

**МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО  
СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ УЗБЕКИСТАН**

**ТАШКЕНСТКИЙ АВТОМОБИЛЬНО-ДОРОЖНЫЙ  
ИНСТИТУТ**

**КАФЕДРА «ВЫСШАЯ МАТЕМАТИКА»**

**ТЕКСТЫ ЛЕКЦИЙ**  
*по курсу*

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ И  
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Ташкент 2007

**М.У.Гафуров, Р.Х.Кенджаев, Ф.М.Закиров**

Теория вероятностей и математическая статистика. 2007 г.  
— 124 стр.

В основу текстов лекций положен семестровый курс теории вероятностей и математической статистики, читаемый авторами в течение ряда лет в Ташкенском автомобильно-дорожном институте. В сборнике основные понятия и факты теории вероятностей и математической статистики первоначально вводятся для конечных схем. Далее утверждения и формулы легко переводятся на общий случай. Приведенные теоретические материалы проиллюстрированы большим числом примеров прикладного содержания.

В сборнике лекций содержатся следующие разделы: пространство случайных событий, случайные величины и их числовые характеристики, предельные теоремы, элементы выборочного пространства, точечное и интервальное статистическое оценивание, элементы корреляционного и регрессионного анализа, а также проверка статистических гипотез.

Настоящий сборник рассчитан для студентов экономических специальностей высших экономических и технических учебных заведений, а также для всех заинтересованных в освоении вероятностных методов решения практических задач.

**Рецензенты:** **А.А.Абдушукуров** — заведующий кафедрой «Теория вероятностей и математическая статистика» Национального Университета Узбекистана, доктор физико-математических наук,

**Х.Валиджанов** — доцент кафедры «Высшая математика» Ташкенского автомобильно-дорожного института, кандидат физико-математических наук

Утверждено на заседании Научно-методического совета  
естественных и инженерных наук ТАДИ (протокол № 7 от 21  
февраля 2007 года)

## О Г Л А В Л Е Н И Е

|     |   |     |
|-----|---|-----|
| 1.  | Предмет теории вероятностей и ее значение для решения экономических, технических задач.<br>Вероятность и ее определение.....      | 4   |
| 2.  | Операции над событиями. Условная вероятность.....   | 12  |
| 3.  | Теоремы сложения и умножения вероятностей.<br>Формулы полной вероятности и Байеса.....  | 16  |
| 4.  | Последовательность независимых испытаний.<br>Локальная и интегральная теоремы Лапласа.....  | 24  |
| 5.  | Дискретные случайные величины. Закон распределения. Виды дискретных распределений.....  | 31  |
| 6.  | Числовые характеристики дискретных случайных величин и их свойства.....   | 38  |
| 7.  | Функции распределения и плотности непрерывных случайных величин, их свойства.....   | 47  |
| 8.  | Числовые характеристики непрерывных случайных величин. Виды непрерывных распределений.....  | 55  |
| 9.  | Закон больших чисел и его практическое значение.<br>Понятие о центральной предельной теореме.....                                 | 64  |
| 10. | Предмет и основные задачи математической статистики. Выборка.....   | 68  |
| 11. | Статистическое распределение выборки. Эмпирическая функция распределения. Полигон и гистограмма.....                              | 73  |
| 12. | Статистическая оценка. Требования, предъявляемые к статистической оценке. Выборочное среднее и выборочная дисперсия.....          | 80  |
| 13. | Интервальные оценки. Доверительный интервал.<br>Доверительные интервалы для неизвестных параметров нормального распределения..... | 88  |
| 14. | Элементы корреляционного и регрессионного анализа...  | 96  |
| 15. | Выборочный коэффициент корреляции и его свойства...   | 104 |
| 16. | Статистические гипотезы и их классификация.<br>Статистический критерий.....   | 109 |
| 17. | Критерии согласия.....  | 121 |
|     | Список литературы.....  | 125 |

## Тема № 1

### **Предмет теории вероятностей и ее значение для решения экономических, технических задач. Вероятность и ее определение**

#### **План:**

1. Предмет теории вероятностей.
2. Краткая история развития теории вероятностей.
3. Значение теории вероятностей для экономических, технических задач.
4. Элементарные события и события.
5. Вероятность и ее определение.
6. Относительная частота.

На протяжении длительного времени человечество изучало и использовало для своей деятельности лишь так называемые детерминистические закономерности. Однако, поскольку случайные события врываются в нашу жизнь помимо нашего желания и постоянно окружают нас, и более того, поскольку почти все явления природы имеют случайный характер, необходимо научиться их изучать и разработать для этой цели методы изучения.

По форме проявления причинных связей законы природы и общества делятся на два класса: детерминированные (предопределенные) и статистические.

Например, на основании законов небесной механики по известному в настоящем положению планет Солнечной системы может быть практически однозначно предсказано их положение в любой наперед заданный момент времени, в том числе очень точно могут быть предсказаны солнечные и лунные затмения. Это пример детерминированных законов.

Вместе с тем не все явления поддаются точному предсказанию. Так, долговременные изменения климата, кратковременные изменения погоды не являются объектами для успешного прогнозирования, т.е. многие законы и закономерности гораздо менее вписываются в детерминированные рамки. Такого рода законы называются

статистическими. Согласно этим законам, будущее состояние системы определяется не однозначно, а лишь с некоторой вероятностью.

Теория вероятностей, как и другие математические науки, возродилась и развилась из потребностей практики. Она занимается изучением закономерностей, присущих массовым случайным событиям.

Теория вероятностей изучает свойства массовых случайных событий, способных многократно повторяться при воспроизведении определенного комплекса условий. Основное свойство любого случайного события, независимо от его природы, — мера, или вероятность его осуществления.

Наблюдаемые нами события (явления) можно подразделить на три вида: достоверные, невозможные и случайные.

*Достоверным* называют событие, которое обязательно произойдет. *Невозможным* называют событие, которое заведомо не произойдет. *Случайным* называют событие, которое может либо произойти, либо не произойти.

Теория вероятностей не ставит перед собой задачу предсказать, произойдет единичное событие или нет, так как невозможно учесть влияние на случайное событие всех причин. С другой стороны, оказывается, что достаточно большое число однородных случайных событий, независимо от их конкретной природы, подчиняется определенным закономерностям, а именно — вероятностным закономерностям.

Итак, *предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.*

Некоторые задачи, относящиеся к массовым случайным явлениям, пытались решать, используя соответствующий математический аппарат, еще в начале XVII в. Изучая ход и результаты различных азартных игр, Б. Паскаль, П. Ферма и Х. Гюйгенс в середине XVII века заложили основы классической теории вероятностей. В своих работах они неявно использовали понятия вероятности и математического ожидания случайной величины. Только в начале XVIII в. Я.Бернулли формулирует понятие вероятности.

Дальнейшими успехами теория вероятностей обязана Муавру, Лапласу, Гауссу, Пуассону и др.

В развитие теории вероятностей огромный вклад внесли русские и советские математики, такие как П.Л. Чебышев, А.А. Марков, А.М. Ляпунов, С.Н. Бернштейн, А.Н. Колмогоров, А.Я. Хинчин, А. Прохоров и др.

Особое место в развитии теории вероятностей принадлежит и узбекистанской школе, яркими представителями которой являются академики В.И. Романовский, С.Х. Сираждинов, Т.А. Сарымсаков, Т.А. Азларов, Ш.К. Фарманов, профессора И.С. Бадалбаев, М.У. Гафуров, Ш.А. Хашимов и др.

Как уже было отмечено, потребности практики, способствовав зарождению теории вероятностей, питали ее развитие как науки, приводя к появлению все новых ее ветвей и разделов. На теорию вероятностей опирается математическая статистика, задача которой состоит в том, чтобы по выборке восстановить с определенной степенью достоверности характеристики, присущие генеральной совокупности. От теории вероятностей отделились такие отрасли науки, как теория случайных процессов, теория массового обслуживания, теория информации, теория надежности, эконометрическое моделирование и др.

В качестве важнейших сфер приложения теории вероятностей можно указать экономические, технические науки. В настоящее время трудно себе представить исследование экономико-технических явлений без моделирований, опирающихся на теорию вероятностей, без моделей корреляционного и регрессионного анализа, адекватности и «чувствительных» адаптивных моделей.

События, происходящие в автомобильных потоках, степень надежности составных частей машин, автокатастрофы на дорогах, различные ситуации в процессе проектирования дорог ввиду их недетерминированности входят в круг проблем, исследуемых посредством методов теории вероятностей.

Основные понятия теории вероятностей — это опыт или эксперимент и события. Действия, которые осуществляются при определенных условиях и обстоятельствах, мы назовем *экспериментом*. Каждое конкретное осуществление эксперимента называется *испытанием*.

Всякий мыслимый результат эксперимента называется *элементарным событием* и обозначается через  $\omega$ . Случайные

события состоят из некоторого числа элементарных событий и обозначаются через  $A, B, C, D, \dots$

Множество элементарных событий таких, что

- 1) в результате проведения эксперимента всегда происходит одно из элементарных событий  $\omega$ ;
- 2) при одном испытании произойдет только одно элементарное событие  $\omega$

называется *пространством элементарных событий* и обозначается через  $\Omega$ .

Таким образом, любое случайное событие является подмножеством пространства элементарных событий. По определению пространства элементарных событий достоверное событие можно обозначить через  $\Omega$ . Невозможное событие обозначается через  $\emptyset$ .

**Пример 1.** Бросается игральная кость. Пространство элементарных событий, отвечающее данному эксперименту, имеет следующий вид:  $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ .

**Пример 2.** Пусть в урне содержатся 2 красных, 3 синих и 1 белый, всего 6 шаров. Эксперимент состоит в том, что из урны вынимаются наудачу шары. Пространство элементарных событий, отвечающее данному эксперименту, имеет следующий вид:  $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ , где элементарные события имеют следующие значения:  $\omega_1$  — появился белый шар;  $\omega_2, \omega_3$  — появился красный шар;  $\omega_4, \omega_5, \omega_6$  — появился синий шар. Рассмотрим следующие события:

$A$  — появление белого шара;

$B$  — появление красного шара;

$C$  — появление синего шара;

$D$  — появление цветного (небелого) шара.

Здесь мы видим, что каждое из этих событий обладает той или иной степенью возможности: одни — большей, другие — меньшей. Очевидно, что степень возможности события  $B$  больше, чем события  $A$ ; события  $C$  — чем события  $B$ ; события  $D$  — чем события  $C$ . Чтобы количественно сравнивать между собой события по степени их возможности, очевидно, нужно с каждым событием связать определенное число, которое тем больше, чем более возможно событие.

Это число обозначим через  $P(A)$  и назовем вероятностью

события  $A$ . Дадим теперь определение вероятности.

Пусть пространство элементарных событий  $\Omega$  является конечным множеством и элементы его суть  $\omega_1, \omega_2, \dots, \omega_n$ . Будем считать, что они являются равновероятными элементарными событиями, т.е. каждое элементарное событие не имеет больше шансов появления, чем другие. Как известно, каждое случайное событие  $A$  состоит из элементарных событий как подмножество  $\Omega$ . Эти элементарные события называются благоприятствующими для  $A$ .

*Вероятность события  $A$  определяется формулой*

$$P(A) = \frac{m}{n}, \quad (1.1)$$

где  $m$  — число благоприятствующих элементарных событий для  $A$ ,  $n$  — число всех элементарных событий, входящих в  $\Omega$ .

Если в примере 1 через  $A$  обозначить событие, состоящее в том, что выпадет четное число очков, то  $P(A) = \frac{3}{6} = \frac{1}{2}$ .

В примере 2 вероятности событий имеют следующие значения:  $P(A) = \frac{1}{6}$ ;  $P(B) = \frac{2}{6} = \frac{1}{3}$ ;  $P(C) = \frac{3}{6} = \frac{1}{2}$ ;  $P(D) = \frac{5}{6}$ .

Из определения вероятности вытекают следующие ее свойства:

1. *Вероятность достоверного события равна единице.*

Действительно, если событие достоверно, то все элементарные события благоприятствуют ему. В этом случае  $m=n$  и, следовательно,

$$P(\Omega) = \frac{m}{n} = \frac{n}{n} = 1.$$

2. *Вероятность невозможного события равна нулю.*

Действительно, если событие невозможно, то ни одно элементарное событие не благоприятствует ему. В этом случае  $m=0$  и, следовательно,

$$P(\emptyset) = \frac{m}{n} = \frac{0}{n} = 0.$$

3. *Вероятность случайного события есть положительное число, заключенное между нулем и единицей.*

Действительно, случайному событию благоприятствует

лишь часть из общего числа элементарных событий. В этом случае  $0 < m < n$ , а значит,  $0 < \frac{m}{n} < 1$  и, следовательно,

$$0 < P(A) < 1.$$

Итак, вероятность любого события удовлетворяет неравенствам

$$0 \leq P(A) \leq 1. \quad (1.2)$$

*Относительной частотой события* называют отношение числа испытаний, в которых событие появилось, к общему числу фактически произведенных испытаний.

Таким образом, относительная частота события  $A$  определяется формулой

$$W(A) = \frac{m}{n}, \quad (1.3)$$

где  $m$  — число появлений события,  $n$  — общее число испытаний.

Сопоставляя определения вероятности и относительной частоты, заключаем: определение вероятности не требует, чтобы испытания производились в действительности; определение же относительной частоты предполагает, что испытания были произведены фактически.

**Пример 3.** Из 80 случайно отобранных одинаковых деталей выявлено 3 бракованных. Относительная частота бракованных деталей равна

$$W(A) = \frac{3}{80}.$$

**Пример 4.** В течение года на одном из объектов было проведено 24 проверки, причем было зарегистрировано 19 нарушений законодательства. Относительная частота нарушений законодательства равна

$$W(A) = \frac{19}{24}.$$

Длительные наблюдения показали, что если в одинаковых условиях производятся опыты, в каждом из которых число испытаний достаточно велико, то относительная частота изменяется мало (тем меньше, чем больше произведено испытаний), колеблясь около некоторого постоянного числа. Оказалось, что это постоянное число есть вероятность появления

события.

Таким образом, если опытным путем установлена относительная частота, то полученное число можно принять за приближенное значение вероятности. Это есть статистическое определение вероятности.

В заключении рассмотрим геометрическое определение вероятности.

Если пространство элементарных событий  $\Omega$  рассматривать как некоторую область на плоскости или в пространстве, а  $A$  как ее подмножество, то *вероятность события  $A$*  будет рассматриваться как отношение площадей или объемов  $A$  и  $\Omega$ , и находится по следующим формулам:

$$P(A) = \frac{S(A)}{S(\Omega)}, \quad (1.4)$$

$$P(A) = \frac{V(A)}{V(\Omega)}. \quad (1.5)$$

### **Вопросы для повторения и контроля:**

1. На какие классы делятся законы природы и общества по форме проявления причинных связей?
2. На какие виды можно подразделить события?
3. Что является предметом теории вероятностей?
4. Что вы знаете об истории развития теории вероятностей?
5. Каково значение теории вероятностей для экономических, технических задач?
6. Что такое эксперимент, испытание, элементарное событие и событие, как они обозначаются?
7. Что называется пространством элементарных событий?
8. Как определяется вероятность события?
9. Какие свойства вероятности вы знаете?
10. Что вы знаете об относительной частоте события?
11. В чем сущность статистического определения вероятности?
12. Каково геометрическое определение вероятности?

## Тема № 2

### Операции над событиями. Условная вероятность

#### План:

1. Операции над событиями.
2. Условная вероятность.

Часто возникает вопрос: насколько связаны два случайных события  $A$  и  $B$  друг с другом, в какой мере наступление одного из них влияет на возможность наступления другого?

В качестве примера связи между двумя событиями можно привести случаи, когда наступление одного из событий ведет к обязательному осуществлению другого или же, наоборот, когда наступление одного события исключает шансы другого.

Если в результате эксперимента события  $A$  и  $B$  не могут наступить одновременно, то они называются *несовместными* событиями, в противном случае *совместными*.

**Пример 1.** Из ящика с деталями наудачу извлечена деталь. То, что она стандартна, исключает ее нестандартность. События «Наудачу извлеченная деталь стандартна» и «Наудачу извлеченная деталь нестандартна» — несовместные.

Если события рассматривать как подмножества пространства элементарных событий, то отношения между событиями можно интерпретировать как соотношения между множествами. Несовместные события — это такие события, которые не содержат общих элементарных событий.

Говорят, что *событие  $A$  влечет за собой событие  $B$* , если в результате эксперимента из наступления события  $A$  обязательно следует наступление события  $B$ , и обозначают это через  $A \subset B$ . Если  $A \subset B$  и  $B \subset A$ , то  $A = B$ .

**Пример 2.** Бросается игральная кость. Событие «выпало 4» влечет за собой событие «выпало четное число очков».

*Суммой двух событий  $A$  и  $B$*  называют событие, состоящее в наступлении события  $A$  или события  $B$ , или обоих этих событий. Оно обозначается через  $A+B$  или  $A \cup B$ . *Суммой нескольких событий* называют событие, которое состоит в наступлении хотя бы одного из этих событий.

**Пример 3.** Из орудия производится два выстрела. Если  $A$  – попадание при первом выстреле, а  $B$  – попадание при втором выстреле, то  $A+B$  – попадание при первом выстреле, или при втором, или в обоих выстрелах.

*Произведением двух событий  $A$  и  $B$  называют событие, состоящее в совместном наступлении событий  $A$  и  $B$ . Оно обозначается через  $AB$  или  $A \cap B$ . Произведением нескольких событий называют событие, состоящее в совместном наступлении всех этих событий.*

**Пример 4.** В ящике содержатся детали, изготовленные заводами №1 и №2. Если  $A$  – появление стандартной детали, а  $B$  – деталь изготовлена заводом №1, то  $AB$  – появление стандартной детали завода №1.

*Противоположное событие для события  $A$  обозначается через  $\bar{A}$ . Оно считается наступившим тогда и только тогда, когда  $A$  не наступает. Иными словами,  $A$  и  $\bar{A}$  – это такие несовместные события, которые вместе образуют достоверное событие, т.е.  $A \cup \bar{A} = \Omega$ .*

**Пример 5.** Попадание и промах при выстреле по цели — противоположные события. Если  $A$  — попадание, то  $\bar{A}$  — промах.

Событие, которое представляет собой наступление события  $A$  и не наступление события  $B$ , называется *разностью событий  $A$  и  $B$* , и обозначается через  $A \setminus B$ .

Два события называют *независимыми*, если вероятность одного из них не зависит от наступления или ненаступления другого. В противном случае эти события называются *зависимыми*.

**Пример 6.** Монета брошена 2 раза. Вероятность появления герба при первом бросании (событие  $A$ ) не зависит от появления герба при втором бросании (событие  $B$ ). В свою очередь, вероятность выпадения герба при втором бросании не зависит от результата первого бросания. Таким образом, события  $A$  и  $B$  — независимы.

Несколько событий называют *парно независимыми*, если любые два из них взаимно независимы.

Пусть  $A$  и  $B$  — два случайных события, причем  $P(B) \neq 0$ . Из определения зависимых событий следует, что вероятность

одного из событий зависит от наступления или ненаступления другого. Поэтому, если нас интересует вероятность события  $A$ , то важно знать, наступило ли событие  $B$ .

Вероятность события  $A$  при условии, что произошло событие  $B$ , называется *условной вероятностью* и обозначается через  $P(A/B)$ .

**Пример 7.** В урне содержится 3 белых и 3 черных шара. Из урны дважды вынимают наудачу по одному шару, не возвращая их в урну. Найти вероятность появления белого шара при втором испытании (событие  $A$ ), если при первом испытании был извлечен черный шар (событие  $B$ ).

Решение. После первого испытания в урне осталось всего 5 шаров, из них 3 белых. Искомая условная вероятность равна

$$P(A/B) = \frac{3}{5}.$$

Выведем теперь формулу условной вероятности. Пусть событиям  $A$  и  $B$  благоприятствуют соответственно  $m$  и  $k$  элементарных событий из  $n$ ; тогда, согласно (1.1), их безусловные вероятности равны  $\frac{m}{n}$  и  $\frac{k}{n}$  соответственно. Пусть событию  $A$  при условии, что событие  $B$  произошло, благоприятствуют  $r$  элементарных событий, тогда, согласно (1.1), условная вероятность события  $A$  равна

$$P(A/B) = \frac{r}{k}.$$

Разделив и числитель, и знаменатель на  $n$ , получим формулу условной вероятности

$$P(A/B) = \frac{r/n}{k/n} = \frac{P(AB)}{P(B)}$$

или

$$P(A/B) = \frac{P(AB)}{P(B)}. \quad (2.1)$$

поскольку событию  $AB$  соответствуют  $r$  элементарных событий и, следовательно,  $\frac{r}{n}$  — его безусловная вероятность.

### Вопросы для повторения и контроля:

1. Какие события называются несовместными, а какие совместными?
2. Что означает выражение «событие  $A$  влечет за собой событие  $B$ » и как оно обозначается?
3. Что называется суммой событий и как оно обозначается?
4. Что называется произведением событий и как оно обозначается?
5. Что такое противоположное событие и как оно обозначается?
6. Что называется разностью событий и как оно обозначается?
7. Какие события называются независимыми, а какие зависимыми?
8. Что такое условная вероятность и какова ее формула?

## Тема № 3

### Теоремы сложения и умножения вероятностей. Формулы полной вероятности и Байеса

#### План:

1. Теоремы сложения вероятностей.
2. Теоремы умножения вероятностей.
3. Формула полной вероятности.
4. Формула Байеса.

Пусть события  $A$  и  $B$  — несовместные, причем вероятности этих событий даны. Как найти вероятность того, что наступит либо событие  $A$ , либо событие  $B$ , т.е. вероятность суммы этих событий  $A+B$ ? Ответ на этот вопрос дает следующая теорема.

**Теорема 3.1 (сложения вероятностей несовместных событий).** *Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий:*

$$P(A + B) = P(A) + P(B). \quad (3.1)$$

Доказательство. Введем обозначения:

$n$  — общее число элементарных событий;

$m_1$  — число элементарных событий, благоприятствующих событию  $A$ ;

$m_2$  — число элементарных событий, благоприятствующих событию  $B$ .

Число элементарных событий, благоприятствующих наступлению либо события  $A$ , либо события  $B$ , равно  $m_1 + m_2$ . Следовательно,

$$P(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

Приняв во внимание, что  $\frac{m_1}{n} = P(A)$  и  $\frac{m_2}{n} = P(B)$ , окончательно получим

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

**Следствие 3.1.** *Вероятность суммы нескольких несовместных событий равна сумме вероятностей этих событий:*

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n). \quad (3.2)$$

**Пример 1.** В урне 30 шаров: 10 красных, 5 синих и 15 белых. Найти вероятность появления цветного шара.

Решение. Появление цветного шара означает появление либо красного, либо синего шара.

Вероятность появления красного шара (событие  $A$ )

$$P(A) = \frac{10}{30} = \frac{1}{3}.$$

Вероятность появления синего шара (событие  $B$ )

$$P(B) = \frac{5}{30} = \frac{1}{6}.$$

События  $A$  и  $B$  несовместны (появление шара одного цвета исключает появление шара другого цвета), поэтому искомая вероятность равна

$$P(A + B) = P(A) + P(B) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

Так как противоположные события вместе образуют достоверное событие, то из теоремы 3.1 вытекает, что

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

поэтому

$$P(\bar{A}) = 1 - P(A). \quad (3.3)$$

**Пример 2.** Вероятность того, что день будет дождливым, равна  $p = 0,3$ . Найти вероятность того, что день будет ясным.

Решение. События «день дождливый» и «день ясный» — противоположные, поэтому искомая вероятность равна

$$q = 1 - p = 1 - 0,3 = 0,7.$$

Из формулы (2.1) получаем следующую теорему.

**Теорема 3.2 (умножения вероятностей зависимых событий).** Вероятность произведения двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило:

$$P(AB) = P(A/B) \cdot P(B). \quad (3.4)$$

**Пример 3.** У сборщика имеется 3 конусных и 7 эллиптических валиков. Сборщик наудачу взял один валик, а

затем второй. Найти вероятность того, что первый из взятых валиков — конусный, а второй — эллиптический.

Решение. Вероятность того, что первый из взятых валиков окажется конусным (событие  $B$ ), равна

$$P(B) = \frac{3}{10}.$$

Условная вероятность того, что второй из валиков окажется эллиптическим (событие  $A$ ), вычисленная в предположении, что первый валик — конусный, равна

$$P(A/B) = \frac{7}{9}.$$

Тогда по формуле (3.4) искомая вероятность равна

$$P(AB) = P(A/B) \cdot P(B) = \frac{7}{9} \cdot \frac{3}{10} = \frac{7}{30}.$$

Теперь перейдем к случаю, когда события  $A$  и  $B$  — независимые, и найдем вероятность произведения этих событий.

Так как событие  $A$  не зависит от события  $B$ , то его условная вероятность  $P(A/B)$  равна его безусловной вероятности  $P(A)$ , т.е.

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

Отсюда вытекает следующая теорема.

**Теорема 3.3 (умножения вероятностей независимых событий).** Вероятность произведения двух независимых событий равна произведению вероятностей этих событий:

$$P(AB) = P(A) \cdot P(B). \quad (3.5)$$

**Следствие 3.2.** Вероятность произведения нескольких независимых событий равна произведению вероятностей этих событий:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n).$$

**Пример 4.** Имеется 3 ящика, содержащих по 10 деталей. В первом ящике 8, во втором 7 и в третьем 9 стандартных деталей. Из каждого ящика наудачу вынимают по одной детали. Найти вероятность того, что все три вынутые детали окажутся стандартными.

Решение. Вероятность того, что из первого ящика вынута стандартная деталь (событие  $A$ ), равна

$$P(A) = \frac{8}{10} = 0,8.$$

Вероятность того, что из второго ящика вынута стандартная деталь (событие  $B$ ), равна

$$P(B) = \frac{7}{10} = 0,7.$$

Вероятность того, что из третьего ящика вынута стандартная деталь (событие  $C$ ), равна

$$P(C) = \frac{9}{10} = 0,9.$$

Так как события  $A$ ,  $B$  и  $C$  — независимые, то искомая вероятность по теореме умножения вероятностей независимых событий равна

$$P(ABC) = P(A) \cdot P(B) \cdot P(C) = 0,8 \cdot 0,7 \cdot 0,9 = 0,504.$$

Теперь перейдем к случаю, когда события  $A$  и  $B$  — совместные, и найдем вероятность суммы этих событий.

**Теорема 3.4 (сложения вероятностей совместных событий).** Вероятность суммы двух совместных событий равна сумме вероятностей этих событий с вычетом вероятности их произведения:

$$P(A + B) = P(A) + P(B) - P(AB). \quad (3.6)$$

**Пример 5.** Вероятности попадания в цель при стрельбе первого и второго орудий соответственно равны:  $p_1 = 0,7$ ;  $p_2 = 0,8$ . Найти вероятность попадания при одном залпе (из обоих орудий) хотя бы одним из орудий.

Решение. Вероятность попадания в цель каждым из орудий не зависит от результата стрельбы из другого орудия, поэтому события  $A$  (попадание первого орудия) и  $B$  (попадание второго орудия) независимы.

Вероятность события  $AB$  (оба орудия дали попадание), равна

$$P(AB) = P(A) \cdot P(B) = 0,7 \cdot 0,8 = 0,56.$$

Искомая вероятность равна

$$P(A + B) = P(A) + P(B) - P(AB) = 0,7 + 0,8 - 0,56 = 0,94.$$

Если независимые события  $A_1, A_2, \dots, A_n$  вместе образуют достоверное событие, то вероятность появления хотя

бы одного из этих событий можно найти по формуле

$$P(A_1 + A_2 + \dots + A_n) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \dots \cdot P(\overline{A_n}) \quad (3.7)$$

**Пример 6.** В типографии имеются 4 машины. Для каждой машины вероятность того, что она работает в данный момент, равна 0,9. Найти вероятность того, что в данный момент работает хотя бы одна машина (событие  $A$ ).

Решение. Вероятность того, что машина в данный момент не работает, равна

$$q = 1 - p = 1 - 0,9 = 0,1.$$

Тогда искомая вероятность равна

$$P(A) = 1 - q^4 = 1 - (0,1)^4 = 0,9999.$$

Говорят, что события  $A_1, A_2, \dots, A_n$  образуют *полную группу событий*, если они несовместны и вместе образуют достоверное событие, т.е.  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ ;  $A_1 + A_2 + \dots + A_n = \Omega$ .

Предположим, что событие  $A$  может наступить только при условии появления одного из событий  $H_1, H_2, \dots, H_n$ , образующих полную группу, которые назовем *гипотезами*. Пусть известны вероятности этих событий и условные вероятности  $P(A/H_i)$ ,  $i = \overline{1, n}$ .

Так как  $A\Omega = A$ , то

$$A = AH_1 + AH_2 + \dots + AH_n.$$

Из несовместности  $H_1, H_2, \dots, H_n$  вытекает несовместность событий  $AH_1, AH_2, \dots, AH_n$ .

Применяя формулу (3.1), имеем

$$P(A) = P(AH_1) + P(AH_2) + \dots + P(AH_n).$$

Согласно формуле (3.4) (так как события  $H_1, H_2, \dots, H_n$  могут быть и зависимыми), заменив каждое слагаемое  $P(AH_i)$  в правой части последнего выражения произведением  $P(A/H_i)P(H_i)$ , получим *формулу полной вероятности*

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i). \quad (3.8)$$

**Пример 7.** Имеется два набора деталей. Вероятность того, что деталь первого набора стандартна, равна 0,8, а второго — 0,9. Найти вероятность того, что взятая наудачу деталь из наудачу

взятого набора — стандартная.

Решение. Обозначим через  $A$  событие «извлеченная деталь стандартна». Деталь может быть извлечена либо из первого набора (событие  $H_1$ ), либо из второго набора (событие  $H_2$ ).

Вероятность того, что деталь будет вынута из первого набора, равна

$$P(H_1) = \frac{1}{2}.$$

Вероятность того, что деталь будет вынута из второго набора, равна

$$P(H_2) = \frac{1}{2}.$$

По условиям задачи  $P(A/H_1) = 0,8$  и  $P(A/H_2) = 0,9$ .

Тогда искомая вероятность находится по формуле полной вероятности и равна

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = 0,5 \cdot 0,8 + 0,5 \cdot 0,9 = 0,85.$$

Пусть теперь для тех же событий, что и при выводе формулы полной вероятности, появилось событие  $A$ , и ставится задача отыскать условные вероятности гипотез  $P(H_k/A)$ ,  $k = \overline{1, n}$ .

Из формулы (2.1) имеем

$$P(H_k/A) = \frac{P(AH_k)}{P(A)}.$$

Далее, из формулы (3.4) получаем

$$P(AH_k) = P(H_k)P(A/H_k).$$

Отсюда и из предыдущего соотношения, применяя формулу полной вероятности, выводим *формулу Байеса*:

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^n P(H_i)P(A/H_i)} \quad (3.9)$$

**Пример 8.** Детали, изготавливаемые цехом завода, попадают для проверки их на стандартность к одному из двух контролеров. Вероятность того, что деталь попадет к первому контролеру, равна 0,6, а ко второму — 0,4. Вероятность того, что годная деталь будет признана стандартной первым контролером, равна

0,94, а вторым — 0,98. Годная деталь при проверке была признана стандартной. Найти вероятность того, что эту деталь проверил первый контролер.

Решение. Обозначим через  $A$  событие, состоящее в том, что годная деталь признана стандартной. Можно сделать два предположения:

- 1) деталь проверил первый контролер (гипотеза  $H_1$ );
- 2) деталь проверил второй контролер (гипотеза  $H_2$ ).

По условиям задачи имеем:

$P(H_1) = 0,6$  (вероятность того, что деталь попадет к первому контролеру);

$P(H_2) = 0,4$  (вероятность того, что деталь попадет ко второму контролеру);

$P(A/H_1) = 0,94$  (вероятность того, что годная деталь будет признана стандартной первым контролером);

$P(A/H_2) = 0,98$  (вероятность того, что годная деталь будет признана стандартной вторым контролером).

Искомую вероятность найдем по формуле Байеса

$$\begin{aligned} P(H_1/A) &= \frac{P(H_1)P(A/H_1)}{P(H_1)P(A/H_1) + P(H_2)P(A/H_2)} = \\ &= \frac{0,6 \cdot 0,94}{0,6 \cdot 0,94 + 0,4 \cdot 0,98} \approx 0,58996. \end{aligned}$$

### **Вопросы для повторения и контроля:**

1. О чем теорема сложения вероятностей несовместных событий и каково ее доказательство?
2. Чему равна вероятность противоположного события?
3. О чем идет речь в теоремах умножения вероятностей зависимых и независимых событий?
4. О чем теорема сложения вероятностей совместных событий?
5. Как можно найти вероятность появления хотя бы одного события?
6. Какие события образуют полную группу событий?
7. Что такое формула полной вероятности и как она выводится?
8. Что такое формула Байеса и как она выводится?

### **Опорные слова:**

Вероятность суммы несовместных событий, вероятность противоположного события, вероятность произведения зависимых событий, вероятность произведения независимых событий, вероятность суммы несовместных событий, вероятность появления хотя бы одного события, полная группа событий, гипотезы, формула полной вероятности, формула Байеса.

## Тема № 4

### Последовательность независимых испытаний. Локальная и интегральная теоремы Лапласа

#### План:

1. Последовательность независимых испытаний.
2. Формула Бернулли.
3. Наивероятнейшее число успехов.
4. Локальная теорема Лапласа.
5. Интегральная теорема Лапласа.
6. Вероятность отклонения относительной частоты от постоянной вероятности.

Пусть производится  $n$  независимых испытаний, в каждом из которых событие  $A$  может либо произойти (успех), либо не произойти (неудача). Будем считать, что вероятность события  $A$  в каждом испытании одна и та же, а именно равна  $p$ . Следовательно, вероятность ненаступления события  $A$  в каждом испытании также постоянна и равна  $q=1-p$ . Такая последовательность испытаний называется *схемой Бернулли*.

В качестве таких испытаний можно рассматривать, например, производство изделий на определенном оборудовании при постоянстве технологических и организационных условий, в этом случае изготовление годного изделия — успех, бракованного — неудача. Эта ситуация соответствует схеме Бернулли, если считать, что процесс изготовления одного изделия не зависит от того, были годными или бракованными предыдущие изделия.

Другим примером является стрельба по мишени. Здесь попадание — успех, промах — неудача.

Поставим своей задачей вычислить вероятность того, что при  $n$  испытаниях событие  $A$  осуществится ровно  $k$  раз и, следовательно, не осуществится  $n-k$  раз, т.е. будет  $k$  успехов и  $n-k$  неудач.

Искомую вероятность обозначим  $P_n(k)$ . Например, символ  $P_5(3)$  означает вероятность того, что в пяти испытаниях событие появится ровно 3 раза и, следовательно, не наступит 2 раза.

Последовательность  $n$  независимых испытаний можно рассматривать как сложное событие, являющееся произведением  $n$  независимых событий. Следовательно, вероятность того, что в  $n$  испытаниях событие  $A$  наступит  $k$  раз и не наступит  $n-k$  раз, по теореме 3.3 умножения вероятностей независимых событий, равна

$$p^k q^{n-k}.$$

Таких сложных событий может быть столько, сколько можно составить сочетаний из  $n$  элементов по  $k$  элементов, т.е.  $C_n^k$ .

Так как эти сложные события несовместны, то по теореме 3.1 сложения вероятностей несовместных событий, искомая вероятность равна сумме вероятностей всех возможных сложных событий. Поскольку же вероятности всех этих сложных событий одинаковы, то искомая вероятность (появление  $k$  раз события  $A$  в  $n$  испытаниях) равна вероятности одного сложного события, умноженной на их число

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

или

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \quad (4.1)$$

Полученную формулу называют *формулой Бернулли*.

**Пример 1.** Вероятность того, что расход электроэнергии на продолжении одних суток не превысит установленной нормы, равна  $p = 0,75$ . Найти вероятность того, что в течение 4 суток из ближайших 6 суток расход электроэнергии не превысит нормы.

Решение. Вероятность нормального расхода электроэнергии на продолжении каждых из 6 суток постоянна и равна  $p = 0,75$ . Следовательно, вероятность перерасхода электроэнергии в каждые сутки также постоянна и равна  $q = 1 - p = 1 - 0,75 = 0,25$ .

Искомая вероятность по формуле Бернулли равна

$$P_6(4) = C_6^4 p^4 q^2 = C_6^2 p^4 q^2 = \frac{6 \cdot 5}{1 \cdot 2} (0,75)^4 \cdot (0,25)^2 = \frac{1215}{4096} \approx 0,297.$$

В ряде задач представляет интерес наивероятнейшее число успехов, т.е. такое число  $\hat{m}$  успехов, вероятность которого самая большая среди вероятностей (4.1). Так как при увеличении  $k$  вероятности (4.1) сначала возрастают, а затем, с определенного

момента, начинают убывать, то для  $\hat{m}$  должны иметь место соотношения

$$P_n(\hat{m}) \geq P_n(\hat{m} - 1) \quad (4.2)$$

и

$$P_n(\hat{m}) \geq P_n(\hat{m} + 1). \quad (4.3)$$

Используя формулу (4.1) и соотношение  $p + q = 1$ , из (4.2) и (4.3) получаем соответственно неравенства

$$(n - \hat{m} + 1)p \geq \hat{m}q \quad (4.4)$$

и

$$(\hat{m} + 1)q \geq (n - \hat{m})p. \quad (4.5)$$

Окончательно получаем, что  $\hat{m}$  лежит в интервале единичной длины:

$$np - q \leq \hat{m} \leq np + p. \quad (4.6)$$

Однако, стоит заметить, что использование формулы Бернулли при больших значениях  $n$  достаточно трудно, так как формула требует выполнения действий над громадными числами.

Например, если  $n = 50$ ,  $k = 30$ ,  $p = 0,1$ , то для отыскания вероятности  $P_{50}(30)$  надо вычислить выражение

$$P_{50}(30) = \frac{50!}{30! \cdot 20!} \cdot (0,1)^{30} \cdot (0,9)^{20}, \quad \text{где} \quad 50! = 30414093 \cdot 10^{57},$$

$$30! = 26525286 \cdot 10^{25}, \quad 20! = 24329020 \cdot 10^{11}.$$

Естественно возникает вопрос: нельзя ли вычислить интересующую нас вероятность, не прибегая к формуле Бернулли? Оказывается можно. Локальная теорема Лапласа и дает асимптотическую формулу, которая позволяет приближенно найти вероятность появления события ровно  $k$  раз в  $n$  испытаниях, если число испытаний достаточно велико.

**Локальная теорема Лапласа.** Если вероятность  $p$  появления события  $A$  в каждом испытании постоянна и отлична от нуля и единицы, то вероятность  $P_n(k)$  того, что событие  $A$  появится в  $n$  испытаниях ровно  $k$  раз, приближенно равна (тем точнее, чем больше  $n$ ) значению функции

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

$$\text{при } x = \frac{k - np}{\sqrt{npq}}.$$

Имеются таблицы, в которых помещены значения функции

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \text{ При этом следует учитывать, что } \phi(-x) = \phi(x),$$

так как функция  $\phi(x)$  четна.

Итак, вероятность того, что событие  $A$  появится в  $n$  независимых испытаниях ровно  $k$  раз, приближенно равна

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \phi(x), \quad (4.7)$$

$$\text{где } x = \frac{k - np}{\sqrt{npq}}.$$

**Пример 2.** Найти вероятность того, что событие  $A$  наступит ровно 80 раз в 400 испытаниях, если вероятность появления этого события в каждом испытании равна 0,2.

Решение. По условию  $n = 400$ ;  $k = 80$ ;  $p = 0,2$ ;  $q = 0,8$ . Воспользуемся формулой (4.7):

$$P_{400}(80) \approx \frac{1}{\sqrt{400 \cdot 0,2 \cdot 0,8}} \cdot \phi(x) = \frac{1}{8} \cdot \phi(x).$$

Вычислим определяемое данными задачи значение  $x$ :

$$x = \frac{k - np}{\sqrt{npq}} = \frac{80 - 400 \cdot 0,2}{8} = 0.$$

По таблице находим  $\phi(0) = 0,3989$ .

Искомая вероятность равна

$$P_{400}(80) = \frac{1}{8} \cdot 0,3989 = 0,04986.$$

Формула Бернулли приводит примерно к такому же результату (выкладки, ввиду их громоздкости, опущены):

$$P_{400}(80) = 0,0498.$$

Пусть теперь требуется вычислить вероятность  $P_n(k_1, k_2)$  того, что событие  $A$  появится в  $n$  испытаниях не менее  $k_1$  и не более  $k_2$  раз (для краткости будем говорить «от  $k_1$  до  $k_2$  раз»). Эта задача решается с помощью следующей теоремы.

**Интегральная теорема Лапласа.** Если вероятность  $p$  наступления события  $A$  в каждом испытании постоянна и отлична от нуля и единицы, то вероятность  $P_n(k_1, k_2)$  того, что событие  $A$  появится в  $n$  испытаниях от  $k_1$  до  $k_2$  раз, приближенно равна определенному интегралу

$$P_n(k_1, k_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{z^2}{2}} dz, \quad (4.8)$$

где  $x_1 = \frac{k_1 - np}{\sqrt{npq}}$  и  $x_2 = \frac{k_2 - np}{\sqrt{npq}}$ .

При решении задач, требующих применения интегральной теоремы Лапласа, пользуются специальной таблицей для интеграла  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ . В таблице даны значения функции  $\Phi(x)$  для  $x \geq 0$ , а для  $x < 0$  воспользуемся нечетностью функции  $\Phi(x)$ , т.е.  $\Phi(-x) = -\Phi(x)$ . Функцию  $\Phi(x)$  часто называют *функцией Лапласа*.

Итак, вероятность того, что событие  $A$  появится в  $n$  независимых испытаниях от  $k_1$  до  $k_2$  раз, равна

$$P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1), \quad (4.9)$$

где  $x_1 = \frac{k_1 - np}{\sqrt{npq}}$  и  $x_2 = \frac{k_2 - np}{\sqrt{npq}}$ .

**Пример 3.** Вероятность того, что организация не прошла проверку налоговой инспекции, равна  $p = 0,2$ . Найти вероятность того, что среди 400 случайно отобранных организаций не прошедших проверку окажется от 70 до 100 организаций.

Решение. По условию  $n = 400$ ;  $k_1 = 70$ ;  $k_2 = 100$ ;  $p = 0,2$ ;  $q = 0,8$ . Воспользуемся формулой (4.9):

$$P_{400}(70, 100) \approx \Phi(x_2) - \Phi(x_1).$$

Вычислим нижний и верхний пределы интегрирования:

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{70 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = -1,25;$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{100 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = 2,5.$$

Таким образом, имеем

$$P_{400}(70, 100) \approx \Phi(2,5) - \Phi(-1,25) = \Phi(2,5) + \Phi(1,25).$$

По таблице значений функции  $\Phi(x)$  находим

$$\Phi(2,5) = 0,4938; \quad \Phi(1,25) = 0,3944.$$

Искомая вероятность равна

$$P_{400}(70, 100) \approx 0,4938 + 0,3944 = 0,8882.$$

В теме № 1 было отмечено, что по статистическому определению вероятности в качестве вероятности можно взять относительную частоту, поэтому представляет интерес оценка разности между ними. Вероятность того, что отклонение относительной частоты  $\frac{m}{n}$  от постоянной вероятности  $p$  по

абсолютной величине не превышает заданного числа  $\varepsilon > 0$ , равна

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right). \quad (4.10)$$

**Пример 4.** Вероятность того, что деталь не стандартна, равна  $p = 0,1$ . Найти вероятность того, что среди случайно отобранных 400 деталей относительная частота появления нестандартных деталей отклонится от вероятности  $p = 0,1$  по абсолютной величине не более, чем на 0,03.

Решение. По условию  $n = 400$ ;  $p = 0,1$ ;  $q = 0,9$ ;  $\varepsilon = 0,03$ .

Требуется найти вероятность  $P\left(\left|\frac{m}{400} - 0,1\right| \leq 0,03\right)$ .

Пользуясь формулой (4.10), имеем

$$P\left(\left|\frac{m}{400} - 0,1\right| \leq 0,03\right) \approx 2\Phi\left(0,03 \cdot \sqrt{\frac{400}{0,1 \cdot 0,9}}\right) = 2\Phi(2).$$

По таблице находим  $\Phi(2) = 0,4772$ . Следовательно,  $2\Phi(2) = 0,9544$ .

Итак, искомая вероятность приближенно равна 0,9544.

Смысл полученного результата таков: если взять достаточно большое число проб по 400 деталей в каждой, то примерно в 95,44 % этих проб отклонение относительной частоты от

постоянной вероятности  $p = 0,1$  по абсолютной величине не превысит  $0,03$ .

### **Вопросы для повторения и контроля:**

1. Что называется схемой Бернулли?
2. Как выводится формула Бернулли?
3. Как находится наиболее вероятное число успехов?
4. О чем идет речь в локальной теореме Лапласа?
5. О чем идет речь в интегральной теореме Лапласа?
6. Как находится вероятность отклонения относительной частоты от постоянной вероятности?

### **Опорные слова:**

Последовательность независимых испытаний, схема Бернулли, формула Бернулли, наиболее вероятное число успехов, локальная теорема Лапласа, вероятность того, что событие  $A$  появится в  $n$  независимых испытаниях ровно  $k$  раз, интегральная теорема Лапласа, вероятность того, что событие  $A$  появится в  $n$  независимых испытаниях от  $k_1$  до  $k_2$  раз, функция Лапласа, вероятность отклонения относительной частоты от постоянной вероятности.

## Тема № 5

### Дискретные случайные величины. Закон распределения. Виды дискретных распределений

#### План:

1. Понятие случайной величины и ее виды.
2. Закон распределения дискретной случайной величины.
3. Некоторые дискретные распределения.

В предыдущих темах неоднократно приводились события, состоящие в появлении того или иного числа. Например, при бросании игральной кости могли появиться числа 1, 2, 3, 4, 5 и 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 величина случайная; числа 1, 2, 3, 4, 5 и 6 есть возможные значения этой величины.

*Случайной величиной* называют величину, которая в результате испытания примет одно и только одно возможное значение, наперед не известное и зависящее от случайных причин, которые заранее не могут быть учтены.

**Пример 1.** Число родившихся мальчиков среди ста новорожденных есть случайная величина, которая имеет следующие возможные значения: 0, 1, 2, ... , 100.

**Пример 2.** Расстояние, которое пролетит снаряд при выстреле из орудия, есть случайная величина. Возможные значения этой величины принадлежат некоторому промежутку  $(a, b)$ .

Так как в результате испытаний происходят элементарные события, то можно связать понятия случайной величины и элементарных событий и дать другое определение случайной величины.

*Случайной величиной* называется функция  $X = X(\omega)$ , определенная на пространстве элементарных событий  $\Omega$ ,  $\omega \in \Omega$ .

**Пример 3.** При подбрасывании двух монет число выпавших гербов  $X$  есть случайная величина, которая может принимать значения 0, 1 и 2. Пространство элементарных событий состоит

из следующих элементарных событий:

$$\omega_1 = \{ГГ\}, \omega_2 = \{РГ\}, \omega_3 = \{ГР\}, \omega_4 = \{РР\}.$$

Тогда  $X$  принимает следующие значения:

$$\begin{aligned} X(\omega_1) = X(ГГ) = 2, & \quad X(\omega_2) = X(РГ) = 1, \\ X(\omega_3) = X(ГР) = 1, & \quad X(\omega_4) = X(РР) = 0. \end{aligned}$$

Случайные величины обозначаются прописными латинскими буквами  $X, Y, Z, \dots$ , а их возможные значения — соответствующими строчными буквами  $x, y, z, \dots$ . Например, если случайная величина  $X$  имеет три возможных значения, то они обозначаются через  $x_1, x_2, x_3$ .

*Дискретной (прерывной)* называют случайную величину, которая принимает отдельные, изолированные возможные значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или бесконечным. В качестве примера таковой можно привести случайную величину из примера 1.

*Непрерывной* называют случайную величину, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Число возможных значений непрерывной случайной величины бесконечно. В качестве примера такой величины можно привести случайную величину из примера 2.

Для задания дискретной случайной величины недостаточно перечислить все возможные ее значения, нужно еще указать их вероятности. С другой стороны, во многих задачах нет необходимости рассматривать случайные величины как функции от элементарного события, а достаточно знать лишь вероятности возможных значений случайной величины, т.е. закон распределения случайной величины.

*Законом распределения вероятностей* или просто *законом распределения дискретной случайной величины* называют соответствие между возможными значениями и их вероятностями; его можно задать в виде таблицы, графика и формулы.

Рассмотрим различные способы задания закона распределения вероятностей на примерах.

При табличном задании закона распределения дискретной случайной величины первая строка таблицы содержит возможные значения, а вторая — их вероятности. Сумма вероятностей во второй строке таблицы должна быть равна 1. В таблице 5.1 задан закон распределения дискретной случайной величины из примера 3.

Т а б л и ц а 5.1

|       |       |       |       |
|-------|-------|-------|-------|
| $x_i$ | 0     | 1     | 2     |
| $p_i$ | 1 / 4 | 1 / 2 | 1 / 4 |

**Пример 4.** В денежной лотерее выпущено 100 билетов. Разыгрывается один выигрыш в 5000 сум, пять выигрышей по 1000 сум и десять выигрышей по 500 сум. Найти закон распределения случайной величины  $X$  — стоимости возможного выигрыша для владельца одного лотерейного билета.

Решение. Напишем возможные значения  $X$ :  $x_1 = 5000$ ,  $x_2 = 1000$ ,  $x_3 = 500$ ,  $x_4 = 0$ . Вероятности этих возможных значений таковы:  $p_1 = 0,01$ ,  $p_2 = 0,05$ ,  $p_3 = 0,1$ ,  $p_4 = 1 - (p_1 + p_2 + p_3) = 0,84$ .

Тогда искомый закон распределения имеет вид

Т а б л и ц а 5.2

|       |      |     |      |      |
|-------|------|-----|------|------|
| $x_i$ | 0    | 500 | 1000 | 5000 |
| $p_i$ | 0,84 | 0,1 | 0,05 | 0,01 |

Для наглядности закон распределения дискретной случайной величины можно изобразить и графически, для чего в прямоугольной системе координат строят точки  $(x_i, p_i)$ , а затем соединяют их отрезками прямых. Полученную фигуру называют *многоугольником распределения*. На рисунке 5.1 приведен многоугольник распределения случайной величины  $X$  из примера 3.

Теперь рассмотрим некоторые дискретные распределения, заданные посредством формул: биномиальное, геометрическое и Пуассона.

Пусть производится  $n$  независимых испытаний, в каждом из которых вероятность наступления события  $A$  (успеха) постоянна

и равна  $p$  (следовательно, вероятность неоявления (неудачи) равна  $q=1-p$ ). Рассмотрим в качестве дискретной случайной величины  $X$  число появлений события  $A$  в этих испытаниях. Возможные значения  $X$  таковы:  $0, 1, 2, \dots, n$ . Вероятности этих возможных значений находятся по формуле Бернулли (4.1):

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

где  $k= 0, 1, 2, \dots, n$ .

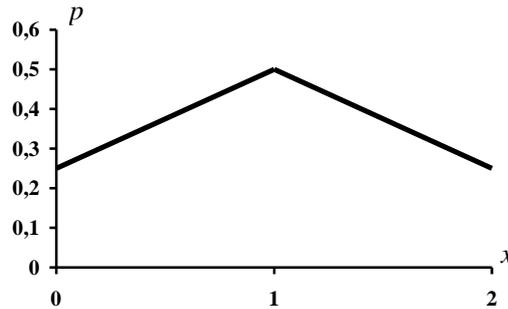


Рис. 5.1.

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

$$(p + q)^n = C_n^n p^n + C_n^{n-1} p^{n-1} q + \dots + C_n^k p^k q^{n-k} + \dots + C_n^0 q^n.$$

Так как  $p + q = 1$ , то сумма вероятностей возможных значений случайной величины равна 1.

Таким образом, биномиальный закон распределения имеет вид

Т а б л и ц а 5.3

|       |       |               |         |                     |         |       |
|-------|-------|---------------|---------|---------------------|---------|-------|
| $x_i$ | $n$   | $n-1$         | $\dots$ | $k$                 | $\dots$ | $0$   |
| $P_i$ | $p^n$ | $n p^{n-1} q$ | $\dots$ | $C_n^k p^k q^{n-k}$ | $\dots$ | $q^n$ |

В качестве примера биномиального распределения можно привести распределение случайной величины из примера 3.

Пусть производятся независимые испытания, в каждом из которых вероятность появления события  $A$  (успеха) равна  $p$  ( $0 < p < 1$ ) и, следовательно, вероятность его неоявления

(неудачи) равна  $q=1-p$ . Испытания продолжаются до первого успеха. Таким образом, если событие  $A$  появилось в  $k$ -м испытании, то в предшествующих  $k - 1$  испытаниях оно не появлялось.

Если через  $X$  обозначить дискретную случайную величину, равную числу испытаний до первого успеха, то ее возможными значениями будут натуральные числа  $1, 2, 3, \dots$

Пусть в первых  $k - 1$  испытаниях событие  $A$  не наступило, а в  $k$ -м испытании появилось. Вероятность этого «сложного события», по теореме 3.3 умножения вероятностей независимых событий, равна

$$P(X = k) = q^{k-1} p. \quad (5.1)$$

*Геометрическим* называют распределение вероятностей, определяемое формулой (5.1), так как полагая в этой формуле  $k = 1, 2, \dots$ , получим геометрическую прогрессию с первым членом  $p$  и знаменателем  $q$  ( $0 < q < 1$ ):

$$p, qp, q^2 p, \dots, q^{k-1} p, \dots$$

Просуммировав бесконечно убывающую геометрическую прогрессию, легко убедиться, что сумма вероятностей возможных значений случайной величины равна 1:

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} pq^{k-1} = p \sum_{k=1}^{\infty} q^{k-1} = p \cdot \frac{1}{1-q} = \frac{p}{p} = 1.$$

Таким образом, геометрический закон распределения имеет вид

Т а б л и ц а 5.4

|       |     |      |         |     |             |     |
|-------|-----|------|---------|-----|-------------|-----|
| $x_i$ | 1   | 2    | 3       | ... | $k$         | ... |
| $p_i$ | $p$ | $qp$ | $q^2 p$ | ... | $q^{k-1} p$ | ... |

**Пример 5.** Из орудия производится стрельба по цели до первого попадания. Вероятность попадания в цель  $p = 0,6$ . Найти вероятность того, что попадание произойдет при третьем выстреле.

Решение. По условию  $p = 0,6$ ,  $q = 0,4$ ,  $k = 3$ . Искомая вероятность по формуле (5.1) равна:

$$P(X = 3) = 0,4^2 \cdot 0,6 = 0,096.$$

Пусть производится  $n$  независимых испытаний, в каждом из которых вероятность появления события  $A$  равна  $p$ . Для определения вероятности  $k$  появлений события в этих испытаниях используют формулу Бернулли. Если же  $n$  велико, то пользуются локальной теоремой Лапласа. Однако она дает большую погрешность, если вероятность события мала ( $p \leq 0,1$ ).

Если сделать допущение, что произведение  $np$  при  $n \rightarrow \infty$  сохраняет постоянное значение, а именно  $np = \lambda$ , то вероятность того, что при очень большом числе испытаний, в каждом из которых вероятность события очень мала, событие наступит ровно  $k$  раз, находится по следующей формуле

$$P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (5.2)$$

Эта формула выражает закон распределения *Пуассона* вероятностей массовых ( $n$  велико) и маловероятных ( $p$  мало) событий. Имеются специальные таблицы для распределения Пуассона.

**Пример 6.** Завод отправил на базу 5000 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равно 0,0002. Найти вероятность того, что на базу придут 3 негодных изделия.

Решение. По условию  $n = 5000$ ,  $p = 0,0002$ ,  $k = 3$ . Найдем  $\lambda$ :

$$\lambda = np = 5000 \cdot 0,0002 = 1.$$

Искомая вероятность по формуле (5.2) равна:

$$P_{5000}(3) = \frac{1^3}{3!} \cdot e^{-1} = \frac{1}{6e} \approx 0,06.$$

### Вопросы для повторения и контроля:

1. Как определяется случайная величина в общем случае и на языке функций?
2. Что такое дискретная случайная величина?
3. Что такое непрерывная случайная величина?
4. Что вы знаете о законе распределения дискретной случайной величины?
5. Что вы знаете о биномиальном законе распределения?
6. Каковы особенности геометрического закона распределения?

7. В каких случаях используют распределение Пуассона?

**Опорные слова:**

Случайная величина, дискретная случайная величина, непрерывная случайная величина, закон распределения дискретной случайной величины, многоугольник распределения, биномиальное распределение, геометрическое распределение, распределение Пуассона.

## Тема № 6

### Числовые характеристики дискретных случайных величин и их свойства

#### План:

1. Математическое ожидание дискретной случайной величины.
2. Свойства математического ожидания.
3. Числовые характеристики рассеяния дискретной случайной величины.
4. Свойства дисперсии.
5. Другие числовые характеристики дискретных случайных величин.

Как мы видели выше, закон распределения полностью характеризует дискретную случайную величину. Однако часто закон распределения неизвестен и приходится ограничиваться числами, которые описывают случайную величину суммарно; такие числа называют *числовыми характеристиками случайной величины*.

К числу важных числовых характеристик относится математическое ожидание. Математическое ожидание приближенно равно среднему значению случайной величины. Для решения многих задач достаточно знать математическое ожидание. Например, если известно, что математическое ожидание числа выбиваемых очков у первого стрелка больше, чем у второго, то первый стрелок в среднем выбивает больше очков, чем второй, и, следовательно, стреляет лучше второго.

*Математическим ожиданием дискретной случайной величины  $X$*  называется сумма произведений всех ее возможных значений на их вероятности и обозначается через  $M(X)$ .

Пусть случайная величина  $X$  принимает значения  $x_1, x_2, \dots, x_n$  с соответствующими вероятностями  $p_1, p_2, \dots, p_n$ . Тогда математическое ожидание  $M(X)$  случайной величины  $X$  определяется равенством

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n. \quad (6.1)$$

Если дискретная случайная величина  $X$  принимает бесконечное множество возможных значений, то

$$M(X) = \sum_{i=1}^{\infty} x_i p_i. \quad (6.2)$$

**Пример 1.** Найти математическое ожидание случайной величины  $X$ , зная закон ее распределения

Т а б л и ц а 6.1

|       |     |     |     |
|-------|-----|-----|-----|
| $x_i$ | 3   | 5   | 2   |
| $p_i$ | 0,1 | 0,6 | 0,3 |

Решение. Искомое математическое ожидание по формуле (6.1) равно

$$M(X) = 3 \cdot 0,1 + 5 \cdot 0,6 + 2 \cdot 0,3 = 3,9.$$

**Пример 2.** Найти математическое ожидание числа появлений события  $A$  в одном испытании, если вероятность события  $A$  равна  $p$ .

Решение. Случайная величина  $X$  — число появлений события  $A$  в одном испытании — может принимать только два значения:  $x_1 = 1$  (событие  $A$  наступило) с вероятностью  $p$  и  $x_2 = 0$  (событие  $A$  не наступило) с вероятностью  $q = 1 - p$ . Искомое математическое ожидание по формуле (6.1) равно

$$M(X) = 1 \cdot p + 0 \cdot q = p.$$

Итак, *математическое ожидание числа появлений события в одном испытании равно вероятности этого события.*

Теперь приведем свойства математического ожидания.

**Свойство 6.1.** *Математическое ожидание постоянной величины равно самой постоянной:*

$$M(C) = C.$$

Доказательство. Будем рассматривать постоянную  $C$  как дискретную случайную величину, которая имеет одно возможное значение  $C$  и принимает его с вероятностью  $p = 1$ . Следовательно,

$$M(C) = C \cdot 1 = C.$$

**Свойство 6.2.** *Постоянный множитель можно выносить за знак математического ожидания:*

$$M(CX) = CM(X).$$

Две случайные величины называются *независимыми*, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величина.

*Произведением независимых случайных величин*  $X$  и  $Y$  называется случайная величина  $XY$ , возможные значения которой равны произведениям каждого возможного значения  $X$  на каждое возможное значение  $Y$ ; вероятности возможных значений произведения  $XY$  равны произведениям вероятностей возможных значений сомножителей.

**Свойство 6.3.** *Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:*

$$M(XY) = M(X) \cdot M(Y).$$

**Следствие 6.1.** *Математическое ожидание произведения нескольких независимых случайных величин равно произведению их математических ожиданий.*

**Пример 3.** Независимые случайные величины  $X$  и  $Y$  заданы следующими законами распределения:

Т а б л и ц а 6.2

|       |     |     |     |
|-------|-----|-----|-----|
| $x_i$ | 5   | 2   | 4   |
| $p_i$ | 0,6 | 0,1 | 0,3 |

Т а б л и ц а 6.3

|       |     |     |
|-------|-----|-----|
| $y_i$ | 7   | 9   |
| $p_i$ | 0,8 | 0,2 |

и

Найти математическое ожидание случайной величины  $XY$ .

**Решение.** Найдем математические ожидания каждой из данных величин:

$$M(X) = 5 \cdot 0,6 + 2 \cdot 0,1 + 4 \cdot 0,3 = 4,4;$$

$$M(Y) = 7 \cdot 0,8 + 9 \cdot 0,2 = 7,4.$$

Случайные величины  $X$  и  $Y$  независимые, поэтому искомое математическое ожидание равно

$$M(XY) = M(X) \cdot M(Y) = 4,4 \cdot 7,4 = 32,56.$$

*Суммой случайных величин*  $X$  и  $Y$  называется случайная величина  $X+Y$ , возможные значения которой равны суммам каждого возможного значения  $X$  с каждым возможным значением  $Y$ ; вероятности возможных значений  $X+Y$  для независимых величин  $X$  и  $Y$  равны произведениям вероятностей слагаемых; для зависимых величин — произведениям вероятности одного

слагаемого на условную вероятность второго.

**Свойство 6.4.** *Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых:*

$$M(X + Y) = M(X) + M(Y).$$

**Следствие 6.2.** *Математическое ожидание суммы нескольких случайных величин равно сумме математических ожиданий слагаемых.*

**Пример 4.** Найти математическое ожидание суммы числа очков, которые могут выпасть при бросании двух игральных костей.

Решение. Обозначим число очков, которое может выпасть на первой кости, через  $X$  и на второй — через  $Y$ . Возможные значения этих величин одинаковы и равны 1, 2, 3, 4, 5 и 6, причем вероятность каждого из этих значений равна  $1/6$ .

Найдем математическое ожидание числа очков, которые могут выпасть на первой кости:

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Очевидно, что и  $M(Y) = \frac{7}{2}$ .

Искомое математическое ожидание равно

$$M(X + Y) = M(X) + M(Y) = \frac{7}{2} + \frac{7}{2} = 7.$$

**Свойство 6.5.** *Математическое ожидание числа появлений события  $A$  в  $n$  независимых испытаниях, в каждом из которых вероятность  $p$  появления события постоянна, равно произведению числа испытаний на вероятность появления события в одном испытании:*

$$M(X) = np.$$

**Пример 5.** Вероятность выявления ошибок в документации при проверке предприятия равна  $p = 0,6$ . Найти математическое ожидание общего числа выявлений ошибок, если будет проведено 10 проверок предприятий.

Решение. Выявление ошибок при каждой проверке не зависит от исходов других проверок, поэтому рассматриваемые события независимы и, следовательно, искомое математическое ожидание равно

$$M(X) = np = 10 \cdot 0,6 = 6 \text{ (выявлений ошибок).}$$

Некоторые случайные величины имеют одинаковые математические ожидания, но различные возможные значения. Рассмотрим, например, дискретные случайные величины  $X$  и  $Y$ , заданные следующими законами распределения:

Т а б л и ц а 6.4

|       |       |      |
|-------|-------|------|
| $x_i$ | -0,01 | 0,01 |
| $p_i$ | 0,5   | 0,5  |

и

Т а б л и ц а 6.5

|       |      |     |
|-------|------|-----|
| $y_i$ | -100 | 100 |
| $p_i$ | 0,5  | 0,5 |

Найдем математические ожидания этих величин:

$$M(X) = -0,01 \cdot 0,5 + 0,01 \cdot 0,5 = 0;$$

$$M(Y) = -100 \cdot 0,5 + 100 \cdot 0,5 = 0.$$

Здесь математические ожидания обеих величин одинаковы, а возможные значения различны, причем  $X$  имеет возможные значения, близкие к математическому ожиданию, а  $Y$  — далекие от своего математического ожидания. Таким образом, зная лишь математическое ожидание случайной величины, еще нельзя судить ни о том, какие возможные значения она может принимать, ни о том, как они рассеяны вокруг математического ожидания.

Другими словами, математическое ожидание полностью случайную величину не характеризует. По этой причине наряду с математическим ожиданием рассматриваются и другие числовые характеристики.

Пусть  $X$  — случайная величина и  $M(X)$  — ее математическое ожидание. *Отклонением случайной величины* называется разность  $X - M(X)$ .

На практике часто требуется оценить рассеяние возможных значений случайной величины вокруг ее среднего значения. Например, в артиллерии важно знать, насколько кучно лягут снаряды вблизи цели, которая должна быть поражена.

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

$$D(X) = M[X - M(X)]^2. \quad (6.3)$$

Для вычисления дисперсии часто бывает удобно воспользоваться следующей формулой:

$$D(X) = M(X^2) - [M(X)]^2. \quad (6.4)$$

**Пример 6.** Найти дисперсию случайной величины  $X$ , которая задана следующим законом распределения:

Т а б л и ц а 6.6

|       |     |     |     |
|-------|-----|-----|-----|
| $x_i$ | 2   | 3   | 5   |
| $p_i$ | 0,1 | 0,6 | 0,3 |

Решение. Математическое ожидание  $M(X)$  равно:

$$M(X) = 2 \cdot 0,1 + 3 \cdot 0,6 + 5 \cdot 0,3 = 3,5.$$

Закон распределения случайной величины  $X^2$  имеет вид:

Т а б л и ц а 6.7

|         |     |     |     |
|---------|-----|-----|-----|
| $x_i^2$ | 4   | 9   | 25  |
| $p_i$   | 0,1 | 0,6 | 0,3 |

Математическое ожидание  $M(X^2)$  равно:

$$M(X^2) = 4 \cdot 0,1 + 9 \cdot 0,6 + 25 \cdot 0,3 = 13,3.$$

Искомая дисперсия равна

$$D(X) = M(X^2) - [M(X)]^2 = 13,3 - (3,5)^2 = 1,05.$$

Дисперсия, как и математическое ожидание, имеет несколько свойств.

**Свойство 6.6.** Дисперсия постоянной величины равна нулю:

$$D(C) = 0.$$

Доказательство. По определению дисперсии,

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

Пользуясь свойством 6.1, получим

$$D(C) = M[C - C]^2 = M(0) = 0.$$

Итак,

$$D(C) = 0.$$

Свойство становится ясным, если учесть, что постоянная величина сохраняет одно и то же значение и рассеяния не имеет.

**Свойство 6.7.** Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(CX) = C^2 D(X).$$

**Свойство 6.8.** Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:

$$D(X + Y) = D(X) + D(Y).$$

**Следствие 6.3.** Дисперсия суммы нескольких независимых случайных величин равна сумме дисперсий этих величин.

**Следствие 6.4.** Дисперсия суммы постоянной величины и случайной величины равна дисперсии случайной величины:

$$D(C + X) = D(X).$$

Доказательство. Величины  $C$  и  $X$  независимы, поэтому по свойству 6.8 имеем

$$D(C + X) = D(C) + D(X).$$

В силу свойства 6.6  $D(C) = 0$ . Следовательно,

$$D(C + X) = D(X).$$

Свойство становится ясным, если учесть, что величины  $X$  и  $X + C$  отличаются лишь началом отсчета и, значит, рассеяны вокруг своих математических ожиданий одинаково.

**Свойство 6.9.** Дисперсия разности двух независимых случайных величин равна сумме их дисперсий:

$$D(X - Y) = D(X) + D(Y).$$

Доказательство. В силу свойства 6.8 имеем

$$D(X - Y) = D(X + (-Y)) = D(X) + D(-Y).$$

По свойству 6.7,

$$D(X - Y) = D(X) + (-1)^2 \cdot D(Y).$$

или

$$D(X - Y) = D(X) + D(Y).$$

**Свойство 6.10.** Дисперсия числа появлений события  $A$  в  $n$  независимых испытаниях, в каждом из которых вероятность  $p$  появления события постоянна, равна произведению числа испытаний на вероятности появления и непоявления события в одном испытании:

$$D(X) = npq.$$

**Пример 7.** ГНИ проводит 10 независимых проверок предприятий, в каждой из которых вероятность выявления

ошибок в документации равна  $p = 0,6$ . Найти дисперсию случайной величины  $X$  — числа выявлений ошибок в документации в этих проверках.

Решение. По условию,  $n = 10$ ,  $p = 0,6$ . Вероятность невыявления ошибок в документации равна

$$q = 1 - 0,6 = 0,4.$$

Искомая дисперсия равна

$$D(X) = npq = 10 \cdot 0,6 \cdot 0,4 = 2,4.$$

Для оценки рассеяния возможных значений случайной величины вокруг ее среднего значения служит также среднее квадратическое отклонение.

*Средним квадратическим отклонением случайной величины  $X$  называется квадратный корень из дисперсии:*

$$\sigma(X) = \sqrt{D(X)}. \quad (6.5)$$

**Пример 8.** Случайная величина  $X$  задана следующим законом распределения:

Т а б л и ц а 6.8

|       |     |     |     |
|-------|-----|-----|-----|
| $x_i$ | 2   | 3   | 10  |
| $p_i$ | 0,1 | 0,4 | 0,5 |

Найти среднее квадратическое отклонение  $\sigma(X)$ .

Решение. Математическое ожидание  $M(X)$  равно:

$$M(X) = 2 \cdot 0,1 + 3 \cdot 0,4 + 10 \cdot 0,5 = 6,4.$$

Математическое ожидание  $M(X^2)$  равно:

$$M(X^2) = 4 \cdot 0,1 + 9 \cdot 0,4 + 100 \cdot 0,5 = 54.$$

Найдем дисперсию:

$$D(X) = M(X^2) - [M(X)]^2 = 54 - (6,4)^2 = 13,04.$$

Искомое среднее квадратическое отклонение равно:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{13,04} \approx 3,61.$$

### Вопросы для повторения и контроля:

1. Что называется числовыми характеристиками случайной величины и какие их виды вы знаете?
2. Что такое математическое ожидание и как оно определяется?

3. Чему равно математическое ожидание числа появлений события в одном испытании и как оно находится?
4. Что вы знаете о 1- и 2-свойствах математического ожидания (свойства 6.1 и 6.2)?
5. Какие случайные величины называются независимыми и что является произведением независимых случайных величин?
6. Как определяется сумма случайных величин?
7. Что вы знаете о 3- и 4-свойствах математического ожидания, а также об их следствиях (свойства 6.3 и 6.4, следствия 6.1 и 6.2)?
8. В чем целесообразность введения других числовых характеристик случайной величины, кроме математического ожидания, и что такое отклонение случайной величины?
9. Что такое дисперсия и как она находится?
10. Что вы знаете о 1- и 2-свойствах дисперсии (свойства 6.6 и 6.7)?
11. Что вы знаете о 3-свойстве дисперсии и его следствиях (свойство 6.8, следствия 6.3 и 6.4)?
12. Что вы знаете о 4-свойстве дисперсии (свойство 6.9)?
13. Чему равны математическое ожидание и дисперсия числа появлений события  $A$  в  $n$  независимых испытаниях (свойства 6.5 и 6.10)?
14. Что такое среднее квадратическое отклонение и как оно определяется?

### **Опорные слова:**

Числовые характеристики случайной величины, математическое ожидание, независимые случайные величины, произведение независимых случайных величин, сумма случайных величин, отклонение случайной величины, дисперсия, среднее квадратическое отклонение.

## Тема № 7

### Функции распределения и плотности непрерывных случайных величин, их свойства

#### План:

1. Функция распределения случайной величины.
2. Свойства функции распределения.
3. Функция плотности непрерывной случайной величины.
4. Свойства функции плотности.

Дискретная случайная величина может быть задана перечнем всех ее возможных значений и их вероятностей. Однако такой способ задания неприменим для непрерывных случайных величин.

Например, рассмотрим случайную величину  $X$ , возможные значения которой сплошь заполняют интервал  $(a, b)$ . Очевидно, что невозможно составить перечень всех возможных значений  $X$ . Поэтому целесообразно дать общий способ задания любых типов случайных величин, для чего вводятся функции распределения вероятностей случайной величины.

Пусть  $x$  — действительное число. Вероятность события, состоящего в том, что  $X$  примет значение, меньшее  $x$ , т.е. вероятность события  $X < x$ , обозначим через  $F(x)$ . Если  $x$  изменяется, то изменяется и  $F(x)$ , т.е.  $F(x)$  — функция от  $x$ .

*Функцией распределения случайной величины  $X$*  называется функция  $F(x)$ , определяющая вероятность того, что случайная величина  $X$  в результате испытания примет значение, меньшее  $x$ , т.е.

$$F(x) = P(X < x). \quad (7.1)$$

Геометрически это равенство можно истолковать так:  $F(x)$  есть вероятность того, что случайная величина примет значение, которое изображается на числовой оси точкой, лежащей левее точки  $x$ .

Теперь рассмотрим свойства функции распределения.

**Свойство 7.1.** *Значения функции распределения*

принадлежат отрезку  $[0, 1]$ :

$$0 \leq F(x) \leq 1. \quad (7.2)$$

Доказательство. Свойство вытекает из определения функции распределения как вероятности: вероятность всегда есть неотрицательное число, не превышающее единицы.

**Свойство 7.2.**  $F(x)$  — неубывающая функция, т.е.:

$$F(x_1) \leq F(x_2), \text{ если } x_1 < x_2. \quad (7.3)$$

**Следствие 7.1.** Вероятность того, что случайная величина примет значение, заключенное в интервале  $(a, b)$ , равна приращению функции распределения на этом интервале:

$$P(a \leq X < b) = F(b) - F(a). \quad (7.4)$$

**Пример 1.** Случайная величина  $X$  задана следующей функцией распределения:

$$F(x) = \begin{cases} 0 & \text{при } x \leq -1 \\ x/4 + 1/4 & \text{при } -1 < x \leq 3. \\ 1 & \text{при } x > 3 \end{cases}$$

Найти вероятность того, что в результате испытания  $X$  примет значение, принадлежащее интервалу  $(0, 2)$ :

$$P(0 < X < 2) = F(2) - F(0).$$

Решение. Так как на интервале  $(0, 2)$ , по условию,

$$F(x) = x/4 + 1/4,$$

то

$$F(2) - F(0) = (2/4 + 1/4) - (0/4 + 1/4) = 1/2.$$

Итак,

$$P(0 < X < 2) = 1/2.$$

**Следствие 7.2.** Вероятность того, что непрерывная случайная величина  $X$  примет одно определенное значение, равна нулю.

**Свойство 7.3.** Если возможные значения случайной величины принадлежат интервалу  $(a, b)$ , то: 1)  $F(x) = 0$  при  $x \leq a$ ; 2)  $F(x) = 1$  при  $x \geq b$ .

Доказательство. 1) Пусть  $x_1 \leq a$ . Тогда событие  $X < x_1$  невозможно (так как значений, меньших  $x_1$ , величина  $X$  по

условию не принимает) и, следовательно, вероятность его равна нулю.

2) Пусть  $x_2 \geq b$ . Тогда событие  $X < x_2$  достоверно (так как все возможные значения  $X$  меньше  $x_2$ ) и, следовательно, вероятность его равна единице.

**Следствие 7.3.** Если возможные значения непрерывной случайной величины расположены на всей числовой оси  $x$ , то справедливы следующие предельные соотношения:

$$\lim_{x \rightarrow -\infty} F(x) = 0; \quad \lim_{x \rightarrow \infty} F(x) = 1. \quad (7.5)$$

График функции распределения непрерывной случайной величины в силу свойства 7.1 расположен в полосе, ограниченной прямыми  $y = 0$ ,  $y = 1$ .

Из свойства 7.2 вытекает, что при возрастании  $x$  в интервале  $(a, b)$ , в котором заключены все возможные значения случайной величины, график имеет вид либо наклона вверх, либо горизонтальный.

В силу свойства 7.3 при  $x \leq a$  ординаты графика равны нулю; при  $x \geq b$  ординаты графика равны единице.

График функции распределения непрерывной случайной величины изображен на рис. 7.1.

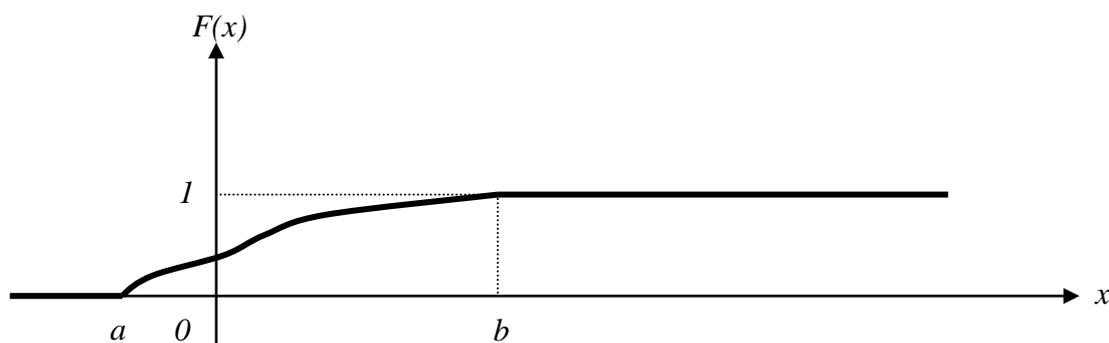


Рис. 7.1.

График функции распределения дискретной случайной величины имеет ступенчатый вид.

**Пример 2.** Дискретная случайная величина  $X$  задана следующим законом распределения:

Т а б л и ц а 7.1

|       |     |     |     |
|-------|-----|-----|-----|
| $x_i$ | 1   | 4   | 8   |
| $p_i$ | 0,3 | 0,1 | 0,6 |

Найти функцию распределения и вычертить ее график.

Решение. Если  $x \leq 1$ , то по свойству 7.3  $F(x) = 0$ .

Если  $1 < x \leq 4$ , то  $F(x) = 0,3$ . Действительно,  $X$  может принять значение 1 с вероятностью 0,3.

Если  $4 < x \leq 8$ , то  $F(x) = 0,4$ . Действительно, если  $x_1$  удовлетворяет неравенству  $4 < x_1 \leq 8$ , то  $F(x_1)$  равно вероятности события  $X < x_1$ , которое может быть осуществлено, когда  $X$  примет значение 1 с вероятностью 0,3 или значение 4 с вероятностью 0,1. Поскольку эти два события несовместны, то по теореме 3.1 вероятность события  $X < x_1$  равна сумме вероятностей  $0,3 + 0,1 = 0,4$ .

Если  $x > 8$ , то по свойству 7.3  $F(x) = 1$ .

Итак, функция распределения аналитически может быть записана так:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 1 \\ 0,3 & \text{при } 1 < x \leq 4 \\ 0,4 & \text{при } 4 < x \leq 8 \\ 1 & \text{при } x > 8 \end{cases}$$

График этой функции приведен на рис. 7.2.

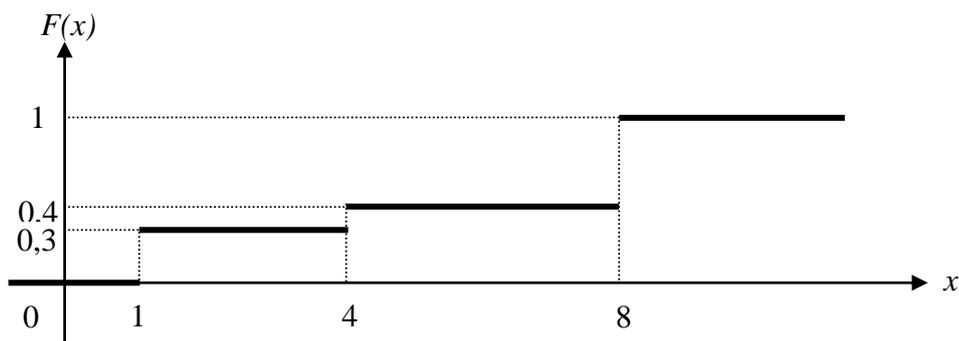


Рис. 7.2.

Непрерывную случайную величину можно также задать, используя другую функцию, которая называется функцией плотности.

*Функцией плотности непрерывной случайной величины  $X$*

называется функция  $f(x)$  — первая производная от функции распределения  $F(x)$ :

$$f(x) = F'(x). \quad (7.6)$$

Отсюда следует, что функция распределения является первообразной для функции плотности. Для описания распределения вероятностей дискретной случайной величины функция плотности неприменима.

Зная функцию плотности, можно вычислить вероятность того, что непрерывная случайная величина примет значение, принадлежащее заданному интервалу.

**Теорема 7.1.** *Вероятность того, что непрерывная случайная величина  $X$  примет значение, принадлежащее интервалу  $(a, b)$ , равна определенному интегралу от функции плотности, взятому в пределах от  $a$  до  $b$ :*

$$P(a < X < b) = \int_a^b f(x) dx. \quad (7.7)$$

Доказательство. Из формулы (7.4) получаем

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

По формуле Ньютона–Лейбница

$$F(b) - F(a) = \int_a^b F'(x) dx = \int_a^b f(x) dx.$$

Таким образом,

$$P(a \leq X < b) = \int_a^b f(x) dx.$$

Так как  $P(a \leq X < b) = P(a < X < b)$ , то получаем

$$P(a < X < b) = \int_a^b f(x) dx.$$

**Пример 3.** Задана функция плотности случайной величины  $X$ :

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ 2x & \text{при } 0 < x \leq 1. \\ 0 & \text{при } x > 1 \end{cases}$$

Найти вероятность того, что в результате испытания  $X$  примет значение, принадлежащее интервалу  $(0,5; 1)$ .

Решение. Искомая вероятность по формуле (7.7) равна

$$P(0,5 < X < 1) = \int_{0,5}^1 2x \, dx = x^2 \Big|_{0,5}^1 = 1 - 0,25 = 0,75.$$

Зная функцию плотности распределения  $f(x)$ , можно найти функцию распределения  $F(x)$  по формуле

$$F(x) = \int_{-\infty}^x f(z) \, dz. \quad (7.8)$$

**Пример 4.** Найти функцию распределения по данной функции плотности:

$$f(x) = \begin{cases} 0 & \text{при } x \leq a \\ 1/(b-a) & \text{при } a < x \leq b. \\ 0 & \text{при } x > b \end{cases}$$

Построить график найденной функции.

Решение. Воспользуемся формулой (7.8). Если  $x \leq a$ , то  $f(x) = 0$ , следовательно,  $F(x) = 0$ . Если  $a < x \leq b$ , то  $f(x) = 1/(b-a)$ , следовательно,

$$F(x) = \int_{-\infty}^x f(z) \, dz = \int_{-\infty}^a 0 \, dz + \int_a^x \frac{1}{b-a} \, dz = \frac{x-a}{b-a}.$$

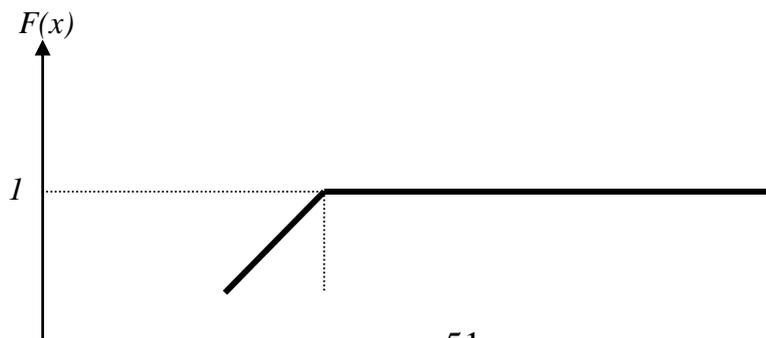
Если  $x > b$ , то

$$F(x) = \int_{-\infty}^a 0 \, dz + \int_a^b \frac{1}{b-a} \, dz + \int_b^x 0 \, dz = \frac{b-a}{b-a} = 1.$$

Итак, искомая функция распределения имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \leq a \\ (x-a)/(b-a) & \text{при } a < x \leq b. \\ 1 & \text{при } x > b \end{cases}$$

График этой функции изображен на рис. 7.3.



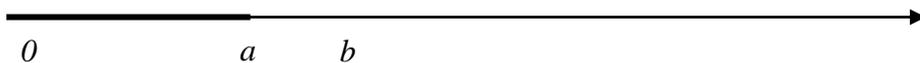


Рис. 7.3.

Приведем два свойства функции плотности.

**Свойство 7.4.** *Функция плотности — неотрицательная функция:*

$$f(x) \geq 0. \quad (7.9)$$

Доказательство. Функция распределения — неубывающая функция, следовательно, ее производная  $F'(x) = f(x)$  — функция неотрицательная.

**Свойство 7.5.** *Несобственный интеграл от функции плотности распределения в пределах от  $-\infty$  до  $\infty$  равен единице:*

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (7.10)$$

### Вопросы для повторения и контроля:

1. Почему целесообразно дать общий способ задания любых типов случайных величин?
2. Что называется функцией распределения случайной величины?
3. Что вы знаете о 1-свойстве функции распределения (свойство 7.1)?
4. Что вы знаете о 2-свойстве функции распределения и его следствиях (свойство 7.2, следствия 7.1 и 7.2)?
5. Что вы знаете о 3-свойстве функции распределения и его следствии (свойство 7.3 и следствие 7.3)?
6. Какими свойствами обладают графики функций распределения непрерывной и дискретной случайной величин?
7. Что называется функцией плотности непрерывной случайной величины и что вы знаете о теореме 7.1?
8. Как можно найти функцию распределения, зная функцию плотности распределения и что вы знаете о свойствах функции плотности (свойства 7.4 и 7.5)?

### Опорные слова:

Функция распределения случайной величины, график функции распределения непрерывной случайной величины, график функции распределения дискретной случайной величины, функция плотности непрерывной случайной величины

## Тема № 8

### Числовые характеристики непрерывных случайных величин. Виды непрерывных распределений

#### План:

1. Числовые характеристики непрерывных случайных величин.
2. Нормальное распределение.
3. Равномерное и показательное распределения.

Как и дискретные случайные величины, непрерывные случайные величины также имеют числовые характеристики. Рассмотрим математическое ожидание и дисперсию непрерывной случайной величины.

Пусть непрерывная случайная величина  $X$  задана функцией плотности  $f(x)$  и возможные значения этой случайной величины принадлежат отрезку  $[a, b]$ .

*Математическим ожиданием непрерывной случайной величины  $X$ , возможные значения которой принадлежат отрезку  $[a, b]$ , называется следующий определенный интеграл*

$$M(X) = \int_a^b x f(x) dx. \quad (8.1)$$

Если возможные значения принадлежат всей числовой оси  $Ox$ , то математическое ожидание имеет следующий вид

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx. \quad (8.2)$$

*Дисперсией непрерывной случайной величины  $X$ , возможные значения которой принадлежат отрезку  $[a, b]$ , называется следующий определенный интеграл*

$$D(X) = \int_a^b [x - M(X)]^2 f(x) dx. \quad (8.3)$$

Если возможные значения принадлежат всей числовой оси  $Ox$ , то дисперсия имеет следующий вид

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x) dx. \quad (8.4)$$

Для вычисления дисперсии более удобны соответственно следующие формулы

$$D(X) = \int_a^b x^2 f(x) dx - [M(X)]^2 \quad (8.5)$$

и

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - [M(X)]^2. \quad (8.6)$$

Свойства математического ожидания и дисперсии дискретных случайных величин сохраняются и для непрерывных случайных величин.

*Среднее квадратическое отклонение непрерывной случайной величины* определяется, как и для дискретной случайной величины, следующим равенством

$$\sigma(X) = \sqrt{D(X)}. \quad (8.7)$$

**Пример 1.** Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины  $X$ , заданной следующей функцией распределения:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ x & \text{при } 0 < x \leq 1. \\ 1 & \text{при } x > 1 \end{cases}$$

Решение. Найдем функцию плотности:

$$f(x) = F'(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ 1 & \text{при } 0 < x \leq 1. \\ 0 & \text{при } x > 1 \end{cases}$$

Найдем математическое ожидание по формуле (8.1):

$$M(X) = \int_0^1 x \cdot 1 \cdot dx = x^2/2 \Big|_0^1 = 1/2.$$

Найдем дисперсию по формуле (8.5):

$$D(X) = \int_0^1 x^2 \cdot 1 \cdot dx - [1/2]^2 = x^3/3 \Big|_0^1 - 1/4 = 1/12.$$

Найдем среднее квадратическое отклонение по формуле (8.7):

$$\sigma(X) = \sqrt{D(X)} = \sqrt{1/12} \approx 0,29.$$

При решении задач, которые выдвигает практика, приходится сталкиваться с различными распределениями непрерывных случайных величин. Функции плотности непрерывных случайных величин называются также *законами распределений*. Наиболее часто встречаются законы нормального, равномерного и показательного распределений.

*Нормальным распределением с параметрами  $a$  и  $\sigma$  ( $\sigma > 0$ )* называется распределение вероятностей непрерывной случайной величины, которое описывается следующей функцией плотности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (8.8)$$

Отсюда видно, что нормальное распределение определяется двумя параметрами:  $a$  и  $\sigma$ . Достаточно знать эти параметры, чтобы задать нормальное распределение.

Отметим вероятностный смысл этих параметров. Итак,  $M(X) = a$ , т.е. *математическое ожидание нормального распределения равно параметру  $a$* , и  $\sigma(X) = \sigma$ , т.е. *среднее квадратическое отклонение нормального распределения равно параметру  $\sigma$* .

Функция распределения нормальной случайной величины имеет вид

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(z-a)^2}{2\sigma^2}} dz. \quad (8.9)$$

*Общим* называется нормальное распределение с произвольными параметрами  $a$  и  $\sigma$  ( $\sigma > 0$ ). *Стандартным* называется нормальное распределение с параметрами  $a = 0$  и  $\sigma = 1$ .

Легко заметить, что функция плотности стандартного нормального распределения имеет следующий вид

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (8.10)$$

Эта функция уже встречалась нам в теме № 4. Ее значения приведены в специальных таблицах в различной литературе по

теории вероятностей и математической статистике.

Вероятность попадания нормальной случайной величины с произвольными параметрами  $a$  и  $\sigma$  в интервал  $(\alpha, \beta)$  можно

найти, пользуясь функцией Лапласа  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ .

Действительно, по теореме 7.1 имеем

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

Введем новую переменную  $z = (x-a)/\sigma$ . Отсюда  $x = \sigma z + a$ ,  $dx = \sigma dz$ . Найдем новые пределы интегрирования. Если  $x = \alpha$ , то  $z = (\alpha-a)/\sigma$ ; если  $x = \beta$ , то  $z = (\beta-a)/\sigma$ .

Таким образом, имеем

$$\begin{aligned} P(\alpha < X < \beta) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{(\alpha-a)/\sigma}^{(\beta-a)/\sigma} e^{-\frac{z^2}{2}} (\sigma dz) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{(\alpha-a)/\sigma}^0 e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_0^{(\beta-a)/\sigma} e^{-\frac{z^2}{2}} dz = \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{(\beta-a)/\sigma} e^{-\frac{z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_0^{(\alpha-a)/\sigma} e^{-\frac{z^2}{2}} dz. \end{aligned}$$

Используя функцию  $\Phi(x)$ , окончательно получим

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \quad (8.11)$$

В частности, вероятность попадания стандартной нормальной случайной величины  $X$  в интервал  $(0, x)$  равна

$$P(0 < X < x) = \Phi(x), \quad (8.12)$$

так как в этом случае  $a = 0$  и  $\sigma = 1$ .

**Пример 2.** 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. Воспользуемся формулой (8.11). По условию,  $\alpha = 10$ ,  $\beta = 50$ ,  $a = 30$ ,  $\sigma = 10$ , 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  $\Phi(2) = 0,4772$ . Отсюда искомая вероятность равна

$$P(10 < X < 50) = 2 \cdot 0,4772 = 0,9544.$$

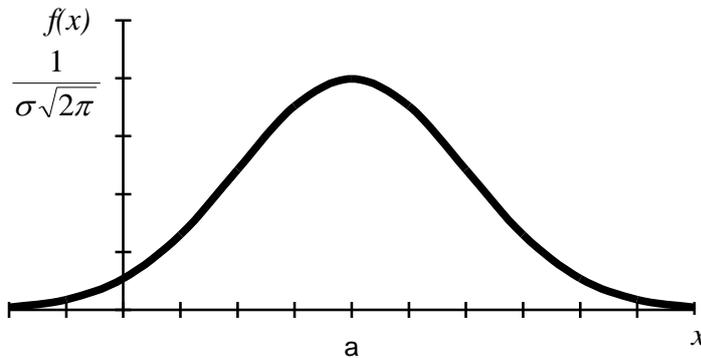


Рис. 8.1.

График функции плотности нормального распределения называется *нормальной кривой* (*кривой Гаусса*). Этот график изображен на рис. 8.1.

*Равномерным распределением на отрезке  $[a, b]$*  называется распределение вероятностей случайной величины  $X$ , все возможные значения которой принадлежат этому отрезку, если ее функция плотности имеет вид

$$f(x) = \begin{cases} 0 & \text{при } x \leq a \\ 1/(b-a) & \text{при } a < x \leq b. \\ 0 & \text{при } x > b \end{cases} \quad (8.13)$$

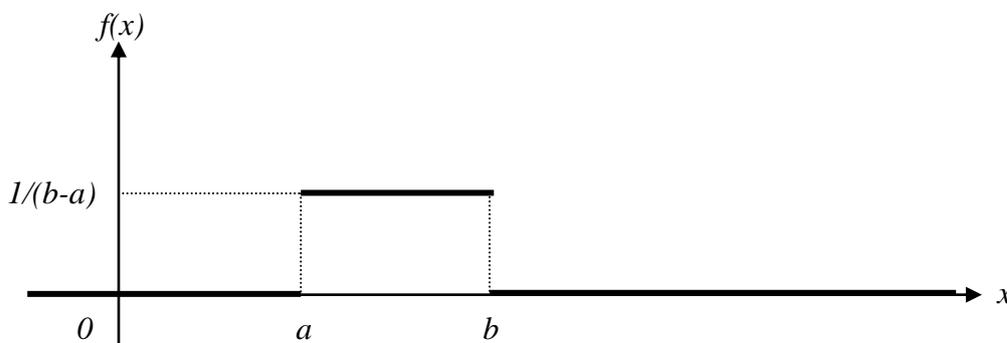


Рис. 8.2.

Функция распределения равномерно распределенной на  $[a, b]$  случайной величины имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \leq a \\ (x-a)/(b-a) & \text{при } a < x \leq b. \\ 1 & \text{при } x > b \end{cases} \quad (8.14)$$

График функции плотности равномерного распределения приведен на рис. 8.2, а график функции распределения — на рис. 7.3.

Вычислим математическое ожидание и дисперсию равномерной случайной величины. По формуле (8.1) имеем

$$M(X) = \int_a^b x f(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Далее, по формуле (8.5) имеем

$$\begin{aligned} D(X) &= \int_a^b x^2 f(x) dx - [M(X)]^2 = \frac{1}{b-a} \int_a^b x^2 dx - \left[ \frac{a+b}{2} \right]^2 = \\ &= \frac{x^3}{3(b-a)} \Big|_a^b - \left[ \frac{a+b}{2} \right]^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}. \end{aligned}$$

Теперь найдем вероятность попадания непрерывной случайной величины  $X$ , распределенной равномерно на  $[a, b]$ , в интервал  $(c, d)$ , принадлежащий  $[a, b]$ .

Используя теорему 7.1 и формулу (8.13), имеем

$$P(c < X < d) = \int_c^d f(x) dx = \int_c^d \frac{1}{b-a} dx = \frac{1}{b-a} \int_c^d 1 \cdot dx = \frac{d-c}{b-a}$$

или

$$P(c < X < d) = \frac{d-c}{b-a}. \quad (8.15)$$

*Показательным (экспоненциальным) распределением* называется распределение вероятностей непрерывной случайной величины  $X$ , которое описывается функцией плотности

$$f(x) = \begin{cases} 0 & \text{при } x < 0 \\ \lambda e^{-\lambda x} & \text{при } x \geq 0 \end{cases}, \quad (8.16)$$

где  $\lambda$  — постоянная положительная величина.

Из определения видно, что показательное распределение определяется одним параметром  $\lambda$ . Найдем функцию распределения показательного закона:

$$F(x) = \int_{-\infty}^x f(z) dz = \int_{-\infty}^0 0 dz + \lambda \int_0^x e^{-\lambda z} dz = 1 - e^{-\lambda x}.$$

Итак,

$$F(x) = \begin{cases} 0 & \text{при } x < 0 \\ 1 - e^{-\lambda x} & \text{при } x \geq 0 \end{cases}. \quad (8.17)$$

Графики функций плотности и распределения показательного закона изображены на рис. 8.3.

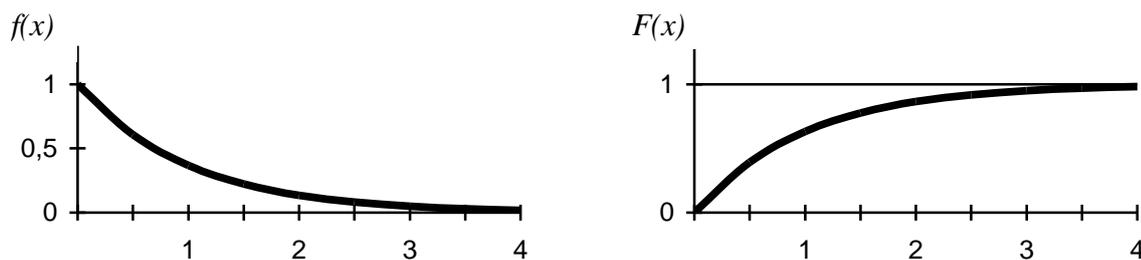


Рис. 8.3.

Найдем вероятность попадания в интервал  $(a, b)$  непрерывной случайной величины  $X$ , которая распределена по показательному закону из формулы (8.17). Используя формулу (7.4), имеем

$$P(a < X < b) = F(b) - F(a) = 1 - e^{-\lambda b} - (1 - e^{-\lambda a})$$

или

$$P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}. \quad (8.18)$$

**Пример 3.** Непрерывная случайная величина  $X$  распределена по показательному закону

$$f(x) = \begin{cases} 0 & \text{при } x < 0 \\ 2e^{-2x} & \text{при } x \geq 0 \end{cases}.$$

Найти вероятность того, что в результате испытания  $X$  попадает в интервал  $(0,3; 1)$ .

Решение. По условию,  $\lambda = 2$ . Воспользуемся формулой (8.18):

$$P(0,3 < X < 1) = e^{-(2 \cdot 0,3)} - e^{-(2 \cdot 1)} = e^{-0,6} - e^{-2} \approx 0,548 - 0,135 \approx 0,41$$

Отметим вероятностный смысл параметра показательного распределения. *Математическое ожидание и среднее квадратическое отклонение показательного распределения*

равны обратной величине параметра  $\lambda$ , т.е.  $M(X) = 1/\lambda$  и  $\sigma(X) = 1/\lambda$ .

**Пример 4.** Непрерывная случайная величина  $X$  распределена по показательному закону

$$f(x) = \begin{cases} 0 & \text{при } x < 0 \\ 5e^{-5x} & \text{при } x \geq 0 \end{cases}.$$

Найти математическое ожидание, среднее квадратическое отклонение и дисперсию случайной величины  $X$ .

Решение. По условию,  $\lambda = 5$ . Следовательно,

$$M(X) = \sigma(X) = 1/\lambda = 1/5 = 0,2;$$

$$D(X) = [\sigma(X)]^2 = 1/\lambda^2 = 1/5^2 = 0,04.$$

### Вопросы для повторения и контроля:

1. Что является математическим ожиданием непрерывной случайной величины?
2. Что является дисперсией непрерывной случайной величины и как она вычисляется?
3. Что называется нормальным распределением?
4. Каков вероятностный смысл параметров нормального распределения?
5. Что такое общее и стандартное нормальные распределения, каковы их функции плотности и распределения?
6. Как находится вероятность попадания нормальной случайной величины в заданный интервал?
7. Что называется равномерным распределением?
8. Как вычисляется математическое ожидание и дисперсия равномерной случайной величины?
9. Как находится вероятность попадания равномерной случайной величины в заданный интервал?
10. Что называется показательным распределением?
11. Как находится вероятность попадания показательной случайной величины в заданный интервал?
12. Каков вероятностный смысл параметра показательного распределения?

## **Опорные слова:**

Математическое ожидание непрерывной случайной величины, дисперсия непрерывной случайной величины, закон распределения, нормальное распределение, общее нормальное распределение, стандартное нормальное распределение, вероятность попадания нормальной случайной величины в заданный интервал, нормальная кривая (кривая Гаусса), равномерное распределение, вероятность попадания равномерной случайной величины в заданный интервал, показательное распределение, вероятность попадания показательной случайной величины в заданный интервал.

## Тема № 9

### Закон больших чисел и его практическое значение. Понятие о центральной предельной теореме

#### План:

1. Закон больших чисел.
2. Центральная предельная теорема.

Как мы видели в предыдущих темах, нельзя заранее уверенно предвидеть, какое из возможных значений примет случайная величина в итоге испытания, потому что это зависит от многих случайных причин, учесть которые невозможно. Однако при некоторых сравнительно широких условиях суммарное поведение достаточно большого числа случайных величин почти утрачивает случайный характер и становится закономерным.

Для практики очень важно знание условий, при выполнении которых совокупное действие очень многих случайных причин приводит к результату, почти не зависящему от случая, так как позволяет предвидеть ход явлений. Эти условия и указываются в теоремах, носящих общее название *закона больших чисел*. К ним относятся теоремы Чебышева и Бернулли.

Теоремы, относящиеся к закону больших чисел, устанавливают условия сходимости среднего арифметического  $n$  случайных величин к среднему арифметическому их математических ожиданий.

Вначале приведем неравенство Чебышева, на которое опираются доказательства вышеназванных теорем.

Если известна дисперсия случайной величины, то с ее помощью можно оценить вероятность отклонения этой величины на заданное значение от своего математического ожидания, причем оценка вероятности отклонения зависит лишь от дисперсии. Соответствующую оценку вероятности дает *неравенство П.Л. Чебышева*:

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}, \quad \varepsilon > 0. \quad (9.1)$$

Из этого неравенства в качестве следствия можно получить следующее неравенство

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}, \quad \varepsilon > 0. \quad (9.2)$$

**Пример 1.** Оценить вероятность отклонения случайной величины  $X$  от своего математического ожидания на величину, превышающую утроенное среднеквадратическое отклонение случайной величины.

Решение. По условию,  $\varepsilon = 3\sigma(X)$ . Учитывая, что  $D(X) = [\sigma(X)]^2$ , из формулы (9.1) получаем

$$P(|X - M(X)| \geq 3\sigma(X)) \leq \frac{D(X)}{9[\sigma(X)]^2} = \frac{1}{9}.$$

**Теорема 9.1 (закон больших чисел в форме Чебышева).** Пусть  $X_1, X_2, \dots, X_n, \dots$  — последовательность независимых случайных величин, дисперсии которых ограничены сверху одним и тем же числом  $c$ :  $D(X_i) \leq c$ ,  $i = 1, 2, \dots$ . Тогда для любого  $\varepsilon > 0$  имеет место:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n M(X_i)\right| < \varepsilon\right) = 1. \quad (9.3)$$

Из этой теоремы вытекает справедливость закона больших чисел для среднего арифметического независимых случайных величин, имеющих одинаковое распределение вероятностей.

**Следствие 9.1.** Пусть  $X_1, X_2, \dots, X_n, \dots$  — последовательность независимых случайных величин, имеющих одно и то же математическое ожидание  $a$ , и дисперсии которых ограничены сверху одним и тем же числом  $c$ :  $D(X_i) \leq c$ ,  $i = 1, 2, \dots$ . Тогда для любого  $\varepsilon > 0$  имеет место:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - a\right| < \varepsilon\right) = 1. \quad (9.4)$$

Закон больших чисел для независимых случайных величин с одинаковым математическим ожиданием отражает сходимость среднего арифметического случайных величин в сериях независимых испытаний к общему математическому ожиданию этих случайных величин.

Таким образом, среднее арифметическое достаточно большого числа независимых случайных величин (дисперсии которых равномерно ограничены) утрачивает характер случайной величины. Объясняется это тем, что отклонения

каждой из этих величин от своих математических ожиданий могут быть как положительными, так и отрицательными, а в среднем арифметическом они погашаются.

Закон больших чисел имеет многочисленные практические приложения. Пусть, например, производится  $n$  независимых измерений некоторой величины, истинное значение которой равно  $a$ . Результат каждого измерения является случайной величиной  $X_i$ . Если измерения выполняются без систематической погрешности, то математическое ожидание случайных величин  $X_i$  можно считать равным истинному значению измеряемой величины,  $M(X_i) = a$ ,  $i = 1, 2, \dots$ . Дисперсию результатов измерений часто можно считать ограниченной некоторым числом  $c$ .

Тогда случайные результаты измерений удовлетворяют условиям теоремы 9.1 и, следовательно, среднее арифметическое  $n$  измерений при большом числе измерений практически не может сильно отличаться от истинного значения измеряемой величины  $a$ . Этим обосновывается выбор среднего арифметического измерений в качестве истинного значения измеряемой величины.

Для относительной частоты успехов в независимых испытаниях справедлива следующая теорема.

**Теорема 9.2 (закон больших чисел в форме Бернулли).**  
*Если в каждом из  $n$  независимых испытаний вероятность  $p$  появления события  $A$  постоянна, то для числа успехов  $m$  в этих испытаниях при любом  $\varepsilon > 0$  имеет место:*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1. \quad (9.5)$$

Рассмотрим последовательность независимых, одинаково распределенных случайных величин  $X_1, X_2, \dots, X_n, \dots$ . Пусть  $M(X_i) = a$ ,  $D(X_i) = \sigma^2$ ,  $i = 1, 2, \dots$ . образуем последовательность  $Y_n$ ,  $n = 1, 2, \dots$ , центрированных и нормированных сумм случайных величин:

$$Y_n = \frac{\sum_{i=1}^n X_i - na}{\sigma \sqrt{n}}. \quad (9.6)$$

Согласно центральной предельной теореме, при достаточно общих предположениях о законах распределения случайных величин  $X_1, X_2, \dots, X_n, \dots$  последовательность функций распределения центрированных и нормированных сумм случайных величин  $Y_n$  при  $n \rightarrow \infty$  сходится для любых  $x$  к функции распределения стандартной нормальной случайной величины.

**Теорема 9.3 (центральная предельная теорема).** Пусть  $X_1, X_2, \dots, X_n, \dots$  — последовательность независимых, одинаково распределенных случайных величин, имеющих конечную дисперсию  $D(X_i) = \sigma^2$ , и пусть  $M(X_i) = a$ ,  $i = 1, 2, \dots$ . Тогда для любого  $x$  ( $-\infty < x < \infty$ ) имеет место:

$$\lim_{n \rightarrow \infty} P(Y_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz. \quad (9.7)$$

### Вопросы для повторения и контроля:

1. О чем идет речь в теоремах, носящих общее название закона больших чисел?
2. Что вы знаете о неравенстве Чебышева?
3. Что утверждает закон больших чисел в форме Чебышева?
4. В чем сущность закона больших чисел и каково его практическое значение?
5. Что утверждает закон больших чисел в форме Бернулли?
6. О чем идет речь в центральной предельной теореме?

### Опорные слова:

Закон больших чисел, неравенство Чебышева, последовательность независимых случайных величин, центрированная и нормированная сумма случайных величин, центральная предельная теорема.

## Тема № 10

### Предмет и основные задачи математической статистики. Выборка

#### План:

1. Задачи математической статистики.
2. Генеральная и выборочная совокупности.
3. Типы выборок, способы отбора.

Основная цель при применении математической статистики состоит в получении выводов о массовых явлениях и процессах по данным наблюдений над ними или экспериментов. Эти статистические выводы относятся не к отдельным испытаниям, а представляют собой утверждения об общих характеристиках этого явления (вероятностях, законах распределения и их параметрах, математических ожиданиях и т.п.) в предположении постоянства условий, порождающих исследуемое явление.

Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении методами теории вероятностей статистических данных — результатов наблюдений.

Первая задача математической статистики — указать способы сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов.

Вторая задача математической статистики — разработать методы анализа статистических данных в зависимости от целей исследования, таких, как:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин и др.;

б) проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

Итак, *предметом математической статистики является*

*создание методов сбора и обработки статистических данных для получения научных и практических выводов.*

Математическая статистика опирается на теорию вероятностей, и ее цель — оценить характеристики генеральной совокупности по выборочным данным.

Если требуется изучить совокупность однородных объектов относительно некоторого признака, характеризующего эти объекты, то естественным является проведение сплошного обследования, т.е. обследование каждого из объектов совокупности относительно этого признака. На практике, однако, проведение сплошного обследования по тем или иным причинам часто бывает невозможным. В таких случаях случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению.

*Выборочной совокупностью* или просто *выборкой* называется совокупность случайно отобранных объектов. *Генеральной совокупностью* называется совокупность объектов, из которых производится выборка. Например, если все студенты Налоговой академии — это генеральная совокупность, то студенты какой-либо группы являются выборкой.

*Объемом* совокупности (выборочной или генеральной) называется число объектов этой совокупности. Например, если из 1000 деталей отобрано для обследования 100 деталей, то объем генеральной совокупности  $N = 1000$ , а объем выборки  $n = 100$ .

При составлении выборки можно поступать двумя способами: после того, как объект отобран и над ним произведено наблюдение, он может быть возвращен либо не возвращен в генеральную совокупность. В зависимости от этого выборки подразделяются на повторные и бесповторные.

*Повторной* называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность. *Бесповторной* называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

Для того, чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности,

т.е. выборка должна быть *репрезентативной* (*представительной*).

В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если каждый объект выборки отобран случайно из генеральной совокупности в предположении, что все объекты имеют одинаковую вероятность попасть в выборку.

Если объем генеральной совокупности достаточно велик, а выборка составляет лишь незначительную часть этой совокупности, то различие между повторной и бесповторной выборками стирается; в предельном случае, когда рассматривается бесконечная генеральная совокупность, а выборка имеет конечный объем, это различие исчезает.

На практике применяются различные способы отбора. Существует отбор, не требующий расчленения генеральной совокупности на части, например, простой случайный бесповторный отбор и простой случайный повторный отбор, а также применяется отбор, при котором генеральная совокупность разбивается на части (типический отбор, механический отбор, серийный отбор).

*Простым случайным* называется такой отбор, при котором объекты извлекаются по одному из всей генеральной совокупности. Если извлеченные объекты возвращаются в генеральную совокупность для участия в последующем отборе, то такой отбор будет простым случайным повторным, в противном случае — простым случайным бесповторным. Например, если требуется определить среднемесячную зарплату по региону, то применяется простой случайный бесповторный отбор, так как зарплата одного и того же человека учитывается только один раз. Если же требуется определить половозрастной, социальный, образовательный состав различных комиссий в каком-либо районе, то отбор является простым случайным повторным, так как один и тот же работник может участвовать в различных комиссиях, и, следовательно, попасть в выборку несколько раз.

*Типическим* называется отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части. Например, если детали изготавливаются на нескольких станках, то отбор производится не из всей

совокупности деталей, произведенных всеми станками, а из продукции каждого станка в отдельности. Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности.

*Механическим* называется отбор, при котором генеральная совокупность механически делится на столько примерно одинаковых по размеру групп, сколько объектов должно войти в выборку, а из каждой группы отбирается объект с одним и тем же номером. Например, если нужно отобрать 20% изготовленных станком деталей, то отбирается каждая пятая деталь; если требуется отобрать 5% деталей, то отбирается каждая двадцатая деталь и т.д. Иногда механический отбор может не обеспечить репрезентативности выборки.

*Серийным* называется отбор, при котором объекты отбираются из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию. Например, если изделия изготавливаются большой группой станков-автоматов, то подвергается сплошному обследованию продукция только нескольких станков. Серийным отбором пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

На практике часто применяется комбинированный отбор, при котором сочетаются указанные выше способы.

### **Вопросы для повторения и контроля:**

1. Какие задачи стоят перед математической статистикой?
2. Какова цель применения математической статистики и в чем ее предмет?
3. Что такое выборочная совокупность (выборка), генеральная совокупность, объем совокупности?
4. Что называется повторной выборкой, бесповторной выборкой и репрезентативной выборкой?
5. Что представляет собой простой случайный отбор и типический отбор?
6. Что представляет собой механический отбор и серийный отбор?

### **Опорные слова:**

Математическая статистика, оценка, проверка статистических гипотез, сбор и обработка статистических данных, выборочная совокупность, выборка, генеральная совокупность, объем совокупности, повторная выборка, бесповторная выборка, репрезентативная выборка, простой случайный бесповторный отбор, простой случайный повторный отбор, типический отбор, механический отбор, серийный отбор, комбинированный отбор.

## Тема № 11

### Статистическое распределение выборки. Эмпирическая функция распределения. Полигон и гистограмма

#### План:

1. Статистическое распределение выборки.
2. Эмпирическая функция распределения.
3. Полигон и гистограмма.

Пусть из генеральной совокупности извлечена выборка. При этом пусть значение  $x_1$  наблюдалось  $n_1$  раз,  $x_2$  —  $n_2$  раз, ...,  $x_k$  —  $n_k$  раз и т.д.;  $\sum n_i = n$  является объемом выборки.

Наблюдаемые значения  $x_i$  называются *вариантами*, а последовательность вариантов, записанных в возрастающем порядке, — *вариационным рядом*. Числа наблюдений  $n_i$  называются *частотами*, а их отношения к объему выборки  $n_i/n = W_i$  — *относительными частотами*.

*Статистическим распределением выборки* называется перечень вариант и соответствующих им частот или относительных частот. Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот. В этом случае в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал. При этом сумма частот должна быть равна объему выборки, а сумма относительных частот — единице.

В теории вероятностей под распределением понимается соответствие между возможными значениями случайной величины и их вероятностями, а в математической статистике — соответствие между наблюдаемыми вариантами и их частотами (относительными частотами).

**Пример 1.** Задано распределение частот выборки объема  $n = 20$ :

Т а б л и ц а 11.1

|       |   |   |    |
|-------|---|---|----|
| $x_i$ | 3 | 5 | 10 |
| $n_i$ | 7 | 8 | 5  |

Написать распределение относительных частот.

Решение. Найдем относительные частоты, для чего разделим частоты на объем выборки:

$$W_1 = 7/20 = 0,35, \quad W_2 = 8/20 = 0,4, \quad W_3 = 5/20 = 0,25.$$

Напишем распределение относительных частот:

Т а б л и ц а 11.2

|       |      |     |      |
|-------|------|-----|------|
| $x_i$ | 3    | 5   | 10   |
| $W_i$ | 0,35 | 0,4 | 0,25 |

К о н т р о л ь:  $0,35 + 0,4 + 0,25 = 1.$

Пусть известно статистическое распределение частот количественного признака  $X$ . Обозначим через  $n_x$  число наблюдений, при которых наблюдались значения признака, меньшие  $x$ , а через  $n$  — общее число наблюдений (объем выборки). Относительная частота события  $X < x$  равна  $n_x/n$ . При изменении  $x$  изменяется и относительная частота, т.е. относительная частота  $n_x/n$  есть функция от  $x$ .

*Эмпирической функцией распределения* (функцией распределения выборки) называется функция  $\bar{F}_n(x)$ , определяющая для каждого значения  $x$  относительную частоту события  $X < x$ , т.е.

$$\bar{F}_n(x) = n_x/n, \quad (11.1)$$

где  $n_x$  — число вариантов, меньших  $x$ ;  $n$  — объем выборки.

Функция  $\bar{F}_n(x)$  называется эмпирической, потому что она находится эмпирическим (опытным) путем.

В отличие от эмпирической функции распределения выборки функция распределения  $F(x)$  генеральной совокупности называется *теоретической функцией распределения*. Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция  $F(x)$  определяет вероятность события  $X < x$ , а эмпирическая функция  $\bar{F}_n(x)$  определяет относительную частоту этого же события.

Из закона больших чисел в форме Бернулли (теорема 9.2) следует, что при больших  $n$  относительная частота события

$X < x$ , т.е.  $\bar{F}_n(x)$  и вероятность этого же события  $F(x)$  мало отличаются одно от другого в том смысле, что

$$\lim_{n \rightarrow \infty} P(|F(x) - \bar{F}_n(x)| < \varepsilon) = 1 \text{ при любом } \varepsilon > 0. \quad (11.2)$$

С другой стороны, из определения функции  $\bar{F}_n(x)$  вытекает, что она обладает всеми свойствами  $F(x)$ :

1) значения эмпирической функции принадлежат отрезку  $[0, 1]$ ;

2)  $\bar{F}_n(x)$  — неубывающая функция;

3) если  $x_1$  — наименьшая варианта, то  $\bar{F}_n(x) = 0$  при  $x \leq x_1$ ; если  $x_k$  — наибольшая варианта, то  $\bar{F}_n(x) = 1$  при  $x > x_k$ .

Отсюда следует целесообразность использования эмпирической функции распределения выборки для приближенного представления теоретической функции распределения генеральной совокупности. Другими словами, эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

**Пример 2.** Построить эмпирическую функцию по данному распределению выборки:

Т а б л и ц а 11.3

|       |   |   |    |
|-------|---|---|----|
| $x_i$ | 1 | 4 | 8  |
| $n_i$ | 9 | 3 | 18 |

Решение. Найдем объем выборки:  $9 + 3 + 18 = 30$ .  
Наименьшая варианта равна 1, следовательно,

$$\bar{F}_n(x) = 0 \text{ при } x \leq 1.$$

Значение  $X < 4$ , а именно  $x_1 = 1$ , наблюдалось 9 раз, следовательно,

$$\bar{F}_n(x) = 9/30 = 0,3 \text{ при } 1 < x \leq 4.$$

Значения  $X < 8$ , а именно  $x_1 = 1$  и  $x_2 = 4$ , наблюдались  $9 + 3 = 12$  раз, следовательно,

$$\bar{F}_n(x) = 12/30 = 0,4 \text{ при } 4 < x \leq 8.$$

Так как наибольшая варианта равна 8, то

$$\bar{F}_n(x) = 1 \text{ при } x > 8.$$

Искомая эмпирическая функция имеет вид

$$\bar{F}_n(x) = \begin{cases} 0 & \text{при } x \leq 1 \\ 0,3 & \text{при } 1 < x \leq 4 \\ 0,4 & \text{при } 4 < x \leq 8 \\ 1 & \text{при } x > 8 \end{cases}$$

График этой функции изображен на рис. 11.1.

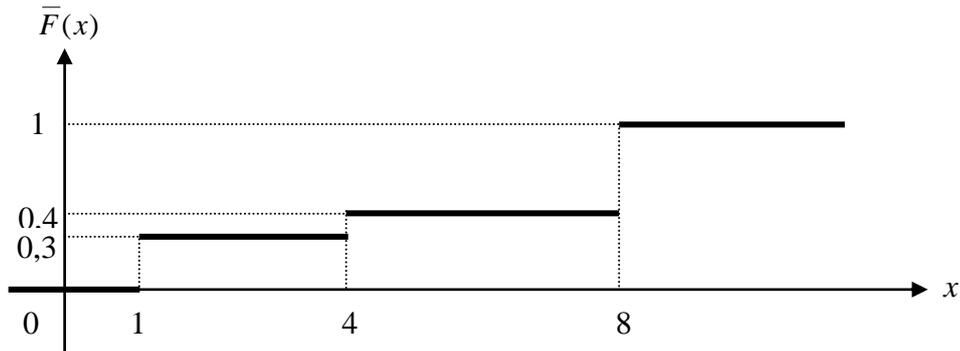


Рис. 11.1.

Статистическое распределение графически можно изобразить различными способами, в частности, в виде полигона и гистограммы.

*Полигоном частот* называется ломаная, отрезки которой соединяют точки  $(x_1; n_1)$ ,  $(x_2; n_2)$ , ...,  $(x_k; n_k)$ . Для построения полигона частот на оси абсцисс откладывают варианты  $x_i$ , а на оси ординат — соответствующие им частоты  $n_i$ . Точки  $(x_i; n_i)$  соединяют отрезками прямых и получают полигон частот.

*Полигоном относительных частот* называется ломаная, отрезки которой соединяют точки  $(x_1; W_1)$ ,  $(x_2; W_2)$ , ...,  $(x_k; W_k)$ . Полигон относительных частот строится аналогичным полигону частот образом. На рис. 11.2 изображен полигон относительных частот следующего распределения:

Т а б л и ц а 11.4

|       |     |     |      |      |
|-------|-----|-----|------|------|
| $x_i$ | 2   | 4   | 6    | 8    |
| $W_i$ | 0,1 | 0,5 | 0,25 | 0,15 |

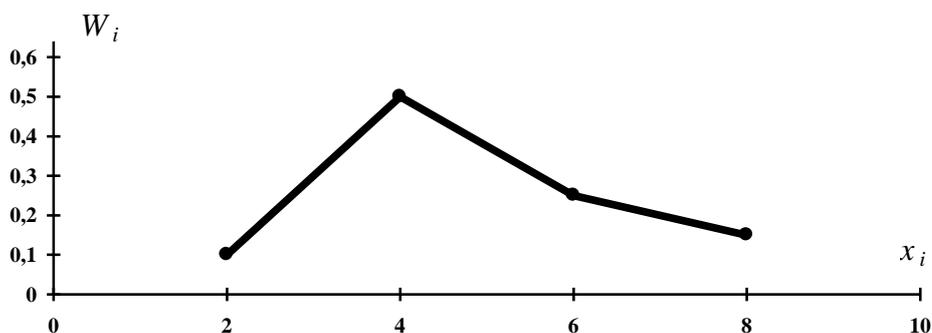


Рис. 11.2.

В случае непрерывного признака целесообразно строить гистограмму, для чего интервал, в котором заключены все наблюдаемые значения признака, разбивается на несколько частичных интервалов длиной  $h$  и для каждого частичного интервала находится  $n_i$  — сумма частот вариантов, попавших в  $i$ -й интервал.

*Гистограммой частот* называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной  $h$ , а высоты равны отношению  $n_i/h$ . Для построения гистограммы частот на оси абсцисс следует отложить частичные интервалы, а над ними провести отрезки, параллельные оси абсцисс на расстоянии  $n_i/h$ .

Площадь  $i$ -го частичного прямоугольника равна  $h n_i/h = n_i$  — сумме частот вариантов  $i$ -го интервала; следовательно, *площадь гистограммы частот равна сумме всех частот, т.е. объему выборки.*

Т а б л и ц а 11.5

| Частичный интервал<br>длиною $h = 5$ | Сумма частот<br>вариант частичного<br>интервала $n_i$ | Плотность частоты<br>$n_i/h$ |
|--------------------------------------|---|------------------------------|
| 5 — 10                               | 4   | 0,8                          |
| 10 — 15                              | 6   | 1,2                          |
| 15 — 20                              | 16  | 3,2                          |
| 20 — 25                              | 36  | 7,2                          |
| 25 — 30                              | 24  | 4,8                          |
| 30 — 35                              | 10  | 2,0                          |
| 35 — 40                              | 4   | 0,8                          |

На рис. 11.3 изображена гистограмма частот распределения,

заданного в табл. 11.5.

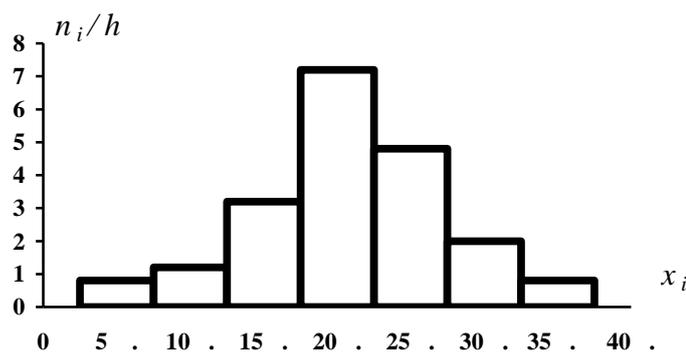


Рис. 11.3.

*Гистограммой относительных частот* называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной  $h$ , а высоты равны отношению  $W_i/h$ . Гистограмма относительных частот строится аналогичным гистограмме частот образом.

Площадь  $i$ -го частичного прямоугольника равна  $h W_i/h = W_i$  — сумме относительных частот вариант  $i$ -го интервала; следовательно, *площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.*

### **Вопросы для повторения и контроля:**

1. Что называется вариантами, вариационным рядом, частотами и относительными частотами?
2. Что такое статистическое распределение выборки и как оно задается, какова разница между распределением в теории вероятностей и распределением в математической статистике?
3. Что такое эмпирическая функция распределения и теоретическая функция распределения?
4. Какими свойствами обладает эмпирическая функция распределения?
5. В чем целесообразность использования эмпирической функции распределения выборки для оценки теоретической функции распределения генеральной совокупности?
6. Что называется полигоном частот и полигоном относительных частот, как они строятся?
7. Что такое гистограмма частот, как она строится и чему равна

площадь гистограммы частот?

8. Что такое гистограмма относительных частот, как она строится и чему равна площадь гистограммы относительных частот?

### **Опорные слова:**

Варианта, вариационный ряд, частота, относительная частота, статистическое распределение выборки, эмпирическая функция распределения, теоретическая функция распределения, полигон частот, полигон относительных частот, гистограмма частот, площадь гистограммы частот, гистограмма относительных частот, площадь гистограммы относительных частот.

## Тема № 12

### Статистическая оценка. Требования, предъявляемые к статистической оценке. Выборочное среднее и выборочная дисперсия

#### План:

1. Статистические оценки параметров распределения.
2. Несмещенные, эффективные и состоятельные оценки.
3. Генеральная и выборочная средние.
4. Генеральная и выборочная дисперсии.

Теория статистического оценивания с точки зрения постановки задачи подразделяется на параметрические и непараметрические случаи.

Если требуется изучить количественный признак генеральной совокупности, то возникает задача оценки параметров, которыми определяется распределение этого признака. Например, если наперед известно, что изучаемый признак распределен в генеральной совокупности нормально, то необходимо оценить (приблизительно найти) математическое ожидание и среднее квадратическое отклонение, так как эти два параметра полностью определяют нормальное распределение.

Обычно имеются лишь данные выборки, например, значения количественного признака  $x_1, x_2, \dots, x_n$ , полученные в результате  $n$  наблюдений, причем эти наблюдения предполагаются независимыми. Через эти данные и выражается оцениваемый параметр. Рассматривая  $x_1, x_2, \dots, x_n$  как независимые случайные величины  $X_1, X_2, \dots, X_n$ , можно сказать, что нахождение статистической оценки неизвестного параметра теоретического распределения равносильно нахождению функции от наблюдаемых случайных величин, которая и дает приближенное значение оцениваемого параметра. Например, для оценки математического ожидания нормального распределения служит функция  $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ , которая является средним арифметическим наблюдаемых значений признака.

Таким образом, *статистической оценкой неизвестного параметра*  $\theta$  теоретического распределения называется функция  $\bar{\theta} = \bar{\theta}(n) = \bar{\theta}(X_1, X_2, \dots, X_n)$  от наблюдаемых случайных величин, которая в определенном статистическом смысле близка к истинному значению этого параметра.

Важнейшими свойствами статистической оценки, определяющими ее близость к истинному значению оцениваемого параметра, являются свойства несмещенности, состоятельности и эффективности.

Пусть  $\bar{\theta}$  — статистическая оценка неизвестного параметра  $\theta$  теоретического распределения. Многократно извлекая из генеральной совокупности выборки объема  $n$ , можно получить оценки  $\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_k$ , которые, вообще говоря, различны между собой. Таким образом, оценку  $\bar{\theta}$  можно рассматривать как случайную величину, а числа  $\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_k$  — как ее возможные значения.

Если оценка  $\bar{\theta}$  дает приближенное значение  $\theta$  с избытком, то каждое найденное по данным выборок число  $\bar{\theta}_i$  ( $i = 1, 2, \dots, k$ ) больше истинного значения  $\theta$ . Ясно, что в этом случае и математическое ожидание (среднее значение) случайной величины  $\bar{\theta}$  больше, чем  $\theta$ , т.е.  $M(\bar{\theta}) > \theta$ . Очевидно, что если  $\bar{\theta}$  дает оценку с недостатком, то  $M(\bar{\theta}) < \theta$ .

Отсюда видно, что использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, приведет к систематическим ошибкам, которые являются неслучайными ошибками, искажающими результаты измерений в одну определенную сторону. По этой причине равенство математического ожидания оценки  $\bar{\theta}$  оцениваемому параметру хотя и не устраняет ошибок ввиду того, что одни значения  $\bar{\theta}$  больше, а другие меньше  $\theta$ , однако гарантирует от получения систематических ошибок, так как ошибки разных знаков будут встречаться одинаково часто.

Статистическая оценка  $\bar{\theta}$  называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру  $\theta$  при любом объеме выборки, т.е.

$$M(\bar{\theta}) = \theta. \quad (12.1)$$

*Смещенной* называется оценка, математическое ожидание которой не равно оцениваемому параметру.

Однако несмещенная оценка необязательно каждый раз дает хорошее приближение оцениваемого параметра. Действительно, возможные значения  $\bar{\theta}$  могут быть сильно рассеяны вокруг своего среднего значения, т.е. дисперсия  $D(\bar{\theta})$  может быть значительной. В этом случае найденная по данным одной выборки оценка может оказаться весьма удаленной от среднего значения  $\bar{\theta}$ , а значит, и от самого оцениваемого параметра  $\theta$ . Если же потребовать, чтобы дисперсия  $D(\bar{\theta})$  была малой, то возможность допустить большую ошибку будет исключена.

Статистическая оценка называется *эффективной*, если при заданном объеме выборки  $n$  она имеет наименьшую возможную дисперсию.

Статистическая оценка  $\bar{\theta}$  называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру  $\theta$ , т.е. для любого  $\varepsilon > 0$

$$P(|\bar{\theta}(n) - \theta| \leq \varepsilon) \rightarrow 1 \text{ при } n \rightarrow \infty. \quad (12.2)$$

Например, если дисперсия несмещенной оценки при  $n \rightarrow \infty$  стремится к нулю, то такая оценка оказывается и состоятельной.

Пусть генеральная совокупность изучается относительно количественного признака  $X$ .

*Генеральной средней*  $\bar{x}_G$  называется среднее арифметическое значений признака генеральной совокупности.

Если все значения  $x_1, x_2, \dots, x_N$  признака генеральной совокупности объема  $N$  различны, то генеральная средняя равна

$$\bar{x}_G = (x_1 + x_2 + \dots + x_N) / N. \quad (12.3)$$

Если же значения признака  $x_1, x_2, \dots, x_k$  имеют соответственно частоты  $N_1, N_2, \dots, N_k$ , причем  $N_1 + N_2 + \dots + N_k = N$ , то в этом случае генеральная средняя равна

$$\bar{x}_G = (x_1 N_1 + x_2 N_2 + \dots + x_k N_k) / N. \quad (12.4)$$

Если рассматривать обследуемый признак  $X$  генеральной совокупности как случайную величину и сопоставлять формулы (12.3) и (12.4) с формулами (6.1) и (6.2), то можно сделать вывод,

что математическое ожидание признака равно генеральной средней этого признака:

$$\bar{x}_G = M(X). \quad (12.5)$$

Пусть теперь для изучения генеральной совокупности относительно количественного признака  $X$  извлечена выборка объема  $n$ .

*Выборочной средней*  $\bar{x}_B$  называется среднее арифметическое наблюдаемых значений признака выборочной совокупности.

Если все значения  $x_1, x_2, \dots, x_n$  признака выборки объема  $n$  различны, то выборочная средняя равна

$$\bar{x}_B = (x_1 + x_2 + \dots + x_n)/n. \quad (12.6)$$

Если же значения признака  $x_1, x_2, \dots, x_k$  имеют соответственно частоты  $n_1, n_2, \dots, n_k$ , причем  $n_1 + n_2 + \dots + n_k = n$ , то в этом случае выборочная средняя равна

$$\bar{x}_B = (x_1 n_1 + x_2 n_2 + \dots + x_k n_k)/n \quad (12.7)$$

или

$$\bar{x}_B = \left( \sum_{i=1}^k x_i n_i \right) / n. \quad (12.8)$$

Убедимся, что выборочная средняя является несмещенной оценкой генеральной средней, т.е. покажем, что математическое ожидание  $\bar{x}_B$  равно  $\bar{x}_G$ . Будем рассматривать  $\bar{x}_B$  как случайную величину и  $x_1, x_2, \dots, x_n$  как независимые, одинаково распределенные случайные величины. Поскольку эти величины одинаково распределены, то они имеют одинаковые числовые характеристики, в частности одинаковое математическое ожидание, которое равно математическому ожиданию признака  $X$  генеральной совокупности.

На основании этого, используя свойство 6.2, следствие 6.2, а также формулы (12.5) и (12.6), получаем

$$M(\bar{x}_B) = \bar{x}_G. \quad (12.9)$$

Используя следствие 9.1, легко показать, что выборочная средняя является и состоятельной оценкой генеральной средней.

Для того, чтобы охарактеризовать рассеяния значений количественных признаков генеральной и выборочной совокупностей вокруг своих средних значений, вводятся сводные

характеристики — соответственно генеральная и выборочная дисперсии, а также средние квадратические отклонения.

*Генеральной дисперсией*  $D_G$  называется среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения  $\bar{x}_G$ .

Если все значения  $x_1, x_2, \dots, x_N$  признака генеральной совокупности объема  $N$  различны, то генеральная дисперсия равна

$$D_G = \left( \sum_{i=1}^N (x_i - \bar{x}_G)^2 \right) / N. \quad (12.10)$$

Если же значения признака  $x_1, x_2, \dots, x_k$  имеют соответственно частоты  $N_1, N_2, \dots, N_k$ , причем  $N_1 + N_2 + \dots + N_k = N$ , то в этом случае генеральная дисперсия равна

$$D_G = \left( \sum_{i=1}^k (x_i - \bar{x}_G)^2 N_i \right) / N. \quad (12.11)$$

*Генеральным средним квадратическим отклонением* называется квадратный корень из генеральной дисперсии:

$$\sigma_G = \sqrt{D_G}. \quad (12.12)$$

**Пример 1.** Генеральная совокупность задана следующей таблицей распределения:

Т а б л и ц а 12.1

|       |   |   |    |   |
|-------|---|---|----|---|
| $x_i$ | 2 | 4 | 5  | 6 |
| $N_i$ | 8 | 9 | 10 | 3 |

Найти генеральную дисперсию и генеральное среднее квадратическое отклонение.

Решение. Найдем генеральную среднюю:

$$\bar{x}_G = \frac{2 \cdot 8 + 4 \cdot 9 + 5 \cdot 10 + 6 \cdot 3}{8 + 9 + 10 + 3} = \frac{120}{30} = 4.$$

Найдем генеральную дисперсию:

$$D_G = \frac{(2-4)^2 \cdot 8 + (4-4)^2 \cdot 9 + (5-4)^2 \cdot 10 + (6-4)^2 \cdot 3}{30} = \frac{54}{30} = 1,8.$$

Найдем генеральное среднее квадратическое отклонение:

$$\sigma_{\Gamma} = \sqrt{D_{\Gamma}} = \sqrt{1,8} \approx 1,34.$$

*Выборочной дисперсией*  $D_B$  называется среднее арифметическое квадратов отклонений наблюдаемых значений признака выборочной совокупности от их среднего значения  $\bar{x}_B$ .

Если все значения  $x_1, x_2, \dots, x_n$  признака выборки объема  $n$  различны, то выборочная дисперсия равна

$$D_B = \left( \sum_{i=1}^n (x_i - \bar{x}_B)^2 \right) / n. \quad (12.13)$$

Если же значения признака  $x_1, x_2, \dots, x_k$  имеют соответственно частоты  $n_1, n_2, \dots, n_k$ , причем  $n_1 + n_2 + \dots + n_k = n$ , то в этом случае выборочная дисперсия равна

$$D_B = \left( \sum_{i=1}^k (x_i - \bar{x}_B)^2 n_i \right) / n. \quad (12.14)$$

*Выборочным средним квадратическим отклонением* называется квадратный корень из генеральной дисперсии:

$$\sigma_B = \sqrt{D_B}. \quad (12.15)$$

**Пример 2.** Выборочная совокупность задана следующей таблицей распределения:

Т а б л и ц а 12.2

|       |    |    |    |   |
|-------|----|----|----|---|
| $x_i$ | 1  | 2  | 3  | 4 |
| $n_i$ | 20 | 15 | 10 | 5 |

Найти выборочную дисперсию и выборочное среднее квадратическое отклонение.

Решение. Найдем выборочную среднюю:

$$\bar{x}_B = \frac{1 \cdot 20 + 2 \cdot 15 + 3 \cdot 10 + 4 \cdot 5}{20 + 15 + 10 + 5} = \frac{100}{50} = 2.$$

Найдем выборочную дисперсию:

$$D_B = \frac{(1-2)^2 \cdot 20 + (2-2)^2 \cdot 15 + (3-2)^2 \cdot 10 + (4-2)^2 \cdot 5}{50} = \frac{50}{50} = 1.$$

Найдем выборочное среднее квадратическое отклонение:

$$\sigma_B = \sqrt{D_B} = \sqrt{1} = 1.$$

Дисперсии удобнее вычислять, используя следующие формулы:

$$D_{\Gamma} = \left( \sum_{i=1}^N x_i^2 \right) / N - (\bar{x}_{\Gamma})^2, \quad (12.16)$$

$$D_{\Gamma} = \left( \sum_{i=1}^k x_i^2 N_i \right) / N - (\bar{x}_{\Gamma})^2, \quad (12.17)$$

$$D_B = \left( \sum_{i=1}^n x_i^2 \right) / n - (\bar{x}_B)^2 \quad (12.18)$$

и

$$D_B = \left( \sum_{i=1}^k x_i^2 n_i \right) / n - (\bar{x}_B)^2. \quad (12.19)$$

Теперь пусть требуется по данным выборки оценить неизвестную генеральную дисперсию  $D_{\Gamma}$ . Выборочная дисперсия  $D_B$  является смещенной оценкой  $D_{\Gamma}$ , так как

$$M(D_B) = \frac{n-1}{n} D_{\Gamma}. \quad (12.20)$$

Если же в качестве оценки генеральной дисперсии принять *исправленную дисперсию*  $s^2$ , которая получается путем умножения  $D_B$  на дробь  $n/(n-1)$ , то она будет несмещенной оценкой генеральной дисперсии. Действительно, учитывая (12.20), имеем

$$s^2 = \frac{n}{n-1} D_B = \frac{n}{n-1} \frac{\left( \sum_{i=1}^k (x_i - \bar{x}_B)^2 n_i \right)}{n} = \left( \sum_{i=1}^k (x_i - \bar{x}_B)^2 n_i \right) / (n-1) \quad (12.21)$$

и

$$M(s^2) = M\left( \frac{n}{n-1} D_B \right) = \frac{n}{n-1} M(D_B) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\Gamma} = D_{\Gamma}. \quad (12.22)$$

### Вопросы для повторения и контроля:

1. Что называется статистической оценкой неизвестного параметра и какими важнейшими свойствами она может обладать?
2. Что такое несмещенная оценка и чем обосновывается ее

введение?

3. Что такое эффективная оценка и в чем необходимость ее ввода?
4. Что называется смещенной оценкой и состоятельной оценкой?
5. Что такое генеральная средняя и по каким формулам она вычисляется?
6. Что называется выборочной средней и по каким формулам она вычисляется?
7. Какой оценкой генеральной средней является выборочная средняя?
8. Что такое генеральная дисперсия и по каким формулам она вычисляется?
9. Что называется выборочной дисперсией и по каким формулам она вычисляется?
10. Что такое генеральное среднее квадратическое отклонение и выборочное среднее квадратическое отклонение, для чего они, а также генеральная и выборочная дисперсии вводятся?
11. По каким формулам удобнее вычислять дисперсии?
12. Что является несмещенной оценкой генеральной дисперсии?

### **Опорные слова:**

Статистическая оценка неизвестного параметра, несмещенная оценка, смещенная оценка, эффективная оценка, состоятельная оценка, генеральная средняя, выборочная средняя, генеральная дисперсия, генеральное среднее квадратическое отклонение, выборочная дисперсия, выборочное среднее квадратическое отклонение, исправленная дисперсия.

## Тема № 13

### Интервальные оценки. Доверительный интервал. Доверительные интервалы для неизвестных параметров нормального распределения

#### План:

1. Точность оценки, надежность, доверительный интервал.
2. Доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратическом отклонении.
3. Доверительный интервал для оценки математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении.
4. Доверительный интервал для оценки среднего квадратического отклонения нормального распределения.

Имеется два способа оценки параметров: точечный и интервальный. Точечные методы указывают лишь точку, около которой находится неизвестный оцениваемый параметр. С помощью интервальных способов можно найти интервал, в котором с некоторой вероятностью находится неизвестное значение параметра.

*Точечной* называется оценка, которая определяется одним числом. При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам. По этой причине при небольшом объеме выборки следует пользоваться интервальными оценками.

*Интервальной* называется оценка, которая определяется двумя числами — концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика  $\bar{\theta}$  служит оценкой неизвестного параметра  $\theta$ . Если  $\delta > 0$  и  $|\theta - \bar{\theta}| < \delta$ , то оценка  $\bar{\theta}$  тем точнее определяет параметр  $\theta$ , чем меньше  $\delta$ . *Точность оценки* характеризуется положительным числом  $\delta$ .

Однако нельзя категорически утверждать, что оценка  $\bar{\theta}$

удовлетворяет неравенству  $|\theta - \bar{\theta}| < \delta$ . Статистические методы позволяют лишь говорить о вероятности, с которой это неравенство осуществляется.

*Надежностью (доверительной вероятностью)* оценки  $\theta$  по  $\bar{\theta}$  называется вероятность  $\gamma$ , с которой осуществляется неравенство  $|\theta - \bar{\theta}| < \delta$ , т.е.

$$P(|\theta - \bar{\theta}| < \delta) = \gamma. \quad (13.1)$$

В качестве  $\gamma$  берется число, близкое к единице.

Из неравенства  $|\theta - \bar{\theta}| < \delta$  легко можно получить двойное неравенство

$$\bar{\theta} - \delta < \theta < \bar{\theta} + \delta. \quad (13.2)$$

Тогда соотношение (13.1) принимает следующий вид

$$P(\bar{\theta} - \delta < \theta < \bar{\theta} + \delta) = \gamma. \quad (13.3)$$

Это соотношение означает следующее: вероятность того, что интервал  $(\bar{\theta} - \delta, \bar{\theta} + \delta)$  включает в себе (покрывает) неизвестный параметр  $\theta$ , равна  $\gamma$ .

Интервал  $(\bar{\theta} - \delta, \bar{\theta} + \delta)$  называется *доверительным интервалом*, который покрывает неизвестный параметр  $\theta$  с заданной надежностью  $\gamma$ .

Пусть количественный признак  $X$  генеральной совокупности распределен нормально, причем среднее квадратическое отклонение  $\sigma$  этого распределения известно. Требуется оценить неизвестное математическое ожидание  $a$  по выборочной средней  $\bar{x}$ . Поставим своей задачей найти доверительные интервалы, покрывающие параметр  $a$  с надежностью  $\gamma$ .

Будем рассматривать выборочную среднюю  $\bar{x}$  как случайную величину  $\bar{X}$  ( $\bar{x}$  изменяется от выборки к выборке) и выборочные значения признака  $x_1, x_2, \dots, x_n$  — как одинаково распределенные случайные величины  $X_1, X_2, \dots, X_n$  (эти числа также изменяются от выборки к выборке). Математическое ожидание каждой из этих величин равно  $a$  и среднее квадратическое отклонение —  $\sigma$ .

Тогда, используя свойство 6.2, следствие 6.2, а также

формулу (12.6), получаем, что параметры распределения  $\bar{X}$  следующие:

$$M(\bar{X}) = a, \quad \sigma(\bar{X}) = \sigma/\sqrt{n}. \quad (13.4)$$

Потребуем, чтобы выполнялось соотношение

$$P(|\bar{X} - a| < \delta) = \gamma, \quad (13.5)$$

где  $\gamma$  — заданная надежность.

Используя формулу (8.11) с заменой  $X$  на  $\bar{X}$  и  $\sigma$  на  $\sigma(\bar{X}) = \sigma/\sqrt{n}$ , нетрудно получить соотношение

$$P(|\bar{X} - a| < \delta) = 2\Phi(\delta\sqrt{n}/\sigma) = 2\Phi(t), \quad (13.6)$$

где  $t = \delta\sqrt{n}/\sigma$ .

Найдя из последнего равенства  $\delta = t\sigma/\sqrt{n}$ , можно написать

$$P(|\bar{X} - a| < t\sigma/\sqrt{n}) = 2\Phi(t). \quad (13.7)$$

Обозначая для общности выборочную среднюю вновь через  $\bar{x}$ , из соотношений (13.5) – (13.7) получаем соотношения

$$\Phi(t) = \gamma/2 \quad (13.8)$$

и

$$P(\bar{x} - t\sigma/\sqrt{n} < a < \bar{x} + t\sigma/\sqrt{n}) = \gamma. \quad (13.9)$$

Значит, с надежностью  $\gamma$  можно утверждать, что доверительный интервал  $(\bar{x} - t\sigma/\sqrt{n}, \bar{x} + t\sigma/\sqrt{n})$  покрывает неизвестный параметр  $a$ , при этом точность оценки равна  $\delta = t\sigma/\sqrt{n}$ , а число  $t$  определяется из равенства (13.8) по таблице функции Лапласа.

**Пример 1.** Случайная величина  $X$  имеет нормальное распределение с известным средним квадратическим отклонением  $\sigma = 3$ . Найти доверительный интервал для оценки неизвестного математического ожидания  $a$  по выборочной средней  $\bar{x}$ , если объем выборки  $n = 36$  и задана надежность оценки  $\gamma = 0,95$ .

Решение. Найдём  $t$ . Из соотношения (13.8) получаем  $\Phi(t) = 0,475$  и по таблице функции Лапласа находим  $t = 1,96$ .

Найдём точность оценки:

$$\delta = t\sigma/\sqrt{n} = (1,96 \cdot 3)/\sqrt{36} = 0,98.$$

Доверительный интервал таков:  $(\bar{x} - 0,98; \bar{x} + 0,98)$ . Например,

если  $\bar{x} = 4,1$ , то доверительный интервал имеет следующие доверительные границы:

$$\bar{x} - 0,98 = 4,1 - 0,98 = 3,12; \quad \bar{x} + 0,98 = 4,1 + 0,98 = 5,08.$$

Далее нам потребуются распределения «хи квадрат» и Стьюдента.

Пусть  $X_i$  ( $i = 1, 2, \dots, n$ ) — нормальные независимые случайные величины, причем математическое ожидание каждой из них равно нулю, а среднее квадратическое отклонение — единице. Тогда сумма квадратов этих величин  $\chi^2 = \sum_{i=1}^n X_i^2$  распределена по закону  $\chi^2$  («хи квадрат») с  $k = n$  степенями свободы.

Функция плотности этого распределения имеет вид

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ \frac{1}{2^{k/2} \Gamma(k/2)} e^{-x/2} x^{(k/2)-1} & \text{при } x > 0 \end{cases}, \quad (13.10)$$

где  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$  — гамма-функция.

Отсюда видно, что распределение «хи квадрат» определяется одним параметром — числом степеней свободы  $k$ .

Далее, пусть  $Z$  — нормальная случайная величина, причем  $M(Z) = 0$ ,  $\sigma(Z) = 1$ , а  $V$  — независимая от  $Z$  случайная величина, которая распределена по закону  $\chi^2$  с  $k$  степенями свободы. Тогда случайная величина

$$T = \frac{Z}{\sqrt{V/k}} \quad (13.11)$$

имеет распределение, которое называется  $t$ -распределением или *распределением Стьюдента* с  $k$  степенями свободы.

Пусть теперь требуется оценить неизвестное математическое ожидание  $a$  количественного признака  $X$  генеральной совокупности, который распределен нормально, по выборочной средней  $\bar{x}$ , когда среднее квадратическое отклонение  $\sigma$  этого распределения **н е и з в е с т н о**. Поставим

своей задачей найти доверительные интервалы, покрывающие параметр  $a$  с надежностью  $\gamma$ .

Рассмотрим случайную величину

$$T = \frac{\bar{X} - a}{s/\sqrt{n}}, \quad (13.12)$$

которая имеет распределение Стьюдента с  $k = n - 1$  степенями свободы. Здесь  $\bar{X}$  — выборочная средняя,  $s$  — «исправленное» среднее квадратическое отклонение,  $n$  — объем выборки.

Функция плотности распределения этой случайной величины равна

$$S(t, n) = B_n \left[ 1 + \frac{t^2}{n-1} \right]^{-n/2}, \quad (13.13)$$

где  $B_n = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)}$ . Отсюда видно, что

распределение случайной величины (13.12) определяется параметром  $n$  — объемом выборки и не зависит от неизвестных параметров  $a$  и  $\sigma$ .

Поскольку  $S(t, n)$  — четная функция от  $t$ , то вероятность осуществления неравенства

$$\left| \frac{\bar{X} - a}{s/\sqrt{n}} \right| < t_\gamma \quad (13.14)$$

определяется на основании теоремы 7.1 из следующей формулы

$$P\left(\left|\frac{\bar{X} - a}{s/\sqrt{n}}\right| < t_\gamma\right) = 2 \int_0^{t_\gamma} S(t, n) dt = \gamma. \quad (13.15)$$

Заменив неравенство (13.14) равносильным ему двойным неравенством, получаем соотношение

$$P(\bar{x} - t_\gamma s/\sqrt{n} < a < \bar{x} + t_\gamma s/\sqrt{n}) = \gamma. \quad (13.16)$$

Итак, пользуясь распределением Стьюдента, мы нашли доверительный интервал  $(\bar{x} - t_\gamma s/\sqrt{n}, \bar{x} + t_\gamma s/\sqrt{n})$ , покрывающий неизвестный параметр  $a$  с надежностью  $\gamma$ . Из специальной таблицы по заданным  $n$  и  $\gamma$  можно найти  $t_\gamma$ .

**Пример 2.** Количественный признак  $X$  генеральной совокупности распределен нормально. По выборке объема  $n = 16$

найденны выборочная средняя  $\bar{x} = 20,2$  и «исправленное» среднее квадратическое отклонение  $s = 0,8$ . Оценить неизвестное математическое ожидание  $a$  при помощи доверительного интервала с надежностью  $\gamma = 0,95$ .

Решение. Найдем  $t_\gamma$ . Пользуясь таблицей, по  $\gamma = 0,95$  и  $n = 16$  находим  $t_\gamma = 2,13$ .

Найдем доверительные границы:

$$\bar{x} - t_\gamma s / \sqrt{n} = 20,2 - 2,13 \cdot 0,8 / \sqrt{16} = 19,774,$$

$$\bar{x} + t_\gamma s / \sqrt{n} = 20,2 + 2,13 \cdot 0,8 / \sqrt{16} = 20,626.$$

Итак, с надежностью 0,95 неизвестный параметр  $a$  заключен в доверительном интервале  $19,774 < a < 20,626$ .

Пусть количественный признак  $X$  генеральной совокупности распределен нормально. Требуется оценить неизвестное генеральное среднее квадратическое отклонение  $\sigma$  по «исправленному» среднему квадратическому отклонению  $s$ . Поставим перед собой задачу найти доверительные интервалы, покрывающие параметр  $\sigma$  с заданной надежностью  $\gamma$ .

Потребуем, чтобы выполнялось соотношение

$$P(|\sigma - s| < \delta) = \gamma \quad (13.17)$$

или равносильное ему соотношение

$$P(s - \delta < \sigma < s + \delta) = \gamma, \quad (13.18)$$

где  $\gamma$  — заданная надежность.

Положив  $\delta/s = q$ , из двойного неравенства

$$s - \delta < \sigma < s + \delta \quad (13.19)$$

получаем неравенство

$$s(1 - q) < \sigma < s(1 + q). \quad (13.20)$$

Для нахождения доверительного интервала, покрывающего параметр  $\sigma$ , остается только найти  $q$ . С этой целью рассмотрим случайную величину

$$\chi = (s/\sigma)\sqrt{n-1}, \quad (13.21)$$

где  $n$  — объем выборки (эта случайная величина обозначена через  $\chi$  ввиду того, что случайная величина  $s^2(n-1)/\sigma^2$  распределена по закону  $\chi^2$  с  $n-1$  степенями свободы).

Функция плотности распределения случайной величины  $\chi$

имеет следующий вид

$$R(\chi, n) = \frac{\chi^{n-2} e^{-\chi^2/2}}{2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right)}. \quad (13.22)$$

Это распределение не зависит от оцениваемого параметра  $\sigma$ , а зависит лишь от объема выборки  $n$ .

Из неравенства (13.20) можно получить неравенство

$$\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)}. \quad (13.23)$$

Умножив все члены этого неравенства на  $s\sqrt{n-1}$ , получаем

$$\frac{\sqrt{n-1}}{1+q} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$$

или

$$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}. \quad (13.24)$$

Воспользовавшись теоремой 7.1, находим, что вероятность осуществления этого неравенства и, следовательно, равносильного ему неравенства (13.20), равна

$$\int_{\sqrt{n-1}/(1+q)}^{\sqrt{n-1}/(1-q)} R(\chi, n) d\chi = \gamma. \quad (13.25)$$

Из этого уравнения можно по заданным  $n$  и  $\gamma$  найти  $q$ . Однако на практике  $q$  находится из специальной таблицы.

Вычислив по выборке  $s$  и найдя по таблице  $q$ , получим искомый доверительный интервал  $(s(1-q), s(1+q))$ , покрывающий неизвестный параметр  $\sigma$  с заданной надежностью  $\gamma$ .

**Пример 3.** Количественный признак  $X$  генеральной совокупности распределен нормально. По выборке объема  $n = 25$  найдено «исправленное» среднее квадратическое отклонение  $s = 0,8$ . Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение  $\sigma$  с надежностью  $\gamma = 0,95$ .

Решение. По специальной таблице по данным  $\gamma = 0,95$  и  $n = 25$  найдем  $q = 0,32$ .

Найдем искомый доверительный интервал:  
 $0,8(1 - 0,32) < \sigma < 0,8(1 + 0,32)$

или

$$0,544 < \sigma < 1,056.$$

### **Вопросы для повторения и контроля:**

1. Какие способы оценки параметров и связанные с ними оценки вы знаете?
2. Что такое точность оценки и надежность (доверительная вероятность)?
3. Что называется доверительным интервалом?
4. Как находится доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратическом отклонении?
5. Что вы знаете о распределениях «хи квадрат» и Стьюдента?
6. Как находится доверительный интервал для оценки математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении?
7. Как находится доверительный интервал для оценки среднего квадратического отклонения нормального распределения?

### **Опорные слова:**

Точечная оценка, интервальная оценка, точность оценки, надежность (доверительная вероятность), доверительный интервал, доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратическом отклонении, распределение «хи квадрат», распределение Стьюдента, доверительный интервал для оценки математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении, доверительный интервал для оценки среднего квадратического отклонения нормального распределения.

## Тема № 14

### Элементы корреляционного и регрессионного анализа

#### План:

1. Виды зависимостей между случайными величинами.
2. Условные средние и выборочные уравнения регрессии.
3. Нахождение выборочного уравнения регрессии по несгруппированным данным.
4. Корреляционная таблица.
5. Нахождение выборочного уравнения регрессии по сгруппированным данным.

Корреляционный анализ и регрессионный анализ являются смежными разделами математической статистики и предназначены для изучения по выборочным данным статистической зависимости случайных величин. Две случайные величины могут быть связаны либо функциональной зависимостью, либо статистической зависимостью, либо быть независимыми.

Если каждому возможному значению случайной величины  $X$  соответствует одно возможное значение случайной величины  $Y$ , то  $Y$  называется *функцией случайного аргумента  $X$* :

$$Y = \varphi(X),$$

а зависимость между случайными величинами  $X$  и  $Y$  называется *функциональной зависимостью*.

Строгая функциональная зависимость реализуется редко, так как обе величины или одна из них подвержены еще действию случайных факторов, причем среди них могут быть и общие для обеих величин, т.е. такие факторы, которые воздействуют как на  $X$ , так и на  $Y$ . В этом случае возникает статистическая зависимость. *Статистической* называется зависимость, при которой изменение одной из величин влечет изменение распределения другой. Частным случаем статистической зависимости является корреляционная зависимость.

Если статистическая зависимость проявляется в том, что при изменении одной из рассматриваемых случайных величин изменяется среднее значение другой случайной величины, то

такая статистическая зависимость называется *корреляционной*.

Приведем пример случайной величины  $Y$ , которая не связана с величиной  $X$  функционально, а связана корреляционно. Пусть  $Y$  — урожай зерна,  $X$  — количество удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т.е.  $Y$  не является функцией от  $X$ . Это объясняется влиянием случайных факторов, таких, как осадки, температура воздуха и др. С другой стороны, средний урожай является функцией от количества удобрений, т.е.  $Y$  связан с  $X$  корреляционной зависимостью.

*Условным средним*  $\bar{y}_x$  называется среднее арифметическое наблюдавшихся значений  $Y$ , соответствующих  $X = x$ . Например, если при  $x_1 = 2$  величина  $Y$  приняла значения  $y_1 = 5$ ,  $y_2 = 6$ ,  $y_3 = 10$ , то условное среднее равно  $\bar{y}_{x_1} = (5 + 6 + 10)/3 = 7$ .

*Условным средним*  $\bar{x}_y$  называется среднее арифметическое наблюдавшихся значений  $X$ , соответствующих  $Y = y$ .

Как видно из определения, условное среднее  $\bar{y}_x$  является функцией от  $x$ ; обозначив эту функцию через  $\bar{f}(x)$ , получим уравнение

$$\bar{y}_x = \bar{f}(x). \quad (14.1)$$

Это уравнение называется *выборочным уравнением регрессии  $Y$  на  $X$* ; функция  $\bar{f}(x)$  называется *выборочной регрессией  $Y$  на  $X$* , а ее график — *выборочной линией регрессии  $Y$  на  $X$* .

Аналогично уравнение

$$\bar{x}_y = \bar{\varphi}(y) \quad (14.2)$$

называется *выборочным уравнением регрессии  $X$  на  $Y$* ; функция  $\bar{\varphi}(y)$  называется *выборочной регрессией  $X$  на  $Y$* , а ее график — *выборочной линией регрессии  $X$  на  $Y$* .

В связи с вышеизложенным возникают две задачи теории корреляции. Первая — нахождение по данным наблюдений параметров функций  $\bar{f}(x)$  и  $\bar{\varphi}(y)$  при условии, что известен их вид. Вторая — оценка силы (тесноты) связи между случайными величинами  $X$  и  $Y$  и установление наличия корреляционной зависимости между этими величинами.

Пусть изучается система количественных признаков  $(X, Y)$ .

В результате  $n$  независимых опытов получены  $n$  пар чисел  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

Найдем по данным наблюдений выборочное уравнение прямой линии регрессии. Для определенности будем искать уравнение

$$\bar{y}_x = kx + b \quad (14.3)$$

регрессии  $Y$  на  $X$ .

Поскольку различные значения  $x$  признака  $X$  и соответствующие им значения  $y$  признака  $Y$  наблюдались по одному разу, то группировать данные нет необходимости. Также нет надобности использовать понятие условной средней, поэтому уравнение (14.3) можно записать следующим образом:

$$y = kx + b. \quad (14.4)$$

Угловым коэффициентом прямой линии регрессии  $Y$  на  $X$  называется *выборочным коэффициентом регрессии  $Y$  на  $X$*  и обозначается через  $\rho_{yx}$ . Следовательно, искомое выборочное уравнение (14.4) прямой линии регрессии  $Y$  на  $X$  следует искать в виде

$$y = \rho_{yx} x + b. \quad (14.5)$$

Нужно найти такие параметры  $\rho_{yx}$  и  $b$ , при которых точки  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , построенные по данным наблюдений, на плоскости  $xOy$  лежали как можно ближе к прямой (14.5).

Для осуществления этого воспользуемся *методом наименьших квадратов*. При использовании этого метода сумма квадратов отклонений  $Y_i - y_i$  ( $i = 1, 2, \dots, n$ ), где  $Y_i$  — вычисленная по уравнению (14.5) ордината, соответствующая наблюдаемому значению  $x_i$ , а  $y_i$  — наблюдаемая ордината, соответствующая  $x_i$ , должна быть минимальной. Так как каждое отклонение зависит от отыскиваемых параметров, то и сумма квадратов отклонений есть функция этих параметров:

$$F(\rho, b) = \sum_{i=1}^n (Y_i - y_i)^2 \quad (14.6)$$

или

$$F(\rho, b) = \sum_{i=1}^n (\rho x_i + b - y_i)^2. \quad (14.7)$$

Для отыскания минимума приравняем нулю соответствующие частные производные:

$$\begin{cases} \frac{\partial F}{\partial \rho} = 2 \sum_{i=1}^n (\rho x_i + b - y_i) x_i = 0 \\ \frac{\partial F}{\partial b} = 2 \sum_{i=1}^n (\rho x_i + b - y_i) = 0 \end{cases} \quad (14.8)$$

Решив эту систему двух линейных уравнений относительно  $\rho$  и  $b$ , найдем искомые параметры:

$$\rho_{yx} = \left( n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i \right) / \left( n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right); \quad (14.9)$$

$$b = \left( \sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i y_i \right) / \left( n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right). \quad (14.10)$$

Аналогично можно найти выборочное уравнение прямой линии регрессии  $X$  на  $Y$ :

$$\bar{x}_y = \rho_{xy} y + c, \quad (14.11)$$

где  $\rho_{xy}$  — выборочный коэффициент регрессии  $X$  на  $Y$ .

**Пример 1.** Найти выборочное уравнение прямой линии регрессии  $Y$  на  $X$  по данным  $n = 5$  наблюдений:

Т а б л и ц а 14.1

|       |      |      |      |      |      |
|-------|------|------|------|------|------|
| $x_i$ | 1,00 | 1,50 | 3,00 | 4,50 | 5,00 |
| $y_i$ | 1,25 | 1,40 | 1,50 | 1,75 | 2,25 |

Решение. Составим следующую расчетную табл. 14.2.

Найдем искомые параметры из соотношений (14.9) и (14.10):

$$\rho_{yx} = (5 \cdot 26,975 - 15 \cdot 8,15) / (5 \cdot 57,5 - 15^2) = 0,202;$$

$$b = (57,5 \cdot 8,15 - 15 \cdot 26,975) / (5 \cdot 57,5 - 15^2) = 1,024.$$

Напишем искомое уравнение прямой линии регрессии  $Y$  на  $X$ :

$$y = 0,202 x + 1,024.$$

Т а б л и ц а 14.2

|                         |                           |                              |                                 |
|-------------------------|---------------------------|------------------------------|---------------------------------|
| $x_i$                   | $y_i$                     | $x_i^2$                      | $x_i y_i$                       |
| 1,00                    | 1,25                      | 1,00                         | 1,250                           |
| 1,50                    | 1,40                      | 2,25                         | 2,100                           |
| 3,00                    | 1,50                      | 9,00                         | 4,500                           |
| 4,50                    | 1,75                      | 20,25                        | 7,875                           |
| 5,00                    | 2,25                      | 25,00                        | 11,250                          |
| $\sum_{i=1}^n x_i = 15$ | $\sum_{i=1}^n y_i = 8,15$ | $\sum_{i=1}^n x_i^2 = 57,50$ | $\sum_{i=1}^n x_i y_i = 26,975$ |

При большом числе наблюдений одно и то же значение  $x$  может встретиться  $n_x$  раз, одно и то же значение  $y$  —  $n_y$  раз, одна и та же пара чисел  $(x, y)$  может наблюдаться  $n_{xy}$  раз. Поэтому данные наблюдений следует группировать, для этого подсчитываются частоты  $n_x, n_y, n_{xy}$ . Все сгруппированные данные записываются в виде таблицы (например, табл. 14.3), которая называется *корреляционной*.

Т а б л и ц а 14.3

| $Y$   | $X$ |    |    |    | $n_y$    |
|-------|-----|----|----|----|----------|
|       | 10  | 20 | 30 | 40 |          |
| 0,4   | 5   | —  | 7  | 14 | 26       |
| 0,6   | —   | 2  | 6  | 4  | 12       |
| 0,8   | 3   | 19 | —  | —  | 22       |
| $n_x$ | 8   | 21 | 13 | 18 | $n = 60$ |

В первой строке корреляционной таблицы 14.3 указаны наблюдаемые значения (10; 20; 30; 40) признака  $X$ , а в первом столбце — наблюдаемые значения (0,4; 0,6; 0,8) признака  $Y$ . На пересечении строк и столбцов находятся частоты  $n_{xy}$  наблюдаемых пар значений признаков.

В последнем столбце записаны суммы частот строк, а в последней строке — суммы частот столбцов. В клетке, расположенной в нижнем правом углу таблицы, помещена сумма всех частот, т.е. общее число всех наблюдений  $n$ . Очевидно, что  $\sum n_x = \sum n_y = n$ .

Теперь определим параметры выборочного уравнения прямой линии регрессии  $Y$  на  $X$  в случае, когда получено

большое число данных (практически для удовлетворительной оценки искомых параметров должно быть хотя бы 50 наблюдений), среди них есть повторяющиеся, и они сгруппированы в виде корреляционной таблицы.

Из системы (14.8) можно получить следующую систему:

$$\begin{cases} \rho_{yx} \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ \rho_{yx} \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases} . \quad (14.12)$$

Для простоты приняв обозначения  $\sum x = \sum_{i=1}^n x_i$ ,

$\sum y = \sum_{i=1}^n y_i$ ,  $\sum x^2 = \sum_{i=1}^n x_i^2$ ,  $\sum xy = \sum_{i=1}^n x_i y_i$  и воспользовавшись

соотношениями  $\bar{x} = \sum x / n$ ,  $\bar{y} = \sum y / n$ ,  $\overline{x^2} = \sum x^2 / n$ ,

$\sum xy = \sum n_{xy} xy$  (в предположении, что пара чисел  $(x, y)$  наблюдалась  $n_{xy}$  раз), из (14.12) получаем

$$\begin{cases} n\overline{x^2} \rho_{yx} + n\bar{x}b = \sum n_{xy} xy \\ \bar{x} \rho_{yx} + b = \bar{y} \end{cases} . \quad (14.13)$$

Второе уравнение системы (14.13) преобразуем к виду  $b = \bar{y} - \bar{x} \rho_{yx}$  и подставив правую часть этого равенства в уравнение  $\bar{y}_x = \rho_{yx} x + b$ , получим следующее соотношение

$$\bar{y}_x - \bar{y} = \rho_{yx} (x - \bar{x}). \quad (14.14)$$

Учитывая соотношения (12.15) и (12.19), найдем из системы (14.13) выборочный коэффициент регрессии  $\rho_{yx}$ :

$$\rho_{yx} = \frac{\sum n_{xy} xy - n\bar{x}\bar{y}}{n(\overline{x^2} - (\bar{x})^2)} = \frac{\sum n_{xy} xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x^2}.$$

Умножим обе части этого равенства на дробь  $\tilde{\sigma}_x / \tilde{\sigma}_y$ :

$$\rho_{yx} \cdot \frac{\tilde{\sigma}_x}{\tilde{\sigma}_y} = \frac{\sum n_{xy} xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x \tilde{\sigma}_y}. \quad (14.15)$$

Обозначим правую часть равенства (14.15) через  $r_B$ :

$$r_B = \frac{\sum n_{xy} xy - n \bar{x} \bar{y}}{n \tilde{\sigma}_x \tilde{\sigma}_y}. \quad (14.16)$$

Тогда из (14.15) получаем

$$\rho_{yx} = r_B \cdot \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}. \quad (14.17)$$

Подставив правую часть этого равенства в (14.14), окончательно получим выборочное уравнение прямой линии регрессии  $Y$  на  $X$  вида

$$\bar{y}_x - \bar{y} = r_B \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x} (x - \bar{x}). \quad (14.18)$$

Аналогично можно найти выборочное уравнение прямой линии регрессии  $X$  на  $Y$ :

$$\bar{x}_y - \bar{x} = r_B \frac{\tilde{\sigma}_x}{\tilde{\sigma}_y} (y - \bar{y}). \quad (14.19)$$

### Вопросы для повторения и контроля:

1. Что изучают корреляционный и регрессионный анализ, как могут быть связаны случайные величины, что такое функция случайного аргумента и функциональная зависимость?
2. Что вы знаете о статистической зависимости и корреляционной зависимости?
3. Что такое условное среднее, выборочное уравнение регрессии, выборочная регрессия, выборочная линия регрессии, и какие две задачи теории корреляции вы знаете?
4. В каком виде ищется выборочное уравнение прямой линии регрессии по несгруппированным данным и что такое выборочный коэффициент регрессии?
5. В чем суть метода наименьших квадратов и как с его помощью находится выборочное уравнение прямой линии регрессии?
6. Что вы знаете о корреляционной таблице?
7. Как находятся параметры выборочного уравнения прямой линии регрессии по сгруппированным данным?

### Опорные слова:

Корреляционный анализ, регрессионный анализ, функция случайного аргумента, функциональная зависимость, статистическая зависимость, корреляционная зависимость, условное среднее, выборочное уравнение регрессии, выборочная регрессия, выборочная линия регрессии, две задачи теории корреляции, выборочное уравнение прямой линии регрессии по несгруппированным данным, выборочный коэффициент регрессии, метод наименьших квадратов, корреляционная таблица, выборочное уравнение прямой линии регрессии по сгруппированным данным.

## Тема № 15

### Выборочный коэффициент корреляции и его свойства

#### План:

1. Корреляционный момент и коэффициент корреляции.
2. Выборочный коэффициент корреляции.
3. Выборочное корреляционное отношение.
4. Свойства выборочного корреляционного отношения.

*Корреляционным моментом*  $\mu_{xy}$  случайных величин  $X$  и  $Y$  называется математическое ожидание произведения отклонений этих величин:

$$\mu_{xy} = M\{[X - M(X)][Y - M(Y)]\}. \quad (15.1)$$

Отсюда легко можно получить соотношение

$$\mu_{xy} = M(XY) - M(X)M(Y). \quad (15.2)$$

*Коэффициентом корреляции*  $r_{xy}$  случайных величин  $X$  и  $Y$  называется отношение корреляционного момента к произведению средних квадратических отклонений этих величин:

$$r_{xy} = \mu_{xy} / (\sigma_x \sigma_y). \quad (15.3)$$

Из соотношения (15.2) вытекает, что корреляционный момент и, следовательно, коэффициент корреляции независимых случайных величин равен нулю.

Две случайные величины  $X$  и  $Y$  называются *коррелированными*, если их коэффициент корреляции отличен от нуля;  $X$  и  $Y$  называются *некоррелированными* величинами, если их коэффициент корреляции равен нулю.

Из вышесказанного следует, что независимые случайные величины всегда являются некоррелированными, а две коррелированные случайные величины также и зависимы. Действительно, если предположить, что коррелированные случайные величины независимы, то для них должно выполняться соотношение  $\mu_{xy} = 0$ , а это противоречит тому, что для коррелированных величин всегда выполняется  $\mu_{xy} \neq 0$ .

С другой стороны, две зависимые случайные величины могут быть как коррелированными, так и некоррелированными;

некоррелированные случайные величины могут быть как зависимыми, так и независимыми.

Если случайные величины  $X$  и  $Y$  независимы, то коэффициент корреляции  $r_{xy} = 0$ ; если  $r_{xy} = \pm 1$ , то случайные величины  $X$  и  $Y$  связаны линейной функциональной зависимостью. Отсюда следует, что коэффициент корреляции измеряет силу (тесноту) линейной связи между  $X$  и  $Y$ .

Величина  $r_B$ , определяемая равенством

$$r_B = \frac{\sum n_{xy} xy - n \bar{x} \bar{y}}{n \tilde{\sigma}_x \tilde{\sigma}_y}, \quad (15.4)$$

называется *выборочным коэффициентом корреляции*. Здесь  $x$  и  $y$  — варианты (наблюдавшиеся значения) признаков  $X$  и  $Y$ ;  $n_{xy}$  — частота пары вариант  $(x, y)$ ;  $n$  — объем выборки (сумма всех частот);  $\tilde{\sigma}_x$ ,  $\tilde{\sigma}_y$  — выборочные средние квадратические отклонения;  $\bar{x}$ ,  $\bar{y}$  — выборочные средние.

Выборочный коэффициент корреляции  $r_B$  является оценкой коэффициента корреляции  $r_{xy}$  генеральной совокупности. Поэтому его можно использовать и для измерения линейной связи между величинами — количественными признаками  $X$  и  $Y$ .

**Пример 1.** Найти выборочное уравнение прямой линии регрессии  $Y$  на  $X$  по данным следующей корреляционной таблицы:

Т а б л и ц а 15.1

| $Y$   | $X$ |    |    |    |    |    | $n_y$     |
|-------|-----|----|----|----|----|----|-----------|
|       | 10  | 20 | 30 | 40 | 50 | 60 |           |
| 15    | 5   | 7  | —  | —  | —  | —  | 12        |
| 25    | —   | 20 | 23 | —  | —  | —  | 43        |
| 35    | —   | —  | 30 | 47 | 2  | —  | 79        |
| 45    | —   | —  | 10 | 11 | 20 | 6  | 47        |
| 55    | —   | —  | —  | 9  | 7  | 3  | 19        |
| $n_x$ | 5   | 27 | 63 | 67 | 29 | 9  | $n = 200$ |

Решение. Сначала вычислим выборочный коэффициент корреляции по формуле (15.4):

$$\begin{aligned}\bar{x} &= (10 \cdot 5 + 20 \cdot 27 + 30 \cdot 63 + 40 \cdot 67 + 50 \cdot 29 + 60 \cdot 9) / 200 = 35,75; \\ \bar{y} &= (15 \cdot 12 + 25 \cdot 43 + 35 \cdot 79 + 45 \cdot 47 + 55 \cdot 19) / 200 = 35,9; \\ \overline{x^2} &= (100 \cdot 5 + 400 \cdot 27 + 900 \cdot 63 + 1600 \cdot 67 + 2500 \cdot 29 + 3600 \cdot 9) / 200 = \\ &= 1400,5; \\ \overline{y^2} &= (225 \cdot 12 + 625 \cdot 43 + 1225 \cdot 79 + 2025 \cdot 47 + 3025 \cdot 19) / 200 = 1395; \\ \tilde{\sigma}_x &= \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{1400,5 - 1278,0625} \approx 11,07; \\ \tilde{\sigma}_y &= \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{1395 - 1288,81} \approx 10,30; \\ \sum n_{xy} xy &= 5 \cdot 10 \cdot 15 + 7 \cdot 20 \cdot 15 + 20 \cdot 20 \cdot 25 + 23 \cdot 30 \cdot 25 + 30 \cdot 30 \cdot 35 + \\ &+ 47 \cdot 40 \cdot 35 + 2 \cdot 50 \cdot 35 + 10 \cdot 30 \cdot 45 + 11 \cdot 40 \cdot 45 + 20 \cdot 50 \cdot 45 + \\ &+ 6 \cdot 60 \cdot 45 + 9 \cdot 40 \cdot 55 + 7 \cdot 50 \cdot 55 + 3 \cdot 60 \cdot 55 = 274350; \\ r_B &= \frac{\sum n_{xy} xy - n \bar{x} \bar{y}}{n \tilde{\sigma}_x \tilde{\sigma}_y} = \frac{274350 - 200 \cdot 35,75 \cdot 35,9}{200 \cdot 11,07 \cdot 10,3} \approx 0,775.\end{aligned}$$

Теперь подставим найденные значения в формулу (14.18) и получим выборочное уравнение прямой линии регрессии  $Y$  на  $X$ :

$$\bar{y}_x - 35,9 = 0,775 \cdot \frac{10,30}{11,07} (x - 35,75)$$

или окончательно

$$\bar{y}_x = 0,721x + 10,12.$$

Если выборка имеет достаточно большой объем и хорошо представляет генеральную совокупность (репрезентативна), то заключение о тесноте линейной зависимости между признаками, полученное по данным выборки, в известной степени может быть распространено и на генеральную совокупность. Например, для оценки коэффициента корреляции  $r_G$  нормально распределенной генеральной совокупности (при  $n \geq 50$ ) можно воспользоваться формулой

$$r_B - 3 \frac{1 - r_B^2}{\sqrt{n}} \leq r_G \leq r_B + 3 \frac{1 + r_B^2}{\sqrt{n}}.$$

Итак, для оценки тесноты линейной корреляционной связи между признаками в выборке служит выборочный коэффициент корреляции. Для оценки тесноты нелинейной корреляционной связи вводится понятие выборочного корреляционного

отношения.

Выборочным корреляционным отношением  $Y$  к  $X$  называется следующее отношение

$$\eta_{yx} = \sigma_{\bar{y}_x} / \tilde{\sigma}_y. \quad (15.5)$$

Здесь

$$\sigma_{\bar{y}_x} = \sqrt{(\sum n_x (\bar{y}_x - \bar{y})^2) / n};$$

$$\tilde{\sigma}_y = \sqrt{(\sum n_y (y - \bar{y})^2) / n},$$

где  $n$  — объем выборки (сумма всех частот);  $n_x$  — частота значения  $x$  признака  $X$ ;  $n_y$  — частота значения  $y$  признака  $Y$ ;  $\bar{y}$  — общая средняя признака  $Y$ ;  $\bar{y}_x$  — условная средняя признака  $Y$ .

Аналогично определяется выборочное корреляционное отношение  $X$  к  $Y$ :

$$\eta_{xy} = \sigma_{\bar{x}_y} / \tilde{\sigma}_x. \quad (15.6)$$

**Пример 2.** Найти  $\eta_{yx}$  по данным следующей корреляционной таблицы:

Т а б л и ц а 15.2

| $Y$         | $X$ |    |    | $n_y$    |
|-------------|-----|----|----|----------|
|             | 10  | 20 | 30 |          |
| 15          | 4   | 28 | 6  | 38       |
| 25          | 6   | —  | 6  | 12       |
| $n_x$       | 10  | 28 | 12 | $n = 50$ |
| $\bar{y}_x$ | 21  | 15 | 20 |          |

Решение. Сначала найдем  $\bar{y}$ ,  $\tilde{\sigma}_y$  и  $\sigma_{\bar{y}_x}$ :

$$\bar{y} = (38 \cdot 15 + 12 \cdot 25) / 50 = 17,4;$$

$$\tilde{\sigma}_y = \sqrt{[38 \cdot (15 - 17,4)^2 + 12 \cdot (25 - 17,4)^2] / 50} \approx 4,27;$$

$$\sigma_{\bar{y}_x} = \sqrt{[10 \cdot (21 - 17,4)^2 + 28 \cdot (15 - 17,4)^2 + 12 \cdot (20 - 17,4)^2] / 50} \approx 2,73.$$

Теперь подставим все эти значения в формулу (15.5) и найдем  $\eta_{yx}$ :

$$\eta_{yx} = \sigma_{\bar{y}_x} / \tilde{\sigma}_y \approx 2,73 / 4,27 \approx 0,64.$$

Перечислим свойства выборочного корреляционного

отношения.

**Свойство 15.1.** Выборочное корреляционное отношение удовлетворяет двойному неравенству

$$0 \leq \eta_{yx} \leq 1.$$

**Свойство 15.2.** Если  $\eta_{yx} = 0$ , то признак  $Y$  с признаком  $X$  корреляционной зависимостью не связан.

**Свойство 15.3.** Если  $\eta_{yx} = 1$ , то признак  $Y$  связан с признаком  $X$  функциональной зависимостью.

**Свойство 15.4.** Выборочное корреляционное отношение не меньше абсолютной величины выборочного коэффициента корреляции:  $\eta_{yx} \geq |r_B|$ .

**Свойство 15.5.** Если выборочное корреляционное отношение равно абсолютной величине выборочного коэффициента корреляции, то имеет место точная линейная корреляционная зависимость.

### **Вопросы для повторения и контроля:**

1. Что называется корреляционным моментом и что называется коэффициентом корреляции?
2. Что такое коррелированные и некоррелированные случайные величины, и какова связь между понятиями зависимости и коррелированности случайных величин?
3. Что вы знаете о выборочном коэффициенте корреляции?
4. Что такое выборочное корреляционное отношение и для чего оно служит?
5. Какие свойства выборочного корреляционного отношения вы знаете?

### **Опорные слова:**

Корреляционный момент, коэффициент корреляции, коррелированные случайные величины, некоррелированные случайные величины, выборочный коэффициент корреляции, выборочное корреляционное отношение.

## Тема № 16

### Статистические гипотезы и их классификация. Статистический критерий

#### План:

5. Статистические гипотезы и их классификация. Ошибки первого и второго рода.
6. Статистический критерий. Критическая область и критические точки.
7. Отыскание критических областей. Мощность критерия.
8. Сравнение двух дисперсий нормальных генеральных совокупностей.
9. Сравнение двух средних нормальных генеральных совокупностей.

Пусть требуется определить закон распределения генеральной совокупности и назовем его  $A$ . Если закон распределения неизвестен, но имеются основания предположить, что он имеет определенный вид, выдвигают гипотезу: генеральная совокупность распределена по закону  $A$ . Таким образом, в этой гипотезе речь идет о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр  $\Theta$  равен определенному значению  $\Theta_0$ , то выдвигают гипотезу:  $\Theta = \Theta_0$ . Таким образом, в этой гипотезе речь идет о предполагаемой величине параметра одного известного распределения.

*Статистической* называется гипотеза о виде неизвестного распределения или гипотеза о параметрах известных распределений. Например, статистическими являются гипотезы:

- 1) генеральная совокупность распределена по закону Пуассона;
- 2) дисперсии двух нормальных совокупностей равны между собой.

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

известных распределений.

*Нулевой (основной)* называется выдвинутая гипотеза  $H_0$ .

*Конкурирующей (альтернативной)* называется гипотеза  $H_1$ , которая противоречит нулевой.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание  $a$  нормального распределения равно 10, то конкурирующая гипотеза может состоять в предположении, что  $a \neq 10$ ; т.е.  $H_0: a = 10$ ;  $H_1: a \neq 10$ .

*Простой* называется гипотеза, содержащая только одно предположение. Например, гипотеза  $H_0$ : математическое ожидание нормального распределения равно 3 ( $\sigma$  известно) — простая.

*Сложной* называется гипотеза, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза  $H: \lambda > 5$  состоит из бесчисленного множества простых гипотез вида  $H_i: \lambda = b_i$ , где  $b_i$  — любое число, большее 5.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость *статистической* (производимой статистическими методами) проверки этой гипотезы. В итоге статистической проверки гипотезы могут быть допущены ошибки.

*Ошибка первого рода* состоит в том, что будет отвергнута правильная гипотеза.

*Ошибка второго рода* состоит в том, что будет принята неправильная гипотеза.

Для проверки нулевой гипотезы используется специально подобранная случайная величина, точное или приближенное распределение которой известно. Эта случайная величина обозначается через  $K$  и называется *статистическим критерием* (или просто *критерием*).

Приведем пример статистического критерия. Если проверяется гипотеза о равенстве дисперсий двух нормальных генеральных совокупностей, то в качестве критерия  $K$  принимается отношение исправленных выборочных дисперсий:

$$F = s_1^2 / s_2^2.$$

*Наблюдаемым значением*  $K_{\text{набл}}$  называется значение критерия, вычисленное по выборкам. Например, если по двум

выборкам найдены исправленные выборочные дисперсии  $s_1^2 = 20$  и  $s_2^2 = 5$ , то наблюдаемое значение критерия  $F$  равно

$$F_{\text{набл}} = s_1^2 / s_2^2 = 20/5 = 4.$$

После выбора определенного критерия множество всех его возможных значений разбивается на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая — при которых она принимается.

*Критической областью* называется совокупность значений критерия, при которых нулевая гипотеза отвергается.

*Областью принятия гипотезы* (областью допустимых значений) называется совокупность значений критерия, при которых нулевая гипотеза принимается.

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

*Критическими точками* (границами)  $k_{кр}$  называются точки, отделяющие критическую область от области принятия гипотезы.

*Правосторонней* называется критическая область, определяемая неравенством  $K > k_{кр}$ , где  $k_{кр}$  — положительное число (рис. 16.1).

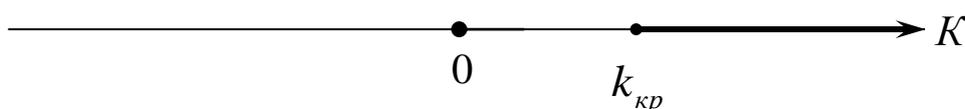


Рис. 16.1.

*Левосторонней* называется критическая область, определяемая неравенством  $K < k_{кр}$ , где  $k_{кр}$  — отрицательное число (рис. 16.2).

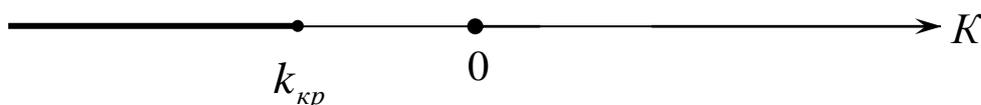


Рис. 16.2.

*Односторонней* называется правосторонняя или левосторонняя критическая область.

*Двусторонней* называется критическая область, определяемая неравенствами  $K < k_1$ ,  $K > k_2$ , где  $k_2 > k_1$ .

В частности, если критические точки симметричны относительно нуля, двусторонняя критическая область определяется неравенствами (в предположении, что  $k_{кр} > 0$ )  $K < -k_{кр}$ ,  $K > k_{кр}$ , или равносильным неравенством  $|K| > k_{кр}$  (рис. 16.3).

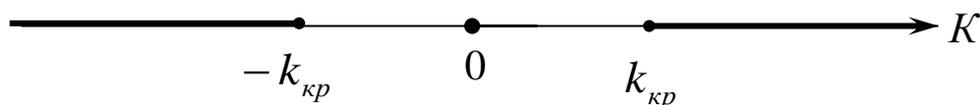


Рис. 16.3.

Для нахождения критической области достаточно найти критическую точку (точки). Для нахождения же такой точки задается достаточно малая вероятность — *уровень значимости*  $\alpha$ . Затем критическая точка  $k_{кр}$  ищется исходя из требования, чтобы при условии справедливости нулевой гипотезы вероятность того, что критерий  $K$  примет значения из критической области, была равна принятому уровню значимости.

Например, для правосторонней критической области должно выполняться соотношение

$$P(K > k_{кр}) = \alpha, \quad (16.1)$$

для левосторонней —

$$P(K < k_{кр}) = \alpha, \quad (16.2)$$

а для двусторонней —

$$P(K < k_1) + P(K > k_2) = \alpha. \quad (16.3)$$

Для каждого критерия имеются соответствующие таблицы, по которым и находится критическая точка, удовлетворяющая требованиям вида (16.1) – (16.3).

Если распределение критерия симметрично относительно нуля и имеются основания выбрать симметричные относительно нуля точки  $-k_{кр}$  и  $k_{кр}$  ( $k_{кр} > 0$ ), то  $P(K < -k_{кр}) = P(K > k_{кр})$ . Учитывая это соотношение, из (16.3) для двусторонней критической области получим соотношение

$$P(K > k_{кр}) = \alpha/2. \quad (16.4)$$

*Мощностью критерия* называется вероятность попадания критерия в критическую область при условии, что справедлива

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

Пусть для проверки гипотезы принят определенный уровень значимости, и выборка имеет фиксированный объем. Если  $\beta$  — вероятность ошибки второго рода, т.е. события «принята нулевая гипотеза, причем справедлива конкурирующая», то мощность критерия равна  $1 - \beta$ .

Пусть мощность  $1 - \beta$  возрастает; следовательно, уменьшается вероятность  $\beta$  совершить ошибку второго рода. Таким образом, чем мощность больше, тем меньше вероятность ошибки второго рода.

Итак, если уровень значимости уже выбран, то критическую область следует строить так, чтобы мощность критерия была максимальной. Это позволит минимизировать ошибку второго рода.

Далее нам потребуется распределение Фишера – Снедекора.

Если  $U$  и  $V$  — независимые случайные величины, распределенные по закону  $\chi^2$  со степенями свободы  $k_1$  и  $k_2$ , то величина

$$F = \frac{U/k_1}{V/k_2} \quad (16.5)$$

имеет распределение, которое называется *распределением  $F$  Фишера – Снедекора* со степенями свободы  $k_1$  и  $k_2$ .

Функция плотности этого распределения имеет вид

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ C_0 \frac{x^{(k_1-2)/2}}{(k_2 + k_1 x)^{(k_1+k_2)/2}} & \text{при } x > 0 \end{cases},$$

где

$$C_0 = \frac{\Gamma\left(\frac{k_1 + k_2}{2}\right) k_1^{k_1/2} k_2^{k_2/2}}{\Gamma(k_1/2)\Gamma(k_2/2)}.$$

Распределение  $F$  определяется двумя параметрами — числами степеней свободы  $k_1$  и  $k_2$ .

Пусть генеральные совокупности  $X$  и  $Y$  распределены нормально. По независимым выборкам с объемами, соответственно равными  $n_1$  и  $n_2$ , извлеченным из этих совокупностей, найдены исправленные выборочные дисперсии  $s_X^2$  и  $s_Y^2$ . Требуется по исправленным дисперсиям при заданном уровне значимости  $\alpha$  проверить нулевую гипотезу, состоящую в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0: D(X) = D(Y). \quad (16.6)$$

Учитывая, что исправленные дисперсии являются несмещенными оценками генеральных дисперсий, т.е.

$$M(s_X^2) = D(X), \quad M(s_Y^2) = D(Y),$$

нулевую гипотезу можно записать так:

$$H_0: M(s_X^2) = M(s_Y^2). \quad (16.7)$$

На практике задача сравнения дисперсий возникает, если требуется сравнить точность приборов, инструментов, самих методов измерений и т.д. Очевидно, предпочтительнее тот прибор, инструмент и метод, который обеспечивает наименьшее рассеяние результатов измерений, т.е. наименьшую дисперсию.

В качестве критерия проверки нулевой гипотезы о равенстве генеральных дисперсий принимается отношение большей исправленной дисперсии к меньшей, т.е. случайная величина

$$F = s_o^2 / s_m^2. \quad (16.8)$$

Величина  $F$  при условии справедливости нулевой гипотезы имеет распределение Фишера – Снедекора со степенями свободы  $k_1 = n_1 - 1$  и  $k_2 = n_2 - 1$ , где  $n_1$  — объем выборки, по которой вычислена большая исправленная дисперсия,  $n_2$  — объем выборки, по которой найдена меньшая исправленная дисперсия.

Критическая область строится в зависимости от вида конкурирующей гипотезы.

Первый случай. Нулевая гипотеза  $H_0: D(X) = D(Y)$ .  
Конкурирующая гипотеза  $H_1: D(X) > D(Y)$ .

В этом случае строится правосторонняя критическая область, исходя из требования, чтобы вероятность попадания критерия  $F$  в эту область в предположении справедливости

нулевой гипотезы была равна принятому уровню значимости:

$$P(F > F_{кр}(\alpha; k_1; k_2)) = \alpha. \quad (16.9)$$

Критическая точка  $F_{кр}(\alpha; k_1; k_2)$  находится по таблице критических точек распределения Фишера – Снедекора.

**Правило 1.** Для того чтобы при заданном уровне значимости проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе  $H_1: D(X) > D(Y)$ , надо вычислить отношение большей исправленной дисперсии к меньшей, т.е.

$$F_{набл} = s_{б}^2 / s_{м}^2, \quad (16.10)$$

и по таблице критических точек распределения Фишера – Снедекора, по заданному уровню значимости  $\alpha$  и числам степеней свободы  $k_1$  и  $k_2$  ( $k_1$  — число степеней свободы большей исправленной дисперсии) найти критическую точку  $F_{кр}(\alpha; k_1; k_2)$ .

Если  $F_{набл} < F_{кр}$  — нет оснований отвергнуть нулевую гипотезу. Если  $F_{набл} > F_{кр}$  — нулевая гипотеза отвергается.

**Пример 1.** По двум независимым выборкам объемов  $n_1 = 12$  и  $n_2 = 15$ , извлеченным из нормальных генеральных совокупностей  $X$  и  $Y$ , найдены исправленные выборочные дисперсии  $s_X^2 = 11,41$  и  $s_Y^2 = 6,52$ . При уровне значимости 0,05 проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий при конкурирующей гипотезе  $H_1: D(X) > D(Y)$

Решение. Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{набл} = 11,41 / 6,52 = 1,75.$$

Конкурирующая гипотеза имеет вид  $D(X) > D(Y)$ , поэтому критическая область — правосторонняя.

По таблице критических точек распределения Фишера – Снедекора, по уровню значимости  $\alpha = 0,05$  и числам степеней свободы  $k_1 = 12 - 1 = 11$  и  $k_2 = 15 - 1 = 14$  находим критическую точку  $F_{кр}(0,05; 11; 14) = 2,56$ .

Так как  $F_{набл} < F_{кр}$ , то нет оснований отвергнуть нулевую

гипотезу о равенстве генеральных дисперсий.

Второй случай. Нулевая гипотеза  $H_0: D(X) = D(Y)$ .  
Конкурирующая гипотеза  $H_1: D(X) \neq D(Y)$ .

В этом случае строится двусторонняя критическая область, исходя из требования, чтобы вероятность попадания критерия  $F$  в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости  $\alpha$ .

Наибольшая мощность критерия (вероятность попадания критерия в критическую область при справедливости конкурирующей гипотезы) достигается тогда, когда вероятность попадания критерия в каждый из двух интервалов критической области равна  $\alpha/2$ .

Если обозначить через  $F_1$  левую границу критической области и через  $F_2$  — правую, то должны иметь место соотношения:

$$P(F < F_1) = \alpha/2, \quad P(F > F_2) = \alpha/2. \quad (16.11)$$

Для обеспечения попадания критерия  $F$  в двустороннюю критическую область с вероятностью, равной принятому уровню значимости  $\alpha$ , в случае конкурирующей гипотезы  $H_1: D(X) \neq D(Y)$  достаточно найти критическую точку  $F_2 = F_{кр}(\alpha/2; k_1; k_2)$ .

**Правило 2.** Для того чтобы при заданном уровне значимости проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе  $H_1: D(X) \neq D(Y)$ , надо вычислить отношение большей исправленной дисперсии к меньшей, т.е. (16.10) и по таблице критических точек распределения Фишера – Снедекора, по заданному уровню значимости  $\alpha/2$  (вдвое меньшем заданного) и числам степеней свободы  $k_1$  и  $k_2$  ( $k_1$  — число степеней свободы большей исправленной дисперсии) найти критическую точку  $F_{кр}(\alpha/2; k_1; k_2)$ .

Если  $F_{набл} < F_{кр}$  — нет оснований отвергнуть нулевую гипотезу. Если  $F_{набл} > F_{кр}$  — нулевая гипотеза отвергается.

**Пример 2.** По двум независимым выборкам объемов  $n_1 = 10$

и  $n_2 = 18$ , извлеченным из нормальных генеральных совокупностей  $X$  и  $Y$ , найдены исправленные выборочные дисперсии  $s_X^2 = 1,23$  и  $s_Y^2 = 0,41$ . При уровне значимости 0,1 проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий при конкурирующей гипотезе  $H_1: D(X) \neq D(Y)$

Решение. Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{\text{набл}} = 1,23/0,41 = 3.$$

Конкурирующая гипотеза имеет вид  $D(X) \neq D(Y)$ , поэтому критическая область — двусторонняя.

По таблице критических точек распределения Фишера – Снедекора, по уровню значимости, вдвое меньшем заданного, т.е. при  $\alpha/2 = 0,05$  и числам степеней свободы  $k_1 = 10 - 1 = 9$  и  $k_2 = 18 - 1 = 17$  находим критическую точку  $F_{\text{кр}}(0,05; 9; 17) = 2,50$ .

Так как  $F_{\text{набл}} > F_{\text{кр}}$ , нулевая гипотеза о равенстве генеральных дисперсий отвергается.

Пусть генеральные совокупности  $X$  и  $Y$  распределены нормально, причем их дисперсии известны. По независимым выборкам с объемами, соответственно равными  $n$  и  $m$ , извлеченным из этих совокупностей, найдены выборочные средние  $\bar{x}$  и  $\bar{y}$ . Требуется по выборочным средним при заданном уровне значимости  $\alpha$  проверить нулевую гипотезу, состоящую в том, что генеральные средние (математические ожидания) рассматриваемых совокупностей равны между собой:

$$H_0: M(X) = M(Y). \quad (16.12)$$

Учитывая, что выборочные средние являются несмещенными оценками генеральных средних, т.е.

$$M(\bar{x}) = M(X), \quad M(\bar{y}) = M(Y),$$

нулевую гипотезу можно записать так:

$$H_0: M(\bar{x}) = M(\bar{y}). \quad (16.13)$$

В качестве критерия проверки нулевой гипотезы о равенстве генеральных средних принимается нормированная нормальная случайная величина

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{D(X)/n + D(Y)/m}}. \quad (16.14)$$

Критическая область строится в зависимости от вида конкурирующей гипотезы.

Первый случай. Нулевая гипотеза  $H_0: M(X) = M(Y)$ .  
Конкурирующая гипотеза  $H_1: M(X) \neq M(Y)$ .

В этом случае строится двусторонняя критическая область, исходя из требования, чтобы вероятность попадания критерия  $Z$  в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости  $\alpha$ .

Поскольку распределение  $Z$  симметрично относительно нуля, то критические точки симметричны относительно нуля, т.е. если обозначить через  $z_{кр}$  правую критическую точку, то  $-z_{кр}$  будет левой критической точкой.

Наибольшая мощность критерия (вероятность попадания критерия в критическую область при справедливости конкурирующей гипотезы) достигается тогда, когда вероятность попадания критерия в каждый из двух интервалов критической области равна  $\alpha/2$ :

$$P(Z < -z_{кр}) = \alpha/2, \quad P(Z > z_{кр}) = \alpha/2. \quad (16.15)$$

Для того, чтобы найти правую границу  $z_{кр}$  двусторонней критической области, достаточно найти значение аргумента функции Лапласа, которому соответствует значение функции, равное  $(1 - \alpha)/2$ :

$$\Phi(z_{кр}) = (1 - \alpha)/2. \quad (16.16)$$

Обозначим значение критерия, вычисленное по данным наблюдений, через  $Z_{набл}$ .

Если  $|Z_{набл}| < z_{кр}$  — нет оснований отвергнуть нулевую гипотезу.

Если  $|Z_{набл}| > z_{кр}$  — нулевая гипотеза отвергается.

Второй случай. Нулевая гипотеза  $H_0: M(X) = M(Y)$ .  
Конкурирующая гипотеза  $H_1: M(X) > M(Y)$ .

В этом случае строится правосторонняя критическая область, исходя из требования, чтобы вероятность попадания критерия  $Z$  в эту область в предположении справедливости

нулевой гипотезы была равна принятому уровню значимости:

$$P(Z > z_{кр}) = \alpha. \quad (16.17)$$

Для того, чтобы найти границу  $z_{кр}$  правосторонней критической области, достаточно найти значение аргумента функции Лапласа, которому соответствует значение функции, равное  $(1 - 2\alpha)/2$ :

$$\Phi(z_{кр}) = (1 - 2\alpha)/2. \quad (16.18)$$

Обозначим значение критерия, вычисленное по данным наблюдений, через  $Z_{набл}$ .

Если  $Z_{набл} < z_{кр}$  — нет оснований отвергнуть нулевую гипотезу.

Если  $Z_{набл} > z_{кр}$  — нулевая гипотеза отвергается.

### **Вопросы для повторения и контроля:**

1. Что вы понимаете под статистической гипотезой? Приведите примеры.
2. Что такое нулевая (основная), конкурирующая (альтернативная), простая, сложная гипотезы?
3. В чем состоят ошибки первого и второго рода, что называется статистическим критерием?
4. Что называется наблюдаемым значением критерия, критической областью, областью принятия гипотезы (областью допустимых значений)?
5. Что такое критические точки (границы), правосторонняя, левосторонняя, односторонняя, двусторонняя критическая области?
6. Что называется уровнем значимости и как находится критическая область?
7. Что такое мощность критерия и как она связана с ошибкой второго рода?
8. Что вы знаете о распределении Фишера – Снедекора?
9. Как сравниваются две дисперсии нормальных генеральных совокупностей в первом случае?
10. Как сравниваются две дисперсии нормальных генеральных совокупностей в втором случае?
11. Как сравниваются два средних нормальных генеральных совокупностей в первом случае?

12. Как сравниваются два средних нормальных генеральных совокупностей в втором случае?

**Опорные слова:**

Статистическая гипотеза, нулевая (основная) гипотеза, конкурирующая (альтернативная) гипотеза, простая гипотеза, сложная гипотеза, ошибка первого рода, ошибка второго рода, статистический критерий, наблюдаемое значение критерия, критическая область, область принятия гипотезы (область допустимых значений), критические точки (границы), правосторонняя критическая область, левосторонняя критическая область, односторонняя критическая область, двусторонняя критическая область, уровень значимости, мощность критерия, распределение Фишера – Снедекора, степени свободы.

## Тема № 17

### Критерии согласия

#### План:

4. Проверка гипотезы о нормальном распределении генеральной совокупности.
5. Критерий согласия Пирсона.
6. Методика вычисления теоретических частот нормального распределения.

Если закон распределения генеральной совокупности неизвестен, но есть основания предположить, что он имеет определенный вид (назовем его  $A$ ), то проверяется нулевая гипотеза: генеральная совокупность распределена по закону  $A$ .

Проверка гипотезы о предполагаемом законе неизвестного распределения производится так же, как и проверка гипотезы о параметрах распределения, т.е. при помощи специально подобранной случайной величины — критерия согласия.

*Критерием согласия* называется критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Одним из критериев согласия является критерий  $\chi^2$  («хи квадрат») К.Пирсона проверки гипотезы о нормальном распределении генеральной совокупности (этот критерий можно применять и для других распределений). Для применения этого критерия будем сравнивать эмпирические (наблюдаемые) и теоретические (вычисленные в предположении нормального распределения) частоты.

Обычно эмпирические и теоретические частоты различаются. Например:

|                        |   |    |    |    |     |    |    |    |   |
|------------------------|---|----|----|----|-----|----|----|----|---|
| эмп. частоты . . . . . | 6 | 13 | 38 | 74 | 106 | 85 | 30 | 10 | 4 |
| теорет. частоты . .    | 3 | 14 | 42 | 82 | 99  | 76 | 37 | 11 | 2 |

Расхождение эмпирических и теоретических частот может быть случайным (незначимым) и объясняется либо малым числом наблюдений, либо способом их группировки, либо другими причинами. С другой стороны, расхождение частот может быть неслучайным (значимым) и объясняется тем, что

теоретические частоты вычислены исходя из неверной гипотезы о нормальном распределении генеральной совокупности.

Критерий Пирсона отвечает на вопрос: случайно ли расхождение эмпирических и теоретических частот? Правда, как и любой критерий, он не доказывает справедливость гипотезы, а лишь устанавливает на принятом уровне значимости ее согласие или несогласие с данными наблюдений.

Пусть по выборке объема  $n$  получено эмпирическое распределение:

|                        |       |       |       |       |       |
|------------------------|-------|-------|-------|-------|-------|
| варианты . . . . .     | $x_i$ | $x_1$ | $x_2$ | . . . | $x_s$ |
| эмп. частоты . . . . . | $n_i$ | $n_1$ | $n_2$ | . . . | $n_s$ |

Допустим, что в предположении нормального распределения генеральной совокупности вычислены теоретические частоты  $n'_i$ . При уровне значимости  $\alpha$  требуется проверить нулевую гипотезу: генеральная совокупность распределена нормально.

В качестве критерия проверки нулевой гипотезы принимается случайная величина

$$\chi^2 = \sum (n_i - n'_i)^2 / n'_i. \quad (17.1)$$

Эта величина случайная, так как в различных опытах она принимает различные, заранее не известные значения. Ясно, что чем меньше различаются эмпирические и теоретические частоты, тем меньше величина критерия (17.1), и, следовательно, он в известной степени характеризует близость эмпирического и теоретического распределений.

При  $n \rightarrow \infty$  закон распределения случайной величины (17.1) независимо от того, какому закону распределения подчинена генеральная совокупность, стремится к закону распределения  $\chi^2$  с  $k$  степенями свободы.

Число степеней свободы находится по равенству  $k = s - 1 - r$ , где  $s$  — число групп (частичных интервалов) выборки;  $r$  — число параметров предполагаемого распределения, которые оценены по данным выборки.

В частности, если предполагаемое распределение — нормальное, то оцениваются два параметра ( математическое ожидание и среднее квадратическое отклонение), поэтому  $r = 2$  и число степеней свободы  $k = s - 1 - r = s - 1 - 2 = s - 3$ .

Если же предполагается, что генеральная совокупность распределена по закону Пуассона, то оценивается один параметр  $\lambda$ , поэтому  $r = 1$  и  $k = s - 2$ .

Построим правостороннюю критическую область, исходя из требования, чтобы вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости:

$$P(\chi > \chi_{кр}^2(\alpha; k)) = \alpha. \quad (17.2)$$

Таким образом, правосторонняя критическая область определяется неравенством  $\chi > \chi_{кр}^2(\alpha; k)$ , а область принятия нулевой гипотезы — неравенством  $\chi < \chi_{кр}^2(\alpha; k)$ .

**Правило.** Для того чтобы при заданном уровне значимости проверить нулевую гипотезу  $H_0$ : генеральная совокупность распределена нормально, надо сначала вычислить теоретические частоты, а затем наблюдаемое значение критерия

$$\chi_{набл}^2 = \sum (n_i - n'_i)^2 / n'_i \quad (17.3)$$

и по таблице критических точек распределения  $\chi^2$ , по заданному уровню значимости  $\alpha$  и числу степеней свободы  $k = s - 3$  найти критическую точку  $\chi_{кр}^2(\alpha; k)$ .

Если  $\chi_{набл}^2 < \chi_{кр}^2$  — нет оснований отвергнуть нулевую гипотезу. Если  $\chi_{набл}^2 > \chi_{кр}^2$  — нулевая гипотеза отвергается.

Сущность критерия согласия Пирсона состоит в сравнении эмпирических и теоретических частот. Ясно, что эмпирические частоты находятся из опыта. Как найти теоретические частоты, если предполагается, что генеральная совокупность распределена нормально? Эту задачу, например, можно решить следующим способом.

1. Весь интервал наблюдаемых значений  $X$  (выборки объема  $n$ ) делится на  $s$  частичных интервалов  $(x_i, x_{i+1})$  одинаковой длины. Затем находятся середины частичных интервалов  $x_i^* = (x_i + x_{i+1})/2$ ; в качестве частоты  $n_i$  варианты  $x_i^*$  принимается число вариантов, которые попали в  $i$ -й интервал. В итоге получается последовательность равноотстоящих вариантов и соответствующих им частот:

$$\begin{array}{ccccccc} x_i^* & x_1^* & x_2^* & \cdots & x_s^* \\ n_i & n_1 & n_2 & \cdots & n_s \end{array}$$

При этом  $\sum n_i = n$ .

2. Вычисляются выборочная средняя  $\bar{x}^*$  и выборочное среднее квадратическое отклонение  $\sigma^*$ .

3. Нормируется случайная величина  $X$ , т.е. переходят к величине  $Z = (X - \bar{x}^*)/\sigma^*$  и вычисляются концы интервалов  $(z_i, z_{i+1})$ :

$$z_i = (x_i - \bar{x}^*)/\sigma^*, \quad z_{i+1} = (x_{i+1} - \bar{x}^*)/\sigma^*,$$

причем наименьшее значение  $Z$ , т.е.  $z_1$ , полагают равным  $-\infty$ , а наибольшее, т.е.  $z_s$ , полагают равным  $\infty$ .

4. Вычисляются теоретические вероятности  $p_i$  попадания  $X$  в интервалы  $(x_i, x_{i+1})$  по равенству ( $\Phi(z)$  — функция Лапласа)

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

и, наконец, находятя искомые теоретические частоты  $n'_i = np_i$ .

### Вопросы для повторения и контроля:

1. Что называется критерием согласия и как применяется критерий Пирсона?
2. По каким причинам различаются эмпирические и теоретические частоты?
3. Какая случайная величина принимается в качестве критерия проверки нулевой гипотезы о нормальном распределении генеральной совокупности и какие ее свойства вы знаете?
4. В чем суть правила проверки нулевой гипотезы о нормальном распределении генеральной совокупности?
5. Каким способом находятя теоретические частоты?

### Опорные слова:

Критерий согласия, критерий Пирсона, эмпирическая частота, теоретическая частота, правило проверки нулевой гипотезы о нормальном распределении генеральной совокупности.

## Список литературы

1. Адиров Т.Ғ, Мамуров Э.Н. Эһтимоллар назарияси ва математик статистикадан маърузалар матни. Т.: ТМИ, 2001 й.
2. Г.М.Булдык. Теория вероятностей и математическая статистика. М.: Наука, 1989 г.
3. Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. М.: «Высшая школа», 1987 г.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. Издание шестое. М.: «Высшая школа», 1998 г.
5. Гмурман В.Е. Эһтимоллар назарияси ва математик статистика. Русча тўлдирилган 4-нашридан тарж. Инж.-экон. институтлари студентлари учун ўқув кўлланма. Т.: Ўқитувчи, 1977 й.
6. В.Е.Гмурман. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для вузов. 3-е изд., перераб. и доп. М.: «Высшая школа», 1979 г.
7. Гмурман В.Е. Эһтимоллар назарияси ва математик статистикадан масалалар ечишга доир кўлланма. Русча тўлдирилган 2-нашридан таржима. Т.: Ўқитувчи, 1980 й.
8. Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике. М.: Изд. ДИС, 1998 г.
9. Колемаев В.А., Калинина В.А. Теория вероятностей и математическая статистика. М.: Инфра-М, 1997 г.
10. Колемаев В.А., О.В.Староверов, В.Б.Турундаевский. Теория вероятностей и математическая статистика: учеб. пособие для экон. спец. вузов. М.: «Высшая школа», 1991 г.
11. Кремер Ш.А. Теория вероятностей и математическая статистика. М.: «Высшая школа», 2001 г.
12. Мамуров Э.Н., Адиров Т.Ғ Эһтимоллар назарияси ва математик статистикадан масалалар ечишга доир кўлланма. Т.: ТМИ, 2000 й.
13. Соатов Ё.У. Олий математика курси. 2-қисм. Т.: Ўқитувчи, 1994 й.
14. Справочник по математике для экономистов. / Под редакцией проф. Ермакова. М.: «Высшая школа», 1987 г.





