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

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.Б.08 Теория вероятностей и математическая статистика

Направление подготовки (специальность) 09.03.01 Информатика и вычислительная техника

Профиль образовательной программы **“**Автоматизированные системы обработки информации и управления**”**

Форма обучения очная

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция № 1 (2 часа).

Тема: «Классическое определение вероятности события. Геометрические вероятности. Относительная частота наступления события и статистическая вероятность. Формулы умножения и сложения вероятностей случайных событий»

1.1.1 Вопросы лекции:

1.Случайные события, их классификация.

2.Вероятность случайных событий, ее интерпретации.

3. Основные теоремы теории вероятностей.

1.1.2 Краткое содержание вопросов:

1. Случайные события, их классификация.

Практика показывает, что в совокупности массы однородных случайных явлений обнаруживаются определенные закономерности. Так, при увеличении числа выстрелов частота попадания в цель стабилизируется, приближаясь к некоторому постоянному числу. При многократном бросании монеты частота выпадения герба (отношение числа выпавших гербов к общему числу бросаний) приближается к числу 0,5.

Чем больше количество рассматриваемых однородных случайных явлений, тем определеннее и отчетливее проявляются присущие им закономерности.

Вот эти специфические закономерности массовых однородных случайных явлений и являются предметом изучения теории вероятностей.

Следует заметить, что вероятностные методы ни в коем случае не противопоставляют себя классическим методам точных наук, но дополняют их, а это позволяет глубже анализировать случайные явления.

В настоящее время нет практически ни одной естественной науки, в которой, так или иначе, не применялись бы вероятностные методы, ведь математические законы теории вероятностей есть отражение реальных законов, объективно существующих в массовых случайных явлениях природы и техники.

Основные понятия

Под испытанием будем понимать опыт, эксперимент, любое действие, приводящее к возникновению определенной совокупности условий. Событием называется результат всякого испытания. Все события делятся на достоверные, невозможные и случайные.

Достоверное событие — это событие, которое обязательно наступает в данном испытании.

Невозможное — это событие, которое никогда не наступает в данном испытании.

Случайное событие – это событие, которое в данном испытании может наступить или не наступить.

Случайные событие называются несовместными в данном испытании, если никакие два из них в этом испытании не могут наступить одновременно.

Случайные события образую полную группу, если являются всеми возможными результатами данного испытания.

Случайные события называются противоположными в данном испытании, если они несовместны и образуют полную группу.

Рассмотрим полную группу равновозможных, несовместных, случайных событий. Такие события будут называться случаями, шансами или исходами.

События называются равновозможными, если нет оснований считать, что одно является более возможным, чем другое.

Случай рассмотренной группы называется благоприятствующим появлению события А, если появление этого случая влечет за собой появление события А.

Например, в урне 8 шаров с цифрами от 1 до 8. Шары 1,2,3 – красные, остальные – белые. Появление шара с 1 (или 2, или 3) есть событие (случай), благоприятствующий появлению красного шара.

Количественная оценка возможности наступления события А в данном испытании называется вероятностью этого события и обозначается Р(А). Существует несколько определений этого понятия. Рассмотрим вначале определение, называемое классическим, проанализируем его слабые стороны, затем перейдем к другим определениям, позволяющим преодолеть указанные недостатки.

2. Вероятность случайных событий, ее интерпретации

Классическое определение вероятности события. Вероятностью события А называется отношение: $P(A)=\frac{m}{n}$,

где m – число благоприятствующих случаев (исходов), а n – число всех возможных случаев (исходов), образующих полную группу равновозможных, несовместных, случайных событий.

Если какому-либо событию благоприятствует все n случаев, образующих полную группу равновозможных, несовместных, случайных событий, то оно является достоверным ($p=1$). Событие, которому не благоприятствует ни один из n случаев, является невозможным ($p=0$).

Следовательно, $0 \le P(A) \le 1$.

Задача. В корзине 8 красных и 12 белых шаров, наудачу извлекают один шар. Какова вероятность того, что он красный? Какова вероятность того, что он белый?

Испытание: извлечение шара из корзины.

Событие А: появление шара красного цвета.

Событие В: появление шара белого цвета.

События А и В – противоположные в данном испытании.

$$P(A)=\frac{m}{n}=\frac{8}{20}=\frac{2}{5}\ ;\ P(B)=1-P(A)=\frac{3}{5}.$$

Ограниченность классического определения вероятности

Классическая формула вероятности события применяется для непосредственного подсчета вероятностей тогда, когда задача сводится к «схеме случаев». Другими словами, классическое определение предполагает, что число элементарных исходов испытания конечно. На практике же часто встречаются испытания, число возможных исходов которых бесконечно, то есть далеко не всякий опыт может быть сведен к «схеме случаев». Следовательно, существует класс событий, вероятности которых нельзя вычислить по классической формуле. Часто так же невозможно представить результат испытания в виде совокупности элементарных исходов или указать основания, позволяющие считать элементарные события равновозможными.

Указанные недостатки могут быть преодолены введением геометрической и статистической вероятностей.

Геометрические вероятности

Геометрической вероятностью называют вероятность попадания наудачу брошенной точки в область (отрезок, часть плоскости, часть пространства).

Пусть отрезок *l* составляет часть отрезка *L*. На отрезок *L* наудачу поставлена точка. Вероятность попадания точки на отрезок *l* пропорциональна длине этого отрезка и не зависит от его расположения относительно отрезка L: $P = \frac{\text{длина l}}{\text{длина L}}$.

Пусть плоская фигура *g* составлять часть плоской фигуры *G*. На *G* наудачу брошена точка. Вероятность попадания брошенной точки на *g* пропорциональна площади этой фигуры и не зависит ни от ее расположения относительно *G*, ни от формы *g*: $P = \frac{\text{площадьg}}{\text{площадьG}}$.

По аналогии через отношение объемов определяется вероятность попадания наудачу брошенной точки в часть пространства.

Задача. Найти вероятность того, что точка, брошенная наудачу, попадет в кольцо, образованное двумя окружностями с радиусами 5 и 10 см.

Площадь кольца (фигура g): $S_g = \pi\left(10^2 - 5^2\right) = 75\,\pi$

$S_G = \pi \cdot 10^2 = 100\,\pi \qquad P = \frac{75\,\pi}{100\,\pi} = 0{,}75$.

Статистическая вероятность события

Введем еще одну количественную оценку возможности появления события в данном испытании, корнями уходящую в опыт, эксперимент.

Относительной частотой наступления события А называется отношение

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

где *n* – число проведенных опытов (испытаний), а *m* – число испытаний, в которых событие А наступило.

Заметим, что классическая формула не требует проведения испытаний в действительности, $P(A)$ вычисляется до опыта. Для нахождения относительной частоты испытания должны быть проведены, либо возможно их проведение, $W(A)$ вычисляют после опыта.

При небольшом числе опытов W носит случайный характер и может изменяться. Например, при 10 бросаниях монеты герб может появиться 2 раза, а может 8 раз.

Но при увеличении числа опытов частота утрачивает случайный характер, случайные причины, влияющие на результат каждого отдельного опыта, взаимно «гасят» друг друга и W приближается к некоторой средней, постоянной величине.

Если в одинаковых условиях производят серии опытов и в каждой серии число испытаний довольно велико, то W обнаруживает свойство устойчивости. В таком случае W или близкое к ней число принимают за статистическую вероятность события.

Все свойства вероятности, вытекающие из классического определения, распространяются и на статистическое определение вероятности события.

Для существования статистической вероятности события требуется:

1) возможность, хотя бы принципиальная, производить неограниченное число испытаний, в каждом из которых событие А наступает или нет;

2) устойчивость относительных частот в различных сериях из достаточно большого числа испытаний.

Например, по данным шведской статистики приводится относительная частота рождения девочек по месяцам года: 0,486; 0,489; 0,490; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,491; 0,482; 0,473.

Значение относительной частоты колеблется около числа 0,482, его можно принять за статистическую вероятность рождения девочки. Статистические данные других стран дают примерно те же значения W и ту же статистическую вероятность.

Рассмотрим другой пример:

Число бросаний монеты	Число появлений герба	W
4040	2048	0,5069
12000	6019	0,5016
24000	12012	0,5005

Данные таблицы показывают, как с увеличением числа испытаний «уточняется» значение относительной частоты.

Недостатком статистического определения является неоднозначность выбора значения относительной частоты при возникновении свойства устойчивости.

При практическом применении вероятностных методов исследования необходимо понимать, принадлежит ли исследуемое случайное явление к категории массовых, для которых выполняется свойство устойчивости частоты и понятие вероятность имеет глубокий практический смысл.

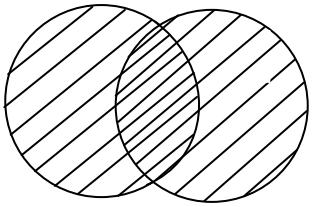
3. Основные теоремы теории вероятностей.

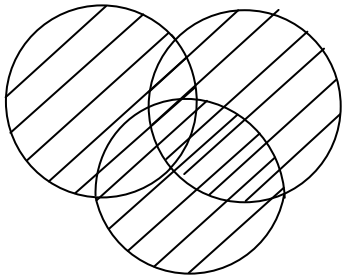
В большинстве практических задач для определения вероятностей событий применяются косвенные методы, позволяющие по известным вероятностям одних событий определять вероятности других. Кроме того, результаты многих испытаний являются сложными, применение классической формулы сразу не представляется возможным, хотя задача и сводится к «схеме случаев». Применение косвенных методов связано с использованием основных теорем теории вероятностей: теоремы сложения вероятностей и теоремы умножения вероятностей.

Но вначале необходимо введение символических операций над событиями.

Суммой двух событий А и В называется новое событие С, состоящее в появлении или события А, или события В, или событий А и В одновременно.

Суммой нескольких событий называется новое событие, состоящее в появлении хотя бы одного из исходных событий.

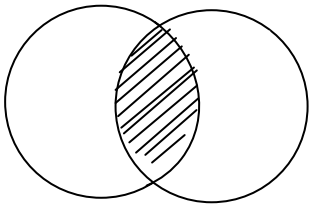


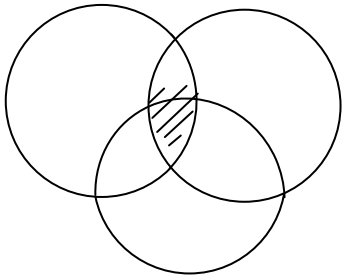


A + B

Произведением (совмещением) двух событий А и В +C овое событие С, состоящее ном появлении события А и события Б одновременно.

Произведением нескольких событий называют новое событие, состоящее в одновременном появлении всех исходных событий.





A · B

Теорема (о сложении вероятностей несовместн

A · B · C
A ∩ B ∩ B

Пусть события A и B несовместны в данном ис ыте), причем вероятности этих событий известны.

Вероятность появления одного из двух несовместных событий, безразлично какого, равна сумме вероятностей этих событий:

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

Формула из теоремы справедлива для любого числа попарно несовместных слагаемых: $P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)$

Задача. Производиться стрельба по области D, состоящей из трех непересекающихся областей (зон). Известны вероятности попадания в каждую зону $P(A_1)=\frac{5}{100}$, $P(A_2)=\frac{10}{100}$, $P(A_3)=\frac{17}{100}$. Найти вероятность попадания в область D.

Событие A – попадание в область D.

$A = A_1 + A_2 + A_3$ (где A_1, A_2, A_3 попарно несовместны)

$$P(A)=P(A_1)+P(A_2)+P(A_3)=\frac{5}{100}+\frac{10}{100}+\frac{17}{100}=\frac{32}{100}.$$

Следствие. Если случайные события $A_1, A_2, \ldots A_n$ образуют полную группу несовместных событий, то справедливо равенство:

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1$$

Случайные события A и B называются совместными, если в данном испытании могут наступить оба этих события, т.е. произойдет совмещение событий A и B.

Событие, заключающееся в совмещении событий A и B, будем обозначать (A и B) или (AB).

Теорема (о сложении вероятностей совместных событий).

Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совмещения: $P(A+B)=P(A)+P(B)-P(A\cdot B)$

Событие A называется независимым от события B, если вероятность появления события A не зависит от того, наступило событие B в данном испытании или нет.

Теорема (об умножении вероятностей независимых событий).

Вероятность совместного появления двух независимых событий равна произведению вероятностей этих событий:

$$P(A\cdot B)=P(A)\cdot P(B).$$

Приведем доказательство теоремы с использованием «схемы урн». Рассмотрим две урны, в каждой из которых n_1 и n_2 шаров соответственно. В 1-ой урне m_1 красных шаров, остальные - черные, во 2-ой урне m_2 красных шаров, остальные – черные. Из каждой урны вынимается по одному шару. Какова вероятность того, что оба вынутых шара красные?

Событие A : вынимание красного шара из 1-ой урны, событие B: вынимание красного шара из 2-ой урны. Эти события независимы.

$$P(A)=\frac{m_1}{n_1};\quad P(B)=\frac{m_2}{n_2}$$

Всех возможных случаев одновременного вынимания по одному шару из каждой урны $n_1 \cdot n_2$. Число случаев, благоприятствующих появлению красных шаров из обеих урн, будет $m_1 \cdot m_2$. Вероятность совмещения событий: $P(A \cdot B) = \frac{m_1 \cdot m_2}{n_1 \cdot n_2} = \frac{m_1}{n_1} \cdot \frac{m_2}{n_2} = P(A) \cdot P(B)$. Что и требовалось доказать.

Замечание. Равенство из теоремы справедливо для любых *n* независимых событий:
$P(A_1 \cdot A_2 \ldots A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$

Замечание. С учетом теоремы об умножении вероятностей теорема о сложении вероятностей совместных событий записывается следующим образом:
$P(A+B) = P(A) + P(B) - P(A \cdot B) = P(A) + P(B) - P(A) \cdot P(B)$,
если события А и В – совместны, но независимы.

Задача. В урне 5 красных, 8 белых и 11 синих шаров. Наудачу извлекают 1 шар. Какова вероятность того, что появится белый или синий шар?

$$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) = \frac{8}{24} + \frac{11}{24} = \frac{19}{24}.$$

Событие А называется зависимым от события В, если вероятность появления события А зависит от того, наступило событие В в данном испытании или нет.

Вероятность события А, найденную при условии, что наступило событие В ($P_B(A)$), будем называть **условной вероятностью** события А при условии В.

Например, в урне 3 белых и 2 черных шара. Наудачу вынимают один шар, затем еще один. Событие В: появление белого шара при первом вынимании; событие А: появление белого шара при втором вынимании. Тогда $Pв(A) = \frac{2}{4} = \frac{1}{2}$.

Теорема (об умножении вероятностей зависимых событий). Вероятность совмещения двух зависимых событий равна произведению вероятности одного из них на условную вероятность второго, вычисленную в предположении, что первое событие наступило:
$P(A \cdot B) = P(B) \cdot P_B(A)$

Приведем доказательство теоремы с использованием «схемы урн».

Всего в урне шаров *n*, где n_1 – белые шары, n_2 – черные шары. Пусть среди n_1 белых шаров n_1 * шаров с отметкой *, остальные – чисто белые. Из урны наудачу вынимается один шар. Какова вероятность того, что это шар белый*?

Событие В: появление белого шара; событие А: появление шара со *. Тогда $P(B) = \frac{n_1}{n}$; $P_B(A) = \frac{n_1 *}{n_1}$ - вероятность появления шара со * при условии, что появился белый шар. Вероятность появления белого шара со * есть $P(A \cdot B)$. Очевидно, что $P(A \cdot B) = \frac{n_1 *}{n}$. Но $\frac{n_1 *}{n} = \frac{n_1}{n} \cdot \frac{n_1 *}{n_1}$, т.е. $P(A \cdot B) = P(B) \cdot P(AB)$

Что и требовалось доказать.

Замечание. Часто формула из последней теоремы служит для определения условной вероятности: $P_B(A) = \frac{P(A \cdot B)}{P(B)}$ ($P(B) \neq 0$)

Замечание. Применим формулу из теоремы об умножении вероятностей зависимых событий к выражению:

$P(B \cdot A) = P(A) \cdot P_A(B)$

$P(A \cdot B) = P(B) \cdot P_B(A)$

Левые части равны. Следовательно, правые также будут равны:

$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A)$.

Задача. В коробке 6 одинаковых занумерованных кубиков. Наудачу по одному извлекают все кубики. Найти вероятность того, что номера извлеченных кубиков появляются в возрастающем порядке.

$$P(A) = \frac{1}{6} \cdot \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{30 \cdot 24} = \frac{1}{720}$$

Задача. Вероятность изготовления годного изделия данным станком равно 0,9. Вероятность появления изделия первого сорта среди годных изделий равна 0,8. Определить вероятность изготовления изделий первого сорта данным станком.

Событие В: изготовление годного изделия; событие А: появление изделия первого сорта. $P(B) = 0{,}9$ $P_B(A) = 0{,}8$ (по условию), тогда

$P(A \text{ и } B) = 0{,}9 \cdot 0{,}8 = 0{,}72$.

1.2 Лекция № 2 (2 часа).

Тема: «Зависимые события. Условная вероятность. Формула полной вероятности события. Вероятности гипотез. Формула Байеса. Повторение испытаний: формулы Бернулли, локальные и интегральные теоремы Лапласа, формула Пуассона, простейший поток событий»

1.2.1 Вопросы лекции:

1. Условная вероятность события. Формула полной вероятности, формула Байеса.

2.Схема повторных испытаний. Формулы Бернулли, Пуассона, Лапласа.

3. Простейший поток событий, его свойства.

1.2.2 Краткое содержание вопросов:

1. Условная вероятность события. Формула полной вероятности, формула Байеса.

Вероятность события А, найденную при условии, что наступило событие В ($P_B(A)$), будем называть **условной вероятностью** события А при условии В.

Например, в урне 3 белых и 2 черных шара. Наудачу вынимают один шар, затем еще один. Событие В: появление белого шара при первом вынимании; событие А: появление белого шара при втором вынимании. Тогда $Pв(A) = \frac{2}{4} = \frac{1}{2}$.

Теорема (формула полной вероятности).

Пусть $B_1, B_2, ..., B_n$ - образуют полную группу несовместных событий, т.е. $\sum_{i=1}^{n} P(B_i) = 1$. Если событие А может осуществляться только при условии совмещении с одним из событий $B_1, B_2, ..., B_n$, то

$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + ... + P(B_n) \cdot P_{B_n}(A)$.

Задача. По цели произведено 3 последовательных выстрела. Вероятность попадания при первом выстреле p_1=0,3; вероятность попадания при втором выстреле p_2=0,6; вероятность попадания при третьем выстреле p_3=0,8. При одном попадании вероятность поражения цели λ_1=0,4; при двух попаданиях – λ_2=0,7; при трех попаданиях – λ_3=1,0. Определить вероятность поражения цели при трех выстрелах?

Решение.

Событие А: поражение цели при трех выстрелах. Рассмотрим полную группу несовместных событий:

B_1: было одно попадание при трех выстрелах;

B_2 : было два попадание при трех выстрелах;

B_3 : было три попадание при трех выстрелах;

B_4 : не было ни одного попадания.

Определим вероятность каждого события:

$$P(B_1) = p_1(1-p_2)(1-p_3) + (1-p_1)p_2(1-p_3) + (1-p_1)(1-p_2)p_3 = 0{,}332$$

$$P(B_2) = p_1 p_2(1-p_3) + p_1(1-p_2)p_3 + (1-p_1)p_2 p_3 = 0{,}468$$

$$P(B_3) = p_1 p_2 p_3 = 0{,}144$$

$$P(B_4) = (1-p_1)(1-p_2)(1-p_3) = 0{,}056.$$

Условные вероятности поражения цели при осуществлении каждого из этих событий:

$$P_{B_1}(A) = 0{,}4; \quad P_{B_2}(A) = 0{,}7; \quad P_{B_3}(A) = 1; \quad P_{B_4}(A) = 0.$$

Подставим все данные в формулу из теоремы:

$$P(A) = P(B_1)\cdot P_{B_1}(A) + P(B_2)\cdot P_{B_2}(A) + P(B_3)\cdot P_{B_3}(A) + P(B_4)\cdot P_{B_4}(A) =$$

$$= 0{,}332\cdot 0{,}4 + 0{,}468\cdot 0{,}7 + 0{,}144\cdot 1 + 0{,}0560 = 0{,}6044.$$

Замечание. Если событие А не зависит от события В, то $P(A) = P_B(A)$. Следовательно, $P(A\cdot B) = P(A)\cdot P(B)$.

Пусть $B_1, B_2, ..., B_n$ - полная группа несовместных событий, $P(B_1), P(B_2), ..., P(B_n)$ - соответствующие вероятности. Событие А может наступить только вместе с каким-либо из событий $B_1, B_2, ..., B_n$, которые мы будем называть гипотезами. Тогда справедлива формула полной вероятности:

$$P(A) = P(B_1)\cdot P_{B_1}(A) + P(B_2)\cdot P_{B_2}(A) + ... + P(B_n)\cdot P_{B_n}(A).$$

Допустим, что событие А уже наступило. Это изменит вероятности гипотез $P(B_1), P(B_2), ..., P(B_n)$. Требуется определить условные вероятности этих гипотез $P_A(B_1), ..., P_A(B_n)$, в предположении, что событие А уже наступило.

Найдем

$$P(A\cdot B_1) = p(B_1)\cdot P_{B_1}(A) = p(A)\cdot P_A(B_1) \Rightarrow P_A(B_1) = \frac{P(A\cdot B_1)}{P(A)} = \frac{p(B_1)\cdot P_{B_1}(A)}{P(A)}$$

Заменим $P(A)$ формулой полной вероятности события:

$$P_A(B_1) = \frac{p(B_1)\cdot P_{B1}(A)}{\sum_{i=1}^{n} P(B_i)\cdot P_{B_i}(A)}$$

Аналогично определяется $P_A(B_2), ..., P_A(B_n)$.

Окончательно получаем формулу Байеса или формулу из теоремы гипотез: $P_A(B_k) = \dfrac{p(B_k)\cdot P_{B_k}(A)}{\sum_{i=1}^{n} P(B_i)\cdot P_{B_i}(A)}$.

Задача. 30% приборов собирает специалист высокой квалификации и 70% - средней квалификации. Надежность работы прибора, собранного специалистом высокой квалификации – 0,9 и надежность работы прибора, собранного специалистом средней квалификации – 0,8. Взятый наудачу прибор оказался надежным. Определить вероятность того, что он собран специалистом высокой квалификации.

Событие А: безотказная работа прибора.

Для проверки прибора возможны гипотезы:

B_1: прибор собран специалистом высокой квалификации;

B_2: прибор собран специалистом средней квалификации.

По условию задачи:

$P_{B_1}(A) = 0{,}9; \quad P_{B_2}(A) = 0{,}8$.

Определим вероятности гипотез B_1 и B_2 при условии, что событие А наступило:

$$P_A(B_1) = \frac{0{,}3\cdot 0{,}9}{0{,}3\cdot 0{,}9 + 0{,}7\cdot 0{,}8} = 0{,}325; \quad P_A(B_2) = \frac{0{,}7\cdot 0{,}8}{0{,}3\cdot 0{,}9 + 0{,}7\cdot 0{,}8} = 0{,}675$$

2. Схема повторных испытаний. Формулы Бернулли, Пуассона, Лапласа.

Рассмотрим методы решения задачи, в которой один и тот же опыт повторяется несколько раз. В результате каждого опыта может появиться или не появиться интересующее нас событие. Причем, нас будет интересовать не результат отдельного опыта, а результат серии опытов, а именно вероятность появления того или иного числа событий в серии независимых опытов (испытаний).

Испытания считаются независимыми, если вероятность появления события $P(A)$ в каждом испытании не зависит от исходов других испытаний.

Пусть проводится *n* независимых испытаний, в каждом из которых событие А может наступить с вероятностью *p* или не наступить с вероятностью *q=1-p*.

Формула Бернулли, Пуассона, теоремы Лапласа

Задача. Вычислить вероятность того, что в *n* испытаниях событие А наступит *k* раз и не наступит *(n-k)* раз, причем последовательность появления события А не важна.

Вероятность этого сложного события по теореме об умножении вероятностей независимых событий определяется как $p^k \cdot q^{n-k}$.

Таких сложных событий может быть столько, сколько можно составить сочетаний C_n^{κ}. Все эти события несовместны, а вероятности их одинаковы, поэтому искомая вероятность определяется по формуле: $P_n(\kappa) = C_n^{\kappa} p^k \cdot q^{n-k}$.

Полученную формулу называют формулой Бернулли.

Задача. Вероятность того, что расход электроэнергии в течение суток не превысит нормы, равна 0,75. Найти вероятность того, что расход электроэнергии не превысит нормы в течение 4 суток из 6.

Испытание: проверка расхода энергии в течение суток, повторяется 6 раз.

А: расход электроэнергии в норме; *p*=0,75; *q=1-p*=0,25.

B: событие A наступило 4 раза в 6 независимых испытаниях.

$P_6(4) = C_6^4 p^4 \cdot q^{6-4} = 15 \cdot (0{,}75)^4 \cdot (0{,}25)^2 = 0{,}30$.

Число k_0 называется наивероятнейшим, если вероятность того, что событие наступит в испытаниях k_0 число раз превышает (или, по крайней мере, не меньше) вероятности остальных возможных исходов испытания.

$np - q \le k_0 \le np + p$,

где n – число испытаний; p – вероятность появления события в одном испытании; q – вероятность не появления события в одном испытании.

Если а) $np - q$ - дробное число, то k_0 – единственное;

б) $n - q$ - целое, то наивероятнейших чисел два k_0 и k_0+1;

в) np - целое, то $k_0 = np$.

Если число независимых испытаний n достаточно велико, то вычисления по формуле Бернулли будут слишком громоздки. В таком случае формулу, хотя и асимптотическую, дает локальная теорема Лапласа.

Заметим, что для частного случая формула была найдена в 1730 году Муавром, а в 1783 году обобщена Лапласом. Поэтому теорему, о которой идет речь, иногда называют теоремой Муавра-Лапласа.

Если производится большое число независимых испытаний, в каждом из которых вероятность наступления события A постоянна и равна $p(p \ne 0,\ p \ne 1)$, то вероятность $P_n(k)$ считается приближенно по формуле:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ - функция Гаусса (табулирована, четная); $x = \frac{k - np}{\sqrt{npq}}$.

Чем больше n, тем точнее будет результат, полученный по формуле из локальной теоремы Лапласа.

Если число проведенных испытаний n очень велико, а вероятность p наступления события A в каждом из n независимых испытаний очень мала, то $P_n(k)$ вычисляется по формуле Пуассона: $P_n(k) \approx \frac{a^{-\lambda} \cdot \lambda^k}{k!}$.

Формула применяется, если параметр $\lambda = n \cdot p < 10$.

Во многих задачах требуется определить вероятность $P_n(k_1 \le k \le k_2)$ того, что событие A наступит не менее k_1 и не более k_2 раз в n независимых испытаниях. Это позволяет сделать интегральная теорема Лапласа.

Если вероятность наступления события A в каждом из n независимых испытаний постоянна и равна $p, (p \ne 0,\ p \ne 1)$, то $P_n(k_1 \le k \le k_2)$ вычисляется по приближенной формуле:

$$P_n(k_1 \le k \le k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right),$$

где $\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{Z^2}{2}}dZ$ - функция Лапласа (табулирована, нечетная, для $x>5 \quad \Phi(x)=0{,}5$).

Задача. Вероятность того, что деталь не прошла проверку ОТК 0,2. Найти вероятность того, что из 400 случайно выбранных деталей непроверенными окажутся от 70 до 100.

Испытание: выбор одной детали, повторяется 400 раз.

A: деталь проверку не прошла; p=0,2; $q=1-p$=0,8.

B: событие A наступило от 70 до 100 раз в 400 независимых испытаниях.

$$P_{400}(70\le k\le 100)\approx \Phi\left(\frac{100-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}}\right)-\Phi\left(\frac{70-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}}\right)\approx 0{,}4938+0{,}3944\approx 0{,}8882$$

3. Простейший поток событий, его свойства

Особое внимание следует обратить на простейший поток событий.

Потоком событий называют последовательность событий, которые наступают в случайные моменты времени. Примеры потоков: поступление вызовов на АТС, поступление вызовов на пункт неотложной медицинской помощи, прибытие кораблей в порт, последовательность отказов элементов некоторого устройства.

Простейшим называют поток, обладающий свойствами стационарности, отсутствием последействия и ординарности.

Свойство стационарности характеризуется тем, что вероятность появления k событий за время длительностью t не зависит от начала отсчета промежутка времени, а зависит лишь от его длительности. Так вероятности появления пяти событий на промежутках времени (1; 4), (6; 9), (8; 11) одинаковой длительности t = 3 единицы времени равны между собой.

Свойство отсутствия последействия характеризуется тем, что вероятность появления k событий на любом промежутке времени не зависит от того, сколько событий появилось до начала рассматриваемого промежутка.

Свойство ординарности характеризуется тем, что вероятность появления двух и более событий пренебрежимо мала, сравнительно с вероятностью появления одного события.

Интенсивностью потока λ называют среднее число событий, которые появляются в единицу времени. Доказано, что если известна постоянная интенсивность потока λ, то вероятность появления k событий простейшего потока за время длительности t определяется формулой:

$$P_t(k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Пример: Среднее число заявок, поступающих на предприятие бытового обслуживания за 1 час, равно трем. Найти вероятность того, что за 2 часа поступит 5 заявок. Предполагается, что поток заявок - простейший.

Решение. По условию λ = 3, t =2, k = 5. Воспользуемся формулой

$$P_t(k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Искомая вероятность того, что за 2 часа поступит 5 заявок, равна

$$P_2(5)=\frac{(6)^5\cdot 0{,}00248}{120}\approx 0{,}268.$$

Пример: Среднее число заявок, поступающих на АТС в одну минуту, равно двум. Найти вероятности того, что за четыре минуты поступит:

а) три вызова;

б) менее трех вызовов;

в) не менее трех вызовов.

Решение, а) По условию $\lambda = 3$, $t = 2$, $k=5$. Воспользуемся формулой: $P_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Подставив данные условия задачи, получим: $P_4(3) = \frac{8^3 \cdot e^{-8}}{3!} = \frac{512 \cdot 0{,}000335}{6} \approx 0{,}03$.

б) Найдем вероятность того, что за четыре минуты поступит менее трех вызовов, т.е. ни одного вызова, или один вызов, или два вызова. Поскольку эти события несовместны, применим теорему суммы несовместных событий:

$$P_4(k<3) = P_4(0) + P_4(1) + P_4(2) = e^{-8} + 8 \cdot e^{-8} \cdot \frac{8^2 \cdot e^{-8}}{2!} = 41 \cdot 0{,}000335 \approx 0{,}01.$$

в) Найдем вероятность того, что за четыре минуты поступит не менее трех вызовов: так как события «поступило менее трех вызовов» и «поступило не менее трех вызовов» - противоположные, то сумма вероятностей этих событий равна единице: $P_4(k<3) + P_4(k\geq 3) = 1$. Поэтому $P_4(k \geq 3) = 1 - P_4(k<3) = 1 - [P_4(0) + P_4(1) + P_4(2)] = 1 - 0{,}01 = 0{,}99$.

1.3 Лекция № 3 (2 часа).

Тема: «Понятие случайной величины примеры. Виды случайных величин. Закон распределения вероятностей. Функция распределения случайных величин. Свойства. Плотность распределения вероятностей. Числовые характеристики: математическое ожидание, свойства; дисперсия, свойства; среднее квадратичное отклонение и его свойства»

1.3.1 Вопросы лекции:

1. Случайные величины, их классификация, закон распределения.

2. Функция распределения, плотность распределения, вероятность попадания в интервал.

3. Числовые характеристики ДСВ.

4. Числовые характеристики НСВ.

1.3.2 Краткое содержание вопросов:

1.. Случайные величины, их классификация, закон распределения.

Рассмотрим событие: появление определения числа очков на грани игральной кости, выпавшей при бросании. При этом может появляться любое из чисел 1,2,3, … 6. Наперед определить число выпавших очков невозможно, поскольку оно зависит от многих случайных причин, которые полностью не могут быть учтены. В этом смысле число очков есть случайная величина, а числа 1,2, … 6 - возможные значения этой величины.

Случайной называют величину, которая в результате испытания принимает одно из всех своих возможных значений, наперед не известное и зависящее от случайных причин, которые заранее не могли быть учтены.

Обозначение: X, Y, Z, … - случайные величины; x, y, z, … - значения случайных величин.

Случайные величины делятся на дискретные (ДСВ) и непрерывные (НСВ).

Значения ДСВ отделены промежутками и могут быть перечислены до проведения испытания. Например, число студентов группы, успешно сдавших экзамен по математике.

Значения НСВ затруднительно перечислить до испытания и отделить друг от друга, проще указать интервал, которому эти значения принадлежат.

Например, скорость ветра в течение суток в данной местности или отклонение размера детали от стандарта.

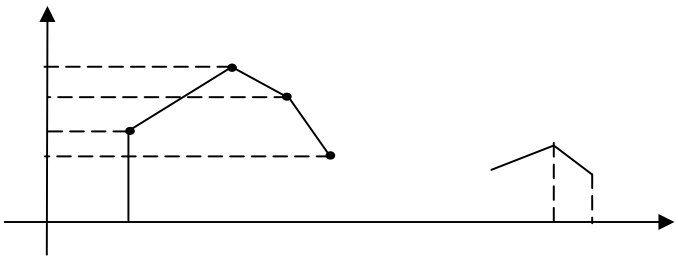
Способы задания ДСВ

Переменная величина X, принимающая в результате испытания одно из конечной или бесконечной последовательности значений $x_1, x_2, \ldots x_k$, называется дискретной, если каждому значению x_k соответствует определенная вероятность p_k того, что переменная величина X примет именно это значение.

Функциональная зависимость вероятности p_k от значения x_k называется законом распределения вероятностей ДСВ *X* (или кратко «закон распределения случайной величины»).

Возможные значения случайной величины	1	2	3		k	
Вероятности этих значений	1	2	3		k	

Закон распределения можно задать графически:



Закон можно задать ан г: $p_k = f(x_k)$.

То, что величина *X* примет одно из значений последовательности $x_1, x_2, \ldots, x_k, \ldots$ есть событие достоверное.

Иначе: $X = x_1, X = x_2, \ldots, X = x_k, \ldots$ - эти события несовместны и образуют полную группу. Следовательно, $\sum_{i=1}^{\kappa} p_i = 1$ (если последовательность конечная) или $\sum_{i=1}^{\infty} p_i = 1$ (если последовательность бесконечная).

Например, пусть ДСВ *X* : число очков, выпадающее на верхней грани игральной кости при ее однократном бросании. Составить закон распределения *X*.

Значение случайной величины x_i , которому соответствует наибольшая вероятность, называется модой случайной величины.

Задача. Вероятность попадания при каждом выстреле p=0,8. Имеется 3 снаряда, стрельба ведется до первого попадания. Составить таблицу распределения числа израсходованных снарядов.

ДСВ X: число израсходованных снарядов.

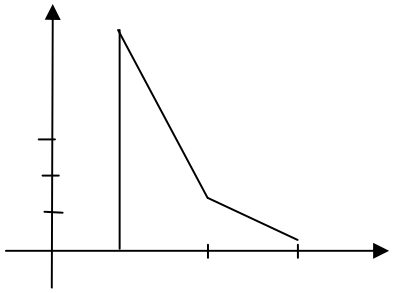
$P(X = x_1)$ - вероятность того, что X примет значение x_1, т.е. вероятность того, что будет израсходован один снаряд;

$P(X = x_2)$ - вероятность того, что будет израсходовано два снаряда;

$P(X = x_3)$ - вероятность того, что будет израсходовано три снаряд (два раза не попали и третий раз – попали; три раза не попали).

$P(X = x_3) = 0{,}2 \cdot 0{,}2 \cdot 0{,}8 + 0{,}2 \cdot 0{,}2 \cdot 0{,}2 = 0{,}2^2(0{,}8 + 0{,}2) = 0{,}2^2$

		2	3
	,8	0,16	0,04



Контроль: 0,8+0,16+0,04=1

x_1=1 – мода случайной величины X.

Пусть производится n независимых испытаний, в каждом из которых событие А может появляться, может не появляться. Вероятность наступления события в каждом испытании постоянна и равна p (q=*1-p* – вероятность не наступления).

Рассмотрим ДСВ X: число появлений события А в этих испытаниях.

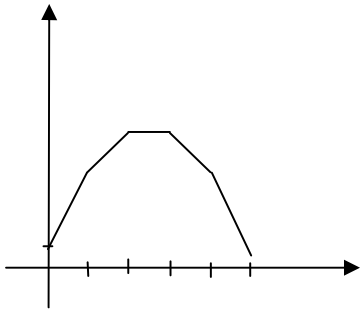
Найдем закон распределения. Т.к. событие А в n испытаниях может не появляться ни разу, 1 раз, 2, … n раз. Следовательно, значения X: 0,1,2, …, n. Для нахождения вероятностей этих значений нужно воспользоваться формулой Бернулли.

Таким образом, формула Бернулли и является аналитическим выражением искомого закона распределения.

Такое распределение, определяемое формулой Бернулли, называется биномиальным, т.к. правую часть формулы Бернулли можем считать общим членом разложения бинома Ньютона.

Изобразить графически биномиальный закон распределения вероятностей случайной величины X при n=5, $p=\frac{1}{2}$, $q=\frac{1}{2}$, где X – число появлений события А в n независимых испытаниях.

$$p_5(0) = \frac{5!}{0!5!} \cdot p^0 \cdot q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$



$$p_5(1)=\frac{5!}{1!4!}\cdot\left(\frac{1}{2}\right)^1\cdot\left(\frac{1}{2}\right)^4=\frac{5}{32}$$

$$p_5(2)=\frac{5!}{2!3!}\cdot\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right)^3=\frac{10}{32}$$

$$p_5(3)=\frac{5!}{3!2!}\cdot\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)^2=\frac{10}{32}$$

$$p_5(4)=\frac{5!}{4!1!}\cdot\left(\frac{1}{2}\right)^4\cdot\frac{1}{2}=\frac{5}{32}\qquad p_5(5)=\frac{5!}{5!0!}\cdot\left(\frac{1}{2}\right)^5\cdot\left(\frac{1}{2}\right)^0=\frac{1}{32}$$

Если число независимых испытаний велико, а вероятность наступления события в каждом испытании очень мала, ($n\cdot m<10$), то вероятность того, что событие А появится *k* раз в *n* испытаниях находится по закону Пуассона.

Такое распределение случайной величины *X* называют распределением Пуассона.

Задача. Завод отправил на базу 5000 изделий. Вероятность того, что в пути изделие повредится равна 0,0002. Составить закон распределения числа испорченных изделий.

ДСВ *X*: число поврежденных изделий среди отправленных.

						000
	,37	,37	,19	,06		

n=5000, *p*=0,0002, *np*=1<10

$$p_{5000}(0)=\frac{\lambda^k\cdot e^{-\lambda}}{k!}=\frac{1^0\cdot e^{-1}}{0!}=\frac{1}{e}\approx 0{,}37\quad p_{5000}(1)=\frac{1^1\cdot e^{-1}}{1!}=\frac{1}{e}\approx 0{,}37$$

$$p_{5000}(2)=\frac{1^2\cdot e^{-1}}{2!}=\frac{1}{2\cdot e}\approx 0{,}19\qquad p_{5000}(3)=\frac{1^3\cdot e^{-1}}{3!}=\frac{1}{6\cdot e}\approx 0{,}06 \text{ и т.д.}$$

Непрерывная случайная величина

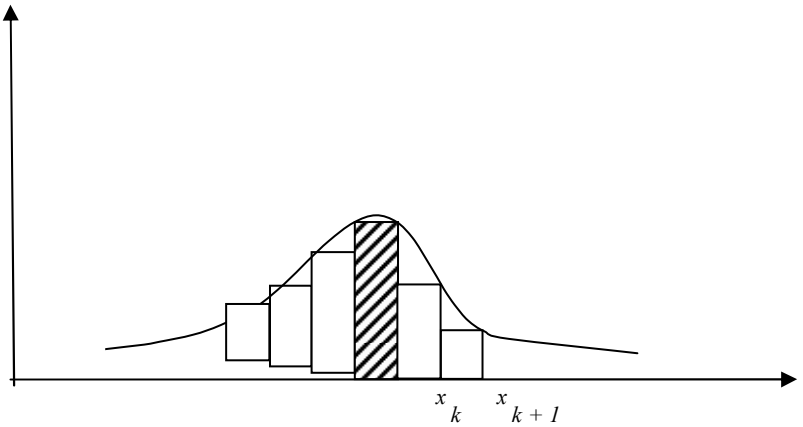
Дадим вначале не совсем точное, но более понятное определение НСВ. Непрерывной называют случайную величину, которая может принимать все (любые) значения из некоторого конечного или бесконечного промежутка.

Рассмотрим НСВ $\overline{x}$, заданную на некотором интервале (a,b)(интервал может быть и бесконечным $(-\infty,+\infty)$). Разделим интервал произвольными точками $x_0, x_1, \ldots, x_n$ на малые интервалы $\Delta x_k = x_{k+1}-x_k$.

Допустим, нам известна вероятность того, что $\overline{x}$ попала на $(x_k; x_{k+1})$. Обозначим эту вероятность $P(x_k<\overline{x}<x_{k+1})$.

Для каждого малого промежутка определим *p* попадания $\overline{x}$ в этот промежуток и изобразим геометрически, т.е. построим соответствующий многоугольник.

Таким образом, получаем ступенчатую ломанную. Проведем плавную кривую, описывающую многоугольники.



Если существует такая функция $y = f(x)$, что $\lim\limits_{\Delta x \to 0} \frac{P(x < \bar{x} < x + \Delta x)}{\Delta x} = f(x)$, то эта функция $f(x)$ называется плотностью распределения вероятностей случайной величины $\bar{x}$ или законом распределения (или плотностью распределения или плотностью вероятности).

Кривая $y = f(x)$ называется кривой распределения вероятностей или кривой распределения.

Механический смысл функции $f(x)$: функция характеризует плотность распределения масс вдоль оси *ox*, т.е. линейную плотность.

2. Функция распределения, плотность распределения, вероятность попадания в интервал

Пусть x – произвольное действительное число. Рассмотрим событие, состоящее в том, что СВ X примет значение, меньшее x.

Вероятность этого события $P(X < x)$ обозначим через $F(x)$.

Функцией распределения называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x, т.е. $F(x) = P(X < x)$.

Геометрически определение означает: $F(x)$ есть вероятность того, что СВ X примет значение, которое изображается на числовой оси точкой, лежащей левее точки x.

Свойства $F(x)$:

1. $0 \le F(x) \le 1$ (из определения).

2. $F(x)$ - неубывающая функция, т.е. если $x_2 > x_1 \Rightarrow F(x_2) \ge F(x_1)$.

Доказательство: Пусть $x_2 > x_1$. Рассмотрим событие: $X < x_2$, оно состоит из двух несовместных событий:

$\tilde{O} < x_1$; $x_1 \le \tilde{O} < x_2 \Rightarrow P(\tilde{O} < x_2) = P(\tilde{O} < x_1) + P(x_1 \le \tilde{O} < x_2)$

$P(\tilde{O} < x_2) - P(\tilde{O} < x_1) = P(x_1 \le \tilde{O} < x_2)$

$F(x_2) - F(x_1) = P(x_1 \le \tilde{O} < x_2) \ge 0$

Следовательно, $F(x_2) - F(x_1) \ge 0 \Rightarrow F(x_2) \ge F(x_1)$

Что и требовалось доказать.

3. Если возможные значения случайной величины принадлежат интервалу (a,b), то, следовательно, $F(x) = 0$ при $x \le a$ и $F(x) = 1$ при $x > b$.

Доказательство: $x_1 \le a \Rightarrow X < x_1$ - невозможное событие. Следовательно,

$P(x < x_1) = 0$.

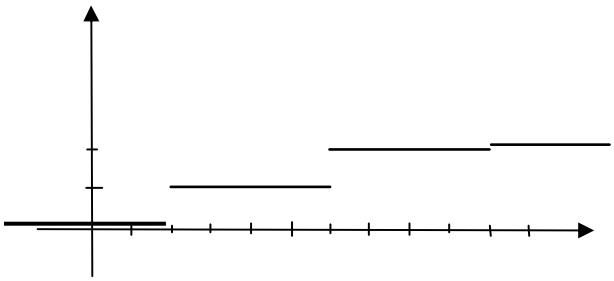
Если $x_2 > b$, то событие $X < x_2$ - достоверное. Следовательно, $P(x < x_2) = 1$.Перейдем к особенностям функции распределения дискретной и непрерывной случайных величин.

Для ДСВ график $F(x)$ имеет разрывный, ступенчатый вид. График расположен в полосе, ограниченной прямыми y=0, y=1.

Задача. Для ДСВ найти $F(x)$ и построить график.

			0
	,5	,4	,1

$$F(x) = \begin{cases} 0 & при \quad x \le 2 \\ 0{,}5 & при \quad 2 < x \le 6 \\ 0{,}9 & при \quad 6 < x \le 10 \\ 1 & при \quad x > 10 \end{cases}$$



Плотность распределения вероятностей являетс з форм закона распределения, но не единственной и не универсальной (только для НСВ).

Свойства плотности:

1. Если все значения случайной величины $\bar{x}$ находятся на (a,b), то

$\int_a^b f(x)dx = 1$ (т.к. достоверно, что значения случайной величины попадут в интервал (a,b)).

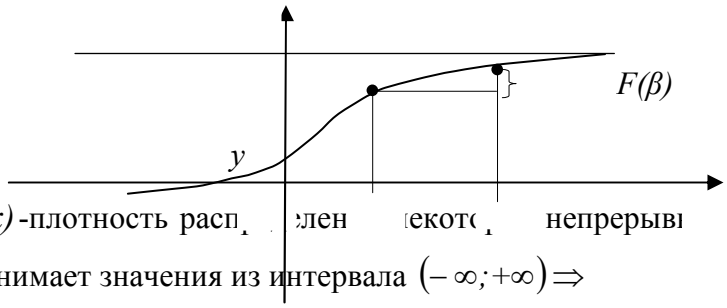
2. $f(x) \ge 0, \forall x \in (a,b)$,

3. Размерность $f(x)$ обратна размерности $\bar{x}$

(что следует из определения).

Вывод: Плотность распределения непрерывной случайной величины полностью задает и определяет непрерывную случайную величину.

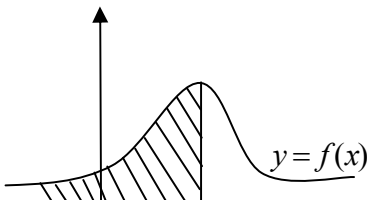
Построим общий вид интегральной кривой, используя свойства $F(x)$:



Пусть $f(x)$ -плотность расп лен екот непрерыв учайной величины $\bar{x}$, которая принимает значения из интервала $(-\infty;+\infty) \Rightarrow$

$$F(x) = P(\bar{x} < x) = P(-\infty < \bar{x} < x) = \int_{-\infty}^{x} f(x)dx$$

Таким образом,



$F(x)=\int_{-\infty}^{x} f(x)dx$ - функция распределения НСВ или интегральная функция. График $F(x)$ называется интегральной кривой распределения.

Теорема. Вероятность попадания случайной величины $\bar{x}$ в заданный интервал $(\alpha;\beta)$ равна приращению функции распределения на этом интервале:

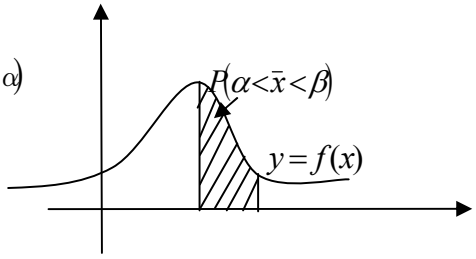
$$P(\alpha<x<\beta)=F(\beta)-F(\alpha).$$

Доказательство:

$$P(\alpha<\bar{x}<\beta)=\int_{-\infty}^{\beta} f(x)dx-\int_{-\infty}^{\alpha} f(x)dx=F(\beta)-F(\alpha)$$

.

Что и требовалось доказать.



3. Числовые характеристики ДСВ

Случайная величина полностью определяется законом распределения.

Однако во многих вопросах практики нет необходимости характеризовать СВ полностью, исчерпывающим образом. Достаточно указать отдельные числовые параметры, характеризующие основные черты распределения. Такие параметры называются числовыми характеристиками случайной величины. Числовые характеристики задают случайную величину косвенно, описывают случайную величину суммарно. В теории вероятностей применяется большое количество числовых характеристик, имеющих различное назначение. Из них рассмотрим только некоторые, наиболее часто встречающиеся характеристики: математическое ожидание, дисперсию и среднее квадратическое отклонение.

Имеется ДСВ X с соответствующим законом распределения:

x				
$p(X$				

Математическим ожиданием ДСВ X (M[x] или m_x) называют сумму произведений всех возможных значений этой величины на вероятности этих значений:

$M[x]=x_1p_1+...+x_np_n=\sum_{k=1}^{n}x_kp_k$, при этом $\sum_{k=1}^{n}p_k=1$.

Если значения случайной величины образуют бесконечную последовательность, то $m_x=\sum_{k=1}^{\infty}x_k\cdot p_k$. Мы будем рассматривать только такие случайные величины, для которых этот ряд сходится.

Замечание. Математическое ожидание случайной величины есть неслучайная (постоянная) величина.

Например, ДСВ задана законом распределения:

	,1	,3	,6

$M[X] = 0{,}3 + 0{,}6 + 2{,}4 = 0{,}9 + 2{,}4 = 3{,}3$.

Задача. Производится один выстрел по объекту. Вероятность попадания *p.* ДСВ *X* – число попаданий. Найти математическое ожидание величины.

Составим закон распределения:

Контроль: *1-p+p=1.* $M[X] = 0 \cdot (1-p) + 1 \cdot p = p$.

Таким образом, математическое ожидание числа появлений события в одном испытании равно вероятности этого события.

Вероятностный смысл M[*X*]: математическое ожидание приближенно равно (чем больше число испытаний, тем точнее) среднему арифметическому наблюдаемых значений случайной величины. На числовой оси возможные значения случайной величины расположены слева и справа от M[*X*]. Поэтому M[*X*] называют центром распределения вероятностей случайной величины (точнее – абсциссой центра).

Свойства математического ожидания:

1. Математическое ожидание постоянной равно самой постоянной $M[c] = c$, где с – ДСВ, которая имеет одно возможное значение с и принимает его с *p*=1. Следовательно, $M[c] = c \cdot 1 = c$.

2. Постоянный множитель можно вынести за знак математического ожидания: $M[cX] = c \cdot M[X]$.

3. $M[X \cdot У] = M[X] \cdot M[У]$, $M[X + У] = M[X] + M[У]$

где величины *X* и *У* – независимы.

Случайные величины *X* и *У* независимы, если закон распределения одной величины не зависит от того, какие возможные значения приняла другая величина.

Последнее свойство распространяется на несколько случайных величин.

Например, независимые случайные величины X и У заданы законами распределения:

p	,7	,2	,1

	,6	,4

$M[X \cdot У]$ - ?

$M[X \cdot У] = M[X] \cdot M[У] = (2{,}1 + 0{,}2 + 0{,}6) \cdot (0{,}6 + 1{,}2) = 2{,}9 \cdot 1{,}8 = 5{,}22$.

Пусть дана случайная величина с соответствующим законом распределения. Обозначим ее математическое ожидание m_x. Рассмотрим разность $X - m_x$, такую случайную величину будем называть центрированной случайной величиной или отклонением.

$$M(X - m_x) = \sum_{k=1}^{n} (x_k - m_k) \cdot p_k = \sum_{k=1}^{n} x_k \cdot p_k - \sum_{k=1}^{n} m_x \cdot p_k = m_x - m_x \sum_{k=1}^{n} p_k = m_x - m_x \cdot 1 =$$

$$= m_x - m_x = 0$$

Таким образом, математическое ожидание центрированной случайной величины равно нулю. Числовой характеристикой рассеивания, разброса значений случайной величины относительно ее математического ожидания является дисперсия. Покажем целесообразность введения дисперсии:

X:

	0,01	,01
p	,5	,5

У:

	100	00
	,5	,5

$$M[X] = -0{,}01 \cdot 0{,}5 + 0{,}01 \cdot 0{,}5 = 0$$

$$M[Y] = -100 \cdot 0{,}5 + 100 \cdot 0{,}5 = 0$$

На рассмотренном примере понятно, что математическое ожидание не полностью характеризует случайную величину. На практике часто требуется оценить рассеяние возможных значений случайной величины вокруг ее среднего значения. Например, насколько кучно лягут снаряды вблизи цели, которая должна быть поражена.

Дисперсией случайное величины *X* называется математическое ожидание квадрата соответствующей центрированной случайной величины:

$$D[X] = M\left[(X - m_x)^2\right] \quad \textit{или} \quad D[X] = \sum_{k=1}^{n} (X_k - m_x)^2 \cdot p_k$$

Средним квадратическим отклонением случайной величины *X* называется характеристика $\sigma_x = \sigma[X] = \sqrt{D[X]}$ или $\sigma[X] = \sqrt{\sum_{k=1}^{n} (x_k - m_x)^2 \cdot p_k}$.

Для вычисления $D[X]$ удобно использовать формулу:

$$D[X] = \sum_{k=1}^{n} (X_k - m_x)^2 \cdot p_k = \sum_{k=1}^{n} x_k^2 \cdot p_k - 2\sum_{k=1}^{n} x_k \cdot m_x \cdot p_k + \sum_{k=1}^{n} m_x^2 \cdot p_k = \sum_{k=1}^{n} x_k^2 \cdot p_k -$$

$$- 2\mathrm{m}_x \sum_{k=1}^{n} x_k \cdot p_k + m_x^2 \sum_{k=1}^{n} p_k = M\left[X^2\right] - 2\mathrm{m}_x \cdot m_x + m_x^2 \cdot 1 = M\left[X^2\right] - m_x^2$$

Следовательно, $D[X] = M\left[X^2\right] - m_x^2$, т.е. дисперсия равна разности математического ожидания квадрата случайной величины и квадрата математического ожидания этой случайной величины.

Задача. Найти $D[X]$ двумя способами:

x	2	3	4
p	0 ,3	0 ,4	0 ,3

Первый способ (по определению):

$$M[X] = 0{,}6 + 1{,}2 + 1{,}2 = 3$$

$$D[X] = (2-3)^2 \cdot 0{,}3 + (3-3)^2 \cdot 0{,}4 + (4-3)^2 \cdot 0{,}3 = 0{,}3 + 0{,}3 = 0{,}6$$

Второй способ (по формуле):

$M[X] = 3$

x^2	4	9	16
p	0,3	0,4	0,3

$$M[X^2] = 1{,}2 + 3{,}6 + 4{,}8 = 9{,}6$$

$$D[X] = M[x^2] - (M[x])^2 = 9{,}6 - 9 = 0{,}6$$

Свойства дисперсии:

1. $D[C] = M[(c-c)^2] = M[0] = 0$, $M[c] = c$
2. $D[CX] = C^2 \cdot D[X]$

$$D[CX] = M[(CX - M[CX])^2] = M[(CX - C \cdot M[X])^2] = M[(C(X - M[X]))^2] =$$
$$C^2 \cdot M[(X - M[X])^2] = C^2 \cdot D[X]$$

3. $D[X+Y] = D[X] + D[Y]$

$$D[X+Y] = M[(X+Y)^2] - (M[X+Y])^2 = M[X^2 + 2XY + Y^2] - (M[X] + M[Y])^2 = M[X^2] +$$
$$+ 2M[X] \cdot M[Y] + M[Y^2] - (M[Y])^2 - (M[Y])^2 - 2M[X] \cdot M[Y] = (M[X^2] - (M[X])^2) +$$
$$+ (M[Y^2] - (M[Y])^2) = D[X] + D[Y]$$

Это свойство распространяется на несколько случайных величин, взаимно независимых.

4. $D[C+X] = D[C] + D[X] = 0 + D[X] = D[X]$
5. $D[X-Y] = D[X] + D[Y]$

$$D[X-Y] = D[X + (-Y)] = D[X] + D[-Y] = D[X] + (-1)^2 D[Y] = D[X] + D[Y]$$

4. Числовые характеристики НСВ.

Теорема. Пусть $f(x)$ - плотность распределения случайной величины $\bar{x}$. Тогда вероятность того, что значение случайной величины $\bar{x}$ попадет в некоторый интервал (α, β), вычисляется следующим образом:

$$P(\alpha < \bar{x} < \beta) = \int_{\alpha}^{\beta} f(x)dx$$

Следовательно, зная плотность распределения случайной величины, мы можем определить вероятность того, что значение случайной величины попало в данный интервал. Геометрически эта вероятность равняется площади соответствующей криволинейной трапеции.

Замечание. В случае непрерывной случайной величины $P(\bar{x} = x_0) = 0$. Действительно, положим $x = x_0$.

Имеем: $P(x_0 < \bar{x} < x_0 + \Delta x) \approx f(x_0) \cdot \Delta x = 0 \Rightarrow \lim\limits_{\Delta x \to 0} P(x_0 < \bar{x} < x_0 + \Delta x) = 0$

Следовательно, $P(\bar{x} = x_0) = 0$. Таким образом, $P(\alpha \le \bar{x} \le \beta) = P(\alpha < \bar{x} < \beta)$, т.к. $P(\alpha \le \bar{x} \le \beta) = P(\bar{x} = \alpha) + P(\alpha < \bar{x} < \beta) + P(\bar{x} = \beta) = 0 + P(\alpha < \bar{x} < \beta) + 0 = P(\alpha < \bar{x} < \beta)$.

По определению функции распределения: $F(x)=\int\limits_{-\infty}^{x} f(x)dx \Rightarrow F'(x)=f(x)$, т.е. производная от функции распределения равна плотности распределения вероятностей. Это равенство выражает связь между $F(x)$ и $f(x)$ непрерывной случайной величины.

Задача. Задана плотность распределения НСВ: $f(x)=\begin{cases}0, & при \quad x\le 0\\ 2x, & при \quad 0<x\le 1\\ 0, & при \quad x>1\end{cases}$

Найти вероятность того, что в результате испытания $\bar{x}$ примет значение, принадлежащее интервалу $(0{,}5;1)$. $P(0{,}5<\bar{x}<1)=2\int\limits_{0{,}5}^{1} xdx = x^2\,|_{0{,}5}^{1}=1-0{,}25=0{,}75$.

того, что в результате испытания $\bar{x}$ примет значение из (0;2).

$P(0<\bar{x}<2)=F(2)-F(0)=\left(\frac{2}{4}+\frac{1}{4}\right)-\frac{1}{4}=\frac{2}{4}=\frac{1}{2}$.

Задача. Случайная величина задана функцией распределения:

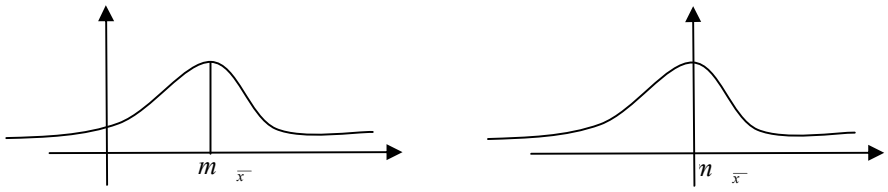
$F(x)=\begin{cases}0, & при \quad x\le 0\\ \frac{1-\cos x}{2}, & при \quad 0<x\le \pi\\ 1, & при \quad x>\pi\end{cases}$ Перейти к другому способу задания.

$$F'(x)=f(x)\Rightarrow \quad f(x)=\begin{cases}0, & при \quad x=0\\ \frac{\sin x}{2}, & при \quad 0<x\le \pi\\ 0, & при \quad x>\pi\end{cases}.$$

Рассмотрим НСВ $\bar{x}$, заданную плотностью распределения $f(x)$. Числовые характеристики НСВ те же, что и для ДСВ: математическое ожидание, дисперсия, среднее квадратическое отклонение.

Математическим ожиданием величины $\bar{x}$ с плотностью распределения $f(x)$ называют $M[\bar{x}]=\int\limits_{-\infty}^{+\infty} x\cdot f(x)dx$ (если $\bar{x}$ принимает значения на $(-\infty;+\infty)$)

Или $M[\bar{x}]=\int\limits_{a}^{b} x\cdot f(x)dx$ (если все возможные значения величины $\bar{x}$ принадлежат промежутку $[a,b]$). $M[\bar{x}]$ является центром распределения вероятностей непрерывной случайной величины $\bar{x}$.



Если кривая распределения симметрична относительно оси *OY,* следовательно, $f(x)$ - четная функция. Значит, $M[\bar{x}] = \int_{-\infty}^{+\infty} x \cdot f(x)dx = 0$. Т.е. в этом случае центр распределения вероятностей совпадает с началом координат.

Дисперсией НСВ $\bar{x}$ называют математическое ожидание квадрата соответствующей центрированной случайной величины.

$D[\bar{x}] = \int_{-\infty}^{+\infty} (\bar{x} - m_{\bar{x}})^2 \cdot f(x)dx$ (аналогично для $[a,b]$).

Средним квадратическим отклонением НСВ $\bar{x}$ называют характеристику: $\sigma[\bar{x}] = \sqrt{D[\bar{x}]} = \sqrt{\int_{-\infty}^{+\infty} (\bar{x} - m_{\bar{x}})^2 \cdot f(x)dx}$.

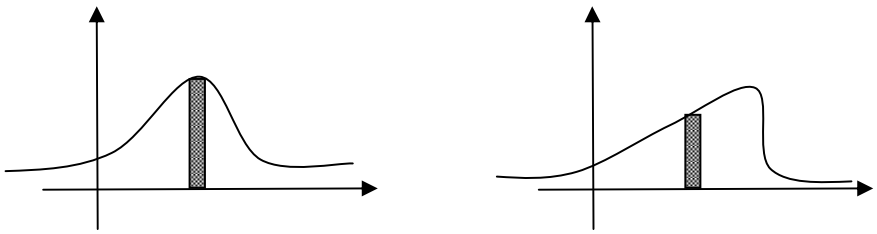
$D[\bar{x}], \sigma[\bar{x}]$ НСВ (как и для ДСВ) характеризуют разброс, рассеяние значений случайной величины относительно $M[\bar{x}]$.

Все свойства $D[x], M[x]$, рассмотренные для ДСВ, справедливы и для НСВ. Для вычисления дисперсии НСВ легко получается следующая формула: $D[\bar{x}] = \int_{-\infty}^{+\infty} x^2 f(x)dx - (M[\bar{x}])^2$ или $D[\bar{x}] = \int_{a}^{b} x^2 f(x)dx - (M[\bar{x}])^2$.

Значение случайной величины, при котором плотность распределения принимает наибольшее значение, называется модой НСВ (M_0). Для НСВ $\bar{x}$, график которой изображен на предыдущем рисунке, мода совпадает с математическим ожиданием.

Число Me называется медианой НСВ, если оно удовлетворяет равенству: $\int_{-\infty}^{Me} f(x)dx = \int_{Me}^{+\infty} f(x)dx = \frac{1}{2}$ или $P(\bar{x} < Me) = P(\bar{x} > Me) = \frac{1}{2}$.

Другими словами, равновероятно, что случайная величина $\bar{x}$ примет значение меньше *Me* или больше *Me*, хотя сама случайная величина $\bar{x}$ может значение *Me* и не принимать.



Геометрически: *Me* – это точка, в которой ордината $f(x)$ делит пополам площадь, ограниченную кривой распределения.

Задача. Найти $D[X], M[X]$ НСВ, заданной функцией распределения:

$$F(x)=\begin{cases}0, & при \quad x\le 0\\ x^2, & при \quad 0<x\le 1\\ 1, & при \quad x>1\end{cases}$$ Найдем $$f(x)=F'(x)=\begin{cases}0, & при \quad x\le 0\\ 2x, & при \quad 0<x\le 1\\ 0, & при \quad x>1\end{cases},$$

$$M[\bar{x}]=\int_0^1 x\cdot 2xdx=2\int_0^1 x^2dx=\frac{2}{3}x^3\Big|_0^1=\frac{2}{3}-0=\frac{2}{3},$$

$$D[\bar{x}]=\int_0^1 x^2\,2xdx-\left(\frac{2}{3}\right)^2=2\cdot\frac{x^4}{4}\Big|_0^1-\frac{4}{9}=\frac{1}{2}-\frac{4}{9}=\frac{1}{18}.$$

1.4 Лекция №4 (2 часа).

Тема: «Законы распределения ДСВ: биноминальный и Пуассона. Законы распределения вероятностей НСВ: равномерное распределение, показательное распределение. Нормальное распределение вероятностей НСВ. Правило трех сигм»

1.4.1 Вопросы лекции:

1. Основные законы распределения ДСВ биномиальный, Пуассона.

2.Основные законы распределения НСВ: равномерный, показательный.

3. Нормальное распределение и его свойства.

1.4.2 Краткое содержание вопросов:

1. Основные законы распределения ДСВ: биномиальный, Пуассона

Задача. Найти математическое ожидание суммы числа очков, которые могут выпасть при бросании трех игральных костей.

Составим закон распределения *X:*

(для *Y* и *Z* аналогично).

$$M[X]=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}=\frac{7}{2}$$

$$M[X+Y+Z]=M[X]+M[Y]+M[Z]=\frac{7}{2}\cdot 3=\frac{21}{2}.$$

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

Теорема. $M[X]=n\cdot p$, где *X* – число появлений события А в *n* независимых испытаниях.

Например, вероятность попадания в цель при стрельбе из орудия 0,6. Найти математическое ожидание общего числа попаданий при 10 выстрелах.

$M[X]=n\cdot p=10\cdot 0{,}6=6$.

Например, вероятность попадания при одном выстреле p=0,2. Определить расход снарядов, обеспечивающих математическое ожидание числа попаданий, равное 5. $n = \frac{M[X]}{p} \Rightarrow n = \frac{5}{0{,}2} = 25$.

Пусть проводится n независимых испытаний, в каждом из которых вероятность появления события А постоянно и равна p, q – вероятность не появления события А в каждом испытании.

Случайная величина X – число появлений событий А в n независимых испытаниях.

Теорема. Дисперсия биномиального распределения с параметрами n и p определяется по формуле: $D[X] = npq$.

Например, производится 10 независимых испытаний, в каждом вероятность появления события А равна 0,7.Найти $D[X]$, где X – число наступлений события А в 10 испытаниях. $q = 0{,}3 \Rightarrow D[X] = npq = 10 \cdot 0{,}7 \cdot 0{,}3 = 2{,}1$.

Пусть проводится n независимых испытаний, в каждом из которых вероятность появления события А постоянно и равна p, q – вероятность не появления события А в каждом испытании.

Случайная величина X – число появлений событий А в n независимых испытаниях.

Теорема. Дисперсия биномиального распределения с параметрами n и p определяется по формуле: $D[X] = npq$.

Например, производится 10 независимых испытаний, в каждом вероятность появления события А равна 0,7.Найти $D[X]$, где X – число наступлений события А в 10 испытаниях. $q = 0{,}3 \Rightarrow D[X] = npq = 10 \cdot 0{,}7 \cdot 0{,}3 = 2{,}1$.

Распределение Пуассона $P(\xi = k) = \frac{a^k}{k!} e^{-a}$, $a>0$, $k=0,1,2,\ldots$

2. Законы распределения вероятностей НСВ

Рассмотрим некоторые наиболее часто встречающиеся распределения НСВ.

Закон равномерного распределения вероятностей НСВ

Рассмотрим $\overline{x}$ с законом равномерного распределения вероятностей.

$f(x)$ такой величины задается следующим образом:

$$f(x) = \begin{cases} 0, & при \quad x \le a \\ c, & при \quad a < x \le b \\ 0, & при \quad x > b \end{cases}.$$

На $(a;b)$ плотность $f(x)$ имеет постоянное значение c, вне этого интервала- равна 0. Такое распределение называется законом равномерной плотности.

$$\int_{-\infty}^{+\infty} f(x)dx = 1 = \int_a^b c dx = c(b-a) \Rightarrow c = \frac{1}{b-a} \Rightarrow b-a = \frac{1}{c}.$$

Интервал $(a;b)$, на котором имеет место равномерное распределение, обязательно конечен.

Определим вероятность того, что случайная величина $\overline{x}$ примет значение, заключенное в $(\alpha; \beta)$: $P(\alpha < \overline{x} < \beta) = \int_\alpha^\beta f(x)dx = \int_\alpha^\beta \frac{1}{b-a} dx = \frac{\beta - \alpha}{b-a}$

Определим интегральную функцию распределения равномерного закона: $F(x)=\int_{-\infty}^{x} f(x)dx$ Если $x \le a$, то $f(x)=0 \Rightarrow F(x)=0$.

Если $a < x \le b$, то $f(x)=\frac{1}{b-a} \Rightarrow F(x)=\int_{a}^{x} \frac{1}{b-a}dx=\frac{x-a}{b-a}$.

Если $b < x$, то $f(x)=0 \Rightarrow \int_{b}^{+\infty} f(x)dx=0 \Rightarrow F(x)=\int_{-\infty}^{x} f(x)dx=\int_{a}^{b} \frac{1}{b-a}dx=\frac{b-a}{b-a}=1$.

$$F(x)=\begin{cases}0, & при \quad x \le a\\ \frac{x-a}{b-a}, & при \quad a < x \le b\\ 1, & при \quad x > b\end{cases}.$$

Задача. При измерении некоторой величины производится округление до ближайшего деления шкалы. Ошибки при округлении есть случайная величина с равномерным распределением вероятностей. Задайте эту величину.

Если *2l* – число некоторых единиц в одном делении шкалы, то плотность распределения этой случайной величины будет иметь вид:

$f(x)=0, \quad x \le -l, f(x)=\frac{1}{2l}, \quad -l < x \le l, f(x)=0, \quad x > l$ Здесь $a=-l, b=l, c=\frac{1}{2l}$

Показательное (экспоненциальное) распределение

НСВ $\overline{x}$ называется распределенной по показательному закону, если она может принимать только неотрицательные значения, а плотность вероятности определяется равенством: $f(x)=\begin{cases}\lambda \cdot e^{-\lambda x}, & x \ge 0\\ 0, & x < 0\end{cases}$

Причем, λ – это параметр распределения, больший 0.

Примеры величин, распределенных по показательному закону: длительность времени безотказной работы элемента; время между появлениями двух последовательных событий простейшего потока с заданной интенсивностью λ (время между двумя сбоями ЭВМ).

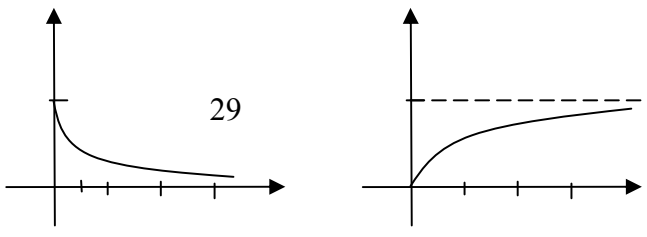
Случайные величины, распределенные показательно, обладают интересным свойством: если промежуток времени, распределенный по показательному закону, уже длился некоторое время, то это никак не влияет на закон распределения оставшейся части промежутка, он остается таким же, как и для всего промежутка.

Определим интегральную функцию $F(x)$: 1. $x < 0 \quad F(x)=0$. 2. $x \ge 0$

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{0} f(x)dx+\int_{0}^{x} f(x)dx=-e^{-\lambda x}\Big|_{0}^{x}=-\left(e^{-\lambda x}-1\right)=1-e^{-\lambda x}.$$

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

Построим графики интегральной и дифференциаль-



ной функций распределения. Для простоты построения возьмем λ=1. $f\left(\frac{1}{2}\right) \approx 0{,}6$ $f(1) \approx 0{,}4$ $f(3) \approx 0{,}1$

$$P(\alpha < \overline{x} < \beta) = \int_{\alpha}^{\beta} \lambda \cdot e^{-\lambda x} dx = -e^{-\lambda x} \Big|_{\alpha}^{\beta} = e^{-\lambda\alpha} - e^{-\lambda\beta}$$

Показательное распределение широко применяется в приложениях теории вероятностей, в частности, в теории надежности, одним из основных понятий этой теории является функция надежности.

Будем называть элементом любое устройство, независимо от его сложности.

Рассмотрим НСВ Т – длительность времени безотказной работы элемента.

Функция распределения Т определяет вероятность отказа элемента за время длительностью *t*: $P(T < t) = F(t)$.

Следовательно, вероятность безотказной работы за то же время:

$P(T \geq t) = 1 - F(t) = R(t)$ определяет функцию надежности.

$R(t) = e^{-\lambda t}; F(t) = 1 - e^{-\lambda t}$ Часто, но не всегда, случайная величина Т имеет показательное распределение.

3. Нормальный закон распределения

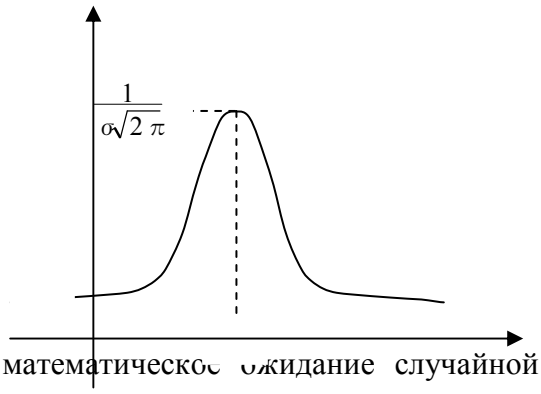
Изучение различных явлений показывает, что многие случайные величины, например, ошибки при измерениях, при стрельбе, величина износа деталей во многих механизмах и т.д., имеют следующую плотность распределения вероятностей:

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

В этом случае говорят, что случайная величина подчинена нормальному закону распределения (или закону Гаусса).

Выражение $(x-a)^2$, присутствующее в формуле плотности, позволяет сделать вывод о том, что кривая нормального распределения симметрична относительно прямой $x=a$.

Найдем $f(a) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{0} = \frac{1}{\sigma\sqrt{2\pi}}$.Кривая нормального распределения или кривая Гаусса.



Определим математическое ожидание случайной величины, подчиненной нормальному закону распределения:

$$m_{\bar{x}} = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left[\frac{x-a}{\sqrt{2}\,\sigma} = t \Rightarrow x = a + \sqrt{2}\,\sigma t,\ dx = \sqrt{2}\,\sigma dt \right] =$$

$$= \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \left(a + \sqrt{2}\,\sigma t\right) \cdot e^{-t^2} \cdot \sqrt{2} \cdot \sigma dt = \frac{1}{\sqrt{\pi}} \left(\int_{-\infty}^{+\infty} a \cdot e^{-t^2} dt + \int_{-\infty}^{+\infty} \sqrt{2}\,\sigma t \cdot e^{-t^2} dt \right) = \frac{a}{\sqrt{\pi}} \cdot \sqrt{\pi} +$$

$$+ \frac{\sqrt{2}\,\sigma}{\sqrt{\pi}} \cdot \int_{-\infty}^{+\infty} t \cdot e^{-t^2} dt = a \Rightarrow m_{\bar{x}} = a$$

Таким образом, значение параметра *a* в формуле плотности равно математическому ожиданию рассмотренной случайной величины. Значит, точка *x=a* – центр распределения вероятностей величины, подчиненной нормальному закону. А т.к. при *x=a* $f(x)$ принимает наибольшее значение, то *a* является модой этой случайной величины.

Кривая плотности $f(x)$ или кривая Гаусса симметрична относительно *x=a*, следовательно,

$$\int_{-\infty}^{a} f(x)dx = \int_{a}^{+\infty} f(x)dx = \frac{1}{2} = \frac{1}{2} \cdot \int_{-\infty}^{+\infty} f(x)dx \Rightarrow Me = a\,.$$

Если в формуле плотности *a*=0, выражение принимает вид:

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

Следовательно, кривая распределения симметрична относительно оси координат О*У* и центр распределения вероятностей совпадает с началом координат.

Форма кривой распределения не зависит от параметра *a*, величина *a* лишь определяет сдвиг кривой распределения вправо $(a > 0)$ или влево $(a < 0)$.

Рассмотрим $\bar{x}$, заданную плотностью нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}\ ;\ m_{\bar{x}} = a = 0\,.$$

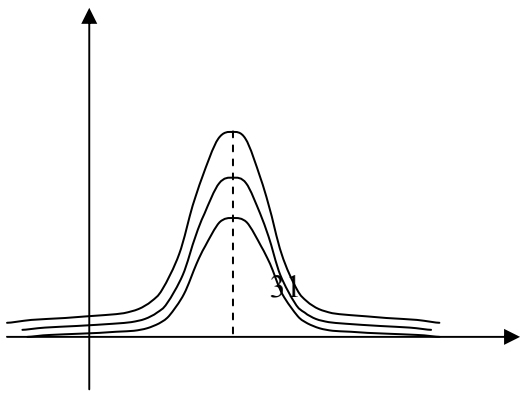
Найдем $D[\bar{x}] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - 0^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}} dx =$

$$= \left[\frac{x}{\sqrt{2}\,\sigma} = t \Rightarrow x = \sqrt{2} \cdot \sigma t,\quad dx = \sqrt{2}\,\sigma dt \right] = \int_{-\infty}^{+\infty} 2\,\sigma^2 t^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{2\sigma^2 t^2}{2\sigma^2}} \sqrt{2} \cdot \sigma dt =$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t \cdot 2\mathrm{t} \cdot e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2 \Rightarrow D[\bar{x}] = \sigma^2 \Rightarrow \sigma[\bar{x}] = \sqrt{D[\bar{x}]} = \sigma$$

Итак, дисперсия равна параметру σ^2 в формуле плотности распределения.

Выясним, как значение σ^2 влияет на форму кривой нормального распределения. Наибольшего значения кривая нормального распределения достигает в точке *a* и равно оно $f(a) = \frac{1}{\sigma\sqrt{2\pi}}$. С возрастанием σ $f(a)$ уменьшается. вдоль положительного направления оси ОУ.



Следовательно, кривая будет более пологой, т.е. сжимается к оси OX. С убыванием σ $f(a)$ увеличивается и кривая становится более «островершинной», т.е. вытягивается. Понятно, что при любых значениях *a* и *σ* площадь, ограниченная нормальной кривой и осью OX, всегда равна 1.

При *a*=0 и *σ*=1 получаем следующее выражение плотности:

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

Такую нормальную кривую называют нормированной.

Определим $P(\alpha<\bar{x}<\beta)$, если $\bar{x}$ задана плотностью нормального распределения:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad P(\alpha<\bar{x}<\beta)=\frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{\alpha}^{\beta}e^{\frac{-(x-a)^2}{2\sigma^2}}dx$$

Преобразуем формулу так, чтобы можно было пользоваться таблицами значений функции Лапласа:

$$x=\alpha\Rightarrow z=\frac{\alpha-a}{\sigma}$$

$$x=\beta\Rightarrow z=\frac{\beta-a}{\sigma}$$

Пользуясь функцией Лапласа, получим:

$$\hat{O}(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_{0}^{x}e^{\frac{-z^2}{2}}dz\Rightarrow P(\alpha<\bar{x}<\beta)=\hat{O}\left(\frac{\beta-a}{\sigma}\right)-\hat{O}\left(\frac{\alpha-a}{\sigma}\right).$$

Например, $\bar{x}$ распределена нормально с параметрами *a*=30, σ=10.

Найдем $P(10<\bar{x}<30)=\hat{O}\left(\frac{50-30}{10}\right)-\hat{O}\left(\frac{10-30}{10}\right)=\hat{O}(2)-\hat{O}(-2)=2\hat{O}(2)=2\cdot 0{,}48\approx 0{,}95$

.

Часто требуется вычислить вероятность того, что случайная величина, распределенная нормально отклонится от математического ожидания *a* по абсолютной величине меньше, чем на заданное положительное число *ε*.

$$P(|\bar{x}-a|<\varepsilon)=\hat{O}\left(\frac{\varepsilon+a-a}{\sigma}\right)-\hat{O}\left(\frac{-\varepsilon+a-a}{\sigma}\right)=\hat{O}\left(\frac{\varepsilon}{\sigma}\right)-\hat{O}\left(\frac{-\varepsilon}{\sigma}\right)=\hat{O}\left(\frac{\varepsilon}{\sigma}\right)+\hat{O}\left(\frac{\varepsilon}{\sigma}\right)=2\cdot\hat{O}\left(\frac{\varepsilon}{\sigma}\right)$$

$$P(|\bar{x}-a|<\varepsilon)=2\cdot\hat{O}\left(\frac{\varepsilon}{\sigma}\right)$$

Например, $\bar{x}$ распределена нормально с параметрами: *a*=20, σ=10.

Найдем $P(|\bar{x}-20|<3)=2\cdot\hat{O}\left(\frac{3}{10}\right)=2\cdot\hat{O}(0{,}3)\approx 2\cdot 0{,}12\approx 0{,}24$.

Пусть $\bar{x}$ распределена нормально, a=0 (для определенности).

Вычислим следующие вероятности:

$$P(|\bar{x}|<\sigma)=2\cdot\hat{O}\left(\frac{\sigma}{\sigma}\right)=2\cdot\hat{O}(1)\approx 2\cdot 0{,}3413\approx 0{,}6826\approx 0{,}683$$

$$P(|\bar{x}|<2\sigma)=2\cdot\hat{O}\left(\frac{2\sigma}{\sigma}\right)=2\cdot\hat{O}(2)\approx 2\cdot 0{,}4772\approx 0{,}954$$

$$P(|\bar{x}| < 3\sigma) = 2 \cdot \hat{O}\left(\frac{3\sigma}{\sigma}\right) = 2 \cdot \hat{O}(3) \approx 2 \cdot 0{,}49865 \approx 0{,}997$$

Вывод: почти достоверно, что случайная величина отклонится от математического ожидания не больше, чем на 3σ. Это предложение называется правилом трех сигм.

На практике правило применяется так: если распределение изучаемой случайной величины неизвестно, но условие об отклонении выполняется, то можно предположить, что указанная величина распределена нормально; в противном случае – она не распределена нормально.

Нормально распределенные случайные величины широко распространены на практике. Это объясняется теоремой, сформулированной и доказанной русским математиком Ляпуновым:

Если случайная величина X – это сумма очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.

Закон Гаусса является предельным законом, к которому приближаются другие законы при типичных условиях.

1.5 Лекция №5 (2 часа).

Тема: «Задачи математической статистики. Статистический материал. Статистические параметры распределения. Статистические оценки параметров распределения»

1.5.1 Вопросы лекции:

1. Статистический материал и его первичная обработка.

2. Эмпирические законы распределения. Полигон частот, гистограмма.

3. Числовые характеристики выборки.

4. Точечные оценки выборочных характеристик.

1.5.2 Краткое содержание вопросов:

1. Статистический материал и его первичная обработка

Предметом математической статистики является изучение случайных величин (или случайных событий) по результатам наблюдений.

Для получения опытных данных необходимо провести обследование соответствующих объектов.

Совокупность всех мысленно возможных объектов данного вида, над которыми проводятся наблюдения с целью получения конкретных значений определённой случайной величины, называется **генеральной совокупностью**.

Генеральную совокупность будем называть **конечной** или **бесконечной** в зависимости от того, конечна или бесконечна совокупность составляющих её элементов.

Часть отобранных объектов из генеральной совокупности называется **выборочной совокупностью** или **выборкой**.

Число *N* объектов генеральной совокупности и число *n* объектов выборочной совокупности будем называть **объёмами генеральной и выборочной совокупности** соответственно.

Для того чтобы по выборке можно было достаточно уверенно судить о случайной величине, выборка должна быть **представительной (репрезентативной)**. Репрезентативность выборки означает, что объекты выборки достаточно хорошо представляют генеральную совокупность. Она обеспечивается случайностью отбора.

Существуют несколько способов отбора, обеспечивающих репрезентативность выборки. Рассмотрим некоторые из них.

После того как сделана выборка, все объекты этой совокупности обследуются по отношению к определённой случайной величине и получают наблюдаемые данные.

Для изучения закономерностей варьирования значений случайной величины опытные данные подвергают обработке.

Операция, заключающаяся в том, что результаты наблюдений над случайной величиной, т.е. наблюдаемые значения случайной величины, располагают в порядке неубывания, называется **ранжированием опытных данных**.

После операции ранжирования опытные данные объединяют в группы так, чтобы в каждой отдельной группе значения случайной величины будут одинаковы.

Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется вариантом (x_i) (**вариантой**), а изменение этого значения – **варьированием**.

Численность отдельной группы сгруппированного ряда наблюдаемых данных называется **частотой** или **весом**(m_i) соответствующей **варианты**.

Отношение частоты данного варианта к общей сумме частот всех вариантов называется **частостью** или долей этой **варианты** (p_i): $p_i = \dfrac{m_i}{\sum\limits_{i=1}^{\upsilon} m_i}$,

где υ – число вариант. Полагая $n = \sum\limits_{i=1}^{\upsilon} m_i$, где n – объём выборки, имеем:

$$p_i = \frac{m_i}{n} .$$

Заметим, что частость p_i – статистическая вероятность появления варианта x_i.

Дискретным вариационным рядом распределения называется ранжированная совокупность вариантов x_i, с соответствующими им частотами m_i или частостями p_i .

Если изучаемая случайная величина является непрерывной, то ранжирование и группировка наблюдаемых значений зачастую не позволяют выделить характерные черты варьирования её значений. Это объясняется тем, что отдельные значения случайной величины могут как угодно мало отличаться друг от друга и поэтому в совокупности наблюдаемых данных одинаковые значения случайной величины могут встречаться редко, а частоты вариантов мало отличаются друг от друга.

Интервальным вариационным рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений величины.

Рассмотрим алгоритм построения интервального ряда.

1. Для построения интервального ряда необходимо определить величину частичных интервалов, на которые разбивается весь интервал варьирования наблюдаемых значений случайной величины. Считая, что все частичные интервалы имеют одну и ту же длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдений. Длину частичного интервала h следует выбрать так, чтобы построенный ряд не был громоздким и в то же время позволил выявить характерные черты изменения значений случайной величины.

2. Найдём размах варьирования ряда R:

$R = x_{наиб} - x_{наим}$

Выберем число интервалов υ (обычно от 7 до 11).

3. Для более точного определения величины частичного интервала можно воспользоваться **формулой Стерджеса**: $h = \frac{R}{1+3{,}322\lg n}$.

Если h – дробное, то за длину частичного интервала следует брать ближайшее целое число, либо ближайшую простую дробь.

4. За начало первого интервала следует брать величину: $x_{нач} = x_{наим} – 0{,}5h$.

5. Конец последнего интервала ($x_{кон}$) должен удовлетворить условию:

$x_{кон} – h \le x_{наиб} < x_{кон}$.

6. Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала h.

7. Определим, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значения случайной величины, большие или равные нижней границе и меньшие верхней границы. Иногда интервальный вариационный ряд для простоты исследования условно заменяют дискретным. В этом случае серединное значение i-го интервала принимают за вариант x_i, а соответствующую интервальную частоту m_i – за частоту этой варианты.

2. Эмпирические законы распределения. Полигон частот, гистограмма

Закон распределения (или просто распределение) случайной величины можно задать различными способами. Например, дискретную случайную величину можно задать с помощью или ряда распределения, или интегральной функции, а непрерывную случайную величину – с помощью или интегральной, или дифференциальной функции. Рассмотрим выборочные аналоги этих двух функций.

В теории вероятностей для характеристики распределения случайной величины X служит интегральная функция распределения $F(x) = P(X<x)$. В дальнейшем, если величина X распределена по некоторому закону $F(x)$, будем говорить, что и генеральная совокупность распределена по закону$F(x)$. Введём выборочный аналог функции $F(x)$.

Пусть имеется выборочная совокупность значений некоторой случайной величины X объёма n и каждому варианту из этой совокупности поставлена в соответствие его частость. Пусть, далее, x – некоторое действительное число, а m_x– число выборочных значений случайной величины X, меньших x. Тогда число m_x/n является частостью наблюдаемых в выборке значений величины X, меньших x, т.е. частостью появления события $X<x$. При изменении x в общем случае будет изменяться и величина m_x/n. Это означает, что относительная частота m_x/n является функцией аргумента x. А т.к. эта функция находится по выборочным данным, полученным в результате опытов, то её называют **выборочной** или **эмпирической.**

Выборочной функцией распределения (или **функцией распределения выборки**) называется функция $F(x)^*$, задающая для каждого значения x относительную частоту события $X<x$.

Итак, по определению, $F(x)^* = m_x/n$, где n – объём выборки, m_x – число выборочных значений случайной величины X, меньших x. В отличие от выборочной функции $F(x)^*$ интегральную функцию $F(x)$ генеральной совокупности называют **теоретической функцией распределения**. Главное различие функций $F(x)$ и $F(x)^*$ состоит в том, что теоретическая функция распределения $F(x)$ определяет вероятность события$X<x$, а выборочная функция – относительную частоту этого события.

Свойство статистической устойчивости частоты, обоснованное теоремой Бернулли, оправдывает целесообразность использования функции $F(x)^*$ при большихn в качестве приближённого значения неизвестной функции $F(x)$.

В заключение отметим, что функция $F(x)$ и её выборочный аналог $F(x)^*$ обладают одинаковыми свойствами. Действительно, из определения функции $F(x)^*$ имеем следующие свойства:

1. $0 \le F^*(x) \le 1$
2. $F^*(x)$ – неубывающая функция.
3. $F^*(-\infty) = 0$, $F(\infty) = 1$.

Такими же свойствами обладает и функция $F(x)$.

Функцию $F^*(x)$ наряду с табличным способом задания можно задать аналитически. В этом случае $F^*(x)$ определяется так:

$$F^*(x) = \begin{cases} 0 & \text{при } x \le x_1, \\ \sum_{l=1}^{i-1} p^*_l & \text{при } x_{i-1} < x \le x_i,\ i = 1, 2, 3, \ldots, v, \\ 1 & \text{при } x > x_v. \end{cases} \qquad (1)$$

Здесь x_v совпадает с $x_{\text{наиб}}$. Частости $\sum_{l=1}^{i-1} p^*_l$ обычно называются **накопленными частостями.**

Для интегральной функции распределения $F(x)$ справедливо приближённое равенство $F(x+\Delta x) - F(x) \approx f(x)\Delta x$, где $f(x)$ – дифференциальная функция распределения или функция плотности вероятности. Из этого равенства следует, что $f(x) \approx (F(x+\Delta x) - F(x))/\Delta x$. Поэтому естественно выборочным аналогом функции $f(x)$ считать функцию

$$f^*(x) = \frac{F^*(x+\Delta x) - F^*(x)}{\Delta x}, \qquad (2)$$

где $F^*(x+\Delta x) - F^*(x)$ – частость попадания наблюдаемых значений случайной величины X в интервал $[x, x+\Delta x]$. Таким образом, значение $f(x)$ характеризует плотность частости на этом интервале.

Пусть наблюдаемые над непрерывной случайной величиной данные представлены в виде интервального вариационного ряда. Полагая, что p^*_1 – частость попадания наблюдаемых значений случайной величины в интервал $[a_i; a_i+h]$, где h – длина частичного интервала, и учитывая равенство (2), для $x \in [a_i; a_i+h]$ запишем $f(x)=p^*_i/h$. Тогда выборочную функцию плотности $f(x)$ можно задать соотношением $f(x) = \begin{cases} 0 & \text{при } x < a_1, \\ p^*_i/h & \text{при } a_i \le x < a_{i+1}, \\ 0 & \text{при } x \ge a_{v+1}, \end{cases}$ $i = 1, 2, 3, \ldots, v$,

где a_{v+1} – конец последнего v-го интервала.

Наблюдаемые данные, представленные в виде вариационного ряда, можно изобразить графически, используя не только функцию $F^*(x)$. К наиболее распространённым видам графического изображения вариационных рядов относятся **полигон** и **гистограмма.** Графическое изображение рядов с помощью полигона или гистограммы позволяет получить наглядное представление о закономерности варьирования наблюдаемых значений случайной величины.

Полигон обычно используют для изображения дискретного вариационного ряд. Для его построения в прямоугольной системе координат наносят точки с координатами $(x_i; m_i)$ или $(x_i; p^*_i)$, где x_i – значение i-го варианта, а m_i (p^*_i) – соответствующие частоты (частости). Затем отмеченные точки соединяют отрезками прямой линии. Полученная ломаная называется **полигоном.**

Если полигон частостей построен по дискретному вариационному ряду дискретной случайной величины, то его называют **многоугольником распределения частостей,** который является выборочным аналогом многоугольника распределения вероятностей. За-

метим, что сумма ординат многоугольника распределения частостей, как и у многоугольника распределения вероятностей, равна 1, т.к. $\sum p^*_i=1$.

Гистограмма служит только для изображения интервальных вариационных рядов. Для её построения в прямоугольной системе координат на осиOx откладывают отрезки, изображающие частичные интервалы варьирования, и на этих отрезках, как на основаниях, строят прямоугольники с высотами, равными частотам или частостям соответствующих интервалов. В результате такой операции получают ступенчатую фигуру, состоящую из прямоугольников, которую называют **гистограммой.**

Для графического изображения интервального вариационного ряда можно использовать полигон, если этот ряд преобразовать в дискретный. В этом случае интервалы заменяют их серединными значениями и ставят им в соответствие интервальные частоты (частости). Для полученного дискретного ряда строят полигон.

3. Числовые характеристики выборки

Построив вариационный ряд и изобразив его графически, можно получить первоначальное представление о закономерностях, имеющих место в ряду наблюдений. Однако на практике зачастую этого недостаточно. Такая ситуация возникает, когда следует уточнить те или иные сведения о ряде распределения или, когда имеется необходимость сравнить два ряда и более. При этом следует сравнивать однотипные вариационные ряды, т.е. такие ряды, которые получены при обработке сравнимых статистических данных.

Сравниваемые распределения могут существенно отличаться друг от друга. Они могут иметь различные средние значения случайной величины, вокруг которых группируются в основном остальные значения, или различаться рассеиванием данных наблюдений вокруг указанных значений и т.д. Поэтому для дальнейшего изучения изменения значений случайной величины используют числовые характеристики вариационных рядов. Поскольку эти характеристики вычисляются по статистическим данным (данным, полученным в результате наблюдений), их обычно называют **статистическими характеристиками** или **оценками.**

Пусть собранный и обработанный статистический материал представлен в виде вариационного ряда. Теперь результаты наблюдений над случайной величиной следует подвергнуть анализу и выявить характерные особенности поведения случайной величины. Для этого удобнее всего выделить некоторые постоянные, которые представляли бы вариационный ряд в целом и отражали присущие изучаемой совокупности закономерности.

Некоторые из этих постоянных отличаются тем, что вокруг них концентрируются остальные результаты наблюдений. Такие величины называются **средними величинами.** К ним относятся среднее арифметическое (среднее выборочное), среднее геометрическое, среднее гармоническое и т.д. Однако эти характеристики не отражают «величину изменчивости» наблюдаемых данных, например, величину разброса значений признака вокруг среднего арифметического. Другими словами, упомянутые средние величины не отражают вариацию.

Для характеристики изменчивости случайной величины, т.е. вариации, служат показатели вариации. К ним относятся размах варьирования R, среднее квадратическое отклонение, дисперсия и т.д.

4. Точечные оценки выборочных характеристик

Точечные оценки параметров статистического распределения

Выборочная характеристика, используемая в качестве приближённого значения неизвестной генеральной характеристики, называется её **точечной статистической оценкой**.

Среднее арифметическое $\overline{\tilde{O}}$ - это точечная статистическая оценка математического ожидания M(X); D*(X)– оценка дисперсии D(X).

«Точечная» означает, что оценка представляет собой число или точку на числовой оси. «Статистическая» означает, что оценка рассчитывается по результатам наблюдений, т.е. по собранной исследователем статистике. Далее слово «статистическая» будет опускаться.

Обозначим через Θ («тэта») некоторую генеральную характеристику (ею может быть и *MX*, и любая другая числовая характеристика случайной величины *X*). Её числовое значение неизвестно, однако предложен некоторый алгоритм или формула вычисления точечной оценки $\Theta_{(n)}$ этой характеристики по результатам $X_1, X_2, \ldots, X_n$ наблюдений величины *X*. Обозначая буквой *f* этот алгоритм, запишем $\Theta^*_{(n)}=f(X_1, X_2, \ldots, X_n)$. (3)

Подставив в (3) вместо $X_1, X_2, \ldots, X_n$ конкретные результаты наблюдений (конкретные числа), получим число, которое и принимают за приближённое значение неизвестной генеральной характеристики Θ. Найти погрешность этого приближения нельзя, поскольку числовое значение характеристики Θ неизвестно. Чтобы ответить на вопрос, хорошо или нет найденное приближение, рассмотрим оценку $\Theta^*_{(n)}$ с других позиций.

Пусть в формуле (3) $X_1, X_2, \ldots, X_n$ – не конкретные числа, а лишь обозначения тех результатов наблюдений, которые мы хотели бы получить. Но результат каждого отдельного наблюдения случайной величины случаен, т.е. $X_1, X_2, \ldots, X_n$– это случайные величины, поэтому и оценка $\Theta^*_{(n)}$ также величина случайная; следовательно, можно говорить о её математическом ожидании ($M(\Theta^*_{(n)})$), дисперсии ($D(\Theta^*_{(n)})$) и законе распределения. Интерпретация оценки $\Theta^*_{(n)}$ как случайной величины позволяет сформулировать свойства, которыми должна была обладать оценка, чтобы её можно было считать хорошим приближением к неизвестной генеральной характеристике. Это свойства состоятельности, несмещённости и эффективности.

Оценка $\Theta^*_{(n)}$ генеральной характеристики Θ называется **состоятельной**, если для любого ε>0 выполняется равенство $\lim_{n\to\infty} P(\left|\Theta^*_{(n)}-\Theta\right|<\varepsilon)=1$. (4)

Поясним смысл равенства (4). Пусть ε – очень малое положительное число. Тогда равенство (4) означает, что чем больше число наблюдений *n*, тем больше уверенность (вероятность) в незначительном по абсолютной величине отклонении оценки $\Theta^*_{(n)}$ от неизвестной характеристики Θ или короче: чем больше объём исходной информации, тем «ближе мы к истине». Если это так, то $\Theta^*_{(n)}$ – состоятельная оценка.

«Хорошая» оценка обязательно должна обладать свойством состоятельности. В противном случае оценка не имеет практического смысла: увеличение объёма исходной информации не будет «приближать нас к истине». Поэтому свойство состоятельности следует проверять в первую очередь.

Оценка $\Theta^*_{(n)}$ генеральной характеристики Θ называется **несмещённой**, если для любого фиксированного числа наблюдений *n* выполняется равенство $M(\Theta^*_{(n)})=\Theta$, (5) т.е. математическое ожидание оценки равно неизвестной характеристике.

Несмещённая оценка $\Theta^*_{(n)}$ характеристики Θ называется **несмещённой эффективной**, если она среди всех прочих несмещённых оценок той же самой характеристики обладает наименьшей дисперсией.

Метод нахождения оценки неизвестного параметра, основанный на требовании максимизации функции правдоподобия, называется **методом максимального правдоподобия**, а найденная этим методом оценка – **оценкой максимального правдоподобия**.

Функции *L* и *lnL*, рассматриваемые как функции параметра λ, достигают максимума при одном и том же значении λ, т.к. *lnL* – монотонно возрастающая функция. Поэтому вместо отыскания максимума функции *L* находят (что удобнее) максимум функции *lnL*. Функция *lnL* называется **логарифмической функцией правдоподобия**.

Для $L(X_1, X_2, \dots X_n; \lambda)= \frac{\lambda^{i=1} e^{-n\lambda}}{\tilde{O}_1!\,\tilde{O}_2!...\tilde{O}_n!}$ логарифмическая функция правдоподобия имеет вид:

$$lnL(X_1, X_2, \dots X_n; \lambda) = ln \frac{\lambda^{\sum_{i=1}^{n} x_1} e^{-n\lambda}}{X_1!\,X_2!...X_n!} = \left(\sum_{i=1}^{n} X_i\right) \ln\lambda - n\lambda - \ln(X_1!) - \ln(X_2!) - \dots - - - \ln(X_n!).$$

Найдём точку максимума этой функции, рассматривая её как функцию параметра λ.

Для этого: найдём производную функции ln$L$по λ: $\frac{\partial \ln L}{\partial \lambda} = -\frac{\sum_{i=1}^{n} X_i}{\lambda} n$; приравняв производную нулю, определим критическую точку – корень полученного уравнения – **уравнения правдоподобия**:

$$\frac{\sum_{i=1}^{n} X_i}{\lambda} - n = 0 \to \lambda_{\hat{e}\eth} = \sum_{i=1}^{n} X_i / n;$$

найдём вторую производную функции ln$L$ и её значение в точке $\lambda_{кр}$:

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \frac{\sum_{i=1}^{n} X_i}{\lambda^2}; \frac{\partial^2 \ln L(\lambda_{\hat{e}\eth})}{\partial \lambda^2} = \frac{\sum_{i=1}^{n} X_i}{\lambda_{\hat{e}\eth^2}} = -\frac{n^2}{\sum_{i=1}^{n} X_i}.$$

Итак, всегда $\lambda_{кр}= \sum_{i=1}^{n} X_i / n$ - это точка максимума функции ln$L$ (или L), поэтому она и является оценкой $\lambda_{мп}$ максимального правдоподобия для неизвестного параметра λ,

т.е. $\lambda^*_{i\ddot{i}} = \sum_{i=1}^{n} X_i / n = \overline{X}$.

1.6 Лекция № 6 (2 часа).

Тема: «Интервальные оценки параметров статистического распределения. Необходимость их введения. Доверительные интервалы. Доверительные вероятности. Доверительные интервалы для оценки математического ожидания нормального распределения. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения»

1.6.1 Вопросы лекции:

1. Интервальные оценки, их свойства.
2. Метод доверительных интервалов при заданных условиях.
3. Метод моментов.

1.6.2 Краткое содержание вопросов:

1. Интервальные оценки параметров статистического распределения. Доверительные вероятности

Вычисляя на основании результатов наблюдений точечную характеристику Θ* неизвестной числовой характеристики Θ, мы понимаем, что величина Θ* является лишь приближённым значением характеристики Θ. Если для большого числа наблюдений точность приближения бывает достаточной для практических выводов (в силу несмещённости, состоятельности и эффективности «хороших» оценок), то для выборок небольшого объёма вопрос о точности оценок очень важен. В математической статистике он решается следующим образом. По сделанной выборке находится точечная оценка Θ* неизвестной характеристики Θ, затем задаются вероятностью γ и по определённым правилам находят такое число ε>0, чтобы выполнялось соотношение

$$P(\underline{\Theta^*-\varepsilon} < \Theta < \underline{\Theta^*+\varepsilon}) = \gamma. \qquad (8)$$

Соотношению (8) тождественно соотношению

$$P(|\Theta^* - \Theta| < \varepsilon) = \gamma, \qquad (9)$$

из которого видно, что абсолютная погрешность оценки Θ не превосходит числа ε. Это верно с вероятностью, равной γ. Число ε называется **точностью оценки** Θ* (чем меньше ε, тем выше точность оценки), числа Θ_1 и Θ_2 называются **доверительными границами**, интервал (Θ_1, Θ_2) – **доверительным интервалом** или **интервальной оценкой** характеристики Θ, вероятность γ называется **доверительной вероятностью** или **надёжностью** интервальной оценки.

В соотношении (8) случайными величинами являются доверительные границы Θ_1 и Θ_2: во-первых, эти границы могут изменяться при переходе от одной выборки к другой хотя бы потому, что при этом изменяется значение оценки Θ*; во-вторых, при фиксированной выборке границы Θ_1 и Θ_2 изменяются при изменении вероятности γ, поскольку ε выбирается в зависимости от γ. Генеральная же характеристики Θ – постоянная величина. Поэтому соотношение (8) следует читать так: «вероятность того, что интервал (Θ_1, Θ_2) накроет характеристику Θ, равна γ»; именно «интервал накроет характеристику», а не «характеристика попадёт в интервал».

Надёжность γ принято выбирать равной 0,95; 0,99; 0,999. Тогда событие, состоящее в том, что интервал (Θ_1, Θ_2) накроет характеристику Θ, будет практически достоверным. Также практически достоверным является событие, состоящее в том, что погрешность оценки Θ* меньше ε, или, иначе, точность оценки Θ* больше ε.

В соотношении (8) границы Θ_1 и Θ_2 симметричны относительно точечной оценки Θ*. Обратим внимание на то, что не всегда удаётся построить границы с таким свойством.

Поскольку довольно часто встречаются нормально распределённые случайные величины, построим интервальные оценки для параметров нормального распределения – математического ожидания a и среднего квадратического отклонения σ.

2. Метод доверительных интервалов при заданных условиях

Обозначим через X случайную величину, имеющую нормальный закон распределения с параметрами a и σ, т.е. $X = N(a, \sigma)$. Будем предполагать, что наблюдения над этой величиной независимы и проводятся в одинаковых условиях, т.е. возможные результаты $X_1, X_2, \ldots, X_n$ этих наблюдений обладают следующими свойствами:

$X_1, X_2, \ldots, X_n$ – независимые случайные величины;
закон распределения любой из величин $X_1, X_2, \ldots, X_n$ совпадает с законом распределения величины X, т.е.

$$X_1 = N(a, \sigma), X_2 = N(a, \sigma), \ldots, X_n = N(a, \sigma). \qquad (10)$$

Интервальная оценка математического ожидания нормального распределения при известной дисперсии

Итак, $X = N(a, \sigma)$, причём математическое ожидание a неизвестно, а дисперсия σ^2 известна. По наблюдениям найдём точечную оценку $\overline{\tilde{O}} = \sum_{i=1}^{n} X_i / n$ математического ожидания a. Зададимся вероятностью γ и попробуем найти такое число ε, чтобы выполнялось соотношение

$$P(\overline{X} - \varepsilon < a < \overline{X} + \varepsilon) = \gamma. \qquad (11)$$

Интервальная оценка математического ожидания такова:

$$(\overline{X} - u_\gamma \sigma / \sqrt{n} \ , \ \overline{X} + u_\gamma \sigma / \sqrt{n}). \qquad (12)$$

Полученный результат имеет следующий смысл: с вероятностью γ можно быть уверенным в том, что интервал (12) накроет среднее математическое ожидание.

3. Метод моментов

Параметрическое оценивание закона распределения

Результаты предварительной обработки наблюдений случайной величины, дополненные сведениями о сущности изучаемого явления, зачастую оказываются достаточными для того, чтобы сформулировать гипотезу о модели закона распределения изучаемой случайной величины, нормальный ли этот закон, биномиальный или какой-либо другой. Используя наблюдения, можно найти оценки параметров предполагаемой модели, т.е. оценки входящих в модель числовых характеристик. Подставив в модель вместо параметров найденные оценки, получим оценку предполагаемой модели закона распределения, которая называется **параметрической**. Оценивание закона распределения, не требующее предварительного выбора его модели и оценивания входящих в неё параметров, называется **непараметрическим.** Примерами непараметрических оценок неизвестного закона распределения являются вариационный ряд, выборочная функция распределения и выборочная плотность распределения.

Пример 3. Дано случайное распределение успеваемости 100 студентов-заочников, сдававших четыре экзамена:

Число сданных экзаменов					
Число студентов				5	0

Здесь случайной величиной является число сданных экзаменов среди четырёх. Обозначим её X. Установим закон распределения этой величины.

Построим сначала его непараметрическую оценку. Величина X – дискретная. Дискретный вариационный ряд, заданный столбцами 2 и 4 табл. 5, даёт непараметрическую оценку закона распределения числа сданных экзаменов среди четырёх сдаваемых.

Теперь сформулируем гипотезу о модели закона распределения случайной величины X – числе сданных экзаменов среди четырёх сдаваемых. Процесс сдачи четырёх экзаменов представим как четыре испытания, относительно которых сделаем следующие допущения:

Таблица 5.

	Чис-ло сданных эк-заменов x_i	Чи сло ст удентов m_i	Ч астость p_i^*	$p_i^{\text{теор}}=$ $=\tilde{N}_4^{x_i}\cdot$ $0{,}88^{x_i}\cdot$ $0{,}12^{4-x_i}$	$=np_i^{\text{теор}}$ $m_i^{\text{теор}}$	($m_i -$ $m_i^{\text{теор}})^2$	($m_i -$ - $m_i^{\text{теор}})^2$: :m_i теор
	2	3	4	5	6	7	8
	0 1 2 3 4	1 1 5 3 35 60	0, 01 0, 01 0, 03 0, 35 0, 60	0,0002 1 0,0060 8 0,0669 1 0,3271 1 0,5996 9	0,02 1 0,60 8 7,32 6,69 1 32,7 11 59,9 69	5, 382 5, 239 0, 001	0, 735 0, 160 0, 000
Итого		$n =$ 100	1, 00	1,0000 0			0, 895

- эти испытания независимы, т.е. вероятность сдачи любым студентом любого экзамена не зависит от того, будет сдано или нет любое количество других экзаменов;

- вероятность сдачи студентом любого отдельно взятого экзамена одна и та же и равна p, а вероятность «несдачи» равна $(1 - p)$.

Конечно, эти допущения могут вызывать некоторые сомнения, но возможно, что они не будут противоречить результатам наблюдений. При этих допущениях мы имеем дело с испытаниями Бернулли и число сданных экзаменов среди четырёх сдаваемых будет иметь биномиальный закон распределения, т.е. вероятность того, что студент сдаст λ экзаменов, равна

$$P(X = x) = C_4^x p^x (1 - p)^{4-x}, \quad x = 0, 1, 2, 3, 4. \qquad (6)$$

Найдём оценку параметра p, входящего в модель (6). В условиях испытаний Бернулли состоятельной, несмещённой и эффективной оценкой вероятности является частость. В рассматриваемом примере p – вероятность того, что студент сдаст экзамен, поэтому частость p^* этого события, учитывая, что имеются сведения об успеваемости 100 студентов, вычисляем следующим образом:

$$p^* = \frac{\text{числоэкзаменов,сданных 100 студентами}}{\text{числоэкзаменов,сдаваемых 100 студентами}} = \frac{\sum_{i=1}^{5} x_i m_i}{4\times 100} = \frac{0\times 1+1\times 1+2\times 3+3\times 35+4\times 60}{100\times 4} = 0{,}88.$$

Так как $\sum_{i=1}^{5} x_i m_i / 100 = \overline{X}$ - это среднее число экзаменов, сданных одним студентом, то p^* можно было бы определить и так:

$$p^* = \frac{\text{среднеечислоэкзаменов,сданныходнимстудентом}}{\text{числоэкзаменов,сдаваемыходнимстудентом}} = \frac{\overline{\tilde{O}}}{4} = 0{,}88.$$

Заметим, что если находить оценку параметра p в модели (6) методом максимального правдоподобия и при этом учесть, что число x_i наблюдалось m_i раз, то мы получили бы для p^* такую же формулу, а именно

$$p^*_{\text{мп}} = \sum_{i=1}^{5} x_i m_i /(4n).$$

Подставив в модель (6) вместо параметра p его оценку p^*, получим параметрическую оценку неизвестного закона распределения числа сданных экзаменов, построенную в предположении, что допустима биномиальная модель

$$P(X = x) = C_4^x 0{,}88^x 0{,}12^{4-x}; \quad x = 0,\ 1,\ 2,\ 3,\ 4. \qquad (7)$$

Теоретические вероятности $p_i^{\text{теор}}$ и частоты $m_i^{\text{теор}}$, вычисленные в предположении, что имеет место модель (7), содержатся в столбцах 5 и 6 табл. 5. Поскольку различия между соответствующими числами столбцов 4 и 5 или между числами столбцов 3 и 6 небольшие, можно сделать предварительное заключение о приемлемости биномиальной модели. Графически это заключение подтверждается рисунком, на котором кривая вероятностей $p_i^{\text{теор}}$ близка к кривой частостей p_i^*.

Метод более глубокого обоснования приемлемости той или иной модели называется **критерием согласия**.

1.7 Лекция №7 (2 часа)

Тема: «Понятие статистической гипотезы. Виды гипотез. Статистический критерий. Критическая область. Мощность критерия. Критерии согласия: критерий Пирсона. Выравнивание рядов»

1.7.1 Вопросы лекции:

1. Статистические гипотезы, ошибки первого и второго рода.
2. Статистические критерии, их виды, мощность критерия.
3. Критерий Пирсона.
4. Выравнивание статистических рядов.

1.7.2 Краткое содержание вопросов:

1. Статистические гипотезы, ошибки первого и второго рода

Под **статистической гипотезой** понимают всякое высказывание о генеральной совокупности (случайной величине), проверяемое по выборке (по результатам наблюдений). Примером статистических гипотез являются следующие высказывания: генеральная совокупность, о которой мы располагаем лишь выборочными сведениями, имеет нормальный закон распределения или генеральная средняя (математическое ожидание случайной величины) равна 5. Не располагая сведениями о всей генеральной совокупности, высказанную гипотезу сопоставляют, по определённым правилам, с выборочными сведениями, и делают вывод о том, можно принять гипотезу или нет. Процедура сопоставления высказанной гипотезы с выборочными данными называется **проверкой гипотезы**.

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

Этап 1. Располагая выборочными данными X_1, X_2, …, X_n и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0, которую называют **основной** или **нулевой**, и гипотезу H_1, **конкурирующую** с гипотезой H_0.

Термин «конкурирующая» означает, что являются противоположными следующие два события:

- по выборке будет принято решение о справедливости для генеральной совокупности гипотезы H_0;
- по выборке будет принято решение о справедливости для генеральной совокупности гипотезы H_1.

Гипотезу H_1 называют также **альтернативной**.

Например, если нулевая гипотеза такова: математическое ожидание равно 5, – то альтернативная гипотеза может быть следующей: математическое ожидание меньше 5, что записывается следующим образом:

H_0 : M(X) = 5; H_1 :M(X)< 5.

Этап 2. Задаются вероятностью α («альфа»), которую называют **уровнем значимости**. Поясним её смысл:

Решение о том, можно ли считать высказывание H_0 справедливым для генеральной совокупности, принимается по выборочным данным, т.е. по ограниченному ряду наблюдений; следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов:

- отвергают гипотезу H_0, или, иначе, принимают альтернативную гипотезу H_1, тогда как на самом деле гипотеза H_0 верна – это **ошибка первого рода**;

- принимают гипотезу H_0, тогда как на самом деле высказывание H_0 неверно, т.е верной является гипотеза H_1 – это **ошибка второго рода**.

Так вот, уровень значимости α – это вероятность ошибки первого рода, т.е. α = $P_{H_0}(H_1)$, (13)

где $P_{H_0}(H_1)$ – вероятность того, что будет принята гипотеза H_1, если на самом деле в генеральной совокупности верна гипотеза H_0. Вероятность α задаётся заранее, разумеется, малым числом, поскольку это вероятность ошибочного заключения, при этом обычно используют некоторые стандартные значения: 0,05; 0,01; 0,005; 0,001. Например, α = 0,05 означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объёма, то в среднем в 5 случаях из 100 мы совершим ошибку первого рода.

Вероятность ошибки второго рода обозначают β, т.е. β = $P_{H_1}(H_0)$, (14)

где $P_{H_1}(H_0)$ – вероятность того, что будет принята гипотеза H_0, если на самом деле верна гипотеза H_1. Зная α, можно найти вероятность β.

Обратим внимание на то, что в результате проверки гипотезы относительно гипотезы H_0 может быть принято и правильное решение. Существует правильное решение двух следующих видов:

- принимают гипотезу H_0, тогда как и в действительности, в генеральной совокупности, она имеет место; вероятность этого решения $P_{H_0}(H_0) = 1 - \alpha$;

- не принимают гипотезу $P_{H_0}(H_0) = 1 - \alpha$ (т.е. принимают гипотезу H_1), тогда как на самом деле гипотеза H_0 неверна (т.е. верна гипотеза H_1); вероятность этого решения $P_{H_1}(H_1) = 1 - \beta$.

Этап 3. Находят величину φ такую, что:

- её значения зависят от выборочных данных X_1, X_2, …, X_n, т.е. для которой справедливо равенство φ = φ(X_1, X_2, …, X_n);

- её значения позволяют судить о «расхождении выборки с гипотезой H_0»;

- и она, будучи величиной случайной в силу случайности выборки X_1, X_2, …, X_n, подчиняется при выполнении гипотезы H_0 некоторому известному, затабулированному закону распределения.

2. Статистические критерии, их виды, мощность критерия.

Величину φ называют **критерием**.

Отметим, что в основе метода построения критерия лежит понятие функции правдоподобия.

Этап 4. Далее рассуждают так. Т.к. значения критерия позволяют судить о «расхождении выборки с гипотезой H_0», то из области допустимых значений критерия φ следует выделить подобласть ω таких значений, которые свидетельствовали бы о существенном

расхождении выборки с гипотезой H_0, и, следовательно, о невозможности принять гипотезу H_0. Подобласть ω называют **критической областью**. Допустим, что критическая область выделена. Тогда руководствуются следующим правилом: если вычисленное по выборке значение критерия φ попадает в критическую область, то гипотеза H_0 отвергается и принимается гипотеза H_1. При этом следует понимать, что такое решение может оказаться ошибочным: на самом деле гипотеза H_0 может быть справедливой. Т.обр., ориентируясь на критическую область, можно совершить ошибку первого рода, вероятность которой задана заранее и равна α. Отсюда вытекает следующее требование к критической области ω:

вероятность того, что критерий φ примет значение из критической области ω, должна быть равна заданному числу α, т.е. $P(\varphi \in \omega) = \alpha$. (15)

Однако критическая область равенством (15) определяется неоднозначно. Действительно, представив себе график функции плотности $f_\varphi(x)$ критерия φ, нетрудно понять, что на оси абсцисс существует бесчисленное множество областей-интервалов таких, что площади построенных на них криволинейных трапеций равны α, т.е. областей, удовлетворяющих требованию (15). Поэтому кроме требования (15) выдвигается следующее требование: критическая область ω должна быть расположена так, чтобы при заданной вероятности α ошибки первого рода вероятность β ошибки второго рода была минимальной.

Возможны три вида расположения критической области (в зависимости от вида нулевой и альтернативной гипотез, вида и расположения критерия φ):

правосторонняя критическая область, состоящая из интервала $(x^{кр}_{пр,\,\alpha}, +\infty)$, где точка $x^{кр}_{пр,\,\alpha}$ определяется из условия $P(\varphi > x^{кр}_{пр,\,\alpha}) = \alpha$ (16)

и называется **правосторонней критической точкой**, отвечающей уровню значимости α;

левосторонняя критическая область, состоящая из интервала $(-\infty, x^{кр}_{лев,\,\alpha})$, где точка $x^{кр}_{лев,\,\alpha}$ определяется из условия $P(\varphi < x^{кр}_{лев,\,\alpha}) = \alpha$ (17)

и называется **левосторонней критической точкой**, отвечающей уровню значимости α;

двусторонняя критическая область, состоящая из следующих двух интервалов: $((-\infty, x^{кр}_{лев,\,\alpha/2})$

По значению критерия φ судят о «расхождении выборочных данных с гипотезой H_0». Естественно, что гипотеза H_0 должна быть отвергнута, если расхождения велики; именно этим объясняется включение в критическую область больших значений критерия φ (больше, чем критическая точка).

Включение же в ряде случаев в критическую область малых значений критерия φ (меньше, чем критическая точка) на первый взгляд противоречит смыслу этой величины. Однако не следует забывать, что φ – случайная величина (она зависит от результатов наблюдений X_1, X_2, …, X_n, которые случайны), поэтому маловероятно появление не только слишком больших, но и слишком малых её значений и их следует включить в критическую область.

Этап 5. В формулу критерия $\varphi = \varphi(X_1, X_2, \ldots, X_n)$ вместо X_1, X_2, …, X_n подставляют конкретные числа, полученные в результате наблюдений, и подсчитывают числовое значение $\varphi_{чис}$ критерия.

Если $\varphi_{чис}$ попадает в критическую область ω, то гипотеза H_0 отвергается и принимается гипотеза H_1. Поступая таким образом, следует понимать, что можно допустить ошибку с вероятностью α.

Если $\varphi_{чис}$ не попадает в критическую область, гипотеза H_0 не отвергается. Но это вовсе не означает, что H_0 является единственно подходящей гипотезой: просто расхождение между выборочными данными и гипотезой H_0 невелико, или, иначе, H_0 не противоречит результатами наблюдений; однако таким же свойством наряду с H_0 могут обладать и другие гипотезы.

3. Критерий Пирсона

Выше рассматривались гипотезы, относящиеся к отдельным параметрам распределения случайных величин, причём модели законов распределения этих величин представлялись известными. Однако во многих практических задачах модель закона распределения заранее не известна и возникает задача выбора модели, согласующейся с результатами наблюдений над случайной величиной. Пусть высказано предположение, что неизвестная функция распределения $F_X(x)$ исследуемой случайной величины X имеет вполне определённую модель $F_{\text{теор}}(x)$, т.е. высказана гипотеза

$$H_0 : F_X(x) = F_{\text{теор}}(x). \qquad (18)$$

В качестве теоретической модели $F_{\text{теор}}(x)$ может быть рассмотрена нормальная, биномиальная или какая-либо другая модель. Это определяется сущностью изучаемого явления, а также результатом предварительной обработки наблюдений над случайной величиной (формой графика вариационного ряда, соотношениями между выборочными характеристиками и т.д.). Критерии, с помощью которых проверяется гипотеза (19), называются **критериями согласия**. Рассмотрим лишь один из них, использующий χ^2-распределение и получивший название **критерия согласия Пирсона**.

Критерий предполагает, что результаты наблюдений сгруппированы в вариационный ряд.

Однако, прежде чем рассматривать сам критерий Пирсона, вспомним параметрическое оценивание закона распределения. Последовательность оценивания такая: формулируют гипотезу о модели закона распределения случайной величины; по результатам наблюдений находят оценки неизвестных параметров этой модели (допустим, что число неизвестных параметров равно l); вместо неизвестных параметров подставляют в модель найденные оценки. В результате предполагаемая модель закона оказывается полностью определённой и, используя её, рассчитывают вероятности $p_i^{\text{теор}} = P(X = x_i)$ того, что случайная величина X примет зафиксированные в наблюдениях значения x_i, i=1, 2, …, ν – 1; эти вероятности называют **теоретическими**. Обратим внимание на следующее обстоятельство: т.к. сумма вероятностей ряда распределения должна быть равна единице, т.е.

$$\sum_i p_i^{\text{òåîð}} = 1, \qquad (19)$$

то полагаем вероятность $p_\nu^{\text{теор}} = 1 - p_1^{\text{теор}} - p_2^{\text{теор}} - \ldots - p_{\nu-1}^{\text{теор}}$.

Обратим внимание на следующее: критерий согласия Пирсона можно использовать только в том случае, когда $m_i^{\text{теор}} \geq 5$, i=1, 2, …, ν. (20)

Поэтому ту группу вариационного ряда, для которой это условие не выполняется, объединяют с соседней и соответственно уменьшают число групп; так поступают до тех пор, пока для каждой новой группы $m_i^{\text{теор}}$ будет не меньше 5. Новое число групп, как и прежде, обозначим символом ν.

Оказывается, что если предполагаемая модель закона распределения действительно имеет место, т.е. верна гипотеза (18), и если к тому же выполняются условия (19) и (20), то величина

$$\varphi = \sum_{i=1}^{\nu} \frac{\left(m_i - m_i^{\text{теор}}\right)^2}{m_i^{\text{теор}}} \qquad (21)$$

будет иметь χ^2-распределение с числом степеней свободы $k = \nu - l - 1$, т.е.

$$\varphi = \sum_{i=1}^{\nu} \frac{\left(m_i - m_i^{\text{òåîð}}\right)^2}{m_i^{\text{òåîð}}} = \chi^2(k = \nu - l - 1),$$

где ν – число (новое) групп вариационного ряда; l – число неизвестных параметров предполагаемой модели, оцениваемых по результатам наблюдений (если все параметры предполагаемого закона известны точно, то l = 0). Величину (21) и называют **критерием согласия χ^2** или **критерием согласия Пирсона**.

Далее поступаем так же, как обычно при проверке гипотез. Задаёмся уровнем значимости α. Зная распределение критерия φ, находим критическую область, как правило, это область правосторонняя, т.е. она имеет вид $(x^{кр}_{пр,\,\alpha}, +\infty)$; найдём числовое значение $\varphi_{чис}$ критерия (21). Если $\varphi_{чис}$ попадает в интервал $(x^{кр}_{пр,\,\alpha}, +\infty)$, то делаем вывод о неправомерности гипотезы H_0 (18); при этом не следует забывать, что этот вывод может оказаться ошибочным (на самом деле в генеральной совокупности гипотеза H_0 (18) имеет место) и вероятность того, что вывод ошибочен, равна α.

Если $\varphi_{чис}$ не попадает в интервал $(x^{кр}_{пр,\,\alpha}, +\infty)$, то гипотеза H_0 (18) не отвергается.

В заключение приведём схему определения точки $x^{кр}_{пр,\,\alpha}$:

$$\left.\begin{array}{r}\alpha \to \gamma = 1-\alpha \\ l,\ \nu \to k = \nu - l - 1\end{array}\right\} \longrightarrow \chi^2_{\gamma} \to x^{кр}_{пр,\,\alpha} = \chi^2_{\gamma}. \qquad (22)$$

4.Выравнивание статистических рядов.

Дано случайное распределение успеваемости 100 студентов-заочников, сдававших четыре экзамена:

Число сданных экзаменов					
Число студентов				5	0

Здесь случайной величиной является число сданных экзаменов среди четырёх. Обозначим её X. Установим закон распределения этой величины.

Построим сначала его непараметрическую оценку. Величина X – дискретная. Дискретный вариационный ряд, заданный столбцами 2 и 4 табл. *, даёт непараметрическую оценку закона распределения числа сданных экзаменов среди четырёх сдаваемых.

Теперь сформулируем гипотезу о модели закона распределения случайной величины X – числе сданных экзаменов среди четырёх сдаваемых. Процесс сдачи четырёх экзаменов представим как четыре испытания, относительно которых сделаем следующие допущения:

Таблица *.

	Число сданных экзаменов x_i	Число студентов m_i	Частость p_i^*	$p_i^{теор} = \tilde{N}_4^{x_i} \cdot 0{,}88^{x_i} \cdot 0{,}12^{4-x_i}$	$m_i^{теор} = np_i^{теор}$	$(m_i - m_i^{теор})^2$	$(m_i - m_i^{теор})^2 : m_i^{теор}$
	2	3	4	5	6	7	8
	0	1	0,01	0,00021	0,021	5,382	0,735
	1	11	0,01	0,00608	0,608	5,239	0,160
	2	5	0,03	0,06691	6,691	0,001	0,000
	3	35	0,35	0,32711	32,711		
	4	60	0,60	0,59969	59,969		
Итого		n = 100	1,00	1,00000			0,895

- эти испытания независимы, т.е. вероятность сдачи любым студентом любого экзамена не зависит от того, будет сдано или нет любое количество других экзаменов;

- вероятность сдачи студентом любого отдельно взятого экзамена одна и та же и равна p, а вероятность «несдачи» равна $(1 - p)$.

Конечно, эти допущения могут вызывать некоторые сомнения, но возможно, что они не будут противоречить результатам наблюдений. При этих допущениях мы имеем дело с испытаниями Бернулли и число сданных экзаменов среди четырёх сдаваемых будет иметь биномиальный закон распределения, т.е. вероятность того, что студент сдаст λ экзаменов, равна

$P(X = x) = C_4^x p^x (1 - p)^{4-x}$, $x = 0, 1, 2, 3, 4$. Найдём оценку параметра p, входящего в модель (6). В условиях испытаний Бернулли состоятельной, несмещённой и эффективной оценкой вероятности является частость. В рассматриваемом примере p – вероятность того, что студент сдаст экзамен, поэтому частость p^* этого события, учитывая, что имеются сведения об успеваемости 100 студентов, вычисляем следующим образом:

$$p^* = \frac{\text{числоэкзаменов,сданных 100 студентами}}{\text{числоэкзаменов,сдаваемых 100 студентами}} = \frac{\sum_{i=1}^{5} x_i m_i}{4 \times 100} = \frac{0 \times 1 + 1 \times 1 + 2 \times 3 + 3 \times 35 + 4 \times 60}{100 \times 4} = 0{,}88.$$

Так как $\sum_{i=1}^{5} x_i m_i / 100 = \overline{X}$ - это среднее число экзаменов, сданных одним студентом, то p^* можно было бы определить и так:

$$p^* = \frac{\text{среднеечислоэкзаменов,сданныходнимстудентом}}{\text{числоэкзаменов,сдаваемыходнимстудентом}} = \frac{\overline{\tilde{O}}}{4} = 0{,}88.$$

Заметим, что если находить оценку параметра p в модели (6) методом максимального правдоподобия и при этом учесть, что число x_i наблюдалось m_i раз, то мы получили бы для p^* такую же формулу, а именно

$$p^*_{\text{мп}} = \sum_{i=1}^{5} x_i m_i / (4n).$$

Подставив в модель (6) вместо параметра p его оценку p^*, получим параметрическую оценку неизвестного закона распределения числа сданных экзаменов, построенную в предположении, что допустима биномиальная модель $P(X = x) = C_4^x 0{,}88^x 0{,}12^{4-x}$; $x = 0, 1, 2, 3, 4$. (**)

Теоретические вероятности $p_i^{\text{теор}}$ и частоты $m_i^{\text{теор}}$, вычисленные в предположении, что имеет место модель(**), содержатся в столбцах 5 и 6 табл. *. Поскольку различия между соответствующими числами столбцов 4 и 5 или между числами столбцов 3 и 6 небольшие, можно сделать предварительное заключение о приемлемости биномиальной модели. Графически это заключение подтверждается рисунком, на котором кривая вероятностей $p_i^{\text{теор}}$ близка к кривой частостей p_i^*.

Метод более глубокого обоснования приемлемости той или иной модели называется **критерием согласия**.

1.8 Лекция № 8, 9 (4 часа).

Тема: «Понятие функциональной, стохастической и корреляционной зависимости. Функция регрессии. Корреляционное отношение. Его свойства, значимость. Линейная функция регрессии. Коэффициент корреляции его»

1.8.1 Вопросы лекции:

1. Виды зависимостей между величинами.

2. Функция регрессии.

3. Корреляционное отношение, коэффициент детерминации. Корреляционная зависимость.

1.8.2 Краткое содержание вопросов:

1. Виды зависимостей между величинами

Условимся обозначать через X независимую переменную, а через Y – зависимую переменную.

Зависимость величины Y от X называется **функциональной**, если каждому значению величины X соответствует единственное значение величины Y. С функциональной зависимостью мы встречаемся, например, в математике, при изучении физических законов. Обратим внимание на то, что если X – детерминированная величина (т.е. принимающая вполне определённые значения), то и функционально зависящая от неё величина Y тоже является детерминированной; если же X – случайная величина, то и Y также случайная величина.

Однако гораздо чаще в окружающем нас мире имеет место не функциональная, а **стохастическая**, или **вероятностная**, **зависимость**, когда каждому фиксированному значению независимой переменной X соответствует не одно, а множество значений переменной Y, причём сказать заранее, какое именно значение примет величина Y, нельзя. Более частое появление такой зависимости объясняется действием на результирующую переменную не только контролируемого или контролируемых факторов (в данном случае таким контролируемым фактором является переменная X), а и многочисленных неконтролируемых случайный факторов. В этой ситуации переменная Y является случайной величиной. Переменная же X может быть, как детерминированной, так и случайной величиной. Следует заметить, что со стохастической зависимостью мы уже сталкивались в дисперсионном анализе.

Допустим, что существует стохастическая зависимость случайной переменной Y от X. Зафиксируем некоторое значение x переменной X. При $X = x$ переменная Y в силу её стохастической зависимости от X может принять любое значение из некоторого множества, причём какое именно – заранее неизвестно. Среднее этого множества называют **групповым генеральным средним** переменной Y при $X = x$ или **математическим ожиданием случайной величины Y, вычисленным при условии, что X = x; это условное математическое ожидание обозначают так: M(Y/X = x)**. Если существует стохастическая зависимость Y от X, то прежде всего стараются выяснить, изменяются или нет при изменении x условные математические ожидания $M(Y/X=x)$. Если при изменении x условные математические ожидания $M(Y/X=x)$ изменяются, то говорят, что имеет место **корреляционная зависимость** величины Y от X; если же условные математические ожидания остаются неизменными, то говорят, что корреляционная зависимость величины Y от X отсутствует.

Функция $\varphi(x)=M(Y/X=x)$, описывающая изменение условного математического ожидания случайной переменной Y при изменении значений x переменной X, называется **функцией регрессии**.

Выясним, почему именно при наличии стохастической зависимости интересуются поведением условного математического ожидания.

Рассмотрим пример. Пусть X – уровень квалификации рабочего, Y – его выработка за смену. Ясно, что зависимость Y от X не функциональная, а стохастическая: на выработку помимо квалификации влияет множество других факторов. Зафиксируем значение x уровня квалификации: ему соответствует некоторое множество значений выработки Y. Тогда $M(Y/X = x)$ – средняя выработка рабочего при условии, что его уровень квалификации равен x, или, иначе говоря, $M(Y/X = x)$ – это норматив выработки при уровне квалификации, равном x. Зная зависимость этого норматива от уровня квалификации, можно

для любого уровня квалификации рассчитать норматив выработки и, сравнив его с реальной выработкой, оценить работу рабочего.

2. Функция регрессии

Обратим внимание на то, что введённые понятия стохастической и корреляционной зависимости относились к генеральной совокупности. Поясним эти понятия числовым примером.

Пример. Допустим, что одновременно изучаются две случайные величины X и Y, или, иначе говоря, двумерная случайная величина (X, Y), которая задана таблицей

	i			
	x_i / y_i	$_1 = 2$	$_2 = 5$	$_3 = 8$
	$y_1 = 0,4$	,15	,12	,03
	$y_2 = 0,8$	,05	,30	,35

Таблицу эту называют **таблицей распределения двумерной величины (X, Y)**; её следует понимать так. Случайная величина X может принять одно из следующих значений: 2, 5 и 8. Случайная величина Y – значения 0,4 и 0,8. Число 0,15 – это вероятность того, что X = 2 и одновременно Y = 0,4, или, иначе говоря, вероятность произведения двух событий; события, состоящего в том, что X = 2, и события, состоящего в том, что Y = 0,4, т.е. $P((X=2)(Y=0,4)) = 0,15$. Аналогично, вероятность $P((X=2)(Y=0,8)) = 0,05$ и т.д. Обратим внимание на следующее: поскольку в табл. 9 указаны все возможные значения величин X и Y, сумма вероятностей, стоящих в таблице, должна быть равна единице: 0,15 + 0,05 + 0,12 + 0,30 + 0,03 + 0,35 = 1.

Прежде чем выяснить тип зависимости величины Y от X, найдём:

а) Закон распределения величины X. Он представлен таблице

x	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$	
P $(X = x)$	0,15 + 0,05 = 0,2	0,12 + 0,30 = 0,42	0,35 + 0,03 = 0,38	= 1

M(X) = 5,54, D(X) = 4,9284

Действительно, например, величина X примет значение, равное 2, только в том случае, когда одновременно с этим величина Y примет значение 0,4 или 0,8, т.е.

$P(X = 2) = P((X = 2)(Y = 0,4)) + P((X = 2)(Y = 0,08)) = 0,15 + 0,05 = 0,2$.

Справа от ряда распределения величины X находятся её математическое ожидание и дисперсия.

б) Закон распределения величины Y. Он имеет вид таблицы

y	$y_1 = 0,4$	$y_2 = 0,8$	
P $(Y = y)$	0,15 + 0,12 + 0,03 = 0,30	0,05 + 0,30 + 0,35 = 0,7	= 1

M(Y) = 0,68, D(Y) = 0,0336

в) Условные законы распределения величины Y, а именно закон распределения величины Y сначала при условии, что X = 2, затем при условии, что X=5, и наконец, при условии, что X = 8.

Итак, пусть X = 2. Тогда условная вероятность

$$P(\mathrm{Y} = 0{,}4/\mathrm{X} = 2) = \frac{P(Y = 0{,}4)(X = 2))}{P(X = 2)} = \frac{0{,}15}{0{,}2} = 0{,}75,$$

а условная вероятность

$$P(\mathrm{Y} = 0{,}8/\mathrm{X} = 2) = \frac{P((Y = 0{,}8)(X = 2))}{P(X = 2)} = \frac{0{,}05}{0{,}2} = 0{,}25.$$

Таким образом, закон распределения величины Y при условии, что X = 2, задан таблицей

y	$y_1 = 0{,}4$	$y_2 = 0{,}8$	
$P(Y = y/X = 2)$	0,75	0,25	= 1

M(Y/X = 2) = 0,4*0,75 + 0,8*0,25 = 0,5, D(Y/X = 2) = 0,03

Справа помещено условное математическое ожидание и значение условной дисперсии. Покажем, как вычисляется условная дисперсия. Общая формула условной дисперсии имеет вид

$$D(Y/X = x) = M[(Y/X = x) - M(Y/X = x)]^2. \qquad (23)$$

$\mathrm{D(Y/X = 2)} = \mathrm{M}\,[(\mathrm{Y/X = 2}) - \mathrm{M(Y/X = 2)}]^2 = \mathrm{M}\,[(\mathrm{Y/X = 2}) - 0{,}5]^2 =$ $=$

$\sum_{i=1}^{2} (y_i - 0{,}5)^2 \cdot \mathrm{P(Y = y_i/X = 2)} = (0{,}4 - 0{,}5)^2 \cdot 0{,}75 + (0{,}8 - 0{,}5)^2 \cdot 0{,}25 = 0{,}03.$

Пусть X = 5. Тогда $\mathrm{P(Y = 0{,}4/X = 5)} = \frac{P((Y = 0{,}4)(X = 5))}{P(X = 5)} = \frac{0{,}12}{0{,}42} = \frac{2}{7}$; $\mathrm{P(Y=0{,}8/X=5)} = \frac{P((Y = 0{,}8)(X = 5))}{P(X = 5)} = \frac{0{,}30}{0{,}42} = \frac{5}{7}$.

Таким образом, закон распределения величины Y при условии, что X = 5, имеет вид таблицы

y	,4	,8	
$P(Y = y/\mathrm{X} = 5)$	/7	/7	= 1

$\mathrm{M(Y/X = 5)} = \frac{24}{35} \approx 0{,}686$, D(Y/X = 5) = 0,03265.

И наконец, при X = 8 ряд распределения задан таблицей.

y	,4	,8	
$P(\mathrm{Y} = y/\mathrm{X} = 8)$	$\frac{3}{38}$	$\frac{35}{38}$	= 1

$\mathrm{M(Y/X = 8)} = \frac{73}{95} \approx 0{,}768$, D(Y/X = 8) = 0,01163

Из таблиц видно, что зависимость Y от X стохастическая, поскольку при каждом фиксированном значении величины X величина Y может быть равной либо 0,4, либо 0,8, причём какому именно из этих чисел она будет равна – сказать заранее нельзя. Ясно прослеживается и корреляционная зависимость величины Y от X, поскольку с изменением

значений x величины X меняются и условные математические ожидания $M(Y/X = x)$. Функция регрессии, т.е. зависимость условного математического ожидания M(Y/X = x) от x, задаётся в виде таблицы

x		5	8
$M(Y/X = x)$	,5	24/35 ≈ 0,686	73/95 ≈ 0,768

3. Корреляционное отношение, коэффициент детерминации. Корреляционная зависимость

Выясним, можно ли измерить степень корреляционной и стохастической зависимости величины Y от X. Ответ проиллюстрируем. Все полученные в примере результаты объединены в таблице.

x_i	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$
$P(X = x_i)$	,2	0,42	0,38
$M(Y/X = x_i)$	,5	0,686	0,768
$D(Y/X = x_i)$	,03	0,03265	0,01163

MY = 0,68, DY = 0,0336

Т.к. X – случайная величина, принимающая значения 2, 5 и 8 с вероятностью 0,2; 0,42 и 0,38, то такими же будут вероятности и условных математических ожиданий, и дисперсий. Т.обр., условное математическое ожидание M(Y/X), так же как и условная дисперсия D(Y/X) – случайные величины.

Обратим также внимание на то, что M(Y), можно вычислить и следующим образом:

$$M(Y) = M[M(Y/X)] = \sum_{i=1}^{3} M(Y/X = x_i)P(X = x_i) = 0,5*0,2 + 0,686*0,42 + 0,768*0,38 = 0,68.$$

Разброс значений величины Y вокруг математического ожидания MY измеряется дисперсией D(Y), или σ_Y^2:

$$\sigma_Y^2 = D(Y) = M(Y - MY)^2.$$

Этот разброс может быть вызван:

- зависимостью величины Y от X (эта зависимость может быть обусловлена не только непосредственным влиянием X на Y, но и наличием случайных факторов, действующих на Y через переменную X);
- зависимостью величины Y от случайных факторов, влияющих только на Y и не влияющих на X; эти факторы называют **остаточными**.

1) Построим показатель разброса значений величины Y, связанного с её зависимостью от фактора X.

Условное математическое ожидание M(Y/X = x) является «представителем игреков», которые имеют место при X = x. Характеристикой разброса условных математических ожиданий M(Y/X = x) относительно M(Y) является дисперсия D[M(Y/X)], или

$$\sigma_\varphi^2 = D[M(Y/X)] = M[M(Y/X) - MY]^2$$

– эта величина и будет показателем разброса значений величины Y, связанного с её зависимостью от фактора X. Найдём:

$$\sigma_\varphi^2 = M[M(Y/X) - MY]^2 = (0,5 - 0,68)^2*0,2 + (0,686 - 0,68)^2*0,42 + (0,768 - 0,68)^2*0,38 = 0,0095.)$$

2) Теперь построим показатель разброса «игреков», связанного с влиянием остаточных факторов.

Зафиксируем какое-либо значение x величины X. Тогда причиной вариации величины Y при X = x будут остаточные факторы, влияющие только на Y и не влияющие на X. Измерителем этой вариации является условная дисперсия D(Y/X = x).При различных же «иксах» характеристикой разброса «игреков», вызванного влиянием на Y остаточных факторов, будет генеральное среднее из условных дисперсий, или, иначе, математическое ожидание условной дисперсии. Эту величину обозначим σ_0^2. Имеем

$$\sigma_0^2 = M[D(Y/X)],$$

где при X = x

$$\sigma_0^2 = M[D(Y/X)] = \sum_{i=1}^{3} D(Y/X = x_i)P(X = x_i) =$$

$$= 0{,}03*0{,}2 + 0{,}03265*0{,}42 + 0{,}01163*0{,}38 = 0{,}0241.)$$

Для вычисленных дисперсий справедливо тождество

$$DY = D[M(Y/X)] + M[D(Y/X)]$$

или

$$\sigma_Y^2 = \sigma_\varphi^2 + \sigma_0^2.$$

Степень стохастической зависимости величины Y от X измеряется **гене-ральным корреляционным отношением**

$$\rho_{Y/X} = +\sqrt{\frac{D[M(Y/X)]}{DY}} = +\sqrt{\frac{\sigma_\varphi^2}{\sigma_Y^2}} \underset{(28)}{=} +\sqrt{1-\frac{\sigma_0^2}{\sigma_Y^2}} = +\sqrt{1-\frac{M[D(Y/X)]}{DY}}\ .$$

Квадрат корреляционного отношения

$$\rho_{Y/X}^2 = \frac{\sigma_\varphi^2}{\sigma_Y^2} = \frac{D[M(Y/X)]}{DY} \underset{(26),\,(25)}{=} \frac{M[M\left(\frac{Y}{X}\right)-MY]^2}{M(Y-MY)^2}$$

называется **генеральным коэффициентом детерминации**; он показывает, какую долю дисперсии величины Y составляет дисперсия условных математических ожиданий, или, иначе говоря, какая доля дисперсии D(Y) объясняется корреляционной зависимостью Y от X.

Свойства генерального корреляционного отношения как измерителя степени корреляционной и стохастической зависимости

1. $\Theta \le \rho_{Y/X} \le 1$.

Действительно, $\rho_{Y/X} \ge 0$; с другой стороны $\sigma_\varphi^2 \le \sigma_Y^2$, поэтому $\rho_{Y/X} \le 1$.

2. Условие $\rho_{Y/X} = 0$ является необходимым и достаточным для отсутствия корреляционной зависимости Y от X, т.е. для того, чтобы M(Y/X) = const при любом значении x величины X.

Следствие. Чем ближе $\rho_{Y/X}$к нулю, тем ближе к нулю D[M(Y/X)], а это означает, что уменьшается разброс условных математических ожиданий M(Y/X = x) относительно MY. Т.обр., чем ближе $\rho_{Y/X}$ к нулю, тем меньше «реакция условного математического ожидания M(Y/X = x) на изменение x», или, иначе говоря, «тем меньше степень корреляционной зависимости Y от X».

И, наоборот, чем «меньше степень корреляционной зависимости Y от X», тем ближе $\rho_{Y/X}$ к нулю.

3. Условие $\rho_{Y/X} = 1$ является необходимым и достаточным для функциональной зависимости величины Y от X.

Достаточность. Пусть $\rho_{Y/X} = 1$. Это в силу означает, что $\sigma_0^2 = 0$ или M[D(Y/X)]=0. Но т.к. дисперсия другой величины неотрицательна, то из последнего равенства следует, что D[M(Y/X = x)] = 0 при любом x, а это означает, что при X = x величина Y остаётся постоянной (принимает единственное значение), т.е. зависимость Y от X – функциональная.

Необходимость. Пусть любому фиксированному значению x величины X соответствует только одно значение величины Y. Это означает, что при любом x дисперсия $D[M(Y/X = x)] = 0$, поэтому и $\sigma_0^2 = M[D(Y/X)] = M(0) = 0$. Но тогда из (28) следует, что $\rho_{Y/X} = 1$.

Следствие. Чем ближе $\rho_{Y/X}$ к единице, тем ближе к нулю $M[D(Y/X)]$, а следовательно, и условные дисперсии $D(Y/X = x)$. Это означает, что при каждом допустимом значении x уменьшается разброс «игреков» относительно $M(Y/X = x)$. Т.обр., чем ближе $\rho_{Y/X}$ к единице, тем меньше при каждом x отличие «игреков» от постоянного числа, равного $M(Y/X = x)$, или, иначе говоря, тем выше степень стохастической зависимости Y от X. И, наоборот, чем выше степень стохастической зависимости Y от X, тем ближе $\rho_{Y/X}$ к единице.

В практических задачах наибольший интерес представляют следующие вопросы:

- существует корреляционная зависимость Y от X или нет, иначе говоря, отлично ли генеральное корреляционное отношение $\rho_{Y/X}$ от нуля или равно нулю;
- если корреляционная зависимость существует, то какой вид имеет функция регрессии (линейный, параболический или какой-либо другой).

Точно ответить на поставленные вопросы можно лишь только в том случае, когда известен закон распределения двумерной величины (X, Y).

Линейная функция регрессии. Генеральный коэффициент корреляции

Допустим, что при изменении x условное математическое ожидание $M(Y/X=x)$ изменяется по линейному закону, т. е. функция регрессии $\varphi(x) = M(Y/X=x)$ линейная:

$$M^{лин}(Y/X = x) = a + bx$$

Найдем для этого случая сначала выражение для параметров a и b линейной функции регрессии, а затем выражение для корреляционного отношения. При этом договоримся используемые обозначения снабжать индексом «лин», что означает «при условии линейной функции регрессии».

Выражения для параметров *a* и *b* и линейной функции регрессии

Обратимся к формуле условной дисперсии. В случае линейной функции регрессии формула принимает вид

$$D^{лин}\left(\frac{Y}{X} = x\right) = M\left[\left(\frac{Y}{X} = x\right) - M^{лин}\left(\frac{Y}{X} = x\right)\right]^2 =$$

$$= M[(Y/X = x) - a - bx]^2$$

Напомним, в общем случае при изменении *x* условная дисперсия $D^{лин}(Y/X=x)$ изменяется. Найдем характеристику разброса «игреков», вызванного влиянием на Yостаточных факторов, она примет вид

$\sigma_0^{2\,лин} = M[D^{лин}(Y/X)] = M(M(Y/X) - a - bX)^2$

— эта величина при фиксированных значениях параметров a и bявляется постоянной.

Принимая во внимание свойство минимальности дисперсии, значения параметре a и bнаходят из условия

$F(a, b) = M(M(Y/X) - a - bX)^2 \to \min$

Окончательный результат такой:

$b = r_{XY}\sigma_Y/\sigma_X$, (50) $a = m_Y - r_{XY}\frac{\sigma_Y}{\sigma_X}m_x$,

гдеm_X=M(X), m_y=M(Y), $r_{XY} = \frac{K_{XY}}{\sigma_X\sigma_Y} = \frac{M[(X-MX)(Y-MY)]}{\sigma_X\sigma_Y}$

Отметим, что выражение $K_{XY} = M[(X - MX)(Y - MY)]$

называют **генеральным корреляционным моментом**, а

$r_{XY} = \frac{K_{XY}}{\sigma_X\sigma_Y} = \frac{M[(X-MX)(Y-MY)]}{\sigma_X\sigma_Y}$

генеральным коэффициентом корреляции.

Подставив найденные значения параметров a и bв линейную функцию регрессии $M^{лин}$ (Y/X=x) = a+bx, получим

$$M^{лин}(Y/X = x) = m_y + r_{XY}\frac{\sigma_Y}{\sigma_X}(x - m_x)$$

Выражение для корреляционного отношения

Напомним вид генерального корреляционного отношения

$$\rho_{Y/X} = +\sqrt{\sigma_ф^2/\sigma_Y^2} = \sqrt{1 - \sigma_o^2/\sigma_Y^2}$$

Установим, как преобразуются выражения для σ_o^2, $\sigma_ф^2$ и $p_{y/x}$, если учесть, что функция регрессии линейна. Согласно формуле (25)

$$\sigma_ф^2 = D[M(Y/X)],$$

имеем:

$$\sigma_ф^{2\,лин} = D[M^{лин}(Y/X)] = D\left[m_Y + r_{XY}\frac{\sigma_Y}{\sigma_X}(X - m_x)\right] = D\left[r_{XY}\frac{\sigma_Y}{\sigma_X}(X - m_x)\right] = r_{XY}^2\frac{\sigma_Y^2}{\sigma_X^2}D(X -$$

mx=rXY2σY2σX2DX=rXY2σY2.

Итак,

$$\sigma_ф^{2\,лин} = D[M^{лин}(Y/X)] = r_{XY}^2\sigma_Y^2$$

Теперь, используя соотношение (27) , определим вид выражения $\sigma^2{}_o^{лин}$=M[$D^{лин}$(Y/X)] при линейной функции регрессии.

Имеем

$$\sigma_o^{2\,лин} = \sigma_Y^2 - \sigma_ф^{2\,лин} = \sigma_Y^2 - r_{XY}^2\sigma_Y^2 = \sigma_Y^2(1 - r_{XY}^2),$$

тогда

$$\sigma_o^{2\,лин} = M[D^{лин}(Y/X)] = (1 - r_{XY}^2)\sigma_Y^2.$$

Для генерального корреляционного отношения получим следующее выражение:

$$\rho_{Y/X}^{лин} = +\sqrt{\frac{\sigma_ф^{2\,лин}}{\sigma_Y^2}} = +\sqrt{\frac{r_{XY}^2\sigma_Y^2}{\sigma_Y^2}} = |r_{XY}|$$

где r_{XY}—коэффициент корреляции.

Свойства коэффициента корреляция как измерителя степени линейности стохастической зависимости

Итак, генеральный коэффициент корреляции

$$r_{XY} = \frac{M[(X - MX)(Y - MY)]}{\sigma_X\sigma_Y}$$

при этом если функция регрессии линейна и имеет вид, то | r_{xy}| совпадает с генеральным корреляционным отношением $P_{Y/X}$. Таким образом, | r_{xy}| - это частный случай $P_{Y/X}$, имеющий место при линейной функции регрессии, поэтому

$$|r_{XY}| \le \rho_{Y/X}$$

и | r_{xy}| будет обладать свойствами корреляционного отношения, «подправленными» с учетом линейности функции регрессии. Сформулируем эти свойства.

1. $|r_{XY}| \le 1$ или $-1 < r_{XY} < 1$.

В отличие от корреляционного отношения, которое не может быть отрицательным, r_{xy}может иметь как знак «+», так и «—»; как видно из, знак « + » означает, что с ростом значений Хувеличивается условное математическое ожидание M(Y/X=x); отрицательное значение r_{xy}говорит о противоположной тенденции.

2. Если корреляционная зависимость Y от X отсутствует, то $r_{xy} = 0$.

Действительно, при отсутствии корреляционной зависимости Y от Хкоэффициент $\rho_{y/x}$=0, поэтому в силу и r_{xy}=0.

Обратное же утверждение верно не всегда: из равенства $r_{xy} = 0$ не всегда следует, что

$\rho_{y/x}=0$, или, иначе говоря, не всегда следует отсутствие корреляционной зависимости. Только в том случае, когда функция регрессии линейна и имеет вид, из равенства $r_{xy}=0$следует отсутствие корреляционной зависимости Y от X. Действительно, подставив $r_{xy} = 0$ в, получим, что $M(Y/X=x)=m_y=const$ при любом x.

3. Условие $|r_{xy}| = 1$ является необходимым и достаточным для существования линейной функциональной зависимости между Y и X.

Коэффициент корреляции симметричен относительно X и Y, т.е. $r_{xy} = r_{yx}$. Если $|r_{xy}| = 1$, то $|r_{yx}| = 1$, поэтому вместо выражения «линейная функциональная зависимость Yот X» мы употребим «линейная функциональная зависимость между Yи X».

Чем ближе $|r_{xy}|$ к единице, тем ближе стохастическая зависимость между величинами X и Y к линейной функциональной, или, иначе говоря, выше степень линейности стохастической зависимости. И, наоборот, чем выше степень линейности стохастической зависимости между величинами X и Y, тем ближе $|r_{xy}|$ к единице.

Метод наименьших квадратов. Линейное уравнение регрессии

Пусть функция регрессии линейная, т. е. M(Y/X)=x = a + bx. Найдем оценки $\hat{a}$ и$\hat{b}$параметров a и b.Критерием нахождения оценок $\hat{a}$ и$\hat{b}$является следующее требование: средняя квадратов отклонений наблюдаемых «игреков» от «игреков», рассчитанных по уравнению Y=a+bx, должна быть минимальной. Запишем это требование в виде формулы

$$\hat{F}(\hat{a},\hat{b}) = \overline{(Y-\hat{Y})^2} \longrightarrow min$$

Метод нахождения значений оценок a и b (в соответствии с требованием962)) называется **методом наименьших квадратов**.

Для результатов $(X_i, У_i)$ наблюдений величины (X,Y) не сгруппированных в корреляционную таблицу, критерий имеет вид

$$\hat{F}(\hat{a},\hat{b}) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}X_i)^2 \longrightarrow \min$$

Если наблюдения сгруппированы в таблицу, то критерий принимает следующий вид:

$$\hat{F}(\hat{a},\hat{b}) = \frac{1}{n}\sum_{i=1}^{v}\sum_{j=1}^{q}(y_i - \hat{a} - \hat{b}x_i)^2 m_{ji} \longrightarrow min$$

Необходимые условия минимума функции F(a, b)образуют систему

$$\begin{cases} \dfrac{\partial F}{\partial a} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} 2\left(Y_i - \hat{a} - \hat{b}X_i\right)(-1) = 0 \\ \dfrac{\partial F}{\partial a} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} 2\left(Y_i - \hat{a} - \hat{b}X_i\right)(-X_i) = 0 \end{cases}$$

которая в результате тождественных преобразований принимает вид

$$\begin{cases} \hat{a} + \hat{b}\overline{X} = \overline{Y} \\ \hat{a}\overline{X} + \hat{b}\overline{X^2} = \overline{YX} \end{cases},$$

где

$$\overline{X} = \sum_{i=1}^{n}\frac{X_i}{n}, \overline{Y} = \sum_{i=1}^{n}\frac{Y_i}{n}, \overline{X^2} = \sum_{i=1}^{n}\frac{X^2{}_i}{n}, \overline{YX} = \sum_{i=1}^{n} Y_i X_i/n$$

Система называется **нормальной системой уравнений**. Решим ее относительно $\hat{a}$ и$\hat{b}$. Из первого уравнения находим $\hat{a} = Y — \hat{b}X$. Подставив это выражение во второе из уравнений, получим

$$\left(\overline{Y} - \hat{b}\overline{X}\right)\overline{X} + \hat{b}\overline{X^2} = \overline{YX},$$

откуда находим

$$\hat{b} = \frac{\overline{YX} - \overline{YX}}{\overline{X^2} - (\overline{X})^2}$$

или, учитывая (60),

$$\hat{b} = \hat{r}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}.$$

Тогда

$$\hat{a} = \overline{Y} - \hat{r}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\overline{X}$$

Подставив выражения для$\hat{a}$и $\hat{b}$в уравнение $Y=\hat{a}+\hat{b}x$, получим

$$\hat{Y} = \overline{Y} + \hat{r}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}(x - \overline{X}).$$

Уравнение называется **выборочным линейным уравнением регрессии.**
Пусть $x=x_i$,тогда

$$\hat{Y}_i = \overline{Y} + \hat{r}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}(x_i - \overline{X})$$

— это оценка условного математического ожидания, вычисляемого по формуле
$M^{лин}\left(\frac{Y}{X} = x_i\right) = M(Y) + \hat{r}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}(x_i - M(X))$.

2 МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1 Практическое занятие № 1 (2 часа).

Тема: «Классическое определение вероятности события. Геометрические вероятности. Относительная частота наступления события и статистическая вероятность. Формулы умножения и сложения вероятностей случайных событий»

2.1.1 Задание для работы:

1. Элементы комбинаторики

2. Непосредственное вычисление вероятности случайного события.

3. Операции над случайными событиями и их свойства. Теоремы о вероятности суммы случайных событий. Теоремы о вероятности суммы произведения

4. Условная вероятность. Формула полной вероятности. Формула Байеса.

5. Схема повторных испытаний. Формула Бернулли. Формула Пуассона. Локальная формулы Лапласа. Интегральная формула Лапласа.

6. Простейший поток событий. Вероятность случайного события с заданной интенсивностью.

2.1.2 Краткое описание проводимого занятия:

1. Элементы комбинаторики

Комбинаторика изучает способы подсчета числа элементов в конечных множествах. Формулы комбинаторики используются при непосредственном вычислении вероятностей. Приведем некоторые сведения.

Соединениями называют различные группы предметов, составленные из каких-либо объ-

ектов.

Элементами называются объекты, из которых составлены соединения. Рассмотрим следующие три вида соединений: перестановки, размещения и сочетания.

Перестановками из n элементов называют *соединения,* содержащие все *n* элементов и отличающиеся между собой лишь порядком элементов. Число перестановок из *n* элементов находится по формуле $P_n = n!$,

где *n!* - произведение натуральных чисел от 1 до *n* включительно, т.е. n!=1·2·3·…·n. Например, $P_6 = 6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$.

Размещениями из n элементов по k в каждом (n ≥ k) называются *такие соединения,* в каждый из которых входит *k* элементов, взятых из данных *n* элементов, и отличающихся друг от друга либо самими элементами, либо порядком их расположения.

Число размещений из *n* элементов по *k* находят по формуле

$A_n^k = n(n-1)(n-2)...(n-k+1)$ или, $A_n^k = \dfrac{n!}{(n-k)!}$

Например, $A_6^4 = 6 \cdot 5 \cdot 4 \cdot 3 = \dfrac{6!}{(6-4)!} = \dfrac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2} = 360$

Сочетаниями из n элементов по k (n>k) называют *соединения,* в каждый из которых входит *k* элементов, взятых из данных *n* элементов и отличающихся друг от друга, по крайней мере, одним элементом. Число сочетаний из *n* элементов по *k* находят по формуле:

$C_n^k = \dfrac{A_n^k}{P_k}$ или $C_n^k = \dfrac{n!}{(n-k)!k!}$.

Для упрощения вычислений при $k > \dfrac{1}{2}n$ полезно использовать следующее свойство сочетаний: $C_n^k = C_n^{n-k}$.

Замечания:

1) по определению $C_n^0 = 1$;

2) для определения числа сочетаний справедливы равенства

$C_n^m = C_n^{n-m}$, $C_{n+1}^{m+1} = C_n^m + C_n^{m+1}$, $C_n^0 + C_n^1 + ... + C_n^n = 2^n$

3) В записанных выше формулах комбинаторики предполагалось, что все *n* элементов различны. Если же некоторые элементы в соединениях повторяются, то в этом случае соединения с повторениями вычисляются по другим формулам.

Пусть среди *n* элементов рассматриваемого множества есть n_1 элементов одного вида, n_2 элементов другого вида и т.д. Число перестановок с повторениями определяется по формуле $P_n(n_1, n_2, ..., n_k) = \dfrac{n!}{n_1! \cdot n_2! \cdot ... \cdot n_k!}$,

где $n_1 + n_2 + ... + n_k = n$.

Число размещений по *m* элементов с повторениями из *n* элементов равно n^m, т.е. $\left(A_n^m\right)_{\text{с повт.}} = n^m$.

Число сочетаний с повторениями из *n* элементов по *m* элементов равно числу сочетаний без повторений из (*n+m-1*) элементов по *m,* т.е.

$\left(C_n^m\right)_{\text{с повт.}} = C_{n+m-1}^m$

4) При решении задач комбинаторики можно использовать следующие правила:

правило суммы. Если некоторый объект *A* может быть выбран из множества объектов *m* способами, а объект *B* может быть выбран *n* способами, то выбрать либо *A,* либо *B* можно *(m + n)* способами.

правило произведения. Если объект *A* можно выбрать из множества объектов *m* способами и после каждого такого выбора объект *B* можно выбрать *n* способами, то пара объектов *(A, B)* в указанном порядке может быть выбрана $m \cdot n$ способами.

2. Непосредственное вычисление вероятности случайного события

Пример 1. В урне 10 одинаковых по размерам и весу шаров, из которых 4 красных и 6 голубых. Из урны извлекается один шар. Какова вероятность того, что извлеченный шар окажется голубым?

Решение. Событие, состоящее в том, что «извлеченный шар оказался голубым», обозначим буквой А. Данное испытание имеет 10 равновозможных элементарных исходов, из которых 6 благоприятствуют появлению события А. По формуле классической вероятности события получим:

$$P(A) = \frac{6}{10} = 0{,}6\,.$$

Пример 2. Среди 25 студентов группы, в которой 10 девушек, разыгрывается 5 билетов лотереи. Найти вероятность того, что среди обладателей билетов окажутся две девушки.

Решение. Пусть А - событие, состоящее в том, что среди обладателей билетов окажутся две девушки. Найдем числа m, n.

Число всех равновозможных случаев распределения 5 билетов среди 25 студентов равно числу сочетаний из 25 элементов по 5, т.е. $C_{15}^5 \cdot$ Число групп по трое юношей из 15, которые могут получить билеты, равно C_{15}^3. Каждая такая тройка может сочетаться с любой парой из десяти девушек, а число таких пар равно C_{10}^2. Следовательно, число групп по 5 студентов, образованных из групп в 25 студентов, в каждую из которых будут входить трое

юношей и две девушки, равно произведению $C_{15}^3 \cdot C_{10}^2$. Это произведение равно числу благоприятствующих случаев распределения пяти билетов среди студентов группы так, чтобы три билета получили юноши и два билета - девушки.

В соответствии с формулой $P(A) = \frac{m}{n}$ находим искомую вероятность

$$P(A) = \frac{C_{15}^3 \cdot C_{10}^2}{C_{25}^5} = \frac{15!}{3!\cdot 12!} \cdot \frac{10!}{2!\cdot 8!} : \frac{25!}{5!\cdot 20!} = \frac{20!\cdot 15!\cdot 10!\cdot 5!}{25!\cdot 12!\cdot 8!\cdot 3!\cdot 2!} = \frac{13\cdot 14\cdot 15\cdot 9\cdot 10\cdot 4\cdot 5}{25\cdot 24\cdot 23\cdot 22\cdot 21\cdot 2} = \frac{13\cdot 5\cdot 3}{23\cdot 22} = \frac{195}{506} \approx 0{,}38$$

Пример 3. В круг вписан квадрат (рис.2). В круг наудачу бросается точка. Какова вероятность того, что эта точка попадет в квадрат?

Решение. Введем обозначения: *R* - радиус круга, *a* - сторона вписанного квадрата, *A* - событие, состоящее в том, что точка попала в квадрат, *S* - площадь круга, S_1 - площадь вписанного квадрата. Известно, что площадь круга $S = \pi R^2$. Сторона вписанного квадрата через радиус описанной окружности выражается

формулой $a = \sqrt{2}R$, поэтому площадь квадрата $S_1 = 2R^2$

Полагая в формуле $P(A) = \frac{S_g}{S_G}$ $S_g = S_1$, $S_G = S$, находим искомую вероятность $P(A) = \frac{2R^2}{\pi R^2} = \frac{2}{\pi} \approx 0{,}637$.

Рис. 2.

Замечание. Выражение стороны квадрата через радиус окружности можно получить следующим образом. Из треугольника ΔKMN по теореме Пифагора будем иметь: $KN^2 + NM^2 = KM^2$, т.е.

$a^2 + a^2 = (2R)^2$, $2a^2 = 4R^2$, $a^2 = 2R^2$, $a = \sqrt{2}R$.

3. Операции над случайными событиями и их свойства. Теоремы о вероятности суммы случайных событий. Теоремы о вероятности суммы произведения

Пример 1. Подбрасываются два игральных кубика. Найти вероятность события A, состоящего в том, что - «сумма выпавших очков не превосходит четырех».

Решение. Событие A - событие, состоящее в том, что есть сумма трех несовместных событий B_2, B_3, B_4. Тогда сумма очков равна соответственно 2, 3, 4. Поскольку $P(B_2) = \frac{1}{36}$, $P(B_3) = \frac{2}{36}$, $P(B_4) = \frac{3}{36}$, по теореме сложения вероятностей несовместных событий получим

$$P(A) = P(B_2) + P(B_3) + P(B_4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}.$$

Замечание. Этот же результат можно было получить, используя непосредственный подсчет вероятности. Действительно, событию A благоприятствуют 6 элементарных исходов: (1,1), (1,2), (2,1), (1,3), (3,1), (2,2). Всего же элементарных исходов, образующих полную группу событий, n= 36, поэтому $P(A) = \frac{6}{36} = \frac{1}{6}$.

Пример 2. Три станка работают независимо. Вероятность того, что в течение смены станок (любой) потребует наладки равна 0,1. Найти вероятность того, что в течение смены хотя бы один станок из трех потребует внимания наладчика.

Решение. Пусть A_k - событие, заключающееся в том, что k- тый по счету станок потребует наладки в течение смены (k = 1, 2,3). Тогда событие $A_1 + A_2 + A_3$ заключается в том, что в течение смены наладки потребует хотя бы один из трех станков. Сначала вычислим вероятность противоположного события $\overline{A_1 + A_2 + A_3}$, заключающегося в том, что все три станка всю смену проработают безотказно. Поскольку $\overline{A_1 + A_2 + A_3} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$, причем события $\overline{A_1}, \overline{A_2}, \overline{A_3}$ независимы, то $P(\overline{A_1 + A_2 + A_3}) = P(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3})$ по теореме умножения вероятностей для независимых событий. По условию $P(A_k)$=0,1,тогда вероятность противоположного события $P(\overline{A_k}) = 1 - P(A_k) = 0{,}9$. Итак, $P(A_1 + A_2 + A_3) = 1 - P(\overline{A_1 + A_2 + A_3})$ и искомая вероятность события будет

$$P(A_1 + A_2 + A_3) = 1 - 0{,}9 \cdot 0{,}9 \cdot 0{,}9 = 0{,}271.$$

Пример 3. Имеются две урны с шариками трех цветов. В первой находятся 2 голубых, 3 красных, 5 зеленых, а во второй - 4 голубых, 2 красных и 4 зеленых. Из каждой урны извлекают по одному шару и сравнивают их цвета. Найти вероятность того, что цвета вынутых шаров одинаковы (событие A).

Решение. Обозначим событие, состоящее в извлечении из первой урны голубого шара, через B_1, красного - C_1, зеленого - D_1 . Аналогичные события для второй урны обозначим соответственно через B_2 , C_2 , D_2 . Событие A наступает в случае B_1B_2 , C_1C_2 или D_1D_2. Они несовместны. Для вычисления искомой вероятности события A применим формулы вероятностей суммы несовместных событий и произведения независимых событий

$P(A) = P(B_1B_2 + C_1C_2 + D_1D_2) = P(B_1B_2) + P(C_1C_2) + P(D_1D_2)$.

Так как независимы события: B_1 и B_2 , C_1 и C_2, D_1 и D_2, то можно пользоваться формулой $P(AB) = P(A)P(B)$ для каждой пары событий:

$P(B_1B_2) = P(B_1)P\{B_2)$,

$P(C_1C_2) = P(C_1)P(C_2)$,

$P(D_1D_2) = P(D_1)P(D_2)$.

Окончательно

$P(A) = P(B_1)P(B_2) + P(C_1)P(C_2) + P(D_1)P(D_2) = 0{,}2 \cdot 0{,}4 + 0{,}3 \cdot 0{,}2 + 0{,}5 \cdot 0{,}4 = 0{,}34$

Пример4. Сколько раз нужно подбросить два игральных кубика, чтобы вероятность выпадения хотя бы один раз двух шестерок была бы больше $\frac{1}{2}$? (Эта задача впервые поставлена французским математиком и писателем де Мере (1610-1684 гг.), поэтому задача называется его именем).

Решение. Пусть событие A_i - «выпадение двух шестерок при *i*-м подбрасывании». Так как с каждой из шести граней первого кубика может выпасть любая из шести граней второго кубика,

то всего равновозможных попарно несовместных событий 6 · 6 = 36. Только одно из них - выпадение шестерки и на первом и на втором кубике - благоприятствуют событию A_i. Следовательно, $P(A_i) = \frac{1}{36}$, откуда $q = 1 - p = 1 - \frac{1}{36} = \frac{35}{36}$.

Подбрасывание игральных кубиков - независимые испытания, поэтому воспользуемся формулой $P(A) = 1 - q^n$, тогда в данном случае получим: $1 - \left(\frac{35}{36}\right)^n > \frac{1}{2}$, или $\left(\frac{35}{36}\right)^n < \frac{1}{2}$.

Решив неравенство, найдем *n*. Логарифмируя обе части неравенства, получим $n \ln\frac{35}{36} < \ln\frac{1}{2}$, откуда $n > \frac{\ln 2}{\ln 36 - \ln 35} = \frac{0{,}6931}{0{,}0284} = 24{,}4$.

Итак, чтобы вероятность выпадения двух шестерок была больше $\frac{1}{2}$, достаточно подбросить кубик не менее 25 раз.

4. Условная вероятность. Формула полной вероятности. Формула Байеса

Пример 1. Слово *nanaxa* составлено из букв разрезной азбуки. Карточки с буквами тщательно перемешаны. Четыре карточки извлекаются по очереди и раскладываются в ряд. Какова вероятность получить при этом слово *nana?*

Решение. Обозначим через *A, B, C, D* соответственно события, состоящие в том, что: извлечена первая, вторая, третья и четвертая буква слова *nana* из набора в 6 букв: *a, a, a, n, n, x*. Найдем вероятности событий: *A, B/A*, *C/AB* , *D/ABC*.

$P(A) = \frac{2}{6} = \frac{1}{3}$; $P(B/A) = \frac{3}{5}$; $P(C/AB) = \frac{1}{4}$; $P(D/ABC) = \frac{2}{3}$.

В соответствии с формулой вероятности произведения зависимых событий при *n*=4 будем иметь:

$P(ABCD) = P(A)P(B/A)P(C/AB)P(D/ABC) = \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{30}$.

Пример 2. В пяти ящиках находятся одинаковые по размерам и весу шары. В двух ящиках - по 6 голубых и 4 красных шара (это ящик состава H_1). В двух других ящиках (состава H_2) - по 8 голубых и 2 красных шара. И в пятом ящике (состава H_3) - 8 красных и 2 голубых шара. Наудачу выбирается ящик, и из него извлекается шар. Какова вероятность того, что извлеченный шар оказался красным?

Решение. Событие, состоящее в том, что «извлечен красный шар» обозначим через *A*. Из условия задачи следует, что $P(H_1) = \frac{2}{5} = 0{,}4$, $P(H_2) = \frac{2}{5} = 0{,}4$, $P(H_3) = \frac{1}{5} = 0{,}2$.

Вероятность вынуть красный шар, если известно, что взят ящик первого состава H_2, будет определяться так:

$$P(A/H_1)=\frac{4}{10}=0{,}4$$

Вероятность извлечь красный шар, если известно, что взят ящик второго состава H_2, будет

$P(A/H_2)=\frac{2}{10}=0{,}2$. Вероятность извлечь красный шар, если известно, что взят ящик третьего состава H_3, будет $P(A/H_3)=\frac{8}{10}=0{,}8$.

При $n = 3$ находим искомую вероятность

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=0{,}4\cdot 0{,}4+0{,}4\cdot 0{,}2+0{,}2\cdot 0{,}8=0{,}4$$

Пример 3. Партия электрических лампочек на 20% изготовлена первым заводом, на 30% - вторым, на 50% - третьим. Вероятность выпуска бракованных лампочек соответственно равны: $q_1 = 0{,}01$, $q_2 = 0{,}005$, $q_3 = 0{,}006$. Найти вероятность того, что наудачу взятая из партии лампочка окажется стандартной.

Решение. Введем обозначения: A - событие, состоящее в том, что «из партии взята стандартная лампочка», H_1 - событие, состоящее в том, что «взятая лампочка изготовлена первым заводом», H_2 - событие, состоящее в том, что «взятая лампочка изготовлена вторым заводом», H_3 - событие, состоящее в том, что взятая лампочка изготовлена «третьим заводом». Найдем условные вероятности

$P(A/H_i)$, (i= 1,2,3) по формуле $P(A/H_i) = 1\text{-}P(\overline{A}/H_i)$, где $\overline{A}$ - событие, противоположное событию A (взята нестандартная лампочка):

$$P(A/H_1)=1-P(\overline{A}/H_1)=1-0{,}01=0{,}99,$$

$$P(A/H_2)=1-P(\overline{A}/H_2)=1-0{,}005=0{,}995,$$

$$P(A/H_3)=1-P(\overline{A}/H_3)=1-0{,}006=0{,}994.$$

Из условия задачи следует, что $P(H_1) = 0{,}2$, $P(H_2) = 0{,}3$, $P(H_3) = 0{,}5$.

Получим по формуле полной вероятности:

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=$$

$$=0{,}2\cdot 0{,}99+0{,}3\cdot 0{,}995+0{,}5\cdot 0{,}994=0{,}9935$$

Пример 4. В пяти ящиках находятся одинаковые по весу и размерам шары. В двух ящиках - по 6 зеленых и 4 красных шара (по ящик состава H_1). В двух других ящиках (состава H_2) - по 8 зеленых и 2 красных шара. В одном ящике (состава H_3) - 2 зеленых и 8 красных шаров. Наудачу выбирается ящик, и из него извлекается шар. Извлеченный шар оказался голубым. Какова вероятность того, что зеленый шар извлечен из ящика первого состава?

Решение. Обозначим через A событие, состоящее в том, что и і ящика извлечен голубой шар. Из условия задачи следует, что

$P(H_1)=\frac{2}{5}=0{,}4$; $P(H_2)=\frac{2}{5}=0{,}4$; $P(H_3)=\frac{1}{5}=0{,}2$.

Вероятность вынуть голубой шар, если известно, что взят ящик состава $H_1,H_2,...,H_3$ соответственно будут равны:

$$P(A/H_1)=\frac{6}{10}=0{,}6;$$

$$P(A/H_2)=\frac{8}{10}=0{,}8;$$

$$P(A/H_1)=\frac{2}{10}=0{,}2.$$

По формуле полной вероятности находим $P(A) = 0{,}4 \bullet 0{,}6 + 0{,}4 \bullet 0{,}8 + 0{,}2 \bullet 0{,}2 = 0{,}6$.

По формуле Бейеса найдем искомую вероятность

$$P(H_1 / A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0{,}4 \cdot 0{,}6}{0{,}6} = 0{,}4 .$$

5. Схема повторных испытаний. Формула Бернулли. Формула Пуассона. Локальная формулы Лапласа. Интегральная формула Лапласа

Пример 1. Частица находится на прямой в начале координат. Под действием случайных толчков частица каждую секунду перемещается вправо (с вероятностью $\frac{1}{3}$) или влево (с вероятностью $\frac{2}{3}$) на единицу масштаба. Найти вероятность того, что через 4 секунды частица вернется в начало координат.

Решение. Через 4 секунды частица вернется в начало координат в том случае, если она переместится ровно два раза вправо (и, значит, два раза влево). По формуле Бернулли найдем вероятность того, что из четырех независимых перемещений частицы ровно два перемещения будут вправо:

$$n = 4 \quad k = 2 \quad p = \frac{1}{3} \quad q = \frac{2}{3}. \quad P4(2) = C_4^2 \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2 = 6 \cdot \frac{1}{9} \cdot \frac{4}{9} = \frac{24}{81} \approx 0{,}296 .$$

Пример 2. К электросети подключено 36 приборов, каждый мощностью 5 киловатт и потребляет в данный момент энергию с вероятностью 0,2. Найти вероятность того, что потребляемая в данный момент мощность:

а) составит ровно 50 киловатт;

б) превзойдет 50 киловатт.

Решение. В случае а) надо найти вероятность того, что из 36 приборов работают ровно 10. Применим локальную теорему Лапласа: $n = 36 \quad k = 10 \quad p = 0{,}2 \quad q = 0{,}8$.

$$\sqrt{npq} = 24 \quad x = \frac{k - np}{\sqrt{npq}} = 1{,}4 . \quad P_{36}(10) = \frac{1}{\sqrt{npq}} \varphi(x) = \frac{1}{2{,}4} \cdot \varphi(1{,}4) = 0{,}0624 .$$

Значение функции локальной функции Лапласа *φ(x)* взято из таблицы приложений.

В случае б) находим вероятность $P_{36}(k \geq 10)$ того, что работают более десяти приборов. Применяем для решения этой части задачи интегральную теорему Лапласа. Находим сначала значения x_1, x_2:

$$x_1 = \frac{k - np}{\sqrt{npq}} = 1{,}4 , \; x_2 = \frac{n - np}{\sqrt{npq}} = 12 /$$

Тогда искомая вероятность будет:

$$P_{36}(k \geq 10) = \Phi(x_2) - \Phi(x_1) = 0{,}5 - 0{,}4192 = 0{,}08808 ,$$

Значения функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{t^2}{2}} dt$ взяты из таблицы приложений.

Пример 3. В нерестовике содержится 200 рыб - производителей вида *A*. Вероятность отдачи икры в искусственных условиях рыбы вида *A* равна $\frac{3}{4}$. Требуется найти вероятность того, что

икру отдадут 150 рыб.

Решение. Вероятность того, что ровно 150 рыб из 200 отдадут икру, найдем, используя локальную теорему Лапласа

$$P_n(k) = \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} = \frac{1}{\sqrt{npq}} \cdot \varphi(x) , \text{ где } x = \frac{k - np}{\sqrt{npq}} .$$

Значение функции $\varphi(x)$ возьмем из таблицы. Находим:

$n = 200$, $npq = 200 \cdot 0{,}75 \cdot 0{,}25 = 150 \cdot 0{,}25 = 0{,}375$.

$m = к = np = 150$ $\quad \sqrt{npq} = 6{,}12$.

$p = 0{,}75$ $\quad x = \dfrac{150-150}{6{,}12} = 0$.

$q = 0{,}25$.

Получим: $P_{200}(150) = \dfrac{1}{6{,}12} \cdot \varphi(0) = \dfrac{0{,}3989}{6{,}12} \approx 0{,}07$.

Пример 4. В партии из 400 деталей 80% - стандартных. Найти границы, в которых с вероятностью 0,9544 заключена доля стандартных деталей.

Решение. Воспользуемся формулой, являющейся частным случаем формулы Муавра-Лапласа

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right),$$

где m/n - доля числа наступивших событий A в n испытаниях,

n - число испытаний,

p - вероятность наступления события A в одном испытании,

ε - величина отклонения доли m/n от вероятности p,

$q = 1 - p$ - вероятность ненаступления события A в одном испытании.

Для данной задачи A - событие, состоящее в том, что деталь стандартная, $n = 400$; $p = 0{,}8$; $q=0{,}2$; $P = 0{,}9544$, величину ε нужно найти.

Итак: $0{,}9544 = 2\Phi\left(\dfrac{\varepsilon\sqrt{400}}{\sqrt{0{,}8 \cdot 0{,}2}}\right) \Leftrightarrow \Phi(\varepsilon \cdot 50) = 0{,}4772$.

По таблице- приложений значений функции Лапласа $Ф(x)$ находим, что $\varepsilon \cdot 50 = 2$, следовательно, $\varepsilon = 0{,}04$. Таким образом, $|m/n - 0{,}8| < 0{,}04$ и $0{,}76 < m/n < 0{,}84$.

6. Простейший поток событий. Вероятность случайного события с заданной интенсивностью

Пример 1. Среднее число заявок, поступающих на предприятие бытового обслуживания за 1 час, равно трем. Найти вероятность того, что за 2 часа поступит 5 заявок. Предполагается, что поток заявок - простейший.

Решение. По условию $\lambda = 3$, $t = 2$, $k = 5$. Воспользуемся формулой

$$P_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Искомая вероятность того, что за 2 часа поступит 5 заявок, равна

$$P_2(5) = \frac{(6)^5 \cdot 0{,}00248}{120} \approx 0{,}268.$$

Пример 2. Среднее число заявок, поступающих на АТС в одну минуту, равно двум. Найти вероятности того, что за четыре минуты поступит:

а) три вызова;

б) менее трех вызовов;

в) не менее трех вызовов.

Решение, а) По условию $\lambda = 3$, $t = 2$, $k=5$. Воспользуемся формулой: $P_t(k) = \dfrac{(\lambda t)^k e^{-\lambda t}}{k!}$

Подставив данные условия задачи, получим: $P_4(3) = \dfrac{8^3 \cdot e^{-8}}{3!} = \dfrac{512 \cdot 0{,}000335}{6} \approx 0{,}03$.

б) Найдем вероятность того, что за четыре минуты поступит менее трех вызовов, т.е. ни одного вызова, или один вызов, или два вызова. Поскольку эти события несовместны, применим теорему суммы несовместных событий:

$$P_4(k<3)=P_4(0)+P_4(1)+P_4(2)=e^{-8}+8\cdot e^{-8}\cdot\frac{8^2\cdot e^{-8}}{2!}=41\cdot 0{,}000335\approx 0{,}01.$$

в) Найдем вероятность того, что за четыре минуты поступит не менее трех вызовов: так как события «поступило менее трех вызовов» и «поступило не менее трех вызовов» - противоположные, то сумма вероятностей этих событий равна единице: $P_4(k<3) + P_4(k\geq 3) = 1$. Поэтому $P_4(k\geq 3)=1-P_4(k<3)=1-[P_4(0)+P_4(1)+P_4(2)]=1-0{,}01=0{,}99$.

2.1.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили основные понятия комбинаторики, теории случайных событий, классификацию случайных событий;
- усвоили основные правила, применяемые в теории случайных событий;
- выработали навыки по вычислению вероятностей случайных событий, их суммы, произведения.
- освоили понятие условная вероятность, формулу полной вероятности, формулы Байеса, Бернулли, Лапласа, Пуассона;
- усвоили основные правила применения формул Байеса, Бернулли, Лапласа, Пуассона, работы с простейшим потоком событий;
- выработали навыки по вычислению вероятностей случайных событий в схеме повторных испытаний, в простейшем потоке с заданной интенсивностью, работы с таблицами функций Гаусса и Лапласа.

2.2 Практическое занятие № 2 (2 часа).

Тема: «Понятие случайной величины примеры. Виды случайных величин. Закон распределения вероятностей. Функция распределения случайных величин. Свойства. Плотность распределения вероятностей. Числовые характеристики: математическое ожидание, свойства; дисперсия, свойства; среднее квадратичное отклонение и его свойства»

2.2.1 Задание для работы:

1.Случайные величины, их классификация. Закон распределения случайной величины. Ряд распределения.

3. Функция распределения. Плотность распределения.

4. Числовые характеристики ДСВ. Числовые характеристики НСВ.

5. Свойства числовых характеристик, их интерпретация.

2.2.2 Краткое описание проводимого занятия:

1. Случайные величины, их классификация. Закон распределения случайной величины. Ряд распределения.

Пример 1. Сырье на завод привозят от 3-х независимо работающих поставщиков на автомашинах. Вероятность прибытия автомашины от первого поставщика равна 0,2; от

второго - 0,3; от третьего - 0,1. Составить закон распределения числа прибывших машин. Найти математическое ожидание *M(X)* , дисперсию *D(X)* и среднее квадратическое отклонение *σ(X)* случайной величины *X.* Найти функцию распределения и построить ее график.
Решение. Для нахождения числовых характеристик дискретной случайной величины *X* - числа прибывших автомашин, необходимо составить закон ее распределения, который в общем вине записывается в виде таблицы так:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Где x_i, - возможные значения дискретной случайной величины *X,* $P_t = P(X = x_i)$ - вероятность того, что случайная величина *X* примет значение x_i, причем $\sum_{i=1}^{n} p_i = 1$. Для данного случая имеем:

x_i	0	1	2	3
p_i	p_1	p_2	p_3	p_4

Надо найти значения вероятностей p_i . Равенство $X = 0$ означает**,** что на завод не прибудет ни одна из трех автомашин. Следовательно: $p_1 = p(X = 0) = 0{,}8 \cdot 0{,}7 \cdot 0{,}9 = 0{,}504$ (по теореме умножения вероятностей независимых событий).
Равенство $X = 1$ означает, что на завод прибудет только одна из трех автомашин. Пользуясь теоремой сложения вероятностей несовместных событий и теоремой умножения независимых событий, найдем значение p_2:
$p_2 = p(X = 1) = 0{,}2 \cdot 0{,}7 \cdot 0{,}9 + 0{,}8 \cdot 0{,}3 \cdot 0{,}9 + 0{,}8 \cdot 0{,}7 \cdot 0{,}1 = 0{,}398.$
Рассуждая аналогично, найдем p_3 и p_4:
$p_3 = 0{,}2 \cdot 0{,}3 \cdot 0{,}9 + 0{,}8 \cdot 0{,}3 \cdot 0{,}1 + 0{,}2 \cdot 0{,}7 \cdot 0{,}1 = 0{,}092,$
$p_4 = 0{,}2 \cdot 0{,}3 \cdot 0{,}1 = 0{,}006.$
Запишем закон распределения:

x_i	0	1	2	3
p_i	0,504	0,398	0,092	0,006

2. Функция распределения. Плотность распределения

Пример. Случайная величина *X* задана интегральной функцией (функцией распределения) *F(X).* Требуется найти:
а) дифференциальную функцию (плотность вероятности);
б) математическое ожидание и дисперсию *X;*
в) построить графики интегральной и дифференциальной функций $F(x) = \begin{cases} 0, & x \le 0 \\ x^2, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$.

Решение. а) между интегральной и дифференциальной функциями непрерывной случайной величины выполняется соотношение *F'({x) = f(x).* В данном случае будем иметь

$$F(x) = \begin{cases} 0, & x \le 0 \\ 2x, & 0 < x \le 1 \\ 0, & x > 1 \end{cases}$$

3. Числовые характеристики ДСВ. Числовые характеристики НСВ

Пример 1
Запишем закон распределения:

x_i	0	1	2	3
p_i	0,504	0,398	0,092	0,006

Для вычисления математического ожидания *M(X)*, дисперсии *P(X)* и среднего квадратического отклонения *σ(X)* воспользуемся формулами приведенными выше:

$M(X) = 0 \cdot 0{,}504 + 1 \cdot 0{,}398 + 2 \cdot 0{,}092 + 3 \cdot 0{,}06 = 0{,}6$,

$D(X) = 0 \cdot 0{,}504 + 1 \cdot 0{,}398 + 4 \cdot 0{,}092 + 9 \cdot 0{,}06 - 0{,}6^2 = 0{,}46$,

$\sigma(X) = 0{,}678$.

Найдем функцию распределения $F(x) = P(X < x)$:

если $x \le 0$, то $F(x) = 0$,

если $0 < x \le 1$, то $F(x) = 0{,}504$,

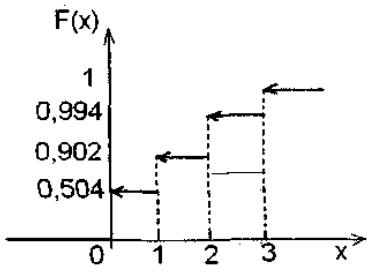
если $1 < x \le 2$, то $F(x) = 0{,}504 + 0{,}398 = 0{,}902$,

если $2 < x \le 3$, то $F(x) = 0{,}902 + 0{,}092 = 0{,}994$,

если $x > 3$, то $F\{x) = 0{,}994 + 0{,}006 = 1$.

Таким образом: $F(x) = \begin{cases} 0 & \text{при } x \le 0 \\ 0{,}504 & \text{при } 0 < x \le 1 \\ 0{,}902 & \text{при } 1 < x \le 2 \\ 0{,}994 & \text{при } 2 < x \le 3 \\ 1 & \text{при } 3 < x \end{cases}$

Построим график этой функции:



Пример 2. Случайная величина *X* задана интегральной функцией (функцией распределения) *F(X).* Требуется найти:

а) математическое ожидание и дисперсию *X;*

б) построить графики интегральной и дифференциальной функций $F(x) = \begin{cases} 0, & x \le 0 \\ x^2, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$.

а) числовые характеристики непрерывной случайной величины определяются по формулам:

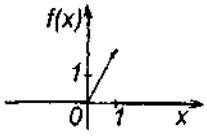
$$M(X) = \int_{-\infty}^{\infty} x f(x)dx, \; D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - (M(X))^2$$

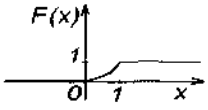
Тогда имеем

$$M(X) = \int_{-\infty}^{0} x \cdot 0dx + \int_{0}^{1} x \cdot 2xdx + \int_{1}^{\infty} x \cdot 0dx = 2\int_{0}^{1} x^2 dx = 2\frac{x^3}{3}\Big|_0^1 = \frac{2}{3}.$$

$$D(X) = \int_{-\infty}^{0} x^2 \cdot 0dx + \int_{0}^{1} x^2 \cdot 2xdx + \int_{1}^{\infty} x^2 \cdot 0dx - \frac{4}{9} = 2\int_{0}^{1} x^2 dx - \frac{4}{9} = 2\frac{x^4}{4}\Big|_0^1 - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \frac{1}{18}$$

б) строим графики функций





4. Свойства числовых характеристик, их интерпретация

Пример. Заданы математическое ожидание a и среднее квадратическое отклонение σ нормально распределенной служимой величины X: $a = 8$, $\sigma = 2$, α=4, $\beta = 14$, $\delta = 6$. Требуется найти:

а) вероятность того, что X примет значение, принадлежащее интервалу (4;14);

б) вероятность того, что абсолютная величина отклонения X-a окажется меньше δ.

Решение. а) вероятность того, что нормально распределенная случайная величина примет значение, принадлежащее интервалу (α; β), равна $P(\alpha < x < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$

$P(4 < x < 14) = \Phi\left(\frac{14-8}{2}\right) - \Phi\left(\frac{4-8}{2}\right) = \Phi(3) + \Phi(2) \approx$ (по таблице значений функции $Ф(x)$≈ 0,4986+0,4772 = 0,9758;

б) вероятность того, что абсолютная величина отклонения меньше положительного числа δ равна $P(|x - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$. В данном случае имеем

$P(|x - 8| < 6) = 2\Phi\left(\frac{6}{2}\right) = 2\Phi(3) \approx 2 \cdot 0{,}4986 = 0{,}9972$.

2.2.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили основные понятия теории случайных величин, классификации, случайных величин, понятие закона, ряда, функции, плотности распределения;
- усвоили основные правила нахождения функции распределения вероятностей, плотности распределения вероятностей;
- выработали навыки по вычислению вероятности попадания в интервал ДСВ, НСВ; числовых характеристик случайных величин; применению свойств числовых характеристик.

2.3 Практическое занятие № 3 (2 часа).

Тема: «Законы распределения ДСВ: биноминальный и Пуассона. Законы распределения вероятностей НСВ: равномерное распределение, показательное распределение. Нормальное распределение вероятностей НСВ. Правило трех сигм»

2.3.1 Задание для работы:

1. Биномиальное распределение, его свойства, числовые характеристики.
2. Распределение Пуассона, его свойства, числовые характеристики.
3. Равномерное распределение, его свойства, числовые характеристики.
4. Показательное распределение, его свойства, числовые характеристики.
5. Нормальное распределение, его свойства, числовые характеристики.

3.3.2 Краткое описание проводимого занятия:

1. Биномиальное распределение, его свойства, числовые характеристики.

Рассмотрим осуществление схемы Бернулли, т.е. производится серия повторных независимых испытаний, в каждом из которых данное событие А имеет одну и ту же вероятность $P(A)=p$, не зависящую от номера испытания. И для каждого испытания имеются только два исхода:

1) событие А – успех; 2) событие $\overline{A}$ - неуспех,

с постоянными вероятностями $P(A)=p \quad и \quad P(\overline{A})=q, \quad q=1-p$

Введем в рассмотрение дискретную случайную величину Х – «число появлений события А при п испытаниях» и найдем закон распределения этой случайной величины. Величина Х может принимать значения $x_0=0, x_1=1, x_2=2, \ldots x_n=n.$

Вероятность $p_k \; k=0,1,\ldots, n$ того, что случайную величину Х примет значение x_k находится по формуле Бернулли

$$p_k = P_n(k) = C_n^k p^k q^{n-k} = \frac{n!}{(n-k)!\cdot k!}\cdot p^k q^{n-k} \qquad (1)$$

Закон распределения дискретной случайной величины, определяемый формулой Бернулли (1), называется **биномиальным законом распределения.** Постоянные **п** и **р** (q=1-p), входящие в формулу (1) называются **параметрами биномиального распределения.** Название «биномиальное распределение» связано с тем, что правая часть в равенстве (1) это общий член разложения бинома Ньютона $(q-p)^n$,т.е.

$$(q-p)^n = q^n - C_n^1 q^{n-1} p - C_n^2 q^{n-2} p^2 - \ldots - C_n^k q^{n-k} p^k - \ldots - p^n \qquad (2)$$

А так как p+q=1, то правая часть равенства (2) равна 1

$$q^n - C_n^1 q^{n-1} p - C_n^2 q^{n-2} p^2 - \ldots - C_n^k q^{n-k} p^k - \ldots - p^n = 1 \qquad (3)$$

Это означает, что

$$\sum_{k=0}^{n} P_n(k) = \sum_{k=0}^{n} C_n^k q^{n-k} p^k = 1 \qquad (4)$$

В равенстве (3) первый член q^n в правой части означает вероятность того, что в п испытаниях событие А не появится ни разу, второй член $C_n^1 q^{n-1} p = npq^{n-1}$ вероятность того, что событие А появится один раз, третий член – вероятность, что событие А появится два раза и наконец, последний член p^n – вероятность того, что событие А появится ровно п раз.

Биномиальный закон распределения дискретной случайной величины представляют в виде таблицы:

		1		k		
	n	$C_n^1 q$		$C_n^k q^{n-}$		n

Основные числовые характеристики биномиального распределения:

1) математическое ожидание $M(X)=np$ (5)

2) дисперсия $D(X) = npq$ (6)

3) среднее квадратическое отклонение $\sigma_X = \sqrt{npq}$ (7)

4) наивероятнейшее число появление события k_0 – это число которому при заданном п соответствует максимальная биномиальная вероятность $P_n(k_0)$

При заданных п и р это число определяется неравенствами

$$np - q \le k_0 \le np + p \quad (8)$$

если число пр+р не является целым, то k_0 равно целой части этого числа, если же пр+р – целое число, то k_0 имеет два значения $k_0' = np - q, k_0' = np - p$

Биномиальный закон распределения вероятностей применяется в теории стрельбы, в теории и практике статистического контроля качества продукции, в теории массового обслуживания, в теории надежности и т.д. Этот закон может применяться во всех случаях, когда имеет место последовательность независимых испытаний.

Пример 1: Проверкой качества установлено, что из каждых 100 приборов не имеют дефекты 90 штук в среднем. Составить биномиальный закон распределения вероятностей числа качественных приборов из приобретенных наугад 4.

Решение: Событие А – появление которого проверяется это – «приобретенный наугад прибор качественный». По условию задачи основные параметры биномиального распределения:

$$n = 4, \quad p = P(A) = \frac{90}{100} = 0,9, \quad q = 1 - p = 1 - 0,9 = 0,1$$

Случайная величина Х – число качественных приборов из взятых 4, значит значения Х - $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$ Найдем вероятности значений Х по формуле (1):

$$p_0 = P_4(0) = q^4 = 0,1^4 = 0,0001$$

$$p_1 = P_4(1) = C_4^1 p^1 q^{4-1} = \frac{4!}{(4-1)!\cdot 1!} \cdot 0,9^1 \cdot 0,1^{4-1} = 4 \cdot 0,9 \cdot 0,001 = 0,0036$$

$$p_2 = P_4(2) = C_4^2 p^2 q^{4-2} = \frac{4!}{(4-2)!\cdot 2!} \cdot 0,9^2 \cdot 0,1^{4-2} = 6 \cdot 0,81 \cdot 0,01 = 0,0486$$

$$p_3 = P_4(3) = C_4^3 p^3 q^{4-3} = \frac{4!}{(4-3)!\cdot 3!} \cdot 0,9^3 \cdot 0,1^{4-3} = 4 \cdot 0,729 \cdot 0,1 = 0,2916$$

$$p_4 = P_4(4) = p^4 = 0,9^4 = 0,6561$$

Таким образом, закон распределения величины Х - число качественных приборов из взятых 4:

	,0001	,0036	,0486	,2916	,6561

Для проверки правильности построения распределения проверим чему равна сумма вероятностей

$$\sum_{k=0}^{4} P_n(k) = 0{,}00001 + 0{,}0036 + 0{,}0486 + 0{,}2916 + 0{,}6561 = \mathbf{1}$$

Ответ: Закон распределения

	,0001	,0036	,0486	,2916	,6561

Пример 2: Применяемый метод лечения приводит к выздоровлению в 95 % случаев. Пятеро больных применяли данный метод. Найти наивероятнейшее число выздоровевших, а так же числовые характеристики случайной величины X – число выздоровевших из 5 больных применявших данный метод.

Решение: Событие A - больной применявший лечение выздоровел, тогда основные параметры биномиального распределения:

$$n = 5; \quad p = P(A) = \frac{95}{100} = 0{,}95; \quad q = 1 - p = 1 - 0{,}95 = 0{,}05$$

По формуле (8) найдем k_0 наивероятнейшее число выздоровевших из 5. Найдем $np + p = 5 \cdot 0{,}95 + 0{,}05 = 4{,}8$ получили не целое число значит k_0 равно целой части, т.е. $k_0 = 4$.

Теперь найдем числовые характеристики X – число выздоровевших из 5 больных применявших данный метод лечения:

1) математическое ожидание по формуле (5) $M(X) = 5 \cdot 0{,}95 = 4{,}75$

2) дисперсия по формуле (6) $D(X) = 5 \cdot 0{,}95 \cdot 0{,}05 = 0{,}2375$

3) среднее квадратическое отклонение по формуле (7) $\sigma_X = \sqrt{0{,}2375} = 0{,}49$

Ответ: $k_0 = 4 \;\; M(X) = 4{,}75 \;\; D(X) = 0{,}2375 \;\; \sigma_X = 0{,}49$

2. Распределение Пуассона, его свойства, числовые характеристики

Приведем примеры, приводящие к случайным величинам, распределенным по закону Пуассона:

· Автоматическая телефонная станция получает в среднем за минуту *A* вызовов. Какова вероятность того, что за данную минуту она получит ровно *M* вызовов? Случайное число вызовов за *Данную минуту* распределено по закону Пуассона.

· Автодорожная инспекция регистрирует количество аврий за неделю на определенном участке дороги. Какова вероятность того, что в течение данной недели произойдет ровно *M* дорожных аварий? Случайное число аварий *За неделю* распределено по закону Пуассона.

Аналогичные примеры можно привести не только для временных интервалов (минута, неделя), но и при учете дефектов дорожного покрытия *На километр пути* или опечаток *На страницу текста*.

Отличительные черты эксперимента, приводящего к распределению Пуассона (на примере временных интервалов):

1. каждый малый интервал времени может рассматриваться как испытание, результатом которого служит либо «успех» - поступление телефонного вызова, либо «неудача». Интервалы столь малы, что может быть только один «успех» в одном интервале, вероятность которого мала и неизменна.

2. Число «успехов» в одном большом интервале не зависит от их числа в другом. То есть попадание «успехов» в неперекрывающиеся интервалы – события независимые, и «успехи» беспорядочно разбросаны по временным промежуткам;

3. Среднее число «успехов» в большом интервале для разных интервалов постоянно на протяжении всего времени.

Число «успехов» на заданном интервале будет случайной величиной, распределенной по закону Пуассона. Случайное число аварий за неделю может принимать значения 0, 1, 2, 3, … (верхнего предела нет). Вероятность того, что случайная величина *X*, распределенная по закону Пуассона примет значение *M*, вычисляется по известной формуле Пуассона:

$$P_m = \frac{a^m}{m!} \cdot e^{-a}$$

, *M* = 0, 1, 2, …

При условии $p \to 0, \quad n \to \infty, \quad np \to \lambda = const$ закон распределения Пуассона является предельным случаем биномиального закона. Так как при этом вероятность р события А в каждом испытании мала, то закон распределения Пуассона называют часто законом редких явлений.

Наряду с "предельным" случаем биномиального распределения закон Пуассона может возникнуть и в ряде других случаев. Так для простейшего потока событий число событий, попадающих на произвольный отрезок времени, есть случайная величина, имеющая пуассоновское распределение. Также по закону Пуассона распределены, например, число рождения четверней, число сбоев на автоматической линии, число отказов сложной системы в "нормальном режиме", число "требований на обслуживание", поступивших в единицу времени в системах массового обслуживания, и др.

Замечание. Если случайная величина представляет собой сумму двух независимых случайных величин, распределённых по закону Пуассона, то она также распределена по закону Пуассона.

Числовые характеристики распределения Пуассона.

Математическое ожидание равно *Дисперсии* и равно параметру распределения *A*: M(X)= *A*, D(X)= *A*.

3. Равномерное распределение, его свойства, числовые характеристики

На практике встречаются случайные величины, о которых заранее известно, что они могут принять какое-либо значение в строго определенных границах, причем в этих границах все значения случайной величины имеют одинаковую вероятность (обладают одной и той же плотностью вероятностей).

Например, при поломке часов остановившаяся минутная стрелка будет с одинаковой вероятностью (плотностью вероятности) показывать время, прошедшее от начала данного часа до поломки часов. Это время является случайной величиной, принимающей с одинаковой плотностью вероятности значения, которые не выходят за границы, определенные продолжительностью одного часа. К подобным случайным величинам относится также и погрешность округления. Про такие величины говорят, что они распределены равномерно, т. е. имеют равномерное распределение.

Определение: Непрерывная случайная величина X имеет равномерное распределение на отрезке [а, в], если на этом отрезке плотность распределения вероятности случайной

величины постоянна, т. е. если дифференциальная функция распределения f(x) имеет следующий вид:

$$f(x)=\begin{cases} c, x\in[a,b]; \\ 0, x\notin[a,b]. \end{cases}$$

Иногда это распределение называют *законом равномерной плотности.* Про величину, которая имеет равномерное распределение на некотором отрезке, будем говорить, что она распределена равномерно на этом отрезке.

Найдем значение постоянной с. Так как площадь, ограниченная кривой распределения и осью *Ox,* равна 1, то

$$\int_{-\infty}^{\infty} f(x)dx=\int_{a}^{b} cdx=c(b-a)=1,$$

откуда $c=1/(b-a)$.

Теперь функцию f(x) можно представить в виде

$$f(x)=\begin{cases} 0 & \text{при } x<a, \\ 1/(b-a) & \text{при } a\leqslant x\leqslant b, \\ 0 & \text{при } x>b. \end{cases}$$

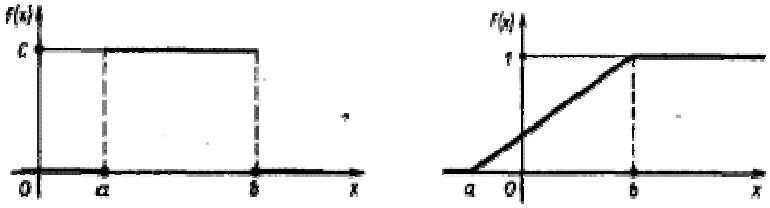
Построим функцию распределения F(x), для чего найдем выражение F(x) на интервале $[a, b]$:

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{a}^{x}\frac{1}{b-a}dt=\frac{t}{b-a}\Big|_{a}^{x}=\frac{x-a}{b-a}.$$

При $x<a$ функция $F(x)=0$ и $F(x)=1$ при $x>b$. Таким образом,

$$F(x)=\begin{cases} 0 & \text{при } x<a, \\ \frac{x-a}{b-a} & \text{при } a\leqslant x\leqslant b, \\ 1 & x>b. \end{cases}$$

Графики функций f(x) и F(x) имеют вид:



Найдем числовые характеристики.

Используя формулу для вычисления математического ожидания НСВ, имеем:

$$MX=\int_{a}^{b} xf(x)dx=\int_{a}^{b}\frac{x}{b-a}dx=\frac{b+a}{2}.$$

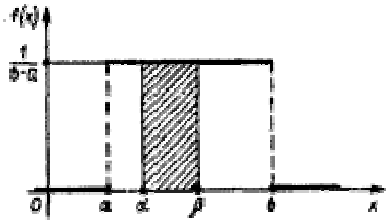
Таким образом, математическое ожидание случайной величины, равномерно распределенной на отрезке $[a, b]$ совпадает с серединой этого отрезка.

Найдем дисперсию равномерно распределенной случайной величины:

$$DX=\int_{-\infty}^{\infty}(x-MX)^2 f(x)dx=\int_{a}^{b}\left(x-\frac{b+a}{2}\right)^2\frac{1}{b-a}dx=\frac{(b-a}{12}$$

откуда сразу же следует, что среднее квадратическое отклонение:

$$\sigma_X=\frac{b-a}{2\sqrt{3}}.$$



Найдем теперь вероятность попадания значения случайной величины, имеющей равномерное распре-

деление, на интервал (α,β), принадлежащий целиком отрезку [*a, b*]:

$$P(\alpha<X<\beta)=\int_{\alpha}^{\beta} f(x)\,dx=\int_{\alpha}^{\beta}\frac{dx}{b-a}=\frac{\beta-\alpha}{b-a},$$

Геометрически эта вероятность представляет собой площадь заштрихованного прямоугольника. Числа *a* и *b* называются *параметрами распределения и* однозначно определяют равномерное распределение.

Пример. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 минут. Найти вероятность того, что пассажир, подошедший к остановке. Будет ожидать очередной автобус менее 3 минут.

Решение:

СВ- время ожидания автобуса имеет равномерное распределение. Тогда искомая вероятность будет равна: $P(0,3)=\frac{3-0}{5-0}=0,6.$

Пример. Ребро куба x измерено приближенно. Причем

$a \le x \le b$

Рассматривая ребро куба как случайную величину, распределенную равномерно в интервале (*a, b)*, найти математическое ожидание и дисперсию объема куба.

Решение: Объем куба- случайная величина, определяемая выражением У= X^3. Тогда математическое ожидание равно:

$$M(X)=\int_{a}^{b} x^3\frac{1}{b-a}dx=\frac{x^4}{4(b-a)}\Big|_a^b=\frac{1}{4}\frac{b^4-a^4}{b-a}=\frac{(b+a)(b^2+a^2)}{4}.$$

Дисперсия:

$$D(X)=\int_{a}^{b} x^6\frac{1}{b-a}dx-[M(X)]^2=\frac{1}{7}\frac{x^7}{b-a}\Big|_a^b-[M(X)]^2=\frac{1}{7}\frac{b^7-a^7}{b-a}-[\frac{(b+a)(b^2+a^2)}{4}]^2.$$

4. Показательное распределение, его свойства, числовые характеристики

Определение: Непрерывная случайная величина X, функция плотности которой задается выражением

$$f(x)=\begin{cases}\mu e^{-\mu x}, x>0;\\ 0, x\le 0,\end{cases}$$

называется случайной величиной, имеющей показательное, или экспоненциальное, распределение.

Величина срока службы различных устройств и времени безотказной работы отдельных элементов этих устройств при выполнении определенных условий обычно подчиняется показательному распределению. Другими словами, величина промежутка времени между появлениями двух последовательных редких событий подчиняется зачастую показательному распределению.

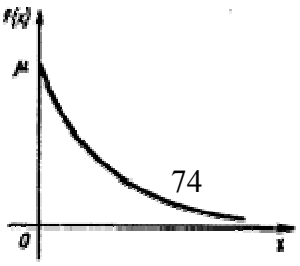
Как видно из формулы , показательное распределение определяется только одним параметром μ.

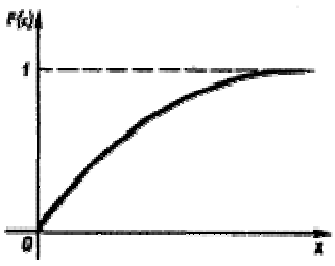
Найдем функцию распределения показательного закона, используя свойства дифференциальной функции распределения:

$$F(x)=\begin{cases}0 \text{ при } x\le 0,\\ \int_{-\infty}^{x} f(t)dt=\int_{0}^{x}\mu e^{-\mu t}dt=1-e^{-\mu x} \text{ при } x>0.\end{cases}$$

Графики дифференциальной и интегральной функций показательного распределения имеют вид:

4.2. Числовые характеристики.





Используя формулы для вычисления математического ожидания, дисперсии и среднего квадратического отклонения нетрудно убедится, что для показательного распределения

$$M(X)=\frac{1}{\mu}, D(X)=\frac{1}{\mu^2}, \sigma(X)=\frac{1}{\mu}\ .$$

Таким образом, для показательного распределения характерно, что среднее квадратическое отклонение численно равно математическому ожиданию.

Найдем вероятность попадания СВ в интервал *(a,b)*:

$$P(a,b)=F(b)-F(a)=e^{-\mu b}-e^{-\mu a}.$$

4.3. Функция надежности.

Пусть некоторое устройство начинает работать в момент времени $t_0 = 0$, а по истечении времени длительностью t происходит отказ. Обозначим через T НСВ - длительность времени безотказной работы устройства. Если устройство проработало безотказно время меньшее t, то, следовательно, за время длительностью t наступит отказ. **Тогда функция распределения $F(t)=P(T<t)=1- e^{-\mu t}$ определяет вероятность отказа устройства за время t.**

Найдем вероятность противоположного события - безотказной работы за время t:

$P(T>t)=1-F(t)=e^{-\mu t}=R(t)$ **. Функция R(t) называется функцией надежности.**

Выясним смысл числовых характеристик и параметра распределения.

Математическое ожидание - это среднее время между двумя ближайшими отказами устройства, а величина обратная математическому ожиданию (параметр распределения)- интенсивность отказов, т.е. количество отказов в единицу времени.

Пример. Время безотказной работы устройства распределено по закону

$$f(t)=0{,}02e^{-0{,}02t}, t\geq 0.$$

Найти среднее время безотказной работы устройства, вероятность того, что устройство не откажет за среднее время безотказной работы. Найти вероятность отказа за время t= 100 часов.

Решение: По условию интенсивность отказов $\mu = 0{,}02$. Тогда среднее время между двумя отказами, т.е. математическое ожидание M(X)=1/0,02=50часов. Вероятность безотказной работы за этот промежуток времени вычислим по функции надежности:

$$R(50)=e^{-0{,}02\cdot 50}=e^{-1}\approx 0{,}37.$$

По функции F(t) вычислим вероятность отказа за время t =100 часов:

$$F(100)=1-e^{-0{,}02\cdot 100}=1-e^{-2}\approx 0{,}86.$$

5. Нормальное распределение, его свойства, числовые характеристики

5.1. Интегральная и дифференциальная функции распределения. Вероятность попадания в заданный интервал.

Одним из наиболее часто встречающихся распределений является нормальное распределение. Оно играет большую роль в теории вероятностей и занимает среди других распределений особое положение. Нормальный закон распределения является предельным законом, к которому приближаются другие законы распределения при часто встречающихся аналогичных условиях.

Если предоставляется возможность рассматривать некоторую случайную величину как сумму достаточно большого числа других случайных величин, то данная случайная величина обычно подчиняется нормальному закону распределения. Суммируемые случайные величины могут подчиняться каким угодно распределениям, но при этом должно выполняться условие их независимости (или слабой зависимости). При соблюдении некоторых не очень жестких условий указанная сумма случайных величин подчиняется при-

ближенно нормальному закону распределения и тем точнее, чем большее количество величин суммируется.

Ни одна из суммируемых случайных величин не должна резко отличаться от других, т. е. каждая из них должна играть в общей сумме примерно одинаковую роль и не иметь исключительно большую по сравнению с другими величинами дисперсию.

Для примера рассмотрим изготовление некоторой детали на станке-автомате. Размеры изготовленных деталей несколько отличаются от требуемых. Это отклонение размеров от стандарта вызывается различными причинами, которые более или менее независимы друг от друга. К ним могут относиться: неравномерный режим обработки детали; неоднородность обрабатываемого материала; неточность установки заготовки в станке; износ режущего инструмента и деталей станков; упругие деформаций узлов станка; состояние микроклимата в цехе; колебание напряжения в электросети и т. д. Каждая из перечисленных и подобных им причин влияет на отклонение размера изготовляемой детали от стандарта. Таким образом, общее отклонение размера, фиксируемое измерительным прибором, является суммой большего числа отклонений, обусловленных различными причинами. Если ни одна из этих причин не является доминирующей, то суммарное отклонение является случайной величиной, имеющей нормальный закон распределения.

Так как нормальному закону подчиняются только непрерывные случайные величины, то это распределение можно задать в виде плотности распределения вероятности.

Определение: Непрерывная случайная величина X имеет нормальное распределение (распределена по нормальному закону), если плотность распределения вероятности f(x) имеет вид $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$.

где a и σ—некоторые постоянные, называемые параметрами нормального распределения.

Функция распределения F(x) в рассматриваемом случае принимает вид

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Параметр a- есть математическое ожидание НСВХ, имеющей нормальное распределение, σ - среднее квадратическое отклонение, тогда дисперсия равна $D(X) = \sigma^2$.

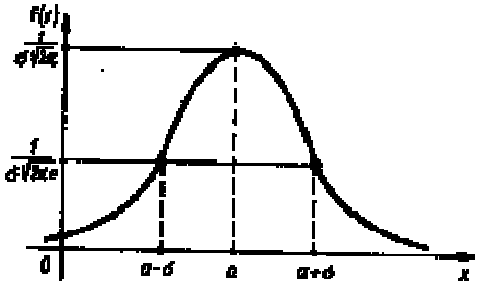
Выясним геометрический смысл параметров распределения a и σ. Для этого исследуем поведение функции f(x). График функции f(x) называется нормальной кривой.

Рассмотрим свойства функции f(x):

1°. Областью определения функции f(x) является вся числовая ось.

2°. Функция f{x) может принимать только положительные значения, т. е. f(x}>0.

3°. Предел функции f(x) при неограниченном возрастании |x| равен нулю, т. е. ось OX является горизонтальной асимптотой графика функции.

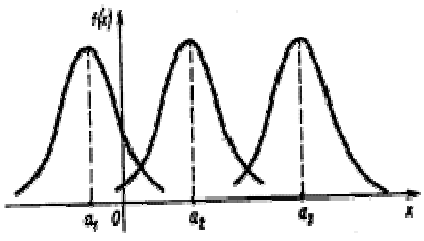


4°. Функция f{x) имеет в точке x = a максимум, равный $\frac{1}{\sigma\sqrt{2\pi}}$.

5°. График функции f(x) симметричен относительно прямой x = a.

6°. Нормальная кривая в точках x = a $\pm\sigma$ имеет перегиб,

$$f(a \pm \sigma) = \frac{1}{\sigma\sqrt{2\pi e}}.$$

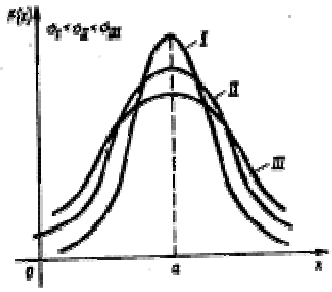


На основании доказанных свойств построим график плотности нормального распределения *f(x)*.

Как видно из рисунка, нормальная кривая имеет колоколообразную форму. Эта форма является отличительной чертой нормального распределения. Иногда нормальную кривую называют *кривой Гаусса.*

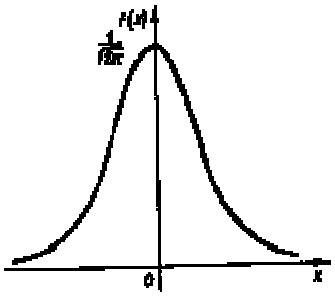
При изменении параметра *a* форма нормальной кривой не изменяется. В этом случае, если математическое ожидание (параметр *a)* уменьшилось или увеличилось, график нормальной кривой сдвигается влево или вправо.

При изменении параметра σ изменяется форма нормальной кривой. Если этот параметр увеличивается, то максимальное значение функции *f(x)* убывает, и наоборот. Так как площадь, ограниченная кривой распределения и осью *Ox,* должна быть постоянной и равной 1, то с увеличением параметра кривая приближается к оси *Ox* и растягивается вдоль нее, а с уменьшением σ кривая стягивается к прямой $x=a$.



Использование формул f(x) и F(x) для практических расчетов затруднительно. Но решение задач по этим формулам можно упростить, если от нормального распределения с произвольными параметрами *a* и σ перейти к нормальному распределению с параметрами $a=0$, $\sigma = 1$.

Функция плотности нормального распределения f(x) с параметрами a=0, σ =1 называется плотностью стандартной нормальной случайной величины и ее график имеет вид:



Функция плотности и интегральная функция стандартной нормальной СВ будут иметь вид:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

Для вычисления вероятности попадания СВ в интервал (α, β) воспользуемся функцией Лапласа: $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{u}e^{-\frac{t^2}{2}}dt$

Перейдем к стандартной нормальной случайной величине $u=\frac{X-a}{\sigma}$.

Тогда $P(\alpha<X<\beta)=P(\frac{\alpha-a}{\sigma}<\frac{X-a}{\sigma}<\frac{\beta-a}{\sigma})=P(u_1,u_2)=\Phi(u_2)-\Phi(u_1).$

Значения функции Φ(u) необходимо взять из таблицы приложений «Таблица значений функции Φ(x)» .

Пример. Случайная величина *X* распределена по нормальному закону. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 30 и 10. Найти вероятность того, что *X* примет значение, принадлежащее интервалу (10, 50).

Решение:

По условию: α =10, β=50, a=30, σ=10, следовательно,

$$P(10<X<50)=\Phi\left(\frac{50-30}{10}\right)-\Phi\left(\frac{10-30}{10}\right)=2\Phi(2).$$

По таблице находим Ф (2) = 0,4772. Отсюда, искомая вероятность:

$P(10 < X < 50) = 2 \cdot 0{,}4772 = 0{,}9544$.

5.2. Вычисление вероятности заданного отклонения

Часто требуется вычислить вероятность того, что отклонение нормально распределенной случайной величины X по абсолютной величине меньше заданного положительного числа δ, т. е. требуется найти вероятность осуществления неравенства $|x - a| < \delta$.

Заменим это неравенство равносильным ему двойным неравенством

$$-\delta < X - a < \delta, \quad \text{или} \quad a - \delta < X < a + \delta.$$

Тогда получим:

$$P(|X - a| < \delta) = P(a - \delta < X < a + \delta) =$$
$$= \Phi\left[\frac{(a+\delta)-a}{\sigma}\right] - \Phi\left[\frac{(a-\delta)-a}{\sigma}\right] = \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right)$$

Приняв во внимание равенство:

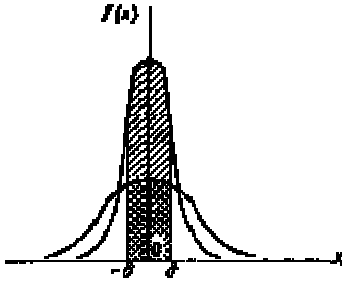
$$\Phi(-\delta/\sigma) = -\Phi(\delta/\sigma)$$

(функция Лапласа—нечетная), окончательно имеем

$$P(|X - a| < \delta) = 2\Phi(\delta/\sigma).$$

Вероятность заданного отклонения равна

На рисунке наглядно показано, что если две случайные величины нормально распределены и $a = 0$, то вероятность принять значение, принадлежащее интервалу (-δ,δ),больше у той величины, которая имеет меньшее значение δ. Этот факт полностью соответствует вероятностному смыслу параметра σ .



Пример. Случайная величина X распределена нормально. Математическое ожидание и среднее квадратическое отклонение X соответственно равны 20 и 10. Найти вероятность того, что отклонение по абсолютной величине будет меньше трех.

Решение: Воспользуемся формулой $P(|X - a| < \delta) = 2\Phi(\frac{\delta}{\sigma})$.

По условию, $\delta = 3, a = 20, \sigma = 10$

тогда $P(|X - 20| < 3) = 2\Phi(\frac{3}{10}) = 2 \cdot 0{,}1179 = 0{,}2358.$

5. Правило трех сигм

Преобразуем формулу $P(|X - a| < \delta) = 2\Phi(\frac{\delta}{\sigma})$. Введем обозначение $t = \frac{\delta}{\sigma}, \delta = t\sigma$. Тогда получим: $P(|X - a| < t\sigma) = 2\Phi(t)$. Если t=3, то

$$P(|X - a| < 3\sigma) = 2\Phi(3) = 2 \cdot 0{,}49865 = 0{,}9973$$

т. е. вероятность того, что отклонение по абсолютной величине будет меньше утроенного среднего квадратического отклонения, равна 0,9973.

Другими словами, вероятность того, что абсолютная величина отклонения превысит утроенное среднее квадратическое отклонение, очень мала, а именно равна 0,0027=1-0,9973. Это означает, что лишь в 0,27% случаев так может произойти. Такие события, исходя из принципа невозможности маловероятных событий, можно считать практически невозможными. В этом и состоит сущность правила трех сигм:

Если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основание предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально.

2.3.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили основные законы распределения ДСВ, НСВ;
- усвоили основные правила нахождения числовых характеристик случайных величин, распределенных по частным законам;
- выработали навыки по вычислению вероятности попадания в интервал ДСВ, НСВ, распределенных по частным законам; по применению свойств специально распределенных случайных величин.

2.4 Практическое занятие № 4 (2 часа)

Тема: «Задачи математической статистики. Статистический материал. Статистические параметры распределения. Статистические оценки параметров распределения»

2.4.1 Задание для работы:

1.Первичная обработка статистических данных.

2. Графическое представление статистических рядов.

3. Эмпирическая функция распределения статистических рядов.

4.Числовые характеристики статистического ряда, их свойства.

2.4.2 Краткое описание проводимого занятия:

1.Первичная обработка статистических данных

Пример1. Записать в виде вариационного и статистического рядов выборку 5, 3, 7, 10, 5, 5, 2, 10, 7, 2, 7, 7, 4, 2, 4. Определить размах выборки.

Решение. В данном случае объем выборки $n = 15$. Упорядочим элементы выборки по величине, получим вариационный ряд 2, 2, 3, 4, 4, 5, 5, 5, 7, 7, 7, 7, 10, 10. Найдем размах выборки ω=10-2= 8. Различными в заданной выборке являются элементы $z_1 = 2$, $z_2 = 3$, $z_3 = 4$, $z_4 = 5$, $z_5 = 7$, $z_6 = 10$; их частоты соответственно равны $n_1 = 3$, $n_2 = 1$, $n_3 = 2$, $n_4 = 3$, $n_5 = 4$, $n_6 = 2$. Статистический ряд исходной выборки можно записать в виде следующей таблицы:

z_i	2	3	4	5	7	10
n_i	3	1	2	3	4	2

Для контроля правильности записи находим $\sum n_i = 15$. При большом объеме выборки ее элементы рекомендуется объединять в группы (разряды), представляя результаты опытов в виде *группированного статистического ряда.* В этом случае интервал, содержащий все элементы выборки, разбивается на k непересекающихся интервалов. Вычисления упрощаются, если эти интервалы имеют одинаковую длину $b \approx \frac{\omega}{k}$. В дальнейшем рассматривается именно этот случай. После того как частичные интервалы выбраны, определяют частоты - количество n_i элементов выборки, попавших в i-й интервал (элемент, совпадающий с верхней границей интервала, относится к следующему интервалу). Получающийся статистический ряд в верхней строке содержит середины z_i интервалов группировки, а в нижней — частоты n_i ($i = 1,2,...,k$).

Наряду с частотами одновременно подсчитываются также накопленные частоты $\sum_{j=1}^{i} n_j$, относительные частоты n_i / n и *накопленные относительные частоты* $\sum_{j=1}^{i} n_j / n$, $i = 1,2,...,k$. Полученные результаты сводятся в таблицу, называемую *таблицей частот группированной выборки.*

Следует помнить, что группировка выборки вносит погрешность в дальнейшие вычисления, которая растет с уменьшением числа интервалов.

Пример 2. Представить выборку 55 наблюдений в виде таблицы частот, разбив имеющиеся данные выборки на семь интервалов группировки. Выборка:

0,3	15,4	17,2	19,2	23,3	18,1	21,9
15,3	16,8	13,2	20,4	16,5	19,7	20,5
14,3	20,1	16,8	14,7	20,8	19,5	15,3
19,3	17,8	16,2	15,7	22,8	21,9	12,5
10,1	21,1	18,3	14,7	14,5	18,1	18,4
13,9	19,1	18,5	20,2	23,8	16,7	20,4
19,5	17,2	19,6	17,8	21,3	17,5	19,4
17,8	13,5	17,8	11,8	18,6	19,1	

В данном случае размах выборки ω=23,8 -10,1 = 13,7; тогда длина интервала группировки будет b = 13,7/7≈2 . В качестве первого интервала возьмем интервал 10 - 12. Результаты группировки сведем в таблицу 1

Таблица 1

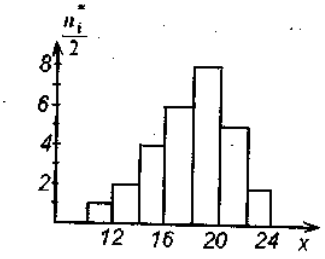
Номер интервала i	Границы интервала	Середина интервала z_i	Частота n_i	Накопленная частота $\sum^{i}$	Относительная частота n_i/n	Накопленная относительная частота
1	10-12	11	2	2	0,0364	0 0364
2	12-14	13	4	6	0,1091	0 1091
3	14-16	15	8	14	0,2546	0 2546
4	16-18	17	12	26	0,4728	0,4728
5	18-20	19	16	42	0,7637	0 7637
6	20-22	21	10	52	0,9455	0,9455
7	22-24	23	3	55	1,0000	1,0000

2. Графическое представление статистических рядов.

Пример 1. Построить гистограмму и полигон частот, а также график эмпирической функции распределения группированной выборки из примера 29.

Решение. По результатам группировки (см. таблицу 1.) строим гистограмму частот (рис. 1). Соединяя отрезками ломаной середины верхних оснований прямоугольников, из которых состоит полученная гистограмма, получаем соответствующий полигон частот (рис. 2).

Так как середина первого интервала группировки $z_1 = 11$, то $F_n^*(x) = 0$ при $x \le 11$. Рассуждая аналогично, находим, что $F_n^*(x) = 1$ при $x > 23$. На полуинтервале (11,23] эмпирическую функцию распределения строим по данным третьего и последнего столбцов таблицы 1.



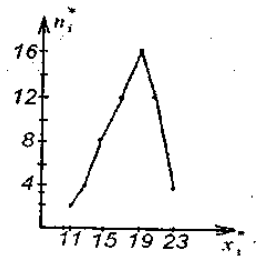


Рис.1 Рис.2

$F_n^*(x)$ имеет скачки в точках, соответствующих серединам интервалов группировки. В результате получаем график $F_n^*(x)$, изображенный на рис. 9.

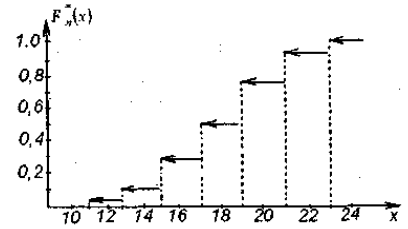


Рис.9

3. Эмпирическая функция распределения статистических рядов

Пусть $(x_1,x_2,...,x_n)$ - выборка из генеральной совокупности с функцией распределения $F_x(x)$. Распределением выборки называется распределение дискретной случайной величины, принимающей значения $x_1,x_2,...,x_n$ с вероятностями $1/n$. Соответствующую функцию распределения называют эмпирической (выборочной) функцией распределения и обозначают $F_n^*(x)$.

Эмпирическую функцию распределения определим по значениям накопленных частот соотношением $F_n^*(x)=\frac{1}{n}\sum_{z_i<x} n_i$, здесь суммируются частоты тех элементов выборки, для которых выполняется неравенство $z_i < x$. Тогда получим, что $F_n^*(x)=0$ при $x \le x^{(1)}$ и $F_n^*(x)=1$ при $x>x^{(n)}$. На промежутке $(x^{(1)}; x^{(n)}]$ $F_n^*(x)$ представляет собой неубывающую кусочно-постоянную функцию.

Аналогично определяем эмпирическую функцию распределения для группированной выборки.

Значение эмпирической функции распределения для статистики определяется следующим утверждением.

Теорема (Гливенко). Пусть $F_n^*(x)$ - эмпирическая функция распределения, построенная по выборке объема n из генеральной совокупности с функцией распределения $F_x(x)$. Тогда для любого $x \in (-\infty,+\infty)$ и любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|F_n^*(x)-F_X(x)\right|<\varepsilon\right)=1 .$$

Таким образом, при каждом x $F_n^*(x)$ сходится по вероятности к $F_x(x)$ и при большом объеме выборки может служить приближенным значением (оценкой) функции распределения генеральной совокупности в каждой точке x.

4.Числовые характеристики статистического ряда, их свойства

Пример 1. Найти формулы, определяющие выборочные математическое ожидание и дисперсию для негруппированной выборки объема n.

Решение. Математическое ожидание дискретной случайной величины определяется по формуле $m_X=\sum_{j=1}^{n} p_j x_j$.

Так как для выборочного распределения $p_j = 1/n$, то $m_X^* = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$.

Аналогично будем иметь выборочную дисперсию

$$D_X^* = \sum_{j=1}^{n}(x_j - \bar{x})^2 p_j = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2 \Big/ n}{n} = \frac{1}{n}\left(\sum_{j=1}^{n} x_j^2 - n\bar{x}^2\right).$$

Выборочной модой d_X^* унимодального (одновершинного) распределения называется элемент выборки, встречающийся с наибольшей частотой.

Выборочной медианой называется число h_X^* , которое делит вариационный ряд на две части, содержащие равное число элементов.

Если объем выборки n — нечетное число (т.е. $n = 2l+1$), то $h_X^* = x^{(l+1)}$, то есть является элементом вариационного ряда со средним номером. Если же $n = 2l$, то $h_X^* = \frac{1}{2}(x^{(l)} + x^{(l+1)})$.

Пример 2. Определить среднее, моду и медиану для выборки 5, 6, 8, 2, 3, 1, 1, 4.

Решение. Представим данные в виде вариационного ряда: 1, 1, 2, 3, 4, 5, 6, 8. Выборочное среднее $\bar{x} = \frac{1}{8}(1+1+2+3+4+5+6+8) = 3{,}75$. Все элементы входят в выборку по одному разу, кроме 1, следовательно, мода $\tilde{d}_X = 1$. Так как $n = 8$, то медиана $\tilde{h}_X = \frac{1}{2}(3+4) = 3{,}5$.

Итак, $\bar{x} = 3{,}75$, $\tilde{d}_X = 1$, $\tilde{h}_X = 3{,}5$.

Для упрощения вычислений выборочных среднего и дисперсии группированной выборки, эту выборку преобразуют так: $u_i = \frac{1}{b}\left(z_i - d_X^*\right)$, $i = 1,2,...,k$, где d_X^* - выборочная мода, а b - длина интервала группировки. Эти соотношения показывают, что в выборку $z_1, z_2, ..., z_n$ внесена систематическая ошибка d_X^*, а результат подвергнут преобразованию масштаба с коэффициентом $k = 1/b$. Полученный в результате набор чисел $u_1, u_2, ..., u_n$ можно рассматривать как выборку из генеральной совокупности $U = \frac{1}{b}\left(x - d_X^*\right)$. Тогда выборочные среднее $\bar{x}$ и дисперсия D_X^* исходных данных связаны со средним $\bar{u}$ и дисперсией D_U^* преобразованных данных следующими соотношениями: $\bar{x} = b\bar{u} + d_X^*$, $D_X^* = b^2 D_U^*$.

Пример 3. Вычислить среднее и дисперсию группированной выборки

Границы интервалов	134-138	138-142	142-146	146-150	150-154	154-158
Частоты	1	3	15	18	14	2

Решение. Длина интервала группировки $b = 4$, значение середины интервала, встречающегося с наибольшей частотой $d_X^* = 148$. Преобразование последовательности середин интервалов выполняется по формуле:

$u_i = \frac{z_i - 148}{4}$, где $i = 1,2,...,6$.

Таблица 2

i	z_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i+1)^2$
1	136	-3	1	-3	9	4
2	140	-2	3	-6	12	3
3	144	-1	15	-15	15	0
4	148	0	18	0	0	18
5	152	1	14	14	14	56
6	156	2	2	4	8	18
	-	-	5	-	5	99

Вычисления сведены в таблицу 2. Последний столбец этой таблицы служит для контроля вычислений при помощи тождества $\sum n_i(u_i+1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + \sum n_i$.

Выполняя вычисления, получим $58 + 2 \cdot (-6) + 23 = 99$.

Полученный результат показывает, что вычисления выполнены правильно. По формулам, данным выше, находим средние значения U

$\bar{u} = \frac{-6}{53} \approx -0{,}113$, $D_U^* = \frac{58 - (-6)^2/53}{53} \approx 1{,}108$.

Далее находим средние данной выборки:

$\bar{x} \approx (-0{,}113) \cdot 4 + 148 \approx 147{,}548$, $D_X^* \approx 4^2 \cdot 1{,}103 \approx 17{,}728$.

2.4.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили первичную обработку статистических данных, ее графическое представление;
- усвоили основные методы нахождения точечных оценок параметров статистического распределения;
- выработали навыки по оценке параметров генеральной совокупности, применению метода доверительных интервалов.

2.5 Практическое занятие № 5 (2 часа)

Тема: «Интервальные оценки параметров статистического распределения. Необходимость их введения. Доверительные интервалы. Доверительные вероятности. Доверительные интервалы для оценки математического ожидания нормального распределения. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения»

2.5.1 Задание для работы:

1. Точечные оценки параметров статистического распределения.

2. Оценки параметров генеральной совокупности. Метод моментов.

3.Метод доверительных интервалов.

2.5.2 Краткое описание проводимого занятия:

1. Точечные оценки параметров статистического распределения

Пример 1. Пусть $x_1,x_2,...,x_n$ - выборка из генеральной совокупности с конечными математическим ожиданием и дисперсией σ^2. Используя метод подстановки, найти оценку m. Проверить свойства несмещенности и состоятельности полученной оценки.

Решение. По методу подстановки в качестве оценки m математического ожидания возьмем математическое ожидание распределения выборки - выборочное среднее. Тогда, получим

$$\tilde{m} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Для проверки несмещенности и состоятельности выборочного среднего как оценки m, рассмотрим эту статистику как функцию выборочного вектора $(X_1,X_2,...,X_n)$. По определению выборочного вектора имеем: $M[X_i]=m$ и $D[X_i]=\sigma^2$, $i = 1,2,...,n$, причем X_i - независимые в совокупности случайные величины.

В данном случае будем иметь

$$M[\overline{X}] = M\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n}\cdot nm = m ,$$

$$D[\overline{X}] = D\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} M[X_i] = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} .$$

Отсюда по определению получаем, что $\overline{X}$ - несмещенная оценка m, и так как $D[\overline{X}] \to 0$ при $n\to\infty$, то в силу теоремы 1 $\overline{X}$ является состоятельной оценкой математического ожидания m генеральной совокупности.

2. Оценки параметров генеральной совокупности. Метод моментов

Для получения оценок неизвестных параметров $\theta_1,\theta_2,...,\theta_s$ распределения генеральной совокупности X используется и метод моментов. Поясним его.

Пусть $f_X(x,\theta_1,\theta_2,...,\theta_s)$ - плотность распределения случайной величины X. Определим с помощью этой плотности S каких-либо моментов случайной величины X, например, первые S начальных моментов, по формулам

$$\alpha_m(\theta_1,...,\theta_s) = M[X^m] = \int_{-\infty}^{+\infty} x^m f_X(x,\theta_1,...,\theta_s)dx , \; m = 1,2,...,S.$$

По выборке наблюдений случайной величины найдем значения соответствующих выборочных моментов:

$$\alpha_m^* = \frac{1}{n}\sum_{i=1}^{n} x_i^m , \; m = 1,2,...,S.$$

Попарно приравнивая теоретические моменты α_m случайной величины X их выборочным значениям α_m^*, получаем систему s уравнений с неизвестными $\theta_1,\theta_2,...,\theta_s$:

$$\alpha_m(\theta_1,...,\theta_s) = \alpha_m^* , \; m = 1,2,...,S.$$

Решая полученную систему относительно неизвестных $\theta_1,\theta_2,...,\theta_s$, находим оценки $\tilde{\theta}_1,...,\tilde{\theta}_s$ неизвестных параметров.

Аналогично находятся оценки неизвестных параметров по выборке наблюдений дискретной случайной величины.

Пример 1. Методом моментов найти оценки неизвестных параметров a и b для Г - распределения с плотностью

$$f_X(x)=\begin{cases}0, & x\le 0\\ \dfrac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}, & x>0\end{cases}.$$

Решение. Для нахождения оценок параметров *a* и *b* по методу моментов воспользуемся начальным моментом первого порядка (математическим ожиданием) и центральным моментом второго порядка (дисперсией):

$$\alpha_1(a,b)=m=\frac{a}{b},\ \mu_2(a,b)=\sigma^2=\frac{a}{\sigma^2}.$$

По выборке $x_1,\dots,x_n$ из генеральной совокупности, имеющей Г-распределение, находим значения соответствующих выборочных моментов:

$$\alpha_1^*=\bar{x}=\frac{1}{n}\sum x_i,\ \mu_1^*=D_X^*=\frac{1}{n}\sum(x_i-\bar{x})^2.$$

Приравнивая соответствующие равенства, получаем следующую систему уравнений:

$\frac{a}{b}=\bar{x}$, $\frac{a}{b^2}=D_X^*$. Решая ее, находим $\tilde{a}=\frac{\overline{x^2}}{D_X^*}$, $\tilde{b}=\frac{\bar{x}}{D_X^*}$.

3. Метод доверительных интервалов

Пример 1. Пусть $x_1,x_2,\dots,x_n$ - выборка из нормально распределенной генеральной совокупности. Найти доверительный интервал для математического ожидания *m* при условии, что дисперсия генеральной совокупности известна и равна σ^2, а доверительная вероятность равна 1-α.

Решение. В качестве оценки математического ожидания *m* возьмем выборочное среднее $\bar{x}=\frac{1}{n}\sum x_i$. Для нормально распределенной генеральной совокупности выборочное среднее является эффективной оценкой *m*. Выборочное среднее $\overline{X}$ в данном случае имеет нормальное распределение $N(m,\sigma/\sqrt{n})$.

Рассмотрим статистику $U=\dfrac{\overline{X}-m}{\sigma/\sqrt{n}}$, имеющую нормальное распределение $N(0,1)$ независимо от значения параметра *m*. Кроме того, *U* как функция *m* непрерывна и строго монотонна. Тогда $P[u_{\alpha/2}<U<u_{1-\alpha/2}]=1-\alpha$, где $u_{a/2}$ и $u_{1-a/2}$ - квантили нормального распределения $N(0,1)$.

Решая неравенство $u_{\alpha/2}<\dfrac{\overline{X}-m}{\sigma/\sqrt{n}}<u_{1-\alpha/2}$ относительно *m,* получим, что с вероятностью 1-α выполняется условие: $\overline{X}-\dfrac{\sigma}{\sqrt{n}}u_{1-\alpha/2}<m<\overline{X}-\dfrac{\sigma}{\sqrt{n}}u_{\alpha/2}$.

Так как квантили нормального распределения связаны соотношением $u_{a/2}=-u_{1-a/2}$, полученный доверительный интервал для *m* можно записать следующим образом: $\overline{X}-\dfrac{\sigma}{\sqrt{n}}u_{1-\alpha/2}<m<\overline{X}+\dfrac{\sigma}{\sqrt{n}}u_{1-\alpha/2}$

Пример 2. При проверке 100 деталей из большой партии обнаружено 10 бракованных деталей.

а) Найти 95 % приближенный доверительный интервал для доли бракованных деталей во всей партии.

б) Какой минимальный объем выборки следует взять для того, чтобы с вероятностью 0,95 можно было утверждать, что доля бракованных деталей по всей партии отличается от частоты появления бракованных деталей в выборке не более чем на 1 %?

Решение:

а) Оценка доли бракованных деталей в партии по выборке равна $\tilde{p} = h = 10/100 = 0{,}1$. По таблице приложений (П1) находим квантиль $u_{1-a/2} = u_{0{,}975} = 1{,}96$. Тогда 95% доверительный
интервал для доли бракованных деталей в партии приближенно имеет вид $0{,}041 < p < 0{,}159$.

б) Представим полученный доверительный интервал в виде неравенства $|h-p| < u_{1-\alpha/2}\sqrt{\frac{h(1-h)}{n}}$, которое выполняется с вероятностью ≈1 - α = 0,95. Так как согласно условию задачи $|h-p| \le 0{,}01$, то для определения n получим неравенство $u_{0{,}975}\sqrt{\frac{h(1-h)}{n}} \le 0{,}01$. Отсюда следует, что $1{,}96\sqrt{\frac{0{,}1(1-0{,}1)}{n}} \le 0{,}01$ и $n \ge (0{,}3\cdot 196)^2 = 3457{,}44$. Итак, минимальный объем выборки $n = 3458$.

2.5.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили первичную обработку статистических данных, ее графическое представление;
- усвоили основные методы нахождения точечных оценок параметров статистического распределения;
- выработали навыки по оценке параметров генеральной совокупности, применению метода доверительных интервалов.

2.6 Практическое занятие № 6 (2 часа)

Тема: «Понятие статистической гипотезы. Виды гипотез. Статистический критерий. Критическая область. Мощность критерия. Критерии согласия: критерий Пирсона. Выравнивание рядов»

2.6.1 Задание для работы:

1. Статистические гипотезы и их виды.
2. Критерии согласия.
3. Оценка параметров неизвестного распределения.
4. Выравнивание рядов.

2.6.2 Краткое описание проводимого занятия:

1. Статистические гипотезы и их виды

Пример 50. По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем, причем выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило $\bar{x} = 9{,}3$ л. Предполагается, что выборка расходов топлива получена из нормально распределенной генеральной совокупности со средним m и дисперсией $\sigma^2 = 4$ л^2 . Используя критерий значимости, проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива.

Решение. Проверяется гипотеза о среднем (m) нормально распределенной генеральной совокупности. Проверку гипотезы проведем по этапам:

1) проверяемая гипотеза H_0: $m = 10$, альтернативная гипотеза H_1: $m < 10$;

2) выберем уровень значимости $\alpha = 0{,}05$;

3) в качестве статистики критерия используем оценку математического ожидания - выборочное среднее $\overline{X}$;

4) так как выборка получена из нормально распределенной генеральной совокупности, выборочное среднее также имеет нормальное распределение с дисперсией $\frac{\sigma^2}{n} = \frac{4}{25}$. При условии, что верна гипотеза H_0, математическое ожидание этого распределения равно 10. Нормированная статистика критерия $U = \frac{\overline{X} - 10}{\sqrt{4/25}}$ имеет нормальное распределение $N(0,1)$.

5) альтернативная гипотеза H_1: $m < 10$ предполагает уменьшение расхода топлива, следовательно, нужно использовать односторонний критерий. Критическая область определяется неравенством $U < u_a$. По таблице приложений П1 находим $u_{0,05} = -u_{0,95} = -1{,}645$;

6) выборочное значение нормированной статистики критерия равно $U = \frac{9{,}3 - 10}{\sqrt{4/25}} = -1{,}75$.

7) статистическое решение: так как выборочное значение статистки критерия принадлежит критической области, гипотеза H_0 отклоняется: следует считать, что изменение конструкции двигателя привело к уменьшению расхода топлива.

Граница $\overline{x_k}$ критической области для исходной статистики X критерия может быть получена из соотношения $\frac{\overline{x_k} - 10}{\sqrt{4/25}} = -1{,}75$, откуда получаем $\overline{x_k}$ = 9,342, т. е. критическая область для статистики X определяется неравенством $\overline{X} < 9{,}342$.

2. Критерии согласия

Пример 1. В первых - двух столбцах таблицы 1 приведены данные об отказах аппаратуры за 10000 часов работы. Общее число обследованных экземпляров аппаратуры n=757, при этом наблюдался отказ:0·427+ 1·235+ 2·72 + 3·21+ 4·1+ 5·1 = 451

Таблица 1

Число отказов, k	Количество случаев, в которых наблюдалось k отказов, n_k	$p_k = \frac{0{,}6^k}{k!} e^{-0{,}6}$	Ожидаемое число случаев с k отказами, np_k
0	427	0,54881	416
1	235	0,32929	249
2	72	0,09879	75
3	21	0,01976	15
4	1	0,00296	2
5	1	0,00036	0
≥6	0	0,00004	0
Сумма	757	-	-

Проверить гипотезу о том, что число отказов имеет распределение Пуассона:

$p_k = P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$, k=0, 1, …, при α=0,01.

Решение. Оценка параметра λ равна среднему числу отказов: $\overline{\lambda} = 451/757 \approx 0{,}6$. По таблице приложений (П3) с λ = 0,6 находим вероятности p_k и ожидаемое число случаев с k отказами (третий и четвертый столбцы таблицы 2).

Для k = 4,5 и 6 значения $np_k < 5$, поэтому объединяем эти строки со строкой для k= 3 . Итак, получаем значения, приведенные в таблице 1.

Таблица 2

k	n_k	np_k	$\frac{(n_k - np_k)^2}{np_k}$
0	*427*	416	0,291
1	235	249	0,787
2	72	75	0,120
≥3	23	17	2,118
-	-	-	$\chi^2_B = 3{,}316$

Так как по выборке оценивался один параметр λ, то l = 1, число степеней свободы равно 4-1-1 = 2. По таблице приложений (П5) находим $\chi^2_{0,99}(2) = 9{,}21$, гипотеза о распределении числа отказов по закону Пуассона принимается.

Пример 2. Проверить гипотезу о нормальном распределении выборки из примера: По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 55 случайно отобранных автомобилей с модернизированным двигателем, причем выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило $\overline{x}$ = 9,3 л. Предполагается, что выборка расходов топлива получена из нормально распределенной генеральной совокупности со средним m и дисперсией $\sigma^2 = 4$ л^2 . Принять $\alpha = 0{,}1$.

Решение. Объем выборки n = 55. Для проверки гипотезы о нормальном распределении нужно найти оценки математического ожидания и дисперсии. Имеем $\tilde{m} = \overline{x} = \frac{1}{n}\sum_{i=1}^{55} x_i \approx 17{,}84$, $\tilde{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{55}(x_i - \overline{x})^2 \approx 8{,}53$. Результаты группировки приведены во втором и третьем столбцах таблицы 3.

Таблица 3

Номер интервала k	Границы интервала Δ_k	Наблюдаемая частота n_k	Вероятность попадания в интервал Δ_k, p_k	Ожидаемая частота, np_k	np_k	n_k - np_k	$\frac{(n_k - np_k)^2}{np_k}$
	- ∞ -						
1	12	2	0 0228	1,254	5,274	0,725	0,010
2	12-14	4	0,0731	4,020			
3	14-16	8	0,1686	9,273	9,273	-1,273	0,175
4	16-18	12	0,2576	14,168	14,168	-2,168	0,332
5	18-20	16	0,2484	13,662	13,662	-2,338	0,400
6	20-22	10	0,1519	8,354	12,633	0,366	0,011
7	22- + ∞	3	0,0778	4,279			
	Сумма	55	1,0001	55	55	-	0,928

В четвертом столбце таблицы 3 приведены вероятности p_k, вычисляемые по формуле:

$$p_k = P[X \in \Delta_k] = \Phi\left(\frac{b_k - \bar{x}}{s}\right) - \Phi\left(\frac{a_k - \bar{x}}{s}\right), k = 1,2,...,7,$$

где $a_к$ и $b_к$ - соответственно нижняя и верхняя границы интервалов, а значения функции *Ф(х)* берутся из таблицы приложений (П1). В пятом столбце приводятся ожидаемые частоты np_k, а в шестом - значения np_k после объединения первых двух и последних двух интервалов.

Так как после объединения осталось $r = 5$ интервалов, а по выборке определены оценки двух параметров, т.е. $l = 2$, то число степеней свободы равно 5-2-1 = 2.

По таблице приложений (П5) находим $\chi^2_{0,90}(2) = 4{,}61$. Выборочное значение статистики критерия равно $\chi^2_В = 0{,}928$, следовательно, гипотеза о нормальном распределении выборки принимается.

3. Оценка параметров неизвестного распределения.

Метод максимального правдоподобия

Метод максимального правдоподобия является одним из наиболее распространенных методов нахождения оценок неизвестных параметров распределения генеральной совокупности. Пусть Х- непрерывная случайная величина с плотностью распределения $f_x(x, \theta)$, зависящей от неизвестного параметра θ, значение которого требуется оценить по выборке объема п. Плотность распределения выборочного вектора $(X_1,X_2,...,X_n)$ можно записать в виде $f_{X_1,...,X_n}(x_1,...,x_n,\theta) = \prod_{i=1}^{n} f_{X_1}(x_i,\theta)$.

Пусть $x_1,x_2,...,x_n$ - выборка наблюдений случайной величины Х, по которой находится оценка неизвестного параметра.

Функцией правдоподобия L(θ) выборки объема п называется плотность выборочного вектора, рассматриваемая при фиксированных значениях переменных $x_1,...,x_п$. Функция правдоподобия является, таким образом, функцией только неизвестного параметра θ, т.е. $L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i,\theta)$.

Аналогично определим функцию правдоподобия выборки дискретной случайной величины Х. Пусть Х- дискретная случайная величина, причем вероятность P[X = x]= p(x, θ) есть функция неизвестного параметра θ. Предполагая, что для оценки параметра θ получена конкретная выборка наблюдений случайной величины Х объема п: $x_1,...,x_п$. Функция правдоподобия L(θ) выборки объема п равна вероятности того, что компоненты выборочного вектора $X_1,...,X_n$ примут фиксированные значения $x_1,...,x_п$, т.е. $L(\theta) = \prod_{i=1}^{n} P[X_i = x_i] = \prod_{i=1}^{n} p(x_i,\theta)$.

Метод максимального правдоподобия состоит в том, что в качестве оценки неизвестного параметра θ принимается значение $\tilde{\theta}$, доставляющее максимум функции правдоподобия. Такую оценку называют МП - оценкой. В случае дискретного распределения наблюдаемой случайной величины Х МП - оценка неизвестного параметра θ есть такое значение $\tilde{\theta}$, при котором вероятность появления данной конкретной выборки максимальна. Аналогичную интерпретацию МП - оценки дают и в случае оценки параметра распределения непрерывной случайной величины.

Для упрощения вычислений, связанных с получением МП -оценок, в некоторых случаях удобно использовать логарифмическую функцию правдоподобия, т.е. ln L(θ).

При выполнении некоторых достаточно общих условий МП - оценки состоятельны, асимптотически эффективны и асимптотически нормально распределены. Последнее оз-

начает, что при увеличении объема выборки п для МП - оценки $\tilde{\theta}_n$ неизвестного параметра θ выполняется условие $\lim\limits_{n\to\infty} P\left[\dfrac{\theta-\tilde{\theta}_n}{D[\tilde{\theta}_n]}<x\right]=\dfrac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-t^2/2}dt=\Phi(x)$.

Если для параметра θ существует эффективная оценка, то метод максимального правдоподобия дает именно эту оценку и другой МП - оценки не существует.

Пример 1. Найти МП - оценки математического ожидания т и дисперсии σ^2 нормально распределенной генеральной совокупности.

Решение. Пусть $x_1,x_2,...,x_п$ - выборка наблюдений случайной величины X с плотностью распределения

$$f_x(x,m,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(x-m)^2}{2\sigma^2}}.$$

Найдем функцию правдоподобия $L(m, \sigma^2)$. Имеем

$$L(m,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(x_i-m)^2}{2\sigma^2}}=\frac{1}{(2\pi)^{n/2}\sigma^n}e^{\frac{\sum(x_i-m)^2}{2\sigma^2}}.$$

Логарифмическая функция правдоподобия отсюда равна

$$\ln L(m,\sigma^2)=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{\sum(x_i-m)^2}{2\sigma^2}.$$

Используя необходимые условия максимума $\ln L(m,\sigma^2)$, получим систему уравнений для нахождения искомых МП - оценок:

$$\frac{\partial\ln L(m,\sigma^2)}{\partial m}=\frac{1}{\sigma^2}\sum(x_i-m)=0,\ \frac{\partial\ln L(m,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum(x_i-m)=0.$$

Из первого уравнения этой системы находим $\tilde{m}=\dfrac{1}{n}\sum x_i=\bar{x}$.

Подставляя полученное значение во второе уравнение, будем иметь $\tilde{\sigma}^2=\dfrac{1}{n}\sum(x_i-\bar{x})^2=D_X^*$.

Отметим, что выборочное среднее $\bar{x}$ является несмещенной и состоятельной оценкой т (см. пример 38), а также эффективной оценкой в случае нормально распределенной генеральной совокупности (убедитесь в этом самостоятельно). Выборочная дисперсия D_X^* является состоятельной и смещенной оценкой σ^2.

Пример 2. Найти МП - оценку параметра X распределения Пуассона.

Решение. Пусть $x_1,...,x_п$ - выборка наблюдений случайной величины X, имеющей распределение Пуассона с неизвестным параметром X, т.е.

$$P[X=x]=\frac{\lambda^k}{x!}e^{-\lambda},$$

где х принимает неотрицательные целочисленные значения, х = 0,1,2. Функция правдоподобия $L(\lambda)$ выборки объема п определяется так: $L(\lambda)=\prod\limits_{i=1}^{n}\dfrac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\dfrac{\lambda^{\sum x_i}}{x_1!x_2!...x_n!}e^{-\lambda_n}$.

Найдем логарифмическую функцию правдоподобия:

$$\ln L(\lambda)=-\ln(x_1!x_2!...x_n!)+(\sum x_i)\ln\lambda-\lambda n.$$

Используя необходимое условие экстремума, получим уравнение для определения МП-оценки:

$\dfrac{d\ln L(\lambda)}{d\lambda}=\dfrac{\sum x_i}{\lambda}-n=0$. Отсюда следует, что $\tilde{\lambda}=\dfrac{1}{n}\sum x_i=\bar{x}$.

Полученная МП - оценка является несмещенной и состоятельной оценкой λ, а также эффективной оценкой этого параметра.

4. Выравнивание рядов

Во всяком статистическом распределении неизбежно присутствуют элементы случайности, связанные с тем, что число наблюдений ограничено, что произведены именно те, а не другие опыты, давшие именно те, а не другие результаты. Только при очень большом числе наблюдений эти элементы случайности сглаживаются, и случайное явление обнаруживает в полной мере присущую ему закономерность. На практике мы почти никогда не имеем дела с таким большим числом наблюдений и вынуждены считаться с тем, что любому статистическому распределению свойственны в большей или меньшей мере черты случайности. Поэтому при обработке статистического материала часто приходится решать вопрос о том, как подобрать для данного статистического ряда теоретическую кривую распределения, выражающую лишь существенные черты статистического материала, но не случайности, связанные с недостаточным объемом экспериментальных данных. Такая задача называется задачей выравнивания (сглаживания) статистических рядов.

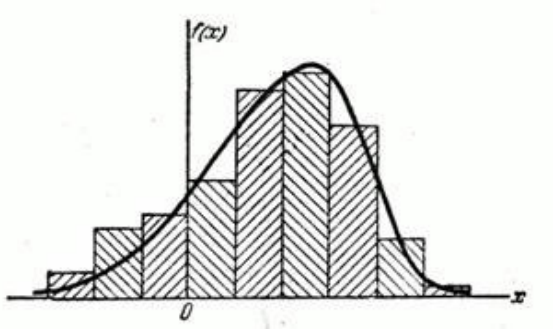


Рис. 1

Задача выравнивания заключается в том, чтобы подобрать теоретическую плавную кривую распределения, с той или иной точки зрения наилучшим образом описывающую данное статистическое распределение (рис. 1).

Задача о наилучшем выравнивании статистических рядов, как и вообще задача о наилучшем аналитическом представлении эмпирических функций, есть задача в значительной мере неопределенная, и решение ее зависит от того, что условиться считать «наилучшим». Например, при сглаживании эмпирических зависимостей очень часто исходят из так называемого принципа или метода наименьших квадратов), считая, что наилучшим приближением к эмпирической зависимости в данном классе функций является такое, при котором сумма квадратов отклонений обращается в минимум. При этом вопрос о том, в каком именно классе функций следует искать наилучшее приближение, решается уже не из математических соображений, а из соображения, связанных с физикой решаемой задачи, с учетом характера полученной эмпирической кривой и степени точности произведенных наблюдений. Часто принципиальный характер функции, выражающей исследуемую зависимость, известен заранее из теоретических соображении, из опыта же требуется получить лишь некоторые численные параметры, входящие в выражение функции; именно эти параметры подбираются с помощью метода наименьших квадратов.

Аналогично обстоит дело и с задачей выравнивания статистических рядов. Как правило, принципиальный вид теоретической кривой выбирается заранее из соображений, связанных с существом задачи, а в некоторых случаях просто с внешним видом статистического распределения. Аналитическое выражение выбранной кривой распределения зависит от некоторых параметров; задача выравнивания статистического ряда переходит в задачу рационального выбора тех значений параметров, при которых соответствие между статистическим и теоретическим распределениями оказывается наилучшим.

Предположим, например, что исследуемая величина X есть ошибка измерения, возникающая в результате суммирования воздействий множества независимых элементарных ошибок; тогда из теоретических соображений можно считать, что величина X подчиняется нормальному закону:

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

и задача выравнивания переходит в задачу о рациональном выборе параметров m и σ в выражении (1).

Бывают случаи, когда заранее известно, что величина X распределяется статистически приблизительно равномерно на некотором интервале; тогда можно поставить задачу о рациональном выборе параметров того закона равномерной плотности

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{при } \alpha<x<\beta,\\ 0 & \text{при } x<\alpha \text{ или } x>\beta\end{cases}$$

которым можно наилучшим образом заменить (выровнять) заданное статистическое распределение.

Следует при этом иметь в виду, что любая аналитическая функция $f(x)$, с помощью которой выравнивается статистическое распределение, должна обладать основными свойствами плотности распределения:

$$\left.\begin{array}{l} f(x)\geq 0;\\ \displaystyle\int_{-\infty}^{\infty} f(x)\,dx=1\end{array}\right\} \quad (2)$$

Предположим, что, исходя из тех или иных соображений, нами выбрана функция $f(x)$, удовлетворяющая условиям (2), с помощью корой мы хотим выровнять данное статистическое распределение; в выражение этой функции входит несколько параметров $a,b,\ldots$; требуется подобрать эти параметры так, чтобы функция $f(x)$ наилучшим образом описывала данный статистический материал. Один из методов, применяемых для решения этой задачи, - это так называемый метод моментов.

Согласно методу моментов, параметры $a,b,\ldots$ выбираются с таким расчетом, чтобы несколько важнейших числовых характеристик (моментов) теоретического распределения были равны соответствующим статистическим характеристикам. Например, если теоретическая кривая $f(x)$ зависит только от двух параметров a и b, эти параметры выбираются так, чтобы математическое ожидание m_x и дисперсия D_x^* теоретического распределения совпадали с соответствующими статистическими характеристиками m_x^* и D_x^*. Если кривая $f(x)$ зависит от трех параметров, можно подобрать их так, чтобы совпали первые три момента и т.д. При выравнивании статистических рядов может оказаться полезной специально разработанная система кривых Пирсона, каждая из которых зависит в общем случае от четырех параметров. При выравнивании эти параметры выбираются с тем расчетом, чтобы сохранить первые четыре момента статистического распределения (математическое ожидание, дисперсию, третий и четвертый моменты). Следует заметить, что при выравнивании статистических рядов нерационально пользоваться моментами порядка выше четвертого, так как точность вычисления моментов резко падает с увеличением их порядка.

Пример. С целью исследования закона распределения ошибки измерения дальности с помощью радиодальномера произведено 400 измерений дальности. Результаты опытов представлены в виде статистического ряда:

l_i	20;	30;	40;	50;	60;	70;	80;	90;
m_i	21	72	66	38	51	56	64	32
p_i^*	0,0	0,1	0,1	0,0	0,1	0 ,140	0,1	0,0

Выровнять статистический ряд с помощью закона равномерной плотности.

Решение. Закон равномерной плотности выражается формулой

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{при } \alpha < x < \beta,\\ 0 & \text{при } x < \alpha \text{ или } x > \beta\end{cases}$$

и зависит от двух параметров α и β. Эти параметры следует выбрать так, чтобы сохранить первые два момента статистического распределения – математическое ожидание m_x^* и дисперсию D_x^*. Из примера n^O 5.8 имеем выражения математического ожидания и дисперсии для закона равномерной плотности:

$$m_x=\frac{\alpha+\beta}{2};$$

$$D_x=\frac{(\beta-\alpha)^2}{12}.$$

Для того, чтобы упростить вычисления, связанные с определением статистических моментов, перенесем начало отсчета в точку $x_0 = 60$ и примем за представителя его разряда его середину. Ряд распределения имеет вид:

$\tilde{x}_i'$	−3;	−2;	−1	−5	5	15	25	35
p_i^*	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,0

где $\tilde{x}_i'$ - среднее для разряда значение ошибки радиодальномера X' при новом начале отсчета. Приближенное значение статистического среднего ошибки X' равно:

$$m_{x'}^*=\sum_{i=1}^{k}\tilde{x}_i' p_i^*=0,26$$

Второй статистический момент величины X' равен:

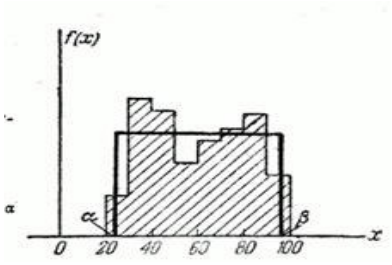
$$\alpha_2^*=\sum_{i=1}^{k}(\tilde{x}_i')^2 p_i^*=447,8$$

, откуда статистическая дисперсия: $D_{x'}^*=\alpha_2^*-\left(m_{x'}^*\right)^2=447,7$. Переходя к прежнему началу отсчета, получим новое статистическое среднее:

$$m_x^*=m_{x'}^*+60=60,26$$

в ту же статистическую дисперсию:

$$D_x^*=D_{x'}^*=447,7.$$



Параметры закона равномерной плотности определяются уравнениями:

$$\frac{\alpha+\beta}{2}=60,26;\quad \frac{(\beta-\alpha)^2}{12}=447,7.$$

Решая эти уравнения относительно α и β, имеем: $\alpha\approx 23,6;\quad \beta\approx 96,9$, Рис.2

откуда $$f(x)=\frac{1}{\beta-\alpha}=\frac{1}{73,3}\approx 0,0136.$$

На рис. 2. показаны гистограмма и выравнивающий ее закон равномерной плотности $f(x)$.

2.6.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили классификацию статистических гипотез;
- усвоили основные правила применения критерия согласия;
- выработали навыки по оценке параметров неизвестного распределения, выравниванию рядов.

2.7 Практическое занятие № 7, 8 (4 часа).

Тема: «Понятие функциональной, стохастической и корреляционной зависимости. Функция регрессии. Корреляционное отношение. Его свойства, значимость. Линейная функция регрессии. Коэффициент корреляции его»

2.7.1 Задание для работы:

1. Виды зависимостей между величинами.

2. Функция регрессии.

3. Корреляционное отношение.

4. Линейная парная регрессия.

5. Коэффициент корреляции, его свойства, значимость.

2.7.2 Краткое описание проводимого занятия:

1. Виды зависимостей между величинами

Между различными явлениями и их признаками необходимо прежде всего выделить два типа связей: функциональную (жестко детерминированную) и статистическую (стохастическую детерминированную).

Связь признака y с признаком x называется функциональной, если каждому возможному значению независимого признака x соответствует одно или несколько строго определенных значений зависимого признака y. Определение функциональной связи может быть легко обобщено для случая многих признаков $x_1,x_2,\ldots,x_n$.

Характерной особенностью функциональных связей является то, что в каждом отдельном случае известен полный перечень факторов, определяющих значение зависимого (результативного) признака, а также точный механизм их влияния, выраженного определенным уравнением.

Стохастическая связь- это связь между величинами, при которых одна из них, случайная величина y, реагирует на изменение другой величины x или других величин x_1, $x_2,\ldots$, x_n, (случайных или неслучайных) изменением закона распределения. Это обуславливается тем, что зависимая переменная (результативный признак), кроме рассматриваемых независимых, подвержена влиянию ряда неучтенных или неконтролируемых (случайных) факторов, а также некоторых неизбежных ошибок измерения переменных. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только указаны с определенной вероятностью.

Характерной особенностью стохастических связей является то, что они проявляются во всей совокупности, а не в каждой ее единице (причем не известен ни полный перечень факторов, определяющих значение результативного признака, ни точный механизм их функционирования и взаимодействия с результативным признаком). Всегда имеет место влияние случайного. Появляющиеся различные значения зависимой переменной - реализации случайной величины.

2. Функция регрессии

Условимся обозначать через X независимую переменную, а через Y – зависимую переменную.

Зависимость величины Y от X называется **функциональной**, если каждому значению величины X соответствует единственное значение величины Y. С функциональной зависимостью мы встречаемся, например, в математике, при изучении физических законов. Обратим внимание на то, что если X – детерминированная величина (т.е. принимающая вполне определённые значения), то и функционально зависящая от неё величина Y тоже является детерминированной; если же X – случайная величина, то и Y также случайная величина.

Однако гораздо чаще в окружающем нас мире имеет место не функциональная, а **стохастическая**, или **вероятностная**, **зависимость**, когда каждому фиксированному значению независимой переменной X соответствует не одно, а множество значений переменной Y, причём сказать заранее, какое именно значение примет величина Y, нельзя. Более частое появление такой зависимости объясняется действием на результирующую переменную не только контролируемого или контролируемых факторов (в данном случае таким контролируемым фактором является переменная X), а и многочисленных неконтролируемых случайный факторов. В этой ситуации переменная Y является случайной величиной. Переменная же X может быть как детерминированной, так и случайной величиной. Следует заметить, что со стохастической зависимостью мы уже сталкивались в дисперсионном анализе.

Допустим, что существует стохастическая зависимость случайной переменной Y от X. Зафиксируем некоторое значение x переменной X. При X = x переменная Y в силу её стохастической зависимости от X может принять любое значение из некоторого множества, причём какое именно – заранее неизвестно. Среднее этого множества называют **групповым генеральным средним** переменной Y при X = x или **математическим ожиданием случайной величины Y, вычисленным при условии, что X = x; это условное математическое ожидание обозначают так: M(Y/X = x)**. Если существует стохастическая зависимость Y от X, то прежде всего стараются выяснить, изменяются или нет при изменении x условные математические ожидания M(Y/X=x). Если при изменении x условные математические ожидания M(Y/X=x) изменяются, то говорят, что имеет место **корреляционная зависимость** величины Y от X; если же условные математические ожидания остаются неизменными, то говорят, что корреляционная зависимость величины Y от X отсутствует.

Функция $\varphi(x)=M(Y/X=x)$, описывающая изменение условного математического ожидания случайной переменной Y при изменении значений x переменной X, называется **функцией регрессии**.

Выясним, почему именно при наличии стохастической зависимости интересуются поведением условного математического ожидания.

Рассмотрим пример. Пусть X – уровень квалификации рабочего, Y – его выработка за смену. Ясно, что зависимость Y от X не функциональная, а стохастическая: на выработку помимо квалификации влияет множество других факторов. Зафиксируем значение x уровня квалификации: ему соответствует некоторое множество значений выработки Y. Тогда M(Y/X = x) – средняя выработка рабочего при условии, что его уровень квалификации равен x, или, иначе говоря, M(Y/X = x) – это норматив выработки при уровне квалификации, равном x. Зная зависимость этого норматива от уровня квалификации, можно для любого уровня квалификации рассчитать норматив выработки и, сравнив его с реальной выработкой, оценить работу рабочего.

Обратим внимание на то, что введённые понятия стохастической и корреляционной зависимости относились к генеральной совокупности. Поясним эти понятия числовым примером.

Пример. Допустим, что одновременно изучаются две случайные величины X и Y, или, иначе говоря, двумерная случайная величина (X, Y), которая задана табл. 1.

Таблица 1.

x_i / y_i	x_1 = 2	x_2 = 5	x_3 = 8
y_1 = 0,4	0,15	0,12	0,03
y_2 = 0,8	0,05	0,30	0,35

Табл. 1 называют **таблицей распределения двумерной величины (X, Y)**;её следует понимать так. Случайная величина X может принять одно из следующих значений: 2, 5 и 8. Случайная величина Y – значения 0,4 и 0,8. Число 0,15 – это вероятность того, что X = 2 и одновременно Y = 0,4, или, иначе говоря, вероятность произведения двух событий; события, состоящего в том, что X = 2, и события, состоящего в том, что Y = 0,4, т.е. P((X=2)(Y=0,4)) = 0,15. Аналогично, вероятность P((X=2)(Y=0,8)) = 0,05 и т.д. Обратим внимание на следующее: поскольку в табл. 9 указаны все возможные значения величин X и Y, сумма вероятностей, стоящих в таблице, должна быть равна единице: 0,15 + 0,05 + 0,12 + 0,30 + 0,03 + 0,35 = 1.

Прежде чем выяснить тип зависимости величины Y от X, найдём:

а) Закон распределения величины X. Он представлен табл. 2.

Таблица 2.

x	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$	
P (X = x)	0,15 + 0,05 = 0,2	0,12 + 0,30 = 0,42	0,35 + 0,03 = 0,38	= 1

M(X) = 5,54, D(X) = 4,9284

Действительно, например, величина X примет значение, равное 2, только в том случае, когда одновременно с этим величина Y примет значение 0,4 или 0,8, т.е.

P(X = 2) = P((X = 2)(Y = 0,4)) + P((X = 2)(Y = 0,08)) = 0,15 + 0,05 = 0,2.

Справа от ряда распределения величины X находятся её математическое ожидание и дисперсия.

б) Закон распределения величины Y. Он имеет вид табл. 3.

Таблица 3.

y	$y_1 = 0,4$	$y_2 = 0,8$	
P (Y = y)	0,15 + 0,12 + 0,03 = 0,30	0,05 + 0,30 + 0,35 = 0,7	= 1

M(Y) = 0,68, D(Y) = 0,0336

в) Условные законы распределения величины Y, а именно закон распределения величины Y сначала при условии, что X = 2, затем при условии, что X=5, и наконец, при условии, что X = 8.

Итак, пусть X = 2. Тогда условная вероятность P(Y = 0,4/X = 2) = 0,75,

а условная вероятность P(Y = 0,8/X = 2) = 0,25.

Таким образом, закон распределения величины Y при условии, что X = 2, задан табл. 4.

Таблица 4

y	$y_1 = 0,4$	$y_2 = 0,8$	

P(Y = y/X = 2)	(,75	(,25	= 1

M(Y/X = 2) = 0,4*0,75 + 0,8*0,25 = 0,5, D(Y/X = 2) = 0,03

Справа помещено условное математическое ожидание и значение условной дисперсии. Покажем, как вычисляется условная дисперсия. Общая формула условной дисперсии имеет вид $D(Y/X = x) = M[(Y/X = x) – M(Y/X = x)]^2$. (23)

Для табл. 4 получаем

$D(Y/X = 2) = M[(Y/X = 2) – M(Y/X = 2)]^2 = M[(Y/X = 2) – 0,5]^2 = P(Y = y_i/X = 2) = (0,4 – 0,5)^2 \cdot 0,75 + (0,8 – 0,5)^2 \cdot 0,25 = 0,03$.

3. Корреляционное отношение

Задача 1. По данным о месячной заработной плате 10 рабочих трех разных профессий (токарь, слесарь и кузнец) вычислены: общая дисперсия заработной платы σ_0^2 =1636 и средняя из внутригрупповых дисперсий $\overline{\sigma^2}$ =1140. Вычислить корреляционное отношение.

Решение. Корреляционное отношение вычисляется по формуле $\eta^2 = \frac{\delta^2}{\sigma_0^2}$. Следовательно, сначала необходимо найти межгрупповую дисперсию $\delta^2 = \sigma_0^2 - \overline{\sigma^2} = 1636 - 1140 = 496$. Подставляя это значение в вышеприведенную формулу, получим: $\eta^2 = \frac{496}{1636} = 0{,}303$.

4. Линейная парная регрессия

Пример 1. Результаты измерения диаметров *(x)* и высот *(y)* 250 сосен записаны в таблицу.

x \	8	9	0	1	2	3	4	5	6	7
5										
0			5	9						
5				8	9	0				
0						2				
5										
0										
5										

Составить уравнения регрессии и найти коэффициент корреляции.

Уравнения регрессии имеют вид:

$$\overline{y_x} = \bar{y} + r_в \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x}), \quad \overline{x_x} = \bar{x} + r_в \cdot \frac{\sigma_y}{\sigma_x}(y - \bar{y}),$$

где $\bar{x}$ и $\bar{y}$ - средние значения величин *x* и *y;*

σ_x и σ_y - средние квадратические отклонения величин *x* и *y;*

$r_в$ - выборочный коэффициент корреляции, вычисляемой по формуле:

$$r_в = \frac{\sum n_{xy} \cdot x \cdot y - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y}.$$

Для определения всех этих величин пользуемся методом произведений. Дополним данную таблицу несколькими строками и столбцами.

На основании метода произведений запишем, что $\bar{x} = \bar{u}h_x + C_x$,

$\bar{y} = \bar{v}h_y + C_y$, где C_x и C_y - значения величин *x* и *y,* имеющих большую частоту;

u, *v* - средние значения условных вариант, вычисление по формулам:

$$\bar{u} = \frac{\sum h_{xi}u_i}{n}; \ \bar{v} = \frac{\sum h_{yj}u_j}{n}; \ u_i = \frac{x_i - C_x}{h_x}; \ v_j = \frac{y_j - C_y}{h_y},$$

где h_x и h_y -разность между соседними значениями *X* и *Y*.

Вычислим $\bar{x}$ и $\bar{y}$: $\bar{x} = \frac{-14}{250} \cdot 5 + 25 = 24{,}7$, $\bar{y} = \frac{11}{250} \cdot 1 + 22 = 22{,}04$.

Пользуясь методом произведений, находим средние квадратические отклонения величин *x* и *y:*

$\sigma_x = h_x\sigma_u$, $\sigma_y = h_y\sigma_v$.

Вычислим σ_u и σ_v по формулам средних квадратических отклонений:

$$\sigma_u = \sqrt{\frac{\sum n_{xi} \cdot u_i^2}{n} - \left(\frac{\sum n_{xi} \cdot u_i}{n}\right)^2}, \quad \sigma_u = \sqrt{\frac{300}{250} - \left(\frac{-14}{250}\right)^2} \approx 1{,}1$$

$$\sigma_v = \sqrt{\frac{\sum n_{yj} \cdot v_j^2}{n} - \left(\frac{\sum n_{yj} \cdot v_j}{n}\right)^2}, \quad \sigma_v = \sqrt{\frac{569}{250} - \left(\frac{11}{250}\right)^2} \approx 1{,}5.$$

Находим: $\sigma_x = 5 \cdot 11 = 5{,}5$, $\sigma_y = 1 \cdot 1{,}5 = 1{,}5$.

При переходе к условным вариантам и коэффициент корреляции имеет вид:

$$r_в = \frac{\sum n_{uv} \cdot u \cdot v - n \cdot \bar{u} \cdot \bar{v}}{n \cdot \sigma_u \cdot \sigma_v}.$$

Для вычисления величины $\sum n_{uv} \cdot u \cdot v$ применяется метод «четырех полей». Строка и столбец, на пересечении которых находится наибольшая частота , делят таблицу на четыре поля. В верхнем углу каждой заполненной клетки, расположенной в одном из четырех полей, записываем соответствующие произведения *uv*, а затем подсчитываем сумму произведений $\sum n_{uv} \cdot u \cdot v$. Определяем, сколько раз n_{uv} встречаются произведения *uv.*

Получаем:

$$\sum n_{uv} \cdot u \cdot v = 1 \cdot 6 + 6 \cdot 4 + 4 \cdot 2 + 3 \cdot 3 + 15 \cdot 2 + 29 \cdot 1 + 12 \cdot 1 + 8 \cdot 2 + 5 \cdot 3 + 3 \cdot 2 + 6 \cdot 4 + 1 \cdot 8 + 3 \cdot 9 + 3 \cdot 12$$
$$+ 1 \cdot 20 + 8 \cdot (-1) + 4 \cdot (-1) + 1 \cdot (-2) = 290$$

Определим коэффициент корреляции $r_в = \frac{290 - 250 \cdot (-0{,}06) \cdot 0{,}04}{250 \cdot 1{,}1 \cdot 1{,}5} = 0{,}7$.

Напишем уравнения регрессий

$$\overline{y_x} = 32{,}04 + 0{,}7 \cdot \frac{1{,}5}{5{,}5}(x - 2 \cdot 24{,}7) \text{ или } \overline{y_x} = 0{,}191x + 17{,}32,$$

$$\overline{x_y} = 24{,}7 + 0{,}5 \cdot \frac{5{,}5}{1{,}5}(y - 2 \cdot 22{,}04) \text{ или } \overline{x_y} = 2{,}57y + 32{,}1.$$

Имея уравнение регрессии, можно вычислить средние значения одной величины при любом значении другой.

5. Коэффициент корреляции, его свойства, значимость

Пример. Найти коэффициент корреляции между производительностью труда Y (тыс. руб.) и энерговооруженностью труда X (кВт) (в расчете на одного работающего) для 14 предприятий региона по следующим данным:

Таблица 1

i	,8	,2	,0	,5	,2	,7	,0	,8	,0	,4	,2	,4	,0	,0
i	,7	,9	,2	,3	,4	,8	,1	,8	0,6	0,7	1,1	1,8	2,1	2,4

Решение. Вычислим необходимые суммы:

$$\sum_{i=1}^{14} x_i = 2{,}8 + 2{,}2 + \ldots + 6{,}0 + 9{,}0 = 64{,}2;$$

$$\sum_{i=1}^{14} x_i^2 = 2{,}8^2 + 2{,}2^2 + \ldots + 6{,}0^2 + 9{,}0^2 = 335{,}26;$$

$$\sum_{i=1}^{14} y_i = 6{,}7 + 6{,}9 + \ldots + 12{,}1 + 12{,}4 = 132{,}9;$$

$$\sum_{i=1}^{14} y_i^2 = 6{,}7^2 + 6{,}9^2 + \ldots + 12{,}1^2 + 12{,}4^2 = 1313{,}95;$$

$$\sum_{i=1}^{14} x_i y_i = 2{,}8 \cdot 6{,}7 + 2{,}2 \cdot 6{,}9 + \ldots + 6{,}0 \cdot 12{,}1 + 9{,}0 \cdot 12{,}4 = 650{,}99.$$

Полагая $n_{ij} = n_i = n_j = 1$, $j = i$ и заменяя $\sum_{i=1}^{l}\sum_{j=1}^{m}$ на $\sum_{i=1}^{n}$, получим

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} =$$

$$= \frac{14 \cdot 650{,}99 - 64{,}2 \cdot 132{,}9}{\sqrt{14 \cdot 335{,}26 - 64{,}2^2}\sqrt{14 \cdot 1313{,}95 - 132{,}4^2}} = 0{,}898,$$

что говорит о тесной связи между переменными.

В примере вычислен коэффициент корреляции r = 0,740. Статистика критерия по

$$t = \frac{0{,}740\sqrt{50-2}}{\sqrt{1-0{,}740^2}} = 7{,}62$$

Для уровня значимости α = 0,05 и числа степеней свободы к = 50 - 2 = 48 находим критическое значение статистики $t_{0,95;48} = 2{,}01$. Поскольку $t > t_{0,95;48}$ коэффициент корреляции между суточной выработкой продукции Y и величиной основных производственных фондов X значимо отличается от нуля.

2.7.3 Результаты и выводы:

В результате проведенного занятия студенты:

- освоили основные виды зависимостей между величинами;
- усвоили основные методы нахождения регрессии;
- выработали навыки по вычислению коэффициентов регрессии и корреляционного отношения.