ФГБОУ ВО НОВОСИБИРСКИЙ ГАУ

Основы теории вероятности и математической статистики

Методическое пособие для практических занятий

для студентов направления подготовки

36.05.01 Ветеринария

Новосибирск 2023

Кафедра математики и физики

Основы теории вероятности и математической статистики.: Метод. пособие. /Новосиб. гос. аграр. ун-т.; сост. : С.Н. Шумарева.-Новосибирск, 2023.- 57 с.

Методическое пособие предназначено для студентов очной, очно-заочной форм обучения по направлению *Ветеринария*. Пособие содержит краткую теорию и примеры решения задач по теории вероятностей и математической статистике. К каждому занятию даны задачи для самостоятельного решения.

Утверждено и рекомендовано к изданию учебно-методическим советом ветеринарного факультета (протокол № от 22.05 2023 г.).

ВВЕДЕНИЕ

Методическое пособие для практических занятий по высшей математике предназначено для студентов аграрных вузов. Согласно программе, пособие содержит следующие разделы: элементы теории вероятностей, статистический анализ результатов исследований, корреляционный и регрессионный анализ.

В начале каждого занятия даются основные определения, соответствующие формулы и методические указания, необходимые для решения задач. В конце каждого раздела даны примеры и задачи для самостоятельного решения. В приложении приводятся табличные данные.

В настоящее время значительно повышаются требования к математической культуре специалистов- аграриев. Для достоверности полученных результатов требуется применение адекватного математического аппарата. Поэтому одной из задач является обучение будущих специалистов правильной интерпретации данных.

Тема 5. ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ*

Теория вероятностей – математическая наука, изучающая закономерности случайных явлений.

Теория вероятностей изучает случайные события и случайные величины.

5.1. Случайное событие

Случайное событие – это любой факт, который в результате испытания может произойти или не произойти. Случайное событие – это результат испытания.

Испытание (опыт, эксперимент) – в этом определении понимается выполнение определенного комплекса условий, в которых наблюдается то или иное явление, фиксируется тот или иной результат. Испытание может проводиться человеком, но может осуществляться и независимо от человека. Человек в этом случае выступает в роли наблюдателя.

События обозначаются начальными прописными (заглавными) буквами латинского алфавита **A, B, C.**

1. *Достоверное событие* – это событие, которое в результате испытания обязательно должно произойти.
2. *Невозможное событие* – это событие, которое в результате испытания вообще не может произойти.

События называются *равновозможными*, если по условиям испытания ни одно из этих событий не является объективно более возможным, чем другое.

События называются *несовместными*, если наступление одного из них исключает появление другого. В противном случае события – *совместные*.

Несколько событий образуют *полную группу*, если в результате опыта обязательно появится хотя бы одно из них.

* Темы 1-4 приведены в части2.

События образующие полную группу событий и являющиеся несовместными и равновозможными, называются *случаями.*

Под противоположным событием $\overline{A}$ понимают событие, которое обязательно должно произойти, если не наступило некоторое событие А ($\overline{A}$ читается «не А»).

5.1.1. Вероятность случайного события

Численная мера степени объективности возможности наступления события называется *вероятностью случайного события*.

Классическое определение вероятности события А:

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

Вероятность события А равна отношению числа случаев, благоприятствующих событию А(m), к общему числу случаев (n).

Свойства вероятности события:

1. $0 \le P(A) \le 1$ для любого события А.
2. Если А - событие невозможное, то P(A)=0.
3. Если А - событие достоверное, то P(A)=1.

♦ Задача 5.1

Лабораторная крыса, помещенная в лабиринт, должна избрать один из пяти возможных путей. Лишь один из них ведет к поощрению в виде пищи. В предположении, что крыса с одинаковой вероятностью изберет любой путь, какова вероятность выбранного пути, ведущего к пище?

Решение: $\frac{1}{5}$.

♦ Задача 5.2

При бросании игральной кости возможно шесть исходов: выпадение 1, 2, 3, 4, 5, 6 очков. Какова вероятность появления четного числа очков?

Решение: $P(A) = \frac{3}{6} = \frac{1}{2}$.

Событию А – «появление четного числа очков» благоприятствуют 3 исхода (2, 4 и 6 очков).

♦ Задача 5.3

Подбрасываются 2 монеты. Какова вероятность, что обе упадут «гербом» кверху?

Решение

Четыре исхода бросания двух монет: ГГ, ГР, РГ, РР.

Пусть событие А – «выпали два герба» - этому событию благоприятствует один исход.

$$P(A) = \frac{m}{n} = \frac{1}{4} = 0{,}25.$$

♦ Задача 5.4

Подбрасываются два игральных кубика, подсчитывается сумма очков на верхних гранях. Что вероятнее – получить в сумме 7 или 8?

Решение

Обозначим события: А – «выпало 7очков», В – «выпало 8 очков».

Событию А благоприятствуют 6 элементарных исходов: (1;6), (2;5), (3;4), (4;3), (5;2), (6;1).

Событию В благоприятствуют 5 элементарных исходов: (2;6), (3;5), (4;4), (5;3), (6;2).

Всех равновозможных исходов $n = 6^2 = 36$.

$P(A) = \frac{6}{36}$,

$P(B) = \frac{5}{36}$.

Итак, P(A) > P(B) получить в сумме 7 очков более вероятное событие, чем получить в сумме 8 очков.

Задачи для самостоятельного решения

1. Из букв слова «дифференциал» наугад выбирается одна буква. Какова вероятность того, что эта буква будет: а) гласной; б) согласной; в) «ч»?
2. В коробке находится 10 шаров: 3 белых и 7 черных. Из нее наугад извлекается один шар. Какова вероятность того, что этот шар будет белым? Черным?
3. В коробке имеется 7 желтых и несколько белых таблеток. Какова вероятность вытащить белую таблетку, если вероятность вытащить желтую таблетку равна $\frac{1}{6}$? Сколько белых таблеток в коробке?
4. Бросают две монеты. Найти вероятность того, что хотя бы на одной монете появится «герб».
5. Бросают две монеты. Найти вероятность того, что ни на одной монете не появится «герб».
6. Набирая номер телефона, абонент забыл одну цифру, и набрал ее наугад. Какова вероятность того, что набранная цифра правильная?

5.1.2. Статистическое определение вероятности

Относительная частота события – это доля тех *фактически проведенных* испытаний, в которых событие А появилось $W = P^*(A) = \frac{m}{n}$. Это опытная экспериментальная характеристика, где *m* – число опытов, в которых появилось событие А; *n* – число всех проведенных опытов.

Вероятностью события называется число, около которого группируются значения частоты данного события в различных сериях большого числа испытаний $P(A) = \lim\limits_{n-\infty} \frac{m}{n}$.

♦ Задача 5.5

Из 982 больных, поступивших в хирургическую больницу за месяц, 275 человек имели травмы. Какова относительная частота поступления больных с этим видом заболевания?

Решение: $P^*(A) = \frac{275}{982}$.

♦ Задача 5.6

При стрельбе по мишени частота попадания $w = 0{,}75$. Найти число попаданий при 40 выстрелах.

Решение: $W = \frac{m}{n} \Rightarrow m = 0{,}75 \cdot 40 = 30$.

Ответ: было получено 30 попаданий.

Задачи для самостоятельного решения

7. Среди 1000 новорожденных оказались 515 мальчиков. Чему равна частота рождения мальчиков?
8. Частота нормального выхода семян W = 0,97. Из высеянных семян взошло 970. Сколько семян было высеяно?
9. Для выявления качества семян было отобрано и высеяно в лабораторных условиях 100 штук, из которых 93 дали нормальный всход. Какова частота нормального всхода семян?
10. Среди 300 пробирок, изготовленных на автоматической линии, оказалось 15 не отвечающих стандарту. Найти частоту появления стандартных пробирок.

5.1.3. Закон сложения вероятностей

Сумма двух событий – это такое событие, при котором появляется хотя бы одно из этих событий (A или B).

Если A и B *совместные* события, то их сумма A + B обозначает наступление события A или события B, или обоих событий вместе.

Если A и B *несовместные* события, то их сумма A + B обозначает наступление или события A или события B.

Вероятность суммы *несовместных* событий равна сумме вероятностей этих событий:

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

Вероятность суммы двух *совместных* событий равна сумме вероятностей этих событий без вероятности их совместного появления:

$$P(A + B) = P(A) + P(B) – P(AB)$$

Несколько событий образуют *полную группу*, если в результате опыта обязательно появится хотя бы одно из них.

Сумма вероятностей дискретных событий, образующих полную группу, равна единице

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

или

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

Сумма вероятностей противоположных событий равна единице:

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

♦ Задача 5.7

Победитель соревнования награждается призом (событие А), денежной премией (событие В), медалью (событие С). Что представляют собой события А + В?

Решение

Событие А + В состоит в награждении победителя или призом или денежной премией, или тем и другим.

♦ Задача 5.8

Турист имеет возможность посетить 3 города: А, В и С. Обозначаем события: А – турист посетит город А;

В – турист посетит город В;

С – турист посетит город С.

В чем заключается событие А + С?

Решение

Турист посетил только один из городов А или С, или он посетил их оба.

♦ Задача 5.9

Вероятность того, что у взрослого пациента все зубы сохранились, равна 0,67. Вероятность того, что некоторые зубы отсутствуют, равна 0,24. Вероятность того, что он беззубый, равна 0,09. Вычислить вероятность того, что у пациента несколько зубов.

Решение: $P(A + B) = P(A) + P(B) = 0,67 + 0,24 = 0,91$.

♦ Задача 5.10

В большой популяции плодовой мушки 25% мух имеют мутацию глаз, 50% - мутацию крыльев, а 40% мух с мутацией глаз имеют мутацию крыльев. Какова вероятность того, что у мухи, неудачу выбранной из этой популяции, окажется хотя бы одна из этих мутаций?

Решение

А – событие, состоящее в том, что случайно выбранная муха имеет мутации глаз. В есть событие, состоящее в том, что случайно выбранная муха имеет мутацию крыльев. Вероятность того, что муха имеет одну или обе мутации:

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

Тогда $$P(A + B) = 0{,}25 + 0{,}5 - 0{,}4 \cdot 0{,}25 = 0{,}65$$

Задачи для самостоятельного решения

11. В коробке 30 таблеток: 10 красных, 5 желтых, 15 белых. Найти вероятность появления цветной таблетки (т.е. или красной или желтой).

12. В ванну, где содержатся 3 рыбы: А, В и С - время от времени помещают кусочки пищи. Каждый раз, когда бросают кусочек, рыбы конкурируют за него. Допустим, что за длительный период было установлено, что А или В добивались успеха в течение $\frac{1}{2}$ времени, а А или С в течение $\frac{3}{4}$ всего времени наблюдения. 1. Какова вероятность того, что добивается успеха рыба А? 2. Какая из рыб накормлена лучше?

13. В некоторую больницу поступают пациенты с четырьмя видами болезней. Многолетние наблюдения показали, что этим группам соответствуют относительные частоты 0,1; 0,4; 0,3; 0,2. Для лечения заболеваний с частотой 0,1 и 0,2 необходимо переливание крови. Какое количество больных следует обеспечить кровью, если в течение месяца поступило 1000 больных?

14. Вероятность попадания в мишень для первого спортсмена 0,85, а для второго – 0,8. Спортсмены независимо друг от друга сделали по одному выстрелу. Найти вероятность того, что в мишень попадет хотя бы один спортсмен.

15. Один стрелок поражает мишень с вероятностью 90%, другой с вероятностью 75%. Найти вероятность поражения цели, если оба стрелка стреляют в нее одновременно. Цель считается пораженной при попадании в нее хотя бы одной из двух пуль.

16. Из колоды в 36 карт наудачу вынимается одна. Какова вероятность того, что будет вынута пика или туз?

17. Брошена игральная кость. Найти вероятность того, что выпадет четное или кратное трем число очков.

18. Консультационный пункт университета получает пакеты с контрольными работами из городов А, В и С. Вероятность получения пакета из города А равна 0,6, а из города В – 0,1. Найти вероятность того, что очередной пакет будет получен из города С.

19. С первого предприятия поступило 200 пробирок, из которых 190 стандартных, а со второго – 300, из которых 280 стандартных. Найти вероятность того, что наудачу взятая пробирка будет стандартной.

5.1.4. Условная вероятность

Условная вероятность события В – вероятность события В, найденная при условии, что событие А произошло. Обозначается Р(В/А).

♦ Задача 5.11

В коробке содержится 3 белых и 3 жёлтых таблетки. Из коробки дважды вынимают наугад по одной таблетке, не возвращая их в коробку. Найти вероятность появления белой таблетки при втором испытании (событие В), если при первом испытании была извлечена жёлтая таблетка (событие А).

Решение

После первого испытания в коробке осталось 5 таблеток, из них 3 белых.

Искомая условная вероятность: $P(B/A) = \frac{3}{5} = 0{,}6.$

♦ Задача 5.12

В коробке находится 8 красных и 6 белых таблеток. Из коробки последовательно без возвращения извлекают 3 таблетки. Найти вероятность того, что все 3 таблетки белые.

Решение

Обозначим: A_1 - первая таблетка белая, A_2 - вторая таблетка белая, A_3 - третья таблетка белая.

$$P(A_1A_2A_3) = P(A_1)\ P(A_2/A_1)\cdot P(A_3/A_1A_2);$$

$$P(A_1) = \frac{6}{14};\quad P(A_2/A_1) = \frac{5}{13};\quad P(A_3/A_1A_2) = \frac{4}{12};$$

$$P(A) = P(A_1A_2A_3) = \frac{6}{14}\cdot\frac{5}{13}\cdot\frac{4}{12} = \frac{5}{91} = 0{,}055.$$

5.1.5. Закон умножения вероятностей

Произведение двух событий – это событие, состоящее в совместном появлении этих событий (А и В).

Событие В называется *независимым* от события А, если появление события А не изменяет вероятности появления события В.

Вероятность появления нескольких *независимых* событий равна произведению вероятностей этих событий:

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

Для *зависимых* событий:

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

Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого, найденную в предположении, что первое событие произошло.

♦ Задача 5.13

Пусть имеются следующие события: А – «из колоды карт вынута дама»; В – «из колоды карт вынута карта пиковой масти». Что представляет собой событие АВ?

Решение: АВ есть событие «вынута дама пик».

♦ Задача 5.14

Найти вероятность совместного появления герба при одном бросании двух монет.

Решение: $P(AB) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$

♦ Задача 5.15

Вероятность того, что у взрослого пациента все зубы сохранились, равна 0,67. Какова вероятность того, что у двух не имеющих отношения друг к другу больных, ожидающих приёма в кабинете стоматолога, есть все зубы?

Решение: $P(A \cdot B) = P(A) \cdot P(B) = 0{,}67 \cdot 0{,}67 = 0{,}45.$

♦ Задача 5.16

Найти вероятность того, что в семьях из двух детей:

1) оба ребёнка – мальчики; 2) оба ребёнка – девочки; 3) старший ребёнок мальчик, а младший – девочка. Вероятность рождения мальчика – 0,515.

Решение

$P(ММ) = P(М) \cdot P(М) = 0{,}515 \cdot 0{,}515 = 0{,}265;$

$P(ДД) = 0{,}485 \cdot 0{,}485 = 0{,}235;$

$P(МД) = 0{,}515 \cdot 0{,}485 = 0{,}25.$

♦ Задача 5.17

Вероятность того, что студент в летнюю сессию сдаст первый экзамен, равна 0,9, второй – 0,9, третий – 0,8. Найти вероятность того, что студентом будут сданы: 1) только второй экзамен; 2) все три экзамена.

Решение

1) $P(B) = P(\overline{A}_1 A_2 \overline{A}_3) = P(\overline{A}_1) \cdot P(A_2) \cdot P(\overline{A}_3) = 0{,}1 \cdot 0{,}9 \cdot 0{,}2 = 0{,}018;$

2) $P(A_1 A_2 A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) = 0{,}9 \cdot 0{,}9 \cdot 0{,}8 = 0{,}648.$

Теорема. Вероятность появления хотя бы одного из событий A_1, A_2, …, A_n, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий $\overline{A}_1$, $\overline{A}_2$, …, $\overline{A}_n$.

♦ Задача 5.18

Вероятность попадания в цель при стрельбе из трёх орудий такова: $P_1 = 0{,}75;\ P_2 = 0{,}8;\ P_3 = 0{,}85.$ Какова вероятность хотя бы одного попадания (событие А) при одном залпе из всех этих орудий?

Решение

$g_1 = 1 - P_1 = 1 - 0{,}75 = 0{,}25;$

$g_2 = 1 - P_2 = 1 - 0{,}8 = 0{,}2;$

$g_3 = 1 - P_3 = 1 - 0{,}85 = 0{,}15;$

$P(A) = 1 - g_1 g_2 g_3;$

$P(A) = 1 - 0{,}25 \cdot 0{,}2 \cdot 0{,}15 = 0{,}9925.$

♦ Задача 5.19

Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8. Найти вероятность того, что при одном залпе в мишень попадёт только один стрелок.

Решение

Вероятность того, что в мишень попадёт первый стрелок и не попадёт второй, равна:

$$P(A_1 \overline{A}_2) = 0{,}7 \cdot (1 - 0{,}8) = 0{,}7 \cdot 0{,}2 = 0{,}14.$$

Вероятность того, что в мишень попадёт второй стрелок и не попадёт первый, равна:

$$P(\overline{A}_1 A_2) = (1 - 0{,}7) \cdot 0{,}8 = 0{,}3 \cdot 0{,}8 = 0{,}24.$$

Вероятность того, что в мишень попадёт только один стрелок, равна сумме этих вероятностей:

$$P(A_1 \overline{A}_2) + P(\overline{A}_1 A_2) = 0{,}14 + 0{,}24 = 0{,}38.$$

♦ Задача 5.20

Сколько должна планировать пара иметь детей, чтобы вероятность хотя бы одного мальчика была выше 90% (вероятность рождения мальчика и девочки – 0,5).

Решение

Пусть вероятность того, что все девочки:

$$P(Д) = \frac{1}{2}\frac{1}{2}\frac{1}{2}\ldots\frac{1}{2} = \left(\frac{1}{2}\right)^n.$$

Вероятность того, что не все девочки:

$$P(\textit{хотя бы один мальчик}) = 1 - \left(\frac{1}{2}\right)^n = 0{,}9.$$

$$0{,}1 = \left(\frac{1}{2}\right)^n; \frac{1}{10} = \frac{1}{2^n}; 2^n = 10 \Rightarrow n \approx 4.$$

Задачи для самостоятельного решения

21. Пусть имеются следующие события: А – «из колоды карт вынут туз»; В – «из колоды карт вынута карта пиковой масти». Что представляет собой событие АВ?

22. Брошены монета и игральная кость. Найти вероятность совмещения событий: «появился герб»; «появилось 6 очков».

23. Найти вероятность того, что при бросании игральной кости выпадет чётное число (событие А) и число, делящиеся на 3 (событие В).
24. В одном аквариуме находятся: 3 белые, 3 красные и 3 голубые рыбки. Трёх случайно выбранных рыбок переносят в другой аквариум. Какова вероятность того, что все 3 рыбки белые?
25. Студент изучает биологию, химию и физику. Он оценивает, что вероятность получить «пятёрку» по этим предметам равна соответственно: $P(Б)=\frac{1}{2}; P(Х)=\frac{1}{3}; P(Ф)=\frac{1}{4}$. Предположим, что оценки студента по трём предметам независимы. Какова вероятность, что он: 1) не получит ни одной «пятёрки»; 2) получит «пятёрку» только по биологии?
26. На стеллаже библиотеки в случайном порядке 7 учебников по менеджменту, из которых три – в переплёте. Было вытащено наудачу 2 учебника. Какова вероятность, что оба учебника будут в переплёте?
27. На лекции по биофизике во втором семестре присутствуют 124 студента. Из них на экзамене по высшей математике в зимнюю сессию получили оценку «отлично» 19 человек, «хорошо» - 50 человек, «удовлетворительно» - 24 и не сдали экзамен 31 человек. Какова вероятность того, что вызванные наугад один за другим два студента из числа присутствующих на лекции не имеют задолженности по высшей математике?
28. Студент пришёл на зачёт, зная из 30 вопросов только 24. Какова вероятность сдать зачёт, если после отказа отвечать на вопрос преподаватель задаёт ещё один вопрос?
29. Вероятность того, что в течение одного рабочего дня возникнет неполадка в определённом медицинском приборе, равна 0,05. Какова вероятность того, что не произойдёт ни одной неполадки за 3 рабочих дня?
30. Три охотника одновременно стреляют в зайца. Шанс на успех первого охотника расценивается как 3 из 5; второго – 3 из 10; наконец, для третьего охотника он составляют лишь 1 из 10. Какова вероятность того, что заяц будет подстрелен?
31. Вероятность того, что в летнюю сессию студент сдаст первый экзамен, равна 0,8, второй – 0,9, третий – 0,8. Какова вероятность того, что он сдаст только первый экзамен?
32. В коробке 3 белых и 3 желтых таблетки. Из коробки дважды вынимают наудачу по одной таблетке, не возвращая их в коробку. Найти вероятность появления белой таблетки при втором испытании (событие В), если при первом испытании была извлечена жёлтая таблетка (событие А).
33. В коробке 8 красных и 6 белых таблеток. Из коробки последовательно без возвращения извлекают 3 таблетки. Найти вероятность того, что все таблетки белые.
34. Колода из 36 карт разложена по мастям. Из каждой масти выбирают по одной карте. Какова вероятность того, что все 4 карты тузы?
35. Вероятность того, что в летнюю сессию студент сдаст первый экзамен, равна 0,8, второй – 0,9, третий – 0,8. Найти вероятность того, что студент сдаст хотя бы один экзамен.

36. Вероятность одного попадания в цель при одном залпе из двух орудий равна 0,38. Найти вероятность поражения цели при одном выстреле первым из орудий, если известно, что для второго орудия эта вероятность равна 0,8.
37. Отдел технического контроля проверяет медицинское изделие на стандартность. Вероятность того, что изделие стандартное, равна 0,9. Найти вероятность того, что из двух проверенных изделий только одно стандартное.

5.1.6. Формула полной вероятности. Формула Байеса

Вероятность события А, которое может произойти при условии осуществления одного из несовместных событий $H_1, H_2, ..., H_n$, образующих полную группу, определяется формулой полной вероятности
$P(A)= P(H_1)P(A/H_1)+ P(H_2)P(A/H_2)+ \ldots +P(H_n)P(A/H_n)$.
Так как изначально неизвестно, какое из событий $H_1, H_2, ..., H_n$ произойдет, то эти события стали называть *гипотезами.*

Формула Байеса применяется, когда событие А, которое может появиться только с одной из гипотез $H_1, H_2, ..., H_n$, произошло и необходимо произвести количественную переоценку *априорных* вероятностей этих гипотез $P(H_1), P(H_2), ..., P(H_n)$, известных до испытания, т.е. найти *апостериорные* (получаемые после проведения испытания) условные вероятности гипотез $P(H_1/A), P(H_2/A), ..., P(H_n/A)$:

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}$$

Или вместо Р(А) используем её значение, вычисленное по формуле полной вероятности:

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(H_i) \cdot P(A/H_i) + ... + P(H_n)P(A/H_n)}.$$

Итак, пусть до опыта имеются гипотезы $H_1, H_2, ..., H_n$. После опыта становится известной информация о результатах опыта, но не полная, а именно: результаты наблюдений показывают, что наступило некоторое событие А.

Считается, что до опыта были известны (*априорные*) вероятности гипотез $P(H_1), P(H_2), ..., P(H_n)$ и *условные* вероятности $P(A/H_1), P(A/H_2), ..., P(A/H_n)$.
Необходимо определить *апостериорные* вероятности гипотез $P(H_1/A), P(H_2/A), ..., P(H_n/A)$.

Значение формулы Байеса состоит в том, что при наступлении события А, т.е. по мере получения новой информации, мы можем проверять и корректировать выдвинутые до испытания гипотезы. Такой подход называется байесовским.

♦ Задача 5.21

Два охотника одновременно стреляют одинаковыми пулями в медведя. В результате медведь был убит одной пулей (событие А). Как охотники должны поделить шкуру убитого медведя, если известно, что вероятность попадания у первого охотника 0,3, а у второго 0,6?

Решение

Воспользуемся формулой Байеса. Определим предварительно гипотезы.

Гипотеза H_1: попал первый охотник, второй промахнулся.

Гипотеза H_2: попал второй, первый промахнулся.

Гипотеза H_3: попали оба охотника.

Гипотеза H_4: оба промахнулись.

Событие А может произойти только тогда, когда произошла либо гипотеза H_1, либо гипотеза H_2, т.е.:

$$P(A/H_1)=1, \qquad P(A/H_3)=0,$$
$$P(A/H_2)=1, \qquad P(A/H_4)=0.$$

Предполагаем, что попадания охотников в медведя не зависят друг от друга. И получаем:

$$P(H_1)=0{,}3(1-0{,}6)=0{,}12;$$
$$P(H_2)=0{,}6(1-0{,}3)=0{,}42;$$
$$P(H_3)=0{,}3\cdot 0{,}6=0{,}18;$$
$$P(H_4)=(1-0{,}3)(1-0{,}6)=0{,}28.$$

Применяем формулу полной вероятности:

$$P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+P(H_3)P(A/H_3).$$

Затем применяем формулу Байеса:

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+P(H_3)P(A/H_3)+P(H_4)P(A/H_4)}$$

$$P(H_1/A)=\frac{0{,}12\cdot 1}{0{,}12\cdot 1+0{,}42\cdot 1+0{,}18\cdot 0+0{,}28\cdot 0}=\frac{2}{9};$$

$$P(H_2/A)=\frac{P(H_2)\cdot P(A/H_2)}{P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)+P(H_4)\cdot P(A/H_4)}$$

$$P(H_2/A)=\frac{0{,}42\cdot 1}{0{,}12\cdot 1+0{,}42\cdot 1+0{,}18\cdot 0+0{,}28\cdot 0}=\frac{7}{9}.$$

Таким образом, при справедливом делении первый охотник должен получить $\frac{2}{9}$ шкуры, т.е. меньше четвёртой части шкуры, в то время как на первый взгляд казалось, что ему причитается $\frac{1}{3}$ шкуры (0,3).

Задачи для самостоятельного решения

38. Имеются два одинаковых ящика с шарами. В первом ящике 2 белых и 1 чёрный шар, а во втором – 1 белый и 4 чёрных шара. Наудачу выбирают один ящик и вынимают из него шар. Какова вероятность того, что извлеченный шар является белым?

39. В чёрном ящике лежит шар, который с равной вероятностью может быть либо чёрным, либо белым. В ящик добавляют белый шар, затем наугад извлекают шар, оказавшийся белым. Какова вероятность того, что оставшийся шар является белым?

При проведении расчётов по формуле Байеса можно воспользоваться следующей подсказкой. Следует диагностировать состояния ящика:

A – белый шар, белый шар;

B – чёрный шар, белый шар.

40. Предположим, что 5% всех мужчин и 0,25% всех женщин – дальтоники. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность того, что это мужчина? (Считать, что мужчин и женщин одинаковое число).

5.1.7. Схема Бернулли

Схемой Бернулли или схемой повторных независимых испытаний с двумя исходами "успех" или "неуспех", называется последовательность n независимых испытаний, в каждом из которых "успех" наступает с одной и той же вероятностью p ≠0 и 1.

Вероятность того, что при n испытаниях "успех" наступит ровно k раз, вычисляется по формуле Бернулли:

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

где

n - число испытаний;

k - число "успехов";

p - вероятность "успеха" в одном испытании;

q = 1 - p - вероятность "неуспеха";

$C_n^k = \frac{n!}{k! \cdot (n-k)!}$ - число сочетаний из n элементов по k.

Задача 5.22

Вероятность заболевания животного во время эпидемии 0,2. Найти вероятность, что из 6 животных 2 заболеют.

Решение

Число животных n = 6, число "успехов" k = 2, p = 0,2, q = 1 – 0,2 = 0,8.

$$P_6(2) = C_6^2 \cdot 0{,}2^2 \cdot 0{,}8^4 = \frac{6!}{2! \cdot (6-2)!} \cdot 0{,}2^2 \cdot 0{,}8^4 = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 4} \cdot 0{,}04 \cdot 0{,}8^4 =$$

$= \frac{5 \cdot 6}{2} \cdot 0{,}04 \cdot 0{,}4 = 0{,}25.$

При больших n использование формулы Бернулли затруднительно, поэтому в этих случаях применяют приближенные формулы, которые следуют из локальной теоремы Лапласа и из теоремы Пуассона.

Выбор формулы для решения задачи на схему Бернулли поможет сделать следующая таблица:

Название формулы	Формула	Когда даст хорошее решение
Формула Бернулли	$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}$	Для всех n и p
Формула, следующая из локальной теоремы Лапласа	$P_n(k) \approx \frac{1}{\sqrt{npg}} \cdot \varphi(x)$ $x = \frac{k - np}{\sqrt{npg}}$;$\varphi(x) = \frac{1}{\sqrt{2\pi}} \ell^{-\frac{x^2}{2}}$	При p>0,1 или np>9
Формула, следующая из теоремы Пуассона	$P_n(k) \approx \frac{e^{-\lambda} \cdot \lambda^k}{k!}$; $\lambda = np$	P ≤0,1; np≤9, n>10

Имеются таблицы, в которых помещены значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \ell^{-\frac{x^2}{2}}$

Свойства функции φ(x):

1) φ(-x) = φ(x);

2) при x > 4 φ(x) ≈ 0.

Задача 5.23

Допустим, укореняют 15 черенков роз. Приживаемость 80%. Найти вероятность того, что из 15 черенков укоренится ровно 12.

Решение.

n = 15; k = 12; p = 0,8; q = 1 – 0,8 = 0,2.

Имеем npq = 15 · 0,8 · 0,4 = 2,4.

$$x = \frac{k - np}{\sqrt{npq}} = \frac{12 - 15 \cdot 0,8}{\sqrt{2,4}} = \frac{12 - 12}{\sqrt{2,4}} = 0;$$

$\varphi(0) = 0,3989;$ $P_{12}(15) = \frac{1}{\sqrt{2,4}} \cdot 0,3989 = 0,2581.$

Интегральная теорема Лапласа

Вероятность того, что в n независимых испытаниях, проведенных по схеме Бернулли, событие наступит не менее k_1 и не более k_2 раз, приближенно равна

$P(k, k_2) \approx \Phi(x_2) - \Phi(x_1)$,

где $\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \int e^{\frac{-t^2}{2}} dt$ - интегральная функция Лапласа.

$x_1 = \frac{k_1 - np}{\sqrt{npg}}$; $x_2 = \frac{k_2 - np}{\sqrt{npg}}$;

Значение функции Лапласа занесены в таблицу.

Если на формировании величин влияет большое число факторов, влияние каждого из них мало и ни один фактор не имеет значительного преимущества перед другими, то эти величины можно отнести к величинам, имеющим нормальный закон распределения, полагая, что их возможные значения не отрицательны.

Задача 5.24

Вероятность того, что подготовка почвы к посеву выполнена с соблюдением требований агротехники, 0,75. Найти вероятность того, что из 100 делянок почва подготовлена к посеву не меньше чем на 70 и не больше чем на 80.

Решение

По условию, p = 0,75; q = 1 – 0,75 = 0,25; n = 100; $K_1 = 70$, $K_2 = 80$.
$P_{100}(70,80) = \Phi(x_2) - \Phi(x_2)$.

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{70 - 0,75 \cdot 100}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = \frac{-5}{4,33} = -1,15;$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{80 - 0,75 \cdot 100}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = \frac{5}{4,33} = 1,15.$$

Таким образом, имеем $P_{100}(70,80) = \Phi(+1,15) - \Phi(-1,15) = \Phi(1,15) + \Phi(1,15) = 2\Phi(1,15)$.
По таблице находим $\Phi(1,15) = 0б3749$.
Искомая вероятность $P_{100}(70,80) = 2 \cdot 0,3749 = 0,7498$.

5.2. Случайные величины

Случайные величина – это величина, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений (какое именно – заранее неизвестно).

Дискретная случайная величина – это случайная величина, которая принимает отдельное изолированное, счетное множество значений.

Непрерывная случайная величина – это случайная величина, принимающая любые значения из некоторого интервала конечного или бесконечного. Понятие непрерывной случайной величины возникает при измерениях.

Случайные величины обозначаются конечными заглавными буквами латинского алфавита X, Y, Z, а их значения – соответствующими строчными буквами x, y, z.

5. 2. 1. Закон распределения случайной величины

Это всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Для *дискретной* случайной величины закон распределения может быть задан в виде *таблицы*, аналитически (в виде формулы) и графически.

Таблица – это простейшая форма задания закона распределения. В ней перечислены в порядке возрастания все возможные значения случайной величины X и соответствующие вероятности.

X_1	X_2	…	X_n
P_1	P_2	…	P_n

$$\sum_{i=1}^{n} p_i = 1$$

Ряд распределения может быть изображен графически, если по оси абсцисс откладывать значения случайной величины, а по оси ординат – соответствующие им вероятности. Соединение образуют ломаную линию. Это многоугольник или полигон распределения вероятностей.

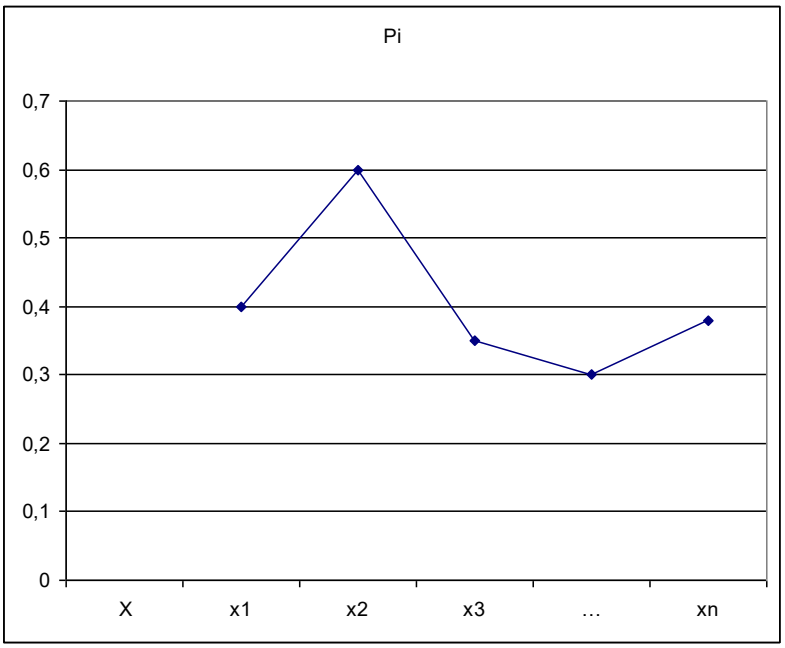


Рис. 5.1. *Полигон распределения вероятностей*

Задача 5.22

Задают ли законы распределения дискретной случайной величины следующие таблицы?

А

X_l	2	3	4	5
P_l	0,1	0,4	0,3	0,2

Б

X_l	6	7	8	9
P_l	0,1	0,2	0,3	0,5

А. Да, так как выполняется условие $\sum p = 1$: 0,1+0,4+0,3+0,2=1

Б. Нет: 0,1+0,2+0,3+0.5≠1

Задача 5.25

Вероятность того, что студент сдаст экзамен по физике, равна 0,7, а по химии – 0,9. Составить закон распределения числа семестровых экзаменов, которые сдаст студент. Построить многоугольник распределения вероятностей.

Решение

Возможные значения X- число сданных экзаменов: 0, 1, 2.

Считаем вероятности:

$P(X=0) = P(\bar{A}_1) \cdot P(\bar{A}_2) = (1 - 0,7) \cdot (1 – 0,9) = 0.3 \cdot 0,1 = 0,03;$

$P(X=1) = P(A_1 \bar{A}_2 + \bar{A}_1 A_2) = P(A_1) P(\bar{A}_2) + P(\bar{A}_1) P(A_2) = 0,7 \cdot 0.1 + 0,3 \cdot 0,9 = 0,34.$

$P(X=2) = P(A_1 A_2) = P(A_1) \cdot P(A_2) = 0,7 \cdot 0,9 = 0,63.$

Ряд распределения имеет вид:

X_l	0	1	2
P_l	0,03	0,34	0,63

Контроль: 0,03+0,34+0,63 = 1.

5.2.2. Функция распределения случайных величин

Функция F(x), выражающая для каждого x вероятность того, что случайная величина X примет значение меньше некоторого фиксированного x, называется *функцией распределения* случайной величины X: $F(x)= P(X<x)$. Ее также называют *интегральной функцией* распределения дискретных и непрерывных случайных величин.

Задача 5.26

Дан ряд распределения случайной величины:

X_l	1	4	5	7
P_l	0,4	0,2	0,2	0,2

Найти и изобразить график ее функции распределения.

Решение

1. Если $x_1 < 1$, $F(x)=0$.
2. Пусть $1<x\leq 4$, (например $x = 2$), $F(x)=P(x=1)=0,4$.

3.Пусть $4<x\leq 5$, (например $x=4,25$),

$F(x)=P(X<x)=P(x=4)=0,4+0,2=0,6$.

4. Пусть $5<x\leq 7$, $F(x)=(P(x=1)+P(x=4)+P(x=5)=0,6+0,2=0,8$.
5. Пусть $x>7$, $F(x)=(P(x=1)+P(x=4)+P(x=5)+P(x=7)=0,8+0,2=1$.

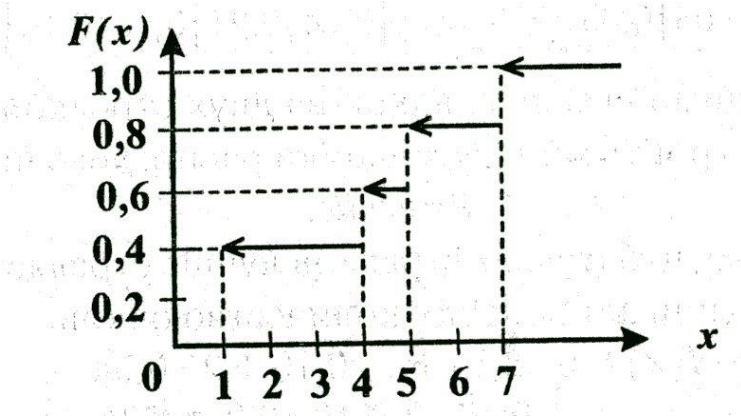


Рис. 5.2. *Функция распределения дискретной случайной величины*

Функция распределения любой дискретной случайной величины есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений.

5.2.3. Числовые характеристики дискретной случайной величины

1. *Математическим ожиданием* M(X) дискретной случайной величины X называется сумма произведений всех ее значений на соответствующие им вероятности:

$$M(X) = \sum_{i-1}^{n} x_i \; p_i$$

Задача 5.27

Известны значения распределения случайных величин X и Y – число очков, выбиваемых первым и вторым стрелками.

X_l	0	1	2	3	4	5	6	7	8	9	10
P_l	0,15	0,11	0,04	0,05	0,04	0,10	0,10	0,04	0,05	0,12	0,20

X_l	0	1	2	3	4	5	6	7	8	9	10
P_l	0,01	0,03	0,05	0,09	0,11	0,24	0,21	0,10	0,10	0,04	0,02

Необходимо выявить, какой из двух стрелков стреляет лучше. Построить многоугольники распределения.

Решение

Очевидно, что из двух стрелкой лучше стреляет тот, кто в среднем выбивает большее количество очков.

M(X) = 0 • 0,15 + 1 • 0,10 + 2 • 0,04 +….

....+ 9 • 0,12 + 10 • 0,2 = 5,36.
M(Y) = 0 • 0,01 + 1 • 0,03 • 0,05 +....
....+ 9 • 0,04 + 10 • 0,02 = 5,36.

То есть среднее число выбиваемых очков у двух стрелков одинаково.

2. *Дисперсия* дискретной случайной величины.

Слово «дисперсия» означает *«рассеяние»:*

$D(X) = M(X - M(X))^2$.

Дисперсией D(X) случайной величины X называется математическое ожидание квадрата ее отклонения от математического ожидания.

$$D(X) = \sum_{i-1}^{n} (x_i - M(X))^2 \; p_i$$

3. *Среднее квадратическое отклонение* σ (стандартное отклонение или стандарт) случайной величины X – это арифметическое значение корня квадратного из ее дисперсии:

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

Задача 5.28

В задаче 5.23 вычислить дисперсию и среднее квадратическое отклонение.

Решение

$$M(X) = 0 \cdot 0{,}03 + 1 \cdot 0{,}34 + 2 \cdot 0{,}63 = 1{,}6;$$

$$D(x) = (0-1{,}6)^2 \cdot 0{,}03 + (1-1{,}6)^2 \cdot 0{,}34 + (2-1{,}6)^2 \cdot 0{,}63 = 0{,}3;$$

$$\sigma = \sqrt{0{,}3} = 0{,}548.$$

5.2.4. Плотность вероятности непрерывных случайных величин

Плотностью вероятности, или *плотностью распределения* $f(x)$ *непрерывной* случайной величины X, называется производная ее функции распределения: $f(x) = F'(x)$.

Ее также называют *дифференциальной* функцией распределения.

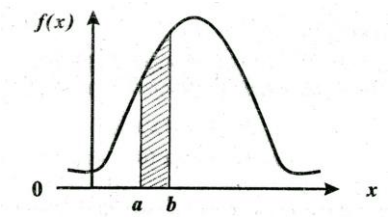


Рис. 5.3. *Плотность распределения*

Свойство плотности вероятности:

1. Неотрицательная функция $f(x) \geq 0$.

2. Площадь фигуры, ограниченной кривой распределения и осью абсцисс, равна 1.

3. Вероятность попадания непрерывной случайной величины в интервал [a, b] равна определенному интервалу от ее плотности в пределах от a до b.

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

Геометрическая интерпретация: полученная вероятность равна площади фигуры, ограниченной сверху кривой распределения и опирающейся на отрезок [a, b].

Непрерывная случайная величина описывается следующими характеристиками:

1. Математическое ожидание:

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

2. Дисперсия: $$D(x) = \int_{-\infty}^{\infty} (x - M(X))^2 \cdot f(x)\, dx.$$

Или $$D(x) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - (M(X))^2.$$

Задача 5.29

Найти математическое ожидание и дисперсию случайной величины X, если плотность распределения:

$$f(x) = \begin{cases} 0, & \text{при } x \le 0 \\ 1, & \text{при } 0 < x < 1 \\ 0, & \text{при } x > 1 \end{cases}$$

Решение

$$M(X) = \int 1x dx = \frac{x^2}{2}\Big|_0^1 = 0{,}5$$

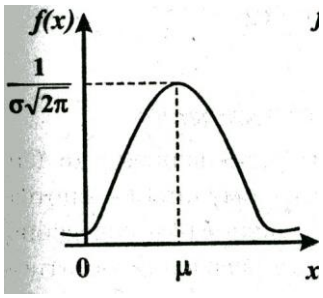
$$D(X) = \int_0^1 x^2 1 dx - (0{,}5)^2 = \frac{x^3}{3}\Big|_0^1 - \frac{1}{4} = 1/12.$$

5.2.5 Нормальный закон распределения

Этот закон наиболее часто встречается на практике. Он является предельным законом, к которому приближаются другие законы распределения. Нормальное распределение является одним из самых важных распределений в статистике. Обычно все сравнивают с нормальным законом распределения.

Непрерывная случайная величина X имеет нормальный закон распределения (закон Гаусса) с параметрами, а и σ ², если ее плотность вероятности имеет вид:

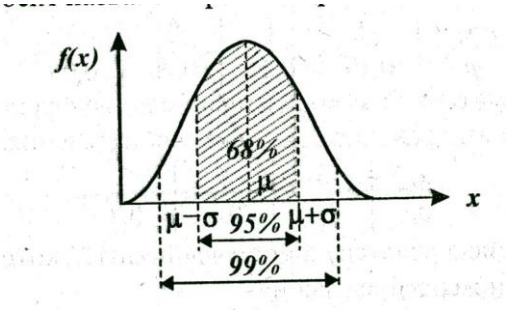
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\ell^{-\frac{(x-\grave{a})^2}{2\sigma^2}}$$



Вероятность того, что нормально распределенная случайная величина x со средним μ и средним квадратическим отклонением σ (стандартное отклонение) находится между (a–σ) и (a+σ), равна 0, 68, т.е. 68% случайной величины x отличается от среднего не более чем на одно стандартное отклонение ±σ.

Вероятность того, что нормально распределенная случайная величина x со средним a и средним квадратическим отклонением σ (стандартное отклонение) находится между (a–2 σ) и (a+ 2 σ) равна 0,95,т.е. 95% случайной величины x отличается от среднего не более чем на одно стандартное отклонение ± 2 σ.

Вероятность того, что нормально распределенная случайная величина x со средним a и средним квадратическим отклонением σ (стандартное отклонение) находится между (a–3 σ) и (a+ σ 3), равна 0, 99, т.е. 99% (практически достоверно). Это свойство носит название *правило трех сигм* (рис. 5.4).



5.4. *Правило трех сигм.*

Для случайной величины X, распределенной по нормальному закону, вероятность попадания ее значений в интервал (α, β) вычисляется по формуле:

$$P(\alpha<X<\beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$,где Φ(x) – функция Лапласа.

Вероятность отклонения нормально распределенной случайной величины от ее математического ожидания менее чем на δ равна:

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$$

Задача 5.30

Вероятность того, что подготовка почвы к посеву выполнена с соблюдением требований агротехники, 0,75. Найти вероятность того, что из 100 делянок почва подготовлена к посеву не меньше чем на 70 и не больше чем на 80.

Решение

По условию, p = 0,75; q = 1 – 0,75 = 0,25; n = 100; K_1 = 70, K_2 =80.

$P_{100}(70,80) = \Phi(x_2) - \Phi(x_2)$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{70 - 0,75 \cdot 100}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = \frac{-5}{4,33} = -1,15;$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{80 - 0,75 \cdot 100}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = \frac{5}{4,33} = 1,15.$$

Таким образом, имеем $P_{100}(70,80) = \Phi(+1,15) - \Phi(-1,15) = \Phi(1,15) + \Phi(1,15) = 2\Phi(1,15)$.

По таблице находим Φ(1,15) = 0б3749.

Искомая вероятность $P_{100}(70,80) = 2 \cdot 0,3749 = 0,7498$.

Задачи для самостоятельного решения

41. Случайная величина X задана законом распределения:

X_l	2	3	10
P_l	0,1	0,4	0,5

Найти математическое ожидание и среднее квадратическое отклонение σ(X). Построить многоугольник распределения.

42. Найти дисперсию случайной величины X, зная закон ее распределения. Построить многоугольник распределения.

X_l	0,1	2	10	20
P_l	0,4	0,2	0,15	0,25

43. Найти дисперсию случайной величины X, зная закон ее распределения. Построить многоугольник распределения.

X_l	-1	1	2	3
P_l	0,48	0,01	0,09	0,42

44. Дискретная случайная величина X имеет закон распределения.

X_l	0,2	0,4	0,6	0,8	1
P_l	0,1	0,2	0,4	p	0,1

Чему равна вероятность p (X=0,8)? Построить многоугольник распределения. Найти математическое ожидание и дисперсию.

45. Составить дифференциальную функцию для нормально распределенной случайной величины и построить её график, если даны её параметры:
1) M(X)=4, σ=0,2; 2) M(X)=-0,5, σ=2; 3) M(X)=3, σ=0,25; 4) M(X)=0, σ=1.

46. Заданы математическое ожидание m и среднее квадратическое σ нормально распределённой случайной величины x. Найти: 1) вероятность того, что x примет значение, принадлежащее интервалу (α,β); 2) вероятность того, что абсолютная величина отклонения |x-a| окажется меньше δ.

1) a=15, σ=2, α=16, β=25, δ=4;
2) a=14, σ=4, α=18, β=34, δ=8;
3) a=13, σ=4, α=15, β=17, δ=6.

Тема 6.СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ

6.1. Основные понятия математической статистики

Математическая статистика - это раздел математики, изучающий приближенные методы отыскания законов распределения и числовых характеристик по результатам эксперимента.

В математической статистике принято выделять два основных направления исследований:

1. Оценка параметров генеральной совокупности.
2. Проверка статистических гипотез (некоторых априорных предположений).

Основными понятиями математической статистики являются: генеральная совокупность, выборка, теоретическая функция распределения.

Генеральная совокупность – это множество всех мыслимых значений наблюдений (объектов), однородных относительно некоторого признака, которые могли быть сделаны. Число всех наблюдений, составляющих генеральную совокупность, называется ее объемом N. Например, популяция представляет собой множество индивидуумов. Изучение целой популяции трудоемко и дорого и, может быть, просто невозможно. Поэтому собирают данные по выборке индивидуумов, которых считают представителями этой популяции, позволяющими сделать вывод относительно этой популяции.

Выборка - это совокупность случайно отобранных наблюдений (объектов) для непосредственного изучения из генеральной совокупности. Объем выборки n. Выборка обязательно должна удовлетворять условию репрезентативности, т. е. давать обоснованное представление о генеральной совокупности. Как сформировать репрезентативную (представительную) выборку? В идеале стремятся получить случайную (рандомизированную) выборку. Для этого составляют список всех индивидуумов в популяции и случайно их отбирают. Но иной раз затраты при составлении списка могут оказаться недопустимыми, и тогда берут приемлемую выборку, например, одну клинику, больницу, и исследуют всех пациентов в этой клинике с данным заболеванием.

Каждый элемент выборки x_i называется *вариантой.* Число наблюдений варианты n_i называется *частотой встречаемости.* Последовательность вариант, записанных в возрастающем порядке, называется *вариационным рядом.*

Статистическое распределение - это совокупность вариант x_i и соответствующих им частот n_i .

Пример 6.1

Задано распределение частот выборки объема 20.

x_i	2	6	12
n_i	3	10	7

Написать распределение относительных частот.

Решение

Найдем относительные частоты. Для этого разделим частоты на объем выборки:

$$\frac{n_1}{n}=\frac{3}{20}=0{,}15 \qquad \frac{n_2}{n}=\frac{10}{20}=0{,}5 \qquad \frac{n_3}{n}=\frac{7}{20}=0{,}35$$

Распределение относительных частот имеет вид:

x_i	2	6	12
ω_i	0,15	0,5	0,35

Контроль: 0,15 + 0,5 + 0,35 = 1.

Для наглядного представления статистического распределения пользуются графическим изображением вариационных рядов: полигоном и гистограммой.

Гистограмма частот – это ступенчатая фигура, состоящая из смежных прямоугольников, построенных на одной прямой, основания которых одинаковы и равны ширине класса, а высота равна или частоте попадания в интервал n_i или относительной частоте $\frac{n_i}{n}$. Ширину интервала i можно определить по формуле Стерджеса:

$$i=\frac{x_{\max}-x_{\min}}{1+3{,}32\lg n}$$

где x_{max} – максимальное; x_{min} – минимальное значения вариант, а их разность - *вариантный размах*; n - объем выборки.

Полигон частот – ломаная линия, отрезки которой соединяют точки с координатами (x_i, n_i).

Пример 6. 2

Построить дискретный вариационный ряд и начертить полигон распределения 45 абитуриентов по числу баллов, полученных ими на приемных экзаменах:

39 41 40 42 41 4042 44 40 43 42 41 43 39 42 41 42 39 41 37 43 41 38 43 42 41 40 41 38 44 40 39 41 40 42 40 41 42 40 43 38 39 41 41 42.

Решение

Для построения вариационного ряда различные значения признака X располагаем в порядке их возрастания и под каждым из этих значений записываем его частоту.

x_i	37	38	39	40	41	42	43	44
n_i	1	3	5	8	12	9	5	27

Построим полигон этого распределения

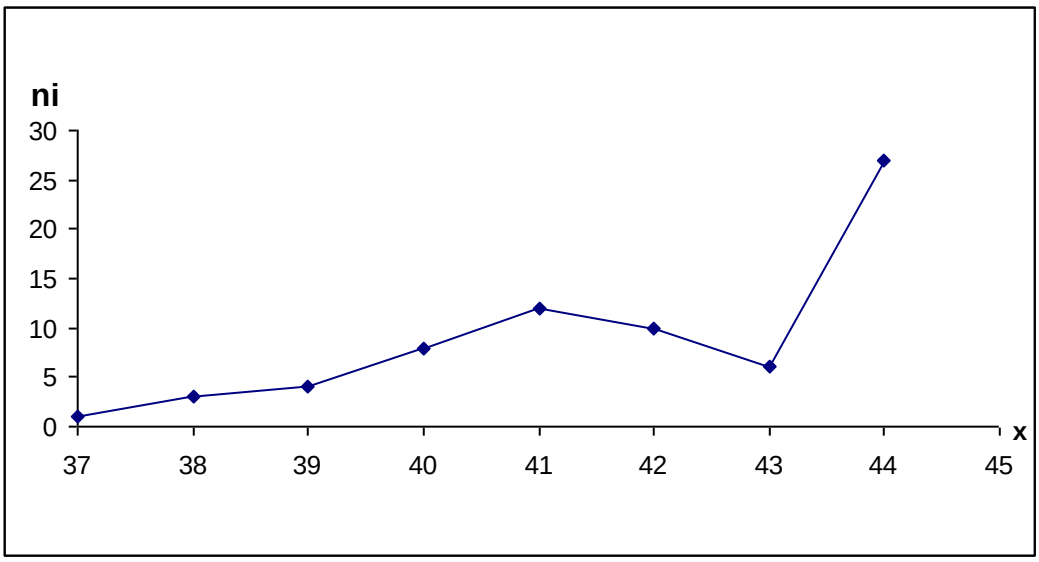


Рис. 6.1. *Полигон частот*

Пример 6. 3

Наблюдения за числом частиц, попавших в счетчик Гейгера, в течение минуты дали следующие результаты:

21 30 39 31 42 34 36 30 28 30 33 24 31 27 31 45 31 34 27 30 48 30 28 30 33 46 43 30 33 28 31 27 31 36 51 34 31 36 34 37 28 30 39 31 42 37 30 31 35 41.

Построить по этим данным интервальный вариационный ряд с равными интервалами (1-ий интервал 20 – 24: 2-ой интервал 24 – 28 и т. д.) и начертить гистограмму.

Решение

Интервал	20 - 24	24 - 28	28 - 32	32 - 36	36 - 40	40 - 44	44 - 48	48 - 52
Частота	1	4	22	8	7	4	2	2

Гистограмма этого распределения имеет вид

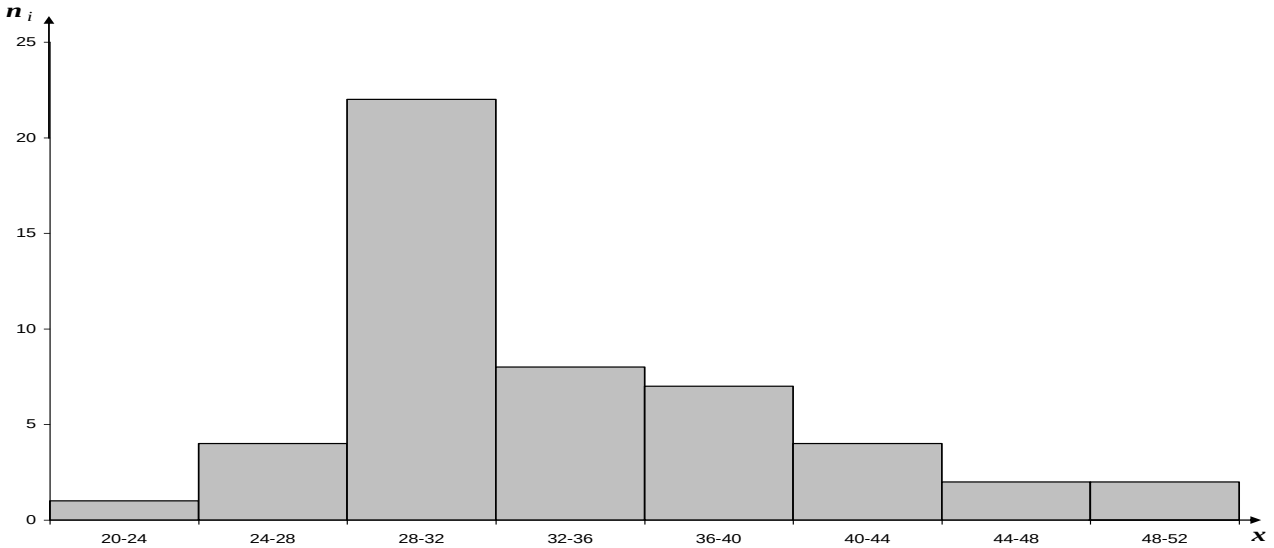


Рис. 6.2. *Гистограмма распределения.*

6.2 Статистические оценки параметров распределения. Выборочные характеристики.

6.2.1. *Характеристика положения*

Выборочная средняя – это среднее арифметическое значение вариант статического ряда

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$$

Мода (M_0) –это такое значение варианты, что предшествующее и следующее за ним значения имеют меньшие частоты встречаемости.

Для распределений дискретной случайной величины мода – это наиболее часто встречающаяся в данной совокупности варианта.

Например, мода распределения

χ_i	16	17	18	20
n_i	5	1	20	6

равна 18.

Медиана M_e – это значение признака, относительно которого ряд распределения делится на 2 равные по объему части.

Например, в распределении

12 14 16 18 20 22 24 26 28

медианой будет центральная варианта, т. е. $M_e = 20$, так как по обе стороны от нее отстоит по 4 варианты.

Для ряда с четным числом членов 6 8 10 12 14 16 18 20 22 24 медианой будет

полусумма его центральных членов, т.е. $M_e = \frac{14+16}{2} = 15$

Пример 6. 4

В выборке взрослых мужчин n = 50 определяли содержание гемоглобина в крови. У n_1 = 30 оно оказалось равным в среднем 70%. Для другой группы мужчин n_2 =20 этот показатель составил 50%. Найти среднюю арифметическую из этих двух средних.

Решение

По формуле:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$$

$$\bar{x} = \frac{1}{50}\left(30 \cdot 70 + 20 \cdot 50\right) = 62\%$$

6.2.2. Характеристика рассеяния вариант вокруг своего среднего

Выборочная дисперсия – среднее арифметическое квадратов отклонения вариант от их среднего значения:

$$D_{â} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x_{â}}\right)^2 \cdot n_i$$

Среднее квадратическое отклонение – это квадратный корень из выборочной дисперсии:

$$\sigma_{â} = \sqrt{D_{â}}$$

Коэффициент вариации CV – это отношение среднего квадратического отклонения к средней величине признака, выраженное в процентах:

$$CV = \frac{S_{â}}{\overline{x_{â}}} \cdot 100\%$$

Коэффициент вариации – это мера относительной изменчивости случайной величины, которая позволяет сравнивать разнородные величины, например, частоту сердечных сокращений (ЧСС уд/мин), артериальное давление (АД, мм рт. ст.) и температуру (t^o, ^{o}C) в единых – процентах.

Вариационный размах $\Delta = x_{\max} - x_{\min}$ - это разность между наибольшим и наименьшим значениями признака.

Пример 6. 5

Выборочная совокупность задана таблицей распределения:

x_i	1	2	3	4
n_i	20	15	10	5

Найти выборочную дисперсию.

Решение

Найдем выборочную среднюю:

$$\overline{x_{â}} = \frac{20\cdot 1 + 15\cdot 2 + 10\cdot 3 + 5\cdot 4}{20+15+10+5} = 2$$

Найдем выборочную дисперсию:

$$D_{â} = \frac{20(1-2)^2 + 15(2-2)^2 + 10(3-2)^2 + 5(4-2)^2}{20+15+10+5} = 1$$

Пример 6. 6

Сравнить 2 варьирующихся признака. Один характеризуется средней $\overline{x_1} = 2{,}4\,кг$ и средним квадратическим отклонением S_1 = 0,58 *кг*, другой – величинами $\overline{x_2} = 8{,}3\,см$ и S_2 =1,57 *см.* Какай признак варьируется сильнее?

Решение

$$CV_1 = \frac{S_1}{\overline{x_1}} \cdot 100\%$$

$$CV_1 = \frac{0{,}58}{2{,}4} \cdot 100\% = 24{,}2\% \qquad CV_2 = \frac{1{,}57}{8{,}3} \cdot 100\% = 18{,}9\%$$

Ответ: первый, так как $CV_1 > CV_2$.

Задания для самостоятельного решения.

1. Найти выборочную среднюю и выборочную дисперсию по данному распределению:

x_i	1	4	8
n_i	5	3	2

2. Найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение, если совокупность задана таблицей распределения.

x_i	2	4	5	6
n_i	8	9	10	3

3. Найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение, если совокупность задана таблицей распределения.

x_i	1	2	3	4
n_i	20	15	10	5

6.3. Оценка параметров генеральной совокупности по ее выборке

Смысл статистических методов заключается в том, чтобы по выборке ограниченного объема n, то есть по некоторой части генеральной совокупности, высказать обоснованное суждение о ее свойствах в целом.

Числовые значения, характеризующие генеральную совокупность, называются *параметрами*. Одна из задач математической статистики – определение параметров большого массива по исследованию его части.

Статистическое оценивание может выполняться двумя способами:

1) *точечная оценка* – оценка, которая дается для некоторой определенной точки;

2) *интервальная оценка* – по данным выборки оценивается интервал, в котором лежит истинное значение с заданной вероятностью.

6.3.1. Точечная оценка параметров генеральной совокупности

Точечная оценка – это оценка, которая определяется одним числом. И это число определяется по выборке. Это функция результатов выборки, и она является точечной оценкой генерального параметра, т. е. принимает только одно значение.

Качество оценки устанавливается по трем свойствам: быть состоятельной, эффективной и несмещенной.

Точечная оценка называется *состоятельной*, если при увеличении объема выборки выборочная характеристика стремится к соответствующей характеристике генеральной совокупности.

Точечная оценка называется *эффективной*, если она имеет наименьшую дисперсию выборочного распределения по сравнению с другими аналогичными оценками.

Точечную оценку называют *несмещенной*, если ее математическое ожидание равно оцениваемому параметру при любом объеме выборки.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя $\overline{x}_в$:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i ,$$

где x_i – варианты выборки; n_i – частота встречаемости варианта x_i; n – объем выборки.

Выборочная средняя является несмещенной оценкой генеральной средней, так как $M(\overline{x}_в) = \overline{x}_{ген}$, т. е. она эквивалентна истинной средней в генеральной совокупности (популяции).

Выборочная дисперсия $S_в^2$ не обладает свойством несмещенности. Это смещенная оценка генеральной дисперсии $\sigma_{ген}^2$.

$M(D_в) = \frac{n-1}{n}\sigma_{ãåí}^2 \neq \sigma_{ãåí}^2$ – это и означает, что выборочная дисперсия $S_в^2$ является смещенной оценкой $\sigma_{ген}^2$.

На практике используют исправленную выборочную дисперсию S^2, которая является несмещенной оценкой дисперсии генеральной совокупности:

$$S^2 = \frac{n}{n-1} \cdot D_B$$

$$S^2 = \frac{1}{n-1} \sum_{i-1}^{k} \left(x_i - \bar{x}_B \right)^2 \cdot n_i$$

Кроме того, в расчетах используют S – исправленное среднее квадратическое отклонение, называемое *стандартным отклонением* в выборке, и ошибку выборочной средней (стандартную ошибку средней) $m_{\bar{x}}$:

$$m_{\bar{x}} = \frac{s}{\sqrt{n}} \ ,$$

которая отражает точность оценки.

Стандартная ошибка уменьшается, т. е. оценка станет более точной, если объем выборки n увеличится и данные имеют небольшое рассеяние S .

Рассмотрим разницу между S – стандартным отклонением в выборке и $m_{\bar{x}}$ – стандартной ошибкой среднего. На первый взгляд, они очень схожи, но их используют в разных целях. Среднее квадратическое отклонение S отражает вариабельность в значениях данных, и его указывают, если надо пояснить изменчивость в наборе данных, разброс данных.

Ошибка выборочной средней $m_{\bar{x}}$ характеризует точность выборочного среднего $\bar{x}_B$ и должна быть указана, если интерес представляет среднее значение выборки.

Пример 6.7

Из генеральной совокупности извлечена выборка объема $n = 50$.

x_i	2	5	10	7
n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

Решение

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^{k} x_i n_i$$

$$\bar{x}_B = \frac{16 \cdot 2 + 5 \cdot 12 + 7 \cdot 8 + 10 \cdot 14}{50} = 5{,}76$$

Пример 6.8

По выборке объема 30 найдена смещенная оценка $S_B^2 = 3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

Решение:

Эта несмещенная оценка равна исправленной дисперсии:

$$S^2 = \frac{n}{n-1} \cdot D_B, \qquad S^2 = \frac{30}{29} \cdot 3 = 3{,}1.$$

Задачи для самостоятельной работы

4. При исследовании клинической оценки тяжести серповидно-клеточной анемии была получена выборка объема 33.
0;0;0;1;1;1;1;1;1;1;1;1;1;1;1;2;2;2;2;3;3;3;3;4;4;5;5;5;5;6;7;9;10;11.
Найдите среднюю, среднее квадратическое отклонение и медиану. Можно ли считать, что выборка извлечена из совокупности с нормальным распределением?

5. Исследуя продолжительность (в секундах) физической нагрузки до развития приступа стенокардии у 12 человек с ишемической болезнью сердца, получили следующие данные:
289;203;359;243;232;210;215;246;224;239;220;211.
Найдите среднюю, среднее квадратическое отклонение, медиану. Можно ли считать, что данная выборка извлечена из совокупности с нормальным распределением?

6. Найдите среднее число очков, выпадающих при бросании игральной кости. Опишите это распределение. Может ли оно быть нормальным?

7. Группа из 50 коров обследована по числу отелов. Получены следующие данные (число отелов):

7	6	1	2	8	7	5	3	5	4
1	1	10	6	4	5	5	3	2	2
2	2	3	5	5	4	6	9	1	1
4	5	3	5	7	8	2	1	6	7
1	2	3	4	4	5	6	7	7	8

Составьте интервальное распределение с 5-8 интервалами. Найдите основные выборочные характеристики: $\bar{x}$, s^2, s, V, $s_{\bar{x}}$; с надежностью 95% указать доверительный интервал для оценки генеральной средней $\overline{x_Г}$.

6.3.2. *Интервальная оценка параметров генеральной совокупности*

Точечные оценки параметров распределения не дают информации о степени близости к соответствующему теоретическому параметру. Поэтому построение интервала, в котором с заданной степенью достоверности будет находиться оцениваемый параметр, является более информативным способом оценивания неизвестных параметров.

Интервальная оценка − это числовой интервал, который определяется двумя числами – границами интервала, содержащий неизвестный параметр генеральной совокупности.

Доверительный интервал – это интервал, в котором с той или иной заранее заданной вероятностью находится неизвестный параметр генеральной совокупности.

Доверительна вероятность p– это такая вероятность, что событие вероятности (1-p)можно считать невозможным $\alpha = 1 - P$ – это уровень значимости. (Обозначения могут быть любыми, часто обозначают наоборот). Обычно в качестве доверительных вероятностей используют вероятности, близкие к 1. Тогда событие, что интервал накроет характеристику, будет практически достоверным. Это p≥ 0,95, p≥ 0,99, p≥ 0,999.

Эти вероятности признаны достаточными для уверенного суждения о генеральных параметрах на основании известных выборочных показателей. Обычно указывают 95% - й доверительный интервал.

Для выборки малого объема $(n < 30)$ нормально распределенного количественного признака x доверительный интервал может иметь вид:

$$\bar{x}_B - m_{\bar{x}} t \le \mu \le \bar{x}_B + m_{\bar{x}} t(P \ge 0{,}95),$$

где μ – генеральное среднее; $\bar{x}_B$ – выборочное среднее; t – нормированный показатель распределения Стьюдента с $(n-1)$ степенями свободы, который определяется вероятностью попадания генерального параметра в данный интервал. Термин «степени свободы» означает, что их можно вычислить как объем выборки минус число ограничивающих условий; $m_{\bar{x}}$ – ошибка выборочной средней.

Для интерпретации доверительного интервала следует помнить, что ширина доверительного интервала зависит от $m_{\bar{x}}$ – ошибки выборочной средней, которая в свою очередь зависит от объема выборки (n) и от изменчивости

данных (S).Если выборка небольшая, то доверительный интервал более широкий, чем в случае выборки большого объема. Широкий доверительный интервал указывает на неточную оценку, а узкий – на точную оценку.

Пример 6.9

Количественный признак x генеральной совокупности распределен нормально. По выборке объема $n=16$ найдены выборочная средняя $\bar{x}_B=20,2$ и среднее квадратическое отклонение $S=0,8$. Определить неизвестное математическое ожидание при помощи доверительного интервала при $\rho \geq 0,95$

Решение

$$\bar{x}_B - \frac{s}{\sqrt{n}}t \leq \mu \leq \bar{x}_B + \frac{s}{\sqrt{n}}t$$

Найдем t из таблицы распределения Стьюдента при уровне значимости $\alpha \leq 0,05$ и числе степеней свободы $f=n-1$; $f=16-1=15$

$t(\alpha \leq 0,05,\ f=15)=2,13$.

Запишем:

$20,2-\frac{0,8}{\sqrt{16}}\cdot 2,13 \leq \mu \leq 20,2+\frac{0,8}{\sqrt{16}}\cdot 2,13 (\rho \geq 0,95)$;

$19,8 \leq \mu \leq 20,6$ при $\rho \geq 0,95$.

Пример 6.10

Имеется выборка объема $n=11$ – это значения систолического давления у мужчин в начальной стадии шока.

x: 127,124,155,129,77,147,65,109,145,141.

С помощью пакета прикладных программ на ЭВМ провести статистическую обработку данных выборки и определить доверительный интервал для генеральной средней при $\rho \geq 0,95$.

Решение:

Пусть расчет на ЭВМ дал: выборочное среднее $\bar{x}_B=122,01$; $m_{\bar{x}}=8,59$.

По таблице распределения Стьюдента найдем:

$t(\alpha \leq 0,05,\ f=11-1=10)=2,23$.

$\mu=\bar{x}_B \pm mt$

$$\mu = 122{,}01 \pm 8{,}59 \cdot 2{,}2 \qquad (P \geq 0{,}95)$$

$$\mu = 122 \pm 19 \qquad (P \geq 0{,}95)$$

Задачи для самостоятельной работы

7. При исследовании частоты дыхания по выборке объема $n = 15$ были получены выборочная средняя $\bar{x}_B = 18{,}5$ и среднее квадратическое отклонение $S = 0{,}6$. Определить интервальную оценку математического ожидания с вероятностью $\rho \geq 0{,}95$.

8. Найдите доверительный интервал для оценки с уровнем доверительной вероятности $\rho \geq 0{,}95$ неизвестного математического ожидания нормального распределения признака x – диаметра эритроцита – генеральной совокупности, если выборочная средняя $\bar{x}_B = 10{,}2$ *мкм,* исправленное выборочное среднее квадратическое отклонение $S = 4$ и объем выборки $n = 16$.

9. Даны результаты измерения длины туловища 40 свиноматок (см):

157	145	159	165	150	154	171	165
163	169	163	168	164	163	150	145
158	168	158	150	143	162	148	147
163	157	157	158	159	164	165	172
157	157	150	165	160	154	158	190

Составить интервальное распределение. Число частичных интервалов 5-9. По данным задачи найти: 1) выборочную среднюю, 2) выборочную дисперсию, 3) среднее квадратическое отклонение, 4) моду, 5) медиану, 6) коэффициент вариации, 7) оценку ошибки выборки.

10. В результате взвешивания отобранных наудачу 50 клубней картофеля получены следующие результаты:

93	209	135	216	206	80	197	134	145	183
251	53	142	120	177	159	111	185	200	191
96	206	138	213	209	77	200	131	148	180
253	50	145	117	180	156	113	181	203	188
152	150	110	118	140	81	120	135	220	144

Составить интервальное распределение. Число частичных интервалов 5-9. По данным задачи найти: 1) выборочную среднюю, 2) выборочную дисперсию, 3)

среднее квадратичное отклонение, 4) моду, 5) медиану, 6) коэффициент вариации, 7) оценку ошибки выборки.

11. Группа из 50 коров обследована по числу отелов. Получены следующие данные (число отелов):

7	6	1	2	8	7	5	3	5	4
1	1	10	6	4	5	5	3	2	2
2	2	3	5	5	4	6	9	1	1
4	5	3	5	7	8	2	1	6	7
1	2	3	4	4	5	6	7	7	8

Составить интервальное распределение. Число частичных интервалов 5-9. По данным задачи найти: 1) выборочную среднюю, 2) выборочную дисперсию, 3) среднее квадратичное отклонение, 4) моду, 5) медиану, 6) коэффициент вариации, 7) оценку ошибки выборки.

12. При уровне вероятности $\gamma = 0{,}95$ требуется установить доверительный интервал среднего значения содержания белка в зернах пшеницы. На основе 100 проб установлено, что выборочная средняя $\bar{x} = 16\%$ и $\sigma = 3{,}5$.

Тема 7.СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

7.1.Терминология

Статистические методы используют для описания данных и для оценки статистической значимости результатов опыта. Сравнивают данные опыта с контролем и т.п. Методы оценки статистической значимости различий называют *критериями*. *Критерий* - это и сам метод и та величина, которая получается в результате его применения. Методов существует множество, но все они построены по одному принципу.

Итак, обратимся ко второму направлению математической статистики – проверке статистических гипотез.

Статистическая гипотеза – это любое предположение о виде неизвестного распределения или о параметрах известных распределений. Статистическая гипотеза – это всякое высказывание о генеральной совокупности, проверяемое по выборке.

Гипотезы будем обозначать буквой *H* с индексами. Будем предполагать, что у нас имеется 2 непересекающиеся гипотезы: H_0 - нулевая гипотеза (или основная); H_1 - альтернативная, или конкурирующая .

Сначала формулируют нулевую гипотезу, то есть предполагают, что исследуемые факторы не оказывают никакого влияния на исследуемую

величину и полученные различия случайны. Нулевая гипотеза (H_0) всегда отвергает эффект.

Затем формулируют альтернативную гипотезу (H_1), которая принимается, если нулевая гипотеза неверна.

Задача проверки статистических гипотез состоит в том, чтобы на основе выборки $x_1, x_2, ..., x_n$ принять (то есть считать справедливой) либо нулевую гипотезу H_0, либо конкурирующую гипотезу H_1.

Для проверки принятой гипотезы используют *статистический критерий* – это правило, позволяющее, основываясь только на выборке $x_1, x_2, ..., x_n$, принять либо отвергнуть нулевую гипотезу H_0.

Значение критерия, полученное из выборки, связывают с уже известным распределением (теоретическим или табличным), которому оно подчиняется, чтобы определить достигнутый уровень значимости (α). Значение α - это максимально приемлемая вероятность, отвергающая нулевую гипотезу, если она верна, и тогда можно сказать, что результаты значимы на 5%-м уровне.

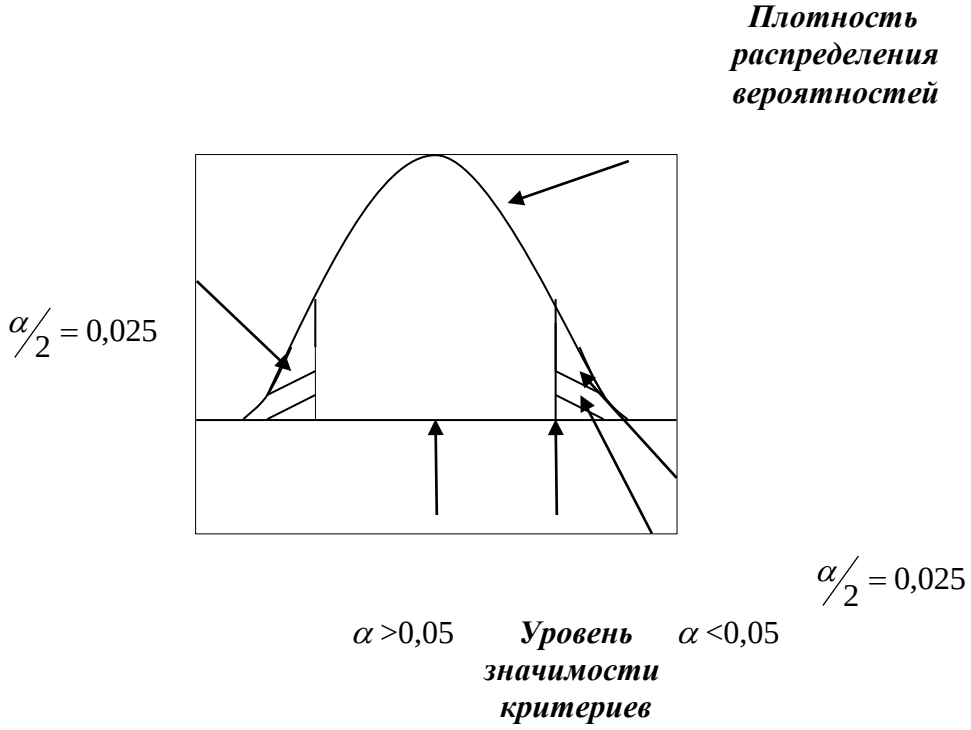


Рис 7.1. Достигнутый уровень значимости $\alpha \le 0{,}05$

Значение α - это площадь обоих «хвостов» на графике распределения вероятностей.

Когда отвергают нулевую гипотезу, то говорят, что результаты эксперимента статистически значимы на уровне $\alpha \le 0{,}05$ (5%).

Если α>0,05, то аргументов недостаточно, чтобы отвергнуть H_0. В этом случае говорят, что результаты статистически незначимы на уровне $\alpha \le 0{,}05$.

Это не означает, что нулевая гипотеза истинна. Просто недостаточно аргументов, чтобы ее отвергнуть.

Различают два вида критериев: параметрические и непараметрические.

Параметрические критерии представляют собой функции параметров данной совокупности и используются, если совокупности, из которых взяты выборки, подчиняются нормальному закону распределения.

Часто данные не подчиняются предложениям, которые лежат в основе этих методов. В этом случае можно использовать непараметрические критерии.

Непараметрические критерии применяются, если нет подчинения распределения нормальному закону. Эти критерии обычно заменяют данные выборок знаками (плюс или минус), рангами (то есть числами 1;2;3; …, описывающими их положение в упорядоченном наборе данных), категориями и т.п. Непараметрический критерий можно использовать, если объем выборки небольшой настолько, что невозможно оценить закон распределения данных. Но непараметрические критерии обладают меньшей мощностью в обнаружении реального эффекта, чем аналогичный параметрический критерий.

7.2.Общая постановка задачи проверки гипотез

1. Формулируют (выдвигают) нулевую гипотезу H_0 об отсутствии различий между группами, об отсутствии существенного отличия фактического распределения от некоторого заданного, например, нормального, экспоненциального и др.

Сущность нулевой гипотезы H_0: разница между сравниваемыми генеральными параметрами равна нулю, и различия, наблюдаемые между выборочными характеристиками, носят случайный характер, то есть эти выборки принадлежат одной генеральной совокупности.

2. Формулируют противоположную нулевой альтернативную гипотезу H_1.

3. Задают уровень значимости α. Уровень значимости α - это вероятность ошибки отвергнуть нулевую гипотезу H_0, если на самом деле эта гипотеза верна. При $\alpha \le 0{,}05$ ошибка возможна в 5% случаев.

4. Для проверки выдвинутой гипотезы используют критерии.

Критерий – это случайная величина K, которая служит для проверки H_0. Эти функции распределения известны и табулированы. Критерий зависит от двух параметров: от числа степеней свободы и от уровня значимости α. Фактическую величину критерия получают по данным наблюдения $K_{набл}$.

5. По таблице определяют критическое значение, превышение которого при справедливости гипотезы маловероятно $K_{крит}(\alpha, f)$.

6. Сравнивают $K_{набл}$ и $K_{крит}(\alpha, f)$.

Если $K_{набл} > K_{крит}(\alpha, f)$, то отвергают H_0 и принимают H_1.

Если $K_{набл} < K_{крит}(\alpha, f)$, то принимают H_0.

Это для параметрических критериев.

Если использованы непараметрические критерии, то наоборот: если $K_{набл} < K_{крит}(\alpha, f)$, то принимают H_0.

7. *Вывод*: различие статистически значимо ($\alpha \le 0{,}05$) или незначимо.

7.3.Проверка гипотез относительно средних

Предположим, что надо сравнивать состояние больных до и после лечения. Для этого сравнивают друг с другом две независимые выборки n_1 и n_2, взятые из нормально распределенных совокупностей с параметрами $M(X_1)$ и $M(X_2)$. Дополнительно предполагаем, что неизвестные генеральные дисперсии равны между собой. По этим выборкам найдены соответствующие выборочные средние $\bar{x}_1$ и $\bar{x}_2$ и исправленные дисперсии S_1^2 и S_2^2. Уровень значимости задан.

1. Нулевая гипотеза H_0: $M(X_1)=M(X_2)$.
2. Конкурирующая гипотеза H_1: $M(X_1)\neq M(X_2)$.
3. Для проверки нулевой гипотезы в этом случае можно использовать критерий Стьюдента сравнения средних.

Величину критерия находим по формуле

$$t_{набл} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1-1)S_{x_1}^2 + (n_2-1)S_{x_2}^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}}$$

Обычно расчет ведется на ЭВМ.

Доказано, что величина $t_{набл}$ при справедливости нулевой гипотезы имеет t – распределение Стьюдента с $f = n_1 + n_2 - 2$ степенями свободы.

4. По таблице находим $t_{крит}\ (\alpha, f = n_1 + n_2 - 2)$.
5. Сравниваем $t_{êðèò}$ и $t_{íàáë}$.

Если $|t_{набл}| < t_{крит}\ (\alpha, f) \Rightarrow H_0$.

Если $|t_{набл}| > t_{крит}\ (\alpha, f) \Rightarrow$ отвергается H_0 и принимается H_1, различие достоверно.

● Пример 7.1

По двум независимым малым выборкам объемов n_1=5 и n_2=6, извлеченным из нормальных генеральных совокупностей X_1 и X_2, найдены выборочные средние $\bar{x}_1$=33, $\bar{x}_2$=2,48. Известно, что генеральные дисперсии примерно равны, то есть $\sigma_{ô_1}^2 = \sigma_{ô_2}^2$. При уровне значимости $\alpha \le 0{,}05$ проверить нулевую гипотезу $H_0: M(X_1)=M(X_2)$, если $t_{набл}$=3,27.

Решение

$t_{êðèò}\ (\alpha \le 0{,}05,\ f = n_1 + n_2 - 2 = 5 + 6 - 2 = 9) = 2.26$.

$t_{íàáë} > t_{êðèò}\ (\alpha, f) \Rightarrow$ отвергаем H_0.

Вывод: генеральные средние различаются значимо ($\alpha \le 0{,}05$).

7.4.Проверка гипотез для дисперсий

Пусть генеральные совокупности X_1 и X_2 распределены нормально. По независимым выборкам объемов n_1 и n_2, извлеченным из этих совокупностей, найдены исправленные выборочные дисперсии $S^2_{x_1}$ и $S^2_{x_2}$. Требуется сравнить эти дисперсии. При заданном уровне значимости α надо проверить нулевую гипотезу о равенстве генеральных дисперсий нормальных совокупностей.

1. $H_0: \sigma^2_{x_1} = \sigma^2_{x_2}$.
2. $H_1: \sigma^2_{x_1} \neq \sigma^2_{x_2}$.
3. В качестве критерия проверки нулевой гипотезы о равенстве генеральных дисперсий используем случайную величину F, равную отношению большей исправленной выборочной дисперсии к меньшей $F_{набл} = \frac{S^2_б}{S^2_м}$.
4. Величина F, при условии справедливости нулевой гипотезы, имеет распределения Фишера – Снедекора со степенями свободы $f_1 = n_1 - 1$ и $f_2 = n_2 - 1$, где $ї_1$ - объем выборки, по которой вычислена большая выборочная дисперсия.

Из таблиц находим $F_{крит}(\alpha, f_1.f_2)$.

5. Сравнивают $F_{крит}$ и $F_{набл}$.

Если $F_{набл} < F_{крит}(\alpha, f_1, f_2) => H_0$, генеральные дисперсии различаются незначимо.

●Пример 7.2

По двум независимым выборкам объемов $ї_1$=12 и $ї_2$=15, извлеченным из нормальных генеральных совокупностей X_1 и X_2, найдены исправленные выборочные дисперсии $S^2_{x_1}$=11,41 и $S^2_{x_2}$=6,52. При уровне значимости $\alpha \le 0{,}05$ проверить нулевую гипотезу о равенстве генеральных дисперсий $H_0: \sigma^2_{x_1} = \sigma^2_{x_2}$.

Решение

Конкурирующая гипотеза: $H_1: \sigma^2_{x_1} > \sigma^2_{x_2}$;

$F_{набл} = \frac{S^2_б}{S^2_м}$; $F_{набл} = \frac{11{,}41}{6{,}52} = 1{,}75$;

$F_{крит}\ (\alpha \le 0{,}05,\ f_1 = n_1 - 1 = 12 - 1 = 11,\ f_2 = 15 - 1 = 14) = 2.57$;

$F_{набл} < F_{крит}(\alpha, f_1, f_2) => H_0$ - нет оснований отвергать нулевую гипотезу о равенстве генеральных дисперсий. Значит, можно применять и критерии Стьюдента для сравнения средних.

Задачи для самостоятельной работы

1. Измерялась высота растений подсолнечника через 15 дней после появления всходов при двух способах возделывания: 1) при весеннем посеве и 2) при осеннем под зиму. Высота растений первой группы (см): 14,5 ;16 ;15; 14; 15,5, второй группы 21;14; 16,5; 19,5; 19. Оценить существенность разности высоты растений.

2. Масса телят при рождении первой группы (кг): 35;39;41;37;43, масса телят второй группы: 41;45;44;39;46. Проверить гипотезу о равенстве двух средних.

3 На двух группах бычков А и В сравнивали влияние на суточный прирост (кг) двух видов кормов: льняного жмыха и сои.

Льняной жмых

x	1,95	2,05	2,11	2,17	2,24	2,52
n	2	2	1	1	1	1

Соя

y	1,74	1,77	1,83	1,86	1,92	2,5
n	1	2	2	2	1	1

Проверить гипотезу о равенстве двух средних.

4. Даны два распределения случайной величины: X - урожайность свеклы, выращенной с применением микроудобрений, и Y – без применения микроудобрений. С надежностью 0,95 установить, значимо или незначимо расхождение между средними значениями. Сделать вывод о влиянии микроудобрений на урожай с учётом данных в следующих таблицах.

x_i	200.. 210	210…220	220…230	230…240	240…250	250…260
n_i	2	4	7	8	6	3

y_i	190…200	200…210	210…220	220…230	230…240	240…250	250..260	260…270
n_i	1	2	4	8	6	5	3	1

Тема 8. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

Корреляционный анализ — это статистический метод, изучающий связь между явлениями, если одно из них входит в число причин, определяющих другое, или, если имеются общие причины, воздействующие на эти явления.

Основная задача — выявление связи между случайными переменными.

Регрессионный анализ—это статистический метод, изучающий зависимость между результативным признаком Y и входной переменной X.

Основная задача — установление формы связи между переменными и изучение зависимости между ними.

8.1. Функциональная и корреляционная зависимости

Функциональная зависимость — это зависимость вида y = *f* (x), когда каждому возможному значению случайной величины X соответствует одно возможное значение случайной величины *Y*. Например, площадь круга *S* однозначно связана с радиусом окружности *R: S* = πR^2.

Корреляционная зависимость—это статистическая зависимость, проявляющаяся в том, что при изменении одной из величин изменяется среднее значение другой:

$$\bar{y} = f\ (x)$$

Например, рост и масса. При одном и том же росте масса различных индивидуумов может быть различна, но между средними значениями этих показателей имеется определенная зависимость.

Установление взаимосвязи между различными признаками и показателями функционирования организма позволяет по изменениям одних судить о состоянии других.

Для изучения корреляционной связи данные о статистической зависимости удобно задавать в виде корреляционной таблицы или в виде двумерной выборки.

X_i	x_1	x_2	…	x_n
Y_i	y_1	y_2	…	y_n

Схема эксперимента следующая: пусть имеется выборка объема *n* из генеральной совокупности *N*. На каждом объекте выборки определяют числовые значения признаков, между которыми требуется установить наличие или отсутствие связи. Таким образом, получают 2 ряда числовых значений.

Для наглядности полученного материала каждую пару можно представить в виде точки на координатной плоскости. По оси абсцисс откладывают значения одного вариационного ряда — x_i, а по оси ординат другого — y_i.

Такое изображение статистической зависимости называется *полем корреляции,* или *корреляционным полем точек.* Оно создает общую картину корреляций.

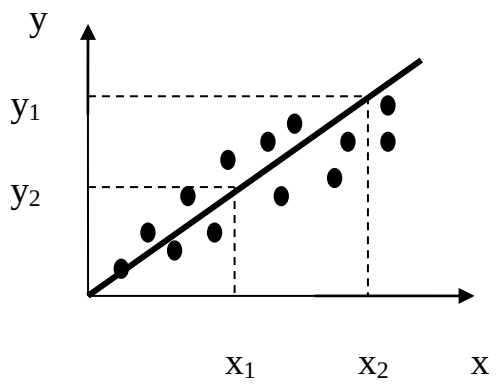


Рис.8.1 Поле корреляции

Если точки группируются вдоль некоторого направления (рис. 8.1), то это говорит о наличии линейной корреляционной связи между признаками.

Если точки распределены равномерно, то линейная корреляционная связь отсутствует.

8.2. Коэффициент линейной корреляции и его свойства

На практике исследователя часто может интересовать не сама зависимость одной переменной от другой, а именно характеристика тесноты связи между ними, которую можно было бы выразить одним числом. Эта характеристика называется *выборочным коэффициентом линейной корреляции* r.

Требования к корреляционному анализу: корреляционный анализ — это метод, используемый, когда данные можно считать случайными и выбранными из совокупностей, распределенных по *нормальному* закону.

Выборочный коэффициент линейной корреляции r характеризует тесноту линейной связи между количественными признаками в выборке:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Если r > 0, то корреляционная связь между переменными прямая, при r < 0 связь обратная.

Свойства коэффициента корреляции r

Они проявляются при достаточно большом объеме выборки n.

1. Коэффициент корреляции принимает значения на отрезке $-1 \le r \le 1$.

В зависимости от того, насколько $|r|$ приближается к 1, различают связи:

► $r < 0{,}3$ — слабая связь;
► $r = 0{,}3 - 0{,}5$— умеренная связь;
►$r = 0{,}5 - 0{,}7$ — заметная (значительная);
►$r = 0{,}7 - 0{,}8$—достаточно тесная;
►$r = 0{,}8 - 0{,}9$ — тесная (сильная);
►$r > 0{,}9$ — очень сильная, то есть чем ближе $|r|$ к 1, тем теснее связь.

2. При $r = 1$ — функциональная зависимость $y = f(x)$.
3. Чем ближе $|r|$ к 0, тем слабее связь.
4. При $r = 0$ линейная корреляционная связь отсутствует.
5. $r_{xy} = r_{yx}$ — случайные переменные симметричные;

x и y могут взаимозаменяться, не влияя на величину r.

♦ **Задача 8.1.**

Построить корреляционное поле точек и вычислить коэффициент корреляции между ростом (X) и массой (У) некоторых животных. Исходные данные приведены в выборке объема $n = 10$.

x_i	31	32	33	34	35	35	40	41	42	46
y_i	7,8	8,3	7,6	9,1	9,6	9,8	11,8	12,1	14,7	13,0

Решение

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Средний рост $\bar{x}$:

$\bar{x} = \frac{\sum x}{n}$, $\bar{x} = \frac{31 + 32 + ... + 46}{10} = \frac{369}{10} = 36{,}9.$

Средняя масса $\bar{y}$:

$\bar{y} = \frac{\sum y}{n}$, $\bar{y} = \frac{7{,}8 + 8{,}3 + ... + 13{,}0}{10} = 10{,}38.$

Находим:

$\sum (x - \bar{x}) \cdot (y - \bar{y}) = (31 - 36{,}9)(7{,}8 - 10{,}38) + \ldots + (46 - 36{,}9)(13 - 10{,}38) = 99{,}9;$

$\sum (x - \bar{x})^2 = (31 - 36{,}9)^2 + (32 - 36{,}9)^2 + \ldots + (46 - 36{,}9)^2 = 224{,}8;$

$\sum (y - \bar{y})^2 = (7{,}8 - 10{,}38)^2 + (8{,}3 - 10{,}38)^2 + \ldots + (13{,}0 - 10{,}38)^2 = 51{,}9.$

Подставим полученные значения в формулу для r:

$$r = \frac{99{,}9}{\sqrt{224{,}8 \cdot 51{,}9}} = 0{,}925$$

Величина r близка к 1, это говорит о тесной связи роста и массы.

8.3. Проверка гипотезы о значимости выборочного коэффициента линейной корреляции

Это ответ на вопрос: существует ли вообще эта связь. Эмпирический коэффициент корреляции, как и любой другой выборочный показатель, служит оценкой своего генерального параметра. Выборочный коэффициент линейной корреляции $r_в$—величина *случайная,* так как он вычисляется по значениям переменных, случайно попавшим в выборку из генеральной совокупности, а значит, как и любая случайная величина, имеет ошибку m_r.

Чтобы выяснить, находятся ли случайные величины *X* и *Y* генеральной совокупности в линейной корреляционной зависимости, надо проверить значимость $r_в$. Для этого проверяют нулевую гипотезу о равенстве нулю коэффициента корреляции генеральной совокупности $H_о$: $r_{ген} = 0$, то есть линейная корреляционная связь между признаками X и Y случайна. Выдвигается альтернативная гипотеза H_1: $r_{ген} \neq 0$, то есть эта линейная корреляционная связь имеется. Задается уровень значимости, например, $\alpha \leq 0{,}05$.

Критерием для проверки нулевой гипотезы является отношение выборочного коэффициента корреляции к своей ошибке:

$$t_{набл} = \frac{r}{m_r},$$

где m_r — ошибка коэффициента корреляции.

Если объем выборки $n < 100$, то $m_r = \sqrt{\frac{1-r^2}{n-2}}$;

если объем выборки $n > 100$, то $m_r = \frac{1-r^2}{\sqrt{n}}$.

Число степеней свободы для проверки критерия равно $f = n - 2$. Гипотезу проверяют по таблицам распределения Стьюдента в соответствии с выбранным уровнем значимости.

По таблице критических точек распределения Стьюдента находим $t_{крит}(\alpha, f)$, определенное на уровне значимости a < 0,05 при числе степеней свободы $f = n - 2$, где n — объем двумерной выборки.

Если $t_{набл} > t_{крит} = \Rightarrow H_1$, отвергают нулевую гипотезу и принимают альтернативную: $r_{ген} \neq 0$, имеется линейная корреляционная связь между признаками.

Если $t_{набл} < t_{крит}$, то нет оснований отвергать нулевую гипотезу, а r_e статистически незначим. Эта связь случайна.

♦ Задача 8.2.

Проверить значимость коэффициента корреляции $r = 0{,}74$ между переменными *X* и Y для выборки объема $n = 50$.

Решение

Проверяется нулевая гипотеза H_0 об отсутствии линейной корреляционной связи между переменными *X* и *Y* в генеральной совокупности H_0; $r_{ген} = 0$.

При справедливости этой гипотезы $t_{набл} = \frac{r}{m_r}$, где ошибка коэффициента корреляции $m_r = \sqrt{\frac{1-r^2}{n-2}}$ и $t_{набл} = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$ имеют распределение Стьюдента с $f = n - 2$ степенями свободы.

Рассчитаем: $t_{набл} = \frac{0{,}74 \cdot \sqrt{50-2}}{\sqrt{1-0{,}74^2}} = 7{,}62$.

По таблицам находим табличное значение *t*-критерия Стьюдента, определенное на уровне значимости $\alpha \le 0{,}05$ и при числе степеней свободы $f = 50 - 2 = 48$, $t_{крит}(\alpha \le 0{,}05; 48) = 2{,}02$.

Поскольку $t_{набл} > t_{крит}$, $7{,}62 > 2{,}02$, коэффициент корреляции значимо отличается от нуля.

Причем это справедливо и для уровня значимости $\alpha \le 0{,}001 (t = 3{,}55)$.

8.4. Выборочное уравнение линейной регрессии. Метод наименьших квадратов

Задача регрессионного анализа состоит в подборе упрощенной аппроксимации связи с помощью математической модели.

Регрессионный анализ имеет в своем распоряжении специальные процедуры проверки, является ли выбранная математическая модель *адекватной* для описания имеющихся данных.

Чаще всего регрессионный анализ используется для *прогноза,* то есть предсказания значений ряда зависимых переменных по известным значениям других переменных.

Выше указывалось, что результаты наблюдений, приведенные в двумерной выборке

x_i	x_1	x_2	x_3	x_4	x_5
y_i	y_1	y_2	y_3	y_4	y_5

можно представить в виде корреляционного поля точек (рис. 8.2), где каждая точка соответствует отдельным **значениям** *x* и *y.*

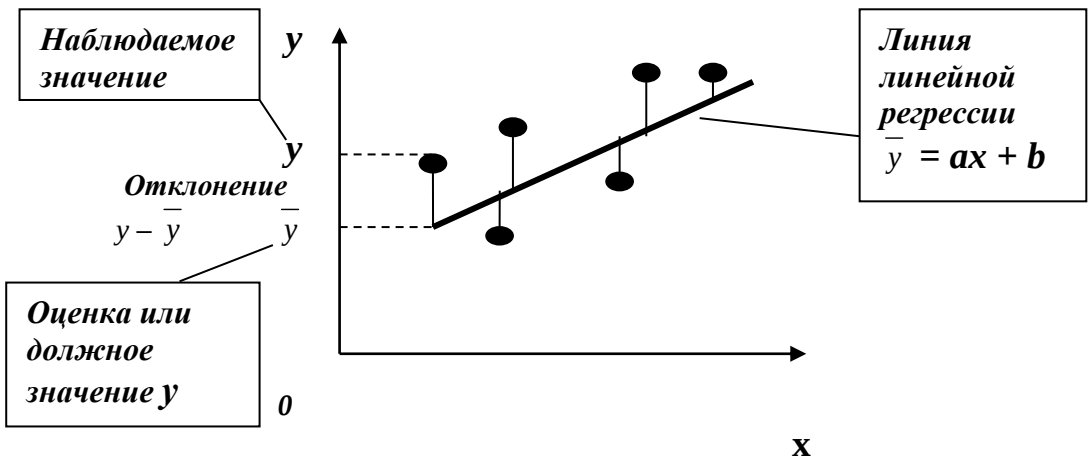


Рис. 8.2. Метод наименьших квадратов

В результате получается диаграмма рассеяния, позволяющая судить о форме и тесноте связи между варьирующими признаками. Довольно часто эта связь может быть аппроксимирована прямой линией (рис. 8.2).

Регрессия — это функция, позволяющая по величине одного признака *X* находить среднее ожидаемое (должное) значение другого признака *Y*, корреляционно связанного с *X*.

В линейной математической модели уравнение линейной регрессии имеет вид

$$\bar{y} = ax + b,$$

где *a* и *b* — параметры линейной регрессии;

a — это коэффициент регрессии, показывающий, насколько в среднем величина одного признака *Y* изменяется при изменении на единицу меры другого признака *X*, корреляционно связанного с *Y*. Чем больше *a* — угловой коэффициент прямой $a = tg\alpha$, тем круче прямая, то есть быстрее изменяется ***Y***.

b — свободный член в уравнении, определяет $\bar{y}$ при $x = 0$.

$\bar{y}$ — это предсказанное (должное) значение ***Y*** для данного *x* при определенных значениях регрессионных параметров.

Параметры линейной регрессии определяют методом наименьших квадратов — это способ подбора параметров регрессионной модели, согласно которому сумма квадратов отклонений вариант от линии регрессии должна быть минимальна:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 \Rightarrow \min$$

Это эффективный метод, позволяющий уменьшить влияние ошибок измерений.

Теперь определяют должные величины $\bar{y}_{должн}$, наносят эти точки и соединяют их прямой линией.

Достоинство корреляционно – регрессионного анализа— наглядное представление о форме и тесноте связи. Регрессия выражает корреляционную зависимость в виде *функционального отношения* и дает более полную информацию.

Задача 8.3

Были произведены измерения общей длины (X) ствола в см и длины его части без ветвей (Y) 10 молодых сосен. Результаты этого измерения представлены в следующей таблице:

X	25	35	45	55	65	75	85	95	105	115
Y	14	18	19	20	23	23	24	26	29	34

Вычислить выборочный коэффициент корреляции и найти выборочное уравнение прямой регрессии Y на X.

Решение

Вычислим выборочный коэффициент корреляции по формуле

$$r_B = \frac{\sum (x_i - \overline{x_b})(y_i - \overline{y_b})}{\sqrt{\sum (x_i - \overline{x_b})^2}\sqrt{\sum (y_i - \overline{y_b})^2}}$$

Для вычисления величин, входящих в формулу, составим вспомогательную таблицу, в которой результаты измерений записаны столбцами. Внизу каждого из этих столбцов вычислены суммы для нахождения средних x_B и y_B. Далее расположены столбцы, в которых вычисляются разности $x_i - x_B$ и $y_i - y_B$, их квадраты и произведения. Значения этих столбцов суммируются, чтобы получить величины, необходимые для подстановки в формулу. Отметим, что суммы в столбцах, в которых вычислены разности $x_i - x_B$ и $y_i - y_B$,будут всегда равны 0.

N	x_i	y_i	$x_i - x_B$	$(x_i - x_B)^2$	$y_i - y_B$	$(y_i - y_B)^2$	$(x_i - x_B)(y_i - y_B)$
1	25	14	-45	2025	-9	81	405
2	35	18	-35	1225	-5	25	175
3	45	19	-25	625	-4	16	100
4	55	20	-15	225	-3	9	045
5	65	23	-5	25	0	0	0
6	75	23	5	25	0	0	0
7	85	24	15	225	1	1	15
8	95	26	25	625	3	9	75
9	105	29	35	1225	6	36	210
10	115	34	45	2025	11	121	495
Σ	700	230	0	8250	0	298	1520

Находим средние x_B и y_B:

$$x_B = \frac{700}{10} = 70, y_B = \frac{230}{10} = 23$$

Из таблицы имеем

$\sum(x_i - \overline{x_B})(y_i - \overline{y_B}) = 1520$, $\sum (x_i - \overline{x_B})^2 = 8250$, $\sum (y_i - \overline{y_B})^2 = 298$

Подставляя эти значения в формулу для вычисления коэффициента корреляции, получим

$$r_B = \frac{1520}{\sqrt{8250} \cdot \sqrt{298}} \approx 0{,}97$$

Таким образом, у выбранных сосен имеет место очень сильная прямая корреляция между общей длиной ствола и длиной его части без ветвей.

Найдём теперь выборочное уравнение прямой регрессии Y на X. Это уравнение имеет вид:

$y - y_в \equiv b \cdot (x - x_B)$, где b - коэффициент регрессии, отражающий интенсивность изменения результативного признака y при изменении x.

Коэффициент регрессии b является величиной с конкретной размерностью и измеряется в тех же единицах признака y.

$$b = \frac{\sum(x_i - x_в)(y_i - y_в)}{\sum(x_i - x_в)^2} = 1520/8250 = 0{,}18$$

Подставляя в выборочное уравнение прямой регрессии Y на X

$\overline{x}_B = 70$, $\overline{y}_B = 23$, $r_B = 0{,}97$,

получим $y - 23 = 0{,}18\,(x - 70)$ или $y - 23 = 0{,}18x - 12{,}6$.

Окончательно, $y = 0{,}18x + 10{,}4$ - искомое уравнение прямой регрессии Y на X (рис. 8.3).

Таким образом, коэффициент регрессии b = 0,18 показывает, что при изменении длины ствола на 1 см длина ствола без ветвей изменится в среднем на 0,18 см.

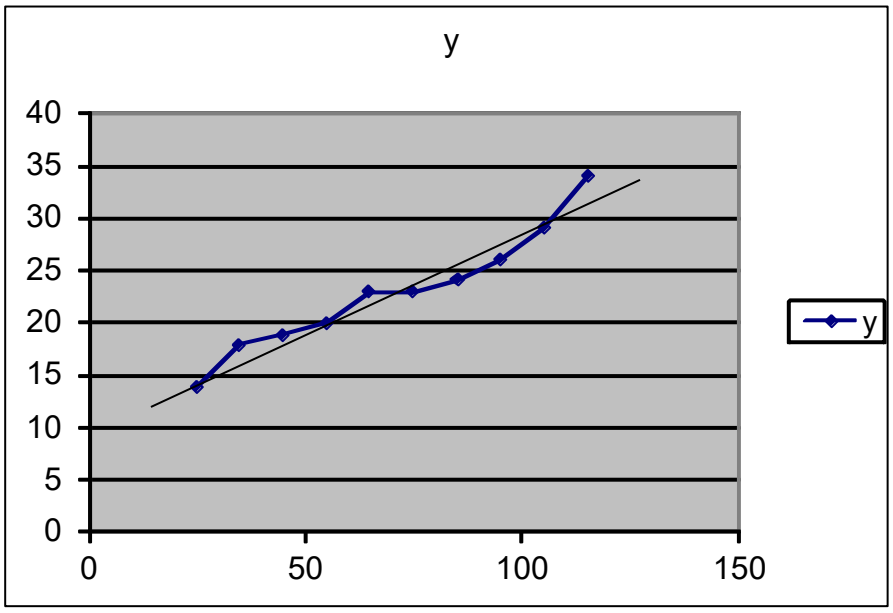


Рис. 8.3 *Эмпирическая линия распределения и* ***линия регрессии***

8.5. Нелинейная регрессия

Если график регрессии $\bar{y} = f$ (x) изображается кривой линией, то это нелинейная регрессия.

Выбор вида уравнения регрессии производится на основании опыта предыдущих исследований, литературных источников, профессионального мнения и визуального наблюдения расположения точек корреляционного поля. Этот очень важный этап анализа называется *спецификацией*.

Наиболее часто встречаются следующие виды уравнений нелинейной регрессии:

$\bar{y} = a_0 + a_1 \cdot x + ... + a_n \cdot x^n$ — полиномиальное уравнение;

$\bar{y} = ax^2 + bx + c$ — уравнение параболы второго порядка;

$\bar{y} = ax^3 + bx^2 + cx + d$ — уравнение параболы третьего порядка;

$\bar{y} = \frac{a}{x} + b$ — гиперболическое уравнение.

Для определения неизвестных параметров регрессии используется метод наименьших квадратов.

Задача для самостоятельной работы

По 10 предприятиям имеются следующие данные:

Выпуск продукции, тыс. руб. в неделю	6,3+ k/10	6,0+ k/10	7,5+ k/10	8,5+ k/10	3,5+ k/10	6,2+ k/10	7,5+ k/10	8,7+ k/10
Потреб-ность в сельхоз-технике, шт.	5	4	6	7	3	4	6	7

Найдите уравнение корреляционной связи (уравнение регрессии) между выпуском продукции и потребностью в сельхозтехнике. Рассчитайте коэффициент корреляции.

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2} dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,52	0,1985	1,04	0,3508	1,56	0,4406	2,16	0,4846
0,01	0,0040	0,53	0,2019	1,05	0,3531	1,57	0,4418	2,18	0,4854
0,02	0,0080	0,54	0,2054	1,06	0,3554	1,58	0,4429	2,20	0,4861
0,03	0,0120	0,55	0,2088	1,07	0,3577	1,59	0,4441	2,22	0,4868
0,04	0,0160	0,56	0,2123	1,08	0,3599	1,60	0,4452	2,24	0,4875
0,05	0,0199	0,57	0,2157	1,09	0,3621	1,61	0,4463	2,26	0,4881
0,06	0,0239	0,58	0,2190	1,10	0,3643	1,62	0,4474	2,28	0,4887
0,07	0,0279	0,59	0,2224	1,11	0,3665	1,63	0,4484	2,30	0,4893
0,08	0,0319	0,60	0,2257	1,12	0,3686	1,64	0,4495	2,32	0,4898
0,09	0,0359	0,61	0,2291	1,13	0,3708	1,65	0,4505	2,34	0,4904
0,10	0,0398	0,62	0,2324	1,14	0,3729	1,66	0,4515	2,36	0,4909
0,11	0,0438	0,63	0,2357	1,15	0,3749	1,67	0,4525	2,38	0,4913
0,12	0,0478	0,64	0,2389	1,16	0,3770	1,68	0,4535	2,40	0,4918
0,13	0,0517	0,65	0,2422	1,17	0,3790	1,69	0,4545	2,42	0,4922
0,14	0,0557	0,66	0,2454	1,18	0,3810	1,70	0,4554	2,44	0,4927
0,15	0,0596	0,67	0,2486	1,19	0,3830	1,71	0,4564	2,46	0,4931
0,16	0,0636	0,68	0,2517	1,20	0,3849	1,72	0,4573	2,48	0,4934
0,17	0,0675	0,69	0,2549	1,21	0,3869	1,73	0,4582	2,50	0,4938
0,18	0,0714	0,70	0,2580	1,22	0,3888	1,74	0,4591	2,52	0,4941
0,19	0,0753	0,71	0,2611	1,23	0,3907	1,75	0,4599	2,54	0,4945
0,20	0,0793	0,72	0,2642	1,24	0,3925	1,76	0,4608	2,56	0,4948
0,21	0,0832	0,73	0,2673	1,25	0,3944	1,77	0,4616	2,58	0,4951
0,22	0,0871	0,74	0,2703	1,26	0,3962	1,78	0,4626	2,60	0,4953
0,23	0,0910	0,75	0,2734	1,27	0,3980	1,79	0,4633	2,62	0,4956
0,24	0,0948	0,76	0,2764	1,28	0,3997	1,80	0,4641	2,64	0,4959
0,25	0,0987	0,77	0,2794	1,29	0,4015	1,81	0,4649	2,66	0,4961
0,26	0,1026	0,78	0,2823	1,30	0,4032	1,82	0,4656	2,68	0,4963
0,27	0,1064	0,79	0,2852	1,31	0,4049	1,83	0,4664	2,70	0,4965
0,28	0,1103	0,80	0,2881	1,32	0,4066	1,84	0,4671	2,72	0,4967
0,29	0,1141	0,81	0,2910	1,33	0,4082	1,85	0,4678	2,74	0,4969
0,30	0,1179	0,82	0,2939	1,34	0,4099	1,86	0,4686	2,76	0,4971
0,31	0,1217	0,83	0,2967	1,35	0,4115	1,87	0,4693	2,78	0,4973
0,32	0,1255	0,84	0,2995	1,36	0,4131	1,88	0,4699	2,80	0,4974
0,33	0,1293	0,85	0,3023	1,37	0,4147	1,89	0,4706	2,82	0,4976
0,34	0,1331	0,86	0,3051	1,38	0,4162	1,90	0,4713	2,84	0,4977
0,35	0,1368	0,87	0,3078	1,39	0,4177	1,91	0,4719	2,86	0,4979
0,36	0,1406	0,88	0,3106	1,40	0,4192	1,92	0,4726	2,88	0,4980
0,37	0,1443	0,89	0,3133	1,41	0,4207	1,93	0,4732	2,90	0,4981
0,38	0,1480	0,90	0,3159	1,42	0,4222	1,94	0,4738	2,92	0,4982
0,39	0,1517	0,91	0,3186	1,43	0,4236	1,95	0,4744	2,94	0,4984
0,40	0,1554	0,92	0,3212	1,44	0,4251	1,96	0,4750	2,96	0,4985
0,41	0,1591	0,93	0,3238	1,45	0,4265	1,97	0,4756	2,98	0,4986
0,42	0,1628	0,94	0,3264	1,46	0,4279	1,98	0,4761	3,00	0,49865
0,43	0,1664	0,95	0,3289	1,47	0,4292	1,99	0,4767	3,20	0,49931
0,44	0,1700	0,96	0,3315	1,48	0,4306	2,00	0,4772	3,40	0,49966
0,45	0,1736	0,97	0,3340	1,49	0,4319	2,02	0,4783	3,60	0,499841
0,46	0,1772	0,98	0,3365	1,50	0,4332	2,04	0,4793	3,80	0,499928
0,47	0,1808	0,99	0,3389	1,51	0,4345	2,06	0,4803	4,00	0,499968
0,48	0,1844	1,00	0,3413	1,52	0,4357	2,08	0,4812	4,50	0,499997
0,49	0,1879	1,01	0,3438	1,53	0,4370	2,10	0,4821	5,00	0,499997
0,50	0,1915	1,02	0,3461	1,54	0,4382	2,12	0,4830		
0,51	0,1950	1,03	0,3485	1,55	0,4394	2,14	0,4838		

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3985	3961	3956	3951	3945	3939	3932	2925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	1878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0335	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0003	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Библиографический список

1. Белько И.В. Теория вероятностей, математическая статистика, математическое программирование: Учебное пособие/ Белько И.В., Морозова И.М., Криштапович Е.А. – М.:НИЦ ИНФРА-М, Нов. знание, 2016. -229 с.
2.Бирюкова Л.Г. Теория вероятностей и математическая статистика: Учебное пособие/ Бирюкова Л.Г., Матвеев В.И., Бобрик Г.И., - 2-е изд. – М.: НИЦ ИНФРА-М, 2017. – 289 с.

3. Гмурман В. К. Теория вероятностей и математическая статистика. – М.: Высшая школа,1977.
4. Гмурман В. К. Руководство и решение задач по теории вероятностей и математической статистике. – М.: Высшая школа,1977.