемых при решении комбинаторных задач.

1. Правило умножения комбинаторики.

2. Правило сложения комбинаторики.

m-элементные подмножества (комбинации) могут отличаться:

- составом элементов;
- порядком следования элементов;
- возможностью повтора элементов в подмножестве;
- объемом подмножества.

В соответствии с этим выделяют следующие виды подмножеств.

1. Размещения

Число всех размещений A_n^m из *n* элементов по *m* (где $m < n$), определяется по формуле:

$$A_n^m = \frac{n!}{(n-m)!}$$

2. Перестановки

Число всех перестановок P_n из n элементов определяется по формуле:

$$P_n = n!$$

Заметим, что перестановки — это частный вид размещений, когда $n = m$*:*

$$P_n = A_n^n$$

3. ***Сочетания***

Число всех сочетаний C_n^m из *n* элементов по *m* (где $m < n$)*,* определяется по формуле:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Рассмотренные выше комбинации относятся к так называемому **выбору без возвращения** — входящие в их состав элементы не повторяются.

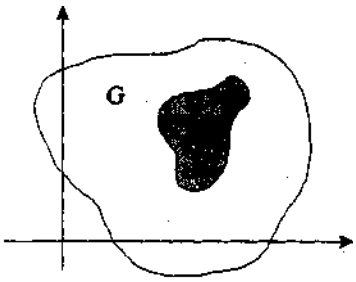
Кроме того, комбинаторика рассматривает и случаи с повторением элементов, входящих в рассматриваемые подмножества — так называемый **выбор с возвращением** — размещения, сочетания и перестановки из *n* элементов по *m,* в которых некоторые элементы (или все) могут быть одинаковыми.

Рассматривая конкретную задачу, необходимо выяснить, каким требованиям удовлетворяют комбинации элементов. Только после этого можно использовать нужные вычислительные формулы, комбинируя их с правилами сложения и умножения комбинаторики.

4. Геометрическое определение вероятности

Классическое определение вероятности основывается на том число всех возможных случаев конечно.

Если распределение возможных исходов испытания непрерывно и бесконечно, то при решении задач используется понятие **геометрической вероятности** — вероятности попадания точки в область (отрезок, часть плоскости и т. д.).



При определении геометрической вероятности полагают, что имеется область G и в ней меньшая область g с квадрируемой границей. На *G* наудачу бросает точка. Событие *A* — попадание точки в область *g*. Вероятность попадания в какую-либо часть *G* пропорциональна мере этой части (обозначим *mes)* и не зависит от ее расположения и формы.

Геометрической вероятностью события *A* называется отношение меры области g, благоприятствующей событию *A* к мере всей области *G:*

$$P(g) = \frac{mes(g)}{mes(G)}.$$

Область, на которую распространяется геометрическая вероятность может быть:

- **одномерной;**
- **двумерной;**
- **трехмерной;**
- ***n*-мерной.**

При этом вероятность попадания случайно взятой точки в область размерности меньшей, чем n, например, в границу области, равна нулю.

5. Статистическое определение вероятности

По данным наблюдений рассчитывают отношение $w_A = \frac{m_a}{n}$, называемое **частостью (относительной частотой,** выборочной долей) события А, где m_A называют частотой события *A.*

Статистической вероятностью события *A* называется частость (относительная частота) *m/n* появления этого события в n произведенных испытаниях:

$P(A) = w = \frac{m}{n}$,

Наблюдаемая частость события *A* почти для любой большой серии указанных испытаний мало отклоняется от некоторой постоянной величины, т. е. проявляется *закон устойчивости* частостей. Поэтому за вероятность случайного события А принимают на практике либо наблюденную частость, либо число, близкое этому значению, т.е. $P(A) \approx w_A$.

1.2 Лекция № 2 (4 часа).

Тема: «Теоремы сложения и умножения вероятностей» (интерактивная форма)

1.2.1 Вопросы лекции:

1. Теорема сложения для несовместных событий
2. Теорема сложения для совместных событий
3. Теоремы умножения вероятностей
4. Формула полной вероятности
5. Формула Байеса

1.2.2 Краткое содержание вопросов:

1. Теорема сложения для несовместных событий

Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$P(A+B) = P(A) + P(B).$

Теорема сложения для n несовместных событий

Вероятность появления одного из нескольких попарно несовместных событий равна сумме вероятностей этих событий.

$$P(A_1 + A_2 + A_3 + \ldots + A_n) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n)$$

$$\text{или } P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)\,.$$

Рассмотрим важные следствия из теоремы сложения для несовместных случайных событий.

Следствие 1.

Сумма вероятностей событий $H_1, H_2, \ldots, H_n$, образующих полную группу попарно несовместных событий, равна 1.

$$P(H_1) + P(H_2) + \ldots + P(H_n) = 1.$$

Следствие 2.

Сумма вероятностей противоположных событий равна 1.

$$P(A) + P(\bar{A}) = 1$$

Это следствие — частный случай следствия 1, так как противоположные события образуют полную группу попарно несовместных событий и сумма их есть событие достоверное.

Выделение этого свойства в отдельное следствие объясняется, прежде всего, большой практической значимостью приведенной формулы.

2. Теорема сложения для двух совместных событий

Вероятность суммы двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного наступления.

$$P(A+B) = P(A) + P(B) - P(AB).$$

Рассуждая аналогично, можно доказать справедливость формулы для суммы трех совместных событий.

Теорема сложения для трех совместных событий

Вероятность суммы трех совместных событий равна

$$P(A+B+C)=P(A)+P(B)+P(C) - P(AB) - P(BC)-P(AC)+P(ABC).$$

Методом полной индукции можно доказать общую формулу для вероятности суммы любого числа совместных событий.

Теорема сложения для n совместных событий

Вероятность суммы n совместных событий равна:

$$P(\sum_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i,j} P(A_i \cdot A_j) + \sum_{i,j,k} P(A_i \cdot A_j \cdot A_k) + ... + (-1)^{n-1} P(A_1 \cdot A_2 \cdot A_3 \cdot ... \cdot A_n)$$

Один из наиболее эффективных приёмов, используемых для вычисления вероятности суммы нескольких совместных событий, — переход к событию противоположному и вычисление его вероятности, а затем применение следствия 2 из теоремы сложения для несовместных событий: $P(A) + P(\bar{A}) = 1$.

Переходя к противоположному событию, и используя следствие и формулу Моргана, можно записать, что:

вероятность суммы n совместных событий:

$P(A_1 + A_2 + A_3 + ... + A_n) = 1 - P(\overline{A_1 + A_2 + ... + A_n}) = 1 - P(\bar{A}_1 \cdot \bar{A}_2 \cdot ... \cdot \bar{A}_n)$

3.Теоремы умножения вероятностей

События А и В называются **зависимыми**, если вероятность одного из них зависит от наступления или не наступления другого. Иначе случайные события называются независимыми.

Несовместные события зависимы, так как появление любого из них обращает в нуль вероятности появления всех остальных.

В случае зависимых событий вводится понятие условной вероятности.

Условной вероятностью P(A|B) события А называется его вероятность, вычисленная при условии, что событие В произошло. Аналогично, через P(B|A) обозначается условная вероятность события В при условии, что А наступило.

Для независимых событий по определению P(A|B) = P(A); P(B|A) = P(B).

Теорема умножения для зависимых событий

Вероятность произведения зависимых событий равна произведению вероятности одного из них на условную вероятность другого, при условии, что первое произошло.

$P(A\cdot B) = P(A)\cdot P(B|A)$ или $P(A\,B) = P(B)\cdot P(A|B)$ (*)

(в зависимости от того, какое событие произошло первым).

Следствие 1.

Теорема умножения для независимых событий.

Вероятность произведения независимых событий равна произведению их вероятностей.

P(AB) = P(A) • P(B).

Следствие 2.

Если А и В независимы, то независимы и пары: (А; $\bar{B}$), ($\bar{A}$;В), ($\bar{A}$; $\bar{B}$).

Следствие 3.

Для n независимых испытаний, в каждом из которых случайное событие А может появиться с вероятностью Р(А) = p, вероятность появления А хотя бы один раз равна:

$$P(B) = 1 - (1 - p)^n.$$

Следствием основных теорем теории вероятностей — теорем сложения и умножения вероятностей — является формула полной вероятности.

4. Формула полной вероятности

Пусть рассматривается полная группа попарно несовместных событий $\{A_i\}_{i=1}$ т.е. выполняются условия $A_1 + A_2 + ... + A_n = \Omega$, $A_i \cdot A_J = \varnothing$, $\forall i \neq j$, и некоторое событие В, которое может осуществиться одновременно только с одним из A_i.

Говорят еще, что об обстановке проведения опыта можно сделать n исключающих друг друга предположений A_i, называемых *гипотезами.*

Формула полной вероятности

Теорема. Вероятность P(B) события B, которое может произойти только при условии появления одного из событий (гипотез) A_1, A_2, … , A_n, образующих полную группу попарно несовместных событий, равна сумме произведений вероятностей каждого из событий A_1, A_2, … , A_n на соответствующие условные вероятности события B:

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B \mid A_i)$$

5. Формула Байеса

Если уже наступило рассматриваемое некоторое событие В, происходящее с одним из событий $\{A_i\}_{i=1}$, образующих полную группу попарно несовместимых событий, причем известны вероятности этих гипотез до испытания $P(A_i)$, а также вероятности, сообщаемые ими событию В: $P(B/A_i)$, то можно рассчитать вероятности гипотез A_i после того, как событие В произошло.

Формула Байеса

Вероятность $P(A_i /B)$, гипотезы A_i , при условии, что событие В произошло:

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

Вероятности гипотез до испытания *$P(A_i)$* называют еще *априорными (доопытными),* а вероятности гипотез *$P(B/A_i)$,* после того как произошло событие *B,* называют *апостериорными (послеопытными).'*

Формула Байеса, таким образом, дает возможность «пересмотреть» вероятности гипотез с учетом наблюденного результата опыта, по мере получения новой информации. Это имеет большое научно-практическое значение.

1.3 Лекция № 3 (2 часа).

Тема: «Повторные независимые испытания» (интерактивная форма)

1.3.1 Вопросы лекции:

1. Формула Бернулли
2. Локальная теорема Муавра-Лапласа
3. Интегральная теорема Муавра-Лапласа
4. Теорема Пуассона

1.3.2 Краткое содержание вопросов:

1. Формула Бернулли

Определение: Если производятся многократные испытания, в которых вероятность появления события А в каждом испытании не меняется в зависимости от исходов других испытаний, то такие испытания называют *независимыми повторными испытаниями.*

Будем далее рассматривать лишь такие независимые испытания, в которых событие А имеет одну и ту же вероятность.

Пусть производится n независимых испытаний, в каждом из которых событие А может появиться или не появиться. Обозначим вероятность события А – р, т.е. Р(А) = р, тогда вероятность ненаступления события А в каждом испытании также постоянна и равна q= 1-р.

Под схемой Бернулли понимают проведение серии в n испытаний, в каждом из которых возможны два исхода: либо наступит событие А, либо не наступит, т.е. произойдет противоположное ему событие, и при этом:

1) все n испытаний независимы;
2) вероятность события А в каждом отдельном испытании постоянна и не меняется от испытания к испытанию.

$$P(A) = p;\ P(\bar{A}) = 1-p = q.$$

Теорема. В случае небольшого числа испытаний n вероятность P_n(k) того, что в n независимых испытания событие А наступит ровно k раз определяется в соответствии с *формулой Бурнулли*:

$$P_n(k) = C_n^k p^k (1-p)^{n-k}$$

или

$$P_n(k) = \frac{n!}{k! \cdot (n-k)!} p^k (1-p)^{n-k}.$$

2. Локальная теорема Муавра – Лапласа

Теорема. При большом числе испытаний n→∞ , вероятности наступления события А в каждом испытании р, отличной от нуля и единицы, и при выполнении условия npq ≥ 20, вероятность P_n(k) того, что в n независимых испытаниях событие А наступит ровно k раз определяется в соответствии с *локальной теоремой Муавра – Лапласа*:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} f(t);\quad t = \frac{k-np}{\sqrt{npq}},$$

где n – число испытаний Бернулли;
k – число испытаний, в которых наступило событие А;
р = Р(А) – вероятность наступления события А в каждом испытании;
q= 1-р – вероятность противоположного события ($\bar{A}$);

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$ - функция Гаусса

Функция Гаусса f(t) представляет собой плотность стандартного нормального закона распределения (будет рассмотрена далее).

Основные свойства f(t), необходимые для применения рассматриваемой теоремы:

1) f(t) – четная функция, т.е. f(t) = f(-t);
2) f(t) – монотонно убывающая функция, т. е. f(t)→0 при t→±∞; при t>5 можно считать f(t) ≈ 0

3. Интегральная теорема Муавра – Лапласа

Теорема. При большом числе испытаний n→∞, вероятности наступления события А в каждом испытании р, отличной от 0 и 1, и при выполнении условия npq ≥ 20, вероятность P_n(k) того, что в n независимых испытаниях событие А наступит от а до b раз, определяется в соответствии с интегральной теоремой Муавра – Лапласа:

$$P_n(a \le k \le b) \approx \Phi(t_2) - \Phi(t_1);$$

$$t_1 = \frac{a - np}{\sqrt{npq}},\ t_2 = \frac{b - np}{\sqrt{npq}},$$

где n – число испытаний Бернулли;

k – число испытаний, в которых наступило событие А;

p = P(A) - вероятность наступления события А в каждом испытании:

q = 1-p – вероятность противоположного события ($\overline{A}$);

$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$ – функция Лапласа.

Функция Лапласа Ф(t) представляет собой функцию стандартного нормального закона распределения и будет более подробно рассмотрена в теме «Нормальный закон распределения».

Отметим основные свойства Ф(t), необходимые для применения данной теоремы:

1) Ф(t) – нечетная функция, т. е. Ф(-t) = - Ф(t).

2) Ф(t) - монотонно возрастающая функция, т. е. Ф(t)→1, при t→∞;при t>5 можно считать Ф(t)≈1 (или 0,5).

Задачи, приводящие к интегральной теореме Муавра – Лапласа

Пусть проводится n – независимых испытаний, в каждом из которых вероятность появления события А постоянна и равна p, 0<p<1.

1) <u>Поставим задачу</u>: найти вероятность того, что отклонение относительной частоты k/n от постоянной вероятности p по абсолютной величине не превышает заданного числа ε ≥ 0. Другими словами, найдем вероятность осуществления неравенства │k/n-p│≤ε.

$P\left(\left|\frac{k}{n} - p\right| \le \varepsilon\right) \approx \Phi(\varepsilon\sqrt{\frac{n}{pq}})$;

2) <u>Наименьшее число испытаний</u>, которое нужно провести, чтобы с вероятностью, равной β, можно было гарантировать, что частота наступления события А отклониться от вероятности p не более чем на ε:

$P\left(\left|\frac{k}{n} - p\right| \le \varepsilon\right) = \beta \to 2\Phi(\varepsilon\sqrt{\frac{n}{pq}}) = \beta \Rightarrow n = pq(\frac{\Phi^{-1}(\beta)}{\varepsilon})^2$

3) При данной вероятности β и числа испытаний n <u>границы возможных изменений</u> отклонения частоты наступления события А от вероятности p

$P\left(\left|\frac{k}{n} - p\right| \le \varepsilon\right) = \beta \to \Phi(\varepsilon\sqrt{\frac{n}{pq}}) = \beta \Rightarrow \varepsilon = \sqrt{\frac{pq}{n}}\,\Phi^{-1}(\beta);$

4) Вероятность того, что k наступлений события А <u>отличается от</u> произведения <u>np </u>(по модулю) не более чем на величину ε≥0

$P(|k-np| \le \varepsilon) \approx \Phi(\frac{\varepsilon}{\sqrt{npq}})$

4. Теорема Пуассона

Теорема. При большом числе испытаний n→∞, постоянной малой вероятности наступления события А в каждом испытании p→0, и при выполнении условия 0,1≤np≤10, вероятность $P_n(k)$ того, что в n независимых испытаниях событие А наступит ровно k раз, определяется в соответствии с теоремой Пуассона:

$P_n(k) \approx \frac{\lambda^k * e^{-\lambda}}{k!}$; λ=np,

где n- число испытаний Бернулли

k – число испытаний, в которых наступило событие А,

λ=np – параметр Пуассоновского распределения, называемый еще *средней интенсивностью*.

Значения функции Пуассона $P_n(k)$ также могут быть определены по таблице «Значение функции Пуассона» при заданных значениях k и λ.

1.4 Лекция № 4 (4 часа).

Тема: «Дискретная случайная величина (ДСВ)» (интерактивная форма)

1.4.1 Вопросы лекции:

1. Определение дискретной случайной величины. Закон распределения
2. Функция распределения ДСВ
3. Основные числовые характеристики ДСВ

4 Основные законы распределения ДСВ: биномиальный, Пуассона, геометрический

1.4.2 Краткое содержание вопросов:

1. Определение дискретной случайной величины. Закон распределения

Случайная величина – это переменная, которая в результате испытания принимает одно из своих возможных значений, причем заранее неизвестно, какое именно, т.к. оно зависит от случая.

Дискретная случайная величина – случайная величина, которая принимает конечное или бесконечное, но счетное число отдельных изолированных значений (т.е. их можно перенумеровать натуральными числами).

Непрерывная случайная величина – это случайная величина, бесконечное и несчетное множество значений которой есть некоторый интервал (конечный или бесконечный) и она сплошь заполняет этот интервал.

Для определения случайной величины необходимо задать ее закон распределения.

Законом распределения СВ называют соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими вероятностями, с которыми случайная величина принимает эти значения. (Др. словами, это соответствие м/у возможными значениями СВ и их вероятностями).

<u>Закон распределения ДСВ можно задать</u>:

1. таблично.

2. графическое представление ряда распределения называется *многоугольником (полиномом) распределения*

3. Третьим способом задания закона распределения ДСВ является интегральная функция распределения или функция накопленных вероятностей.

2. Функция распределения ДСВ

Функция распределения (интегральная функция)F(x) определяет для каждого возможного значения *x* вероятность того, что случайная величина X примет значения, меньшее *x*: F(x) = $\sum_{i=xi<x} p(X = xi)$.

<u>Свойства интегральной функции распределения ДСВ</u>

1) Функция распределения может принимать любые значения от 0 до 1, т.к. по определению является вероятностью;

2) Интегральная функция распределения является неубывающей;

3) Функция распределения любой дискретной случайной величины есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующим возможным значениям случайной величины и равны вероятностям

этих значений. Сумма всех скачков равна 1. Эта функция кусочно постоянна на интервалах, на которых нет ее значений.

4) Интегральная функция распределения ДСВ непрерывна слева

Условия непрерывности функции:

$$\lim_{\substack{x\to x0 \\ x>x0}}[F(x) - F(x_0)] = 0$$

$$\lim_{\substack{x\to x0 \\ x<x0}}[F(x) - F(x_0)] = P(X = x_0).$$

5) Вероятность попадания ДСВ в интервал [a;b) равна приращению функции распределения в этих точках:

$$P(a \le X \le b) = F(b) - F(a).$$

6) Если всевозможные значения случайной величины X принадлежат интервалу $(x_{min};\ x_{max})$, то

$$F(x) = 0, \text{ при } x \le x_{min}$$

$$F(x) = 1, \text{ при } x \ge x_{max}$$

7) Если всевозможные значения ДСВ X расположены на всей числовой оси OX , то

$F(-\infty) = \lim_{x\to-\infty} F(x) = 0$ (как вероятность невозможного события)

$F(+\infty) = \lim_{x\to+\infty} F(x) = 1$ (как вероятность достоверного события)

3. Основные числовые характеристики ДСВ

Математическое ожидание M(x) – это число, характеризующее среднее значение случайной величины X.

Математическим ожиданием ДСВ называют сумму произведений всех ее возможных значений на их вероятности, т. е.

$$M(x) = \sum_{i=1}^{n} x_i p_i = x_1p_1 + x_2p_2 + \cdots + x_np_n$$

Свойства математического ожидания

1) Математическое ожидание постоянной величины С – const равно этой величине:

$$M(C) = C.$$

2) Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX) = C\,M(X)$$

3) Математическое ожидание алгебраической суммы n случайных величин x_1, x_2, …x_n равно сумме математических ожиданий этих СВ:

$$M(X_1 \pm X_2 \pm \ldots \pm X_n) = M(X_1) \pm M(X_2) \pm \ldots \pm M(X_n).$$

4) Математическое ожидание произведения n случайных величин x_1, x_2, …x_n равно произведению математических ожиданий этих СВ:

$$M(X_1 * X_2 * \ldots * X_n) = M(X_1) * M(X_2) * \ldots * M(X_n).$$

Пусть производится n независимых испытаний, в каждом из которых вероятность появления события А постоянна и равна p.

5) Теорема: Математическое ожидание M(x) числа появлений события А в n независимых испытания равно произведению числа испытаний на вероятность появления события в каждом испытании:

$$M(x) = np.$$

Зная математическое ожидание мы не можем судить как значения СВ рассеяны вокруг M(X). Другими словами M(X) полностью не характеризует случайную величину. Рассмотрим показатель дисперсия.

Дисперсия характеризует разброс или рассеяние значений СВ около ее математического ожидания.

Дисперсия – это математическое ожидание квадрата отклонения СВ от ее математического ожидания:

$$D(x) = M[X - M(x)]^2.$$

Формула упрощенного вычисления дисперсии имеет вид:

$$D(x) = M(x^2) - (M(x))^2$$

Свойства дисперсии самостоятельно

Среднее квадратическое отклонение случайной величины определяется как корень квадратный из дисперсии:

$$\sigma(x) = \sqrt{D(x)}$$

Среднее кв. отклонение было введено как дополнительная характеристика рассеяния значений СВ вокруг ее математического ожидания и, в отличие от дисперсии, она имеет размерность, совпадающую с размерностью СВ.

Мода $M_o(X)$распределения – это значение СВ, имеющее наиболее вероятное значение.

Медиана $M_e(X)$ - это значение СВ, которое делит таблицу распределения на две части т.о., что вероятность попадания в одну из них равно 0,5

4. Основные законы распределения ДСВ: биномиальный, Пуассона, геометрический

а) Биномиальное распределение

Пусть производится n независимых испытаний, вероятность наступления события А в каждом испытании постоянна P(A_i) = p (ненаступления, соответственно, q=1-p)

Рассмотрим в качестве ДСВ X число появлений события А в этих испытаниях. Очевидно, что событие А в этих испытаниях может не появиться, появится 1,2,3….m,…,n раз. Так как случайное событие подчиняются схеме Бернулли, то вероятность наступления события А может быть вычислена по формуле Бернулли:

$$P_n(k) = C_n^k p^k q^{n-k} \quad (*)$$

Формула (*) – аналитическое выражение Биномиального закона.

Дискретная случайная величина имеет биномиальный закон распределения, если она принимает целочисленные значения 0,1,2,…n с вероятностями, вычисленными по формуле Бернулли.

Для ДСВ, описываемой биномиальным законом распределения:

$$M(x) = np;$$
$$D(x) = npq.$$

б) распределение Пуассона

Условия, что и при Биномиальном законе распределения, НО вероятность появления события А мала, p≤0,1.

Произведение np = λ – сохраняет постоянное значение.

Закон распределения Пуассона - закон распределения ДСВ X, представляющей собой число k наступления события А в заданном промежутке или пространства при заданной интенсивности np = λ.

ДСВ имеет закон распределения Пуассона с параметром λ, если она принимает целочисленные неотрицательные значения: 0,1,2,..m,..,n, … (бесконечное, но счетное множество значений) с вероятностями, вычисляемыми по формуме Пуассона:

$$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

λ=np – параметр распределения Пуассона.

Его еще называют *законом редких событий*.

M(x) = λ, D(x) = λ.

в) Геометрическое распределение

Пусть проводится n независимых испытаний, вероятность появления события А в каждом из них равна p(0<p<1). Испытания заканчиваются как только появится событие А. Таким образом, если событие А появилось в k-м испытании, то в k-1 оно не появлялось.

Обозначим Х – ДСВ – число испытаний, которое нужно провести до первого появления события А, в k испытании вероятность Х равна pq^{k-1}

ДСВ имеет геометрический закон распределения, если она принимает целочисленные значения 1,2,… с вероятностями, вычисленными по формуле геометрической прогрессии

$$P(X=k) = q^{k-1}p.$$

1.5 Лекция № 5 (4 часа).

Тема: «Непрерывная случайная величина (НСВ)» (интерактивная форма)

1.5.1 Вопросы лекции:

1. Функция распределения непрерывной случайной величины
2. Функция плотности вероятностей НСВ
3. Основные числовые характеристики НСВ
4. Основные законы распределения НСВ: равномерный, экспоненциальный, нормальный

1.5.2 Краткое содержание вопросов:

1. Функция распределения непрерывной случайной величины

НСВ - это случайная величина, бесконечное и несчетное множество значений, которой есть некоторый интервал (конечный или бесконечный) и она сплошь заполняет этот интервал.

Функция распределения НСВ X F(x) непрерывна в любой точке и имеет всюду (кроме, возможно, конечного числа точек) непрерывную производную.

Теорема. Вероятность любого отдельного взятого значения НСВ равно нулю.

Свойства интегральной функции распределения НСВ

1. Функция распределения может принимать значения от 0 до 1, т.к. определяется вероятностью

$$0 \le F(x) \le 1$$

2. Интегральная функция распределения является не убывающей

$$F(x_2) \ge F(x_1), \text{ если } x_2 > x_1$$

3. Для НСВ (согласно теореме)

$$P(x_1 \le X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X \le x_2) = P(x_1 < X < x_2)$$

4. Вероятность попадания НСВ в интервал [$x_1;x_2$) равна приращению функции распределения в этих точках:

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1)$$

5. Если все возможные значения СВ x принадлежат интервалу ($x_1;x_2$), то

$$F(x) = 0, \text{при } x \le x_1$$

$$F(x) = 1, \text{ при } x > x_2$$

6. Если все возможные значения НСВ x расположены на всей числовой оси OX, то

$$\lim_{x \to -\infty} F(x)=0$$

$$\lim_{x\to+\infty} F(x)=1.$$

2. Функция плотности вероятностей НСВ

Рассмотрим вероятность попадания случайной точки на элементарный участок [x; Δx] длины Δx НСВ X, имеющей непрерывную и дифференцируемую функцию распространения F(x) на этом участке.

По 4 свойству функции распределения:

$$P(x\le X< x+\Delta x)=F(x+\Delta x)-F(x).$$

Определим теперь отношение этой вероятности к длине участка, т.е среднюю вероятность, приходящуюся на единицу длины рассматриваемого участка, и рассмотрим придел при $\Delta x\to 0$

$$\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x).$$

Эта функция, характеризующая плотность, с которой распределяются значения НСВ в точке, и была названа функцией плотности распределения или функцией плотности вероятностей.

Плотность вероятности (плотностью распределения дифференциальной функцией) СВ x называется $f(x)$, являющаяся первой производной интегральной функции распределения

$$f(x)=F'(x)$$

Под элементом вероятности для СВ x понимается величина f(x)dx, с точностью до бесконечно малых величин высшего порядка отражающая вероятность попадания случайной точки x в элементарный отрезок dx, примыкающий к точке x.

Свойства функции плотностей вероятностей:

1. Функции плотности вероятностей принимает только неотрицательные значения как производная неубывающей функции распространения $F(x)$:

$$f(x)\ge 0$$

2. Вероятность попадания НСВ x в интервал x_1 до x_2 равна определенному интегралу от функции плотности вероятностей в этих пределах:

Зная плотность распределения, можно вычислить вероятность того, что НСВ примет значение, принадлежащее интервалу (x_1, x_2).

$$P(x_1\le X\le x_2)=F(x_2)-F(x_1)\overset{\text{по формуле Ньютона–Лейбница}}{=}\int_{x_1}^{x_2}F'(x)dx=\int_{x_1}^{x_2}f(x)dx$$

Таким образом, геометрически полученный результат можно истолковать так: вероятность того, что НСВ примет значение принадлежащий интервалу (a,b), равна площади криволинейной трапеции, ограниченной осью OX, кривой распределения $f(x)$ и прямыми $x=a$, $x=b$.

<u>**Определение**</u>: График плотности распределения называют кривой распределения.

Т.о. зная плотность распределения, можно найти функцию распределения.

3. Функция распространения НВС равна интегралу от функции плотности вероятностей в пределах от -∞ до x:

4. Интеграл в бесконечных приделах от функции плотности вероятностей равен 1. (как сумма вероятностей всех возможных значений СВ x):

Геометрически: вся площадь криволинейной трапеции ограниченная осью OX и кривой распределения равна 1.

3.Основные числовые характеристики НСВ

1. Математическое ожидание НСВ:

$$M(x)=\int_{-\infty}^{+\infty} x\, f(x)dx$$

2. Дисперсия НСВ:

$$D(x)=\int_{-\infty}^{+\infty} ((x-M(x))^2 f(x)dx$$

3. Среднее квадратичное отклонение:

$\sigma_x = \sqrt{D(x)}$

4. Основные законы распределения НСВ: равномерный, экспоненциальный, нормальный

Плотности распределения НСВ называют также <u>*законами распределений*</u>.

1) ***Равномерный закон распределения***

<u>Определение</u>: НСВ X имеет равномерный закон распределения на отрезке [a; b], если ее плотность вероятности f(x) постоянна на этом отрезке и равна нулю вне его, т.е.

$$f(x)=\begin{cases} 0, \text{ при } x < a \\ 1/(в-а), \text{ при } a \le x \le b \\ 0, \text{ при } x > b \end{cases} \qquad C= 1/(a+в)$$

По определению функции распределения и свойству плотности вероятности $F(x)=\int_{-\infty}^{x} f(t)dt$: $F(x)=\begin{cases} 0, \; при \; x < a \\ \dfrac{x-a}{b-a}, \; при \; a \le x \le b \\ 1, \; при \; x > b \end{cases}$

<u>*Числовые характеристики СВ*</u>, распределенной по равномерному закону распределения

$$M(x) = \frac{a+b}{2}$$

$$D(x)= \frac{(b-a)^2}{12}$$

2) **Показательный (экспоненциальный) закон распределения**

<u>Определение</u>: НСВ X имеет показательный (экспоненциальный) закон распределения с параметром λ >0, если ее плотность вероятности имеет вид:

$$f(x)=\begin{cases} \lambda e^{-\lambda x}, при \; x \ge 0 \\ 0, \; при \; x < 0 \end{cases}$$

Функция распределения СВ X, распределенной по экспоненциальному закону распределения

$$F(x)=\begin{cases} 1-e^{-\lambda x}, при \; x \ge 0 \\ 0, \; при \; x < 0 \end{cases}.$$

Числовые характеристики НСВ X, распределенной по экспоненциальному закону распределения:

$$M(x)=\frac{1}{\lambda}; \quad D(x)=\frac{1}{\lambda^2}.$$

Таким образом, математическое ожидание и среднее квадратичное отклонение показательного распределения равны между собой .

Замечание 1. Если на практике изучается показательно распределенная СВ, причем параметр λ неизвестен, то $\lambda \approx \frac{1}{\bar{x}}$ (и при этом неизвестно M(x), то принимают $M(x) \approx \bar{x}$).

Замечание 2. Если есть основания предположить, что НСВ X имеет показательное распределение, то находят $\bar{x}$ и σ и сравнивают их. Если оценки окажутся близкими одна к другой, то данные наблюдений подтверждают гипотезу о показательном распределении.

3) Нормальный закон распределения

Главная особенность нормального закона распределения состоит в том, что он является *предельным законом,* к которому с ростом числа наблюдений стремятся другие распределения.

Определение. НСВ X имеет нормальный закон распределения (закон Гаусса) с параметрами μ и σ (обозначают X ⊂ N(μ; σ)), если ее плотность вероятности имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ- математическое ожидание X;

σ^2 – дисперсия X;

σ – среднее квадратичное отклонение.

Свойства функции плотности вероятности нормального закона распределения:

1. f(x) >0 существует при любых действительных *x*;
2. $f(x) \xrightarrow{x \to \pm\infty} 0$;
3. Кривая плотности нормального закона распределения симметрична относительно прямой x=μ;
4. Максимальное значение f(x) принимает в т. x_0 =μ, при этом $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$;
5. Кривая плотности нормального закона распределения имеет 2 точки перегиба с координатами ($\mu \pm \sigma$; $\frac{1}{\sigma\sqrt{2\pi e}}$)

Если $\sigma = const$, и меняется параметр μ, т.е. центр симметрии распределения, то нормальная кривая будет смещаться: вдоль оси абсцисс, не меняя формы.

Если $\mu = const$ и меняется σ, то меняется ордината максимальная кривой - $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$.

Таким образом, параметр μ характеризует положение центра, а параметр σ - форму кривой плотности вероятности.

Функция распределения случайной величины, распределенной по нормальному закону.

Определение: Функция распределения СВ X, имеющее нормальный закон распределения с параметром μ и σ, определяется по формуле: $F(x) = \frac{1}{2} + \Phi(t),\ t = \frac{x-\mu}{\sigma}$

Ф(t)- функция Лапласа

Геометрически функция распределения представляет собой площадь под нормальной кривой на интервале $(-\infty; x)$. Как видим, она состоит из 2-х частей: первой - $(-\infty; \mu)$ равной ½ и второй, на интервале (μ;x), равной 1/2Ф(t).

Определение: Нормальный закон распределения СВ X с параметром $\mu = 0, \sigma = 1$ (N (0;1)) называется *стандартным или нормальным*.

Свойства СВ, распределенной по нормальному закону распределения

1. Если равна $X \subset N(\mu;\sigma)$, то вероятность попадания СВ X в интервал [x_1; x_2] равна

$$P(x_1 \le x \le x_2) = \Phi(t_2) - \Phi(t_1),$$

$$t_1 = \frac{x_1-\mu}{\sigma};\ t_2 = \frac{x_2-\mu}{\sigma}$$

2. Вероятность того, что отклонение случайной величины X, от её математического ожидания μ не превысит величину $\varepsilon > 0$ (по абсолютной величине), равна

$$P(|X-\mu| \le \varepsilon) = 2\Phi(\frac{\varepsilon}{\sigma}).$$

Если заменить ε на σ, 2σ, 3σ , то получим «правило трех сигм»

3. Правило трех сигм

Если случайная величина $X \subset N(\mu;\sigma)$, то практически достоверно, что её значения заключены в интервале $(\mu - 3\sigma;\ \mu + 3\sigma)$ (Вероятность выброса равна 0,0027)

1.6 Лекция № 6 (4 часа).

Тема: «Закон больших чисел. Понятие о методе Монте-Карло. Цепи Маркова» (интерактивная форма)

1.6.1 Вопросы лекции:

1. Сущность закона больших чисел. Основные теоремы закона больших чисел
2. Центральная предельная теорема Ляпунова
3. Сущность метода Монте-Карло. Оценка погрешности метода Монте-Карло
4. Правила разыгрывания ДСВ и НСВ

1.6.2 Краткое содержание вопросов:

1. Сущность закона больших чисел. Основные теоремы закона больших чисел

В широком смысле под законом больших чисел понимается свойство устойчивости массовых явлений, состоящее в том, что средний результат действия большого числа случайных явлений практически перестает быть случайным и может быть предсказан с достаточной определенностью.

В узком смысле под законом больших чисел понимают совокупность теорем, устанавливающих факт приближения средних характеристик, полученных по результатам большого числа наблюдений, к некоторым постоянным величинам.

Пусть дана последовательность случайных величин $\{X_i\}_{i=1}^{\infty} = X_1, X_2, \ldots, X_n, \ldots$; а случайные величины $\{Y_i\}_{i=1}^{\infty} = Y_1, Y_2, \ldots, Y_n, \ldots$- представляют собой заданные симметрические функции от первых n членов последовательности $\{X_i\}_{i=1}^{\infty} : Y_n = f_n(X_1, X_2, ..., X_n)$. Тогда если существует последовательность чисел $\mu_1, \mu_2, \ldots, \mu_n, \ldots$,такая, что для любого ε > 0 выполняется $\lim_{n\to\infty} P\{| Y_n - \mu_n |< \varepsilon\} = 1$, то говорят, что последовательность $\{X_i\}_{i=1}^{\infty}$ подчиняется закону больших чисел.

1. Лемма Маркова

Если случайная величина X не принимает отрицательных значений, то для любого положительного числа τ выполняется:

$$P(X \geq \tau) \leq \frac{M(X)}{\tau}$$

Так как события X ≥ τ и X< τ противоположные, то заменяя выражение P(X ≥ τ) выражением 1 - P(X < τ), придем к другой форме неравенства Маркова.

2. Неравенство Чебышева

Для любой случайной величины X, имеющей конечную дисперсию, при каждом ε > 0 имеет место неравенство:

$$P\{| X - M(X) |\leq \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

Учитывая, что события $| X - M(X) |\leq \varepsilon$ и $| X - M(X) |> \varepsilon$ противоположные, то неравенство Чебышева можно записать в другом виде

$$P\{| X - M(X) |> \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

Неравенство Чебышева применимо для любых случайных величин. В первом случае оно устанавливает *нижнюю* границу, во втором *верхнюю* границу вероятности рассматриваемого события.

Замечание. Если M(X) > τ или D(X) > ε^2, то правые части неравенств Маркова и Чебышева будут либо больше 1, либо отрицательными. Это означает, что применение указанных неравенств в этих случаях приведет к тривиальному результату: вероятность события меньше числа, превосходящего 1 или вероятность события больше отрицательного числа. Но такой вывод очевиден и без использования данных неравенств.

3. Теорема Чебышева (общий случай)

Если $X_1, X_2, \ldots, X_n, \ldots$— последовательность попарно независимых случайных величин, имеющих конечные дисперсии, ограниченные одной и той же постоянной С, т. е. $DX_1 < C$, $DX_2 < C$, ..., $DX_n < C$, ..., то при неограниченном увеличении числа n средняя арифметическая случайных величин сходится по вероятности к средней арифметической их математических ожиданий:

$$\lim_{n\to\infty} P\left\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)|< \varepsilon\right\} = 1$$

Смысл формулировки «*сходимость по вероятности*» означает, что с увеличением числа n вероятность неравенства $\left|\frac{1}{n}\sum X_i - \frac{1}{n}\sum M(X_i)\right| < \varepsilon$ стремиться к 1, т.е. это неравенство будет выполняться в подавляющем числе случаев, хотя в отдельном случае оно может и не выполняться.

Смысл теоремы Чебышева: при достаточно большом числе n случайных величин X_i практически достоверно, что их средняя $X = \frac{\sum X_i}{n}$ – величина случайная, как угодно мало отличается от неслучайной величины $\frac{\sum M(X_i)}{n}$, т.е. практически перестает быть случайной.

4. Теорема Чебышева (частный случай)

Если $X_1, X_2, \ldots, X_n, \ldots$ - последовательность наблюдений случайной величины X, имеющая конечную дисперсию, и одинаковое математическое ожидание то, каково бы ни было ε > 0:

$$\lim_{n\to\infty} P\left\{ \left| \frac{1}{n}\sum_{i=1}^{n} X_i - M(X) \right| < \varepsilon \right\} = 1$$

5. Теорема Бернулли

Пусть m — число наступления события А в серии n независимых испытаний, а p — есть вероятность наступления события в каждом из испытаний. Тогда частость события при неограниченном увеличении числа n сходится по вероятности к вероятности p этого события в отдельном испытании:

$$\lim_{n\to\infty} P\left\{ \left| \frac{m}{n} - p \right| < \varepsilon \right\} = 1$$

Смысл теоремы Бернулли: при большом числе n повторных независимых испытаний практически достоверно, что частость (статистическая вероятность) события m/n – величина случайная, как угодно мало отличается от неслучайной величины p – вероятности события, т.е. практически перестает быть случайной.

Теорема Бернулли дает теоретическое обоснование замены неизвестной вероятности события его частостью, или статистической вероятностью, полученной в n повторных независимых испытаниях, проводимых при одном и том же комплексе условий.

Непосредственным обобщение теоремы Бернулли является теорема Пуассона, когда вероятности события в каждом испытании различны.

6. Теорема Пуассона

Пусть m — число наступления события А в серии n независимых испытаний, а p_i — есть вероятность наступления события в i-м испытании. Тогда частость события при неограниченном увеличении числа n сходится по вероятности к средней арифметической вероятностей p_i события в отдельных испытаниях:

$$\lim_{n\to\infty} P\left\{ \left| \frac{m}{n} - \frac{1}{n}\sum_{i=1}^{n} p_i \right| < \varepsilon \right\} = 1$$

2. Центральная предельная теорема Ляпунова

Выше были рассмотрены различные формы закона больших чисел, которые, при всем своем разнообразии, утверждают одно: сходимость по вероятности тех или иных случайных величин к определенным постоянным.

Но этим не ограничиваются закономерности, возникающие в результате суммарного действия случайных величин. Оказывается, при некоторых условиях совокупное действие случайных величин приводит к нормальному закону распределения.

Все формы центральной предельной теоремы посвящены другому аспекту — установлению условий, при которых возникает самый распространенный в случайных явлениях нормальный закон распределения. Важнейшее место занимает теорема Ляпунова.

<u>Теорема Ляпунова</u>

Рассмотрим n независимых случайных величин $X_1, X_2, \ldots, X_n, \ldots$, удовлетворяющих условиям:

1) все величины имеют определенные математические ожидания и конечные дисперсии *D(X_i);*

2) ни одна из величин не выделяется резко среди остальных по своим значениям.

Тогда при неограниченном возрастании $n \to \infty$ распределение случайной величины — $\frac{1}{n}\sum_{i=1}^{n} X_i$, приближается к нормальному закону.

Таким образом, имеем следующую асимптотическую формулу:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < t\sqrt{\frac{\overline{D(X)}}{n}}\right) = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx = \Phi(t),$$

где $\overline{D(X_i)} = \frac{1}{n}\sum_{i=1}^{n} D(X_i)$

3. Сущность метода Монте-Карло. Оценка погрешности метода Монте-Карло

В 1949 г. Н. Метрополис и С. Улам систематизировано изложили метод Монте-Карло. Название метода связано с названием города Монте-Карло, где игорных домах игра в рулетку – одно из простейших устройств для получения случайных чисел, на использовании которой основан этот метод.

Случайными числами называют возможные значения r непрерывной случайной величины R, распределенной равномерно в интервале (0; 1).

<u>Сущность метода Монте-Карло</u>:

Требуется найти значение *a* некоторой изучаемой величины. Для этого выбирают такую СВ X, математическое ожидание которой равно *a*: M(X) = *a*.

Теория этого метода указывает, как наиболее целесообразно выбрать СВ X, как найти ее значения.

Отыскание возможных значений СВ X называют «разыгрыванием СВ».

На практике невозможно найти точную оценку математического ожидания.

Верхняя граница δ допускаемой ошибки с заданной вероятностью γ определяется исходя из равенства: $P(|\overline{X} - a| \le \delta) = \gamma$.

Возможны частные случаи определения δ.

4. Разыгрывание ДСВ и НСВ

Пусть требуется разыграть ДСВ X, т.е. получить последовательность ее возможных значений x_i (i=1,…, n), зная закон распределения X.

Правило: Для того, чтобы разыграть ДСВ, заданную законом распределения, надо:

1. разбить интервал (0; 1) оси or на n частичных интервалов: Δ_1 – (0; p_1); Δ_2 – (p_1; p_1+p_2); …; Δ_n – ($p_1+p_2+\ldots+p_n$; 1);
2. выбрать (например, из таблицы случайных чисел) случайное число r_j;
3. Если r_j попало в частичный интервал Δ_i, разыгрываемая дискретная случайная величина приняла возможное значение x_i.

<u>Разыгрывание НСВ</u>

Пусть требуется разыграть НСВ X, т.е. получить последовательность ее возможных значений x_i (i=1,…, n), зная функцию распределения F(x).

<u>Теорема</u>: Если r_i – случайное число, то возможное значение x_i, разыгрываемой НСВ X с заданной функцией распределения F(x), соответствующее r_i, является корнем уравнения $F(x_i)= r_i$.

Правило: Для того, чтобы найти возможное значение x_i НСВ X, зная ее функцию распределения F(x), надо выбрать случайное число r_i, приравнять его функции распределения и решить относительно x_i уравнение $F(x_i)= r_i$.

Приближенное разыгрывание НСВ, распределенной по:

а) равномерному закону распределения в интервале (a; b):

$$x_i = (b-a)\cdot r_i + a;$$

б) экспоненциальному закону распределения с параметром λ>0:

$x_i = -\frac{1}{\lambda}\ln(1-r_i)$ или $x_i = -\frac{1}{\lambda}\ln r_i$;

в) нормальному закону распределения с M(X) = a, D(X) = σ^2:

$$x_i = \sum_{i=1}^{12} r_i - 6 = S_i - 6.$$

1.7 Лекция № 7 (4 часа).

Тема: «Статистическое оценивание параметров распределения» (интерактивная форма)

1.7.1 Вопросы лекции:

1. Математическая статистика и ее методы
2. Общие сведения о выборочном методе
3. Понятие оценки параметров.
4. Методы нахождения оценок
5. Оценка параметров генеральной совокупности по собственно-случайной выработке
6. Интервальные оценки параметров генеральной совокупности

1.7.2 Краткое содержание вопросов:

1. Математическая статистика и ее методы

Математическая статистика занимается обработкой результатов эксперимента.

Методы математической статистики можно разделить на *описательные* и *аналитические*

Аналитические методы позволяют на основании выборочных наблюдений сделать статистически значимые выводы о наличии закономерностей для всей совокупности. Аналитические методы обычно основываются на соответствующих вероятностных моделях, предполагающих нормальное (или другое известное) распределение совокупности изучаемого признака, и составляют методы параметрической статистики.

Другим классом аналитических методов являются методы непараметрической статистики, которые не опираются на нормальное распределение (или любое другое) и не используют его свойства.

Методы математической статистики позволяют обоснованно осуществлять сбор и группировку статистических сведений и строить оптимальные математико-статистические модели массовых явлений, изменчивость которых обусловлена рядом факторов.

2. Общие сведения о выборочном методе

Объектом изучения математической статистики являются генеральные совокупности, которые исследуются на основе выборки. Задачи математической статистики практически сводятся к обоснованному суждению об объективных свойствах генеральной совокупности по результатам случайной выборки.

Под *генеральной совокупностью* понимается совокупность всех мыслимых наблюдений, которые могли бы быть произведены при данном реальном комплексе условий.

Понятие генеральной совокупности в определенном смысле аналогично понятию случайной величины (закону распределения вероятностей), так как полностью обусловлены определенным комплексом условий.

Выборочной совокупностью или выборкой называется та часть объектов, которая отобрана для непосредственного изучения из генеральной совокупности.

Число объектов в выборке или в генеральной совокупности называется их объемом.

Генеральная совокупность может иметь как конечный так и бесконечный объем.

Выборку можно рассмотреть как некий эмпирический аналог генеральной совокупности.

Для эффективного применения математико-статистических методов анализа необходимо, чтобы выборочная совокупность отражала структуру и особенности распределения признаков генеральной совокупности. Таким образом, выборка должна быть *репрезентативной*. Это достигается случайностью отбора, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными.

Ошибки репрезентативности всегда имеют место быть, однако они могут быть заранее оценены и сведены к минимуму посредством правильной организации выборки.

Различают следующие виды выборок: собственно-случайная, типическая, механическая, серийная.

Способы образования выборки: повторный отбор, бесповторный отбор.

Важнейшей задачей выборочного метода является оценка параметров (характеристик) генеральной совокупности по данным выборки.

Теоретическую основу применимости выборочного метода составляет закон больших чисел, согласно которому при неограниченном увеличении объема выборки практически достоверно, что случайные выборочные характеристики как угодно близко приближаются (сходятся по вероятности) к определенным параметрам генеральной совокупности.

3. Понятие оценки параметров

Пусть распределение признака X – генеральной совокупности описывается законом распределения F(x, Θ), который содержит известный параметр Θ. (например : λ, в случае распределения Пуассона или экспоненциальном распределении, или $\bar{x}$ и σ – нормального распределения). Для вычисления параметра Θ исследовать все элементы генеральной совокупности не предоставляется возможным, поэтому извлекается из генеральной совокупности выборка $x_1, x_2, \ldots, x_n$, и по ее результатам судят о параметре Θ.

Определение: Оценкой параметра Θ называют некоторую функцию результата наблюдений $\tilde{\theta}_n = \tilde{\theta}_n$ $(x_1, x_2, \ldots, x_n)$ с помощью которой судят о значении параметра Θ.

Поскольку $x_1, x_2, \ldots, x_n$ – случайные величины то и оценка $\tilde{\theta}_n$ (в от оцениваемого параметра Θ- величины неслучайной, детерминированной) является случайной величиной, зависящей от закона распределения случайной величины X и числа наблюдений n.

Основная задача теории оценивания состоит в том, чтобы произвести выбор оценки $\tilde{\theta}_n$ параметра Θ, позволяющей получить наилучшее приближение к оцениваемому параметру.

Например: если Θ - математическое ожидание СВ X т.е генеральной средней $\bar{x}_0$ то в качестве оценки $\tilde{\theta}_n$ можно взять выборочную среднюю, M_o , M_e , $(x_{min} + x_{max})/2$ и.т.д.

Назвать «наилучшей» оценкой такую, которая наиболее близка к истинному значению оцениваемого параметра, невозможно, т.к $\tilde{\theta}_n$ – случайная величина, поэтому невозможно предсказать индивидуальное значение оценки в данном частном случае. Так что о качестве оценки следует судить не по индивидуальным ее значениям, а лишь по распределению ее значений в большом количестве испытаний, т.е по выборочному распределению оценки.

Выбор той или иной функции в качестве оценки параметра Θ проводиться с учетом удовлетворения следующих требований.

Свойства точечных оценок :

1. *Несмещенность*. Статистическая оценка $\tilde{\theta}_n$ параметра называется несмещенной, если ее математическое ожидание равно оцениваемому параметру, т.е. $M(\tilde{\theta}_n)=\Theta$

Если оценка является смещенной, то смещение определяется как $B_n=M(\tilde{\theta}_n)=\Theta$

Требования несмещенности гарантирует отсутствие систематических ошибок при оценивании ($M(\tilde{\theta}_n)>\Theta$ или $M(\tilde{\theta}_n)<\Theta$

2. *Состоятельность*. Оценка $\tilde{\theta}_n$ параметра Θ называется состоятельной, если она удовлетворяет закону больших чисел, т.е сходиться по вероятности к оцениваемому параметру Θ :

$$\lim_{n\to\infty} P\left(\left|\tilde{\theta}_n - \theta\right| < 1\right) = 1$$

или

$$\tilde{\theta}_n \xrightarrow[n\to\infty]{} \theta$$

В данном случае оправдывается увеличение объема выборки.

3.*Эффективность*. Несмещенная оценка $\tilde{\theta}_n$ параметра Θ называется эффективной, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра Θ, вычисленных по выборкам одного и того же объема n.

В качестве статистических оценок параметров генеральной совокупности желательно использовать оценки, удовлетворяющие одновременно 3 –м требованиям. Однако достичь этого удается не всегда. Может оказаться, что для простаты расчетов целесообразно использовать незначительно смещенные оценки или оценки, обладающие большой дисперсией по сравнению с эффективными оценками.

Основные точечные оценки

Наиболее часто использующиеся точечные оценки предоставлены в таблице 1.

Таблица 1. – Точечные оценки основных параметров распределения

Оцениваемый параметр генеральной совокупности	Оценка его выборочная точечна	
	Простая выборка	Сгруппированные данные
Генеральная средняя или математическое ожидание μ	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum n_i}$
Генеральная		

дисперсия σ^2	$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$ $S^2=\overline{x^2}-(\bar{x})^2$	$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 n_i}{\sum n_i}$ $S^2=\frac{1}{n}\sum_{i=1}^{k} {x_i}^2 n_i-\bar{x}$
Исправленная дисперсия	$\hat{S}=\frac{n}{n-1}S^2$	
Генеральное среднее квадратическое отклонение σ	$S=\sqrt{S^2}$	$S=\sqrt{S^2}$
Начальные моменты l-го порядка	$\tilde{V}_k=\frac{1}{n}\sum_{i=1}^{n} {x_i}^l$	$\tilde{V}_k=\frac{1}{n}\sum_{i=1}^{k} {x_i}^l n_i$
Центральные моменты l-го порядка	$\tilde{\mu}_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^l$	$\tilde{\mu}_k=\frac{\sum(x_i-\bar{x})^l n_i}{\sum n_i}$

4. Методы нахождения оценок

1. Метод моментов, предложенный К. Пирсоном: определенное количество выборочных моментов (начальных $\tilde{V}_k$ или центральных $\tilde{\mu}_k$, или тех и других) приравнивается к соответствующим теоретическим моментам распределения (V_k или M_k) случайной величины X.

Оценки метода моментов обычно состоятельны, однако по эффективности они не являются «наилучшими», их эффективности e $(\tilde{\theta}_n)$ часто значительно ниже единицы. Метод моментов часто используются на практике, т.к приводит к сравнительно простым вычислениям .

2. Метод наибольшего правдоподобия (Р. Фишер)

Пусть $x_1, x_2, \ldots, x_n$ - случайная выборка из генеральной совокупности X, $f(x; \Theta_1;\Theta_2;\ldots \Theta_k)$ – функция вероятностей (для ДСВ) или плотность(для НСВ).

$\Theta_1;\Theta_2;\ldots \Theta_k$ параметры закона распределения , подлежащие оцениванию по случайной выборке.

МНП состоит в том, что *в качестве оценки* неизвестного параметра θ распределения СВ X выбирается то его значение, при котором *полученное значение* $x_1, x_2, \ldots, x_n$ *выборки имеет наибольшую вероятность* (в случае ДСВ) *или плотность* (в случае НСВ).

Функция правдоподобия наблюдений $x_1, x_2, \ldots, x_n$ представляет собой:

- <u>В случае ДСВ</u> – вероятность получить в качестве первого элемента выборки значение x_1 ,второго – значение x_2, … n-го значение x_n , т.е

$L(x_1, x_2, \ldots, x_n ; \Theta_1;\Theta_2;\ldots \Theta_k)=P(X_1=x_1\ X_2=x_2 \ldots X_n=x_n\ \Theta_1;\Theta_2;\ldots \Theta_k)$

т.к. наблюдения независимы, то $L(x_1,x_2,\ldots,x_n; \Theta_1;\Theta_2;\ldots\Theta_k) =\prod_{i=1}^{n} P(xi;\ \Theta1;\Theta2;\ldots\Theta k)$.

- <u>В случае НСВ</u> – совместная n мерная плотность вероятности, описывающая закон распределения вероятности n наблюдений.

Т.к. отдельные наблюдения независимы, то $L(x_1, x_2, \ldots, x_n ; \Theta_1;\Theta_2;\ldots \Theta_k) =\prod_{i=1}^{n} F(xi;\ \Theta1;\Theta2;\ldots\Theta k\ )$

За оценки наибольшего правдоподобия принимают такие значения $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k$, которые максимизируют функцию правдоподобия из условия:

$L(x_1, x_2, \ldots, x_n; \tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k) = \max L(x_1, x_2, \ldots, x_n; \Theta_1; \Theta_2; \ldots \Theta_k)$

Т.е. оценки $\tilde{\theta}_k$ таковы, что имеющиеся наблюдения x_1, x_2 ,…, x_n являются наиболее правдоподобными .

Для упрощения вычислений максимизируют $\ln L$. Для отыскания оценки параметров $\tilde{\theta}_k$ надо решить (уравнение) систему правдоподобия , получаемую приравниваем частных производных нулю по параметрам

$\frac{d \ln L}{d\theta}=0$

Найденную т. максимума $\tilde{\theta}_k$ принимают в качестве оценки наибольшего правдоподобия параметра Θ .

Метод дает оценки, которые состоятельны (смещенными) распределены асимптотически нормально (при больших значениях n приближенно нормальны) и имеют наименьшую дисперсию по сравнению с другими оценками. Особенно полезен метод для малых выборок.

3. Метод наименьших квадратов

Оценка определяется из условия минимизации суммы квадратов отклонений выборочных данных от определяемой оценки.

Например, для нахождения оценки генеральной средней $\theta = \overline{x_0}$ по методу наименьших квадратов $\tilde{\theta}_n$.

Преимущества : 1) не требует знание закона распределения выборочных данных; 2) достаточно хорошо разработан в плане вычислительной реализации.

5. Оценка параметров генеральной совокупности по собственно-случайной выработке

Оценка генеральной доли

Пусть генеральная совокупность содержит N элементов, из которых M обладает некоторым признаком A. Тогда оценкой генеральной доли $\mathrm{P} = \frac{M}{N}$ является его статистический аналог – выборочная доля $w = \frac{m}{n}$. (Для повторной и бесповторной выборки).

Теорема: Выборочная доля $w = \frac{m}{n}$ повторной и бесповторной выборок есть несмещенная, состоятельная и эффективная оценка генеральной доли $\mathrm{P} = \frac{M}{N}$.

Оценка генеральной средней

Теорема: Выборочная средняя $\bar{x}$ повторный и бесповторный выборок есть несмещенная, состоятельная и эффективная оценка генеральной средней $\bar{x}_0$.

Оценка генеральной дисперсии

Теорема: Выборочная дисперсия S^2 повторный и бесповторной выборок есть смещенная и состоятельная оценка генеральной дисперсии.

$$M\,(s^2) = \frac{n-1}{n}\sigma^2, \text{т.е } S^2 - \text{ смещенная оценка } \sigma^2$$

т.к. $\frac{n-1}{n} < 1$ и $M(S^2) < \sigma^2$, то выборочная дисперсия (в среднем, полученная по разным выборкам) занижает генеральную дисперсию.

Поэтому, заменяя σ^2 на s^2 мы допускаем систематическую погрешность в меньшую сторону. Чтобы ликвидировать этот недостаток вводят поправку на значение $\frac{n}{n-1}$.

$\hat{S}^2 = \frac{n}{n-1} S^2$ -исправленная выборочная дисперсия

$\hat{S}^2$- является несмещенной и состоятельной оценкой генеральной дисперсии.

6. Интервальные оценки параметров генеральной совокупности.

Пусть $\tilde{\theta}$ – точечная оценка неизвестного параметра θ– генеральной совокупности. Однако $\tilde{\theta}$ – это лишь приближенное значение θ и для выборки малого объема может существенно отличается от θ.

Чтобы получить представление о точности и надежности оценки $\tilde{\theta}n$ параметра θ используют интервальную оценку параметра.

Определение: *Доверительным интервалом* $\left[\tilde{\theta} - \Delta;\ \tilde{\theta} + \Delta\right]$ для параметра θ называется такой интервал, относительно которого можно утверждать с определенной, близкой к единице, вероятностью γ, что он содержит неизвестное значение параметра θ.

Также доверительный интервал называют интервальной оценкой параметра θ.

Верить γ называется *доверительной вероятностью*, *уровнем доверия* или *надежностью оценки*.

Наибольшее отклонение Δ оценки $\tilde{\theta}$ от оцениваемого параметра θ, которое возможно с заданной доверительной вероятностью γ называется *предельной ошибкой выборки*.

Ошибка Δ является *ошибкой репрезентативности* (случайной).

Общий подход к построению интервальной оценки параметра θ на основе случайной выборки $x_1, x_2.., x_n \in X$- генеральная совокупность состоит в определении границ интервала удовлетворяющего условию:

$$P\{\theta_{\min} \le \theta \le \theta_{\max}\} = \gamma .$$

1. Построение доверительно интервала для генеральной средней.

а) При известной дисперсии σ^2 - генеральной совокупности.

Будем рассматривать $\bar{x}$ как случайную величину $\bar{X}$ ($\bar{x}$ изменяется от выборки к выборке) определяется по выборке.

Рассмотрим n независимых выборок одинаково нормально распределенных т. е. $x_1, x_2.., x_n$, с математическим ожиданием $M(x) = \bar{x}_0$ и средним квадратическим отклонением σ.

Вероятность того, что отклонение выборочной средней от генеральной средней не превзойдет число $\Delta > 0$ (по абсолютной величине) равна:

$$P(|X - \overline{x_0}| < \Delta) = \Phi\left(\frac{\Delta}{\sigma}\right) = \gamma$$

Если определить t (с заданной надежностью от вероятности), то доверительный интервал можно определить:

$$\bar{x} - \Delta \le \mu \le \bar{x} + \Delta$$

Из (*) видно, что: 1) при возрастании n, Δуменьшается, и, следовательно, точность оценки увеличивается;

2) при увеличении надежности γ, t увеличивается и увеличивается Δ.

Поясним *смысл*, который имеет заданная надежность (вероятность). Надежность, (вероятность) $\gamma = 0{,}95$ указывает, что если проведено достаточно большое число выборок, то 95 % из них определяет такие доверительные интервалы, в которых параметр действительно заключен; лишь в 5% случаев он может выйти за границы доверительного интервала.

<u>б) Доверительный интервал для μ при неизвестной дисперсии σ^2.</u>

Случайная величина, определенная по данным выборки $t = \frac{\bar{x}-\mu}{s} \cdot \sqrt{n-1}$ имеет распределение Стьюдента (t - распределение) с v= n-1 степенями свободы и вероятности γ.

$$P\left(\left|\frac{\bar{x}-\mu}{S} \cdot \sqrt{n-1}\right| < t_\alpha\right) = P\left(\bar{x} - t_\alpha \frac{S}{\sqrt{n-1}} < \mu < \bar{x} + t_\alpha \frac{S}{\sqrt{n-1}}\right) = 1-\alpha$$

t_α - значение обратной функции распределения Стьюдента (t-распределение) соответствующее v = n – 1 степеням свободы и вероятности γ/

$$t_\alpha = St^{-1}(\alpha = 1-\gamma; V = n-1)$$

$$\boxed{\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta \quad , \quad \Delta = t_\alpha \frac{S}{\sqrt{n-1}}}$$

<u>Определение:</u> Число степеней свободы *v* определяется как общее число n наблюдений (вариантов) случайной величины X минус число уравнений *l* связывающих эти наблюдения, т. е. V = n – *l*.

Например: $t = \frac{\bar{x}-\mu}{s} \cdot \sqrt{n-1}$ V = n – 1

Наблюдения связаны одним уравнением $\bar{x} = \frac{\sum x_i}{n}$

<u>2) Интервальные оценки доли или вероятности p</u>

Пусть в n независимых испытаний некоторое событие A, вероятность появления которого в каждом испытании равна p, поступило m раз; $0 \leq m \leq n$

Точечная оценка генеральной доли используется w = $\frac{m}{n}$ – частность.

а) <u>Доверительный интервал для p при достаточно больших</u> n(n> 30)

$$q = 1-p \qquad \Delta_w = t_\gamma \sqrt{\frac{p(1-P)}{n}} = t_\gamma \sqrt{\frac{\sigma_n}{n}}$$

б) – <u>по малой выборке.</u>

Если доля признака в генеральной совокупности равна *p*, то <u>вероятность</u> <u>того</u>, что в повторной выборке объема n m элементов обладают этим признаком определяется по формуле Бернулли:

$$P_n(m) = C_n^m p^m q^{n-m}, \qquad q = 1-p$$

т. к. при p≠ 0,5 биноминальное распределение несимметрично, то в качестве доверительного интервала для p берут такой интервал $(p_1; p_2)$

$$\sum_{i=m}^{n} C_n^i p_1^i (1-p_1)^{n-1} = \frac{\gamma}{2}; \quad \sum_{i=1}^{m} C_n^i p_2^i (1-p_2)^{n-1} = \frac{\gamma}{2}$$

который решается приближенно.

Теорема: Вероятность того, что отклонение выборочной средней и выборочной средней (или доли) от генеральной средней (доли) не произойдет число $\Delta > 0$ (по абсолютной величине), равно

$$P(|\bar{x} - \overline{x_0}| \le \Delta) = \Phi(t) = \gamma \qquad P(|w - p| \le \Delta) = \Phi(t) = \gamma$$

$$t = \frac{\Delta}{\sigma_x} \qquad t = \frac{\Delta}{\sigma_w}$$

Определение: Среднее квадратическое отклонение выборочной средней $\sigma_{\bar{x}}$ и выборочной доли σ_w собственно-случайной выборки называется средней квадратической (стандартной) ошибкой выборки.

$$\Delta = t \cdot \sigma_x\ ,\ \Delta = t \cdot \sigma_w$$

Если задать t, то интервальные оценки (доверительные интервалы) по формулам

$$\bar{x} - \Delta \le \overline{x_0} \le \bar{x} + \Delta$$

$$w - \Delta \le p \le w + \Delta$$

Таблица 2 – Формулы средних квадратических ошибок выборки

Оцениваемый параметр	Повторная выборка	Бесповторная выборка
Средняя	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \approx \sqrt{\frac{S^2}{n}}$	$\sigma_{\bar{x}} \approx \sqrt{\frac{S^2}{n}(1 - \frac{n}{N})}$
Доля	$\sigma_w = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{w(1-w)}{n}}$	$\sigma_w = \sqrt{\frac{w(w-1)}{n}(1 - \frac{n}{N})}$

1.8 Лекция № 8 (4 часа).

Тема: «Статистическая проверка статистических гипотез» (интерактивная форма)

1.8.1 Вопросы лекции:

1. Статистическая гипотеза. Виды гипотез
2. Ошибки первого и второго рода
3. Статистический критерий проверки нулевой гипотезы. Критическая область. Область принятия решений
4. Мощность критерия
5. Критерий Вилкоксона и проверка гипотезы об однородности двух выборок
6. Критерий согласия хи-квадрат Пирсона

1.8.2 Краткое содержание вопросов:

1. Статистическая гипотеза. Виды гипотез

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений генеральных или выборочных совокупностей.

Например: 1) генеральная совокупность распределена по закону Пуассона;

2) дисперсии двух нормальных совокупностей равны между собой.

Наряду с выдвинутой гипотезой рассматривают и противоречащую (альтернативную) ей. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая (альтернативная) гипотеза.

Нулевой (основной) называют выдвинутую гипотезу H_0, которую нужно проверить.

Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Например: если нулевая гипотеза состоит в предположении, что математическое ожидание М нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что $M \neq 10$. (H_0: $M = 10$; H_1: $M \neq 10$.)

Различают гипотезы, которые содержат только одно и более одного предположений.

Простой называют гипотезу, содержащую только одно предположение. Например: если λ - параметр показательного распределения, то гипотеза H_0: $\lambda = 5$ - простая.

Сложной называют гипотезу, состоящую из конечного или бесконечного числа простых гипотез. Например: сложная гипотеза H_0: $\lambda>5$ состоит из бесчисленного множества простых вида H_i: $\lambda = b_i$, где b_i - любое число, большее 5.

2. Ошибки первого и второго рода

- *Ошибка первого рода* состоит в том, что будет отвергнута правильная гипотеза.
- *Ошибка второго рода* состоит в том, что будет принята неправильная гипотеза.

Последствия этих ошибок могут оказаться весьма различными.

Замечание: Правильное решение может быть принято также в двух случаях:

1) гипотеза принимается, причем и в действительности она правильная;

2) гипотеза отвергается, причем и в действительности она неверна.

Таблица 4 – Принятие решений о гипотезе H_0

Гипотеза H_0		согласно критерию	
		принимается	отвергается
на самом деле	верна	Правильное решение $p(H_0/H_0)=1-\alpha$	Ошибка 1-го рода $p(H_1/H_0)=\alpha$
	не верна	Ошибка 2-го рода $p(H_0/H_1)=\beta$	Правильное решение $p(H_1/H_1)=1-\beta$

Вероятность совершить ошибку первого рода (т.е. ошибку, заключающуюся в том, что нулевая статистическая гипотеза отклоняется в то время, как она верна) называется уровнем значимости α.

Ошибка первого рода равна вероятности того, что выборочная характеристика (статистика) критерия попадает в критическую область, когда H_0 верна.

Наиболее часто уровень значимости принимают равным 0.05 или 0.01. Например, если принят уровень значимости, равный 0,05, то это означает, что в пяти случаях из ста мы рискуем допустить ошибку первого рода (отвергнуть правильную гипотезу).

3. Статистический критерий проверки нулевой гипотезы. Критическая область. Область принятия гипотезы

Основу критерия представляет специально составленная выборочная характеристика K (статистика), точное или приближенное распределение которой известно. Эту величину обозначают через U или Z , если она распределена нормально, F по закону Фишера - Снедекора, T - по закону Стьюдента, $\chi 2$ - по закону "хи-квадрат"и т. д. Поскольку вид распределения во внимание приниматься не будет, обозначим эту величину в целях общности через K .

Статистическим критерием называют однозначно определённое правило, устанавливающее условия, при которых проверяемую гипотезу следует либо отвергнуть, либо не отвергнуть.

Значение случайной величины К, вычисленное по выборке называют наблюдаемым значением $К_{набл}$.

Каждый критерий разбивает всё множество возможных значений статистики К на два непересекающихся подмножества (области).

- *критическая область* – это совокупность значений критерия, при которых нулевую гипотезу отвергают.

- *область принятия гипотезы* – совокупность значений критерия, при которых нулевую гипотезу принимают.

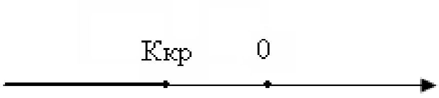
Основной принцип проверки статистических гипотез: если наблюдаемое значение статистики $К_{набл}$ принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Критическими точками (границами) $К_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают следующие виды критической области:

1. *Правосторонней* называют критическую область, определяемую неравенством $К > К_{кр}$, где $К_{кр} > 0$;

2. *Левосторонней* называют критическую область, определяемую неравенством $К < К_{кр}$, где $К_{кр} < 0$.



Односторонней называют правостороннюю или левостороннюю критическую область.

Двусторонней называют критическую область, определяемую неравенствами $К < К1$, $К > К2$, где $К2 > К1$. Если критические точки симметричны, то $|К| > К_{кр}$.

Для каждого критерия имеются соответствующие таблицы, по которым и находят точку удовлетворяющую требованию:

$P(К > К_{кр}) = \alpha$ – правосторонняя

$P(К < К_{кр}) = \alpha$ – левосторонняя

$P(К < К_{кр}) + P(К > К_{кр}) = \alpha$ – двусторонняя, если симметрична обл. то $P(К < К_{кр}) = \alpha/2$

Т.е. для нахождения $К_{кр}$ – критической точки исходят из требования, чтобы при условии справедливости нулевой гипотезы, вероятность того, что критерий К примет значение, большее $К_{кр}$, была равна принятому уровню значимости.

Замечание 1. Если известно $К_{кр}$ и вычислено по данным выборки $K_{набл}$ и если $K_{набл} > К_{кр}$, то нулевую гипотезу отвергают, если $K_{набл} <$ Ккр – то нет оснований отвергнуть нулевую гипотезу.

Замечание 2. $K_{набл}$ может оказаться больше $К_{кр}$, но потому, что H_0 ложна, а по другим причинам (малый объём выборки, недостатки методики эксперимента и т.д.). В этом случае отвергнув правильную нулевую гипотезу, совершают ошибку первого рода. Вероятность этой ошибки равна α.

Поскольку вероятность события $K > K_{кр}$ мала (α– малая вероятность), такое событие при справедливости нулевой гипотезы, в силу принципа практической невозможности маловероятных событий, в единичном испытании не должно наступить. Если все же оно произошло, т. е. наблюдаемое значение критерия оказалось больше $K_{кр}$, то это можно объяснить тем, что нулевая гипотеза ложна и, следовательно, должна быть отвергнута.

4. Мощность критерия

Мы строили критическую область, исходя из требования, чтобы вероятность попадания в нее критерия была равна α при условии, что нулевая гипотеза справедлива.

Целесообразно ввести в рассмотрение вероятность попадания критерия в критическую область при условии, что нулевая гипотеза неверна и, следовательно, справедлива конкурирующая.

Мощностью критерия называют вероятность правильного отклонения неверной нулевой гипотезы H_0, т.е. вероятность $1\text{-}\beta$ не совершить ошибку второго рода.

Другими словами, мощность критерия – это вероятность того, что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.

Критическую точку (область) целесообразно построить так, чтобы мощность критерия была максимальной.

Пусть вероятность ошибки второго рода (принять неверную гипотезу H_0) равна β, (т.е. верна H_1), то мощность критерия равна $1\text{-}\beta$.

Пусть мощность $1\text{-}\beta$ возрастает, тогда уменьшается вероятность β совершить ошибку второго рода (принять неверную гипотезу H_0 или отвергнуть верную H_1) т.е. чем мощность критерия больше, тем меньше вероятность ошибки второго рода.

Замечание. Мощность критерия – вероятность того, что ошибка второго рода будет не допущена.

Замечание. С уменьшением α возрастает вероятность ошибки β второго рода: принять H_0, когда она не верна.

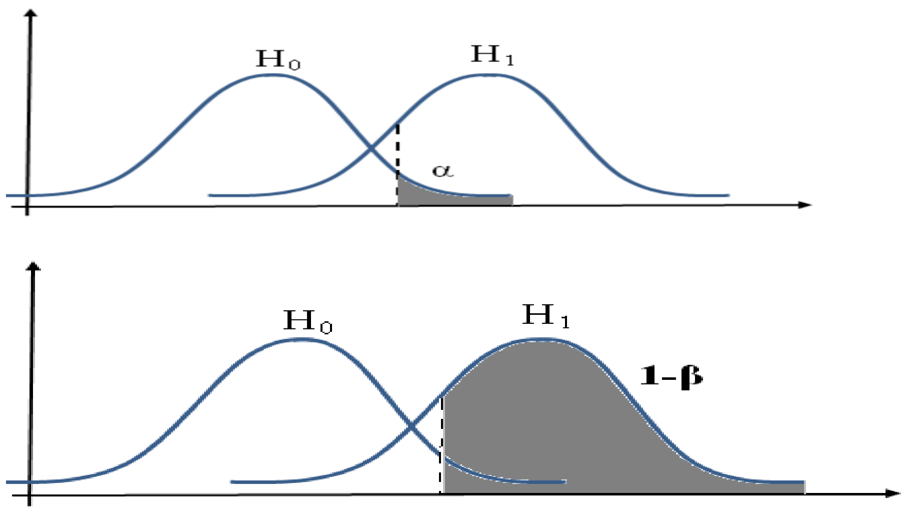


Рисунок 1. Уровень значимости α и мощность критерия $1\text{-}\beta$

5. Критерий Вилкоксона и проверка гипотезы об однородности двух выборок

Критерий Вилкоксона (предложен в 1945 году, в 1947 году Манн и Уитни обобщили критерий на выборки различного объема) служит для проверки однородности двух независимых выборок: $x_1, x_2, \ldots, x_{n1}$ и $y_1, y_2, \ldots, y_{n2}$. достоинство этого критерия состоит в том, что он применим к случайным величинам, распределения которых неизвестны; требуется лишь, чтобы величины были непрерывными.

Если выборки однородны, то считают, что они извлечены из одной и той же генеральной совокупности и, следовательно, имеют одинаковые, причем неизвестные, непрерывные функциираспределения $F_1(x)$ *и* $F_2(x)$.

H_0: $F_1(x)=F_2(x)$ - при всех значениях аргумента x функции распределения равны между собой.

H_1: $F_1(x)\neq F_2(x)$, $F_1(x)<F_2(x)$ (X>Y), $F_1(x)>F_2(x)$ (то X<Y).

Предположим, что $n_1\leq n_2$.

1. Если объем обеих выборок не превосходит 25.

Правило 1. для уровня значимости α=2Q и H_1: $F_1(x)\neq F_2(x)$ надо:

1) расположить варианты обеих выборок в возрастающем порядке в виде одного вариационного ряда и вычислить для этого ряда $W_{набл}$ – сумма порядковых номеров вариант первой выборки;

2) найти по таблице критических значений Вилкоксона нижнюю критическую точку $w_{нижн.кр.}(Q;\ n_1,\ n_2)$, где Q=α/2;

3) найти верхнюю критич. точку по формуле $w_{верх.кр.}=(n_1+n_2+1)n_1-\ w_{нижн.кр.}$.

Если $W_{набл}<\ w_{нижн.кр.}$ или $W_{набл}>\ w_{верх.кр}$ – нулевую гипотезу отвергают.

Если $w_{нижн.кр.}<W_{набл}<\ w_{верх.кр}$ – нет оснований отвергнуть нулевую гипотезу.

Правило 2. При H_1: $F_1(x)>F_2(x)$ надо найти по таблице критических значений Вилкоксона нижнюю критическую точку $w_{нижн.кр.}(Q;\ n_1,\ n_2)$, где Q=α.

Если $W_{набл}>\ w_{нижн.кр.}$ – нет оснований отвергнуть нулевую гипотезу.

Если $W_{набл}<\ w_{нижн.кр.}$ – нулевую гипотезу отвергают.

Правило 3. При H_1: $F_1(x)<F_2(x)$ надо найти по таблице критических значений Вилкоксона верхнюю критическую точку $w_{верх.кр.}=(n_1+n_2+1)n_1-\ w_{нижн.кр.}$, где Q=α.

Если $W_{набл}<\ w_{верх.кр.}$ – нет оснований отвергнуть нулевую гипотезу.

Если $W_{набл}<\ w_{верх.кр.}$ – нулевую гипотезу отвергают.

2. Если объем хотя бы одной выборки превосходит 25.

1) При конкурирующей гипотезе $F_1(x)\neq F_2(x)$ нижняя критическая точка

$$w_{нижн.кр.}(Q;\ n_1,\ n_2)=\frac{(n_1+n_2+1)n_1-1}{2}-z_{кр}\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}, \qquad (*)$$

где Q=α/2; $z_{кр}$ находят по таблице функции Лапласа по равенству $\Phi(z_{кр})=(1-\alpha)/2$. В остальном правило 1) п.1 сохраняется.

2) При конкурирующих гипотезах $F_1(x)<F_2(x)$, $F_1(x)>F_2(x)$ нижнюю критическую точку находят по формуле (*), положив Q=α; $z_{кр}$ находят по таблице функции Лапласа по равенству $\Phi(z_{кр})=(1-2\alpha)/2$. В остальном правило сохраняется.

6. Критерий согласия хи-квадрат Пирсона

Одной из важнейших задач математической статистики является установление по эмпирическим данным теоретического закона распределения СВ.

Как бы хорошо не было подобран теоретический закон распределения, между эмпирическим и теоретическим распределениями неизбежны расхождения. Возникает вопрос: объясняются ли эти расхождения только случайными обстоятельствами, связанными с ограниченным числом наблюдений или способом группировки, или они существенны и связаны с тем, что теоретическое значение распределения подобрано неудачно. Для ответа на этот вопрос и служат *критерии согласия*.

Критерий согласия – статистический критерий, предназначенный для проверки гипотезы о предполагаемом законе неизвестного распределения.

Критерий согласия основан на использовании различных мер расстояний между анализируемой эмпирической функцией $\tilde{F}(x)$ распределения, определенной по выборке и функцией F(x) распределения генеральной совокупности X.

Выдвигается нулевая гипотеза:

H_0: $w_1 = p_1;\ w_2 = p_2;\ \ldots,\ w_i = p_i,\ \ldots,\ w_n = p_n$

о том, что исследуемая СВ X подчиняется определенному закону распределения.

p_i - вероятность принятия случайной величиной i-го значения.

Критерий согласия состоит в том, что выбирается некоторая случайная величина (статистика) T_n, являющаяся мерой расхождения (рассогласования) между рядом наблюдений и предполагаемым теоретическим распределением.

При проверке нулевой гипотезы заранее задается уровень значимости $\alpha(\alpha = 0{,}1; 0{,}05; 0{,}01)$.

На основании выборки вычисляется наблюдаемая величина $T_{набл}$, которая затем сравнивается с $T_{табл(\ крит)}$. Если $T_{набл} > T_{кр}$, то нулевая гипотеза отвергается. Если же $T_{набл} < T_{кр}$ – H_0 не отвергается, в этом случае отклонения от предполагаемого теоретического закона распределения считаются незначительными, т. е. данные наблюдения не противоречат гипотезе о виде распределения.

Замечание: Можно осуществит проверку гипотезы о виде распределения с помощью критерия согласия в другом порядке. По наблюдаемому значению ($T_{набл}$) определить, пользуясь соответствующей таблицей, вероятность $\alpha_{набл} = P(T > T_{набл})$ Если $\alpha_{набл} > \alpha$, то отклонения значимы, т. е. гипотеза отвергается, если же $\alpha_{набл} < \alpha$, то гипотеза не отвергается.

$\alpha_{набл}$ близкие к 1 указывают нерепрезентативность выборки.

χ^2- критерий Пирсона (имеет наибольшее применение). В качестве меры расхождения $T_{набл}$ берется величина χ^2, равная сумме квадратов отклонений частностей (статистических. вероятностей)$\omega_i = \frac{m_i}{n_i}$ от гипотетических p_i , рассчитанных по предполагаемому распределению.

$$\chi^2 = \sum \frac{(\frac{m_i}{n} - p_i)^2}{p_i} \cdot n, \qquad (*)$$

$m_i^T = np_i$ – «теоретическая частота»

m_i - эмпирические частоты $m_i = w_i \cdot n$

Доказано, что при $n \to \infty$ СВ (*) стремится к закону распределения с v степенями свободы; $v=l-r-1$, r- число параметров предполагаемого теоретического закона распределения (нормальное распределение r=2 (μ, σ), Пуассона r=1 (λ)). l- число интервалов эмпирического распределения (вариационного ряда)

Схема применения критерия χ^2 для проверки гипотезы H_0

1. Рассчитывают теоретические (частоты) вероятности;
2. Определяется мера расхождения эмпирических и теоретических частот χ^2;
3. Для выбранного уровня значимости α по таблице критических значений χ^2 распределения находят критическое значение $\chi^2_{\alpha,v}$ при числе степеней свободы *v=l-r-1;*
4. Если $\chi^2_{набл} > \chi^2_{\alpha,v}$, то H_0 – отвергается;

$\chi^2_{\text{набл}} \le \chi^2_{\alpha,v}$ - гипотеза H_0 принимается (не противоречит опытным данным.)

<u>Замечание:</u> Необходимо, чтобы в каждом интервале было не менее 5 наблюдений, иначе имеет смысл объединить соседние интервалы.

(Крамер стр. 377 пример)

Расчет теоретических частот

1. Производят группировку наблюдаемых значений X (выборорки объема n) на l групп (интервалов) одинаковой длины.

Находят середины интервалов $x_i' = \dfrac{x_i + x_{i+1}}{2}$. Частота интервала определяется путем подсчета количества единиц, попавших в него.

В итоге получаем последовательность равностоящих вариант и соответствующих им частот.

x_1'	x_2'	…	x_s'
n_1	n_2	…	n_3

$$\sum_{i=1}^{s} n_i = n$$

2. Вычисляют $\overline{x}'$, σ.

3. Нормируют случайную величину X, т. е. $t = \frac{x_i - \bar{x}'}{\sigma'}$, вычисляют концы интервалов $(t_i;\ t_{i+1})$

4. Вычисляют теоретические вероятности попадания X в интервалы $(x_i; x_{i+1})$ по функции Лапласа $p_i = P(x_i < x < x_{i+1}) = \Phi(t_{i+1}) - \Phi(t)$

5. Теоретические частоты $m_i' = np_i$

1.9 Лекция № 9 (2 часа).

Тема: «Дисперсионный анализ» (интерактивная форма)

1.9.1. Вопросы лекции:

1. Понятие о дисперсионном анализе
2. Однофакторный дисперсионный анализ
3. Сравнение нескольких средних методом дисперсионного анализа

1.9.2 Краткое содержание вопросов:

1. Понятие о дисперсионном анализе

Дисперсионный анализ предназначен для проверки зависимости нормально распределенной случайной величины У-результативного признака, от факторных признаков или факторов.

Модели дисперсионного анализа классифицируются в зависимости от числа факторов на однофакторные, двухфакторные, и. т. д комплексы.

При этом среди факторов может быть как случайные, так и неслучайные величины.

2. Однофакторный дисперсионный анализ

Пусть требуется проверить влияние на результативный количественный, нормально распределенный признак одного фактора F, который имеет р постоянных уровней : F_j, $j=1, \dots ,p$.

Наблюдаемые значения результативного признака У на каждом уровне F_j обозначим через y_{ij} , где $i = 1,2, \dots , n_j$ – число наблюдений У на уровне F_j .

Наблюдаемые значения результативного признака обычно представляются в виде таблицы наблюдений

$$\begin{pmatrix} y_{11}\ y_{12}\ \dots\ y_{1p} \\ y_{21}\ y_{22}\ \dots\ y_{2p} \\ y_{31}\ y_{32}\ \dots\ y_{3p} \\ \dots\ \dots\ \ \dots\ \dots \\ y_{n_1 1}\ y_{n_2 2}\dots y_{n_p p} \end{pmatrix}, i=1, \dots , n_j;\ j=1, \dots , p.$$

Однофакторная дисперсионная модель имеет вид:

$$y_{ij}= \mu+\alpha_j + \varepsilon_{ij} \qquad j=\overline{1,p} \ , i=1,2 \dots , n_j$$

y_{ij} –наблюдаемые значения результативного признака, полученного на j-м уровне фактора F_j с порядковым номером i.

μ- генеральная средняя комплекса

α_j – эффект, обусловленный влиянием j-го уровня фактора, т.е. F_j.

Фактор F_j может принимать фиксированные значения (Например : номер станка, вид удобрения) и иметь случайные уровни, т. е. измеряется количественно, где F_j – случайные величины.

ε_{ij} - случайные величины (остатки), отражающие влияние на У всех неконтролируемыхфакторов, т.е. вызванное вариацией переменной внутри отдельного уровня.

Основные предпосылки дисперсионного анализа:

1.Математическое ожидание возмущения $\varepsilon_{ij}=0$ для любых *i*, т.е. $\mu(\varepsilon_{ij})=0$

2.Возмущения ε_{ij} взаимно независимы

3.Дисперсия возмущения ε_{ij} (или переменной y_{ij}) постоянна для $\forall$ i,j , т.е.

$$D(\varepsilon_{ij})=\sigma^2$$

4.Возмущение ε_{ij} (или переменная y_{ij}) имеет нормальный закон распределения $N(0; \sigma^2)$

Введем обозначения:

$\bar{y}_j = \frac{1}{k_j}\sum_{i=1}^{y} y_{ij}$ - групповые средние (средние для уровня F_j).

$\bar{y}= \frac{1}{N}\sum \bar{y}_j \cdot n_j=\frac{1}{N}\sum_{j=1}^{p}\sum_{j=1}^{n_j} y_{ij}$ - общая средняя, где $N= \sum_{j=1}^{p} n_j$.

$Q_F=\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(\bar{y}_j-\bar{y}\right)^2=\sum_{i=1}^{p_j}\left(\bar{y}_j-\bar{y}\right)^2 n_j$ – факторная сумма квадратов отклонений (сумма квадратов отклонений групповых средних от общей средней).

$Q_{ост}= \sum_{i=1}^{p}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}\right)^2$- остаточная (внутригрупповая) сумма квадратов отклонений (сумма квадратов отклонений наблюдений от групповых средних).

$Q_{общ}= \sum_{i=1}^{p}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}\right)^2$ – общая сумма квадратов отклонений.

После ряда преобразований (самостоятельно)

$(y_{ij} - \bar{y})$=$(\bar{y}_j - \bar{y})$ + $(y_{ij} - \bar{y}_j)$

Обе части тождества возводят в квадрат и производят суммирование по индексам i и j.

$\sum_{j=1}^{p}\sum_{i=1}^{n_j}(\overline{y_{ij}} - \bar{y})^2 = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(\bar{y}_j - \bar{y})^2 + \sum_{j=1}^{p}\sum_{i=1}^{n_j}(\overline{y_{ij}} - \bar{y}_j)^2 + 2\sum_{i=1}^{n}(\bar{y}_i - -\bar{y})\sum_{i=1}^{n}(y_{ij} - y_j)$ и.т.д.

получим следующее основное тождество дисперсионного анализа:

$Q_{общ}$ = Q_F + $Q_{ост}$

в котором заключена основная идея дисперсионного анализа; общая вариация результативного признака складывается из 2х компонент : (уровня фактора и всех неучтенных факторов) : Q_F ,характеризующей изменчивость, обусловленную различиями между уровнями фактора и $Q_{ост}$, характеризующей одинаковую для всех уровней F вариацию, под воздействием неучтенных факторов.

В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а средние квадраты, являющиеся несмещенными оценками соответствующих дисперсий, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы.

Напомним, что число степеней свободы определяется как общее число наблюдений минус число связывающих их уравнений.

Поэтому для $\hat{S}_F^2$ – оценки межгрупповой дисперсии, число степеней свободы k_1=p-1, т.к. при его расчете используют p групповых средних, связанных между собой одним уравнением $\bar{y} = \frac{1}{N}\sum \bar{y}_i n_i$

Для $\hat{S}_{ост}^2$ - оценки внутригрупповой дисперсии (несмещенной), k_2 =pn – p, т.к. используют все N наблюдений, связанных меду собой p уравнениями $\bar{y}_i = \frac{\sum y_{iy}}{n_j}$.

Схему дисперсионного анализа представим в таблице:

Таблица 1- Однофакторный дисперсионный анализ

Виды дисперсий	Число степеней свободы df	Сумма квадратов	Средний квадрат
Межгрупповая (факторная)	p-1	$Q_F = \sum_{i=1}^{p_j}(\bar{y}_j - \bar{y})^2$	$\widehat{S_F} = \frac{Q_F}{p-1}$
Внутригрупповая (остаточная)	N-p	$Q_{ост} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_i)^2$	$\widehat{S_{ост}} = \frac{Q_{ост}}{N-p}$
Общая	N-1	$Q_{ост} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$	$\widehat{S_{общ}} = \frac{Q_{общ}}{N-1}$

Основная гипотеза дисперсионного анализа состоит в утверждении, что уровни фактора F не влияют на изменение результативного признака У.

Для проверки нулевой гипотезы H_0 вычисляется

$$F_{набл} = \frac{R_F/(p-1)}{R_{ост}/(N-p)} = \frac{\widehat{S_F^2}}{\hat{S}_{ост}^2}$$

Если $F_{набл} \leq F_{кр}$ (α, p-1, N – p) , где $F_{кр}$ находится на таблице F распределения на уровне значимости и числа степеней свободы v_1=p-1,v_2=N-p , то гипотеза не отвергается. Из этого следует, что влияние фактора F на результирующий признак не доказано.

Если, $F_{набл} \geq F_{кр}$ то гипотеза отвергается с вероятностью ошибки, равной α; фактор существенно (значимо) влияет на результативный признак.

3.Сравнение нескольких средних методом дисперсионного анализа

Иногда дисперсионный анализ применяется, чтобы установить однородность нескольких совокупностей, (дисперсии этих совокупностей одинаковы, то в этом смысле совокупности однородны). Однородные же совокупности можно объеденить в одну и тем самым получить о ней более полную информацию, следовательно, более надежные выводы.

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_p$

Для того, что бы проверить нулевую гипотезу о равенстве групповых средних нормальных совокупностей с одинаковыми дисперсиями, достаточно проверить по критерию F нулевую гипотезу о равенстве факторной и остаточной дисперсии:

рассчитать

$$t_{набл} = \frac{(\overline{y_1} - \overline{y_2})}{\sqrt{R_{ост}/(N-p)}} = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

При выполнении гипотезы H_0 рассматриваемая статистика имеет t- распределение Стьюдента: $t_{кр}$ (α, N-p).

Если $|t_{набл}| > t_{кр}$ (α, N-p), то гипотеза H_0 отвергается.

$F=t^2$

Замечание. Отклонение от основных предпосылок дисперсионного анализа – нормальности распределения последуемой переменной и равенства дисперсий в ячейках (если оно не чрезмерно) не сказывается существенно на результатах дисперсионного анализа при равном числе наблюдений в ячейках.

1.10 Лекция № 10 (2 часа).

Тема: «Корреляционный анализ» (интерактивная форма)

1.10.1. Вопросы лекции:

1. Функциональная, статистическая и корреляционная зависимости
2. Коэффициент корреляции
3. Проверка статистической значимости коэффициента корреляции и его интервальная оценка
4. Корреляционное отношение и индекс корреляции

1.10.2 Краткое содержание вопросов:

1. Функциональная, статистическая и корреляционная зависимости.

Во многих задачах требуется установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других величин. Рассмотрим зависимость Y от одной случайной (или неслучайной) величины X.

Две случайные величины могут быть связаны либо функциональной зависимостью, либо зависимостью другого рода, называемой статистической, либо быть независимыми.

Статистической (или ***стохастической, вероятностной***) называют зависимость, при которой изменение одной из величин влечет изменение распределения другой.

Статистическая зависимость между двумя переменными при которой каждому значению одной переменной соответствует определенное (среднее значение) условное математическое ожидание другой называется ***корреляционной***.

Возникновение понятия статистической связи объясняется тем, что зависимая переменная подвержена влиянию ряда неконтролируемых или неучтенных факторов, а также тем, что измерение значений переменных неизбежно сопровождается случайными ошибками.

Статистические связи между переменными можно изучать методами корреляционного и регрессионного анализа.

Корреляционный анализ является одним из методов статистического анализа взаимозависимости нескольких признаков. В настоящее время он определяется как метод, применяемый тогда, когда данные наблюдений можно считать случайными и выбранными из генеральной совокупности, распределенной по многомерному нормальному закону.

Основная задача корреляционного анализа – выявление связи между случайными переменными и оценка ее тесноты.

2. Коэффициент корреляции.

Рассмотрим генеральную совокупность с двумя признаками X и Y, совместное распределение которых задано плотностью двумерного нормального закона распределения:

$$f(X,Y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[\frac{1}{2\sqrt{1-\rho^2}}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)\right]$$

определяемого пятью параметрами: MX = μ_x, MY = μ_y, DX = σ_x^2, DY = σ_y^2, ρ.

Парный коэффициент корреляции ρ — показатель, который характеризует тесноту линейной связи между случайными величинами X и Y и определяется как:

$$\rho=M\left[\frac{X-\mu_x}{\sigma_x}\cdot\frac{Y-\mu_y}{\sigma_y}\right].$$

Свойства парного коэффициента корреляции ρ:

1) $-1 < \rho < 1$ – ρ принимает значения на отрезке [-1; 1], при этом, чем ближе |p| к единице, тем теснее связь между признаками X и Y;

2) при |p| = 1 корреляционная связь между признаками X и Y представляет собой линейную функциональную зависимость (y = f(x));

3) при p = 0 линейная корреляционная связь между X и Готсутствует (но это не означает невозможность наличия между ними нелинейной связи);

4) p < 0 указывает на наличие обратной зависимости между переменными X и Y (при увеличении одной переменной другая уменьшается);

5) p > 0 указывает на наличие прямой зависимости между переменными X и Y (при увеличении (уменьшении) одной переменной другая тоже возрастает (уменьшается);

6) если все значения признаков увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции не изменится.

Коэффициент детерминации ρ^2 (равный квадрату коэффициента корреляции) указывает долю дисперсии одной случайной величины, обусловленную вариацией другой.

Соответственно, $(1 - \rho^2)$, показывает долю дисперсии случайной величины, объясняемой не включенными в рассматриваемую двумерную модель факторами.

Пусть из генеральной совокупности (X, Y) взята случайная выборка объемом n: (x_1, y_1), $(x_2; y_2)$, ..., (x_i, y_i),..., (x_n, y_n).

1. Если объем выборки n невелик, то статистические характеристики генеральной совокупности вычисляются непосредственно по ряду наблюдений (x_v, y_v).

x_1	x_2	…	x_i	…	x_n
y_1	y_2	…	y_i	…	y_n

n- объем выборки.

2. Если выборка из генеральной совокупности велика, то ряд наблюдений преобразовывается к двумерному вариационному ряду, представляемому в виде таблицы, называемой корреляционной.

В первой строке в возрастающем порядке расположены варианты x_i , а в первом столбце — варианты y_i. На пересечении столбца x_i- и строки y_i находится частота m_{ij}, обозначающая число точек выборки, значения признаков у которых равны $(x_i; y_j)$, где i = 1, 2, ..., k, j = 1, 2, ..., l.

	x_1	…	x_i	…	x_k	Итого по y (m_y)
y_1	m_{11}	…	m_{i1}	…	m_{k1}	m_{*1}
…	…	…	…	…	…	…
y_j	m_{1j}	…	m_{ij}	…	m_{kj}	m_{*j}
…	…	…	…	…	…	…
y_l	m_{1l}	…	m_{il}	…	m_{kl}	m_{*l}
Итого по x (m_x)	m_{1*}	…	m_{i*}	…	m_{k*}	n

В строке m_x помещены частоты одномерного вариационного ряда x, полученные путем суммирования значений m_{ij} в x_i-м столбце: $m_{i*} = \sum_{j=1}^{l} m_{ij}$

В столбце m_y помещены частоты ряда y: $m_{*j} = \sum_{i=1}^{k} m_{ij}$

Наконец, n — объем выборки.

Основные формулы для вычисления оценок статистических характеристик генеральной совокупности для различных объемов выборок приведены в табл. 2.

Таблица 2. Точечные оценки параметров двумерной корреляционной модели

Оцениваемый параметр генеральной совокупности	Его выборочная точечная оценка	
	По простым выборкам (n мало, данные не группированы)	По сгруппированным данным (n велико, данные сгруппированы)
Генеральная средняя X, μ_x	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_{i*}$

Генеральная средняя У, μ_y	$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$	$\bar{y} = \frac{1}{n}\sum_{j=1}^{l} y_j m_{*j}$
Генеральная средняя X^2, $M(X^2)$	$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_{i*}$
Генеральная средняя Y^2, $M(Y^2)$	$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$	$\overline{y^2} = \frac{1}{n}\sum_{j=1}^{l} y_j^2 m_{*j}$
Генеральная средняя XY, M(XУ)	$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_j$	$\overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j m_{ij}$
Генеральные дисперсии σ_x^2, σ_y^2	$S_x^2 = \overline{x^2} - \bar{x}^2, \; S_y^2 = \overline{y^2} - \bar{y}^2$	
Генеральный коэффициент корреляции ρ	Выборочный коэффициент корреляции $r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y}$	

Выборочный коэффициент детерминации r^2 (равен квадрату коэффициента корреляции) и указывает долю дисперсии одной случайной величины, обусловленную вариацией другой.

Соответственно, $(1 — r^2)$, показывает долю остаточной дисперсии случайной величины, объясняемой не включенными в рассматриваемую двумерную модель факторами.

Таким образом, дисперсию каждой их изучаемых переменных можно представить в виде суммы двух составляющих — регрессионной (обусловленной влиянием другой переменной) и остаточной (определяемой не включенными в рассматриваемую двумерную модель остаточными факторами). Остаточную дисперсию называют еще условной — т. е. вычисленной при условии фиксации другой переменной модели и обозначают подобно условной вероятности.

Итак, остаточная (условная) дисперсия переменной Y:

$$S_{y/x}^2 = S_y^2(1 - r^2)\,;$$

остаточная (условная) дисперсия переменной X:

$$S_{y/x}^2 = S_x^2(1 - r^2)\,.$$

3. Проверка значимости генерального коэффициента корреляции и его интервальная оценка.

В двумерной модели параметрами связи являются коэффициент корреляции ρ или его квадрат — коэффициент детерминации ρ^2.

На практике о тесноте зависимости между случайными переменными приходится судить не на основе истинных параметров связи (генеральных характеристик), а их выборочных аналогов, которые являются случайными величинами. Поэтому возникает вопрос, действительно, ли полученные значения объясняется наличием существующей линейной корреляционной зависимостью между переменными X и Y, или являются следствием случайности отбора переменных в выборку.

В двумерной модели достаточно проверить значимость коэффициента корреляции, т. е. на уровне значимости а проверить гипотезу H_0: $\rho = 0$.

Проверка значимости генерального коэффициента корреляции

I способ. С использованием таблицы Фишера—Иейтса .

По таблице Фишера—Иейтса для уровня значимости α и числа степеней свободы n- 2 находится критическое значение $r_{кр} = r_{табл}$ (α; n- 2).

Затем сравниваются наблюдаемое значение коэффициента корреляции — выборочный коэффициент корреляции r — с полученным критическим $r_{кр}$.

Гипотеза отвергается и генеральный коэффициент корреляции считается значимым, если наблюдаемое значение $r_{набл} = r$ по абсолютной величине окажется больше критического, т. е. если $|r_{набл}| > r_{кр}$.

В противном случае гипотеза H_0: $\rho = 0$ не отвергается и генеральный коэффициент корреляции считается незначимым.

II способ. С использованием распределения Стьюдента.

Основывается на расчете статистики $t_{набл} = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$, которая при истинности гипотезы H_0: $\rho = 0$ имеет t-распределение с n - 2 степенями свободы.

Гипотеза H_0: $\rho = 0$ отвергается и генеральный коэффициент корреляции считается значимым, если $|t_{набл}| > t_{кр}$, где $t_{кр} = t(\alpha; n-2)$ определяется по таблице t-распределения Стьюднта при уровне значимости α и числе степеней свободы n - 2.

Для значимых параметров связи находят интервальные оценки, которые с заданной надежностью содержат истинные значения параметров.

Интервальная оценка генерального коэффициента корреляции

Точечная оценка генерального коэффициента корреляции ρ — выборочный коэффициент корреляции r — является лишь приближенным его значением и при малых объемах выборки может давать существенные ошибки. Представляет интерес построение интервальной оценки генерального коэффициента корреляции ρ, куда он будет попадать с заданной надежностью γ.

$$P(\rho_{\min} \le \rho \le \rho_{\max}) = \gamma .$$

При определении с надежностью у доверительного интервала для ρ сначала используют Z-преобразование Фишера и определяют значение статистики Z_r по таблицам «Z – преобразования Фишера» или по формуле:

$$Z_r = \frac{1}{2}\ln\frac{1+r}{1-r} .$$

Затем устанавливают интервальную оценку для Z_r :

$$Z_r - \frac{t_\gamma}{\sqrt{n-3}} \le Z \le Z_r + \frac{t_\gamma}{\sqrt{n-3}} ,$$

где t_γ вычисляют по таблице интегральной функции Лапласа из условия $\Phi(t_\gamma) = \gamma$.

Обратный переход от Z_r к ρ осуществляют также по таблице Z-преоб-разования, после использования которой получают интервальную оценку для ρ с надежностью γ:

$$Z^{-1}(Z_r^{\min}) \le \rho \le Z^{-1}(Z_r^{\max})$$
$$\rho_{\min} \le \rho \le \rho_{\max}$$

При выборе ρ_{min} и ρ_{max} следует учитывать, что Z-функция — нечетная, т. е. Z(-r) = - Z(r).

Построение интервальной оценки генерального коэффициента корреляции р двумерной модели с заданной надежностью у можно представить в виде алгоритма.

4. Корреляционное отношение и индекс корреляции.

Рассмотренный выше коэффициент корреляции является полноценным показателем тесноты связи лишь в случае линейной зависимости между переменными. Однако часто возникает необходимость в достоверном показателе тесноты связи при любой форме зависимости.

Для получения такого показателя вспомним правило сложения дисперсий: $Q_{общ} = Q_F + Q_{ост}$

где $Q_{общ}$ — общая дисперсия переменной;

$Q_{ост}$— средняя групповых дисперсий или остаточная дисперсия;

Q_F — межгрупповая дисперсия

Остаточной дисперсией измеряют ту часть колеблемости Y, которая возникает из-за изменчивости неучтенных факторов не зависящих от X. Межгрупповая дисперсия выражает ту часть вариации Y, которая обусловлена изменчивостью X. Величина

$$\eta_{yx} = \sqrt{\frac{Q_F}{Q_{общ}}}$$

называется ***эмпирическим корреляционным отношением*** Y по X.

Чем теснее связь, тем большее влияние на вариацию переменной Y оказывает изменчивость X по сравнению с неучтенными факторами, тем выше η_{yx}.

Величина η_{yx}^2, называется ***эмпирическим коэффициентом детерминации***, показывает, какая часть общей вариации Y обусловлена вариацией X.

Основные свойства корреляционного отношения (при достаточно большом объеме выборки n):

1) Корреляционное отношение есть неотрицательная величина, не превосходящая 1: $0 < \eta < 1$.

2) Если $\eta = 0$, то корреляционная связь отсутствует.

3) Если $\eta = 1$, то между переменными существует функциональная зависимость.

4) $\eta_{yx} \neq \eta_{xy}$, т.е. в отличие от коэффициента корреляции r (для которого $r_{yx} = r_{xy} = r$) при вычислении корреляционного отношения существенно, какую переменную считать независимой, а какую — зависимой.

Эмпирическое корреляционное отношение η_{yx} является показателем рассеяния точек корреляционного поля относительно эмпирической линии регрессии, выражаемой ломаной, соединяющей значения $\bar{y}_i$ Однако в связи с тем, что закономерное изменение $\bar{y}_i$ нарушается случайными зигзагами ломаной, возникающими вследствие остаточного действия неучтенных факторов, η_{yx} преувеличивает тесноту связи. Поэтому наряду с η_{yx} рассматривается показатель тесноты связи R_{yx}, характеризующий рассеяние точек корреляционного поля относительно линии регрессии ух.

Показатель R_{yx}, получил название ***теоретического корреляционного отношения или индекса корреляции*** Y по X.

$$R_{yx} = \sqrt{\frac{Q_F}{Q_{общ}}} = \sqrt{1 - \frac{Q_{ост}}{Q_{общ}}}$$

где дисперсии $Q_{общ}$ и $Q_{ост}$ определяются по формулам, в которых групповые средние $\bar{y}_i$ заменены условными средними y_{xi}, вычисленными по уравнению регрессии.

Достоинством рассмотренных показателей η и R является то, что они могут быть вычислены при любой форме связи между переменными. Хотя η и завышает тесноту связи по сравнению с R, но для его вычисления не нужно знать уравнение регрессии. Корреляционные отношения η и R связаны с коэффициентом корреляции r следующим образом:

$$0 \le |r| \le R \le \eta \le 1$$

В случае линейной модели индекс корреляции R_{yx} равен коэффициенту корреляции r (по абсолютной величине).

Коэффициент детерминации R^2, равный квадрату индекса корреляции (для парной линейной модели — r^2), показывает долю общей вариаций зависимой переменной, обусловленной регрессией или изменчивостью объясняющей переменной.

Чем ближе R^2 к единице, тем лучше регрессия аппроксимирует эмпирические данные, тем теснее наблюдения примыкают к линии регрессии. Если $R^2 = 1$, то эмпирические точки (x, y) лежат на линии регрессии и между переменными Y и X существует линейная функциональная зависимость. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных, и линия регрессии параллельна оси абсцисс .

Расхождение между η^2 и R^2 (или r^2) может быть использовано для проверки линейности корреляционной зависимости.

Проверка значимости корреляционного отношения η основана на том, что статистика

$$F = \frac{\eta^2 (n-m)}{(1-\eta^2)(m-1)}$$

(где m — число переменных) имеет F-распределение Фишера—Снедекора с $k_1 = m - 1$, $k_2 = n - m$ степенями свободы. Поэтому η значимо отличается от нуля, если $F > F(\alpha, k_1, k_2)$, где $F(\alpha, k_1, k_2)$— табличное значение F-критерия на уровне значимости α при числе степеней свободы $k_1 = m - 1$, $k_2 = n – m$.

1.11 Лекция № 11 (2 часа).

Тема: «Регрессионный анализ» (интерактивная форма)

1.11.1. Вопросы лекции:

1. Понятие и предпосылки регрессионного анализа
2. Виды регрессионных моделей. Оценка параметров парной линейной регрессии
3. Проверка статистической значимости уравнения регрессии в целом и его параметров

1.11.2 Краткое содержание вопросов:

1. Понятие и предпосылки регрессионного анализа

Регрессионный анализ – метод исследования зависимости результативного признака Y (случайной величины) от нескольких случайных величин X1, X2,…,Xk, называемых *факторами* или *регрессорами.*

Важно различать корреляционные и регрессионные связи.

Основная задача регрессионного анализа – установление формы и изучение зависимости между переменными.

Уравнение регрессии может быть представлено:

$Y_i=f(x_i)+\varepsilon_i$, где ε_i - случайная переменная (возмущение), характеризующая отклонение фактических значений результативной переменной от функции регрессии.

Линейная парная регрессионная модель имеет вид:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (*)$$

Основные предпосылки регрессионного анализа

1.Зависимая переменная y_i (возмущение ε_i) является случайной величиной, а объясняющая переменная x_i- неслучайной;

2.Математическое ожидание возмущения ε_i равно нулю: $M(\varepsilon_i)=0$;

3.Дисперсия зависимой переменной y_i постоянна (для всех i) и равна $D(\varepsilon_i)=\sigma^2$;

4.Переменные y_i и y_j (возмущение ε_i и ε_j) не коррелированны $M(\varepsilon_i;\ \varepsilon_j)=0$, для любого $i\neq j$.

5.Зависимая переменная y_i (возмущение ε_i) является нормально распределенной случайной величиной.

2. Виды регрессионных моделей. Оценка их параметров

I. Различают однофакторные (с одной независимой переменной) и многофакторные (с несколькими независимыми переменными) модели.

Предположим, что для оценки параметров линейной функции регрессии взята выборка n пар переменных (x_i,y_i) i=1, … , n.

Оценкой модели (*) по выборке является уравнение регрессии $\hat{y}=b_0+b_1x$.

При *однофакторных* моделях связь часто называют парной, а коэффициент регрессии b_1- *полным*. Он показывает на сколько единиц изменяется зависимая переменная *y* при изменении независимой переменной *x* на единицу своего измерения.

Параметры b_0, b_1 определяются на основе МНК согласно которому неизвестные параметры b_0 и b_1, выбираются таким образом, чтобы сумма квадратов отклонений эмпирических данных результативного признака от значений ŷ, найденные по уравнению регрессии была минимальной:

$$S=\sum (y_i-\hat{y})^2=\sum(y_i-b_0-b_1x_i)^2 \xrightarrow{b_0,b_1} \min$$

Дифференцируя S по b_0 и b_1 и приравнивая к нулю ее частные производные, получим:

$$\begin{cases} \dfrac{dS}{db_0} = 2\sum (y_i - b_0 - b_1 x_i) = 0 \\ \dfrac{dS}{db_1} = 2\sum (y_i - b_0 - b_1 x_i)\cdot x_i = 0 \end{cases}$$

После преобразования получим систему нормальных уравнений:

$$\begin{cases} b_0 n + b_1 \sum x_i = \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum y_i x_i \end{cases}$$

решая которую относительно b_0 и b_1 получаем:

$$\begin{cases} b_1 = \dfrac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} \\ b_0 = \dfrac{1}{n}\sum y_i - b_1 \dfrac{1}{n}\sum x_i \end{cases} \quad \text{или} \quad \begin{cases} b_1 = \dfrac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x^2} \\ b_0 = \bar{y} - b_1\bar{x} \end{cases}.$$

Многофакторная модель с линейной связью признаков имеет вид:

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$.

Коэффициенты регрессии b_j при факторных признаках называется <u>чистыми.</u>

Коэффициенты чистой регрессии показывает, на сколько единиц изменится результативный признак при изменении соответствующего факторного признака на единицу своего измерения, при условии, что другие факторы, включенные в уравнение, зафиксированы на одном уровне.

Регрессионный анализ дополняют расчетом показателя силы корреляционной связи – коэффициента эластичности.

Коэффициент эластичности показывает на сколько % изменится в среднем значение результативного признака при изменении факторного признака на 1 % .

$$\bar{Э}_j = b_j \frac{\bar{x}_j}{\bar{y}}.$$

Величина указывает на силу реакции результативного признака на относительный прирост фактора.

Если $|\bar{Э}_j| \le 1$ – эластичность слабая;

$|\bar{Э}_j| = 1$ - эластичность средняя;

$|\bar{Э}_j| \ge 1$ - эластичность высокая.

Во множественной корреляции по коэффициентам эластичности проводится сравнительный анализ силы влияния факторов на результативный признак.

II. Различают <u>линейную и нелинейную</u> форму связи.

Нелинейная форма парной модели может быть выражена любым видом криволинейной функции:

- $\hat{y} = b_0 + b_1 \ln x$ – полулогарифмическая функция
- $\hat{y} = b_0 + b_1^{\,x}$ – показательной
- $\hat{y} = b_0 x^{b_1}$ – степенной
- $\hat{y} = b_0 + b_1 \dfrac{1}{x}$ - гиперболической и т.д

Многофакторная модель с нелинейной формой связи может выглядеть различным образом, например:

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2^2$.

Для правильного выбора формы связи необходимо провести глубокий теоретический анализ, а затем подкрепить свои логические предположения графическим методом, методом статистических группировок, сопоставлением параллельных рядов.

3. Проверка статистической значимости уравнения регрессии в целом и его параметров

Так как выборочные характеристики являются случайными величинами, то проверим значимость уравнения регрессии и его параметров.

Значимость уравнения регрессии в целом оценивается с поищью F критерия Фишера и на основе дисперсии анализа.

1. Для проверки значимости уравнения регрессии, т.е гипотезы H_0: $\beta_1=0$ (Для множественной регрессии H_0: $\beta_1=\beta_2=\ldots=\beta_k$), используют критерий основанный на статистике:

$$F_{\text{набл}} = \frac{Q_R / k}{Q_{\text{ост}} /(n-2)} = \frac{S_R^2}{S_{\text{ост}}^2},$$

При истинности гипотезы H_0 приведенная статистика имеет F распределение с числом степеней свободы $v_1=k$, $v_2=n-2$.

Q_R- сумма квадратов отклонений, обусловленных регрессией,

$$Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 .$$

$S_R^2 = \dfrac{Q_R}{k}$ - дисперсия, обусловленная регрессией.

Qост = сумма квадратов отклонений обусловленнфя остаточными, не включенными в модель факторами.

$$Q_{\text{ост}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 .$$

$S_{\text{ост}}^2 = \dfrac{Q_{ост}}{n-k-1}$ - остаточная дисперсия.

Результаты расчетов, как правило, оформляют в виде таблицы

Таблица 1- Дисперсионный анализ

Компоненты дисперсии	Число степеней свободы df	Сумма квадратов	Дисперсии	F критерий	
				фактический	табличный
Регрессия	k	$Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$S_R^2 = \frac{Q_R}{k}$	$F_{\text{набл}} = \frac{S_R^2}{S_{\text{ост}}^2}$	$F_{\text{кр}}(\alpha; v_1; v_2)$
Остаточная	*n-k-1*	$Q_{\text{ост}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$S_{\text{ост}}^2 = \frac{Q_{ост}}{n-k-1}$		
Общая	*n-1*	$Q_{\text{общ}} = \sum_{i=1}^{n}(y_i - \bar{y})^2$			

Табличное значение находим по таблице « F –распределения Фишера» при уровне значимости α и числе степеней свободы $v_1=k$, $v_2=n-k-1$, $F_{\text{кр}}$ (α, $v_1=k$, $v_2=n-k-1$).

Если $F_{\text{набл}} > F_{\text{крит}}$, то уравнение регрессии значимо (достоверно) с вероятностью ошибки α. Гипотеза H_0 – отвергается. В противном случае гипотеза не отвергается.

2.Для оценки значимости коэффициентов (параметров) регрессии используется критерий t- нормального распределения (большие выборки) или t- Стьюдента (малые выборки).

H_0: $\beta_1=0$

H_1: $\beta_1 \neq 0$

1)Рассчитывается средняя (стандартная) ошибка выборки для коэффициента регрессии

$$m_{b_1} = \sqrt{\frac{S^2_{\text{ост}}}{\sigma_x^2 \cdot n}} \text{ или } m_{b_1} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot (n-2)}}$$

2)Рассчитывается $t_{\text{набл}} = \dfrac{b_1}{m_{b_1}}$

3) Сравнивается $t_{\text{набл}}$ с $t_{\text{крит}}$ (α, $v=n-2$)

Если $|t_{\text{набл}}| > t_{\text{крит}}$, то отвергают H_0 и с вероятностью $\gamma=1-\alpha$ принимают альтернативную гипотезу о значимости коэффициента регрессии b_1;

$|t_{\text{набл}}| < t_{\text{крит}}$ – гипотеза H_0 принимается, с вероятностью ошибки α.

Интервальные оценки параметров регрессии

Если гипотеза H_0: $\beta=0$ отвергается, то представляет интерес определение с надежностью γ интервальных оценок параметров регрессии.

-для коэффициента регрессии β_1

$b_1 - t_\alpha m_{b_1} \le \beta_1 \le b_1 + t_\alpha m_{b_1}$;

-для интервала предсказания в точке $x= x_{n+1}$

$$\tilde{y}_{n+1} \in \left[\hat{y}_{n+1} \pm t_\alpha \cdot \widehat{S} \cdot \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \right]$$

t_α - определяются по таблице распределения Стьюдента.

При $n > 30$ можно пользоваться - «Нормальный закон распределения».

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

3.1 Практическое занятие №1 (4 часа).

Тема: «Случайные события. Вероятность события»

1 Задание для работы:

Вопросы к занятию:

1. Теория вероятностей как наука. Краткая историческая справка
2. Случайные события и их виды
3. Операции над событиями
4. Классическое определение вероятности
5. Комбинаторика. Правила сложения и умножения комбинаторики
6. Размещения, перестановки сочетания без повторений
7. Геометрическое определение вероятности
8. Статистическое определение вероятности

Типовые тесты (для контроля усвоения знаний):

1. Математическая наука, изучающая закономерности случайных явлений:

+а) теория вероятностей;

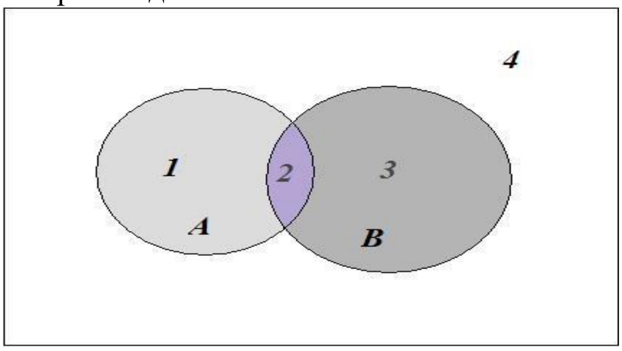
б) комбинаторика;

в) математическая статистика;
г) выборочное наблюдение;
д) дисперсионный анализ.

2. Любой факт, который может либо произойти, либо не произойти при выполнении некоторого комплекса условий:
+а) случайное событие;
б) случайная переменная;
в) вероятность события;
г) независимые испытания;
д) элемент комбинаторики

3. Из колоды карт извлекается карта случайным образом. Возможны следующие случайные события: А – извлечение пиковой дамы; В – извлечение карты черной масти; С – извлечение карты красной масти; D – извлечение короля. Совместные события:
а) В и С;
+б) А, В;
в) А, D;
г) С, D;
д) А, С.

4. Произведением событий А и В является область:



ОТВЕТ: 2

5. Отношение числа благоприятствующих событию А исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу - … определение вероятности:
ОТВЕТ: классическое

6. Соответствие между термином и видом вероятности: 1) отношение числа благоприятствующих исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу; 2) частость (относительная частота) *m/n* появления события в n произведенных испытаниях; 3) отношение меры области g, благоприятствующей событию *A* к мере всей области *G*
1 а) классическая;
2 б) статистическая;
3 в) геометрическая.

7. Сколькими способами можно отобрать 3 студентов из 10 для прохождения практики в трех различных предприятиях:
а) 3;
б) 10;

в) 120;
+г) 720.

8. Любые упорядоченные множества, состоящие из одних и тех же n различных элементов и отличающиеся только порядком их расположения называются:
Ответ: перестановки
Ответ: перестановками

9. При испытании партии приборов относительная частота годных приборов оказалась равной 0,8. Если всего было проверено 100 приборов, то число годных приборов составило:
Ответ: 80

Типовые задачи:

1. Рассмотрим случайные события – выпадение определенного числа на верхней грани, которые могут произойти при бросании простого 6-гранного игрального кубика.

Обозначения случайных событий:

А – выпадение числа 2, В – выпадение числа 3, С – выпадение нечетного числа, D – выпадение любого из чисел 1, 3 или 5.

Какие события являются достоверными, невозможными, несовместными, совместными, равносильными и равновозможными?

2. Брошены две игральные кости. Найти вероятность, что а) на гранях кубиков выпадут одинаковые цифры; б) сумма выпавших очков четная, при этом на грани хотя бы одной из костей появится тройка; в) сумма выпавших очков равна восьми, а разность четырем; г) сумма выпавших очков равна шести, если известно, что их разность равна четырем.

3. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера, которые затем тщательно перемешаны. Найти вероятность того, что наудачу извлеченный кубик имеет окрашенных граней а) одну; б) две; в) три; г) хотя бы одну.

4. Сколькими способами можно случайным образом из 25 лучших студентов курса выбрать 2-х для поездки в Англию и Америку?

5. Из 50 сотрудников фирмы 30 человек владеют английским языком. Для участия в международной конференции случайным образом отбирается 5 человек. Какова вероятность того, что а) все выбранные сотрудники владеют английским языком; б) 3 человека из 5 выбранных знают английский язык; в) хотя бы один владеет английским языком.

6. Для изучения качества произведенных деталей техническим отделом было обследовано 200 деталей, из которых 10 оказались бракованными. Найти вероятность изготовления бракованных деталей.

2 Краткое описание проводимого занятия:

1. Входной контроль
2. Решение задач по теме занятия. При решении задач необходимо акцентировать внимание правильном применении формулы расчета вероятности.
3. Самостоятельная работа по решению задач
4. Проверка подготовленности студентов к занятию

3 Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.2 Практическое занятие №2 (4 часа).

Тема: «Теоремы сложения и умножения вероятностей»

1 Задание для работы:

Вопросы к занятию:

1. Теорема сложения для несовместных событий
2. Теорема сложения для совместных событий
3. Теорема умножения независимых событий
4. Теорема умножения зависимых событий
5. Формула полной вероятности
6. Формула Байеса

Типовые тесты (для контроля усвоения знаний):

1. Если А и В несовместные события, то Р(А+В) равно:
+а) Р(А)+Р(В);
б) Р(А)+Р(В) – Р(АВ);
в) Р(А)+Р(В) – P_A(В);
г) Р(А)+P_A(В);
д) Р(А)+P_A(В) – P_B(А).

2. Вероятность произведения двух … событий А и В равна произведению их вероятностей:
а) совместных;
б) несовместных;
в) зависимых;
+г) независимых;
д) условных.

3. Вероятность события А, вычисленная при условии, что событие В произошло, называется:
Ответ: условная
Ответ: условной

4. Если события А и В независимы, то вероятность произведения событий Ā и В вычисляется по формуле:
а) Р(А)·Р(В);
б) Р(Ā)·Р(В) – Р(АВ);
+в) Р(Ā)·Р(В);
г) Р(Ā)·P_A(В);
д) Р(Ā)· P_B(А).

5. В урне 2 белых и 3 черных шара. Из урны наудачу извлекают 2 шара без возвращения. Вероятность того, что оба шара белые равна:
а) 0,4;
б) 1,0;
+в) 0,1;
г) 0,2;
д) 0,3.

6. Формула, позволяющая «пересмотреть» вероятности гипотез с учетом наблюденного результата опыта, по мере получения новой информации:
ОТВЕТ: Байеса
Ответ: Байес

7. Вероятность P(B) события B, которое может произойти только при условии появления одного из событий (гипотез) A_1, A_2, … , A_n, образующих полную группу попарно несовместных событий, вычисляется по формуле … - $P(B)=\sum_{i=1}^{n}P(A_i)\cdot P(B\mid A_i)$.
а) Байеса;
+б) полной вероятности;
в) Лапласа;
г) Бернулли;
д) Пуассона.

8. Вероятность наступления событий, образующих полную группу, $P(A_1+A_2+...+A_n)$ равна:
+а) 1;
б) 0;
в) от 0 до 1;
г) ∞;
д) $-\infty$.

9. Если события A и B зависимы, то вероятность произведения событий $\overline{A}$ и $\overline{B}$ вычисляется по формуле:
а) P(A)·P(B);
+б) $P(\bar{A})\cdot P(\overline{B}) - P(\overline{A}\ \overline{B})$;
в) $P(\overline{A})\cdot P(\overline{B})$;
г) $P(\bar{A})\cdot P_A(\overline{B})$;
д) $P(\overline{B})\cdot P_B(A)$.

10. В ящике 21 деталь, из которых 11 окрашены в синий цвет. При перевозке была утерена одна деталь. После этого наудачу извлекается одна деталь. Условная вероятность того, что извлечена синяя деталь, если была утеряна деталь синего цвета равна:
Ответ: 0,5

Типовые задачи:

1. Рабочий обслуживает три станка, работающих независимо друг от друга. Вероятность того, что в течение часа первый станок не потребует внимания рабочего равна 0,9; второй – 0,8; третий – 0,85. Какова вероятность того, что в течение часа :
а) ни один станок не потребует внимания рабочего;
б) все три станка потребуют внимания рабочего;
в) какой-нибудь один станок потребует внимания рабочего;
г) хотя бы один станок потребует внимания рабочего.

2. В цехе работают семь мужчин и три женщины. По табельным номерам наудачу отобраны три человека. Найти вероятность того, что все отобранные лица окажутся мужчинами.

3. Саженец яблони приживается с вероятностью 0,6; груши – 0,5; винограда – 0,4. В саду было посажено по одному дереву каждого вида. Прижилось два саженца. Какое событие при этом более вероятно: саженец винограда прижился или саженец винограда не прижился.

2 Краткое описание проводимого занятия:

1. Устный опрос и / или тестирование
2. Письменно решение задач

3 Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.3 Практическое занятие №3 (2 часа).

Тема: «Повторные независимые испытания»

1 Задание для работы:

Вопросы к занятию:

1. Формула Бернулли
2. Локальная теорема Муавра-Лапласа
3. Интегральная теорема Муавра-Лапласа
4. Вероятность отклонения относительной частоты от постоянной вероятности в независимых испытаниях
5. Теорема Пуассона

Типовые тесты (для контроля усвоения знаний):

1. При небольшом числе n испытаний вероятность того, что в n независимых испытаниях событие А наступит ровно k раз определяется:
+а) по формуле Бернулли;
б) по локальной теореме Муавра-Лапласа;
в) по интегральной теореме Муавра-Лапласа;
г) по теореме Пуассона;
д) по формуле Байеса.

2. Теорема Пуассона применяется для вычисления вероятности того, что в n независимых испытаниях событие А наступит ровно k раз, если выполняется условие:
а) $npq \geq 20$;
б) $\boldsymbol{p \to 1}$;
+в) $\boldsymbol{p \to 0}$;
г) $k \leq 10$;
д) $0 < p < 1$.

3. Для вычисления $P_n(k)$ по локальной теореме Муавра-Лапласа используется функция:
а) стандартного нормального закона распределения;
+б) плотности стандартного нормального закона распределения;
в) Z – распределения Фишера;
г) биномиального распределения;
д) t – распределения Стьюдента.

4. Для вычисления $P_n(a \leq k \leq b)$ по интегральной теореме Муавра-Лапласа используется функция:
+а) стандартного нормального закона распределения;
б) плотности стандартного нормального закона распределения;
в) Z – распределения Фишера;

г) биномиального распределения;
д) t – распределения Стьюдента.

5. Некоторый стрелок попадает в цель с вероятностью 0,6, он собирается произвести 6 выстрелов. Вероятность того, что он попадет в цель 3 раза, равна:
а) 0,002;
+б) 0,276;
в) 0,5;
г) 0,216;
д) 0,996.

Типовые задачи:

Вероятность того, что посетитель продуктового магазина купит молоко равна 0,7. Определите вероятность того, что а) из 10 покупателей ровно 4 купят молоко; б) из 80 покупателей ровно 50 купят молоко; в) из 100 покупателей от 65 до 80 человек купят молоко.

2 Краткое описание проводимого занятия:

1. Письменный опрос и / или тестирование
2. Проверка вопросов для самостоятельного изучения
3. Проверка письменной подготовки к занятию
4. Решение задач.

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.4 Практическое занятие № 4 (4 часа).

Тема: «Дискретная случайная величина»

1 Задание для работы:

Вопросы к занятию:

1. Определение ДСВ. Закон распределения случайной величины
2. Способы задания закона распределения ДСВ
3. Свойства интегральной функции распределения ДСВ
4. Основные числовые характеристики ДСВ
5. Одинаково распределенные взаимно независимые случайные величины
6. Биномиальный закон распределения ДСВ
7. Распределение Пуассона
8. Геометрическое распределение

Типовые тесты (для контроля усвоения знаний):

1. Какая из следующих случайных величин является дискретной:
+а) число вкладчиков банка;
б) рентабельность предприятия;
в) средняя зарплата работников;
+г) количество станков цеха;
д) темп роста ВВП.

2. Способы задания закона распределения дискретной случайной величины:

+а) ряд распределения;
+б) функция накопленных вероятностей;
в) нормальный закон распределения;
г) корреляционная функция;
д) функция плотности вероятностей.

3. Чему равно значение функции распределения *F(X<1)* для случайной величины X, принимающей значение меньшее 1, для данного ряда распределения

x_i	0	1	2	3	4
p_i	0,1	0,2	0,4	0,2	0,1

Ответ: 0,1

4. Для интегральной функции распределения дискретной случайной величины X справедливо свойство:
+а) функция распределения может принимать любые значения от 0 до 1;
б) функция распределения непрерывна в любой точке;
в) функция распределения является невозрастающей;
г) функция распределения при $x \le x_{\min}$ равна 1, при $x \ge x_{\max}$ равна 0;
+д) если значения X расположены на всей числовой оси, то $F(+\infty)=1$.

5. Распределение, описывающее распределение вероятностей случайной величины X, представляющей собой число испытаний, которое нужно провести до первого появления события А, которое имеет одну и ту же вероятность p:
а) биномиальное;
б) Пуассона;
+в) геометрическое;
г) равномерное;
д) Бернулли.

6. Закон редких событий – это распределение:
а) биномиальное;
+б) Пуассона;
в) геометрическое;
г) равномерное;
д) Бернулли.

7. Законы распределения дискретной случайной величины:
+а) биномиальное;
+б) пуассоновское;
+в) геометрическое;
г) нормальное;
д) равномерное;
е) экспоненциальное.

Типовые задачи:

1. Случайная величина X задана законом распределения:

X	1	2	3	4
p	0,2	0,4	0,3	0,1

Найти функцию распределения случайной величины X и построить ее график. Рассчитать математическое ожидание, дисперсию, среднее квадратическое отклонение.

2. Вероятность опоздания студента на первую пару равна 0,4. Написать биномиальный закон распределения числа опозданий студента из четырех дней занятий. Построить график распределения вероятностей. Найти функцию распределения числа опоздавших поездов и построить ее график. Рассчитать математическое ожидание, дисперсию и среднеквадратическое отклонение.

3. В приемное время врача-педиатра посещает в среднем 7 человек в час. Составить закон распределения числа пациентов, посетивших педиатра в течение часа.

2 Краткое описание проводимого занятия:

1. Письменно решение задач
2. Проверка самостоятельного изучения вопросов
3. Проверка подготовки к занятиям
4. Устный опрос и / или тестирование

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.5 Практическое занятие №5 (4 часа).

Тема: «Непрерывная случайная величина»

1 Задание для работы:

Вопросы к занятию:

1. Функция распределения НСВ и ее свойства
2. Определение функции плотности вероятностей и ее свойства
3. Вероятностный смысл плотности распределения
4. Основные числовые характеристики НСВ
5. Равномерный закон распределения
6. Экспоненциальный закон распределения
7. Нормальный закон распределения
8. Влияние параметров нормального закона распределения на форму нормальной кривой
9. Свойства случайной величины, распределенной по нормальному закону распределения

Типовые тесты (для контроля усвоения знаний):

1. Законы распределения непрерывной случайной величины:
а) биномиальное;
б) пуассоновское;
в) геометрическое;
+г) нормальное;
+д) равномерное;
+е) экспоненциальное.

2. Кривой распределения называют:
+а) график плотности распределения;
б) график функции распределения;
в) многоугольник распределения;
г) линию регрессии;
д) корреляционное поле.

3. Площадь криволинейной трапеции, ограниченной осью OX, кривой распределения и прямыми $x = a,\ x = b$, есть геометрическая интерпретация:

+а) $\int_{a}^{b} f(x)dx$;

+б) $F(b) - F(a)$;

+в) $P(a \le x \le b)$;

г) $M(X^2)-(M(X))^2$;

д) $\int_{-\infty}^{\infty} \tilde{o} f(x)dx$.

4. Утверждение верное о функции распределения непрерывной случайной величины:

+а) всюду непрерывна и имеет непрерывную производную, кроме, возможно, некоторых точек;

б) разрывная ступенчатая функция, скачки которой происходят в точках эксремума;

в) функция распределения при $x \le x_{\min}$ равна 1, при $x \ge x_{\max}$ равна 0;

г) функция распределения есть производная первого порядка функции плотности распределения;

+д) может принимает любые значения от 0 до 1.

Типовые тесты (для контроля усвоения знаний):

5. Функция распределения случайной величины: $F(x) = \begin{cases} 0, & при\ \ x < 0, \\ 1 - e^{-5x}, & при\ \ x \ge 0 \end{cases}$, тогда вероятность того, что случайная величина примет значение, равное 1, т.е. P(x=1) равно:

+а) 0;

б) $1 - e^{-5}$;

в) e^{-5};

г) 1;

д) 5.

6. Вероятность попадания непрерывной случайной величины в отрезок [-1; 2] равна 0,25, т.е. $P(-1 \le x \le 2) = 0{,}25$. Определить $P(-1 < x < 2)$:

а) 0;

+б) 0,25;

в) 1;

г) 2;

д) 0,75.

7. Для экспоненциального закона распределения справедливо свойство:

а) M(X) = D(X);

+б) M(X) = σ(X);

в) M(X) = λ;

г) D(X) = σ(X);

д) M(X) = λσ(X).

8. Каким числом параметров задается нормальный закон распределения:

а) 0;

б) 1;

+в) 2;

г) 3;

д) 4.

9. Нормальный закон распределения с математическим ожиданием, равным нулю, и средним квадратическим отклонением, равным единице, называется:
а) начальным;
б) элементарным;
+в) стандартным;
г) единичным;
д) нулевым.

Типовые задачи:

1. Задана дифференциальная функция распределения случайной величины X:

$$\boldsymbol{f(x)} = \begin{cases} \mathbf{0,\ при\ x \le 3} \\ \mathbf{\frac{1}{2},\ при\ 3 < x \le 5} \\ \mathbf{0,\ при\ x > 5} \end{cases}$$

Рассчитать: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X; 2) вероятность того, что в результате испытаний случайная величина примет значение большее 3 и меньшее 5. 3) Построить график функции F(x).

2. Случайная величина X равномерно распределена на отрезке [0; 2]. Построить функцию распределения случайной величины X. Рассчитать: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X; б) вероятность того, что в результате испытаний случайная величина примет значение большее 0,5 и меньшее 1,5. Построить график функции F(X).

3. При сортировке случайные значения веса зерна распределены нормально со средним значением 0,15 г. и средним квадратическим отклонением 0,03 г. Нормальные всходы дают зерна, вес которых более 0,10 г.

Определить: а) процент семян, от которых следует ожидать нормальные всходы;
б) величину, которую не превзойдет вес отдельного зерна с вероятностью 0,99.

2 Краткое описание проводимого занятия:

1. Проверка подготовки к занятию самостоятельно изученных вопросов
2. Устный опрос и / или тестирование
3. Самостоятельная работа по решению задач
4. Решение задач

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.6 Практическое занятие № 6 (4 часа).

Тема: «Закон больших чисел. Понятие о методе Монте-Карло и цепях Маркова»

1 Задание для работы:

Вопросы к занятию:

1. Сущности закона больших чисел

2. Основные теоремы закона больших чисел
3. Центральная предельная теорема Ляпунова
4. Сущность метода Монте-Карло
5. Оценка погрешности метода Монте-Карло
6. Правила разыгрывания ДСВ и НСВ
7. Правила разыгрывания полной группы событий
8. Приближенное разыгрывание нормальной случайной величины
9. Цепь Маркова
10. Однородная цепь Маркова. Переходные вероятности. Матрица перехода
11. Равенство Маркова

Типовые тесты (для контроля усвоения знаний):
1. Теорема, устанавливающая теоретическое обоснование замены неизвестной вероятности события его частостью или статистической вероятностью:
+а) Бернулли;
б) Чебышева;
+в) Пуассона;
г) Ляпунова;
д) Колмогорова.

2. Теорема, устанавливающая при определенных условиях, устойчивость средней арифметической:
а) Бернулли;
+б) Чебышева;
в) Пуассона;
г) Ляпунова;
д) Колмогорова.

3. Теорема, устанавливающая условия, при которых случайная величина распределена по нормальному закону распределения:
а) Бернулли;
б) Чебышева;
в) Пуассона;
+г) Ляпунова;
д) Колмогорова.

Типовые задачи:
1. Разыграть шесть опытов по схеме Бернулли: опыт состоит из четырех испытаний, в каждом из которых вероятность появления события А равна 0,5.
2. Разыграть пять возможных значений нормальной случайной величины в параметрами $a=10$, $\sigma=2$.

2 Краткое описание проводимого занятия:
1. Проверка подготовленности студентов к занятию
2. Устный опрос
3. Письменный опрос
4. Письменно решение задач.

Результаты и выводы:
Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.7 Практическое занятие № 7 (4 часа).

Тема: «Статистическое оценивание параметров распределения»

1 Задание для работы:

Вопросы к занятию:

1. Математическая статистика и ее методы
2. Общие сведения о выборочном методе
3. Статистическое распределение выборки и эмпирическая функция распределения
4. Понятие оценки параметров
5. Свойства точечных оценок
6. Некоторые характеристики вариационного ряда
7. Методы нахождения оценок
8. Оценка параметров генеральной совокупности по собственно-случайной выборке
9. Интервальные оценки параметров генеральной совокупности

Типовые тесты (для контроля усвоения знаний):

1. Для обработки результатов эксперимента используют методы математической статистики:
+а) описательные и аналитические;
+б) параметрические и непараметрические;
в) вероятностные и классические;
г) дискретные и интервальные;
д) вариационные и атрибутивные.

2. Для изображения непрерывного вариационного ряда используют:
+а) гистограмму;
б) полигон;
в) секторные диаграммы;
г) картограмму;
д) кумуляту.

3. Значения признака вариационного ряда называются:
+а) вариантами;
б) частотами;
в) частостями;
г) моментами;
д) весами.

4. Соответствие между наблюдаемыми вариантами и их частотами или частостями называется:
+а) статистическим распределением;
+б) вариационным рядом;
в) функцией распределения;
г) оценкой параметра;
д) согласованностью распределений.

5. Статистическая оценка $\tilde{\Theta}$ параметра Θ называется … , если ее математическое ожидание равно оцениваемому параметру Θ .
+а) несмещенной;

б) состоятельной;
в) эффективной;
г) независимой;
д) достаточной

6. Начальный момент первого порядка равен:
а) дисперсии;
б) среднему квадратическому отклонению;
+в) средней арифметической;
г) средней геометрической;
д) среднему линейному отклонению.

7. Метод нахождения оценок, заключающийся в приравнивании определенного количества выборочных моментов к соответствующим теоретическим моментам распределения случайной величины:
+а) моментов;
б) наибольшего правдоподобия;
в) наименьших квадратов;
г) взвешенный метод наименьших квадратов;
д) выборочный.

8. Метод определения параметра Θ, заключающийся в минимизации суммы квадратов отклонений выборочных данных от определяемой оценки Θ:
а) моментов;
б) наибольшего правдоподобия;
+в) наименьших квадратов;
г) взвешенный метод наименьших квадратов;
д) главных компонент.

9. Интервал $\left[\tilde{\Theta}-\Delta;\ \tilde{\Theta}+\Delta\right]$параметра Θ, который с определенной близкой к единице вероятностью, содержит значение параметра Θ, называется:
а) группировочным;
+б) доверительным;
в) прогнозным;
г) ошибочным;
д) предсказания.

10. На контрольных испытаниях n=25 электроламп найдено, что средний срок службы ламп равен 950 часов. Предположив, что срок службы электроламп распределен по нормальному закону с $\sigma_{ген}$=15 ч и надежностью γ=0,8664 предельная ошибка выборки составит:
+а) 4,5;
б) 1,5;
в) 0,9;
г) 6,19;
д) 2,6.

Типовые задачи:

1. На предприятии для анализа производительности труда случайным образом было отобрано 40 человек. Получены следующие данные:

Произведено изделий в час	10-14	14-18	18-22	22-26	26-30

Количество работников	5	8	15	10	2

Найти: а) среднее число изделий, произведенных в час, дисперсию, среднее квадратическое отклонение и коэффициент вариации;
б) моду и медиану;
в) коэффициент асимметрии и эксцесса.

2. Проведено исследование коммерческих фирм по затратам на рекламу в год. Для этого случайным образом отобрано 50 фирм. Результаты представлены в таблице:

Расходы на рекламу, тыс. руб.	Менее 20	20-40	40-60	60-80	80-100	100 и более
Число фирм	3	5	9	16	13	4

Найти: а) границы, в которых с вероятностью 0,95 заключен средний размер расходов на рекламу;
б) вероятность, с которой средний расход на рекламу будет отличаться от выборочной средней не более чем на 5 тыс. руб.

2 Краткое описание проводимого занятия:

1. Проверка уровня освоения студентами вопросов для самостоятельного изучения
2. Устный опрос и тестирование
3. Решение задач

3.8 Практическое занятие № 8 (4 часа).

Тема: «Статистическая проверка статистических гипотез»

1 Задание для работы:

Вопросы к занятию:

1. Статистическая гипотеза. Виды гипотез
2. Ошибки первого и второго рода
3. Статистический критерий проверки нулевой гипотезы. Критическая область. Область принятия решений
4. Мощность критерия
5. Сравнение двух средних нормальных генеральных совокупностей с неизвестными дисперсиями
6. Критерий Бартлетта
7. Критерий Кочрена
8. Критерий Вилкоксона и проверка гипотезы об однородности двух выборок
9. Критерий согласия хи-квадрат Пирсона.

Типовые тесты (для контроля усвоения знаний):

1. Статистическими являются гипотезы:
+а) средняя выработка рабочих цеха за месяц равна 121%;
+б) число вызовов скорой помощи распределено по закону Пуассона;
в) студент на экзамене получит положительную оценку;
г) стрелок при выстреле попадет в цель;
д) завтра температура воздуха составит 20°С.

2. При проверке статистической гипотезы уровнем значимости α называется вероятность:
+а) совершить ошибку первого рода;
б) совершить ошибку второго рода;
в) принять правильное решение;
г) справедливости нулевой гипотезы;
д) справедливости конкурирующей гипотезы.

3. Соответствие между понятиями и их разновидностями: 1) гипотеза; 2) ошибка; 3) критическая область; 4) значение критерия
1 а) простые и сложные;
2 б) первого и второго рода;
3 в) односторонние и двухсторонние;
4 г) наблюдаемые и критические.

4. Критерий, основанный на использовании различных мер расстояний между анализируемой эмпирической функцией распределения, определенной по выборке и функцией F(x) распределения генеральной совокупности X.
+а) согласия Пирсона;
б) Вилкоксона;
в) Кочрена;
г) Стьюдента;
д) Фишера.

5.Однородность двух выборок проверяют с помощью критерия:
а) Пирсона;
+б) Вилкоксона;
в) Кочрена;
г) Стьюдента;
д) Фишера.

Типовые задачи:

Для изучения среднего возраста работников предприятия было отобрано 100 работников случайным образом. Результаты исследования представлены в таблице:

Возраст, лет	Менее 25	25-30	30-35	35-40	40-45	45-50	50-55	Более 55
Число работников	6	8	12	18	19	17	14	6

Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении случайной величины X – возраст работника с эмпирическим распределением выборки.

2 Краткое описание проводимого занятия:

1. Письменный опрос
2. Решение задач

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.9 Практическое занятие № 9 (2 часа).

Тема: «Дисперсионный анализ»

1 Задание для работы:

Вопросы к занятию:

1. Понятие о дисперсионном анализе
2.Модель однофакторного дисперсионного анализа
3. Основные предпосылки дисперсионного анализа
4. Основное тождество дисперсионного анализа
5. Таблица дисперсионного анализа
6. Сравнение нескольких средних методом дисперсионного анализа

Типовые тесты (для контроля усвоения знаний):

1. Вид статистического анализа, который раскладывает общую вариацию результативного показателя на составляющие: вариацию обусловленную влиянием вариации фактора и вариацию обусловленную влиянием неучтенных факторов:
+а) дисперсионный;
б) регрессионный;
в) корреляционный;
г) вариационный;
д) выборочный.

2. Дисперсия, характеризующая вариацию результативного признака, обусловленную вариацией факторного признака:
а) общая;
б) остаточная;
+в) факторная;
г) внутригрупповая;
д) полная.

3. Основное тождество дисперсионного анализа:
+а) $Q_{общ} = Q_{ост} + Q_{факт}$;
б) $Q_{ост} = Q_{общ} + Q_{факт}$;
в) $Q_{факт} = Q_{ост} - Q_{общ}$;
г) $Q_{общ} = Q_{ост} / Q_{факт}$;
д) $Q_{ост} = Q_{фкт} / Q_{общ}$.

4. Вариация себестоимости продукции составляет 220 рубля, вариация себестоимости продукции, обусловленная влиянием вариации производительности труда, составляет 170 руб. Тогда вариация себестоимости продукции, сформировавшаяся под влиянием неучтенных факторов, составляет:
а) 390;
+б) 50;
в) 1,29;
г) 0,23
д) 4,4.

5. Проверка основной гипотезы дисперсионного анализа проводится с помощью критерия:
а) Стьюдента;
+б) Фишера;
в) нормального распределения;

г) Колмогорова;
д) Пирсона.

Типовые задачи:

В процессе изучения влияния типа станка на сменную выработку рабочего получены следующие данные:

Тип станка		
Тип 1	Тип 2	Тип 3
54	58	64
58	62	62
57	64	63

Предполагая, что фактор типа станка имеет фиксированные уровни, а сменная выработка рабочего есть случайная величина, имеющая нормальный закон распределения, требуется:

а) проверить на уровне значимости α=0,05 существенность влияния ипа станка на сменную выработку рабочего;

б) проверить на уровне значимости α=0,01 гипотезу об общей средней H_0: μ=60.

2 Краткое описание проводимого занятия:

1. Проверка уровня освоения студентами вопросов для самостоятельного изучения
2. Устный опрос
3. Самостоятельное решение задач
4. Решение задач

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.10 Практическое занятие № 10 (2 часа).

Тема: «Корреляционный анализ»

1 Задание для работы:

Вопросы к занятию:

1. Функциональная, статистическая и корреляционная зависимости
2. Расчет коэффициента корреляции и детерминации, их интерпретация
3. Корреляционное отношение и индекс корреляции
4. Проверка статистической значимости коэффициента корреляции: с использованием таблиц Фишера-Иейтса, распределения Стьюдента
5. Интервальная оценка генерального коэффициента корреляции

Типовые тесты (для контроля усвоения знаний):

1. Основная задача … анализа – выявление связи между случайными переменными и оценка ее тесноты:
+а) корреляционного;
б) регрессионного;
в) дисперсионного;
г) вариационного;
д) выборочного.

2. Отношение факторной дисперсии к остаточной называется:
а) эмпирическим корреляционным отношением;
б) теоретическим корреляционным отношением;
+в) коэффициентом детерминации;
г) индексом корреляции;
д) линейным коэффициентом корреляции.

3. Линейный коэффициент корреляции, который указывает на наибольшую тесноту связи?
а) r =0,80;
б) r = -0,45;
+в) r = -0,85;
г) r = 0;
д) r = 0,5.

4. При изучении зависимости производительности труда от квалификации рабочих получили коэффициент детерминации равный 45,5%, это значит, что:
+а) фактор квалификации рабочих объясняет 45,5% вариации производительности труда;
б) зависимость между показателями умеренная и прямая;
в) при увеличении квалификации рабочих на 1 % производительность труда повысится на 45,5%;
г) наблюдается сильная вариация показателей;
д) фактор квалификации рабочих объясняет 54,5% вариации производительности труда.

5. Проверка нулевой гипотезы о значимости коэффициента корреляции H_0: r = 0, проводится с помощью:
+а) критерия Стьюдента;
б) критерия Фишера;
+в) таблицы Фишера-Иейтса;
г) Z – преобразования Фишера;
д) критерия Колмогорова.

Типовые задачи:

На основе выборочных данных о производительности труда (X), измеряемой в млн. руб. на человека, и себестоимости продукции (У), измеряемой в тыс. руб. на единицу продукции, полученных с однотипных предприятий за месяц:

X	5	4	3	20	10	15
У	7	10	12	2	5	4

Найти: а) выборочный коэффициент корреляции между X и У;

б) проверить значимость генерального коэффициента корреляции при α=0,05;

в) с надежностью γ=0,95 найти границы доверительного интервала генерального коэффициента корреляции.

2 Краткое описание проводимого занятия:

1. Проверка уровня освоения студентами вопросов для самостоятельного изучения
2. Устный опрос
3. Самостоятельное решение задач
4. Решение задач

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

3.11 Практическое занятие № 11 (2 часа).

Тема: «Регрессионный анализ»

1 Задание для работы:

Вопросы к занятию:

1. Понятие и предпосылки регрессионного анализа
2. Виды регрессионных моделей. Оценка параметров парной линейной регрессии
3. Проверка статистической значимости уравнения регрессии в целом и его параметров
4. Простейшие случаи криволинейной корреляции

Типовые тесты (для контроля усвоения знаний):

1. Регрессионная модель вида $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_2^2 + \varepsilon_i$ называется:
а) линейной однофакторной;
б) линейной многофакторной;
в) нелинейной однофакторной;
г) криволинейной многофакторной;
д) парной линейной.

2. Проверка значимости уравнения регрессии осуществляется на основе:
а) t - критерия Стьюдента;
б) множественного коэффициента корреляции;
в) F - критерия Фишера;
г) средней ошибке аппроксимации;
д) таблиц Фишера-Иейтса.

3. Метод определения параметров линейного уравнения регрессии:
а) моментов;
б) наименьших квадратов;
в) систем нормальных уравнений;
г) наибольшего правдоподобия;
д) наименьшего правдоподобия.

Типовые задачи:

На основе выборочных данных о производительности труда (X), измеряемой в млн. руб. на человека, и себестоимости продукции (У), измеряемой в тыс. руб. на единицу продукции, полученных с однотипных предприятий за месяц:

X	5	4	3	20	10	15
У	7	10	12	2	5	4

Найти: а) параметры линейного уравнения регрессии между X и У;

б) проверить значимость уравнения регрессии в целом и его параметров при α=0,05;

2 Краткое описание проводимого занятия:

1. Проверка уровня освоения студентами вопросов для самостоятельного изучения

2. Тестирование
3. Решение задач

Результаты и выводы:

Усвоение студентами знаний и закрепление навыков по теме практического занятия.